NEW APPROACHES TO POPULATION GENETICS IN THE LIGHT OF LOAD AT MANY LOCI

By

Joseph Matheson

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF ECOLOGY AND EVOLUTIONARY BIOLOGY

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2022

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by: Joseph Matheson, titled: New approaches to population genetics in the light of load at many loci
and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Joanna Masel — Date: Mar 4, 2022
Joanna Masel

David Enard — Date: Mar 4, 2022
David Enard

Aurelio Figueredo — Date: Mar 4, 2022
Aurelio J. Figueredo

J. Bruce Walsh — Date: Mar 4, 2022
J. Bruce Walsh

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Joanna Masel — Date: Mar 4, 2022
Joanna Masel
Dissertation Committee Chair
Department of Ecology & Evolutionary Biology

Acknowledgements

I'd like to thank my wife Kara, who has no deleterious mutations. I'd like to thank my advisor, Joanna Masel, for years of mentorship and discussion. Her talent for identifying important skills and knowledge and developing a structured plan to achieve them has helped me become a better scientist, and person, than I thought was possible when I first started. I'd like to thank my committee members, Bruce Walsh, David Enard, and AJ Figueredo, for providing a massive pool of knowledge on a wide array of subjects that I have shamelessly drawn from in the research that follows. I'd also like to thank some other scientists who have contributed time and ideas to improve the research in this dissertation, including Ryan Gutenkunst, Moises Exposito-Alonso, Jason Bertram, and Kevin Thornton. I'd like to thank the faculty and staff of the Ecology and Evolutionary Biology department for creating a supportive environment to work. Lastly, I simply must thank the wonderful friends I've made among the fellow graduate students here in EEB, who have commiserated with me at Friday happy hours, listened to my practice talks, played Dungeons and Dragons with me, come to my wedding, and altogether made these past few years some of the best.

Table of Contents

Abstract

Hermann Muller first used the term ‘genetic load’ in 1950 to refer to the burden imposed on populations by deleterious genetic variants, but fears about heritable deleterious traits proliferating have haunted the newborn field of population genetics from the beginning. Classical population genetic theory seemed to suggest that load could build up to dangerous levels in a variety of situations: rapid adaptation, high mutation rates, or sufficiently weak natural selection. Different formulations of load problems have been proposed and then apparently solved, but the solutions have not all stood the test of time. In this dissertation, I will highlight some places where established population genetic theory fails to sufficiently capture the effects of load, especially mildly deleterious load dispersed across many linked loci.

I first investigate Haldane’s Dilemma, an implausibly strict speed limit on adaptation proposed by Haldane based on his analysis of the load that builds up during the process of substituting one allele for another. This dilemma was apparently solved by noting that Haldane’s calculations used the wrong type of load, but substantial confusion in the literature on the topic has left a more accurate formulation of the dilemma still unsolved. I apply this formulation to an *Arabidopsis* dataset tracking seed production, seed survival, and juvenile plant survival across eight different environments. In this dataset, the dilemma poses no real threat because a much higher fraction of deaths contribute to adaptation than had been anticipated by previous estimates.

I then investigate the effects of removing deleterious load on neutral variation, a phenomenon known as background selection. The prevailing analytical approach assumes no linkage between deleterious mutations as a simplification. There are theoretical reasons to expect that this assumption will be wrong, and given the sheer quantity of deleterious mutations entering populations, violations of this assumption are likely not trivial. I find that when simulating a genome with realistic degrees of mutation and recombination, background selection has much stronger effects on neutral diversity than the prevailing theory predicts.

I last investigate accumulation of small-effect deleterious load, a load problem which Kondrashov (1995) suggests should have killed us 'one hundred times over'. One standard answer to this problem — synergistic epistasis between deleterious mutations — has been contradicted by empirical evidence. The second answer — beneficial mutations counteracting deleterious load — is promising, but previous theoretical descriptions again ignored the effects of linkage. I built a novel simulation model to investigate accumulating load in the presence of rare beneficial mutations. I find that rare large-effect beneficial mutations can asymmetrically counteract deleterious load. However, whether there are enough beneficial mutations left over for populations to adapt depends on the value of beneficial parameters for which we still lack good empirical estimates.

I show in a variety of cases that new approaches which account for the effects of deleterious load across many linked loci are able to contribute new answers to foundational questions in population genetics.

Chapter 1: Introduction

1.1. Load

The term ‘genetic load’ refers to a reduction in fitness in a population due to deleterious genetic variants. This term was popularized by Hermann Muller (1950), although historian Amir Teicher (2018) argues that the origins of the term date to late 19th-century German psychiatry, which used the term to describe the possible heritability of mental conditions. This earlier usage highlights a feature of genetic load which has haunted it ever since: the fear that load might increase to dangerous levels and lead to extinction.

To translate this fear into population genetics, load needs to be compared to some benchmark fitness. The most intuitive definition of load is the reduction in average fitness of a real population (with many segregating deleterious variants) compared to the fitness of a hypothetical population without any deleterious variants. This definition also applies to cases where a population is adapting via new beneficial mutations, with the hypothetical fully-adapted population being the benchmark (Haldane 1957). The speed at which populations increase in fitness is also related to load, although in this case the appropriate benchmark is the fittest individual in the population, not the fitness of a hypothetical perfect population (Ewens 1970; Desai and Fisher 2007).

The interpretation of any load quantity relies strongly on the benchmark used, and so different types of load have been described, e.g. substitutional load (Kimura 1960), lag load (Maynard Smith 1976), mutation load (Muller 1950), segregation load (Kimura 1960), etc. Unfortunately, usage of these terms has not been consistent in the literature, which has been a persistent source of confusion in each of the load questions I explore. In each case that follows, I will clarify the relevant type of load and try to disentangle any prior misunderstandings on the topic.

1.2. Classic load paradoxes

The first mathematical descriptions of load returned results that could be interpreted as seriously concerning for the fate of populations. Every deleterious mutation which arises persists in the population contributing load for some time before being eliminated. Mutations with larger effects are purged faster, producing the surprising result that the amount of load contributed by deleterious mutations doesn't depend on their effect size, only the rate at which they arise (Haldane 1937). Multiplying together the effects of all deleterious mutations suggest that fitness should be reduced by a factor of $1 - e^{-U_d}$ where U_d is the total genome-wide deleterious mutation rate (Kimura *et al.* 1963). Estimates of U_d in humans put it around 2 (Lesecque *et al.* 2012), and estimates for *Drosophila* are similar (Haag-Liautard *et al.* 2007). Plugging in those numbers suggests that humans have a load of 0.87, or that our fitness is only 13% of what it could be without deleterious mutations!

Other ways of describing load produced similarly concerning results. Haldane (1957) proposed interpreting load in terms of 'selective deaths', or the number of individuals who needed to die without passing on their genes in order to effect the elimination of a deleterious allele or the fixation of a beneficial one. Again, he found the surprising result that the number of deaths did not depend on the selective effects of the alleles, only their starting frequencies. Since the elimination of a new deleterious mutation requires a minimum of one selective death, this seems to suggest that populations should be unable to eliminate deleterious mutations as fast as they appear if more than one deleterious mutation arises per newborn individual. Unfortunately, newborn humans have an average of two new deleterious mutations!

Haldane (1957) also calculated the amount of selective deaths accumulated over the course of beneficial substitutions, which is typically several times the population size. By estimating how many selective deaths a typical population could safely absorb, he was able to estimate how fast new beneficial mutations could complete their substitutions. Unfortunately, his estimate was that populations could only safely

substitute one new beneficial allele every 300 generations, much slower than the observed rate of substitutions! This disconnect between Haldane's estimated speed limit and the actual rate of substitutions was the original evidence used by Kimura to propose that most substitutions must be neutral (Kimura 1968).

More modern formulations of load have been no more optimistic. Deleterious mutations with sufficiently small effect sizes will have their frequency determined primarily by the effects of drift, rather than selection, which results in these small deleterious mutations contributing more load than others — a phenomenon that geneticists have been aware of for a long time (Kimura *et al.* 1963). However, the full magnitude of this problem wasn't articulated until the size of the human genome was known. Kondrashov (1995) modeled each site in the genome as having effectively four alleles, one for each possible nucleotide, with one allele beneficial and the other three deleterious. With around 10^9 sites in the genome and given that mutations with effect sizes around 10^{-5} are typically small enough to avoid selection, then if even 10% of sites have alleles with these small effect sizes, the resulting load will be on the order of 100 lethal equivalents. Other studies which similarly model beneficial and deleterious alleles as having symmetric effects, but allow for larger beneficial variants, still find that populations like humans shouldn't be able to survive (Goyal *et al.* 2012).

These articulations seem bad, but they may in fact be underestimating the potential damage of small deleterious mutations. Kondrashov (1995) assumed nucleotide changes as the only mutations which could occur, but other types of mutations exist, and might not be limited to the number of sites in the genome (e.g. addition of useless DNA). Since small deleterious mutations have appreciable chances of fixing, this implies a constant accumulation of load from fixations of small deleterious mutations.

1.3. Classic load solutions

So why aren't we "dead one hundred times over", in Kondrashov's words (1995), or at least more worried about our genetic load? Each of the load problems in the preceding section are generally agreed to be solved by modern geneticists, but often for different reasons.

One way of responding to load problems is by pointing out that load is often quantified by comparison to a hypothetical mutation-free individual. Since such an individual has almost certainly never existed in any species, how concerned we should be about a given quantity of load depends on assumptions about what such an individual might look like. If even a perfect human could only produce 10-12 offspring on average, then reducing our fitness to 13% of that is a serious concern, but if it's plausible that a completely mutation-free human could average 100 offspring, then this load is much less concerning!

Another way of responding is by focusing on what type of fitness is involved in load. Haldane seemed to consistently assume that load was absolute, and based his estimates of how much load species could maintain on how many offspring the species could afford to lose (Haldane 1937, 1957). However, if load manifests primarily in traits affecting relative fitness (e.g. mating success, intraspecific competition for resources), then high load may be devastating for individuals, but not populations (Agrawal and Whitlock 2012). A similar solution arises if fitness is determined by the value of some higher-order trait relative to an optimum, and load represents the distance of this trait from the optimum. In this case, new mutations in a population far from the optimum have a much higher chance of being beneficial than they would in a population very close to the optimum, and this keeps populations fairly close to the optimum, implying low load (Charlesworth 2013).

One of the most common responses to load problems has been to note that these load formulations assume that the load contributed by deleterious mutations is independent of the genetic background they land on. But epistatic interactions between deleterious mutations might be common and represent a possible solution to mutation load concerns. Specifically, synergistic epistasis between deleterious

mutations (perhaps representing a situation where fit individuals can tolerate a few deleterious mutations quite well, but unfit individuals find the next deleterious mutation suffered to be a serious burden) will result in high-load individuals being purged by selection more effectively. Synergistic epistasis solves the problem of small deleterious mutations (since they can be effectively cleared), as well as selective deaths problems (since selective deaths can reliably clear multiple deleterious mutations at once).

A last solution, which has received much less attention than the preceding ones, suggests that load can be partially absorbed with reproductive excess. This solution was originally proposed in response to Haldane's arguments about selective deaths (Felsenstein 1971; Nei 1971), but it can also help offset load formulated in other ways. The reason this solution is underexplored is likely due to the fact that the most common models in population genetics, Wright-Fisher models, assume infinite reproductive excess (Bertram and Masel 2019).

1.4. Issues with classic solutions

Unfortunately, these classic solutions to load problems have serious flaws. This is especially concerning since these flaws have not yet produced a substantial revisiting of load arguments.

One crucial flaw in arguments which suggest that load is primarily in relative or stabilizing fitness is that a change to a DNA sequence should be far more likely to simply reduce functionality than to slightly modify a higher-level trait or exclusively affect mating success. Looked at from a molecular perspective, unconditionally deleterious mutations likely represent the vast majority of load. The fact that load in other types of fitness is less dangerous doesn't address the threat posed by this more common load.

Another flaw with these arguments and ones which argue that high load relative to a perfect individual is no serious threat are that they assume stable load. Some of the load problems (selective deaths and small deleterious mutations) involve a failure to eliminate load as fast as it arises, implying continuous degradation of fitness. Differing assumptions about the fitness of a mutationless individual or the fraction

of mutations affecting traits under relative or stabilizing selection only serve to slow the expected accumulation of load. Constantly accumulating load should still eventually hit a level that populations cannot survive.

Synergistic epistasis between deleterious mutations should successfully solve all of the preceding load problems and does not have either of the flaws above. Unfortunately, empirical studies of epistasis in *de novo* mutations have not observed net synergistic epistasis (Kouyos *et al.* 2007). Epistasis is indeed quite common, but antagonistic epistasis (which would make load problems worse!) is just as common as synergistic epistasis, and the net results on fitness average out very close to multiplicative (Kouyos *et al.* 2007).

1.5. Explanation of dissertation format

The next chapter is a summary of two submitted first-author preprints and a draft of a third first-author manuscript not yet submitted. Each paper investigates aspects of different load problems. The first is focused on arguments based on selective deaths, clarifying several lines of reasoning which have been confused in the literature, and investigating unaddressed arguments related to reproductive excess using a dataset of seed production, seed survival, and juvenile survival in *Arabidopsis*. The second points out a disagreement between theoretical approximations and simulations of the effect of clearing deleterious mutations on linked neutral variation. The last draft focuses on a novel solution to the problem of accumulating deleterious load: compensating for the unavoidable accumulation with rare, large-effect beneficial mutations.

My role in each of the papers will be explained at the end of each summary section.

Chapter 2: Present Study

2.1. Haldane's cost of selection imposes a mild constraint on adaptation

2.1.1. Haldane's Dilemma

Haldane's Dilemma is the name for a speed limit on adaptation based on comparing the number of selective deaths required per adaptive substitution to estimates of the amount of selective deaths a population can sustain, as mentioned in Chapter 1 (Van Valen 1963). Reasoning about this speed limit has been confused by the fact that the term 'substitution load' has been used to refer to arguments based on selective deaths, load in populations undergoing adaptation, and reproductive excess required to sustain adaptation. Haldane's original argument improperly applied load relative to a perfectly adapted population to determine the amount of selective deaths incurred by a population, when in this case the relevant load is relative to the most adapted individual in the population (Ewens 1970, 2004). Since appropriate load comparisons resulted in no significant speed limit, the consensus which emerged was that speed limits are not relevant. This consensus missed the fact that an alternative way of expressing speed limits based on reproductive excess available to use for selective deaths remained unaddressed (Felsenstein 1971; Nei 1971).

2.1.2. *Arabidopsis* dataset

We apply reasoning based on reproductive excess to a dataset which tracked thousands of *Arabidopsis* plants and seeds from 517 different genotypes across 8 different environmental treatments (Exposito-alonso *et al.* 2019). Seed and seedling survival, juvenile survival, and seed production as adults was tracked for every plant in the experiment. The experiment used a 2 x 2 x 2 design for environmental treatments, with plants exposed to all combinations of high or low water, high or low density, and normal or high temperatures. This dataset uniquely allows for the precise calculation of selective deaths incurred

in the process of adapting to new environments, by allowing the identification of the fittest genotype in each environment and comparison of the number of deaths sustained by every genotype. We calculate numbers of selective deaths, the fraction of all deaths which are selective, reproductive excess, and speed limits implied by those reproductive excesses.

2.1.3. Conclusions

We find that many more deaths are selective in *Arabidopsis* than was assumed in previous discussions of speed limits. Haldane assumed that around 10% of deaths could be selective, and while later authors disagreed with him in a variety of ways, the 10% figure was largely accepted (Haldane 1957; Kimura 1968; Felsenstein 1971; Nei 1971; Maynard Smith 1976). In this experiment, however, only the most idyllic environmental conditions stay near 10% selective deaths, with all other environmental conditions seeing much higher proportions of selective deaths, up to 95% in the harshest environmental condition.

This high proportion of selective deaths poses no problems for *Arabidopsis*, due to unsurprisingly high reproductive excess in this weedy annual plant species. Even in the harshest environmental condition, the speed limit to adaptation implied by reproductive excess is biologically insignificant. However, we do find that raw reproductive excess decreases a hundredfold between the environmental conditions, reaching its lowest in two of the harshest conditions. So although populations may be able to sustain much higher proportions of selective deaths than Haldane expected, species with less fecundity than *Arabidopsis* may exhaust their reproductive excess and fail to adapt rapidly enough in the harsh conditions where adaptation is most necessary.

2.1.4. Contributions to the work presented in Appendix A

The inspiration for the work presented here and in Appendix A was discussion between Joanna Masel and Moises Exposito-Alonso. Moises Exposito-Alonso provided the *Arabidopsis* dataset. All data analysis

was performed by me and the first draft of the paper was written by me. Substantial editing and choice of presentation of data was done in discussion between me and Joanna Masel.

2.2. Background selection theory overestimates effective population size for high mutation rates

2.2.1. Background selection models of effective population size

The neutral theory of molecular evolution has produced an elegant mathematical framework of simple models capable of predicting patterns of genetic variation in a wide array of situations (Kimura 1968; Kimura and Ohta 1971; Charlesworth 2009). A crucial parameter in these simple models is the population size, and a key factor which makes these models so flexible is the ability to absorb a wide array of additional effects simply by modifying the population size parameter into an 'effective' population size. This value is the population size required for an idealized population to have the same value for some statistic of interest as the real population, and describing populations with an effective population size allows for the use of simple neutral models in populations with many complex processes affecting genetic diversity (Charlesworth 2009; Masel 2011; Jensen *et al.* 2019).

One such complex process is the removal of neutral variation linked to deleterious mutations — a process known as background selection. Standard models of background selection incorporate this removal as a small reduction in the effective population size, and have calculated an analytical approximation for this reduction suggesting that the degree to which deleterious mutations reduce effective population size depends primarily on the deleterious mutation rate and recombination rate (Hudson and Kaplan 1995; Nordborg *et al.* 1996).

2.2.2. Problem of linkage disequilibrium between deleterious mutations

One crucial assumption this analytical approximation makes is that deleterious mutations are linked only to nearby neutral variants, and not to each other. This assumption is reasonable if deleterious mutations are reliably cleared before another deleterious mutation linked to the same neutral variant arises, such as if deleterious mutation rates are low. However, deleterious mutation rates are high (Lesecque *et al.* 2012), there are theoretical reasons to believe that deleterious mutations should develop negative linkage disequilibrium (Barton and Otto 2005), and data suggests that many deleterious mutations are linked to each other (Koch *et al.* 2013). Linkage disequilibrium enhances the loss of neutral variation in the somewhat different case of selective sweeps (Barton 1998), suggesting that widespread violation of this assumption could be a serious problem for analytical approximations.

2.2.3. Our model

We take advantage of the recent development of computationally efficient forward-time population genetic simulation software to investigate this problem (Thornton 2014, 2019; Kelleher *et al.* 2018). We simulate the evolution of whole genomes with realistically high deleterious mutation rates and allow linkage to emerge naturally under a realistic recombination process. We simulate large populations over evolutionary time for a variety of deleterious mutation rates, census population sizes, and strengths of background selection against deleterious variants and calculate effective population sizes for the resulting populations. We then compare these 'observed' effective population sizes to the expectation from the analytical approximation.

2.2.4. Effective population size is much lower than expected from analytical approximations at high mutation rates

We find that at realistically high mutation rates (>1 per genome), the effective population sizes observed in our simulations are substantially lower than expected from analytical approximations. At low mutation

rates, analytical approximations and simulations produce similar effective population sizes. This effect does not depend on the census population size across a fivefold range of sizes, and depends only moderately on the strength of selection against deleterious mutations across a tenfold range of strengths. Restricting the occurrence of deleterious mutations to certain regions of the genome (a rough approximation of genes) has mild and inconsistent effects on neutral diversity.

2.2.5. Contributions to the work presented in Appendix B

Initial simulation results which led to this paper were found by me. Coding of the fwdpy11 simulations was done by me, with gracious assistance from Kevin Thornton. First draft of the paper was written by me. Substantial editing and decisions about final data presentation were done by me and Joanna Masel.

2.3. Beneficial mutations can asymmetrically counteract accumulation of mutation load from small deleterious mutations

2.3.1. Preamble

The following section will be a draft manuscript which is intended to be submitted as both a bioRxiv preprint and journal manuscript soon, exploring a potential solution to the problem of accumulating load. There will therefore be some repetition in the next section from the Introduction. The inspiration for this paper came from conversations between Joanna Masel and Jason Bertram. Jason Bertram also found the Fenwick tree algorithm which enables the simulation framework in the paper. I wrote the simulation code used in the paper, generated the results, and wrote the first draft of the paper. Substantial editing and decisions about final data presentation were done by me and Joanna Masel.

2.3.2. Draft paper

Introduction

The average human begins life with upwards of a hundred new mutations not found in their parents (Lynch 2010), of which an average of two are deleterious (Lesecque, Keightley, & Eyre-Walker 2012). Mutation rates on this order are not unique to humans (e.g. Haag-Liautard *et al*. 2007) and some argue the deleterious mutation rate is as high as 10 per individual (Kondrashov 2017). Geneticists have long been worried about the effects of the resulting "mutation load" on health (Crow 1997, Muller 1950). Classical population genetics theory predicts that deleterious mutations reduce fitness from 1 (maximum relative fitness for a mutationless individual) to e^{-U_d}, which means that human fitness is reduced to only 13% of what it could be without deleterious mutations. Even worse, since removing a single deleterious mutation requires one 'selective death', mutation rates above one should result in accumulation of load. How is it that populations such as humans persist in the face of such high deleterious mutation rates?

Some of the concerns raised by mutation load have been resolved. Firstly, mutation load is typically defined as $L = W_{max} - \overline{W}$, where W_{max} represents the fitness of a completely mutationless individual. Since this hypothetical deleterious-mutation-free individual has almost certainly never existed, mutation load concerns depend on assumptions about this individual's fitness. If a mutation-free human could average a hundred offspring, then reducing human fitness to 13% of that poses no real threat.

Secondly, load is much lower than in classical theory if most deleterious mutations are in traits under stabilizing selection (Charlesworth 2013). In a model like this, all mutations modify the value of a higher-level trait and load is determined by the distance between this value and some optimum value. This implies that all mutations are deleterious in a perfect genome, but many mutations may be beneficial in a genome far from the optimum. At equilibrium, this results in load estimates that are quite small (only about 5% for humans) (Charlesworth 2013). But at the molecular level at which new mutations actually occur, a DNA change in a protein is far more likely to simply reduce functionality than to slightly modify

a higher-order trait, suggesting that unconditionally deleterious mutations represent a substantial portion, if not the vast majority, of new mutations. Load contributed by unconditionally deleterious mutations still presents a substantial threat to populations.

Thirdly, high load is no threat to population survival if load is primarily in intrinsically relative fitness traits, such as mating success or intraspecific competition for resources (Agrawal and Whitlock 2012). Defining load in terms of relative, rather than absolute, fitness means that the appropriate W_{max} is the fittest individual in the population, not a mutationless individual. High load would represent large differences in competitive ability between members of a population, not a threat to population survival. Again, however, unconditionally deleterious mutations are likely to be more common than mutations that only affect mating success.

Lastly, load could be cleared faster than it arises even for $U_d > 1$ if epistasis among deleterious mutations was on average synergistic. Unfortunately, empirical data has not supported this case, suggesting that epistasis among new deleterious mutations instead averages to multiplicative (Kouyos, Silander, & Bonhoeffer 2007).

Existing explanations leave a key mutation load paradox unresolved: small deleterious mutations fix at a constant rate, creating an endless series of deterioration. Some proportion of these mutations may affect traits under stabilizing selection or relative fitness traits, but this only serves to slow the rate of constant mutational degradation. The outstanding nature of this problem is an issue for studies inferring load in empirical populations. Such studies typically compare observed data to simulations, but are often forced to periodically re-normalize simulated fitness data to cosmetically remove ongoing degradation (e.g. Fig. S2 to Fig. 2 in (Simons, Turchin, Pritchard, & Sella 2014)).

The speed of this degradation depends on the selective effects of deleterious mutations which can reach fixation. In a simple Wright-Fisher model, the size of the population would determine the size of deleterious mutation capable of reaching fixation, sometimes referred to as a 'drift barrier' (Ohta 1973;

Sung *et al.* 2012). All other effects on the drift barrier in real populations are typically summarized by an 'effective' population size: the required size for a simplified population to have the same statistic of interest as the real population. The most common statistic of interest used to describe an effective population size is heterozygosity at neutral sites, known as the coalescent effective population size, but the drift barrier defines a different effective population size.

Approached from a molecular perspective, there are many loci each with alleles, which in practice can be simplified to two alleles, one beneficial and one deleterious. In this view, all mutations are reversible, and so beneficial and deleterious effect sizes are closely tied. If these effects are small, then most sites are found in the deleterious state, leading to a load of one hundred "lethal equivalents", prompting the expression that we should have 'died one hundred times over' (Kondrashov 1995). Even if the effect sizes are large enough for selection to favor the beneficial variant, realistically high mutation rates and low effective population sizes allow deleterious mutations to swamp out the much lower fraction of beneficial mutations. For example, given $U = 2$, $s = 0.01$, and $N = 10{,}000$, an analytic approximation suggests that more than 30% of new non-neutral mutations would need to be beneficial to counteract deleterious load (using equation 10 in Goyal et al., 2012), which is not plausible.

Our hypothesis instead argues that that population persistence in the face of mutational degradation is accomplished asymmetrically, by rare, large-effect beneficial mutations recovering fitness lost to many small-effect deleterious fixations. A conceptual example of this hypothesis could be a protein accumulating fixations of small deleterious mutations which slightly inhibit its ability to fold, but compensating for this with the introduction of a novel chaperone protein. Additionally, an empirical example of this pattern of asymmetric adaptation to deleterious load has been observed in influenza (Koelle and Rasmussen 2015). While new beneficial mutations should have symmetric effect sizes to new deleterious mutations, the distribution of effect sizes of mutations *conditional on fixation* should be strongly asymmetrical (Kimura 1962). Small-effect deleterious mutations are the most dangerous precisely because no other deleterious mutations fix, but large-effect beneficial mutations, while rare,

have larger fixation probabilities than other beneficials. Indeed, a previous study comparing fitness gain from beneficial fixations with fitness lost from deleterious fixations found that populations remained stable down to a critical effective population size of barely over 100 (Whitlock 2000).

However, this optimistic result ignored the effects of linkage disequilibrium. With the large amount of segregating mutations expected with high mutation rates, the probability of fixation of beneficial mutations will be sharply reduced due to clonal interference (Hill and Robertson 1966), meaning that this result overstated the ability of beneficial mutations to ameliorate load. In other words, the flux of fixations of beneficial mutations is no longer their mutation rate times the census population size times the probability, proportional to the selection coefficient, that they would fix if evolving independently. Instead, this fixation probability is reduced both by clonal interference (negative LD with other beneficial mutations) and by background selection (positive LD with deleterious mutations). These same factors in reverse also increase the fixation probability of deleterious mutations, in a phenomenon sometimes interpreted as a change in drift barrier effective population size.

We model evolution of load in populations with census population sizes that give rise to a realistic level of human neutral diversity, allowing linkage disequilibrium to emerge appropriately. This is in contrast to methods that hold a product such as sN constant, and rescale N to be smaller and s to be larger in order to accelerate computation (Haller and Messer 2017). We employ novel simulation techniques to overcome the computational challenges of such an approach, including the use of binary indexed trees to allow both birth-death and selection processes to occur in O(log N) time, as well as the use of 'linkage blocks' to minimize computational costs from many segregating mutations. We measure fitness trajectories in populations over time to determine whether beneficial mutations arrest degradation even when linkage disequilibrium restricts load. We additionally measure the effect sizes of mutations conditional on fixation to determine the degree of asymmetry between fixed beneficial and deleterious mutations.

Our goal is to determine whether beneficial mutations are sufficient to recover fitness lost to deleterious mutations in the crucial case of realistic mutation rates and linkage disequilibrium. Our metric is fitness

flux, i.e. the average change in fitness in the population. If asymmetric adaptation is sufficient to explain population persistence in the face of accumulating deleterious mutations, then we expect to see positive fitness fluxes even in populations with conservatively low estimates for the rates of beneficial mutations and their effect sizes.

Methods

We performed individual-based forward-time simulations written in C with the following basic structure:

Each individual has two characteristics: a genome, and a fitness value derived from it. The genome of an individual is represented as an array of L non-recombining 'linkage blocks', which is divided into 23 chromosomes, each with two copies. Each block consists of a floating-point variable l_j, calculated by multiplying together the fitness effects of each mutation on that block, such that $l_j = \prod_i (1 + s_i)$. Each individual's fitness w_i is then calculated similarly by multiplying together all linkage blocks, such that $w_i = \prod_{j=1}^{L} (l_j)$. Recombination occurs only at hotspots between linkage blocks via crossing-over events between homologous chromosomes. We simulate exactly two recombination events per chromosome per meiosis, matching data for humans (Pardo-Manuel De Villena and Sapienza 2001), although we don't explicitly simulate a centrosome. Representing a genome as a set of 'linkage blocks' is a good approximation of population genetics in non-microbial species (Neher 2013). Appropriate values of L in humans is in the range of 10^5-10^6 (Wall and Pritchard 2003; Belmont *et al.* 2005; Altshuler *et al.* 2008; Coop *et al.* 2008; Pratto *et al.* 2018), although for computational efficiency we typically use values of L of a few thousand. This simplification should overestimate the effect of linkage between selected mutations, which is conservative with respect to the ability of beneficial mutations to counteract load.

After simulating recombination's contribution to a new gamete, we next simulate mutation. New deleterious mutations appear with a genome-wide deleterious mutation rate U_d = 2, and their fitness effects are drawn from a distribution based on a large empirical study of Europeans (Kim *et al.* 2017). They estimated a gamma distribution for the population-scaled selection coefficient $2N_e s$ with mean

-224.33, shape parameter $\alpha = 0.169$ and scale parameter $\beta = 1327.4$. After drawing a value of $2N_e s$ from this distribution, we rescale to s using their inferred $N_e = 11{,}823$. New beneficial mutations appear with a genome-wide beneficial mutation rate U_b with fitness effects drawn from an exponential distribution with mean s_b. We explore a range of values for U_b and s_b that we consider *a priori* plausible: $U_b \sim 0.0001\text{-}0.01$ and $s_b \sim 0.001\text{-}0.01$.

We simulate a Moran model with constant population size *N*. An individual chosen uniformly at random dies each time step and is replaced by a child produced by two parents, who are chosen with probability proportional to their fitness w_i. Each generation consists of *N* time steps. The fitnesses of the population are stored in an unsorted array — in a naïve implementation, births and deaths would be trivial computations, but sampling from the fitness array would be O(*N*). The current fastest forward-time genetic simulation tools for large population sizes (e.g. fwdpy (Thornton 2014, 2019)) preprocess cumulants each generation in a Wright-Fisher model; this speeds up sampling from the fitness array, and while the processing algorithm is O(*N*), it only needs to be performed once per generation. We instead use a binary indexed tree (Fenwick 1994) to sample fitnesses efficiently according to the cumulative probability distribution — both updating and sampling from the tree is O(log *N*). Our scheme is expected to have similar efficiency but allows us increased modeling flexibility, such as the future expansion of this approach to absolute fitness and more complex life history models (Bertram and Masel 2019), which will allow better treatment of reproductive compensation (Ober *et al.* 1999). The population is initialized with completely mutationless individuals, so each simulation begins with a 'burn-in' phase where variation increases to stable levels, as seen in Figure 1. We end the burn-in phase 500 generations after a linear regression of the variance in fitness for a rolling window of 200 generations produces a slope less than $0.007/N$. We chose this threshold to be both arbitrarily close to zero, and scale with population size, since the variance in fitness also scales with population size. As expected from the dynamics of a population undergoing regular sweeps, this burn-in phase is O(log *N*) generations (~900-1300 generations for the range of *N* tested).

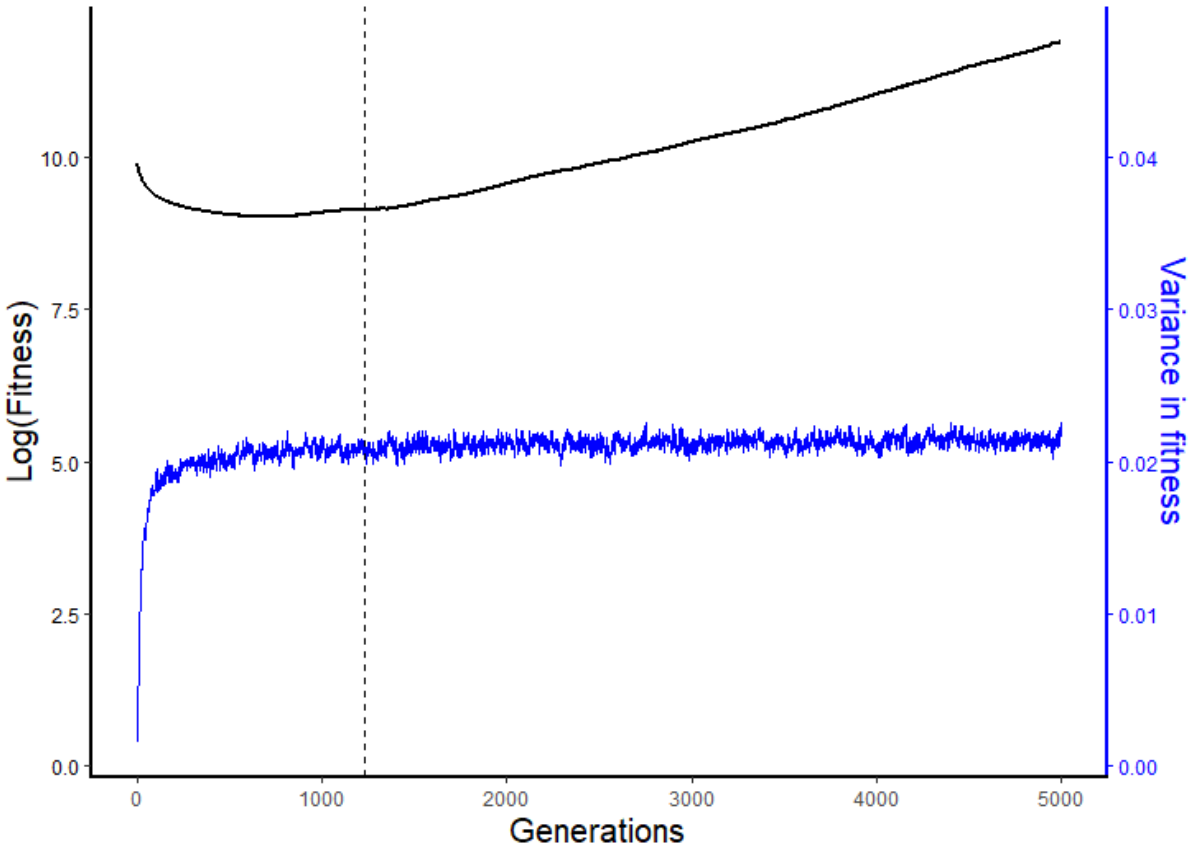

Figure 1. The burn-in phase ends once variance in fitness plateaus. Plotted is the log mean fitness (black line) and variance in fitness (blue line) each generation for a single population, where N = 20,000, each individual has 23 chromosomes with 50 linkage blocks per chromosome, a beneficial mutation rate of 0.002, and a mean beneficial effect size of 0.0025. Dashed line indicates the end of the burn-in phase, in generation 1227. The slope of log mean fitness after the burn-in phase describes the net fitness flux in the population over time.

The main output of an individual simulation is the net fitness flux of the population. To calculate this, we regress the log mean population fitness at the end of each generation after burn-in against the number of generations since the end of the burn-in phase and use the slope of this regression as the net fitness flux per generation. In Figure 1, this is the slope of the black line after the dashed line indicating the end of the burn-in phase. After the end of the burn-in phase, the expected rate of beneficial and deleterious fixations, as well as their expected effect sizes, will be constant. Given multiplicative fitness effects, this means that population fitness change will be linear on a log scale in all simulations.

To determine the parameter values that yield zero net fitness flux for Figure 2, we determined which combinations of our two beneficial parameters U_b and s_b, beneficial mutation rate and mean effect size, would be precisely enough to counteract fitness lost to deleterious mutations. With fixed values of U_b, we ran individual simulations where we varied the value of s_b to determine the value of s_b that produces a given net fitness flux for the set U_b. We use an algorithm (modified from Numerical Recipes in C, Second Edition) that brackets the target net fitness flux (finding values of s_b which produce higher and lower net fitness fluxes), and then uses a bisection method to iteratively run simulations with values of s_b halfway between the brackets, updating either the upper or lower bracket depending on whether the halfway s_b results in a lower or higher net fitness flux than the target. This runs until it finds a value of s_b that is within a margin of error of the target net fitness flux (we use plus or minus 0.00005).

Although the census population size *N* is a parameter of our model, the effective population size N_e emerges over the course of a given simulation. Our simulations can record the genetic history of the population using the tree-sequence recording tools from the tskit package (Kelleher *et al.* 2018), which allows us to retroactively add neutral mutations after each simulation. We calculate N_e using neutral heterozygosity. The choice of neutral mutation rate will not affect estimated N_e; we arbitrarily chose 1.0×10^{-6} per linkage block, or 1.15×10^{-4} per haploid genome. We find that a census population size of 20,000 results in $N_e \sim 7500$, on the order of effective population sizes inferred for ancestral human populations (Tenesa *et al.* 2007).

Tree sequence recording also allows us to identify all non-neutral mutations which have reached fixation, and thus to determine the degree of asymmetry in the effect sizes of those mutations that go on to fix. Note that without tree sequence recording, this information would be inaccessible due to the way we store fitness information on linkage blocks. However, using tree-sequence recording for all non-neutral mutations increases the computation time of simulations. For results which require many simulations

where net fitness flux is the main output, we typically do not use tree sequence recording to maximize computational efficiency.

Results

The threshold for a positive mean population fitness flux depends primarily on the mean beneficial effect size, not on the beneficial mutation rate (Figure 2; black). The black line in Figure 2 shows the values for which there is exactly zero change in fitness, and the blue line shows the parameters that produce a 10% increase in fitness every 100 generations, e.g. to track changes in the external environmental. In the absence of environmental change, population persistence is possible for $\bar{s_b} > \sim 0.002$ and $U_b > 0.001$. While there is great uncertainty in the true values of these parameters, this range seems entirely plausible.

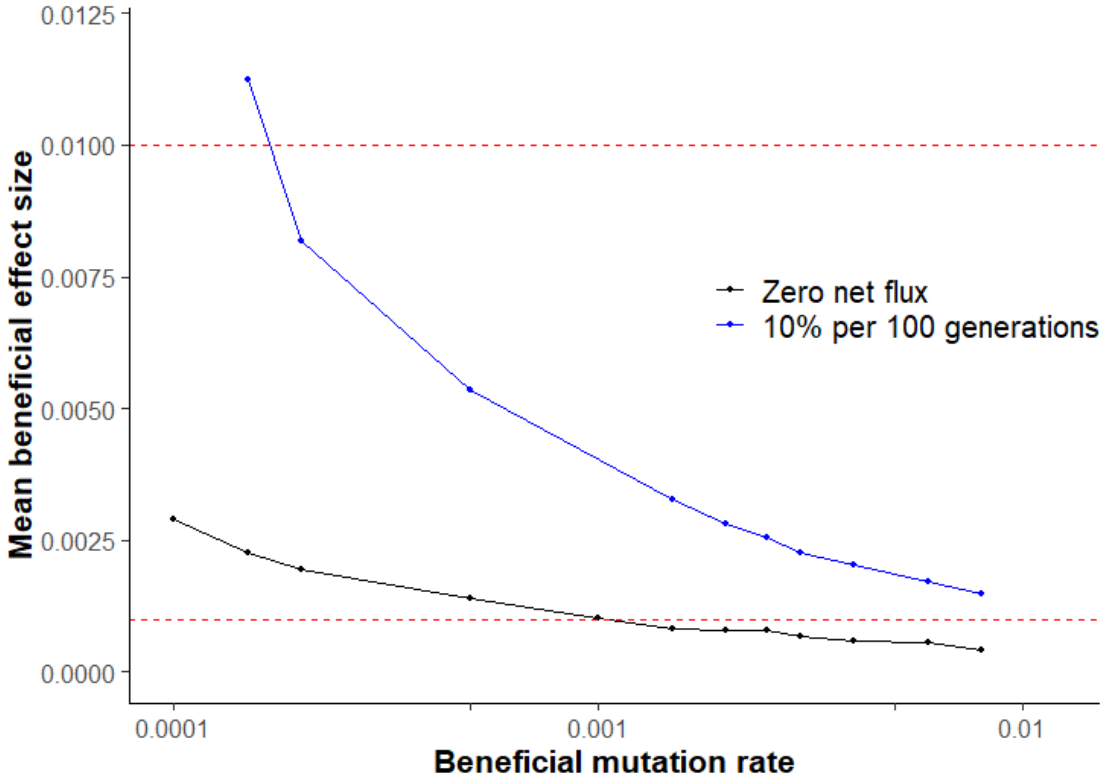

Figure 2. Beneficial mutations are sufficient to counteract deleterious mutations across the plausible range of beneficial parameters. Black line shows combinations of beneficial mutation parameters that produce zero net fitness flux. Blue line shows combinations which produce a 10% increase in fitness every 100 generations. All populations simulated with N = 20,000, genome-wide deleterious mutation rate of 2, and 23 chromosomes with 50 linkage blocks per chromosome. Any combinations of beneficial mutation rate and mean effect size below the black line produce net degradation. Red dashed lines show plausible upper and lower estimates of the mean effect size of new beneficial mutations in humans.

The reason that such low beneficial mutation rates are sufficient for population persistence is that each beneficial mutation that fixes has a much greater magnitude selection coefficient than each deleterious mutation that fixes (Figure 3). Beneficial fixations are larger on average than new beneficial mutations, and deleterious fixations are much smaller on average than new deleterious mutations. Even in simulations that improve in fitness on average, deleterious fixations outnumber beneficial fixations.

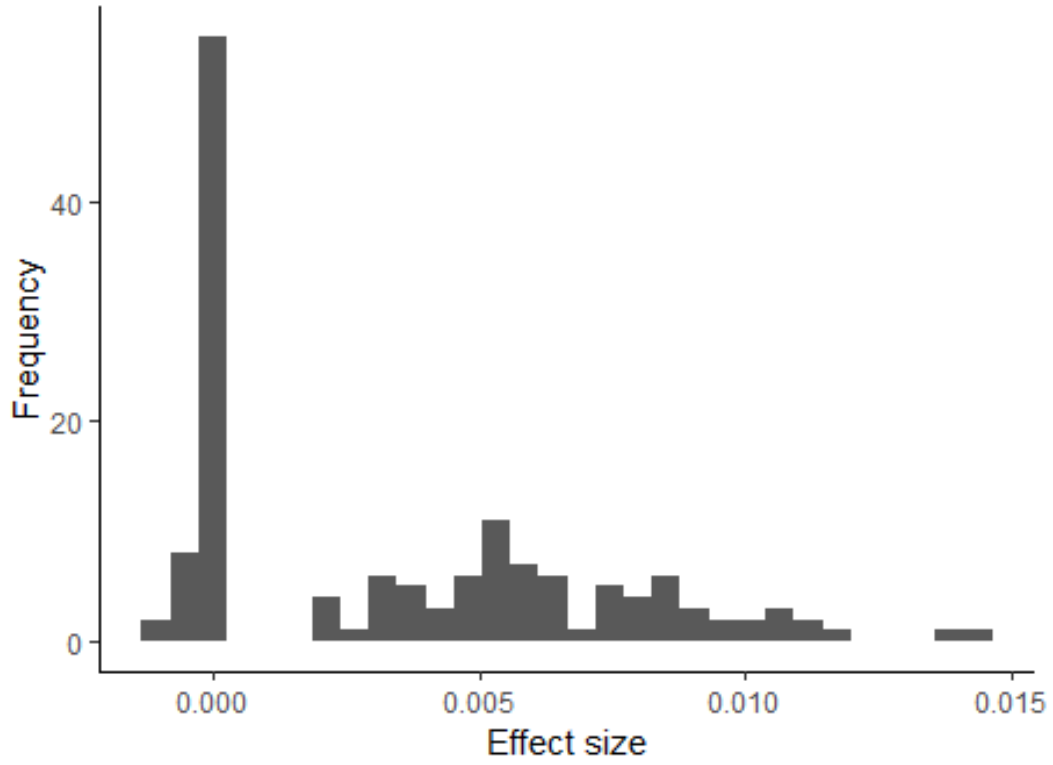

Figure 3. Effect sizes of fixed beneficial and deleterious mutations are strongly asymmetrical. Shown is a histogram of the distribution of effect sizes of fixed mutations after 5000 generations, in a population of N = 20,000 with individuals having 23 chromosomes, 50 linkage blocks per chromosome, with a beneficial mutation rate of 0.002 per generation and mean beneficial effect size of 0.0025.

Discussion

Our results provide a new answer to the paradox posed by the accumulation of small-effect deleterious mutations under realistically high mutation rates. Previous answers either attempted to explain what prevents small-effect deleterious mutations from accumulating or modeled an equilibrium under which continuous deterioration was impossible. Our alternative answer allows that they do in fact accumulate. This does not lead to population deterioration because a smaller number of beneficial fixations of greater

size successfully counteracts many more small-effect deleterious fixations. This success is consistent across a wide range of plausible beneficial parameters.

This asymmetry parallels known features of molecular adaptation. While mutations that, for instance, jeopardize the stable folding of a protein unavoidably accumulate in individual proteins, many such deleterious fixations can be ameliorated at once by the evolution of chaperones to stabilize proteins. A pattern of many small mutations, each of which cannot be effectively cleared, being counteracted by compensatory mutations with global effects has previously been predicted by drift barrier theory (Lynch 2010b; Rajon and Masel 2011; Sung *et al.* 2012; Xiong *et al.* 2017).

Drift barrier theory typically emphasizes the causal importance of effective population size (the eponymous barrier) on producing a ratcheting effect that leads to increased molecular and organismal complexity. See Figure 4 for an illustrative causal diagram. First, the drift barrier determines genome size, based on whether mutations resulting in small increases in genome size have fitness effects above or below the drift barrier. Larger genomes have higher mutation rates, accelerating the input of mutations which increase genome size. This feedback loop forces the evolution of novel molecular processes (adding functional complexity) to compensate for lost fitness. Closing the loop all the way back to effective population size requires the additional argument that functional complexity ultimately decreases effective population size, perhaps due to increases in body size or ecological specialization. However, the feedback loop between mutation rate and genome size should be sufficient to force increased complexity on its own.

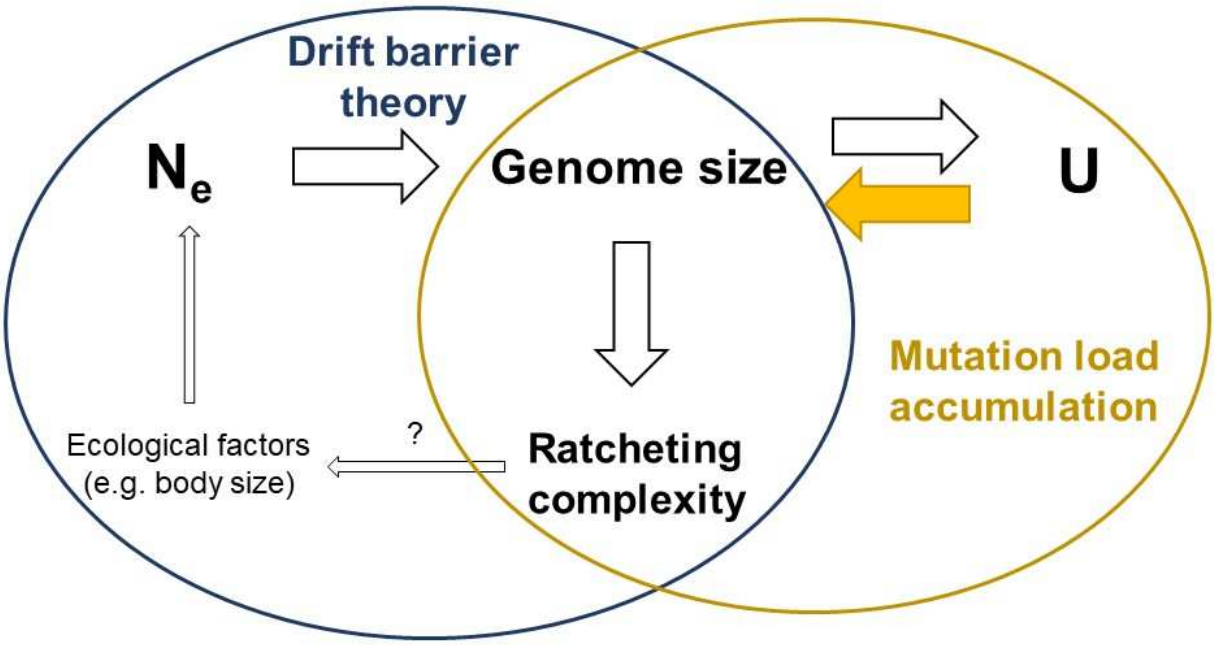

Figure 4. Changes in either effective population size or deleterious mutation rate can drive a feedback loop of ratcheting complexity. Decreasing effective population size lowers the drift barrier, allowing mutations with larger fitness effects to drift to fixation. A sufficiently large deleterious mutation rate could also inundate populations with deleterious mutations. In either case, if some such deleterious mutations expand the size of the genome, this will increase the genome-wide mutation rate and start a feedback loop. The only way out of this feedback loop is novel compensatory mutations adapting global solutions to many small deleterious problems, thus increasing molecular complexity. If increased complexity results in other ecological changes (increased body size, specialization, etc.) which lead to decreased population sizes, this creates a further feedback loop.

Our results suggest a shift in perspective, placing causal emphasis on high mutation rates. Sufficiently high mutation rates can lead to accumulation of small mutations (such as those increasing genome size) even in species with large population sizes, starting the feedback loop of increasing genome size and mutation rates. Increased mutation rate also provides a greater input of novel beneficial mutations, assisting the ratcheting of molecular complexity. In addition, background selection against deleterious

mutations reduces effective population sizes, and this effect can be substantial at high mutation rates (as in chapter 2, or using equation 4 in Charlesworth 2012 for U > 1). Note that this feedback loop creates a ratchet of increasing complexity without requiring that complexity itself loops back to decrease effective population sizes.

Although we show that a small input of large-effect beneficial mutations are sufficient to explain the persistence of populations in the face of a much larger influx of accumulating small-effect deleterious mutations, we don't claim that this is the only possible solution. The prevailing solution, synergistic epistasis, is seriously flawed, due to empirical tests of *de novo* deleterious mutations in bacteria and eukaryotic microbes which consistently show epistatic effects between mutations, but do not show synergistic epistasis on average (Elena and Lenski 1997; Kouyos *et al.* 2007). On the other hand, segregating deleterious alleles in humans and flies show negative linkage disequilibrium (Sohail *et al.* 2017; Lee 2022), potentially a signature of synergistic epistasis. Similarly, experimental evolution studies consistently find diminishing returns epistasis (synergistic epistasis at the other end of the fitness scale) between new beneficial mutations (e.g. Barrick *et al.* 2009). One way to reconcile these observations would be to hypothesize that mutational effects are multiplicative (or antagonistic) between functional modules while being synergistic within modules (Rice 1998; Wei and Zhang 2019). This could potentially purge mutation load fast enough to avoid degradation while being consistent with empirical results showing no tendency towards synergistic epistasis on average.

Alternatively, sexual selection can assist with purging load, since assortative mating results in increased variance in fitness, similar to synergistic epistasis. Intuitively, this effect would be stronger under mating systems which allow for more skewed distributions of offspring, but some studies suggest that assortative mating might also be able to prevent mutational degradation even under perfect monogamy with mutual mate choice (Hooper and Miller 2008)., Another phenomenon that could combat load accumulation is hybridization between isolated populations. The small deleterious mutations drifting to fixation in each population should be independent, and so hybridization or introgression would introduce reversions of

many, if not all, of the fixed deleterious mutations. Of course, such hybrids would bring similar amounts of new deleterious mutations from the other population, so that the effect of hybridization would be to suddenly increase the number of sites with segregating deleterious variants without necessarily changing average fitness. This would allow for the production (via recombination) of haplotypes with more extreme fitness differences, increasing variance in fitness and potentially purging load more efficiently. Although none of these explanations has been fully explored in this context, our simulation structure offers a way to test these explanations individually or in concert.

Understanding how mutation load might be stabilized is a precondition for addressing a separate, long-standing concern of geneticists: that load might be increasing in modern humans because of recent changes to human lifestyles or technology. For example, if mutation rate, beginning already at a critically high level, increases further due to increased paternal age, or if selection against deleterious mutations is relaxed due to modern medicine, the perception has been that load should increase, potentially with disastrous consequences (Crow 1997; Lynch 2016).

Intriguingly, preliminary results suggest that increasing the mutation rate leads populations to adapt faster, rather than slower. To estimate the likely recent increase in human mutation rate, we note that mean paternal age in the U.S. increased from 27.4 to 30.9 years of age between 1972 and 2015 (Khandwala *et al.* 2017), which is expected to correspond to a 12 percent increase in mutation rate (Kong *et al.* 2012). When we increase the mutation rate for both deleterious and beneficial mutations by 10 percent for a reference population with $N = 10{,}000, U_b = 0.002, \bar{s}_b = 0.0025$ and all other parameters the same as in Figure 2, we find that the populations with increased mutation rates take 207 generations to increase their fitness by 10%, compared to 217 generations for the reference population.

To properly address historical load concerns, however, would require a model that describes steady state load in the context of absolute fitness. This would allow for fitness (and therefore load) to have absolute meaning in terms of births and deaths, for load to reach a steady state (which it does not do in our current model), and for load accumulation to threaten extinction by mutational meltdown (Lynch *et al.* 1995).

Under this scenario, it would be possible to measure the difference in steady state load under increased mutation rate or relaxed selection and relate differences in load to extinction risk or offspring produced, both of which are likely better measurements of the health effects of load than changes in speed of adaptation.

Ultimately, we have shown that populations are only able to survive the constant accumulation of mildly deleterious load by acquiring large effect beneficial mutations. Mutation load may therefore not present a paradox with respect to population persistence, but this does still suggest that load plays a crucial evolutionary role. Only some populations, like bacteria, are able to solve load problems and still retain simple, efficient genomes. Previous hypotheses have focused on the size of such populations as the crucial divider between being able to avoid load with a small, simple genome or being forced into ratcheting molecular complexity to solve load problems. But large bacterial populations also have deleterious mutation rates below 1, which could equally explain how they can maintain low load. Whether due to increased mutation rates or lowered population sizes, we argue that any population subject to the onslaught of small deleterious mutations will be forced to the other solution: constantly finding innovative molecular solutions to stay ahead of perpetual degradation. The pressure of mutation load might therefore be a primary driver behind molecular complexity across the entire tree of life.

Chapter 3: Conclusions

Of course, populations do in fact manage to adapt to new environments and have not yet collapsed due to mutation load, so the fact that this dissertation concludes that mutation load problems have solutions is not a massive breakthrough in the field. However, I have shown in this dissertation that previous approaches were in fact unable to find satisfactory solutions to some mutation load problems, suggesting that these approaches are missing some important features about population genetics.

One feature of many of the models I discuss is the general tactic of trying to summarize genome-wide patterns of load by extrapolating from single-locus models. Many approaches to Haldane's Dilemma rightly and wrongly involved extrapolating the load involved in substitution at a single locus to many loci. Analytical approximations of the effect of background selection on neutral diversity multiply together the effects of many single deleterious mutations, assuming linkage equilibrium between the deleterious mutations. Previous models of load accumulation, including ones with beneficial mutations, assumed linkage equilibrium between all mutations, allowing for simply multiplying single-locus effects by the mutation rate to beneficial or deleterious mutations.

The main conclusion of this dissertation is that extrapolating single-locus models fails precisely in the case most relevant to mutation load: high rates of deleterious mutations entering populations with moderate population sizes. Such populations can expect to have genomes rife with segregating deleterious mutations and persistent linkage disequilibrium between mutations under selection, both features empirically observed in human populations and other similar species. Unfortunately, these features are extremely difficult to mathematically model and previous analytical approaches were forced to make some sort of simplifying assumptions for tractability.

However, we now have a new tool at our disposal to investigate these sorts of populations: forward-time simulations. For the first time, it's computationally feasible to simulate sufficiently large genomes in sufficiently large populations to capture the effects of deleterious load at many potentially linked loci. I

demonstrate that these new tools allow more accurate descriptions of neutral diversity under background selection and a demonstration that beneficial mutations are able to asymmetrically counteract small-effect deleterious mutations, but these are only some of the possible effects of mutation load at many linked loci.

Literature Cited

Agrawal A. F., and M. C. Whitlock, 2012 Mutation Load: The Fitness of Individuals in Populations Where Deleterious Alleles Are Abundant. Annu. Rev. Ecol. Evol. Syst 43: 115–35. https://doi.org/10.1146/annurev-ecolsys-110411-160257

Altshuler D., M. J. Daly, and E. S. Lander, 2008 Genetic mapping in human disease. Science (80-.). 322: 881–888. https://doi.org/10.1126/science.1156409

Awadalla P., J. Gauthier, R. A. Myers, F. Casals, F. F. Hamdan, *et al.*, 2010 Direct measure of the de novo mutation rate in autism and schizophrenia cohorts. Am. J. Hum. Genet. 87: 316–324. https://doi.org/10.1016/j.ajhg.2010.07.019

Barrick J. E., D. S. Yu, S. H. Yoon, H. Jeong, T. K. Oh, *et al.*, 2009 Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature 461: 1243–1247. https://doi.org/10.1038/nature08480

Barton N. H., 1998 The effect of hitch-hiking on neutral genealogies. Genet. Res. 72: 123–133. https://doi.org/10.1017/S0016672398003462

Barton N. H., and S. P. Otto, 2005 Evolution of recombination due to random drift. Genetics 169: 2353–2370. https://doi.org/10.1534/genetics.104.032821

Belmont J. W., A. Boudreau, S. M. Leal, P. Hardenbol, S. Pasternak, *et al.*, 2005 A haplotype map of the human genome. Nature 437: 1299–1320. https://doi.org/10.1038/nature04226

Bertram J., K. Gomez, and J. Masel, 2017 Predicting patterns of long-term adaptation and extinction with population genetics. Evolution (N. Y). https://doi.org/10.1111/evo.13116

Bertram J., and J. Masel, 2019 Density-dependent selection and the limits of relative fitness. Theor. Popul. Biol. 129: 81–92. https://doi.org/10.1016/j.tpb.2018.11.006

Bertram J., 2021 Allele frequency divergence reveals ubiquitous influence of positive selection in Drosophila. PLOS Genet. 17: e1009833. https://doi.org/10.1371/journal.pgen.1009833

Buffalo V., and G. Coop, 2020 Estimating the genome-wide contribution of selection to temporal allele frequency change. Proc. Natl. Acad. Sci. U. S. A. 117: 20672–20680. https://doi.org/10.1073/pnas.1919039117

Charlesworth B., M. T. Morgan, and D. Charlesworth, 1993 The Effect of Deleterious Mutations on Neutral Molecular Variation. Genetics 134: 1289–1303.

Charlesworth B., 2009 Effective population size and patterns of molecular evolution and variation. Nat. Rev. Genet. 10: 195–205. https://doi.org/10.1038/nrg2526

Charlesworth B., 2012 The Effects of Deleterious Mutations on Evolution at Linked Sites. Genetics 190: 5–22. https://doi.org/10.1534/genetics.111.134288

Charlesworth B., 2013 Why we are not dead one hundred times over. Evolution (N. Y). 67: 3354–3361. https://doi.org/10.1111/evo.12195

Comeron J. M., 2014 Background Selection as Baseline for Nucleotide Variation across the Drosophila Genome. PLoS Genet. 10: e1004434. https://doi.org/10.1371/journal.pgen.1004434

Comeron J. M., 2017 Background selection as null hypothesis in population genomics: Insights and challenges from drosophila studies. Philos. Trans. R. Soc. B Biol. Sci. 372: 20160471. https://doi.org/10.1098/rstb.2016.0471

Conrad D. F., M. Jakobsson, G. Coop, X. Wen, J. D. Wall, *et al.*, 2006 A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. Nat. Genet. 38: 1251–1260. https://doi.org/10.1038/ng1911

Coop G., X. Wen, C. Ober, J. K. Pritchard, and M. Przeworski, 2008 High-resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans. Science

(80-.). 319: 1395–1398. https://doi.org/10.1126/science.1151851

Crow J. F., 1958 Some Possibilities for Measuring Selection Intensities in Man. Hum. Biol. 30: 1–13.

Crow J. F., 1997 The high spontaneous mutation rate: is it a health risk? Proc. Natl. Acad. Sci. U. S. A. 94: 8380–6. https://doi.org/10.1073/pnas.94.16.8380

Cutter A. D., and B. A. Payseur, 2013 Genomic signatures of selection at linked sites: unifying the disparity among species. Nat. Rev. Genet. 14: 262–274. https://doi.org/10.1038/nrg3425.Genomic

Desai M. M., and D. S. Fisher, 2007 Beneficial mutation-selection balance and the effect of linkage on positive selection. Genetics 176: 1759–1798. https://doi.org/10.1534/genetics.106.067678

Elena S. F., and R. E. Lenski, 1997 Test of synergistic interactions among deleterious mutations in bacteria. Nature 390: 395–398. https://doi.org/10.1038/37108

Elyashiv E., S. Sattath, T. T. Hu, A. Strutsovsky, G. McVicker, *et al.*, 2016 A Genomic Map of the Effects of Linked Selection in Drosophila. PLoS Genet. 12: e1006130. https://doi.org/10.1371/journal.pgen.1006130

Enard D., L. Cai, C. Gwennap, and D. A. Petrov, 2016 Viruses are a dominant driver of protein adaptation in mammals. Elife 5: e12469. https://doi.org/10.7554/eLife.12469

Eöry L., D. L. Halligan, and P. D. Keightley, 2010 Distributions of selectively constrained sites and deleterious mutation rates in the hominid and murid genomes. Mol. Biol. Evol. 27: 177–192. https://doi.org/10.1093/molbev/msp219

Ewens W. J., 1970 Remarks on the Substitutional Load. Theor. Popul. Biol. 139: 129–139.

Ewens W., 2004 *Mathematical Population Genetics*. Springer US, New York.

Exposito-alonso M., O. Bossdorf, R. Nielsen, and D. Weigel, 2019 Natural selection on the Arabidopsis thaliana genome in present and future climates. Nature 573: 126–129.

https://doi.org/10.1038/s41586-019-1520-9

Felsenstein J., 1971 On the Biological Significance of the Cost of Gene Substitution. Am. Nat. 105: 1–11.

Felsenstein J., 1974 The evolutionary advantage of recombination. Genetics 78: 737–756. https://doi.org/10.1093/genetics/83.4.845

Fenwick P. M., 1994 A new data structure for cumulative frequency tables. Softw. Pract. Exp. 24: 327–336. https://doi.org/10.1002/spe.4380240306

Fisher R., 1930 *The Genetical Theory of Natural Selection*.

Galeota-Sprung B., P. Sniegowski, and W. Ewens, 2020 Mutational Load and the Functional Fraction of the Human Genome. Genome Biol. Evol. 12: 273–281. https://doi.org/10.1093/gbe/evaa040

Galtier N., 2016 Adaptive Protein Evolution in Animals and the Effective Population Size Hypothesis. PLoS Genet. 12: 1–23. https://doi.org/10.1371/journal.pgen.1005774

Gilbert K. J., F. Pouyet, L. Excoffier, and S. Peischl, 2020 Transition from Background Selection to Associative Overdominance Promotes Diversity in Regions of Low Recombination. Curr. Biol. 30: 101-107.e3. https://doi.org/10.1016/j.cub.2019.11.063

Goyal S., D. J. Balick, E. R. Jerison, R. A. Neher, B. I. Shraiman, *et al.*, 2012 Dynamic Mutation–Selection Balance as an Evolutionary Attractor. Genetics 191: 1309–1319. https://doi.org/10.1534/genetics.112.141291

Haag-Liautard C., M. Dorris, X. Maside, S. Macaskill, D. L. Halligan, *et al.*, 2007 Direct estimation of per nucleotide and genomic deleterious mutation rates in Drosophila. Nature 445: 82–85. https://doi.org/10.1038/nature05388

Haldane J., 1937 The Effect of Variation on Fitness. Am. Nat. 71: 337–349.

Haldane J., 1957 The Cost of Natural Selection. Genetics 55: 511–524.

Haller B. C., and P. W. Messer, 2017 SLiM 2: Flexible, interactive forward genetic simulations. Mol. Biol. Evol. 34: 230–240. https://doi.org/10.1093/molbev/msw211

Hill W., and A. Robertson, 1966 The effect of linkage on limits to artificial selection. Genet. Res. 8: 269–294. https://doi.org/10.1017/S001667230800949X

Hill W. G., and A. Robertson, 1968 Linkage disequilibrium in finite populations. TAG Theor. Appl. Genet. 38: 226–231. https://doi.org/10.1007/bf01245622

Hinds D. A., A. P. Kloek, M. Jen, X. Chen, and K. A. Frazer, 2006 Common deletions and SNPs are in linkage disequilibrium in the human genome. Nat. Genet. 38: 82–85. https://doi.org/10.1038/ng1695

Hooper P. L., and G. F. Miller, 2008 Mutual mate choice can drive costly signaling even under perfect monogamy. Adapt. Behav. 16: 53–70. https://doi.org/10.1177/1059712307087283

Hudson R. R., and N. L. Kaplan, 1994 Gene Trees with Background Selection, pp. 140–153 in *Non-neutral Evolution: Theories and Molecular Data*, edited by Golding B. Springer US, Boston, MA.

Hudson R. R., and N. L. Kaplan, 1995 Deleterious background selection with recombination. Genetics 141: 1605–1617.

Jensen J. D., B. A. Payseur, W. Stephan, C. F. Aquadro, M. Lynch, *et al.*, 2019 The importance of the Neutral Theory in 1968 and 50 years on: a response to Kern and Hahn 2018. Evolution (N. Y). 73: 111–114. https://doi.org/10.1111/evo.13650

Johri P., C. F. Aquadro, M. Beaumont, B. Charlesworth, L. Excoffier, *et al.*, 2021 Statistical Inference in Population Genomics. biorxiv 2021.10.27.466171. https://doi.org/10.1101/2021.10.27.466171

Keightley P. D., and S. P. Otto, 2006 Interference among deleterious mutations favours sex and recombination in finite populations. Nature 443: 89–92. https://doi.org/10.1038/nature05049

Kelleher J., A. M. Etheridge, and G. McVean, 2016 Efficient Coalescent Simulation and Genealogical

Analysis for Large Sample Sizes. PLoS Comput. Biol. 12: e1004842. https://doi.org/10.1371/journal.pcbi.1004842

Kelleher J., K. R. Thornton, J. Ashander, and P. L. Ralph, 2018 Efficient pedigree recording for fast population genetics simulation. PLoS Comput. Biol. 14: e1006581.

Kern A. D., and M. W. Hahn, 2018 The Neutral Theory in Light of Natural Selection. Mol. Biol. Evol. 35: 1366–1371. https://doi.org/10.1093/molbev/msy092

Khandwala Y. S., C. A. Zhang, Y. Lu, and M. L. Eisenberg, 2017 The age of fathers in the USA is rising: An analysis of 168 867 480 births from 1972 to 2015. Hum. Reprod. 32: 2110–2116. https://doi.org/10.1093/humrep/dex267

Kim Y., and W. Stephan, 2000 Joint effects of genetic hitchhiking and background selection on neural variation. Genetics 155: 1415–1427. https://doi.org/10.1093/genetics/155.3.1415

Kim B., C. Huber, and K. Lohmueller, 2017 Inference of the Distribution of Selection Coefficients for New Nonsynonymous Mutations Using Large Samples. Genetics 206: 345–361. https://doi.org/10.1534/genetics.116.197145/-/DC1.1

Kimura M., 1960 Optimum Mutation Rate and Degree of Dominance as Determined by the Principle of Minimum Genetic Load. Genetics 57: 21–34.

Kimura M., 1962 On the probability of fixation of mutant genes in a population. Genetics 47: 713–719. https://doi.org/10.1007/BF01418872

Kimura M., T. Maruyama, and J. F. Crow, 1963 The mutation load in small populations. Genetics 48: 1303–1312.

Kimura M., 1968 Evolutionary Rate at the Molecular Level. Nature 217: 624–626.

Kimura M., 1969 The number of heterozygous nucleotide sites maintained in a finite population due to

steady flux of mutations. Genetics 61: 893–903.

Kimura M., and J. F. Crow, 1969 Natural selection and gene substitution. Genet. Res. 13: 127–141. https://doi.org/10.1017/S0016672300002846

Kimura M., and T. Ohta, 1971 On the rate of molecular evolution. J. Mol. Evol. 1: 1–17. https://doi.org/10.1007/BF01659390

King J. L., and T. H. Jukes, 1969 Non-Darwinian Evolution. Science (80-.). 164: 788–798.

Koch E., M. Ristroph, and M. Kirkpatrick, 2013 Long range linkage disequilibrium across the human genome. PLoS One 8: e80754. https://doi.org/10.1371/journal.pone.0080754

Koelle K., and D. A. Rasmussen, 2015 The effects of a deleterious mutation load on patterns of influenza A/H3N2's antigenic evolution in humans. Elife 4: e07361. https://doi.org/10.7554/eLife.07361

Kondrashov A. S., and J. F. Crow, 1993 A molecular approach to estimating the human deleterious mutation rate. Hum Mutat 2: 229–234. https://doi.org/10.1002/humu.1380020312

Kondrashov A. S., 1995 Contamination of the genome by very slightly deleterious mutations: why have we not died 100 times over? J. Theor. Biol. 175: 583–594. https://doi.org/10.1006/jtbi.1995.0167

Kondrashov A. S., 2017 *Crumbling Genome: The Impact of Deleterious Mutations on Humans*. John Wiley & Sons, Inc, Hoboken, NJ.

Kong A., M. L. Frigge, G. Masson, S. Besenbacher, P. Sulem, *et al.*, 2012 Rate of de novo mutations and the importance of father-s age to disease risk. Nature 488: 471–475. https://doi.org/10.1038/nature11396

Kouyos R. D., O. K. Silander, and S. Bonhoeffer, 2007 Epistasis between deleterious mutations and the evolution of recombination. Trends Ecol. Evol. 22: 308–315. https://doi.org/10.1016/j.tree.2007.02.014

Kulathinal R. J., S. M. Bennett, C. L. Fitzpatrick, and M. A. F. Noor, 2008 Fine-scale mapping of recombination rate in Drosophila refines its correlation to diversity and divergence. Proc. Natl. Acad. Sci. U. S. A. 105: 10051–10056. https://doi.org/10.1073/pnas.0801848105

Lee Y. C. G., 2022 Synergistic epistasis of the deleterious effects of transposable elements. Genetics 220. https://doi.org/10.1093/genetics/iyab211

Lesecque Y., P. D. Keightley, and A. Eyre-Walker, 2012 A resolution of the mutation load paradox in humans. Genetics 191: 1321–1330. https://doi.org/10.1534/genetics.112.140343

Lohmueller K. E., A. Albrechtsen, Y. Li, S. Y. Kim, T. Korneliussen, *et al.*, 2011 Natural Selection Affects Multiple Aspects of Genetic Variation at Putatively Neutral Sites across the Human Genome. PLoS Genet. 7: e1002326. https://doi.org/10.1371/journal.pgen.1002326

Lynch M., J. Conery, and R. Burger, 1995 Mutational Meltdowns in Sexual Populations. Evolution (N. Y). 49: 1067–1080.

Lynch M., 2010a Rate, molecular spectrum, and consequences of human mutation. Proc. Natl. Acad. Sci. U. S. A. 107: 961–8. https://doi.org/10.1073/pnas.0912629107

Lynch M., 2010b Evolution of the mutation rate. Trends Genet. 26: 345–352. https://doi.org/10.1016/j.tig.2010.05.003

Lynch M., 2016 Mutation and human exceptionalism: Our future genetic load. Genetics 202: 869–875. https://doi.org/10.1534/genetics.115.180471

Machado H. E., A. O. Bergland, R. Taylor, S. Tilk, E. Behrman, *et al.*, 2021 Broad geographic sampling reveals the shared basis and environmental correlates of seasonal adaptation in Drosophila. Elife 10. https://doi.org/10.7554/eLife.67577

Masel J., 2011 Genetic drift. Curr. Biol. 21: R837–R838. https://doi.org/10.1016/j.cub.2011.08.007

Maynard Smith J., 1968 “Haldane’s dilemma” and the rate of evolution. Nature 219: 1114–1116. https://doi.org/10.1038/2191114a0

Maynard Smith J., and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. (Camb). 23: 23–35. https://doi.org/10.1017/S0016672308009579

Maynard Smith J., 1976 What Determines the Rate of Evolution? Am. Nat. 110: 331–338.

McCandlish D. M., and A. Stoltzfus, 2014 Modeling evolution using the probability of fixation: History and implications. Q. Rev. Biol. 89: 225–252. https://doi.org/10.1086/677571

McEvoy B. P., J. E. Powell, M. E. Goddard, and P. M. Visscher, 2011 Human population dispersal “Out of Africa” estimated from linkage disequilibrium and allele frequencies of SNPs. Genome Res. 21: 821–829. https://doi.org/10.1101/gr.119636.110

McVicker G., D. Gordon, C. Davis, and P. Green, 2009 Widespread genomic signatures of natural selection in hominid evolution. PLoS Genet. 5: e1000471. https://doi.org/10.1371/journal.pgen.1000471

Muller H., 1950 Our load of mutations. Am. J. Hum. Genet. 2: 111–176. https://doi.org/10.1007/BF00139458

Murphy D., E. Elyashiv, G. Amster, and G. Sella, 2021 Broad-scale variation in human genetic diversity levels is predicted by purifying selection on coding and non-coding elements. bioRxiv 2021.07.02.450762. https://doi.org/10.1101/2021.07.02.450762

Neher R. A., 2013 Genetic draft, selective interference, and population genetics of rapid adaptation. Annu. Rev. Ecol. Evol. Syst 44: 195–215. https://doi.org/10.1146/annurev-ecolsys-110512-135920

Nei M., 1971 Fertility Excess Necessary for Gene Substitution in Regulated Populations. Genetics 68: 169–184.

Nordborg M., B. Charlesworth, and D. Charlesworth, 1996 The effect of recombination on background selection. Genet. Res. (Camb). 67: 159–174.

Ober C., T. Hyslop, and W. W. Hauck, 1999 Inbreeding effects on fertility in humans: Evidence for reproductive compensation. Am. J. Hum. Genet. 64: 225–231. https://doi.org/10.1086/302198

Ohta T., 1973 Slightly Deleterious Mutant Substitutions in Evolution. Nature 246: 96–98.

Pardo-Manuel De Villena F., and C. Sapienza, 2001 Recombination is proportional to the number of chromosome arms in mammals. Mamm. Genome 12: 318–322. https://doi.org/10.1007/s003350020005

Pennings P. S., S. Kryazhimskiy, and J. Wakeley, 2014 Loss and Recovery of Genetic Diversity in Adapting Populations of HIV. PLoS Genet. 10: e1004000. https://doi.org/10.1371/journal.pgen.1004000

Pratto F., K. Brick, P. Khil, F. Smagulova, G. V. Petukhova, *et al.*, 2018 Erratum Recombination initiation maps of individual human genomes (Science DOI: 10.1126/science.1256442). Science (80-.). 362. https://doi.org/10.1126/science.aav6294

Rajon E., and J. Masel, 2011 Evolution of molecular error rates and the consequences for evolvability. Proc. Natl. Acad. Sci. 108: 1082–1087. https://doi.org/10.1073/pnas.1012918108

Remine W. J., 2005 Cost theory and the cost of substitution — a clarification. J. Creat. 19: 113–125.

Remine W., 2006 More Precise Calculations of the Cost of Substitution. Creat. Res. Sci. Q. 43.

Rice W. R., 1998 Requisite mutational load, pathway epistasis, and deterministic mutation accumulation in sexual versus asexual populations. Genetica 102/103: 71–81.

Roach J. C., G. Glusman, A. F. A. Smit, C. D. Huff, P. T. Shannon, *et al.*, 2010 Analysis of Genetic Inheritance in a Family Quartet by Whole Genome Sequencing. Science (80-.). 328: 636–639.

https://doi.org/10.1126/science.1186802.Analysis

Sella G., D. A. Petrov, M. Przeworski, and P. Andolfatto, 2009 Pervasive Natural Selection in the Drosophila Genome? PLoS Genet. 5: e1000495. https://doi.org/10.1371/journal.pgen.1000495

Simons Y. B., M. C. Turchin, J. K. Pritchard, and G. Sella, 2014 The deleterious mutation load is insensitive to recent population history. Nat. Genet. 46: 220–224. https://doi.org/10.1038/ng.2896

Sohail M., O. A. Vakhrusheva, J. H. Sul, S. L. Pulit, L. C. Francioli, *et al.*, 2017 Negative selection in humans and fruit flies involves synergistic epistasis. Science (80-.). 356: 539–542. https://doi.org/10.1126/science.aah5238

Sung W., M. S. Ackerman, S. F. Miller, T. G. Doak, and M. Lynch, 2012 Drift-barrier hypothesis and mutation-rate evolution. Proc. Natl. Acad. Sci. 109: 18488–18492. https://doi.org/10.1073/pnas.1216223109/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1216223109

Takahata N., 1993 Allelic genealogy and human evolution. Mol. Biol. Evol. 10: 2–22. https://doi.org/10.1093/oxfordjournals.molbev.a039995

Teicher A., 2018 Caution, overload: The troubled past of genetic load. Genetics 210: 747–755. https://doi.org/10.1534/genetics.118.301093

Tenesa A., P. Navarro, B. J. Hayes, D. L. Duffy, G. M. Clarke, *et al.*, 2007 Recent human effective population size estimated from linkage disequilibrium. Cold Spring Harb. Lab. Press Hum. 2: 520–526. https://doi.org/10.1101/gr.6023607.8

Thornton K. R., 2014 A c++ template library for efficient forward-time population genetic simulation of large populations. Genetics 198: 157–166. https://doi.org/10.1534/genetics.114.165019

Thornton K. R., 2019 Polygenic adaptation to an environmental shift: Temporal dynamics of variation under Gaussian stabilizing selection and additive effects on a single trait. Genetics 213: 1513–1530.

https://doi.org/10.1101/505750

Torres R., M. G. Stetter, R. D. Hernandez, and J. Ross-Ibarra, 2020 The Temporal Dynamics of Background Selection in Nonequilibrium Populations. Genetics 214: 1019–1030.

Uricchio L. H., D. A. Petrov, and D. Enard, 2019 Exploiting selection at linked sites to infer the rate and strength of adaptation. Nat. Ecol. Evol. 3: 977–984. https://doi.org/10.1038/s41559-019-0890-6

Valen L. Van, 1963 Haldane's Dilemma, Evolutionary Rates, and Heterosis. Am. Nat. 97: 185–190.

Vucetich J., T. Waite, and L. Nunney, 2017 Fluctuating Population Size and the Ratio of Effective to Census Population Size. Evolution (N. Y). 51: 2017–2021.

Wall J. D., and J. K. Pritchard, 2003 Haplotype blocks and linkage disequilibrium in the human genome. Nat. Rev. Genet. 4: 587–597. https://doi.org/10.1038/nrg1123

Wang J., and M. C. Whitlock, 2003 Estimating effective population size and migration rates from genetic samples over space and time. Genetics 163: 429–446. https://doi.org/10.1093/genetics/163.1.429

Wei X., and J. Zhang, 2019 Patterns and Mechanisms of Diminishing Returns from Beneficial Mutations. Mol. Biol. Evol. 36: 1008–1021. https://doi.org/10.1093/molbev/msz035

Whitlock M. C., 2000 Fixation of new alleles and the extinction of small populations: drift load, beneficial alleles, and sexual selection. Evolution (N. Y). 54: 1855–1861. https://doi.org/10.2144/05384CI01

Wright S., 1938 Size of population and breeding structure in relation to evolution. Science (80-.). 87: 430–431.

Xiong K., J. P. Mcentee, D. J. Porfirio, and J. Masel, 2017 Drift barriers to quality control when genes are expressed at different levels. Genetics 205: 397–407. https://doi.org/10.1534/genetics.116.192567

Appendix A: Haldane's cost of selection imposes a mild constraint on adaptation, with a high proportion of deaths in *A. thaliana* being selective

Available at bioRxiv, doi: https://doi.org/10.1101/2021.11.01.466728

INTRODUCTION

During adaptation, new alleles need to be substituted for old alleles across an entire population. This means that all individuals with the old alleles need to die, and individuals with new alleles must produce enough offspring to replenish the population. These requirements put a limit on the speed at which substitutions can happen. If substitutions need to happen rapidly at many sites, this limit could be prohibitive. Haldane (1957) was the first to use this reasoning to propose a rough estimate of the maximum speed at which substitutions could proceed. This speed limit, later known as Haldane's dilemma (Van Valen 1963), motivated the development of Kimura's neutral theory (Kimura 1968). However, the logic behind Haldane's dilemma and the use of Haldane's specific estimate to motivate neutral theory have been challenged on multiple counts (Ewens 1970; Maynard Smith 1968; Felsenstein 1971; Kern and Hahn 2018). In these discussions, conceptually distinct approaches to quantifying the issue (and criticisms of those quantifications) are often described using identical terms, which apart from being confusing, leaves unresolved the critical question: what is the upper limit on the speed of adaptation and does it matter for natural populations?

Haldane made two somewhat different arguments in his seminal 1957 paper, muddying the waters from the beginning. In the first argument, he defined "selective deaths" as deaths which contribute to a change in allele frequency. This dramatic term is somewhat misleading, as selection that acts by reducing fertility is mathematically equivalent to selection that acts through mortality, and so selective deaths can result from losing potential offspring as well as from literal deaths. He defined the 'cost of selection' as the number of selective deaths required for a substitution (i.e. allele fixation). He calculated this cost as the

integral of $sNp(1-p)$ over the course of a sweep during which allele frequency p rises from p_0 to close to 1 (Figure 1A). This includes only changes in p that are due to selection, and not those due to drift. For modest s, most deaths during this time will not be selective deaths with respect to this single locus.

This cost is typically expressed as a multiple of the population size N, assuming a constant population size. Haldane calculated the cost of selection for a single locus where an adaptive allele begins at low frequency p_0. In the haploid asexual case, the number of selective deaths required for full substitution is N times a factor $D = -\ln(p_0)+O(s)$, where s is the difference in fitness between the two alleles. For appropriately small s and p_0 (Haldane suggests $s < \frac{1}{3}$ and $p_0 = 10^{-4}$), the first term will be much larger than the second, meaning that D is nearly independent of the selection coefficient. For other assumptions about ploidy, dominance, and degree of inbreeding, the dependence of D on p_0 is different, but s remains unimportant unless close to 1 (Haldane 1957). Violating the assumption of constant population size does not change the calculation of D, but requires adjusting the value of N.

In a representative case, such as $p_0 = 10^{-4}$ at a diploid autosomal locus with no dominance, $D = 18.4$. Haldane therefore estimated conservatively that 20-30 N selective deaths are likely to be typical for an adaptive substitution, after considering extreme cases where D can be significantly smaller (selection coefficients close to 1) or significantly larger (fully recessive beneficial alleles).

For his second argument, Haldane quantified the relationship between the speed of adaptation and the fitness reduction due to adaptation using the following model. Assume there are x loci undergoing adaptation such that the i^{th} locus reduces population fitness from its post-substitution value by d_i. The fitness of the population is reduced by a factor of $\prod_{i=1}^{x}(1-d_i)$ or about $e^{-\sum_{i=1}^{x} d_i}$. Haldane claimed (incorrectly) that this implies $N\sum d_i$ selective deaths per generation. With $30N$ selective deaths required to complete an adaptive substitution, it would take $\frac{30}{\sum d_i}$ generations per substitution. During this time, mean fitness will be reduced from 1 (the fitness of a population after all substitutions complete) to $e^{-30/n}$,

where *n* is the number of generations a single substitution takes. This means that the faster substitutions happen, the more population fitness is reduced.

This second argument of Haldane's is a load argument. Load is a reduction in fitness in a population relative to a reference optimal genotype (Figure 1B). When the reduction in fitness is measured relative to an ideal genotype, it is referred to as lag load (Maynard Smith 1976). The term lag load is inspired by adaptation to a changing environment, where new mutations are required to keep up (Bertram, Gomez, and Masel 2017). The concept can be extended to a static environment where innovative new adaptive alleles reveal the possibility of an even better optimal genotype. Reasoning with respect to an upper limit on lag load was later used by Kimura and Ohta (1971) to argue that nucleotide substitution rates exceed an upper limit on the speed of adaptation.

Haldane argues *a priori* that species could probably only sustain a lag load of about 10% for any serious length of time, leading him to suggest that the maximum speed of adaptation is likely around one substitution every 300 generations. He seems to have interpreted this load as corresponding directly to differences in absolute fitness, such that a 10% lag load implies that 10% of all deaths (or differences in fecundity) are selective. His critical *a priori* premise is that populations could not withstand more than 10% selective deaths, resulting in this speed limit on adaptation, later termed Haldane's dilemma (Van Valen 1963). Here we will revisit this *a priori* assumption about the tolerable fraction of selective deaths, by directly quantifying this fraction using data on the survival and reproduction of different *Arabidopsis thaliana* genotypes.

The fact that species accumulate amino acid substitutions far faster than Haldane's limit was the original evidence supporting neutral theory (Kimura 1968). Kimura and Ohta plugged in estimates of the actual rate of substitution in mammalian lineages as n in Haldane's equation $L = e^{-30/n}$, which produced an excessively large lag load (1971). Although still a load argument, their argument was subtly different from Haldane's, arguing that a high lag load implies that typical individuals would need to have a biologically implausible fraction of their offspring die (Kimura and Ohta 1971).

Ewens challenged the derivation of Haldane's Dilemma by pointing out that all these calculations improperly base their limits on the lag load, when in populations with many loci undergoing substitution, what matters is fitness relative to the most fit individual present (Ewens 1970). Haldane's calculation only gives the reduction in fitness relative to an ideal genotype. In a population with many substitutions occurring at once, the likelihood that even a single individual has the ideal genotype, i.e. has the superior allele at each and every locus, is vanishingly small (Figure 1B). Ewens points out that the relevant load is that of mean population fitness relative to the fitness of the fittest individual actually present at any given time (Ewens 1979). More recent travelling wave theories have rediscovered the importance of this quantity, which has come to be known as the "lead" (Desai and Fisher 2007).

One approach to quantify the maximum number of adaptive substitutions, based only on the fitness of individuals actually present, is to begin with the variance in fitness during simultaneous adaptive substitutions (Ewens 1970; Kimura 1969). In the case of many substitutions at once, the variance in fitness is approximately s/n, where s is the selective advantage of an adaptive allele and n is the number of generations between fixation events. The fittest genotype likely to be present can be estimated using the statistics of extreme values. For a population of size 10^6, the relationship between s and n is given by $4.9\sqrt{s/n} = 0.1$, where 4.9 is how many standard deviations above the mean the most extreme fitness value is, and 0.1 is the maximum difference in fitness (following Haldane's *a priori* premise). For $s = 0.01$, n is around 20, much less than Haldane's estimate of 300, and the speed limit becomes even faster for lower values of s. Note that modern traveling wave theory instead derives the lead directly from s, N, and the beneficial mutation rate U, and obtains the fitness variance from that (Desai and Fisher 2007), rather than relying on our ability to directly measure fitness and its variance as an input to the calculation.

Ewens' approach follows from a past emphasis in population genetics on the relationship between variance in fitness and the rate at which populations adapt (Fisher 1930; Crow 1958; Ewens 1979). Once previous load arguments are appropriately stated in terms of the lead, rather than in terms of lag load, it becomes clear that their associated speed limits are unlikely to present a serious issue for populations. In

other words, many simultaneous substitutions do not imply an implausibly large lead. Reframing load problems to instead be about variance has similarly been applied to other deleterious mutation load arguments, e.g. as pertaining to an upper limit on the functional fraction of the human genome (Galeota-Sprung *et al.* 2020). However, although a variance-based approach puts to rest the lag load-based argument for a speed limit, it doesn't address Haldane's first line of argument as discussed above: the cost of natural selection in terms of selective deaths.

Maynard Smith made a different argument against speed limits, claiming that the reason Haldane's dilemma is not a problem is pervasive synergistic epistasis. Synergistic epistasis might solve Haldane's dilemma by increasing differences in fitness above those expected from differences in the numbers of beneficial mutations, thereby making selective deaths more likely to count towards multiple substitutions at once (Maynard Smith 1976). A persistent source of confusion has been that in his model of truncation selection, Maynard Smith also made the shift from Haldane's absolute fitness to a more standard population genetic relative fitness, and hence from lag load to lead (Maynard Smith 1976). The fact that Haldane's dilemma did not arise in this model might therefore be due to reasons put forth by Ewens, rather than due to epistasis.

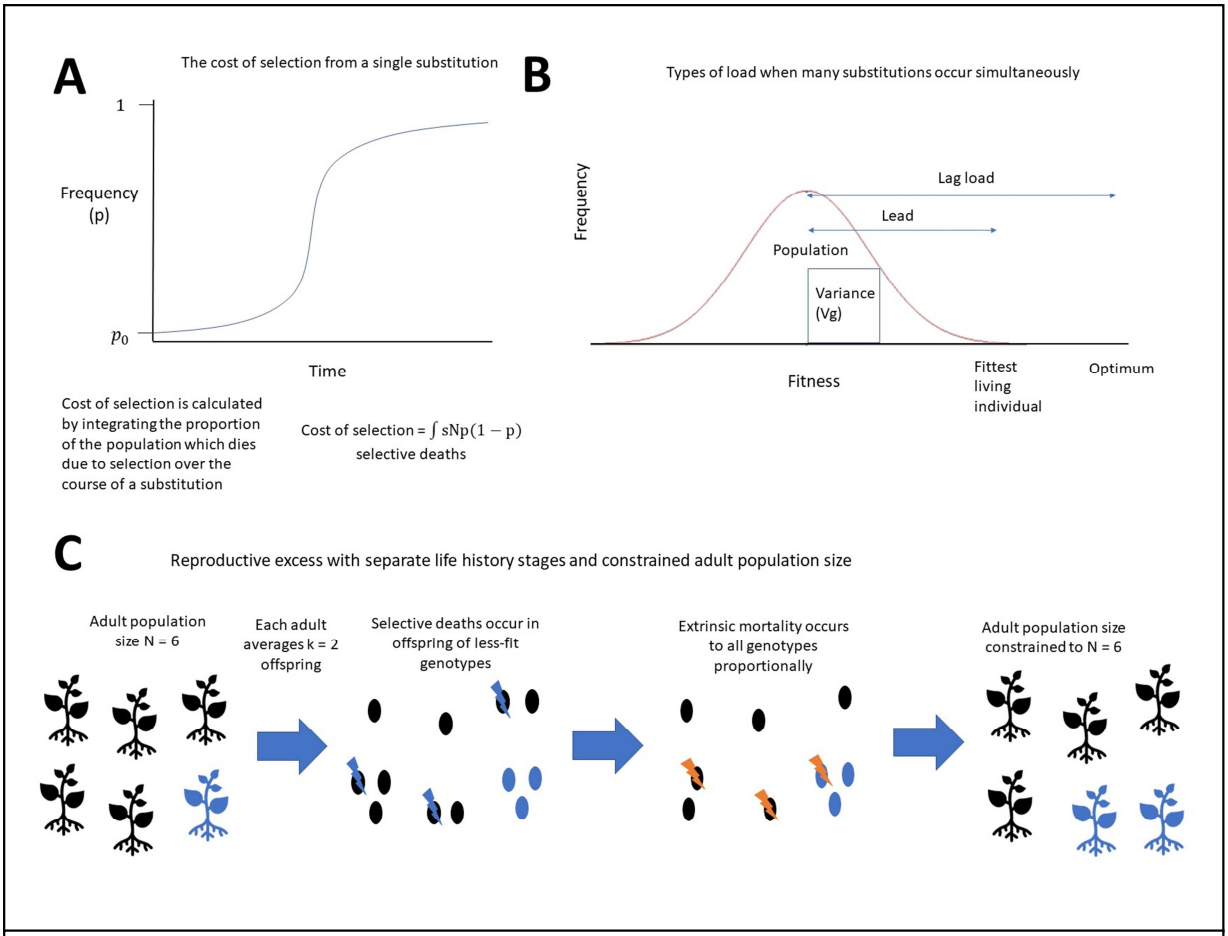

Figure 1. Three different types of arguments have been used to argue for limitations on the speed of adaptation. A) The cost of selection is the multiple of the population size which must die over time to complete a substitution. B) Load arguments calculate the reduction in mean fitness of a population from what it could be. C) With finite reproductive excess, populations have an upper limit on how many selective deaths per generation they can sustain, which sets an upper bound to how fast substitutions can occur.

A further source of confusion is that different papers use the term "substitutional load", which we avoid here, to mean very different things. 'Substitutional load' has been used to refer to what we here call the lag load (Kimura and Ohta 1971), the cost of selection (Kimura 1968), the lead (Maynard Smith 1976), the number of offspring that the most fit genotype must produce (Ewens 1979), the sum of lag load across all generations involved in a substitution (Nei 1971; Kimura 1960), and even more broadly to refer to variance rather than load-based arguments when made in the context of similar questions (Ewens 1970).

This confusion in terminology has obscured the consequences of formulating Haldane's dilemma in different ways.

Selective deaths can occur at any life history stage, which allows the cost of selective deaths to be paid without necessitating a drop in adult population size. The development of Haldane's selective deaths argument therefore requires explicit consideration of reproductive excess (Felsenstein 1971; Nei 1971). This is because the adult population size need not be affected by selective deaths at the juvenile stage, if reproductive excess is sufficient, although this excess is potentially subject to exhaustion during sufficiently rapid adaptation. The Wright-Fisher and related relative fitness models, which dominate population genetics, implicitly assume inexhaustible reproductive excess (Bertram and Masel 2019). However, real populations have a finite reproductive excess, leading to exponential population decline when exhausted. Unfortunately, discussion of the relevance of reproductive excess for Haldane's arguments has been so absent from the literature that a creationist author tried to reintroduce the concept of reproductive excess in 2005, submitting a manuscript to *Theoretical Population Biology* before publishing his discussion of the cost of selection in the *Journal of Creation* and *Creation Research Society Quarterly* (Remine 2005; Remine 2006).

The speed limit implied by the exhaustion of reproductive excess has been calculated for an asexual population with adult population size N (Felsenstein 1971; Nei 1971). Each adult has fertility k, producing exactly k offspring prior to juvenile deaths (Figure 1C). The population undergoes substitutions all with the same selection coefficient s and beginning at the same frequency p_0 (e.g. $1/2N$ for *de novo* mutations in a diploid population, or larger for standing genetic variants which become adaptive due to an environmental change). The raw reproductive excess is defined as $(k-1)N$, with $k > 1$. Some of this reproductive excess is lost to non-selective mortality; in these models, non-selective mortality is not set extrinsically but derived from population size regulation occurring subsequent to selective mortality (Figure 1C). Extrinsic mortality prior to selective mortality can be folded into a lower value of k; the number of selective deaths then available for adaptation is still given by $(k-1)N$,

although this no longer represents raw reproductive excess once k is modified to include extrinsic mortality. The minimum number of generations between substitutions, n, that can be sustained without reproductive excess dropping below 0 (i.e. k dropping below 1), is given by $-\ln(p_0)/\ln(k)$. For Haldane's estimates of $p_0 = 10^{-4}$ and $k = 1.1$, $n = 97$ generations. This limit is more permissive than Haldane's, but much slower than the rosy limits predicted by variance- or lead-based arguments. Importantly, this speed limit calculation is not subject to the same criticisms as Haldane's dilemma, since it doesn't rely on lag load. Reproductive excess directly produces this limit on the rate of substitution, past which the population size will exponentially decrease.

A complication introduced by moving to multiple life history stages is that the integral of selective deaths in Haldane's first argument needs to be interpreted relative to the appropriate denominator, which is the population size at that life history stage (Kimura and Crow 1969). I.e., a population cannot pay the cost through juvenile selective deaths in proportion to the much smaller adult population size.

Felsenstein's and Nei's formulations of Haldane's dilemma rest on estimates of how much reproductive excess is available to a species, while Haldane's cost integration depends on the proportion of deaths which are selective. But it is unclear from these formulations how to quantitatively estimate either reproductive excess or the proportion of deaths that are selective. For example, salmon produce huge numbers of offspring, the vast majority of which fail to become adults, while elephants produce far fewer offspring and devote resources to ensuring their offspring survive to adulthood. This suggests that salmon have a larger raw reproductive excess, but this doesn't guarantee more reproductive excess available to selection, nor does it resolve what fraction of juvenile deaths are selective in either species. No empirical estimates of this fraction exist.

Even within a species, environmental factors might affect both the proportion of selective deaths and the reproductive excess available. Adverse environmental conditions seem likely to result in a higher

proportion of juvenile deaths being selective, resulting in faster adaptation. However, adversity might also reduce raw reproductive excess, which might reduce the reproductive excess available for selection.

To obtain the first empirical estimates of these numbers, we use data from Exposito-Alonso et al. (2019), who counted or estimated every plant grown and seed produced of *Arabidopsis thaliana* cultivars from 517 different genotypes under one season of distinct environmental conditions. This enables us to count both the overall reproductive excess and the proportions of deaths that are selective at each life history stage. This allowed, for the first time, direct application of Haldane's selective death arguments to empirical data.

METHODS

Data source

Exposito-Alonso et al. (2019) used a 2 × 2 × 2 design, with the three treatments being climate, water availability, and adult density, applied over one season to 517 different, fully homozygous plant genotypes. For climate, plants were grown in greenhouses in either Tübingen, Germany, near the center of the species range of *A. thaliana*, or in Madrid, Spain, at the southern edge of the range. Plants were all artificially watered. A high-water treatment matched soil moisture levels near the greenhouse in Germany, and a low-water treatment matched soil moisture levels near the greenhouse in Spain. To generate high adult density, thirty seeds of the same genotype were planted per pot. For low density, several seeds (~10) were planted per pot, enough to ensure that at least one seed would germinate, but few enough that the seeds were unlikely to inhibit each other pre-germination. To avoid any competition between adult plants, only one plant, chosen at random, was retained after germination. We refer to each treatment with a three-letter abbreviation: M or T for Madrid or Tübingen, L or H for low or high water, and I or P for a single individual plant or a population of thirty plants per pot. For example, the treatment with thirty seeds per pot grown in Madrid with high water would be abbreviated as MHP.

Selective deaths across life history stages

The experiment captures three demographic stages: starting seeds, seedlings, adults, and then a new generation of seeds produced by adults. The experiment begins with seeds which germinate into seedlings. In the low-density treatment, where exactly one seedling is retained after germination, we do not have access to data on selective seed deaths vs. this random bottleneck, and we exclude this life history transition from analysis. Any subsequent deaths of seedlings before the end of the experiment are recorded as seedling deaths. In contrast, in the high-density treatment, any of the thirty seeds which fail to survive to the end are counted as deaths, whether due to seed death before germination or to subsequent seedling death, and so we combine selective death calculations across these two life history stages. This means that across the two life history stages at which juvenile plants can die (as seeds before germination and as seedlings), only one set of juvenile deaths is recorded in each density treatment, but they are not comparable. They are combined seed and seedling deaths in the high-density case, and seedling deaths alone in the low-density case. In both density treatments, adult plants produce some amount of seeds before the end of the experiment.

Adult plants which survived to the end of the experiment were recorded as non-selective deaths, representing the fact that *A. thaliana* is an annual plant. To calculate reproductive excess, from the seeds which survived to the end of the experiment, we subtracted the number of seeds of each genotype needed to restart the experiment for a hypothetical second generation (30 seeds per pot per genotype for the high-density conditions, 10 seeds per pot per genotype for the low-density conditions). The reproductive excess was recorded as non-selective deaths, with these subtracted numbers not counted as either selective deaths in the numerator or total deaths in the denominator of the proportion of deaths that were selective.

Our interest is in variation among genotypes, not among replicates. For this reason, we calculate death rates and seed production rates for each genotype-environment combination by averaging across all replicates. This gives appropriate comparisons of a suboptimal genotype to the highest performing genotype, which is required to calculate selective deaths. We wish to calculate selective deaths due to seed or seedling mortality, as well as selective ‘deaths’ due to reduced fecundity (the two are

mathematically equivalent, despite no literal deaths occurring in the second life history stage). Deaths observed during the experiment in the highest performing genotype are scored as non-selective deaths. Similarly, we take the most fecund genotype for each treatment to be the reference point to which to compare genotypes that are less fecund in that environment. Conceptually (ignoring the complication treated in the next section), the selective deaths for a genotype-environment combination are given as follows:

$$\text{Selective deaths in seed survival} = n(d_i - d_{best})$$

$$\text{Selective } '\text{deaths}'\text{in differential seed production} = a(b_{best} - b_i)$$

where n is the number of seeds of that genotype planted, d_i is the genotype's average death rate, d_{best} is the average death rate of the genotype with the lowest death rate in that environment, a is the number of seedlings which survive to produce seeds in that genotype, b_i is the genotype's average seed production rate, and b_{best} is the average seed production rate of the genotype with the highest seed production in that environment. All other deaths, including adult plants and reproductive excess, were considered non-selective. The proportion of deaths that are selective in an environmental condition is then simply calculated as total selective deaths divided by total deaths (selective and non-selective).

Correction for extreme value bias

However, the estimated best fitness is subject to extreme value bias, leading to over-estimation of the number of selective deaths. I.e., the best genotype observed for each treatment is likely not only to be a superior genotype, but also to have overperformed its own long-term average fitness by chance. The more uncertainty in genotype fitness estimates, relative to variance in fitness among genotypes, the worse the extreme value bias problem.

We performed simulations to estimate and correct for extreme value bias in the fitness of the best genotype. Our general approach was to resample a set of 'true' average fitnesses for genotypes, add measurement error, and calculate the magnitude of the bias as the difference between the resulting best

'observed' fitness and the 'true' best fitness. We estimate the 'true' distribution of genotype fitnesses from the observed distributions.

For the proportion of seeds that survive to adulthood, we fit a beta distribution, separately for each of the eight environmental conditions. For each genotype in each of 10,000 simulations per environmental treatment, we then sampled the number of surviving plants from a binomial distribution specified by the 'true' genotype fitness and the number of seeds planted. This simulated number of surviving plants was used to calculate the 'observed' fitness of the genotype. The procedure was similar for fecundity, where we fitted a normal and a lognormal distribution, and used whichever of the two distributions was the better fit for each environmental condition, determined by visual comparison. We then sampled numbers of seeds produced by each genotype from a Poisson distribution with mean equal to the genotype's expected fecundity and used this number of seeds to calculate 'observed' fitness.

Most genotypes had five to seven replicates, but some had as few as one replicate (see the Data Availability section for a link to the GitHub page which has a data file with this information). In our simulated datasets, we assigned each genotype the same number of replicates as the genotype with that index number in the experimental data. In doing so, we assumed that the number of replicates in the experimental data was independent of the genotype's fitness.

In each simulation, we recorded the difference between the best 'observed' fitness and the 'true' fitness for that genotype, then averaged these differences across the 10,000 simulations for each treatment to obtain estimated bias. We then adjusted our estimate of the number of selective deaths so that $\text{Total adjusted selective deaths} = \sum_{i=1}^{517} n_i(d_i - (d_{best} + bias))$, where $bias$ is the degree to which the death rate of the best genotype is expected to be underestimated. Adjustments to selective deaths of seeds or seedlings are shown in Table 1. Table 2 shows corresponding adjustments in the case of fecundity.

Note that we have double counted sampling error, by estimating true variance among genotypes from empirical data subject to sampling errors, and then adding sampling error back in again on top. However,

because the sampling variance is typically orders of magnitude smaller than the among-genotype variance, we expect this to have no qualitative impact on our results.

Environmental condition	Among-genotype variance	Average sampling variance	'Observed' maximum survival rate	'True' maximum survival rate	Unadjusted proportion of deaths selective	Adjusted proportion of deaths selective
MLI	0.0618	0.000106	1	0.999	0.664	0.664
MHI	0.0144	Close to zero	1	1	0.0847	0.0847
TLI	0.0337	0.00266	1	0.981	0.564	0.545
THI	0.0139	Close to zero	1	1	0.109	0.109
MLP	0.0113	0.00104	0.876	0.868	0.96	0.953
MHP	0.0433	0.000227	0.979	0.971	0.564	0.556
TLP	0.023	0.000363	0.947	0.941	0.633	0.628
THP	0.0195	0.000784	0.871	0.858	0.391	0.378

Table 1: Adjusting for extreme value bias has little effect on the estimated proportion of selective deaths at the life history stages of survival of seeds or seedlings to adulthood. Among-genotype variance in survival is much larger than sampling variance, averaged across 10,000 simulations. The proportion of deaths that are selective is adjusted by using the difference between observed and true maximum survival rates as our estimate of bias. In some environmental conditions, both the 'true' and 'observed' fittest genotypes nearly always had survival of 1, resulting in effectively no sampling variance or bias in maximum survival rate.

Across all environmental treatments, adjusting the estimate of the fitness of the best genotype for extreme value bias leads to negligible changes in the estimates of the proportion of deaths which are selective. For

fecundity under the MHP treatment, the predicted bias even drops below zero, so we used the unadjusted values for that treatment.

Environmental condition	Among-genotype variance	Sampling variance	Observed maximum fecundity	True maximum fecundity	Unadjusted proportion of deaths selective	Adjusted proportion of deaths selective
MLI	12,356,463	30,450	22,188	22,186	0.8731	0.8731
MHI	12,215,481	29,710	28,975	28,962	0.3834	0.3830
TLI	1,837,291	11,760	8,982	8,971	0.5877	0.5869
THI	4,492,312	18,880	17,390	17,375	0.4215	0.4207
MLP	1,986,315	7,497	12,288	12,286	0.9591	0.9588
MHP	14,460,181	30,950	16,277	16,277	0.9474	N/A
TLP	812,421	7,016	6,394	6,394	0.9591	0.9590
THP	70,690	937	1,757	1,747	0.5853	0.5809

Table 2: Adjusting for extreme value bias has little effect on the estimated proportion of selective 'deaths' at the life history stage of seed production. Among-genotype variance in fecundity is much larger than sampling variance, averaged across 10,000 simulations. The proportion of 'deaths' that are selective is adjusted by using the difference between the observed and true maximum fecundities as our estimate of bias. In one environmental condition, MHP, the bias was predicted to be in the wrong direction, and we do not adjust.

Pairwise genotype comparisons

For every possible pair of genotypes, we repeat the analysis above to estimate selective deaths and the proportion of deaths which are selective, using the better genotype of the pair as the 'best' genotype in the calculation of selective deaths. With only two genotypes, we do not adjust for extreme value bias. Since we have whole-genome information for these inbred lines, we also calculated the total number of SNP differences between every pair of genotypes (Hamming distance, number of allele differences out of 1,353,386 biallelic SNPs) using the calculation engine in PLINK v1.9.

RESULTS

Surprisingly, in most environmental conditions, most deaths are selective in experimental competitions between *A. thaliana* genotypes (Table 3). In some environmental conditions and life history stages, as many as 95% of deaths are selective. Only with high water availability and a single seed per pot do seedling deaths stay near Haldane's 10% estimate, while all other environmental conditions and life history stages have selective deaths far in excess of 10%.

Environmental condition	Raw reproductive excess	Reproductive excess used by selection	Proportion of seedling deaths selective	Proportion of combined seed and seedling deaths that are selective	Proportion of deaths due to reduced fecundity that are selective
MLI	3,882	2578	0.664		0.8731
MHI	18,323	1552	0.0847		0.3830
TLI	4,862	2650	0.545		0.5869
THI	10,926	1191	0.109		0.4207
MLP	315	300		0.953	0.9588

MHP	1,629	906		0.556	0.9474
TLP	290	182		0.628	0.9590
THP	914	345		0.378	0.5809

Table 3: *Arabidopsis thaliana* has high reproductive excess, and high proportions of selective deaths. The proportions of deaths which are selective are adjusted for extreme value bias as shown in Tables 1 and 2. Raw reproductive excess is calculated as $k - 1$ where k is the number of seeds produced per reproductively mature adult. We estimate how much of this raw reproductive excess would be available for selection by calculating how many of these seeds would result in selective deaths, assuming similar proportions of selective deaths observed in each environmental condition. Note that since raw reproductive excess already includes seed production, those proportions are not included in the calculation of reproductive excess available for selection.

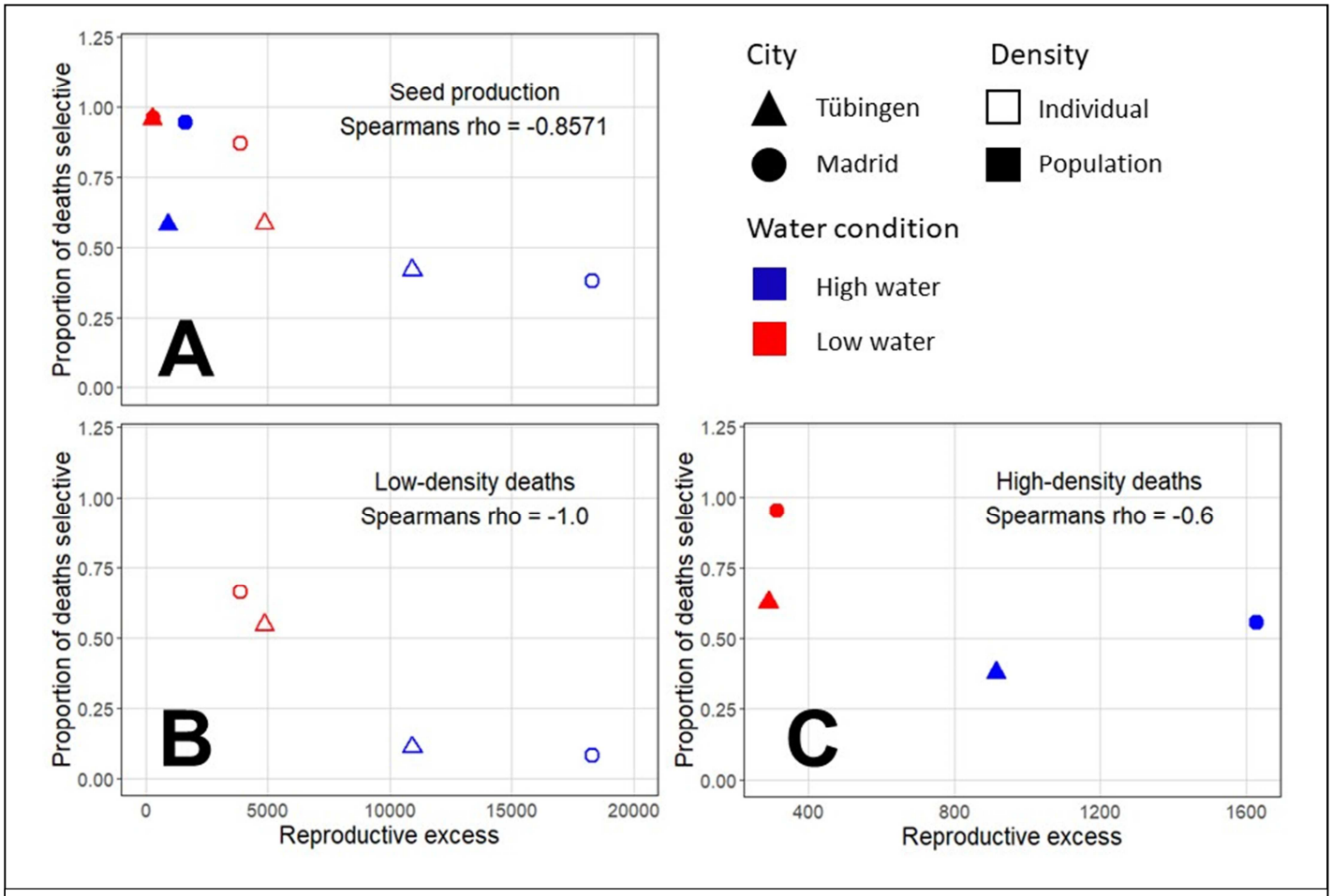

Figure 2. Environmental conditions resulting in lower reproductive excess have higher proportions of deaths that are selective. Low- and high-density treatments are shown on different graphs for seedlings. We performed a Spearman's test for correlation between proportion of deaths selective and raw reproductive excess for each life history stage separately. Fisher's combined probability test on the resulting three p-values showed a significant correlation, with $p = 0.015$.

Comparing environmental treatments allows us to assess the impact of adverse environmental conditions on reproductive excess and the proportion of deaths that are selective. *A priori*, we expect the Madrid climate, low water availability, and thirty plants per pot to inflict harsher and more adverse environmental conditions. We expect adversity to decrease raw reproductive excess, but potentially to increase the proportion of deaths that are selective. While we cannot compare seed deaths at low density to seed plus seedling deaths at high density, we can note that low-density treatments have a smaller proportion of selective deaths associated with fecundity and higher reproductive excess. High-water treatments have a

substantially lower proportion of selective deaths than low-water treatments, and higher reproductive excess. Additionally, reproductive excess and proportions of selective deaths are inversely related across all environmental conditions (Figure 2, p=0.015).

The proportion of selective 'deaths' during fecundity is consistently higher than that of seedling deaths or combined seed and seedling deaths, by 20-30 percentage points in most environmental conditions. This suggests that *A. thaliana* pays more of the cost of selective deaths in lowered fecundity than increased seed or seedling death. The main exception to this pattern occurs in the harshest environmental condition, MLP, where both life history stages have around 95% selective deaths. In this condition, a single genotype achieved perfect survival and a (different) single genotype achieved massive fecundity, with all other genotypes lagging far behind. This condition seems to show selective deaths close to an absolute maximum, which is not incompatible with the conclusion that selective deaths more commonly occur from lowered fecundity where possible.

We quantify extrinsic mortality as the death rate of the highest-fitness genotype, after correcting for extreme value bias (best seen in columns 4 and 5 of Table 1). We expected *a priori* that harsher environmental conditions would have higher extrinsic mortality. However, this wasn't the case. The highest-fitness genotype had perfect survival, or close to it, under most environmental conditions, even after correcting for bias (although we note that the four low-density conditions are uninformative with

respect to seed deaths). Of the conditions with any extrinsic mortality to speak of, the most extrinsic mortality was seen in the THP environmental condition, which is not one of the harsher conditions.

Life history stage	Seedling deaths		Combined seed and seedling deaths		Fecundity	
Environmental condition	Spearman's rho	p-value	Spearman's rho	p-value	Spearman's rho	p-value
MLI	0.096	2.2e-16			0.029	2.2e-16
MHI	0.0004	0.88			0.014	4.0e-7
TLI	-0.0024	0.39			0.041	2.2e-16
THI	-0.0056	0.042			-0.0056	0.041
MLP			0.055	2.2e-16	0.046	2.2e-16
MHP			0.011	4.9e-5	-0.026	2.2e-16
TLP			-0.0081	0.0031	-0.058	2.2e-16
THP			-0.0085	0.002	-0.0029	0.29

Table 4: Genetic distance between genotypes does not correlate with proportion of deaths selective in competitions between genotypes. Visualization of genetic distance versus proportion of deaths selective for each pair of genomes in every environmental condition and life history stage is available in Supplementary Figures 1-3.

One might expect that greater genetic diversity would lead to a higher proportion of deaths being selective. To test this, we repeated our analysis on every pair of genotypes, as though they were the only two genotypes in the experiment. Although some statistically significant correlations were observed in some environmental conditions and life history stages, the direction of correlation was evenly split between negative and positive (as seen in Table 4), and the highest R^2 value observed in any environmental condition was 0.0092, for seedling deaths in the MLI environment, which we deem biologically insignificant.

DISCUSSION

Even in the most forgiving of environmental conditions, most *A. thaliana* deaths in these experiments were selective deaths. The smallest proportion of selective deaths observed among the environmental conditions tested was 8.47%, which falls close to Haldane's 10% guess for the upper limit a population could sustain. In adverse environmental conditions, however, up to 95% of deaths were selective, while producing no shortage of offspring for the next generation. Similarly, the raw reproductive excess and reproductive excess available for selection in every environmental condition are high enough that even extremely rapid adaptation would not jeopardize population viability.

Previous discussion of limitations on the speed of adaptation has been stymied by confusion in the theoretical literature over what the limitations on the rate of adaptation are and why. Disparate usage of the term 'substitutional load' in the literature, as well as the variety of underlying lines of reasoning involved, has made this topic unnecessarily opaque. One aspect of our current work is simply to clarify the variety of lines of reasoning that produce limits on the rate of adaptation. Some of those lines of reasoning have already been resolved, while some, despite going all the way back to Haldane, have not. Our empirical analysis puts to rest lines of reasoning about reproductive excess and selective deaths that were not previously addressed.

Criticisms of Haldane focused on load, and often diverted to alternative approaches such as the emphasis of Fisher (1930) on genetic variance. Interestingly, the concept of relative load was later reinvented as the "lead", with v, the speed of adaptation, derived directly from U, the beneficial mutation rate, N, and s (Desai and Fisher 2007). The reason this solution was not available to Haldane was that population genetics had not yet begun to treat origination processes (McCandlish and Stoltzfus 2014). Instead of treating a steady input of beneficial new mutations, Haldane considered a scenario in which environmental change activates beneficial variants within standing genetic variation. Indeed, their initial frequency $-\ln(p_0)$ is the primary factor in determining the cost (and inversely, maximum speed) of

substitution. In super-sexual populations, the rate of beneficial substitutions is UNs, and in asexual populations the rate is approximately $\frac{2s \ln[Ns]}{\ln^2\left[s/U_b\right]}$, with s as the dominant term (Desai and Fisher 2007).

Despite persistent confusion as to the reason(s), a prevailing historical consensus nevertheless emerged regarding Haldane's Dilemma, namely that there is no real limitation to the speed of adaptation. Haldane's original line of reasoning has been resolved: he used a formulation of load based incorrectly on an absolute rather than relative standard for fitness. Correcting this mistake yields an extremely permissive limit to the rate of adaptation. Similar arguments based on variance in fitness, rather than on differences in fitness (i.e. load), produce equally permissive limits.

However, this consensus overlooks a crucial type of limitation which, as pointed out by Nei 1971 and Felsenstein 1971, depends on the reproductive excess required for substitutions. If we accepted Haldane's 10% guess as an estimate for reproductive excess available for selection (following Nei 1971 and Felsenstein 1971, where $k = 1.1$), then the rate of substitution would be severely limited. This estimate comes from assuming that only a small fraction of the raw reproductive excess in a species is available to be used for selective deaths, thus reducing k well below the raw reproductive excess.

But our results suggest that besides having a great deal of raw reproductive excess, *Arabidopsis* is also able to use almost all available reproductive excess in adverse conditions. There is thus no serious limitation to adaptation from reproductive excess in a fecund species such as *A. thaliana*. The lowest reproductive excess available for selection observed in an environmental condition in this study is hundreds of times larger than 10%, and some environmental conditions have ten times more reproductive excess available for selection.

Of course many species are less fecund than *Arabidopsis*, and Haldane (1957) originally used mammals as example. The part of our results that is perhaps most likely to generalize is that so much of the reproductive excess was available to selection, a much higher proportion than Haldane had guessed. Here

we show for the first time that it is possible to estimate these properties empirically — future comparative work in this area would be of great interest.

Although no environmental condition in this study had too little available reproductive excess to limit how fast *A. thaliana* could adapt, the most adverse environmental conditions had far less raw reproductive excess. Less fecund species than *Arabidopsis* might be unable to sustain a similar, two orders of magnitude reduction in their raw reproductive excess. This means that species with sufficient reproductive excess for rapid adaptation in permissive environments might not be able to adapt quickly in the adverse environments in which adaptation is most necessary. Partially counteracting this is our observation that in adverse environmental conditions, *A. thaliana* used significantly more of its raw reproductive excess for selection.

There are a number of caveats in extrapolating from these controlled experiments to natural populations. Some genotypes in this experiment might have been unrealistically badly adapted to some environmental conditions, inflating our estimates of selective deaths. Controlled environmental conditions might make extrinsic mortality in the best adapted genotypes unnaturally low, also inflating our estimates of raw reproductive excess. Even with these caveats, the fact that such a large proportion of deaths can be selective without coming close to exhausting reproductive excess suggests that *A. thaliana* faces no significant limit to the rate of adaptation.

This experiment combined greater genetic diversity of *Arabidopsis* than would normally be observed in a single population of this asexual species. Our results on this particular experiment could therefore perhaps best be interpreted in terms of the potential for rapid adaptation in the face of substantial migration among populations adapted for different environments, such that environmental change in one location could rapidly lead to the invasion of a genotype already adapted for similar conditions elsewhere. Concerns about the excessive genetic diversity in our experiment are mitigated by our unexpected finding that the genetic distance between genotypes is unrelated to the proportion of deaths which would be selective in a competition between genotypes. This is unexpected because we would have expected that more distantly

related genotypes would experience larger selective differences, and therefore larger proportions of deaths would be selective.

The historical significance of Haldane's arguments about limitations to adaptation is that they were convincingly used to support neutral theory. They were framed as a dilemma because data on the average rate of substitutions seemed to violate estimated speed limits. The development of neutral theory resolved this apparent dilemma by suggesting that most substitutions are neutral and do not count against the speed limit. However, this historical argument is now on troubled ground, because recent literature argues that the fraction of substitutions explained by adaptation can be close to 50% or even higher in some species (Sella *et al.* 2009; Galtier 2016; Uricchio *et al.* 2019), that as much as 37% of allele frequency changes are attributable to adaptation (Buffalo and Coop 2020). For example, recent experiments have shown consistently rapid, pervasive adaptation between seasons in *Drosophila* (Machado et al. 2021; Bertram 2021). There are other resolutions, e.g. these estimates include substitutions of neutral alleles via hitchhiking. It is nevertheless curious that the empirical collapse of historical arguments for neutral theory has not led to a reevaluation of related arguments by Haldane. Here we revise Haldane's arguments for the modern era, finding that Haldane's revised arguments are compatible with empirical evidence for abundant adaptation, while still posing upper limits that might matter in some contexts.

Appendix B: Background selection theory overestimates effective population size for high mutation rates

Available at bioRxiv, doi: https://doi.org/10.1101/2022.01.11.475913

INTRODUCTION

The neutral theory of molecular evolution postulates that i) most genetic diversity observed in natural populations is neutral with respect to an organism's fitness (Kimura 1968; King and Jukes 1969), and ii) dynamics are well-described by models of a single neutral locus in a population of a specified "effective" population size (Ewens 2004; Charlesworth 2009; Masel 2011; Kern and Hahn 2018). This elegant mathematical framework has since been expanded to incorporate migration among populations (Wang and Whitlock 2003), temporal changes in the effective population size (Wright 1938; Vucetich *et al.* 2017), a threshold for neutrality that varies among populations (Ohta 1973), and the effects of selection against deleterious mutations on linked neutral variants (Charlesworth *et al.* 1993). Approximating complex genomic dynamics with single-locus neutral models is extraordinarily powerful, but how accurate is this approximation? Here we focus on the case of background selection at realistically many sites.

The original version of neutral theory modeled one neutral locus in a population of constant size, idealized to obey either Wright-Fisher or Moran dynamics. With more complex demographic histories, an effective population size N_e can be defined as the size of an idealized population that produces the same value of a chosen population statistic (Charlesworth 2009). The statistic usually chosen is genetic diversity (expected heterozygosity) to produce the coalescent effective population size. In an idealized population, genetic diversity depends only on the mutation rate at that locus and the census size of the population (Kimura 1969).

Not long after neutral theory was proposed, it became clear that it was incompatible with some patterns of molecular evolution, in particular the independence of rates of evolution from generation time (Ohta 1973). Neutral theory was therefore replaced by nearly neutral theory. Neutral theory considers mutations that either are strictly neutral or so deleterious that they can be ignored. Nearly neutral theory retains the binary distinction between rapidly purged versus neutral mutations, but allows the ratio of mutations across these two types to vary among species. This ratio is determined by another effective population size, sometimes referred to today as the drift barrier effective population size (Ohta 1973; Sung *et al.* 2012). Nearly neutral theory still derives genome-wide patterns of diversity from models of many independent single loci.

The problem with this binary distinction is that slightly deleterious mutations are purged only slowly from populations. During this removal process, they depress genetic variation at linked sites in the genome, a phenomenon known as background selection (Charlesworth *et al.* 1993). The depression in genetic variation caused by background selection is typically modeled as a decrease in the coalescent effective population size for the neutral loci linked to deleterious variants (Hudson and Kaplan 1995; Lohmueller *et al.* 2011; Comeron 2014). In a population with no recombination, the coalescent effective population size would decrease from N_e to $f_0 N_e$, where f_0 is the equilibrium frequency of individuals with no deleterious mutations, because any neutral variants linked to deleterious variants would be doomed (Charlesworth *et al.* 1993).

Recombination can decouple neutral variants from deleterious variants, resulting in less depression of variation under background selection (Cutter and Payseur 2013). For a single neutral locus linked to a single locus where deleterious mutations with fixed effect size s occur at rate u per diploid individual per generation, and with recombination between the loci occurring at rate r, heterozygosity at the neutral locus is reduced by a factor $F \approx 1 - \frac{us}{2(s+r)^2}$ (see Figure 1A) (Hudson and Kaplan 1994). This result can be straightforwardly extended to any number of deleterious sites linked to the focal neutral site by assuming that mutation and recombination rates are uniform across a genomic window, and that there is

linkage equilibrium among deleterious variants e.g. because multiple significantly linked deleterious mutations are not present at the same time (see Figure 1B) (Hudson and Kaplan 1995; Nordborg *et al.* 1996). In this case, the ratio of observed N_e (based on heterozygosity) to N at a focal neutral site in a genomic window of any size is given by $\frac{N_e}{N} = e^{-U_w/2s+R_w}$, where U_w is now the total diploid deleterious mutation rate across the entire window, and R_w is now the total recombination rate between the ends of the window (Hudson and Kaplan 1995). Since $R_w \gg s$, this can be approximated as $\frac{N_e}{N} = e^{-U_w/R_w}$, producing the result that the effect of background selection on N_e can be approximated with a single factor that depends only on rates of deleterious mutation and recombination. With the same assumptions, similar results can be obtained in the case where both deleterious and beneficial mutations occur (Kim and Stephan 2000).

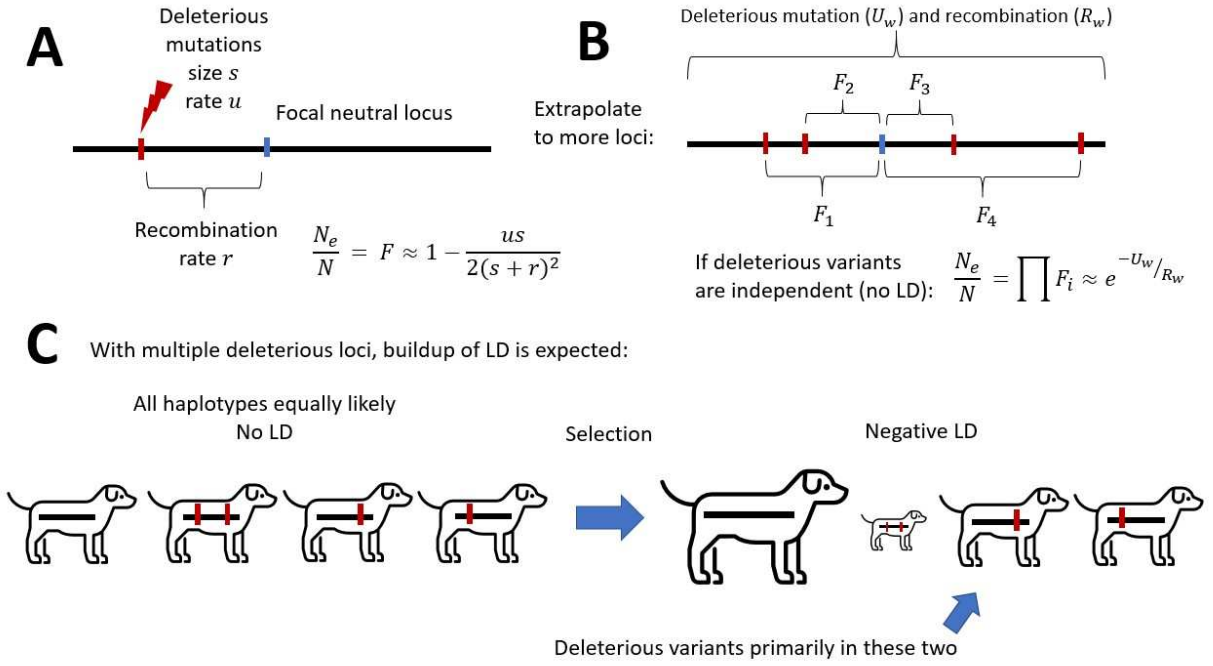

Figure 1. Analytical approximations of the effects of background selection on neutral diversity depend on an unjustified assumption of linkage equilibrium among deleterious variants. (A) The reduction in variation at a focal neutral locus linked to a single deleterious locus can be solved as a function of the deleterious mutation rate, selective effect of deleterious mutations, and recombination rate between the

two loci. (B) This result can be extended to a focal neutral locus in a genomic window with a deleterious mutation rate and recombination rate specified across the whole window rather than at a single site. This requires the assumption of linkage equilibrium among deleterious variants. (C) However, linkage disequilibrium is expected between deleterious variants. Left: for two deleterious loci, we initialize all four haplotypes at linkage equilibrium. Selection will then tend to destroy positive linkage disequilibria by removing haplotypes with both deleterious variants and promoting haplotypes with no deleterious variants. Right: after selection, deleterious variants will therefore be over-represented by haplotypes with one or the other variant but not both, producing net negative linkage disequilibrium.

But an excess of negative linkage disequilibrium is expected among selected mutations (Barton and Otto 2005; Keightley and Otto 2006), violating a key assumption of the Hudson and Kaplan (Hudson and Kaplan 1995) model. Each new mutation is born into either positive or negative linkage disequilibrium with each previously circulating mutation, and these disequilibria tend to be amplified by subsequent random genetic drift when physical linkage keeps recombination low. The two kinds of disequilibria will cancel out across a statistical average when only mutation and drift are considered, but selection quickly eliminates positive disequilibria, either by removing haplotypes with multiple deleterious variants or by fixing haplotypes with multiple beneficial variants (see Figure 1C), leaving an excess of negative disequilibria (Hill and Robertson 1968; Felsenstein 1974). Linkage disequilibrium has been found to enhance the loss of neutral variation in the somewhat different case of selective sweeps (Barton 1998). This suggests that Hudson and Kaplan's approximation might similarly underestimate the effect of background selection on neutral diversity.

Violation of the assumption of linkage equilibrium among deleterious variants might not be a trivial matter given the sheer quantity of deleterious mutations entering populations. For example, the average human is estimated to begin life with an average of two new deleterious mutations not present in either parent (Lesecque *et al.* 2012), and high deleterious mutation rate estimates are not unique to humans (Haag-Liautard *et al.* 2007). Some argue that deleterious mutation rates are even higher, closer to 10 new deleterious mutations per person in humans (Kondrashov 2017).

Here we perform a multi-locus simulation using the fwdpy11 package (Thornton 2014, 2019) that efficiently handles large numbers of non-neutral mutations in relatively large census size populations (Haller and Messer 2017). The recent addition of tree-sequence recording (Kelleher *et al.* 2018) to fwdpy11 additionally allows the calculation of a coalescent effective population size without needing to explicitly simulate and track neutral mutations. This approach allows us to compare the 'observed' coalescent effective population size from our high-deleterious-mutation-rate simulations to the analytic expectations from (Hudson and Kaplan 1995)'s model that assumes linkage equilibrium among deleterious mutations. More broadly, this can inform whether the general approach of single locus models of a neutral mutation are appropriate for a population subject to background selection under realistically high deleterious mutation rates and the resulting linkage disequilibrium among deleterious mutations.

METHODS

Multi-locus simulations

All simulations were written in Python using fwdpy11 (Thornton 2014, 2019). We simulated populations of N diploid individuals undergoing selection against deleterious mutations using a standard Wright-Fisher model for $10N$ generations. Each individual's genome was made up of 23 chromosomes of length 100 under an infinite-sites model (i.e., all floating-point numbers between 0 and 100 on each chromosome are potential loci). Recombination occurs by crossing-over exactly twice per chromosome, matching data for humans (Pardo-Manuel De Villena and Sapienza 2001), although we allow recombination to occur anywhere rather than explicitly simulating a centrosome or recombination hotspots.

Deleterious mutations occur with genome-wide rate U, and have fixed selection coefficients s. In the "no genes" condition, they are located uniformly at random along the chromosomes, while in the "genes" condition they occur only in "genes". We simulate 1,000 genes, accounting for 10 percent of the genome,

interspersed at regular intervals throughout the genome. These parameters were not chosen to be representative of any particular species, but simply to capture the qualitative consequences of clustering among sites subject to background selection.

A recent study of a large sample of modern European humans estimated a gamma distribution of fitness effects of new mutations with mean $sN_e = -224.33$ and $N_e = 23{,}646$, implying a mean s≈ -0.01 (Kim *et al.* 2017). In our main results, we simplify to use a constant s = -0.01 to avoid complications from slightly deleterious mutations with sN_e near 1. We also explore higher and lower values of s, and the complete distribution with the reference mean.

While our forward-time simulations track only deleterious mutations, we compute genetic diversity and hence effective population size by using tree-sequence recording (Kelleher *et al.* 2018) during the simulation, which allows neutral mutations to be projected backwards onto the genealogical histories of different genomic regions. Neutral mutations occur uniformly at random on the entire genome at an arbitrary rate 10^{-4} per genomic 'unit', for a total rate of 0.23 per genome. This low value provides sufficient resolution of N_e at low computational cost. We use msprime (Kelleher *et al.* 2016) to calculate neutral diversity θ on the resulting tree sequence, and then calculate the effective population size for a simulation using $\theta = 4N_e\mu$ and solving for N_e.

We simulated census population sizes *N* ranging from 2000 to 10,000. This is compatible with the range of inferred estimates for human effective population sizes (Takahata 1993; Tenesa *et al.* 2007; McEvoy *et al.* 2011). In all cases, we compared the values of $\frac{N_e}{N}$ calculated from the neutral diversity in the multi-locus simulations to the expected $\frac{N_e}{N}$ ratio given from the two-locus model.

RESULTS

Multi-locus simulations produce a lower N_e than expected from two-locus analytical predictions (Figure 2). The discrepancy becomes marked when the genome-wide mutation rate U is high, specifically when $U > 1$, as is estimated to be the case for humans (Lesecque *et al.* 2012).

This main result is independent of *N* and *s*. A five-fold change in census population size *N* has no significant effect on the $\frac{N_e}{N}$ ratio (Figure 2A). Larger selective effect sizes only slightly increase the degree to which $\frac{N_e}{N}$ drops with *U* (Figure 2B). Simulating a full distribution of effect sizes (Kim *et al.* 2017) instead of a single s_d value for new deleterious mutations has no impact on the effect of *U* on $\frac{N_e}{N}$ (Figure 2C).

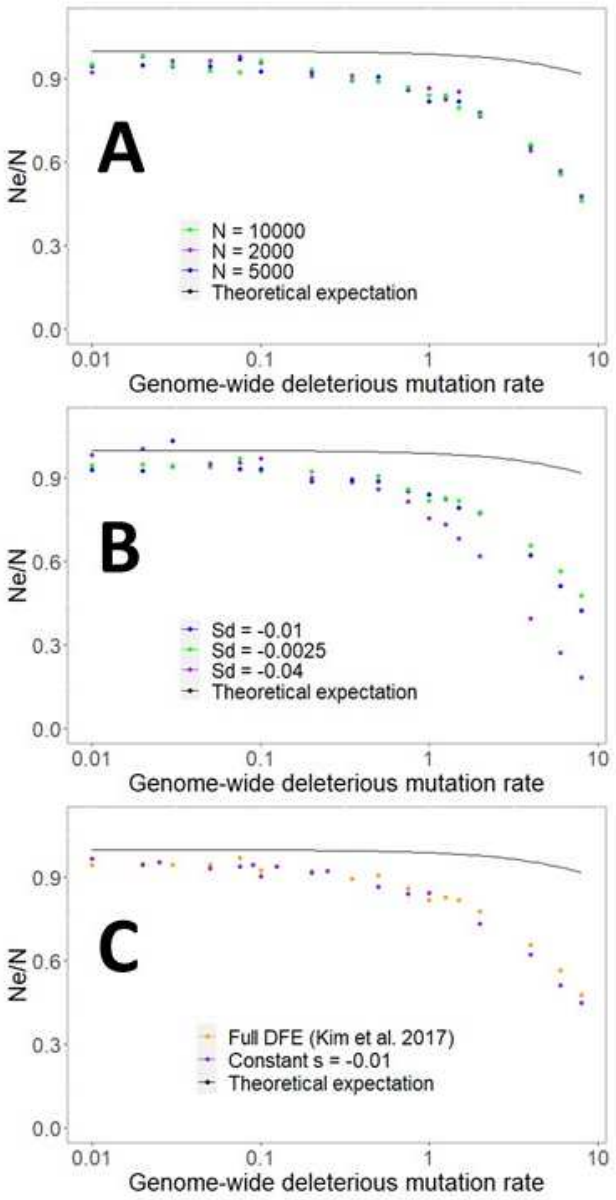

Figure 2. At high deleterious mutation rates, effective population sizes are much lower in multi-locus simulations than in analytic approximations which assume linkage equilibrium. Solid black line is the analytical expectation (Hudson and Kaplan 1995). Where not shown, $s_d = -0.01$ and $N = 5000$. (A) Census population size has no effect on $\frac{N_e}{N}$ ratio across a five-fold change in census population size. (B) Larger selective effects of deleterious mutations results in greater reduction of N_e at high mutation rates. (C) The relationship between deleterious mutation rate and N_e is the same whether the effect size of new deleterious mutations is constant versus drawn from a full distribution of effect sizes (Kim *et al.* 2017) with the same mean value of -0.01. Each point represents a single simulation — we chose to allocate computation to a denser grid of parameter values rather than to replicates of the same parameter values.

The simulations above assume deleterious mutations occur uniformly at random across the genome. A more realistic scenario would be for deleterious mutations to be clustered within a functional subset of the genome. We modify our simulations to model genomes where only 10% of the genome is made up of 'genes' subject to deleterious mutations. Concentrating deleterious mutations into more tightly linked 'genes' has mild and inconsistent effects (Figure 3, red vs. blue).

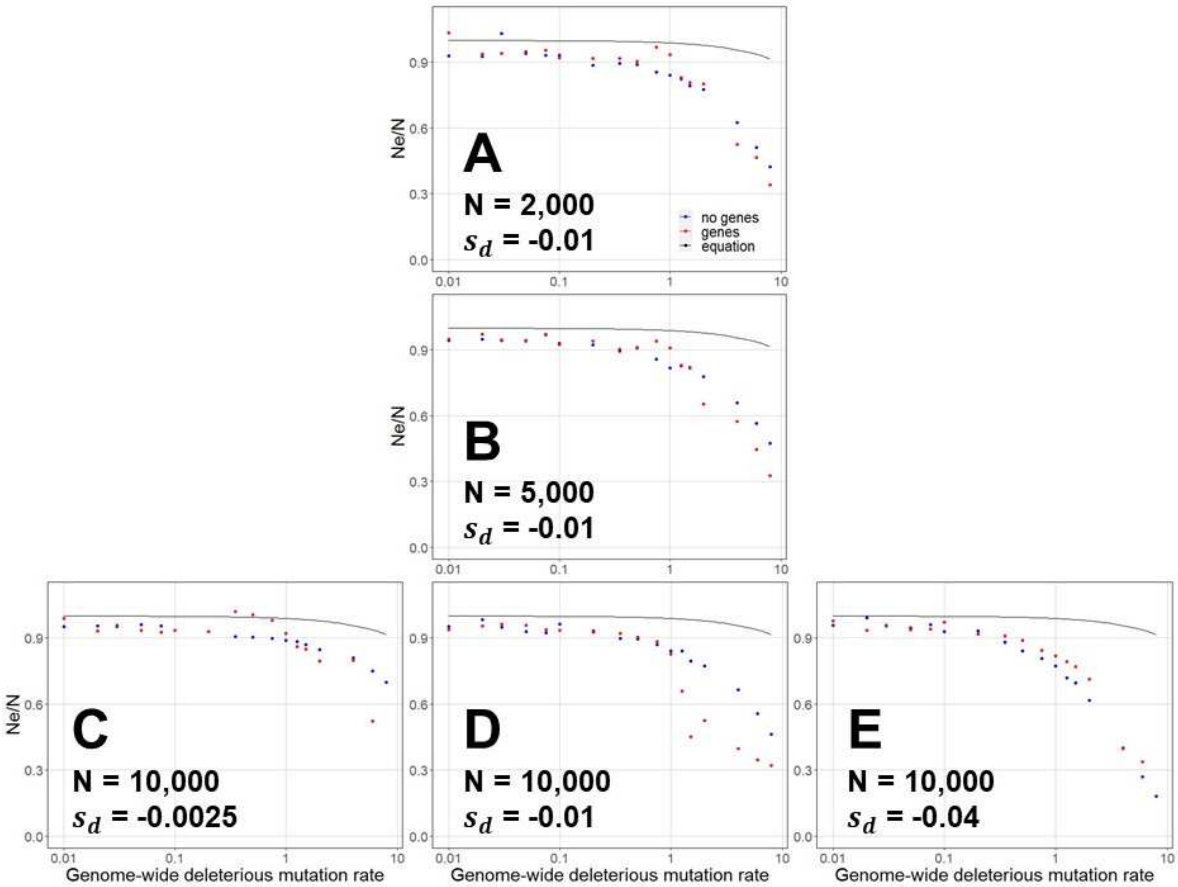

Figure 3. Concentrating deleterious mutations into 'genes' has little effect on background selection. The strongest effect is seen at high population size, moderate strength of selection, and high mutation rate, where background selection clustered in genes reduces neutral diversity more than uniformly distributed background selection does. In all panels, the solid black line is the theoretical expectation from two-locus analytical approximations (Hudson and Kaplan 1995), blue dots are simulations where deleterious mutations occur uniformly at random on the genome, and red dots are simulations where deleterious mutations are clustered into 'genes'. Vertical comparison explores different census population sizes, and horizontal comparison explores different selective effects.

DISCUSSION

When deleterious mutation rates are realistically high, multi-locus simulations of evolution in the presence of background selection produce less neutral diversity than expected from analytic models that treat each deleterious mutation independently. This finding does not depend on the census population size, depends little on the selection coefficient characterizing sites under purifying selection, and depends little on the degree to which such sites are clustered near one another along chromosomes.

The disagreement between simulations and analytic models is substantial only with high deleterious mutation rates producing a significant number of tightly linked deleterious mutations subject to linkage disequilibrium. However, high deleterious mutation rates (Kondrashov and Crow 1993; Awadalla *et al.* 2010; Eöry *et al.* 2010; Lynch 2010a; Roach *et al.* 2010; Kondrashov 2017) and widespread linkage disequilibrium (Conrad *et al.* 2006; Hinds *et al.* 2006; Koch *et al.* 2013) are both well established empirically. The effect of background selection on neutral diversity can thus be presumed to be larger than predicted by approximations that assume linkage equilibrium.

Our multilocus simulations neglect some population features known to affect neutral diversity (e.g. adaptive evolution (Maynard Smith and Haigh 1974) and temporal changes in population size (Torres *et al.* 2020)), and simplify others, (e.g. variation in dominance coefficients among deleterious variants (Gilbert *et al.* 2020) and heterogeneity in recombination rates (Kulathinal *et al.* 2008)). The purpose of our simulations is to isolate the effects of background selection with high mutation rates, rather than to accurately reflect the genetics of specific populations. Our simplified simulations nonetheless produce substantial disagreements with previous analytic methods, posing a serious challenge to the validity of those approximations. Incorporating additional complications into the model has more potential to strengthen rather than to weaken the broader case that simple analytic approximations are insufficient.

This phenomenon may underly some strange results in studies which attempt to fit analytical models of background selection to empirical data. Elyashiv et al (2016) used a maximum likelihood framework to fit a model of background selection based on Hudson and Kaplan (1995) to heterozygosity at synonymous sites (a proxy for neutral variation) in *Drosophila melanogaster*. Input assumptions into this model included i) a map of recombination rates, ii) locations of sites expected to be under purifying selection, and (optionally) iii) locations of substitutions (i.e. possible sweeps) observed in the *D. melanogaster* lineage since the common ancestor with *D. simulans* based on a multiple sequence alignment of three *Drosophila* species. This generated inconsistent results, whereby the deleterious mutation rate across all exons was inferred to be several times larger than the mean deleterious mutation rate across the whole genome. Since these inconsistencies occurred even after adding a model of classic selective sweeps, the authors argue that the data are best explained by partial and soft sweeps. A similar, implausibly high deleterious mutation rate was inferred in humans by McVicker *et al*. (2009), although a more recent study produced consistent deleterious mutation rate estimates albeit still with 75-85% of deleterious mutations under background selection inferred to occur at non-coding sites (Murphy *et al.* 2021). However, correctly accounting for the larger effects of background selection found here, rather than using Hudson and Kaplan (1995), would presumably not require as high a deleterious mutation rate to account for observed data. In general, such effects should be accounted for prior to discarding a model of background selection alone.

Debates about "neutral theory" have in recent years focused on whether patterns of genetic diversity can be explained by a combination of genetic drift, demography, and background selection (Lohmueller *et al.* 2011; Comeron 2014; Jensen *et al.* 2019), or whether these causes are insufficient and observed patterns indicate pervasive adaptation (Sella *et al.* 2009; Kern and Hahn 2018; Uricchio *et al.* 2019; Buffalo and Coop 2020). Our current work does not address this dispute about the relative importance of background selection vs. hitch-hiking. We instead exclude beneficial mutations in order to focus on models of

background selection, extending them into the parameter regime of realistically high genome-wide deleterious mutation rates at which multi-locus effects can become significant.

However, our work does relate to whether it is appropriate to consider models of background selection to be a part of "neutral theory". While this might seem like a strange proposition, given that background selection obviously from its very name involves selection, proponents of "neutral theory" (Jensen *et al.* 2019) now include in their definition of the theory not only the direct effects of slightly deleterious mutations that were first treated by nearly neutral theory, but even the effects of that selection on linked neutral sites.

Behind these odd semantics is a substantial claim that background selection is a straightforward expansion of Kimura's original neutral theory. This claim is based on the argument that Kimura's original neutral models are still useful because the effects of background selection on neutral diversity can be captured by simply modifying the effective population size in a one-locus neutral model. Our results add to the body of evidence that this is not so simple. We find that background selection removes more neutral diversity than expected from previous two-locus models; this larger effect might broaden the scope of phenomena that background selection could explain. But we also find that there is no simple way to derive a value for an effective population size, which is highly sensitive to the genome-wide deleterious mutation rate U.

We can distinguish three criteria influencing judgements of any evolutionary theory, whether it be Kimura's original neutral theory, Ohta's nearly neutral theory, background selection theory, or hitch-hiking theory. First, a theory's predictions must fit empirical patterns of genetic diversity — failure to meet this standard is what led to the replacement of Kimura's neutral theory by Ohta's nearly neutral theory.

Second, those predictions should be grounded in biologically reasonable and parsimonious assumptions. What these are can be subject to dispute, e.g. whether adaptive mutations should be excluded from

"baseline models" (Comeron 2017; Johri *et al.* 2021) vs. whether independent lines of evidence so conclusively support widespread adaptation (Pennings *et al.* 2014; Enard *et al.* 2016) such that adaptive mutations must be considered until sufficient power for their exclusion has been proved. For our own somewhat different purposes, we take the empirical evidence for high deleterious mutation rates and widespread linkage disequilibrium to be broadly accepted, and hence their consideration to be required.

The third criterion is that it is wonderful when predictions can be derived from a simple mathematical framework such as Kimura's one-locus models. However, our results cast serious doubt on the validity of one-locus effective population size approximations of background selection. Instead, we suggest that given high deleterious mutation rates, models not just of hitchhiking but also of background selection need to incorporate less elegant multilocus effects of selection.