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

Joseph Matheson[1], Moises Exposito-Alonso[2], Joanna Masel[1]

[1]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721, USA

[2]Departments of Plant Biology & Global Ecology, Carnegie Institution for Science, Stanford University, Stanford, California 94305, USA

Corresponding author: Joanna Masel

ABSTRACT

Haldane's Dilemma refers to the concern that the need for many "selective deaths" to complete a substitution creates a speed limit to adaptation. However, discussion of this concern has been marked by confusion over which features of substitutions produce limits to the speed of adaptation, what those limits are, and the consequences of violating speed limits. The term 'substitution load' has been particularly unhelpful in this regard. Here we distinguish different lines of reasoning that lead to speed limits, including one line of reasoning which has not yet been fully addressed. We then apply these lines of reasoning to a dataset measuring survival and fecundity of 517 different genotypes of *Arabidopsis thaliana* grown in eight different environmental conditions. We estimate highly permissive limits to the speed of adaptation in all environmental conditions. We also estimate that much higher proportions of deaths contribute to adaptation than were anticipated during historical discussions of speed limits.

Keywords: substitutional load, adaptation rate, genetic load

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 ith 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, Sniegowski, and Ewens 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.

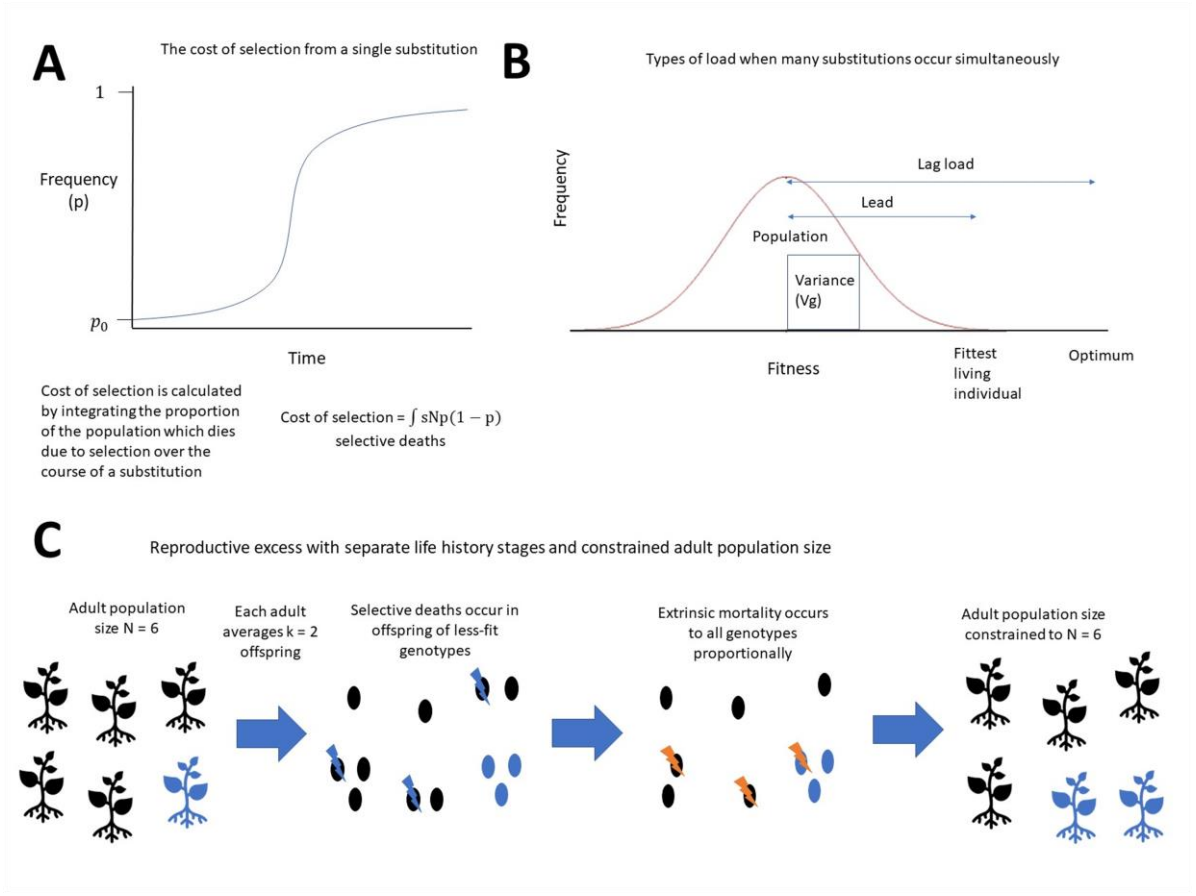

Figure 1. Three different types of arguments have been used to argue for limitations on the speed of adaptation. A) The cost of selection is the multiple of the population size which must die over time to complete a substitution. B) Load arguments calculate the reduction in mean fitness of a population 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 (Nei 1971; Felsenstein 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 (Nei 1971; Felsenstein 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 genotypes, we also calculated the total number of SNP differences between any pair (Hamming distance, number of allele differences out of 1,353,386 biallelic SNPs) using 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

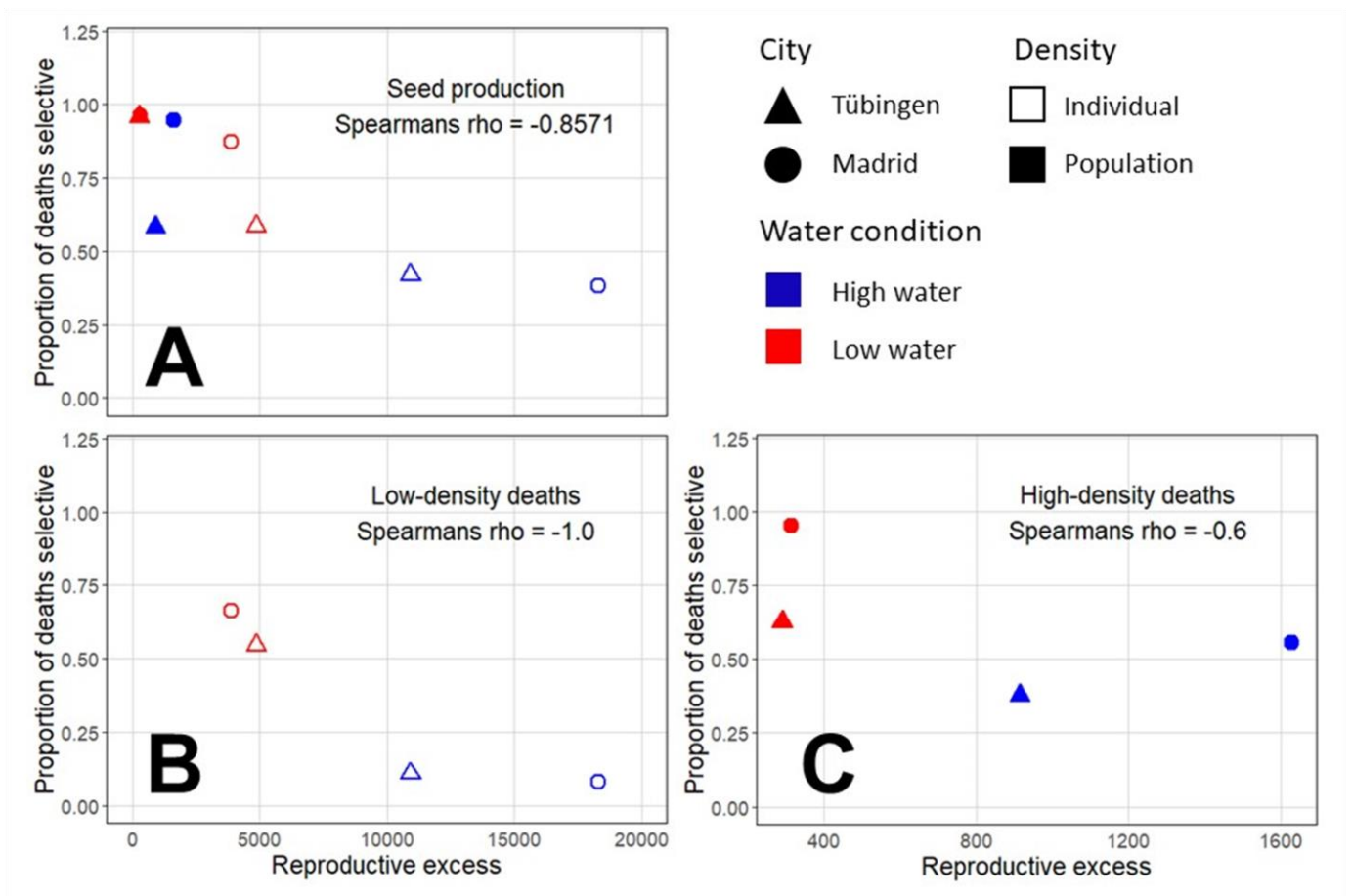

Figure 2. Environmental conditions resulting in lower reproductive excess have higher proportions of deaths that are selective. Low- and high-density treatments are shown on different graphs for seedlings. We performed a Spearman's test for correlation between proportion of deaths selective and raw reproductive 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.

reproductive excess already includes seed production, those proportions are not included in the calculation of reproductive excess available for selection.

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.

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

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.

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 (Galtier 2016; Sella et al. 2009; Uricchio, Petrov, and Enard 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.

DATA AVAILABILITY

All analysis was performed in R, and our code is available on GitHub (www.github.com/josephmatheson/selective_deaths)

ACKNOWLEDGEMENTS

This work was supported by the John Templeton Foundation [62028]. We thank Jason Bertram for helpful discussions.

AUTHOR CONTRIBUTIONS

Conception of the paper by Joanna Masel and Moises Exposito-Alonso. Design and interpretation by Joseph Matheson and Joanna Masel. Data provided by Moises Exposito-Alonso. Data analysis by Joseph Matheson. Manuscript drafted by Joseph Matheson and substantively revised by Joseph Matheson and Joanna Masel.

CONFLICT OF INTEREST

The authors declare no conflicting interests.

BIBLIOGRAPHY

Bertram, J., K. Gomez, and J. Masel. 2017. "Predicting Patterns of Long-Term Adaptation and Extinction with Population Genetics." *Evolution*. https://doi.org/10.1111/evo.13116.

Bertram, Jason. 2021. "Allele Frequency Divergence Reveals Ubiquitous Influence of Positive Selection in Drosophila." *PLOS Genetics* 17 (9): e1009833. https://doi.org/10.1371/journal.pgen.1009833.

Bertram, Jason, and Joanna Masel. 2019. "Density-Dependent Selection and the Limits of Relative Fitness." *Theoretical Population Biology* 129: 81–92. https://doi.org/10.1016/j.tpb.2018.11.006.

Buffalo, Vince, and Graham Coop. 2020. "Estimating the Genome-Wide Contribution of Selection to Temporal Allele Frequency Change." *Proceedings of the National Academy of Sciences of the United States of America* 117 (34): 20672–80. https://doi.org/10.1073/pnas.1919039117.

Crow, James F. 1958. "Some Possibilities for Measuring Selection Intensities in Man." *Human Biology* 30 (1): 1–13.

Desai, Michael M., and Daniel S. Fisher. 2007. "Beneficial Mutation-Selection Balance and the Effect of Linkage on Positive Selection." *Genetics* 176 (3): 1759–98. https://doi.org/10.1534/genetics.106.067678.

Ewens, W J. 1970. "Remarks on the Substitutional Load." *Theoretical Population Biology* 139: 129–39.

Ewens, Warren. 1979. *Mathematical Population Genetics*.

Exposito-alonso, Moises, Oliver Bossdorf, Rasmus Nielsen, and Detlef Weigel. 2019. "Natural Selection on the Arabidopsis Thaliana Genome in Present and Future Climates." *Nature* 573 (7772): 126–29. https://doi.org/10.1038/s41586-019-1520-9.

Felsenstein, Joseph. 1971. "On the Biological Significance of the Cost of Gene Substitution." *American Naturalist* 105 (941): 1–11.

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

Galeota-Sprung, Benjamin, Paul Sniegowski, and Warren Ewens. 2020. "Mutational Load and the Functional Fraction of the Human Genome." *Genome Biology and Evolution* 12 (4): 273–81. https://doi.org/10.1093/gbe/evaa040.

Galtier, Nicolas. 2016. "Adaptive Protein Evolution in Animals and the Effective Population Size Hypothesis." *PLoS Genetics* 12 (1): 1–23. https://doi.org/10.1371/journal.pgen.1005774.

Haldane, JBS. 1957. "The Cost of Natural Selection." *Genetics* 55 (3): 511–24.

Kern, Andrew D, and Matthew W Hahn. 2018. "The Neutral Theory in Light of Natural Selection." *Molecular Biology and Evolution* 35 (6): 1366–71. https://doi.org/10.1093/molbev/msy092.

Kimura, M. 1968. "Evolutionary Rate at the Molecular Level." *Nature* 217: 624–26.

———. 1969. “The Number of Heterozygous Nucleotide Sites Maintained in a Finite Population Due to Steady Flux of Mutations.” *Genetics* 61 (4): 893–903.

Kimura, Motoo. 1960. “Optimum Mutation Rate and Degree of Dominance as Determined by the Principle of Minimum Genetic Load.” *Genetics* 57 (1): 21–34.

Kimura, Motoo, and James F. Crow. 1969. “Natural Selection and Gene Substitution.” *Genetical Research* 13 (2): 127–41. https://doi.org/10.1017/S0016672300002846.

Kimura, Motoo, and Tomoko Ohta. 1971. “On the Rate of Molecular Evolution.” *Journal of Molecular Evolution* 1 (1): 1–17. https://doi.org/10.1007/BF01659390.

Machado, Heather E., Alan O. Bergland, Ryan Taylor, Susanne Tilk, Emily Behrman, Kelly Dyer, Daniel K. Fabian, 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.

Maynard Smith, J. 1968. “‘Haldane’s Dilemma’ and the Rate of Evolution.” *Nature* 219 (5159): 1114–16. https://doi.org/10.1038/2191114a0.

———. 1976. “What Determines the Rate of Evolution?” *American Naturalist* 110 (973): 331–38.

McCandlish, David M., and Arlin Stoltzfus. 2014. “Modeling Evolution Using the Probability of Fixation: History and Implications.” *Quarterly Review of Biology* 89 (3): 225–52. https://doi.org/10.1086/677571.

Nei, Masatoshi. 1971. “Fertility Excess Necessary for Gene Substitution in Regulated Populations.” *Genetics* 68: 169–84.

Remine, Walter. 2006. “More Precise Calculations of the Cost of Substitution.” *Creation Research Science Quarterly* 43 (September).

Remine, Walter J. 2005. “Cost Theory and the Cost of Substitution — a Clarification.” *Journal of Creation* 19 (1): 113–25. http://creation.com/images/pdfs/tj/j19_1/j19_1_113-125.pdf.

Sella, Guy, Dmitri A Petrov, Molly Przeworski, and Peter Andolfatto. 2009. “Pervasive Natural Selection in the Drosophila Genome?” *PLoS Genetics* 5 (6). https://doi.org/10.1371/journal.pgen.1000495.

Uricchio, Lawrence H, Dmitri A Petrov, and David Enard. 2019. “Exploiting Selection at Linked Sites to Infer the Rate and Strength of Adaptation.” *Nature Ecology & Evolution* 3 (6): 977–84. https://doi.org/10.1038/s41559-019-0890-6.

Valen, Leigh Van. 1963. “Haldane’s Dilemma, Evolutionary Rates, and Heterosis.” *American Naturalist* 97 (894): 185–90.