

# Relaxed selection can speed the evolution of complex adaptations

Jeremy Draghi<sup>1,2</sup>, C. Brandon Ogbunugafor<sup>3,4</sup>, Luis Zaman<sup>5,6</sup>, and Todd L. Parsons<sup>7</sup>

<sup>1</sup>Dept. of Biological Sciences, Virginia Tech, Blacksburg, VA USA 24060

<sup>2</sup>Global Change Center, Virginia Tech, Blacksburg, VA USA 24060

<sup>3</sup>Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520

<sup>4</sup>Santa Fe Institute, Santa Fe, NM 87501

<sup>5</sup>Department of Ecology and Evolutionary Biology, University of Michigan 48109

<sup>6</sup>Center for the Study of Complex Systems, University of Michigan 48109

<sup>7</sup>Sorbonne Université, Laboratoire de Probabilités, Statistique et Modélisation (LPSM), CNRS UMR 8001, 75005 Paris, France

July 9, 2024

## Abstract

Natural selection drives adaptive evolution and removes deleterious mutations; these effects are countervailing when a complex adaptation requires mutations that are initially deleterious when they arise, but beneficial in combination. While many models of this dynamic consider how genetic drift or other influences can aid valley crossing by weakening selection, we lack a general, analytical treatment of when relaxed selection might speed this type of adaptation. Here we use simulation and analysis to show that relaxed selection is generally favorable for valley-crossing when adaptive pathways require more than a single deleterious step. We also demonstrate that spatial heterogeneity in selection pressures could, by relaxing selection, allow populations to cross valleys much more rapidly than expected. These results relate to several applications of evolutionary theory to complex systems ranging from host-pathogen evolution to search algorithms in computer science.

# Introduction

Ever since Sewall Wright introduced the fitness landscape in 1931, the metaphor of mountains as genetic regions of high fitness separated by valleys of low fitness has been the canonical framework for understanding how epistatic (i.e., non-additive) interactions between genes impact evolutionary dynamics (1; 2; 3). From the beginning this framework has been applied to the origins of complex traits—those that seemingly require multiple, well-tuned parts to deliver adaptive benefits. Using the fitness landscape metaphor, adaptations in which several individually deleterious mutations improve fitness when combined are visualized by peaks separated by valleys. Quantifying the speed of valley-crossing may shed light on the origins of complex adaptations throughout the history of life (4), but has also been applied to contemporary problems like evolution in somatic cancers (5; 6). These same questions are also essential for predicting evolution in several applications that rely on the assumption that valley-crossing will be difficult: the use of multiple simultaneous treatments against bacterial infections (7; 8) and chronic viral infections like HIV (9), and the use of multiple resistance genes against plant pathogens (10).

Though valley crossing can in principle occur via the sequential fixation of each deleterious mutation, this pathway is a virtual impossibility in large populations when intermediates steps are significantly deleterious (11). Wright’s shifting balance theory avoided this problem by focusing on small populations which were hypothesized to occasionally shift to a different peak with the aid of genetic drift (3). The relevance of this model for evolution has been vigorously debated (see (12)), but more recent theory has shown that large populations can readily cross valleys and do so without substantial drops in mean fitness. Rather than sequential fixation, a lineage carrying a deleterious mutation can acquire subsequent mutations before the deleterious intermediates become very common (11; 5). This process may be slow, on average, but not gradual: distributions of crossing times are approximately exponential (13), reflecting the waiting time for a series of events that is rare but proceeds quickly once initiated.

Selection plays a unique, dual role in crossing adaptive valleys, manifesting as both the constraining and the driving forces. Consequently, it is not at all obvious whether more intense selection would speed or slow the process, and several models suggest that crossing is faster when selection is weakened or opposed. In Wright’s models, genetic drift played this countervailing role. Other candidate mechanisms that enable valley crossing by lessening the influence of natural selection include random phenotypic variation (14; 15), spatial (16; 17) or temporal (18; 19) heterogeneity in the strength of selection, high genetic variation for fitness (20; 21), non-Mendelian transmission biases (22), low density (23), and subdivided populations with local competition (24) or cooperative social interactions (25; 26). But despite these many examples and detailed mathematical analyses of some aspects of valley crossing, we still lack clear criteria for when lessened or relaxed selection speeds the arrival of complex adaptations.

Here we use analytical methods and simulations to show that relaxed selec-

tion can speed valley-crossing when multiple deleterious steps separate peaks. We then examine the effects of one generic cause of relaxed selection in nature—spatial heterogeneity, in which populations spread across multiple environments experience relaxed selection on environment-specific traits. We model this scenario and show that populations faced with multiple, independent valley-crossing problems can surmount all of them in parallel, much more readily than would be extrapolated from the behavior of a specialist population in a single environment.

## Model & Results

### I. Valley crossing and relaxed selection

We define adaptive valleys following the terminology of (27). As shown in Figure 1,  $k$  mutations are required to realize a fitness benefit  $s$  relative to the wild-type. Each mutation individually lowers fitness by  $\delta$  outside of the context of the  $k$ -mutant genotype (Figure 1). For simplicity, we follow (27) in modeling additive effects, such that the fitness  $W_i$  for  $0 < i < k$  is  $1 - \delta i$ , not  $(1 - \delta)^i$ . This choice is an excellent approximation for the multiplicative case if  $\delta$  and  $k$  are both small.

We model asexual reproduction in haploids. Each offspring carries a Poisson-distributed number of mutations with mean  $\mu$ , and the total mutations carried by a genotype is capped at  $k$ . For simplicity, back-mutation is not allowed, and the forward mutation rate does not depend on the number of mutations carried by the parent.

Waiting times for valley-crossing have been analyzed systematically across a number of papers (e.g., (28; 27)). However, almost all such results focus on two-step crossings ( $k = 2$ ) in the case where  $\delta$  is substantial relative to the strength of drift, or instead consider more than two steps but with effectively neutral intermediates (29; 30; 31). A few studies formally consider cases where  $k > 2$  for deleterious intermediates, but have not analyzed the consequences of relaxed selection for crossing time (27; 32; 33). In the supplemental material, we analyze waiting times for arbitrary  $k$  for both the Moran and Wright-Fisher models. Below we sketch our main results and compare the resulting predictions to simulations.

First, as a point of comparison we examine the effects of relaxed selection on  $k = 2$  valleys. For a Moran process, we can approximate the waiting time until a genotype with  $k$  mutations arises as the product of the number of first-step mutations per generation,  $N\mu$ , the expected lineage size created by each such mutation,  $\delta^{-1}$ , and the mutation rate from one to two mutations,  $\mu$ . This calculation assumes that deleterious steps are truly deleterious ( $\delta \gg N^{-1}$ ) and mutations relatively rare ( $\mu \ll 1$ ), ensuring that the first-step mutation remains rare. Once the second mutant arises, its chance to fix is well-approximated as  $\frac{s}{1+s}$  for large  $N$ . Relaxed selection, achieved by multiplying both  $\delta$  and  $s$  by the same constant, is therefore expected to have little effect on the arrival time of the  $k$ -mutant for  $k = 2$  as this constant will largely cancel out of the

expected waiting time when  $s \ll 1 + s$ . Further, the expected waiting time until the  $k$ -mutant has achieved a substantial frequency will only increase with weaker selection. In short, we do not expect relaxed selection to have a large or consistent effect on mean crossing times when  $k = 2$ .

For larger valleys ( $k > 2$ ), these dynamics are qualitatively different. Figure 2 depicts landscapes of the probability of crossing valleys of for  $k = 2, 3$ , and 4. For  $k = 2$ , weakening selection (movement along the diagonal lines toward the lower-left corner) has little effect on this probability, as the contour lines are nearly parallel to this vector of movement. But for higher  $k$ , there are large regions of parameter space in which weakening selection substantially raises the probability of crossing valleys.

We analyzed these dynamics in both the Moran and Wright-Fisher models, deriving approximations for the arrival time of the first mutant that will be ancestral to a successful  $k$ -mutant lineage (see Supplemental Materials §1.2). Under the assumption that the lineages seeded by deleterious intermediates are expected to remain small ( $\sqrt{\mu} \ll \delta$ ) and that the  $k$ -mutant is strongly favored, we can generate accurate, simple formula for the rate at which valleys are crossed. When  $\delta \gg \sqrt{\mu \frac{s}{1+s}}$ , we derive a particularly intuitive expression for the  $k = 3$  waiting time in the Moran model (see Supplemental Materials, Equations S.14–16):

$$t_1 \approx \frac{2\delta^2(1+s)}{N\mu^3s}. \quad (1)$$

This expression can be decomposed into the flux of first-step mutations per generation ( $N\mu$ ), the number of second-step mutations they give rise to ( $\frac{\mu}{\delta}$ ), the number of  $k$ -mutants arising from each of those ( $\frac{\mu}{2\delta}$ ), and the fixation probability of each of those  $k$ -mutants ( $\frac{s}{1+s}$ ). Together, this expression illustrates why relaxed selection speeds valley-crossing for larger valleys: because there are multiple deleterious steps, adaption is accelerated if  $\delta$  is reduced, even if  $s$  is reduced by the same amount. In general, we find a dependency of order  $\delta^{k-1}$ , indicating that the benefits of relaxed selection increase with  $k$  (see Supplemental Materials §1.2.2).

When  $s$  is relatively large compared to  $\mu$ , we expect that the vast majority of the waiting time for a successful  $k$ -mutant will be spent waiting for the first-step ancestor described by Eq. 1. Therefore, we can take  $t_1$  as a good approximation of  $t_k$ , the arrival time of the  $k$ -mutant (see Supplementary Materials §4). Figure 3 plots simulated observations of these arrival times, along with time,  $t_{\text{inv}}$ , at which the  $k$ -mutant has reached a frequency of 0.1. Relaxing selection, modeled as multiplying both  $s$  and  $\delta$  by the fraction indicated on the x-axis, substantially lowers the waiting time. For the Moran model, eq. 1 accurately captures this trend as long as  $\delta \gg \sqrt{\mu \frac{s}{1+s}}$ , visualized as the transition from a solid to a dotted line. When selection is very weak,  $t_1$  and  $t_{\text{inv}}$  separate noticeably, mitigating any further effect of decreasing selection on valley-crossing times. In the Supplemental Materials, we derive similar formula for the Wright-Fisher model (see Supplementary Materials §4). Moreover, we show that relaxed selection is

generally expected to speed the arrival of the  $k$ -mutant as long as  $k > 2$  (see Supplemental Materials §1.3).

## II. Valley crossing and spatial heterogeneity

Relaxed selection can arise when a population is spread over multiple environments, and a trait, or the genes underlying it, are not under selection in some of those environments. To explore how this form of relaxed selection could impact valley crossing, we simulated populations distributed among  $m$  environments. In each environment, we assumed that fitness was determined by a set of loci expressed uniquely in that one environment. Alleles expressed within each environment were subject to a valley-crossing scenario, exactly as described above. For example, with  $m = 5$  and  $k = 3$  a population would be spread across five environments at any one time and faced with five genetically distinct valley-crossing problems, represented by a genome of fifteen total loci. Each valley-crossing dynamic would then be allowed to evolve via separate mutations, with the fitness of an individual being determined by these mutations and their environment.

Within this basic scenario of spatial heterogeneity, competition and selection can operate in several different ways. An individual organism may experience only one environment within their lifespan (coarse-grained heterogeneity) or may visit many environments (fine-grained heterogeneity). In the former case we uniformly assigned each individual at birth to one of the  $m$  environments, eventually filling each environment with  $N/m$  organisms. In the latter, we assigned a fitness to each individual by averaging its fitness across all  $m$  habitats. Separately, selection can occur by global competition (also called hard selection) or can be localized to each environment (soft selection). To simulate the former, we chose  $N$  individuals for the next generation in proportion to assigned fitnesses; to simulate soft selection, we first standardized fitnesses within each environment by dividing them by the mean in that environment. We therefore model four different combinations of these decisions, which represent extremes on spectra bridging coarse- and fine-grained heterogeneity and global and local selection.

Figure 4 illustrates that populations can cross valleys much more readily when spread across multiple environments. While local competition is associated with slightly more facile valley-crossing, the main effect of spatial heterogeneity is largely independent of the details of selection. Note that these probabilities are averaged across each environment, such that the number of valleys crossed is the product of  $p_{\text{inv}}$  and the number of environments. For example, a population spread across ten environments typically crosses more than four valleys; in the same span of time, a population experiencing only a single environment would have about a 10% chance to cross that single valley.

## Discussion

Empirical fitness landscapes motivate the need for quantitative predictions of how populations with specific characteristics, and in particular environments, can navigate their ways to higher fitness (34; 35; 2). Strictly uphill paths are often evident in cases of evolution under strong, focused sources of selection (36; 37). In other cases, valleys may be present but can be skirted by evolution along ridges (1): mutations comprising the complex adaptation can be acquired in sequences of substitutions, or in particular environments, in which each step is individually beneficial, following paths that are lengthy but still entirely uphill (e.g., (38; 39)). More generally, though, fitness landscapes may have properties that make evolutionary optimization a *hard* problem in computational terms (40). In practical applications of valley-crossing theory like the acquisition of multi-drug resistance in HIV (19), we must study the causes of evolutionary events that are rare precisely because we have sought to design treatments that are hard for evolution to undermine. The example of resistance to treatment cocktails in HIV illustrates that understanding valley crossing is particularly important in applied contexts in which we have constructed situations that frustrate simple adaptation. In this study, we look at two general approaches to slowing adaptation—adaptive valleys and relaxed selection caused by spatial heterogeneity. Surprisingly, these two obstacles to adaptive evolution interact antagonistically, partially canceling each other’s constraining effect when certain conditions are met. When adaptive valleys are wider than the minimum size of  $k = 2$ , and organisms experience orthogonal valley-crossing problems in distinct environments, then for a broad range of parameters environmental heterogeneity speeds up valley crossing.

Virus evolution, in particular, has a long history of applications of theoretical population genetics to adaptation on rugged fitness landscapes like those modeled in this study. This is partly because viruses embody the properties (e.g., large population sizes and high mutation rates) that facilitate rapid adaptation to diverse environments (41). More recently, the origin of the SARS-CoV-2 variants has drawn attention to the relevance of rapid evolution across epistatic landscapes (42). For example, the SARS-CoV-2 Omicron (B.1.1.529) variant contains several mutations that are likely to be deleterious individually but beneficial in combination (43). If, as is hypothesized, Omicron originated in chronic infection in an immunocompromised host, this environment likely facilitated the fixation of mutations (corresponding to immune escape and/or increased transmission) that were otherwise too deleterious individually to permit valley crossing in immunocompetent individuals (44; 45); indeed, relaxed selection is explicitly hypothesized as an important driver of Omicron’s emergence (46).

Similar ideas apply to the use of vaccines, which may owe some of their therapeutic longevity to the varied antibodies they induced across the host population, creating a spatially heterogeneous set of selective pressures for pathogens (47). Experimental evidence in prokaryotes demonstrates the importance of a diversity of immune responses (48) and models have analyzed the benefits of

distributing varied vaccines for fighting infections (49). Yet, our results suggest that presenting pathogens with a varied set of evolutionary challenges may not be as effective a constraint on escape as one might hope. In SARS-CoV-2, different components of viral fitness show strong epistasis among mutations that might facilitate escape (50; 46; 51). Moreover, host antibodies targeting one virus epitope have limited effects on binding to other epitopes, suggesting a heterogeneous selective environment for virus mutants (51). Our results suggest that while epistasis, and the resulting fitness valleys, may constrain viral evolution, the fact that diverse virus epitopes are targeted by distinct antibodies across hosts may weaken, rather than strengthen, this constraint.

Complimenting our results, recent work modeling pathogen populations (23) has noted that a type of relaxed selection can arise from low contact rates between infected individuals, allowing deleterious mutants to persist without being immediately out-competed by coinfecting wild-type strains, consequently facilitating valley crossing. Future reconciliations between evolutionary theory and the dynamics of infectious diseases could more clearly identify epidemiological scenarios that promote or prevent the emergence of variants by identifying the population genetic conditions where valley crossing and immune escape (via adaptation on rugged fitness landscapes) is more likely.

While our findings are relevant to real world problems in infectious diseases, the implications are broader. Valley-crossing in evolution is formally very similar to the problem of trapping at local minima in energy landscapes or premature convergence in evolutionary computing. In part, this can be understood as a dilemma of exploration vs exploitation: the greater the strength or stringency of natural selection, the more the evolutionary process is slanted toward exploiting already-discovered high-fitness genotypes, rather than exploring the potential offspring of genotypes with lower fitnesses in the present. In the context of directed evolution of macromolecules, researchers have connected the benefits of relaxed selection to the magnitude of epistasis (52), mirroring the epistasis central to the valley-crossing examples studied here.

The field of evolutionary computation in computer science has long examined analogous problems, where the search for optimal solutions in high-dimensional rugged parameter spaces led researchers to algorithmic implementations of Darwinian evolution. The problems being solved often require multi-objective selection, in which several sub-problems are simultaneously optimized and sophisticated algorithms are used to ensure progress is made along all fronts. Recently, these multi-objective algorithms have been tested for their efficacy in directed evolution experiments using microbial populations and show great promise (53; 54). The balance of adaptation being spread across multiple objectives, allowing each to afford some genetics drift on the other fronts, is a feature of these multiobjective approaches that we argue is contributing to their success. Other concepts rooted in computer science but with interdisciplinary applications include deception (55). Deceptive landscapes contain features that lead search processes away from global optima, as was recently shown to be the case for some antibiotic-resistance landscapes (56). These examples point to the largely unexplored potential for cross-disciplinary progress on best to navigate



rugged landscapes.

Other surprising effects have been investigated using these more computational and engineering approaches, some of which can be attributable to the mechanism we investigate here. For example, Dolson and Ofria observed that evolutionary hotspots—regions of the physical world that produced a disproportionately large amount of novel variation—often occurred in regions of high resource overlap (57). As in our models, such diversity of adaptive opportunities could increase the effects of drift, and thus allow populations to cross more or larger fitness valleys. Recent computational work by Bohm et al. shows how transient relaxed selection, which is driven by the lagging increase in mean fitness of an adapting population, leads to temporary increases in exploration of the fitness landscape (21). Pushing the limit of relaxed selection further, *novelty search* is a new selection algorithm that has gained substantial popularity and success, where rather than rewarding any single (or multiple) objective, individuals are selected based on the uniqueness of their phenotype (58). It is counter-intuitive that jettisoning the evaluation of success actually improves the search for optimal solutions, but our results perhaps offer a simple explanation: relaxed selection allows valley crossing that would otherwise inhibit evolutionary exploration.

Understanding the dynamics of valley-crossing in adaptive landscapes has become a pragmatic, multi-disciplinary problem. Classically, fitness landscapes have helped depict evolution as a set of genotypes competing in a single environment. Our results here help to emphasize the importance of considering environmental variation: when populations are spread across environments, each with its own fitness landscape, our expectations for important dynamics like valley crossing are substantially different. While we have focused here on a simplified picture, without pleiotropy, of how environment-specific landscapes interact. Future work could continue to explore how interrelated landscapes across environments, in which mutations can have pleiotropic effect, shape the diversity of trajectories and likelihood of progress toward higher fitness peaks.

## Acknowledgements

This work was partially supported by NSF (USA) Division of Environmental Biology award 2147101 to JAD and TLP, the NSF Division of Environmental Biology Award Number 2142720 to C.B.O, and the NSF Division of Environmental Biology Award Number 1813069 to L.Z.

## Methods

Individual-based simulations were used to track evolution of distinct genotypes in populations of fixed size, with reproduction occurring with overlapping (Moran) or discrete (Wright-Fisher) generations. In both types of simulations, we tracked the lineages of all genotypes such that we could unambiguously es-



establish the origination time of each mutation in a successful lineage. Matching the assumptions of the analytical model, we assumed that the chance of a mutational change from  $i$  to  $i + 1$  mutations was independent of  $i$ . However, unlike the assumptions of the analysis we allowed for the possibility of multiple mutations in a single reproductive event. Selection affected fecundity in both model. Simulations were coded in R and C.

## Data sharing plan

All code and simulation data will be made available on Data Dryad prior to publication.

## References

- [1] I Fragata, A Blanckaert, MAD Louro, DA Liberles, C Bank, Evolution in the light of fitness landscape theory. *Trends in ecology & evolution* **34**, 69–82 (2019).
- [2] C Bank, Epistasis and adaptation on fitness landscapes. *Annual Review of Ecology, Evolution, and Systematics* **53**, 457–479 (2022).
- [3] S Wright, The shifting balance theory and macroevolution. *Annual review of genetics* **16**, 1–20 (1982).
- [4] CH Martin, JA McGirr, EJ Richards, M St. John, How to investigate the origins of novelty: insights gained from genetic, behavioral, and fitness perspectives. *Integrative Organismal Biology* **1**, obz018 (2019).
- [5] Y Iwasa, F Michor, MA Nowak, Stochastic tunnels in evolutionary dynamics. *Genetics* **166**, 1571–1579 (2004).
- [6] N Vasan, J Baselga, DM Hyman, A view on drug resistance in cancer. *Nature* **575**, 299–309 (2019).
- [7] B Raymond, Five rules for resistance management in the antibiotic apocalypse, a road map for integrated microbial management. *Evolutionary applications* **12**, 1079–1091 (2019).
- [8] M Tyers, GD Wright, Drug combinations: a strategy to extend the life of antibiotics in the 21st century. *Nature Reviews Microbiology* **17**, 141–155 (2019).
- [9] DE Goldberg, RF Siliciano, WR Jacobs, Outwitting evolution: fighting drug-resistant tb, malaria, and hiv. *Cell* **148**, 1271–1283 (2012).
- [10] S Gandon, et al., Building pyramids against the evolutionary emergence of pathogens. *Proceedings of the Royal Society B* **291**, 20231529 (2024).

- [11] DM Weinreich, L Chao, Rapid evolutionary escape by large populations from local fitness peaks is likely in nature. *Evolution* **59**, 1175–1182 (2005).
- [12] JA Coyne, NH Barton, M Turelli, Perspective: a critique of sewall wright’s shifting balance theory of evolution. *Evolution* **51**, 643–671 (1997).
- [13] R Durrett, D Schmidt, Waiting for two mutations: with applications to regulatory sequence evolution and the limits of darwinian evolution. *Genetics* **180**, 1501–1509 (2008).
- [14] M Kirkpatrick, Quantum evolution and punctuated equilibria in continuous genetic characters. *The American Naturalist* **119**, 833–848 (1982).
- [15] SA Frank, Natural selection. ii. developmental variability and evolutionary rate. *Journal of evolutionary biology* **24**, 2310–2320 (2011).
- [16] L Hadany, Adaptive peak shifts in a heterogenous environment. *Theoretical population biology* **63**, 41–51 (2003).
- [17] S Moreno-Gamez, et al., Imperfect drug penetration leads to spatial monotherapy and rapid evolution of multidrug resistance. *Proceedings of the National Academy of Sciences* **112**, E2874–E2883 (2015).
- [18] MC Whitlock, Founder effects and peak shifts without genetic drift: adaptive peak shifts occur easily when environments fluctuate slightly. *Evolution* **51**, 1044–1048 (1997).
- [19] AF Feder, KN Harper, CJ Brumme, PS Pennings, Understanding patterns of hiv multi-drug resistance through models of temporal and spatial drug heterogeneity. *Elife* **10**, e69032 (2021).
- [20] MC Whitlock, Variance-induced peak shifts. *Evolution* **49**, 252–259 (1995).
- [21] C Bohm, VR Ragusa, C Ofria, RE Lenski, C Adami, Reduced selection during sweeps lead to adaptive momentum on rugged landscapes. *bioRxiv* pp. 2024–04 (2024).
- [22] MM Osmond, SP Otto, Fitness-valley crossing with generalized parent–offspring transmission. *Theoretical Population Biology* **105**, 1–16 (2015).
- [23] P Cárdenas, V Corredor, M Santos-Vega, Genomic epidemiological models describe pathogen evolution across fitness valleys. *Science advances* **8**, eabo0173 (2022).
- [24] NL Komarova, L Shahriyari, D Wodarz, Complex role of space in the crossing of fitness valleys by asexual populations. *Journal of The Royal Society Interface* **11**, 20140014 (2014).
- [25] NL Komarova, E Urwin, D Wodarz, Accelerated crossing of fitness valleys through division of labor and cheating in asexual populations. *Scientific reports* **2**, 917 (2012).

- [26] U Obolski, O Lewin-Epstein, E Even-Tov, Y Ram, L Hadany, With a little help from my friends: cooperation can accelerate the rate of adaptive valley crossing. *BMC evolutionary biology* **17**, 1–10 (2017).
- [27] DB Weissman, MM Desai, DS Fisher, MW Feldman, The rate at which asexual populations cross fitness valleys. *Theoretical population biology* **75**, 286–300 (2009).
- [28] Y Iwasa, F Michor, NL Komarova, MA Nowak, Population genetics of tumor suppressor genes. *Journal of theoretical biology* **233**, 15–23 (2005).
- [29] J Schweinsberg, Waiting for m mutations. *Electronic Journal of Probability* **13** (2008).
- [30] R Durrett, D Schmidt, J Schweinsberg, A waiting time problem arising from the study of multi-stage carcinogenesis. (2009).
- [31] DB Saakian, AS Bratus, CK Hu, Crossing fitness canyons by a finite population. *Physical Review E* **95**, 062405 (2017).
- [32] M Lynch, A Abegg, The rate of establishment of complex adaptations. *Molecular biology and evolution* **27**, 1404–1414 (2010).
- [33] SR Proulx, The rate of multi-step evolution in moran and wright–fisher populations. *Theoretical population biology* **80**, 197–207 (2011).
- [34] JAG De Visser, J Krug, Empirical fitness landscapes and the predictability of evolution. *Nature Reviews Genetics* **15**, 480–490 (2014).
- [35] CB Ogbunugafor, CS Wylie, I Diakite, DM Weinreich, DL Hartl, Adaptive landscape by environment interactions dictate evolutionary dynamics in models of drug resistance. *PLoS computational biology* **12**, e1004710 (2016).
- [36] DM Weinreich, NF Delaney, MA DePristo, DL Hartl, Darwinian evolution can follow only very few mutational paths to fitter proteins. *science* **312**, 111–114 (2006).
- [37] HH Chou, HC Chiu, NF Delaney, D Segrè, CJ Marx, Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* **332**, 1190–1192 (2011).
- [38] EM Quandt, et al., Fine-tuning citrate synthase flux potentiates and refines metabolic innovation in the lenski evolution experiment. *Elife* **4**, e09696 (2015).
- [39] B Szappanos, et al., Adaptive evolution of complex innovations through stepwise metabolic niche expansion. *Nature communications* **7**, 11607 (2016).
- [40] A Kaznatcheev, Computational complexity as an ultimate constraint on evolution. *Genetics* **212**, 245–265 (2019).

- [41] SF Elena, R Sanjuán, Virus evolution: insights from an experimental approach. *Annu. Rev. Ecol. Evol. Syst.* **38**, 27–52 (2007).
- [42] BF Nielsen, et al., Host heterogeneity and epistasis explain punctuated evolution of sars-cov-2. *PLoS computational biology* **19**, e1010896 (2023).
- [43] DP Martin, et al., Selection analysis identifies clusters of unusual mutational changes in omicron lineage ba. 1 that likely impact spike function. *Molecular Biology and Evolution* **39**, msac061 (2022).
- [44] C Chaguza, et al., Accelerated sars-cov-2 intrahost evolution leading to distinct genotypes during chronic infection. *Cell Reports Medicine* **4** (2023).
- [45] CA Smith, B Ashby, Antigenic evolution of sars-cov-2 in immunocompromised hosts. *Evolution, Medicine, and Public Health* **11**, 90–100 (2023).
- [46] A Moulana, et al., Compensatory epistasis maintains ace2 affinity in sars-cov-2 omicron ba. 1. *Nature Communications* **13**, 7011 (2022).
- [47] DA Kennedy, AF Read, Why does drug resistance readily evolve but vaccine resistance does not? *Proceedings of the Royal Society B: Biological Sciences* **284**, 20162562 (2017).
- [48] S van Houte, et al., The diversity-generating benefits of a prokaryotic adaptive immune system. *Nature* **532**, 385–388 (2016).
- [49] DV McLeod, LM Wahl, N Mideo, Mosaic vaccination: How distributing different vaccines across a population could improve epidemic control. *Evolution Letters* **5**, 458–471 (2021).
- [50] L Witte, et al., Epistasis lowers the genetic barrier to sars-cov-2 neutralizing antibody escape. *Nature communications* **14**, 302 (2023).
- [51] A Moulana, et al., The landscape of antibody binding affinity in sars-cov-2 omicron ba. 1 evolution. *Elife* **12**, e83442 (2023).
- [52] BA Alpay, MM Desai, Effects of selection stringency on the outcomes of directed evolution. *bioRxiv* pp. 2024–06 (2024).
- [53] A Lalejini, E Dolson, AE Vostinar, L Zaman, Artificial selection methods from evolutionary computing show promise for directed evolution of microbes. *Elife* **11**, e79665 (2022).
- [54] FI Arias-Sanchez, B Vessman, A Haym, G Alberti, S Mitri, Artificial selection optimizes pollutant-degrading bacterial communities. *bioRxiv* pp. 2023–07 (2023).
- [55] MJ Eppstein, CB Ogbunugafor, Quantifying deception: A case study in the evolution of antimicrobial resistance in *Proceedings of the Genetic and Evolutionary Computation Conference 2016*. pp. 101–108 (2016).

- [56] CB Ogbunugafor, MJ Eppstein, Competition along trajectories governs adaptation rates towards antimicrobial resistance. *Nature ecology & evolution* **1**, 0007 (2016).
- [57] E Dolson, C Ofria, Spatial resource heterogeneity creates local hotspots of evolutionary potential in *Artificial Life Conference Proceedings*. (MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info ...), pp. 122–129 (2017).
- [58] J Lehman, KO Stanley, Abandoning objectives: Evolution through the search for novelty alone. *Evolutionary computation* **19**, 189–223 (2011).

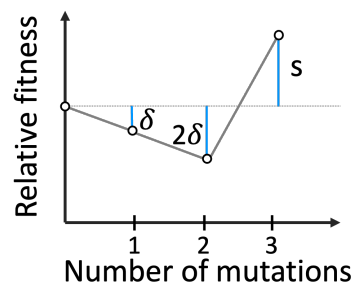


Figure 1: Illustration of the basic fitness landscape for the  $k = 3$  case.

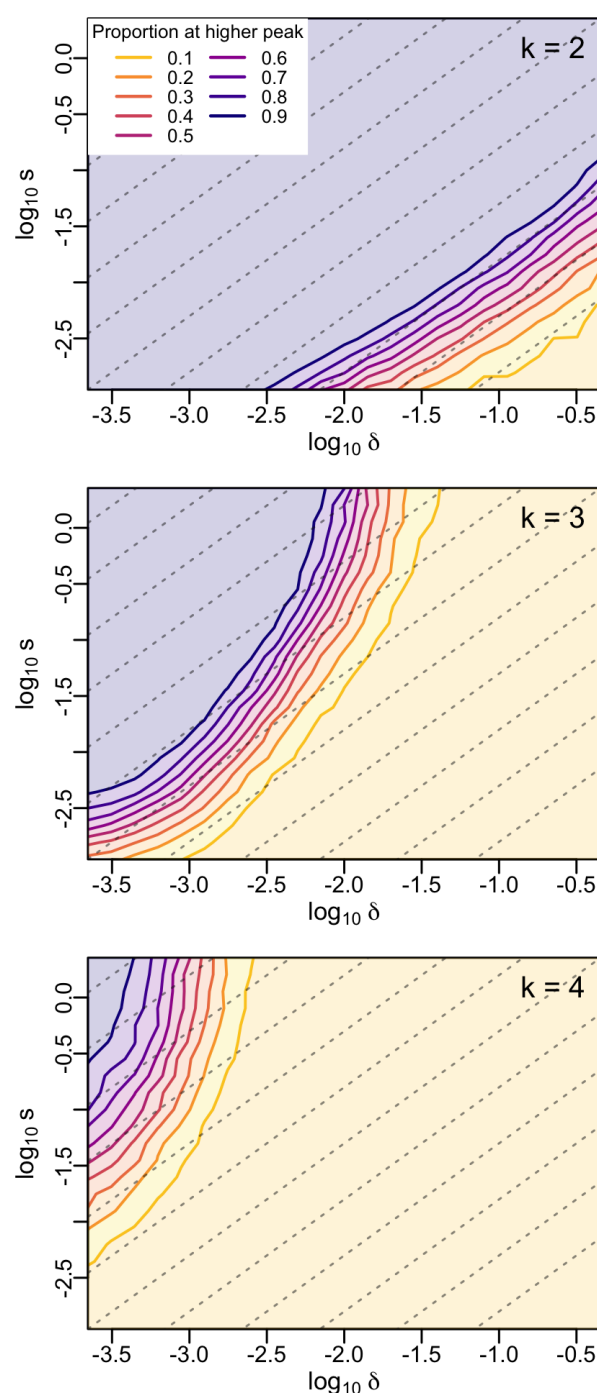


Figure 2: Probability of the  $k$ -mutant to arise and invade (reach a frequency of at least 0.01) within 5000 generations for three values of  $k$ . The Wright-Fisher model was simulated with  $N = 500,000$  and  $\mu = 0.00005$  across more than three orders of magnitude for both  $\delta$  and  $s$ . B. Diagonal lines show proportional change in both  $\delta$  and  $s$ ; movement along these lines from the top-right to the bottom-left corresponds to weakening selection. A minimum of two hundred replicate simulations were performed for each point in a  $21 \times 21$  grid of conditions for each value of  $k$ .



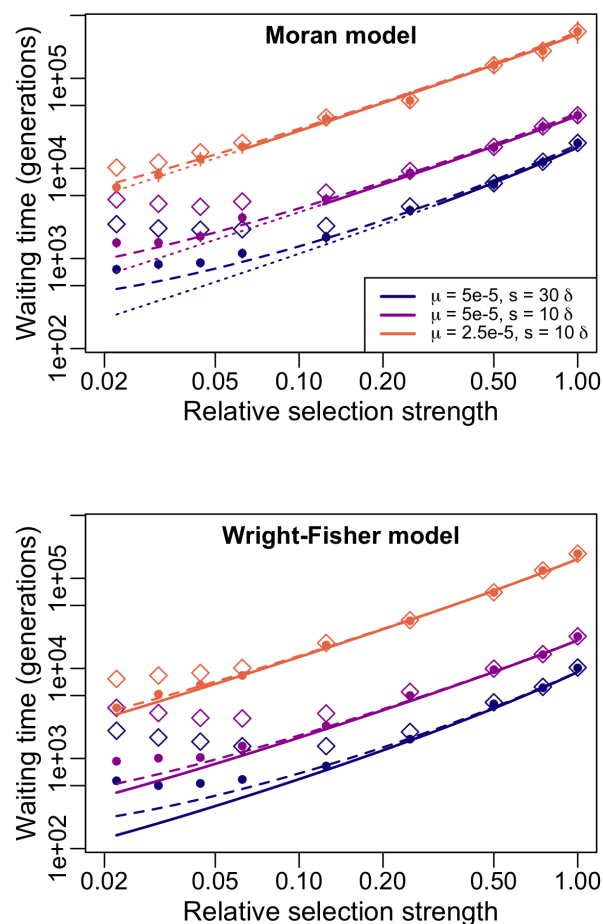


Figure 3: Mean waiting times of valley-crossing for  $k = 3$  in the Moran (*top*) and Wright-Fisher (*bottom*) models. Filled points show the means of  $t_1$  from replicate simulations; open diamonds show the corresponding mean of  $t_{inv}$ . (Top) Solid lines indicate predictions from Eq. 1; the line is dotted when the ratio of  $\delta$  to  $\sqrt{\mu \frac{s}{1+s}}$  is below five, indicating that we expect the approximation to begin to fail. Dashed lines are calculated iteratively from Supplemental Materials Eq. S.5. (Bottom) Solid lines indicate predictions from the approximations given by Eq. S.40; dashed lines show the more complete expression given by iterating Eq. S.34. Confidence intervals, determined by bootstrapping, are visible where they exceed the size of plotting symbols.

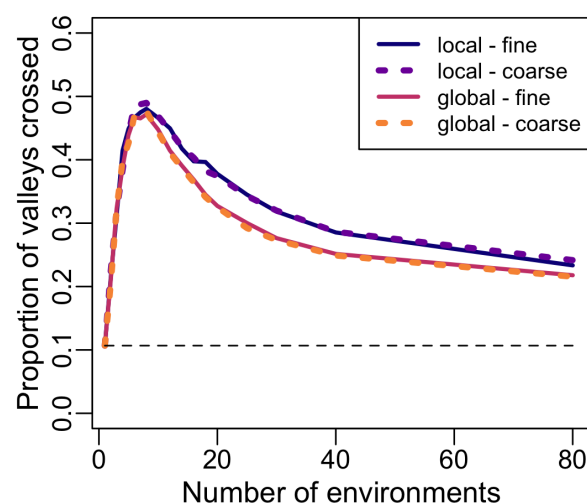


Figure 4: Probability of the  $k = 3$  mutant to arise and invade (reach a frequency of at least 0.01) within 5000 generations, averaged over each environment the population experiences (x-axis). Each population is spread across  $m$  environments, with three loci per environment to create an independent  $k = 3$  valley for each environment. The Wright-Fisher model was simulated with  $N = 504000$  and  $\mu = 0.00005$  (the population size was chosen so that it was divisible by each tested value of  $m$ ). The dashed line shows the expected probability for a single environment, as determined by simulation. At least three hundred replicates are averaged for each treatment.

# Supplementary Material: Relaxed selection can speed the evolution of complex adaptations

Jeremy Draghi, C. Brandon Ogbunugafor, Luis Zaman, Todd L. Parsons

July 9, 2024

In what follows, we obtain estimates for the valley crossing time for a Moran model [13], allowing for a valley of arbitrary width (*i.e.*, an arbitrary number of mutations,  $k$ , in which each is deleterious but the combination of all  $k$  is beneficial. We will focus on a model with **fecundity selection** (*i.e.*, each epistatically negative mutation acquired reduces the individual's birth rate), but proceeding identically, one can obtain a similar result for a model with **mortality selection** (*i.e.*, when epistatically negative mutations increase an individual's death rate).

After first specifying our model (§1), we present an approximation by a birth-and-death process with immigration [10] in §1.1. We then adapt a heuristic argument presented in [17] for the waiting time for  $m$  mutations in a neutral Moran model to our case (§1.2). While we content ourselves with an informal approach, these arguments can be rigorously justified using coupling arguments similar to those in [17, 6]. By considering various regimes, we are able further simplify our expressions (§1.2.1, 1.2.2). Similar results appear in [19], obtained via a different and arguably more complicated argument. Using the analytical expression thus obtained, we show that relaxing selection reduces the expected valley-crossing time (§1.3).

We then illustrate the generality of our approach by applying it to the Wright-Fisher model (§2), obtaining what we believe to be the first estimates for that model. Interestingly, because we are working in a regime when the diffusion approximation is not appropriate (see *e.g.*, [14] for a comparison of the exact Wright-Fisher model, its diffusion approximation, and an approximating branching process), we find that, unlike many quantities of population genetic interest, the expected time to cross a fitness valley is *not* identical in the two models, even after one accounts for the factor of two difference in the effective population sizes for the Moran and Wright-Fisher models [7, §3.7].

Finally, we provide evidence for the two key assumptions underlying our approach, that mutant individuals remain rare until the valley is crossed (§3), and that the time to valley crossing is dominated by the wait until the arrival of the first mutant individual that will be the ancestor of the adapted  $k$ -mutant that will go on to fix (§4).

We start from a general formulation, making simplifications as needed to obtain tractable analytical results. Equation (S.16) specializes to Equation [1] in the Main Text when one takes  $k = 3$  and  $i = 1$ .

## 1 Moran Model

We assume that population grows and changes according to a Moran model in a finite population of size  $N$  [13]. The individuals are grouped into types  $i = 0, \dots, k$ , the type indicating the number of

loci at which the individual carries a mutation, and we let  $X_i(t)$  denote the number of individuals of type  $i$  at time  $t$ .

Selection can be modelled in the Moran framework in a number of asymptotically-equivalent ways; here we assume *fecundity selection*. Individuals of type  $i$  are assumed give birth at rate  $b_i$ . When a birth occurs, the offspring then displaces an individual chosen uniformly at random from the population (which, as formulated below, may include their parent). By assumption, individuals carrying  $0 < i < k$  mutations are less fit than the wild-type,  $b_0 > b_1 > \dots > b_{k-1}$ , whereas an individual carrying  $k$  mutations enjoys a positive epistatic effect, and is more fit  $b_k > b_0$ .

If, in addition, we assume that the offspring of an individual with  $i$  mutations has  $j$  mutations with probability  $\mu_{i,j} \ll 1$  ( $\mu_{j,j} = 0$ ), then type  $i$  individuals are replaced by type  $j$  individuals at total rate

$$b_j X_j \left(1 - \sum_{l \neq i} \mu_{j,l}\right) \frac{X_i}{N} + \sum_l b_l X_l \mu_{l,j} \frac{X_i}{N} = \frac{X_i}{N} \left[ b_j X_j \left(1 - \sum_{l \neq i} \mu_{j,l}\right) + \sum_l b_l X_l \mu_{l,j} \right]$$

Thus, the number of type  $i$  individuals decreases by 1 at rate

$$\sum_{j \neq i} \left[ b_j X_j \left(1 - \sum_{l \neq i} \mu_{j,l}\right) + \sum_l b_l X_l \mu_{l,j} \right] \frac{X_i}{N} = \left[ \sum_{j \neq i} b_j X_j + b_i X_i \sum_{j \neq i} \mu_{i,j} \right] \frac{X_i}{N} \quad (\text{S.1})$$

and increases by 1 at rate

$$\begin{aligned} b_i X_i \left(1 - \sum_l \mu_{i,l}\right) \sum_{j \neq i} \frac{X_j}{N} + \sum_l b_l X_l \mu_{l,i} \sum_{j \neq i} \frac{X_j}{N} \\ = \left[ b_i X_i \left(1 - \sum_l \mu_{i,l}\right) + \sum_l b_l X_l \mu_{l,i} \right] \left(1 - \frac{X_i}{N}\right) \end{aligned} \quad (\text{S.2})$$

## 1.1 Birth-and-Death Process Approximation

By rescaling time, we can always take  $b_0 = 1$ ,  $b_i = 1 - \delta_i$  for  $0 < \delta_1 < \dots < \delta_{k-1}$ , and  $b_k = 1 + s$  for  $s > 0$ . Moreover, suppose the number of  $i$  individuals is negligible compared to  $N$  whereas  $X_0 \sim N$ . This is assured by deleterious selection for individuals  $0 < i < k$  mutations (see §3). In general, this is *not* the case for individuals with  $k$  mutations. However, we will only need the approximation up until the appearance of the first individual with  $k$  mutations that is destined to go on to fixation; up until this individual appears,  $X_k \ll N$  (see §3.3).

Under these assumptions, for  $0 < i$  (S.1) becomes

$$X_i \left[ 1 - \frac{X_i}{N} - \sum_{\substack{0 < j < k \\ j \neq i}} \frac{\delta_j X_j}{N} + \mathbb{1}_{i \neq k} \frac{s X_k}{N} + \left( \mathbb{1}_{i \neq k} (1 - \delta_i) \frac{X_i}{N} + \mathbb{1}_{i=k} (1 + s) \frac{X_k}{N} \right) \sum_{j \neq i} \mu_{i,j} \right] \sim X_i$$

where  $\mathbb{1}_{i \neq k}$  and  $\mathbb{1}_{i=k}$  are indicator functions (*i.e.*,  $\mathbb{1}_{i \neq k}$  is 1 if  $i \neq k$  and 0 if  $i = k$ , the opposite for  $\mathbb{1}_{i=k}$ ). For  $0 < i < k$ , (S.2) becomes

$$\left[ (1 - \delta_i) X_i - \sum_l (1 - \delta_i) \mu_{i,l} X_i + \sum_l \mu_{l,i} X_l - \sum_{l \neq 0} \delta_l \mu_{l,i} X_l \right] \left(1 - \frac{X_i}{N}\right) \sim (1 - \delta_i) X_i + N \mu_{0,i},$$

and, similarly, for  $i = k$ , (S.2) is approximately  $(1 + s)X_k + N\mu_{0,k}$ .

Thus, if  $N\mu_{0,i} \ll 1$ , we may approximate  $X_i$  by a simple birth-and-death process [10] with per-capita birth rate  $1 - \delta_i$  if  $i < 0 < k$ , and  $1 + s$  if  $i = k$ , and per capita death rate 1. If, on the other hand,  $\mu_{0,i}$  is of order  $\mathcal{O}(1)$ , then we need to use a birth-and-death process with immigration [10] with the same birth and death rates and immigration rate  $N\mu_{0,i}$ .

## 1.2 Valley Crossing Rate

We adapt the heuristic presented in [17] to estimate the expected waiting time to cross a fitness valley. We do so by computing the probability that a given type 1 individual (*i.e.*, one carrying the mutant allele at the first locus) is **ancestral** to a successful adapted individual. Here we define **success** by stipulating that the lineage of that individual must be destined to fix; a successful adapted individual has crossed the fitness valley. We will refer to this type 1 individual as **ancestral** or as the **ancestor**. Below, we will argue that the time from the birth of the ancestor to the birth of the selected mutant destined to fix is negligible with respect the wait for the ancestor to appear. Thus, expected valley-crossing time, defined as the wait until the  $k$ -mutant appears, is approximately equal to the waiting time to the birth of this ancestor.

We henceforth assume that the probability of multiple mutations is negligible, so that an offspring born to an individual carrying  $i$  mutations carries  $i + 1$  mutations with probability  $\mu_i$  and  $i$  mutations with probability  $1 - \mu_i$ , and no other mutations are possible (*i.e.*,  $\mu_{i,j} = \mu_i$  if  $j = i + 1$ , and  $N\mu_{i,j} \ll 1$  otherwise). We thus approximate the number of type 1 individuals by a birth-and-death process with immigration, and all other types  $i > 0$  by birth-and-death processes.

As we observed above, prior to the selective sweep that follows the arrival of the mutant destined to fix, the population is almost entirely composed of wild type individuals, and using our approximations above, type 1 individuals are being produced at constant rate

$$N\mu_0$$

*i.e.*, they are being produced at the times of a Poisson process with this rate, and the times between arrivals are exponentially distributed with rate  $N\mu_0$ . Below, we will compute the probability,  $p_1$ , that a given mutant is the ancestor. Then, using the thinning property of the Poisson process, an ancestor is produced at rate  $N\mu_0 p_1$ , the waiting time until the it appears is exponentially distributed with rate  $N\mu_0 p_1$ , and the mean time to the arrival of the ancestor is  $\frac{1}{N\mu_0 p_1}$ .

### 1.2.1 Ancestral Recursion

We compute  $p_1$  via recursion. As we saw above, we can approximate the Moran process by a birth-and-death process with type-dependent per-capita birth and death rates  $b_i$  and  $d_i$ :

$$b_i = 1 - \delta_i \quad \text{and} \quad d_i = 1.$$

Let  $p_i$  be the probability that an individual carrying mutant alleles at  $i$  loci is ancestral to the successful individual, and consider the first event that happens to that individual during their lifetime (either a birth or death). The first event is a birth with probability

$$\frac{b_i}{b_i + d_i}$$

and a death with probability

$$\frac{d_i}{b_i + d_i}.$$

Clearly, if the first event is a death, the individual cannot be ancestral to a selected individual. If the first event is a birth, then the focal individual is ancestral if either it or its offspring is ancestral. Because our simple Markov model has no aging (*i.e.*, in the event that the first event is a birth, the probabilities that the next event is a birth or a death are identical to those for the first event, *etc.*) the parent still has probability  $p_i$  of being ancestral after the first birth. On the other hand, there are two scenarios in which the offspring is ancestral: either it was born without mutation – with probability  $1 - \mu_i$  – and it has  $i$  mutations like its parent, and thus probability  $p_i$  of being ancestral, or, a mutation occurred – with probability  $\mu_i$ , and the offspring carries  $i + 1$  mutant alleles, and thus has probability  $p_{i+1}$  of being ancestral. Combining these, the offspring has probability

$$p'_i = (1 - \mu_i)p_i + \mu_i p_{i+1}, \quad (\text{S.3})$$

of being ancestral. Thus, the probability that at least one of the parent or the offspring is ancestral is

$$1 - (1 - p_i)(1 - p'_i) = p_i + p'_i - p_i p'_i.$$

Combining all of the above, we see that

$$p_i = \frac{b_i}{b_i + d_i} (p_i + p'_i - p_i p'_i) \quad (\text{S.4a})$$

$$= \frac{b_i}{b_i + d_i} [(2 - \mu_i(1 - p_{i+1}))p_i - (1 - \mu_i)p_i^2 + \mu_i p_{i+1}]. \quad (\text{S.4b})$$

Rearranging, we have

$$b_i(1 - \mu_i)p_i^2 + (d_i - b_i + b_i\mu_i(1 - p_{i+1}))p_i - b_i\mu_i p_{i+1} = 0,$$

which may be solved via the quadratic equation to express  $p_i$  as a function of  $p_{i+1}$ :

$$p_i = \frac{\sqrt{(d_i - b_i + b_i\mu_i(1 - p_{i+1}))^2 + 4b_i^2\mu_i(1 - \mu_i)p_{i+1}} - (d_i - b_i + b_i\mu_i(1 - p_{i+1}))}{2b_i(1 - \mu_i)}.$$

In particular, given an initial condition  $p_k$ , we can iterate this to obtain  $p_1$ .

Specializing to the case  $b_i = 1 - \delta_i$  and  $d_i = 1$  gives us a slightly simpler expression,

$$p_i = \frac{\sqrt{(\delta_i + (1 - \delta_i)\mu_i(1 - p_{i+1}))^2 + 4\mu_i(1 - \delta_i)^2(1 - \mu_i)p_{i+1}} - (\delta_i + (1 - \delta_i)\mu_i(1 - p_{i+1}))}{2(1 - \delta_i)(1 - \mu_i)}. \quad (\text{S.5})$$

**p<sub>k</sub>:** Now, by assumption, individuals carrying  $k$  mutations are again adapted: they have birth rate  $b_k = 1 + s$  and death rate  $d_k = 1$ . Then, using Moran's expression for the fixation probability [13], the probability that an individual with  $k$  mutations is ancestral to the whole population is

$$p_k = \frac{1 - \frac{d_k}{b_k}}{1 - \left(\frac{d_k}{b_k}\right)^n} = \frac{1 - \frac{1}{1+s}}{1 - \left(\frac{1}{1+s}\right)^n} \sim \frac{s}{1 + s}. \quad (\text{S.6})$$

**p<sub>k-1</sub>:** Substituting (S.6) into (S.5) yields

$$p_{k-1} = \frac{\sqrt{\left(\delta_{k-1} + \frac{(1-\delta_{k-1})\mu_{k-1}}{1+s}\right)^2 + \frac{4\mu_{k-1}(1-\mu_{k-1})s}{1+s}} - \left(\delta_{k-1} + \frac{(1-\delta_{k-1})\mu_{k-1}}{1+s}\right)}{2(1-\delta_i)(1-\mu_{k-1})}. \quad (\text{S.7})$$

giving us a starting point for the recursion.

Unfortunately, the iteration quickly becomes unwieldy. If, however, we assume in addition that  $\mu_i \ll \delta_i \ll 1$ , then discarding all terms except those of highest order, (S.5) and (S.7) simplify to

$$p_i \sim \frac{\sqrt{\delta_i^2 + 4\mu_i p_{i+1}} - \delta_i}{2},$$

and

$$p_{k-1} \sim \frac{\sqrt{\delta_{k-1}^2 + 4\mu_{k-1} \frac{s}{1+s}} - \delta_{k-1}}{2}.$$

If, moreover,  $\sqrt{\mu_{k-1} \frac{s}{1+s}} \ll \delta_{k-1}$ , we have, using the binomial series  $\sqrt{1+x} = 1 + \frac{1}{2}x + \mathcal{O}(x^2)$ ,

$$p_{k-1} \sim \frac{\delta_{k-1}}{2} \left[ \sqrt{1 + \frac{4\mu_{k-1}}{\delta_{k-1}^2} \frac{s}{1+s}} - 1 \right] \sim \frac{\mu_{k-1}}{\delta_{k-1}} \frac{s}{1+s}.$$

On the other hand, if  $\sqrt{\mu_{k-1} \frac{s}{1+s}} \gg \delta_{k-1}$ , then

$$p_{k-1} \sim \sqrt{\mu_{k-1} \frac{s}{1+s}}.$$

*N.B.*, when  $\sqrt{\mu_{k-1} \frac{s}{1+s}} \sim \delta_{k-1}$ , these last simplifications are not possible.

**p<sub>k-2</sub>:** Now, consider  $p_{k-2}$ :

$$p_{k-2} \sim \begin{cases} \frac{1}{2} \left[ \sqrt{\delta_{k-2}^2 + 4\mu_{k-2} \frac{\mu_{k-1}}{\delta_{k-1}} \frac{s}{1+s}} - \delta_{k-2} \right] & \text{if } \sqrt{\mu_{k-1} \frac{s}{1+s}} \ll \delta_{k-1}, \text{ and} \\ \frac{1}{2} \left[ \sqrt{\delta_{k-2}^2 + 4\mu_{k-2} \sqrt{\mu_{k-1} \frac{s}{1+s}}} - \delta_{k-2} \right] & \text{if } \sqrt{\mu_{k-1} \frac{s}{1+s}} \gg \delta_{k-1}. \end{cases}$$

If we assume that all the  $\delta_i$  have the same order of magnitude:

$$\delta_i \sim \delta_j, \quad (\text{S.8})$$

then  $\sqrt{\mu_{k-1} \frac{s}{1+s}} \ll \delta_{k-1}$  implies that  $\sqrt{\mu_{k-1} \frac{s}{1+s}} \ll \delta_{k-2}$  and

$$\mu_{k-2} \frac{\mu_{k-1}}{\delta_{k-1}} \frac{s}{1+s} \ll \mu_{k-2} \delta_{k-1} \sim \mu_{k-2} \delta_{k-2} \ll \delta_{k-2}^2,$$

so the first case reduces, as above, to

$$p_{k-2} \sim \frac{\mu_{k-2}}{\delta_{k-2}} \frac{\mu_{k-1}}{\delta_{k-1}} \frac{s}{1+s} \quad \text{if } \sqrt{\mu_{k-1} \frac{s}{1+s}} \ll \delta_{k-1}.$$



If, on the other hand,  $\delta_{k-1} \ll \sqrt{\mu_{k-1} \frac{s}{1+s}}$ , we need to consider two possibilities, either

$$\sqrt[4]{\mu_{k-1} \mu_{k-2}^2 \frac{s}{1+s}} \ll \delta_{k-2} \quad (\text{S.9})$$

(*N.B.*, since, by assumption (S.8),  $\delta_{k-2} \sim \delta_{k-1}$ , we also have  $\delta_{k-2} \ll \sqrt{\mu_{k-1} \frac{s}{1+s}}$ ) or,

$$\delta_{k-2} \ll \sqrt[4]{\mu_{k-1} \mu_{k-2}^2 \frac{s}{1+s}}. \quad (\text{S.10})$$

If we assume (S.9), then

$$p_{k-2} \sim \frac{\mu_{k-2}}{\delta_{k-2}} \sqrt{\mu_{k-1} \frac{s}{1+s}}$$

whereas assuming (S.10), we have

$$p_{k-2} \sim \sqrt[4]{\mu_{k-2}^2 \mu_{k-1} \frac{s}{1+s}}.$$

**$p_{k-l}$ ,  $l = 1, \dots, k-1$ :** Proceeding similarly, we find that if for some  $i$ ,  $i = 1, \dots, l+1$

$$\sqrt[2^i]{\prod_{j=1}^i \mu_{k-j}^{2^{j-1}} \frac{s}{1+s}} \ll \delta_{k-l} \ll \sqrt[2^{i-1}]{\prod_{j=1}^{i-1} \mu_{k-j}^{2^{j-1}} \frac{s}{1+s}}, \quad (\text{S.11})$$

(*N.B.*, by our assumption (S.8), if this asymptotic inequality holds for any  $\delta_m$ ,  $m = 1, \dots, k-1$ , then it holds for all of them) then

$$p_{k-l} \sim \prod_{j=i}^l \frac{\mu_{k-j}}{\delta_{k-j}} \sqrt[2^{i-1}]{\prod_{j=1}^{i-1} \mu_{k-j}^{2^{j-1}} \frac{s}{1+s}}, \quad (\text{S.12})$$

where we take  $\prod_{j=i}^l \frac{\mu_{k-j}}{\delta_{k-j}} = 1$  if  $i > l$ , and  $\sqrt[l]{x} = x$ . In particular, the waiting time until the arrival of the first adapted individual has mean

$$t_1 \approx \frac{1}{N \mu_0 p_1} = \left( N \frac{\prod_{j=i}^k \mu_{k-j}}{\prod_{j=i}^{k-1} \delta_{k-j}} \sqrt[2^{i-1}]{\prod_{j=1}^{i-1} \mu_{k-j}^{2^{j-1}} \frac{s}{1+s}} \right)^{-1}. \quad (\text{S.13})$$

### 1.2.2 Additive and Multiplicative Selection

Simplifying to additive ( $\delta_i = \delta i$ ) or multiplicative selection ( $1 + \delta_i = (1 + \delta)^i$ ), so  $\delta_i \sim \delta i$ , since, by assumption  $\delta_i \ll 1$ ), and assuming that all mutation rates are equal ( $\mu_i \equiv \mu$ ), we see that if

$$\mu^{1-2^{-i}} \sqrt[2^i]{\frac{s}{1+s}} \ll \delta \ll \mu^{1-2^{-i+1}} \sqrt[2^{i-1}]{\frac{s}{1+s}}, \quad (\text{S.14})$$

then, taking  $0! = 1$  and, as previously,  $\sqrt[l]{x} = x$ ,

$$p_{k-l} \sim \frac{(k-l-1)!}{(k-i)!} \frac{\mu^{l-i+2-2^{-i+1}}}{\delta^{l-i+1}} \sqrt[2^{i-1}]{\frac{s}{s+1}}, \quad (\text{S.15})$$

and

$$t_1 \sim \frac{1}{N\mu_0 p_1} = \left( N \frac{1}{(k-i)!} \frac{\mu^{k-i+2-2^{i+1}}}{\delta^{k-i}} 2^{i-1} \sqrt{\frac{s}{1+s}} \right)^{-1}. \quad (\text{S.16})$$

Equation [1] in the Main Text is obtained by taking  $k = 3$  and  $i = 1$ .

### 1.3 Relaxing Selection Can Accelerate Adaptation

Now, suppose that all forms of selection are relaxed by a factor of  $x \in (0, 1)$ , *i.e.*, we replace  $\delta_i$  by  $\delta_i x$  and  $s$  by  $sx$  above. This gives a probability of being ancestral that depends on  $x$ :

$$p_1(x) = \prod_{j=i}^{k-1} \frac{\mu_{k-j}}{\delta_{k-j}} 2^{i-1} \sqrt{\prod_{j=1}^{i-1} \mu_{k-j}^{2^{j-1}} s} x^{i-k+2^{1-i}} (1+sx)^{-2^{1-i}}, \quad (\text{S.17})$$

To see the effect on adaptation time, note that *increasing*  $p_1$  has the consequence of *decreasing* the mean adaptation time  $t_1$ , (S.13). Now,

$$\begin{aligned} \frac{\partial p_1}{\partial x} &= \prod_{j=i}^{k-1} \frac{\mu_{k-j}}{\delta_{k-j}} 2^{i-1} \sqrt{\prod_{j=1}^{i-1} \mu_{k-j}^{2^{j-1}} s} \frac{\partial}{\partial x} x^{i-k+2^{1-i}} (1+sx)^{-2^{1-i}} \\ &= \prod_{j=i}^{k-1} \frac{\mu_{k-j}}{\delta_{k-j}} 2^{i-1} \sqrt{\prod_{j=1}^{i-1} \mu_{k-j}^{2^{j-1}} s} \\ &\quad \times \left( (i-k+2^{1-i}) x^{i-k-1+2^{1-i}} (1+sx)^{-2^{1-i}} - 2^{1-i} x^{i-k+2^{1-i}} (1+sx)^{-1-2^{1-i}} \right) \\ &= \prod_{j=i}^{k-1} \frac{\mu_{k-j}}{\delta_{k-j}} 2^{i-1} \sqrt{\prod_{j=1}^{i-1} \mu_{k-j}^{2^{j-1}} s} x^{i-k+2^{1-i}} (1+sx)^{-2^{1-i}} \left( \frac{i-k+2^{1-i}}{x} - 2^{i-1} \frac{1}{1+sx} \right) \\ &= p_1(x) \left( \frac{i-k+2^{1-i}}{x} - 2^{i-1} \frac{1}{1+sx} \right) \end{aligned}$$

which is negative (so increasing with *decreasing*  $x$ ) provided

$$k-i+2^{1-i} \frac{x}{1+sx} \geq 2^{1-i},$$

Now, for  $0 < x < 1$ ,

$$0 < 2^{1-i} \frac{x}{1+sx} < 2^{1-i},$$

so the inequality holds provided  $k \geq i+1$ . Note that  $k \geq i$ , whereas when  $i = k$ , the  $\delta_i$  are in their smallest possible range (*cf.* (S.11)) and

$$p_1 = 2^{k-1} \sqrt{\prod_{j=1}^{k-1} \mu_{k-j}^{2^{j-1}} \frac{s}{1+s}}$$

is independent of the  $\delta_i$ , so relaxing selection only decreases the probability that an adapted individual eventually fixes.

## 2 Wright-Fisher Model

The above reasoning translates readily to the Wright-Fisher model. The latter is characterized by discrete, non-overlapping generations, where the pedigree of the next generation is determined by a fitness-weighted sampling of the previous generation and mutations that occur independently in each offspring. Where previously (§1.1) we approximated the numbers of individuals of rare types via a birth-and-death process, here, we use a Galton-Watson branching process [18] with a Poisson distributed offspring number. The latter allows us to once again write a recurrence equation for the probability,  $p_1$  that a given individual is ancestral to an adapted individual that carries  $k$  mutations and is destined to fix.

### 2.1 Poisson-Galton-Watson Approximation

We again assume a population of fixed size  $N$ . In this discrete-generation formulation we can implement fecundity selection by assuming that an individual carrying  $i < k$  mutations is “chosen” to be the parent of an individual in the next generation with weight  $w_i \leq 1$  (*N.B.*, carrying fewer than  $k$  mutations is deleterious). Thus, if the  $j^{\text{th}}$  individual (we assign an arbitrary ordering to the population) carries  $i_j$  mutations and has  $X_j$  offspring in the next generation, the joint distribution of the  $X_j$  is multinomial:

$$\mathbb{P}\{(X_1, \dots, X_N) = (x_1, \dots, x_N)\} = \frac{N!}{x_1! \cdots x_N!} \left( \frac{w_{i_1}}{\sum_{\ell=1}^N w_{i_\ell}} \right)^{x_1} \cdots \left( \frac{w_{i_N}}{\sum_{\ell=1}^N w_{i_\ell}} \right)^{x_N}. \quad (\text{S.18})$$

Using the grouping property of the multinomial distribution,  $X_j$  is binomially distributed:

$$\mathbb{P}\{X_j = x_j\} = \frac{N!}{x_j!(N-x_j)!} \left( \frac{w_{i_j}}{\sum_{\ell=1}^N w_{i_\ell}} \right)^{x_j} \left( 1 - \frac{w_{i_j}}{\sum_{\ell=1}^N w_{i_\ell}} \right)^{N-x_j}. \quad (\text{S.19})$$

Now, letting  $\bar{w}$  denote the population mean fitness,

$$\bar{w} = \frac{1}{N} \sum_{\ell=1}^N w_{i_\ell}, \quad (\text{S.20})$$

we have

$$\mathbb{P}\{X_j = x_j\} = \frac{N!}{x_j!(N-x_j)!} \left( \frac{w_{i_j}}{N\bar{w}} \right)^{x_j} \left( 1 - \frac{w_{i_j}}{N\bar{w}} \right)^{N-x_j} \quad (\text{S.21})$$

$$= \frac{N(N-1) \cdots (N-x_j+1)}{N^{x_j}} \frac{1}{x_j!} \left( \frac{w_{i_j}}{\bar{w}} \right)^{x_j} \left( 1 - \frac{w_{i_j}}{N\bar{w}} \right)^{N-x_j}. \quad (\text{S.22})$$

Taking the large population limit, we get

$$\lim_{N \rightarrow \infty} \mathbb{P}\{X_j = x_j\} = \frac{1}{x_j!} \left( \frac{w_{i_j}}{\bar{w}} \right)^{x_j} e^{-\frac{w_{i_j}}{\bar{w}}}, \quad (\text{S.23})$$

*i.e.*, the  $j^{\text{th}}$  individual has a Poisson-distributed number of offspring with rate equal to their relative fitness,  $\frac{w_{i_j}}{\bar{w}}$ . A similar calculation shows that for any arbitrarily chosen  $k$  individuals ( $k$  fixed and

not growing with  $N$ ), their offspring numbers will be approximately independent Poisson random variables with rates equal to the relative fitness of the individuals.

While this gives us a characterization of an individual's reproductive output, unless the population is completely neutral, an individual's relative fitness will depend on the state of the whole population. If, however, as before, we restrict our attention to the situation where the number of individuals carrying one or more mutations is negligible compared to  $N$ , then

$$\bar{w} = \frac{1}{N} \sum_{\ell=1}^N w_{i_\ell} \sim w_0 \quad (\text{S.24})$$

as  $N \rightarrow \infty$  and thus, to first approximation, the  $j^{\text{th}}$  individual produces offspring with Poisson rate  $\frac{w_{i_j}}{w_0}$  that is independent of the composition of the population. Thus, if  $N \gg 1$ , we may approximate a small subpopulation of the Wright-Fisher process by a Galton-Watson branching process with Poisson distributed offspring.

Without loss of generality, we can take  $w_0 = 1$ , provided we scale the  $w_i$  appropriately, taking  $w_i = 1 - \delta_i$  for  $0 < i < k$  and  $w_k = 1 + s$ . As previously, we can model additive selection by taking  $\delta_i = \delta i$  and multiplicative selection by taking

$$w_i = (1 - \delta)^i = 1 - \delta i + \mathcal{O}(\delta^2) \quad (\text{S.25})$$

for  $0 < i < k$ , for some  $\delta \ll 1$ .

## 2.2 Galton-Watson Recurrence

As before, we wish to compute  $p_i$ , the probability that an individual carrying  $i$  mutations is ancestral to a mutation destined to fix. We will again do this recursively. We first compute the recurrence relation for an arbitrary Galton-Watson process. Then, in the next section, we will specialize to the Poisson-Galton-Watson process derived above.

No individual survives more than one generation, so an individual is ancestral to an individual destined to sweep if and only if one of its offspring is ancestral to the selected individual. As before, an offspring may or may not acquire an additional mutation, and we assume that the probability of multiple mutations is negligible ( $\mu_{i,j} = \mu_i$  if  $j = i + 1$ , and is zero otherwise). Thus, as previously  $p'_i = (1 - \mu_i)p_i + \mu_i p_{i+1}$  (S.3) is the probability that some given offspring of an individual carrying  $i$  mutations is ancestral.

To obtain the probability that at least one offspring is ancestral, we use the probability generating function for the offspring distribution: if  $X$  is the (random) number of offspring, then

$$f(z) = \sum_{x=0}^{\infty} \mathbb{P}\{X = x\} z^x. \quad (\text{S.26})$$

(for a Poisson distribution with rate  $\lambda$ , we have  $f(z) = e^{\lambda(z-1)}$ ). Conditional upon  $X = x$  the probability that *no* offspring is ancestral is  $(1 - p'_i)^x$ , whence the unconditional probability that no offspring is ancestral is

$$\sum_{x=0}^{\infty} \mathbb{P}\{X = x\} (1 - p'_i)^x = f(1 - p'_i), \quad (\text{S.27})$$

so that the probability that at least one is ancestral is  $1 - f(1 - p'_i)$ , yielding the recurrence

$$p_i = 1 - f(1 - p'_i). \quad (\text{S.28})$$

## 2.3 Solution in the Poisson Case

We now focus on the case of a Poisson-distributed number of offspring, corresponding to our approximation to the Wright-Fisher model (§2.1). In this case, (S.28) becomes

$$p_i = 1 - e^{-w_i p'_i}, \quad (\text{S.29})$$

where  $p'_i$  is given by (S.3), and as above, we adopt the convention that  $w_0 = 1$ , so that each individual's relative fitness is simply  $w_i$ .

This can be solved exactly using the Lambert  $W$ -function [2].  $W(z)$  is the multi-function satisfying the relation  $W(z)e^{W(z)} = z$ . Two branches of the  $W$ -function are real valued,  $W_0(z)$  and  $W_{-1}(z)$ , which map  $[-e^{-1}, \infty)$  to  $[-1, \infty)$  and  $[-e^{-1}, 0)$  to  $(-\infty, -1]$ , respectively. The two branches meet at the point  $(-e^{-1}, 1)$ . These real-valued branches satisfy a one-sided inverse identity:

$$\begin{aligned} W_0(ze^z) &= z \quad \text{if } z \geq -1, \\ W_{-1}(ze^z) &= z \quad \text{if } z \leq -1. \end{aligned} \quad (\text{S.30})$$

Moreover,

$$W_i(-Ae^{-A+z}) = W_i(-Ae^{-A}) + \frac{W_i(-Ae^{-A})}{1 + W_i(-Ae^{-A})}z + \frac{1}{2} \frac{W_i(-Ae^{-A})}{(1 + W_i(-Ae^{-A}))^3}z^2 + \mathcal{O}(z^3). \quad (\text{S.31})$$

see [16, Equation (2.40)]. If  $A \leq 1$ , using (S.30), this reduces to

$$W_i(-Ae^{-A+z}) = -A + \frac{A}{1-A}z + \frac{1}{2} \frac{A}{(1-A)^3}z^2 + \mathcal{O}(z^3). \quad (\text{S.32})$$

To use the  $W$ -function to solve the recurrence, we rearrange (S.29) as:

$$1 - p_i = e^{-w_i p'_i} = e^{-w_i((1-\mu_i)p_i + \mu_i p_{i+1})} \quad (\text{S.33a})$$

$$(p_i - 1)e^{w_i(1-\mu_i)p_i} = -e^{-w_i\mu_i p_{i+1}} \quad (\text{S.33b})$$

$$(p_i - 1)e^{w_i(1-\mu_i)(p_i-1)} = -e^{-w_i(1-\mu_i+\mu_i p_{i+1})} \quad (\text{S.33c})$$

$$w_i(1-\mu_i)(p_i-1)e^{w_i(1-\mu_i)(p_i-1)} = -w_i(1-\mu_i)e^{-w_i(1-\mu_i+\mu_i p_{i+1})} \quad (\text{S.33d})$$

Now,  $0 < w_i \leq 1$ ,  $0 \leq p_i \leq 1$ , and  $0 < \mu_{i,i+l} \ll 1$ , whence  $w_i(1-\mu_i)(p_i-1) \geq -1$ . Applying  $W_0$  to both sides of (S.33d), using (S.30) and (S.31) we conclude that

$$w_i(1-\mu_i)(p_i-1) = W_0\left(-w_i(1-\mu_i)e^{-w_i(1-\mu_i+\mu_i p_{i+1})}\right),$$

so

$$p_i = 1 + \frac{1}{w_i(1-\mu_i)}W_0\left(-w_i(1-\mu_i)e^{-w_i(1-\mu_i)-\mu_i w_i p_{i+1}}\right) \quad (\text{S.34})$$

which we can iterate and simplify as above.

Using (S.32), we can approximate (S.34) as

$$p_i = \frac{\mu_i w_i p_{i+1}}{1 - w_i(1 - \mu_i)} + \frac{1}{2} \frac{(\mu_i w_i p_{i+1})^2}{(1 - w_i(1 - \mu_i))^3} + \mathcal{O}(\mu_i^3) \quad (\text{S.35})$$

since  $-w_i(1 - \mu_i) \geq -1$ .

### 2.3.1 Initial condition $p_k$

To compute the initial value  $p_k$ , we note that a selectively favoured individual fails to establish itself if and only if all of its offspring fail to establish. Let  $p$  denote the probability of successful establishment and let  $q = 1 - p$ . Letting  $f(z)$  be the probability generating function for the offspring distribution, as above (S.26), we must have  $q = f(q)$ . In the Poisson case, this gives us  $q = e^{w_k(1-q)}$ . As before, we can solve this using the Lambert  $W$ -function. After rearranging, this is

$$-qw_k e^{-qw_k} = -w_k e^{-w_k}.$$

This always has the trivial solution  $q = 1$ , since  $-qw_k = -w_k \leq -1$ , but also a non-trivial solution with  $-qw_k \geq -1$  (cf. (S.30)):

$$q = -\frac{1}{w_k} W_0(-w_k e^{-w_k}). \quad (\text{S.36})$$

(recall  $w_k = 1 + s$  for  $s > 0$ ).

Assuming  $s \ll 1$ , we can find an elementary approximation to the Lambert  $W$ -function: let  $\mathcal{E}(z) = ze^z$ , so  $W(z)$  is  $\mathcal{E}^{-1}(z)$ . Then, the  $n^{\text{th}}$  derivative of  $\mathcal{E}(z)$  is

$$\mathcal{E}^{(n)}(z) = (n + z)e^z,$$

so, the second order Taylor polynomial for  $\mathcal{E}(z)$  at  $z = -1$  is

$$\mathcal{E}_2(z) = -e^{-1} + \frac{1}{2}e^{-1}(z + 1)^2.$$

We can invert  $\mathcal{E}_2(z)$  using the quadratic formula: setting  $y = \mathcal{E}_2(z)$  and expanding, we see that

$$z^2 + 2z - (1 + 2ey) = 0,$$

whence

$$\mathcal{E}_2^{-1}(y) = -1 \pm \sqrt{2(1 + ey)}$$

and

$$\mathcal{E}_2^{-1}(ze^z) = -1 \pm \sqrt{2(1 + ze^{z+1})}, \quad (\text{S.37})$$

the two roots corresponding to the two real branches of the Lambert  $W$  function:  $+$  gives the  $i = 0$  branch, which takes value in  $[-1, \infty)$ , whereas  $-$  gives the  $i = -1$  branch, which takes values in  $(-\infty, 1]$ .

Applying this approximation using  $z = -w_k = -(1 + s)$ , we find

$$q = -\frac{1}{w_k} W_0(-w_k e^{-w_k}) \approx \frac{-1 + \sqrt{2(1 - (1 + s)e^{-s})}}{1 + s}. \quad (\text{S.38})$$

Expanding the numerator using the binomial series, we find  $q = 1 - \frac{2s}{1+s} + \mathcal{O}(s^2)$ , whence

$$p_k = p = \frac{2s}{1 + s} + \mathcal{O}(s^2). \quad (\text{S.39})$$

*N.B.*, this approximation to the fixation probability of an allele with selective advantage  $s$  was first obtained in [8], without reference to the exact solution via the Lambert  $W$ -function. The factor

of two difference between the Wright-Fisher (S.39) and Moran models (S.6) is a consequence of a similar factor of two in the effective population sizes [7, §3.7].

Iterating (S.35) starting from  $p_k = \frac{2s}{1+s}$  and assuming as previously, that  $\mu_i \ll \delta_i \ll 1$ , we get that to lowest order,

$$\begin{aligned} p_{k-l} &= p_k \prod_{j=1}^l \frac{\mu_{k-j} w_{k-j}}{1 - w_{k-j}(1 - \mu_{k-j})} \\ &= \frac{2s}{1+s} \prod_{j=1}^l \frac{\mu_{k-j}(1 - \delta_{k-j})}{\delta_{k-j} + \mu_{k-j}(1 - \delta_{k-j})} \\ &\sim \frac{2s}{1+s} \prod_{j=1}^l \frac{\mu_{k-j}}{\delta_{k-j}}, \end{aligned} \quad (\text{S.40})$$

which corresponds to the case when  $i = 1$  in the Moran model (S.11), (S.12). Interestingly, unlike the Moran model, there are no other regimes that need to be considered. As previously, the waiting time to the first ancestor is  $t_1 \sim \frac{1}{N\mu_{0,1}p_1}$ , and our argument in §1.3 shows that relaxing selection decreases the mean time to adaptation in the Wright-Fisher model as well.

### 3 Justifying the Assumption that Mutants are Rare

Our arguments have been based upon the assumption that mutant individuals are rare, and that their numbers are negligible as compared to the number of wild-type individuals. We argued that this was a consequence of negative selection. Here, we briefly discuss some results on the maximum population size of birth-and-death and branching processes that show that provide evidence for that claim and quantify the strength of selection necessary. A rigorous proof would require additional coupling arguments to formalize the comparison between the Moran and Wright-Fisher models and the corresponding approximations; we omit these in the interest of brevity.

#### 3.1 Moran case

We will consider the total progeny of a given 1-mutant. To simplify our presentation, in what follows we will measure time from the birth of that individual. We let  $X(t)$  denote the total progeny of our focal individual alive at time  $t$  (so  $X(0) = 1$ ) and let

$$M = \sup_{t \geq 0} X(t) \quad (\text{S.41})$$

be the maximum number of progeny alive at any given time. Here, we present an estimate of the cumulative distribution function (c.d.f.) of  $M$ ,  $\mathbb{P}\{M \geq m\}$ .

We start with a simple observation: if  $T = \inf\{t : X(t) \in \{0, m\}\}$  then  $M \geq m$  if and only if  $X(T) = m$ . If  $X(t)$  is a birth-and-death process with birth and death rates  $b$  and  $d$  respectively, then this probability is well known (see *e.g.*, [13], [9, §3.7], or [5, Theorem 6.1]):

$$\mathbb{P}\{X(T) = m | X(0) = k\} = \frac{a^k - 1}{a^m - 1}. \quad (\text{S.42})$$

where  $a = \frac{d}{b}$ .



We cannot directly apply this to our process, where individuals have different birth rates,  $b_i = 1 - \delta_i$ , according to the number of mutations that they carry. We can, however, use it to obtain an upper bound: all individuals have birth rate no greater than  $b_1 = 1 - \delta_1$  and death rate one, and thus will be fewer in number than the birth-and-death process,  $X(t)$ , with those rates and  $X(0) = 1$ . For this  $X(t)$ ,  $a = \frac{1}{1-\delta_1}$ , so

$$a^m = e^{m \ln a} = e^{-m \ln(1-\delta_1)} = e^{m(\delta_1 + \mathcal{O}(\delta_1^2))},$$

and we see that the probability that  $M \geq m$ ,

$$\mathbb{P}\{M \geq m\} = \frac{a-1}{a^m-1}, \quad (\text{S.43})$$

is exponentially small provided  $m \gg \frac{1}{\delta_1}$ . Thus, to prevent the number of descendants from ever reaching size  $\mathcal{O}(N)$ , we need that  $\delta_1 \gg \frac{1}{N}$ . Note that

$$\begin{aligned} \mathbb{E}[M] &= \sum_{m=1}^{\infty} \mathbb{P}\{M \geq m\} = \sum_{m=1}^{\infty} \frac{a-1}{a^m-1} = \sum_{m=1}^{\infty} \frac{1}{a^{m-1} + \dots + a + 1} \\ &\leq \sum_{m=1}^{\infty} \frac{1}{a^{m-1}} = \frac{a}{a-1} = \frac{1}{\delta_1}, \end{aligned} \quad (\text{S.44})$$

with approximate equality when  $a \gg 1$  ( $\delta_1 \ll 1$ ). In particular, for very small  $\delta_1$  and large  $\mu_0$ —leading to a number of overlapping lineages—the rare mutant approximation can fail.

### 3.2 Wright-Fisher case

The argument for the Wright-Fisher case is almost identical to §3.1. We will bound the total number of mutant progeny of a single individual by a Poisson-Galton-Watson branching process  $X(t)$  where all individuals have fitness  $w_1 = 1 - \delta_1$ , and  $X(0) = 1$ .

Define  $M$  and  $T$  as previously (where now time is measured in generations). The result (S.42) continues to hold for the Galton-Watson process (see [12, §3]) provided there exists  $a > 1$  such that  $f(a) = a$ , where  $f(z)$  is the p.g.f. for the offspring distribution (see (S.26) above) of an individual with one mutation (hence  $m = f'(1) < 1$ ).

Just as we used the Lambert  $W$  function to compute the extinction probability  $q$  for the beneficial mutation, we can use it to compute  $a$  for a  $\text{Poisson}(w_1)$  offspring distribution. Recall that for this distribution, the p.g.f. is  $f(z) = e^{w_1(z-1)}$ , so  $a = f(a)$  if and only if

$$-w_1 a e^{-w_1 a} = -w_1 e^{-w_1},$$

which has two solutions,  $a = 1$  and

$$a = -\frac{1}{w_1} W_{-1}(-w_1 e^{-w_1}). \quad (\text{S.45})$$

In particular, if  $w_1 = 1 - \epsilon$  for some small  $\epsilon$ , then using (S.37) above, we have

$$a \approx \frac{-1 - \sqrt{2(1 - (1 - \epsilon)e^{-\epsilon})}}{1 - \epsilon} \approx 1 + 2\epsilon \quad (\text{S.46})$$

Now suppose  $\epsilon = \epsilon_N \rightarrow 0$  as  $N \rightarrow \infty$ . Then,

$$a^{m_N} = e^{m_N \ln a} = e^{m_N \ln(1+2\epsilon_N)} \approx e^{2m_N \epsilon_N},$$

so the probability that  $M \geq m_N$  tends to 0 as  $N \rightarrow \infty$  if  $m_N \gg \frac{1}{\epsilon_N}$ , and to 1 if  $m_N \ll \frac{1}{\epsilon_N}$ , and to a non-zero constant if  $m_N \propto \frac{1}{\epsilon_N}$ . In particular, the number of deleterious mutants could reach  $\mathcal{O}(N)$  if  $\epsilon_N = \mathcal{O}(\frac{1}{N})$ , *i.e.*, if selection is weak. This gives us the lower bound on the strength of selection as discussed above.

### 3.3 Conditioning on Not Fixing is Equivalent to Assuming Negative Selection

The above arguments bound the number of  $i$ -mutants,  $i < k$ , for which the mutations are deleterious. We must also, however, bound the number of  $k$ -mutants that do not go on to fixation, which, due to epistasis, have selective advantage  $s$ . These lines will eventually go on to extinction, but we must ensure they remain small.

This, fortunately, follows almost immediately from our previous work. *Conditioned upon extinction*, a birth-and-death process with birth rate  $b$  and death rate  $d$ ,  $b > d$ , is equal in law (*i.e.*, has the same probability distribution) as a birth-and-death process with birth rate  $d$  and death rate  $b$  (see [15, Lemma B.2]). Applying this to lineages of  $k$ -mutants conditioned on extinction, we have  $b = 1 + s$ ,  $d = 1$ , and we can take  $a = \frac{b}{d} = 1 + s > 1$  in (S.43) to estimate the maximum number of  $k$ -mutants.

Similarly, *conditioned upon extinction*, a Galton-Watson process with p.g.f  $f(z)$  for the offspring distribution and extinction probability  $q < 1$  (recall  $f(q) = q$ ) is equivalent to a Galton-Watson process with offspring p.g.f.  $g(z) = \frac{1}{q}f(qz)$  (see *e.g.*, [1, Theorem I.8.1]). Thus,  $g(a) = a$  for  $a = \frac{1}{q} > 1$ . Applying this to the  $k$  mutants, we have  $q = \frac{2s}{1+s} < 1$  so taking  $a = \frac{1+s}{2s} > 1$  in (S.43) gives us a bound on their maximum number.

## 4 Bounding the Adaptation Time

We now turn our attention to estimating the expected waiting time from the birth of the ultimate ancestor carrying 1 mutation to the birth of the adapted individual. There is a direct line of descent between the 1-ancestor and the adapted individual. In what follows, we will refer to individuals on this line as **ancestral**, and those carrying  $i$  mutations as  **$i$ -ancestors**. Thus an  $i$  ancestor gives birth at rate  $i$ , and its offspring carries  $i + 1$  mutations with probability  $\mu_i$ .

Imagine following the genealogical tree between the ultimate ancestor and adapted individual: we follow the branch corresponding to each ancestral individual until they give birth to their successor in the line of descent, after which we follow the successor's branch. Over its lifespan, an  $i$ -ancestor gives births at the times of a Poisson process with rate  $b_i$ , but most of these will not be ancestral: the probability that a given offspring is ancestral is  $p'_i$  (S.3). By the thinning property of the Poisson process (see *e.g.*, [3, §3.1]), the  $i$ -ancestor gives birth to would-be ancestors (only the first will be truly ancestral) as a Poisson process with rate  $b_i p'_i$  *i.e.*, the waiting time birth of next ancestor is exponentially distributed with rate  $b_i p'_i$ . Again, the probability that one of these ancestors is an  $i + 1$ -ancestor is  $\mu_i$ , further thinning the process, so that the waiting time from the first  $i$ -ancestor to the first  $i + 1$  ancestor is exponentially distributed with rate  $b_i p'_i \mu_i$ .

The total adaptation time is the sum of these exponential waiting times, but we must *condition upon the fact that the adapted individual appears before the line of descent goes extinct*. This

is a difficult, and possibly intractable problem. Here, we will content ourselves with a coarse approximation: we will simplify the processes by *underestimating* the number of mutants, which we do by considering a birth-and-death process  $X(t)$  with death rate 1 and birth rate  $b = b_{k-1} = 1 - \delta_{k-1}$  and  $X(0) = 1$  (compare §3.1), and we will assume that all mutation rates are equal  $\mu_i \equiv \mu$ . We then have  $p'_i \equiv p'$ .

The *unconditioned* waiting time to the adapted individual, say  $T_a$ , is then a sum of  $k$  exponentially distributed random variables with rate  $\lambda = bp'\mu$ , *i.e.*, an Erlang or Gamma distributed random variable with probability density function (p.d.f)<sup>1</sup>

$$\frac{\lambda^k}{(k-1)!} t^{k-1} e^{-\lambda t}. \quad (\text{S.47})$$

The total line of descent, on the other hand, has lifespan  $L$  with cumulative distribution function (c.d.f.) [11, §3]

$$P_0(\ell) = \mathbb{P}\{L \leq \ell\} = \frac{e^{(b-1)\ell} - 1}{be^{(b-1)\ell} - 1}. \quad (\text{S.48})$$

*N.B.*,  $b < 1$ , so  $P_0(\infty) = 1$ .

Now, suppose  $L = \ell$ . The density of  $T_a$  conditional upon  $T_a \leq \ell$  is

$$\frac{\lambda^k t^{k-1} e^{-\lambda t} \mathbb{1}_{[0, \ell)}(t)}{\gamma(k, \lambda \ell)}. \quad (\text{S.49})$$

where  $\mathbb{1}_{[0, \ell)}(t)$  is the indicator function of the interval  $[0, L)$ <sup>2</sup> and  $\gamma(a, z)$  is the **incomplete gamma function** [4, §8.2.1]:

$$\gamma(a, z) = \int_0^z t^{a-1} e^{-t} dt, \quad (\text{S.50})$$

the latter arising as

$$\begin{aligned} \mathbb{P}\{T_a \leq \ell\} &= \frac{\lambda^k}{(k-1)!} \int_0^\ell t^{k-1} e^{-\lambda t} dt \\ &= \frac{1}{(k-1)!} \int_0^{\lambda \ell} t^{k-1} e^{-t} dt = \frac{1}{(k-1)!} \gamma(k, \lambda \ell). \end{aligned}$$

Integrating (S.49) against  $P'_0(\ell)$ , (S.48), gives us the p.d.f. of  $T_a$  conditional on the random lifespan  $L$  (as opposed to the fixed value  $L = \ell$ ):

$$\int_t^\infty \frac{\lambda^k t^{k-1} e^{-\lambda t}}{\gamma(k, \lambda \ell)} P'_0(\ell) d\ell. \quad (\text{S.51})$$

We can use this to estimate the expected adaptation time,  $\mathbb{E}[T_a]$ :

$$\begin{aligned} \mathbb{E}[T_a] &= \int_0^\infty t \int_t^\infty \frac{\lambda^k t^{k-1} e^{-\lambda t}}{\gamma(k, \lambda \ell)} P'_0(\ell) d\ell dt \\ &= \int_0^\infty \int_0^\ell \frac{\lambda^k t^k e^{-\lambda t}}{\gamma(k, \lambda \ell)} dt P'_0(\ell) d\ell \\ &= \frac{1}{\lambda} \int_0^\infty \frac{\gamma(k+1, \lambda \ell)}{\gamma(k, \lambda \ell)} P'_0(\ell) d\ell. \end{aligned}$$

<sup>1</sup>Informally, a random variable  $X$  has p.d.f.  $f(x)$  if  $\mathbb{P}\{X \in [x, x+dx]\} = f(x) dx$ .

<sup>2</sup> $\mathbb{1}_{[0, \ell)}(t) = 1$  if  $t \in [0, \ell)$ , and is zero otherwise.

This latter integral does not appear to be analytically tractable, but we can obtain a suitable bound by noting that for integer values of  $k$ , [4, §8.4.7],

$$\gamma(k+1, z) = k! \left( 1 - e^{-z} \sum_{j=0}^k \frac{z^j}{j!} \right). \quad (\text{S.52})$$

In particular,

$$\lim_{z \rightarrow \infty} \frac{\gamma(k+1, z)}{\gamma(k, z)} = \frac{\Gamma(k+1)}{\Gamma(k)} = k, \quad (\text{S.53})$$

and  $\frac{\gamma(k+1, z)}{\gamma(k, z)} \leq k$  for all  $z$ , whereas for  $z \leq k+1$ ,

$$\frac{\gamma(k+1, z)}{\gamma(k, z)} \leq \frac{k}{k+1} z. \quad (\text{S.54})$$

Thus,

$$\begin{aligned} \mathbb{E}[T_a] &\leq \frac{1}{\lambda} \left[ \int_0^{\frac{k+1}{\lambda}} \frac{\lambda k}{k+1} \ell P'_0(\ell) d\ell + \int_{\frac{k+1}{\lambda}}^{\infty} k P'_0(\ell) d\ell \right] \\ &= \frac{k}{k+1} \left[ \frac{k+1}{\lambda} P_0\left(\frac{k+1}{\lambda}\right) - \int_0^{\frac{k+1}{\lambda}} P_0(\ell) d\ell \right] + \frac{k}{\lambda} [P_0(\infty) - P_0\left(\frac{k+1}{\lambda}\right)] \\ &= \frac{k}{\lambda} - \frac{k}{k+1} \int_0^{\frac{k+1}{\lambda}} P_0(\ell) d\ell \\ &= \frac{k}{\lambda} - \frac{k}{k+1} \int_0^{\frac{k+1}{\lambda}} \frac{e^{(b-1)\ell} - 1}{be^{(b-1)\ell} - 1} d\ell \\ &= \frac{k}{\lambda} - \frac{k}{(k+1)(1-b)} \int_{e^{-\frac{(k+1)(1-b)}{\lambda}}}^1 \frac{1-u}{u(1-bu)} du \\ &= \frac{k}{\lambda} - \frac{k}{(k+1)(1-b)} \int_{e^{-\frac{(k+1)(1-b)}{\lambda}}}^1 \frac{1}{u} + \frac{1-b}{1-bu} du \\ &= \frac{k}{b(k+1)} \left[ -\ln(1-b) + \ln\left(1 - be^{-\frac{(k+1)(1-b)}{\lambda}}\right) \right]. \end{aligned}$$

Recalling that  $b = 1 - \delta_{k-1}$  and  $\lambda = bp'\mu \ll 1$ , the latter is approximately

$$\frac{k}{(1 - \delta_{k-1})(k+1)} \ln(\delta_{k-1}^{-1}) - \frac{k}{k+1} e^{\frac{-(k+1)\delta_{k-1}}{(1-\delta_{k-1})p'\mu}} \quad (\text{S.55})$$

where the latter term is negligible compared to the first.

Finally, it is a standard result (see *e.g.*, [5, Theorem 6.3]) that the fixation time for an allele with selection coefficient  $s$  in a population size  $n$  is approximately  $\frac{2}{s} \ln n$ .

# References

- [1] K. B. Athreya and P. Ney. *Branching processes*. Springer-Verlag, Berlin, 1972.
- [2] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth. On the Lambert  $W$  function. *Adv. Comput. Math.*, 5(1):329–359, Dec 1996.
- [3] D. R. Cox and V. Isham. *Point processes*, volume 12. CRC Press, 1980.
- [4] *NIST Digital Library of Mathematical Functions*. <http://dlmf.nist.gov/>, Release 1.0.27 of 2020-06-15. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds.
- [5] R. Durrett. *Probability Models for DNA Sequence Evolution*. Springer, New York, 2nd edition, 2009.
- [6] R. Durrett, D. Schmidt, and J. Schweinsberg. A waiting time problem arising from the study of multi-stage carcinogenesis. *Ann. Appl. Prob.*, pages 676–718, 2009.
- [7] W. J. Ewens. *Mathematical Population Genetics*. Springer, New York, 2nd edition, 2004.
- [8] J. B. S. Haldane. A mathematical theory of natural and artificial selection. Part V: Selection and mutation. *Proc. Camb. Phil. Soc.*, 23:838–844, 1927.
- [9] S. Karlin and H. M. Taylor. *A First Course in Stochastic Processes*. Academic Press, New York, 2nd edition, 1975.
- [10] D. G. Kendall. On some modes of population growth leading to R. A. Fisher’s logarithmic series distribution. *Biometrika*, 35(1/2):6–15, 1948.
- [11] D. G. Kendall. On the generalized “Birth-and-Death” process. *Ann. Math. Stat.*, 19(1):1–15, 1948.
- [12] T. Lindvall. On the maximum of a branching process. *Scand. J. Stat.*, pages 209–214, 1976.
- [13] P. A. P. Moran. Random processes in genetics. *Proc. Camb. Phil. Soc.*, 54:60–71, 1958.
- [14] S. P. Otto and M. C. Whitlock. Fixation probabilities and times. In *eLS*. John Wiley & Sons, Ltd, 2013.
- [15] T. L. Parsons. Invasion probabilities, hitting times, and some fluctuation theory for the stochastic logistic process. *J. Math. Biol.*, 77(4):1193–1231, 2018.
- [16] T. L. Parsons and D. J. D. Earn. Analytical approximations for the phase plane trajectories of the SIR model with vital dynamics. to appear in *SIAM J. Appl. Math.*; <https://cnrs.hal.science/hal-04178969>, 2023.
- [17] J. Schweinsberg. The waiting time for  $m$  mutations. *Electron. J. Probab.*, 13:1442–1478, 2008.
- [18] H. W. Watson and F. Galton. On the probability of the extinction of families. *J. Anthropol. Inst. GB & Ireland*, 4:138–144, 1875.
- [19] D. B. Weissman, M. M. Desai, D. S. Fisher, and M. W. Feldman. The rate at which asexual populations cross fitness valleys. *Theor. Popul. Biol.*, 75(4):286–300, 2009.