

Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Mahan Ghafari

Date

Quantitative analysis of adaptive evolution

By

Mahan Ghafari
Master of Science
Physics

Daniel Weissman
Advisor

Gordon Berman
Committee Member

Minsu Kim
Committee Member

Katia Koelle
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Quantitative analysis of adaptive evolution

By

Mahan Ghafari
B.Sc., Sharif University of Technology, 2015

Advisor: Daniel Weissman, Ph.D.

An abstract of
A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science
in Physics
2018

Abstract

Quantitative analysis of adaptive evolution

By Mahan Ghafari

In this work, we first perform a quantitative analysis on the original data from the Luria-Delbrück fluctuation experiment. We compare the performance of the Darwinian model of evolution to the Lamarckian model and a combined model that allows both Darwinian and Lamarckian mechanisms. We also consider the possibility of neither model fitting the experiment. Using a Bayesian model selection approach, we show that although the experiment does, indeed, favor the Darwinian over pure Lamarckian evolution, it does not rule out the combined model and, hence, cannot completely rule out Lamarckian contributions to evolution. Next, we mainly focus on complex adaptations involving three neutral mutations. We show that large populations can cross them rapidly via lineages that acquire multiple mutations while remaining at low frequency. Plateau-crossing is fastest for very large populations. At intermediate population sizes, recombination can greatly accelerate adaptation by combining independent mutant lineages to form triple-mutants. For more frequent recombination, such that the population is kept near linkage equilibrium, we extend our analysis to find simple expressions for the expected time to cross plateaus of arbitrary width.

Quantitative analysis of adaptive evolution

By

Mahan Ghafari
B.Sc., Sharif University of Technology, 2015

Advisor: Daniel Weissman, Ph.D.

A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science
in Physics
2018

Acknowledgement

I would like to express the deepest appreciation to my advisor, Daniel Weissman, who has been incredibly supportive throughout my postgraduate studies. I have been extremely lucky to have an advisor who has taught me how to be a professional researcher and who has given me so many wonderful opportunities to do the research that I enjoy.

I must express my gratitude to my mentors, Fereydoon Family and Bruce Levin, for generously sharing their invaluable experience and time with me and showing me, by their example, what a good scientist (and human being) should be.

Completing this work would have been all the more difficult were it not for the support and friendship provided by the members of the Physics and Biology Departments at Emory University. I am indebted to them for their help.

And I thank the two amazing women of my life, my mother and wife, for their continued support and encouragement. They are the most important people in my world and I dedicate this thesis to them.

Contents

1	Introduction	1
2	Luria-Delbrück, revisited: The classic experiment does not rule out Lamarckian evolution	2
2.1	Introduction	2
2.2	Methods	5
2.2.1	Models and Notational preliminaries	5
2.2.2	Computational models	11
2.2.3	Quality of fit	12
2.2.4	Comparing models	13
2.2.5	Statistical power of the tests	15
2.3	Results	17
2.3.1	Experiment 22	18
2.3.2	Experiment 23	19
2.4	Discussion	21
3	The expected time to cross extended fitness plateaus	25
3.1	Introduction	25
3.2	Model	27
3.3	Results	27

3.3.1	Rare recombination, $r \ll s$	31
3.3.2	Frequent recombination, $r \gg s$	33
3.4	Analysis	37
3.4.1	Rare recombination ($r \ll s$)	37
3.4.2	Frequent recombination ($r \gg s$)	47
3.5	Discussion	51
3.A.	Appendix	55
3.A.1	Small populations with rare recombination	55

References **63**

List of Figures

2.1	Schematics of the two theories tested by the Luria-Delbrück experiment	6
2.2	Data with a known Darwinian and Lamarckian mutation rates	17
2.3	Posterior likelihood of the Darwinian and Lamarckian for Exp. #22	20
2.4	Experimental data for Exp. #22 and ML fit for Combined model	21
2.5	Posterior likelihood of the Darwinian and Lamarckian for Exp. #23	22
2.6	Experimental data for Exp. #23 and ML fit for Combined model	23
3.1	A visualization of the fitness landscape	28
3.2	Schematic diagram of the asymptotic dynamics for $K = 3$ fitness plateau	30

3.3	Expected time for a rarely recombining population to cross the plateau	34
3.4	Expected time for a rapidly recombining population to cross the plateau	36
3.5	Typical plateau-crossing dynamics for a very large population	41
3.6	Typical plateau-crossing dynamics for a moderately large population .	46
A.1	8 different ways of plateau-crossing for a small population with $r \ll s$	57

List of Tables

1	Symbol definitions	28
2	Rate of plateau-crossing for asymptotic regimes	29

1 Introduction

Evolution is inherently a random process and, therefore, any quantitative prediction is inevitably a statistical statement. One of the key roles that physicists can play is in framing and developing such quantitative models [Fisher et al., 2013]. In this work, we discuss two fundamental, classic problems in evolution, namely the origins of variation and complexity, and re-examine them quantitatively. Having a stronger computational power and new statistical methods at our disposal, we go back and reconsider some of the neglected or dismissed processes of Lamarckian evolution and irreducible complexity. In the first chapter of this thesis, we perform the quantitative analysis missing in the Luria-Delbrück paper and put three models of adaptation to test by finding out how well they fit the data. Traditionally, it has been assumed that adaptations requiring a fitness valley or plateau crossing are extremely unlikely because it would be much slower for populations to cross them rather than accumulate beneficial mutations. Therefore, quantitative analyses on complex adaptations has largely been dismissed. In the second chapter, we analyze how sexual populations can acquire complex adaptations requiring three individually neutral mutations. This thesis is based on the work done in collaboration with Daniel Weissman, Ilya Nemenman, Caroline Holmes, Varun Saravanan, and Anzar Abbas. Chapter 2 was published in Holmes et al. [2017], and Chapter 3 was originally disseminated in Ghafari and Weissman [2018].

2 Luria-Delbrück, revisited: The classic experiment does not rule out Lamarckian evolution

2.1 Introduction

From the dawn of evolutionary biology, two general mechanisms, Darwinian and Lamarckian, have been routinely considered as alternative models of evolutionary processes. The *Darwinian hypothesis* posits that adaptive traits arise continuously over time through spontaneous mutation, and that evolution proceeds through natural selection on this already existing variation. In contrast, the *Lamarckian hypothesis* proposes that adaptive mutations arise in response to environmental pressures. The Nobel Prize winning fluctuation test by Luria and Delbrück [1943] is credited with settling this debate, at least in the context of evolution of phage-resistant bacterial cells.

Luria and Delbrück realized that the two hypotheses would lead to different variances (even with the same means) of the number of bacteria with any single adaptive mutation. Specific to the case of bacteria exposed to a bacteriophage, this would result in different distributions of the number of surviving bacteria, cf. Fig. 2.1. In the Darwinian scenario, there is a possibility of a phage-resistance mutation arising in generations prior to that subjected to the phage. If this mutation happens many generations earlier, there will be a large number of resistant progeny who will survive (a “jackpot” event). However, there will be no survivors if the mutation does not exist in the population at the moment the phage is introduced. If the same ex-

periment were repeated many times, the variance of the number of survivors would be large. In contrast, in the Lamarckian scenario, the distribution of the number of survivors is Poisson. Indeed, each occurring mutation (and hence each survivor) happens with a small probability, independent of the others. This would result in the usual square-root scaling of the standard deviation of the number of survivors, a much smaller spread than in the Darwinian case.

To test this experimentally, Luria and Delbrück let the cells grow for a few generations, exposed them to a phage, plated the culture, and then counted the number of emergent colonies, each started by a single resistant, surviving bacterium. They found that the distribution of the number of survivors, as measured by the number of colonies grown after plating, was too heavy-tailed to be consistent with the Poisson distribution. They concluded then that the bacteria must evolve using the Darwinian mechanism. They could not derive an analytical form of the distribution of survivors in the Darwinian model, so that their data analysis was semi-quantitative at best. In particular, they could only establish that the Darwinian model fits the data better than the Lamarckian/Poissonian one, but they could not quantify *how good* the fit is.

Potentially even more importantly, the original paper contrasted only two scenarios: pure Lamarckian and pure Darwinian ones. However, it is possible that both processes have a role in bacterial evolution, as is abundantly clear now in the epoch of epigenetics and, especially, CRISPR-Cas bacterial immunity, which is essentially Lamarckian [Koonin and Wolf, 2009, Jablonka and Lamb, 2002, Jablonka and Raz, 2009, Barrangou et al., 2007, Koonin and Wolf, 2016]. In addition, stress can increase

the (undirected) mutation rates, so that more mutations arise at the time of the challenge and then get selected, which will also appear Lamarckian: Luria-Delbrück experiments cannot distinguish between induced directed and undirected mutations. Thus ruling out a significant Lamarckian contribution to evolution through either of these or any other mechanism would require us to show not only that the Darwinian model explains the data better than the Lamarckian one, but also that the Darwinian model is more likely than the *Combination model*, which allows for both types of evolutionary processes. Evolution could also proceed through an entirely different mechanism, so that neither of the proposed models explain the data. Distinguishing between these possibilities requires evaluating whether a specific model fits the data well, rather than which of the models fits the data better.

Unlike Luria and Delbrück in 1943, we have powerful computers and new statistical methods at our disposal. Distributions that cannot be derived analytically can be estimated numerically, and model comparisons can be done for models with different numbers of parameters. In this paper, we perform the quantitative analysis missing in the Luria-Delbrück paper and use their original data to evaluate and compare the performance of three models: Darwinian (D), Lamarckian (L), and Combination (C) models. The comparison is somewhat complicated by the fact that both the L and D models are special cases of the C model, so that C is guaranteed to fit not worse than either L or D. We use Bayesian Model Selection [MacKay, 1992, Mackay, 2003], which automatically penalizes for more complex models (such as C) to solve the problem. We conclude that, while the L model is certainly inconsistent with the data, D and C explain the data about equally well when this penalty for complexity is accounted for. Thus the Lamarckian contribution to evolution cannot be ruled

out by the 1943 Luria and Delbrück data. Further, while D and C fit equally well, neither provide a quantitatively good fit to one of the two primary experimental data sets of Luria and Delbrück’s paper, suggesting that the classic experiment may have been influenced by factors or processes not considered.

Even though we will show that our analysis of the Luria-Delbrück data does not rule out Lamarckian contributions to evolution, our goal is not to challenge the well-established knowledge that this particular system (T1 phage interacting with *E. coli*) is largely purely Darwinian [Lederberg and Lederberg, 1952, Newcombe, 1949, Taylor, 1963, Hayes, 1964, Hantke and Braun, 1978]. Instead our goal is to limit ourselves to the original 1943 data, even though additional experiments have been performed many times since then, and to ask: Does this data actually tell us what every textbook says, and namely that the 1943 experiment has ruled out Lamarckian evolution in favor of Darwinian?

2.2 Methods

2.2.1 Models and Notational preliminaries

There have been many theoretical attempts, with varying degrees of success, to find closed-form analytical expressions of the distribution of mutants under the Darwinian—but not the combined—scenario (the Luria-Delbrück distribution). These have followed several different modeling assumptions [Lea and Coulson, 1949, Armitage, 1952, Mandelbrot, 1974, Bartlett, 1978, Sarkar, 1991, Zheng, 2007, Angerer, 2001], with good reviews by Zheng [1999] and Ycart [2013]. For example, models

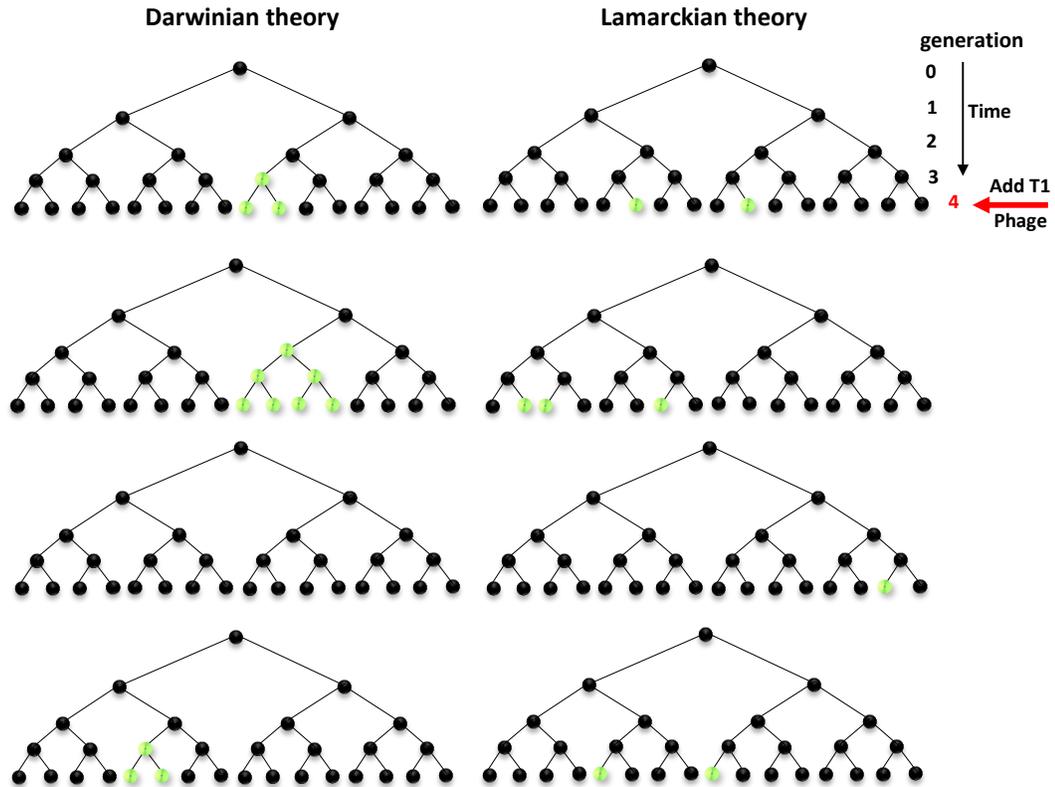


Figure 2.1: Schematics of the two theories tested by the Luria-Delbrück experiment. Here black dots denote bacteria susceptible to the bacteriophage, and green dots resistant bacteria. Each tree represents one realization of the experiment, which starts with a single bacterium (top). The bacterium then divides for several generations, and the phage is introduced into the culture at the last generation (bottom row of each tree). Darwinian theory (left column) of evolution predicts that mutations happen spontaneously throughout the experiment. Thus different repeats of the experiment (different trees) will produce very broadly distributed numbers of survivors (from 0 to 4 in this example). In contrast, in the Lamarckian case (right column), mutations only occur when the phage is introduced, so that the standard deviation of the number of survivors in different repeats (from 1 to 3 in this example) scales as the square root of their mean, which is much smaller than in the Darwinian case.

with constant and synchronous division times [Sarkar, 1991, Zheng, 2007] or many different distributions of generation times, in some cases allowing for different growth

rates for the wild-type and the mutant populations [Koch, 1982, Jones et al., 1994, Jaeger and Sarkar, 1995], have been proposed to find the distribution of survivors. In this work, we do not advance these analytical treatments. However, we present one such analysis, mainly to introduce notation and to illustrate complications of using analytical expressions for our statistical analysis.

We follow Haldane’s modeling hypotheses [Sarkar, 1991] and assume that (i) normal cells and mutants have the same fitness until the phage is introduced, (ii) all cells undergo synchronous divisions, (iii) no cell dies before the introduction of the phage, (iv) mutations occur only during divisions, with each daughter becoming a mutant independently (D case), or only when the phage is introduced (L case), and (v) backwards mutations are negligible. With these assumptions, the D and the C models are able to produce very good fits to the experimental data (see below), which suggests that relaxation of these assumptions and design of more biologically realistic models is unnecessary in the context of these experiments.

For the subsequent analyses, let N_0 be the initial number of wild-type, phage-sensitive bacteria, and g be the number of generations before the phage is introduced, so that the total number of bacteria after the final round of divisions is $N = 2^g N_0$, and the total that have ever lived is $2N - N_0$. We use θ_D to denote the probability of an adaptive (Darwinian) mutation during a division, and θ_L to denote the probability of an adaptive Lamarckian mutation when the phage is introduced. With this, and discounting the probability of another mutation in the already resistant progeny, the mean number of adaptive Darwinian mutations at generation g is $m_D = \theta_D(2N - N_0)$, and the mean number of adaptive Lamarckian mutations is $m_L = \theta_L N$. The number

of survivors in the L model is Poisson-distributed with the parameter m_L :

$$P_L(k|\theta_L, N_0) = \frac{e^{-m_L} m_L^k}{k!} \quad (2.1)$$

For the D model, there are multiple ways to have a certain number of resistant bacteria, k , in the population of size N before introducing the phage. For instance, there are four ways to have 5 resistant bacteria (i. e., $k = 5$): (i) One mutation occurs 2 generations before the phage introduction (where the total living population is $N/4$ at that generation), resulting in 4 resistant progenies in the last generation, and one more mutation at the last generation, making a total of 5 resistant cells before the phage introduction. This is the most likely scenario with probability $P_5^{(i)} = (1 - \theta_D)^{(2N - N_0) - 8} \theta_D^2 \binom{N/4}{1} \binom{N-4}{1}$, where $(2N - N_0)$ is the total number of cells that have ever lived in the entire experiment, so that $(2N - N_0) - 8$ is the total number that have ever lived without mutating. The θ_D^2 factor indicates that a total of two mutations have occurred in the population. The first choose factor denotes the number of independent mutational opportunities 2 generations before the phage introduction, and the second one denotes the number of mutational opportunities in the last generation. (ii) Two mutations occur 1 generation before the phage introduction and one more mutation in the last generation. This is less likely than (i) with probability $P_5^{(ii)} = (1 - \theta_D)^{(2N - N_0) - 7} \theta_D^3 \binom{N/2}{2} \binom{N-4}{1}$, where, there are a total of 7 mutant that have ever lived in the history of the experiment and 3 mutational events before the introduction of the phage. (iii) One mutation occurs 1 generation before the phage introduction and 3 more mutations in the last generation. This is less likely than (ii) with probability $P_5^{(iii)} = (1 - \theta_D)^{(2N - N_0) - 6} \theta_D^4 \binom{N/2}{1} \binom{N-2}{3}$. (iv)

Five mutations occur in the last generation before introducing the phage. This is the least likely scenario with probability $P_5^{(iv)} = (1 - \theta_D)^{(2N - N_0) - 5} \theta_D^5 \binom{N}{5}$. In general, for an arbitrary number of resistant cells, k , let Π_k denote the set of sequences (a_0, a_1, \dots) that satisfy $k = \sum_0^\infty a_s 2^s$, such that $a_s \in \mathbb{Z}_{\geq 0}$. This condition captures all the possible sequences of $\{a_s\} \in \Pi_K$ that produce k number of resistant cells. For instance, in case (i) the corresponding sequence is $\{a_2 = 1, a_1 = 0, a_0 = 1\}$, and in case (ii) it is $\{a_1 = 2, a_0 = 1\}$. Then, following Haldane's approach [Sarkar, 1991], we can write $P_D(k)$, the probability of finding k resistant cells given the Darwinian model of evolution, as

$$P_D(k|\theta_D, N_0) = \sum_{\{a_s\} \in \Pi_K} (1 - \theta_D)^{(2N - N_0) - \sum_{s=0}^\infty a_s (2^{s+1} - 1)} \theta_D^x \times \prod_{s=0}^\infty \binom{\frac{N}{2^s} - \sum_{n=s+1}^\infty a_n (2^{n-s})}{a_s}, \quad (2.2)$$

where $x \equiv \sum \{a_s\} a_s$ and the probability $P_D(k|\theta_D, N_0)$ is summed over all the possible sequences $\{a_s\} \in \Pi_K$ that produce the number k ; in the case of $P_{k=5}$ mentioned earlier, $P_D(5|\theta_D, N_0) = P_5^{(i)} + P_5^{(ii)} + P_5^{(iii)} + P_5^{(iv)}$.

For the Combination model, both the L and the D processes contribute to generating survivors. Thus we write the distribution of the number of survivors in this case as a convolution

$$P_C(k|\theta_L, \theta_D, N_0) = \sum_{k'=0}^k P_D(k'|\theta_D, N_0) P_L(k - k'|\theta_L, N_0). \quad (2.3)$$

Further analytical progress on the problem is hindered by additional complica-

tions. First, in actual experiments, the initial number of bacteria in the culture N_0 is random and unknown. We view it as Poisson-distributed around the mean that one expects to have, denoted as $\Pi(N_0|\bar{N}_0)$. This gives:

$$P_{D/L/C}(k|\theta_D, \theta_L, \bar{N}_0) = \sum_{N_0=0}^{\infty} P_{D/L/C}(k|N_0)\Pi(N_0|\bar{N}_0). \quad (2.4)$$

Finally, in some of the Luria-Delbrück experiments, they plated only a fraction r the entire culture subjected to the phage. This introduced additional randomness in counting the number of survivors after the plating, k_p , which we again model as a Poisson distribution with the mean rk , $\Pi(k_p|rk)$ [Stewart et al., 1990, Stewart, 1991, Montgomery-Smith et al., 2016], resulting in the overall distribution of survivors:

$$P_{D/L/C}(k_p|\theta_D, \theta_L, \bar{N}_0) = \sum_k^{\infty} \Pi(k_p|rk)P_{D/L/C}(k). \quad (2.5)$$

Equation (2.5) illustrates the main complication of using analytical results for statistical inference studies: it is not computationally efficient, involving multiple nested (and infinite) sums. Alternative approaches (e. g., [Lea and Coulson, 1949]) represent the distribution as recursive expressions, through the inverse Fourier transform of the characteristic functional, or through low-order moments. These expressions are also not easy to calculate, or are hard to compare to the experimental probability distribution in the inference step due to additional complications, such as dilution or the Lamarckian contribution in the C model. One can try to develop an efficient algorithm for evaluating the probability of the number of survivors for the C model, similar to the ones that have already been developed for the D case [Ma et al., 1992, Sarkar et al., 1992], but this is not a trivial task. Instead, since our focus is on

the inference and not on the analytics, we chose a simpler approach: sampling the distribution using Monte-Carlo techniques.

2.2.2 Computational models

Our simulations assume that each culture begins with a Poisson-distributed number of bacteria, with a mean number of 135, as in the original paper. The bacteria were modeled as dividing in discrete generations for a total of $g = 21$ generations. Both of the numbers are easily inferable from the original paper using the known growth rate and the final cell density numbers. Cells divide synchronously, and each of the daughters can gain a resistance mutation at division with the probability θ_D , which is nonzero in C and D models. Daughters of resistant bacteria are themselves resistant. Non-resistant cells in the final generation are subjected to a bacteriophage, which induces Lamarckian mutations with probability θ_L , nonzero in C and L models. We note again that this total number of Lamarckian-mutated cells is Poisson-distributed with the mean θ_L times the number of the remaining wild type bacteria.

To speed up simulations of the Darwinian process, we note that the total number of cells that have ever lived is $N_t = 2N_02^g - N_0$. Thus the total number of Darwinian mutation attempts is Poisson distributed, with mean $N_t\theta_D$. We generate the number of these mutations with a single Poisson draw and then distribute them randomly over the multi-generational tree of cells, marking every offspring of a mutated cell as mutated. We then correct for overestimating the probability of mutations due to the fact that the number of mutation attempts in each generation decreases if there are already mutated cells there. For this, we remove original mutations (and unmark

their progenies) at random with the probability equal to the ratio of mutated cells in the generation when the mutation appeared to the total number of cells in this generation. Note that since mutations are rare, such unmarking is not very common in practice, making this approach substantially faster than simulating mutations one generation at a time.

To estimate $P_C(k_p|\theta_D, \theta_L)$, we estimate this probability on a 41x41 grid of values of θ_D and θ_L . For each pair of values of these parameters, we perform $n = 30,000,000$ simulation runs (see below for the explanation of this choice) starting with a Poisson-distributed number of initial bacteria, then perform simulations as described above, and finally perform a simulated Poisson plating of a fraction of the culture if the actual experiment we analyze had such plating. We measure the number of surviving bacterial cultures k_p in each simulation run and estimate $P_C(k_p|\theta_D, \theta_L)$ as a normalized frequency of occurrence of this k_p across runs, $f_C(k_p|\theta_D, \theta_L)$. The Darwinian case is evaluated as $P_C(k_p|\theta_D, \theta_L = 0)$, and the Lamarckian case as $P_C(k_p|\theta_D = 0, \theta_L)$.

2.2.3 Quality of fit

In the original Luria and Delbrück publication [Luria and Delbrück, 1943], no definitive quantitative tests were done to determine the quality of fit of either of the model to the data. We can use the estimated values of $P_C(k_p|\theta_D, \theta_L)$ for this task. Namely, Luria and Delbrück have provided us not with frequencies of individual values of k_p , but with frequencies of occurrence of k_p within bins of $x \in (0, 1, 2, 3, 4, 5, 6 - 10, 11 - 20, 21 - 50, 51 - 100, 101 - 200, 201 - 500, 501 - 1000)$. By summing $f_C(k_p|\theta_D, \theta_L)$ over $k_p \in x$, we evaluate $f_C(x|\theta_D, \theta_L)$, which allows us to

write the probability that each experimental set of measurements, $\{n_x\}$, came from the model:

$$P(\{n_x\}|\theta_D, \theta_L) = \mathcal{C} \prod_x \left(\frac{n_x}{\sum_x n_x} \right)^{f_C(x|\theta_D, \theta_L)}, \quad (2.6)$$

where \mathcal{C} is the usual multinomial normalization coefficient. This probability can also be viewed as the likelihood of each parameter combination, and the peak of the probability gives the usual Maximum Likelihood estimation of the parameters [Nelson, 2015]. To guarantee that the estimated value of the likelihood has small statistical errors, we ensured that each of the (θ_D, θ_L) combinations has 30,000,000 simulated cultures. Then, at parameter combinations close to the maximum likelihood, each bin x has at least 10,000 samples. Correspondingly, at these parameter values, the sampling error in each bin is smaller than 1%.

Finally, to evaluate the quality of fit of a model, rather than to find the maximum likelihood parameter values, we calculate empirically the values of $\log_{10} P(\{n_x^*\}|\theta_D, \theta_L)$ for each parameter combination, where $\{n_x^*\}$ are synthetic data generated from the model with θ_D, θ_L . Mean and variance of $\log_{10} P$ gives us the expected range of the likelihood if the model in question fits the data *perfectly*.

2.2.4 Comparing models

In comparing the L, the D, and the C models, we run into the problem that C is guaranteed to have at least as good of a fit as either D or L since it includes both of them as special cases. Thus in order to compare the models quantitatively, we need to penalize C for the larger number of parameters (two mutation rates)

compared to the two simpler models. To perform this comparison, we use Bayesian model selection [MacKay, 1992, Mackay, 2003], which automatically penalizes for such model complexity.

Specifically, Bayesian model selection involves calculation of probability of an entire model family $M = \{L, D, C\}$ rather than of its maximum likelihood parameter values:

$$P(M|\{n_x\}) \propto \int d\vec{\theta}_M P(\vec{\theta}_M|\{n_x\}, M)P(M) \quad (2.7)$$

where the posterior distribution of $\vec{\theta}$ is given by the Bayes formula,

$$P(\vec{\theta}_M|\{n_x\}, M) \propto P(\{n_x\}|\vec{\theta}_M, M)P(\vec{\theta}_M|M), \quad (2.8)$$

and $P(\{n_x\}|\vec{\theta}_M, M)$ comes from Eq. (2.6). Finally, $P(M)$ and $P(\vec{\theta}_M|M)$ are the *a priori* probabilities of the model and the parameter values within the model, which we specify below.

The integral in Eq. (2.7) is over as many dimensions as there are parameters in a given model. Thus while more complex models may fit the data better at the maximum likelihood parameter values, a smaller fraction of the volume of the parameter space would provide a good fit to the data, resulting in an overall penalty on the posterior probability of the model. Thus posterior probabilities $P(M|\{n_x\})$ can be compared on equal footing for models with different number of parameters to say which specific model is *a posteriori* more likely given the observed data. Often the integral in Eq. (2.7) is hard to compute, requiring analytical or numerical approximations. However, here we already have evaluated the likelihood of combinations of

(θ_L, θ_D) over a large grid, so that the integral can be computed by direct summation of the integrand at different grid points.

To finalize computation of posterior likelihoods, we must now define the *a priori* distributions $P(M)$ and $P(\vec{\theta}_M|M)$. We choose $P(C) = P(M) = P(L) = 1/3$, indicating our ignorance about the actual process underlying biological evolution. The choice of $P(\vec{\theta}_M|M)$ is tricky, as is often the case in applications of Bayesian statistics. We point out that the experiment was designed so that the number of surviving mutants is almost always 1 or less, for a population with $\approx 0.25 \times 10^8$ individuals, which indicated that *a priori* both θ_L and θ_D are less than 4×10^{-9} . Further, we assume that, for the combined model, $P(\vec{\theta}_C) = P(\theta_L)P(\theta_D)$. Beyond this, we do not choose one specific form of $P(\vec{\theta}_C)$, but explore multiple possibilities to ensure that our conclusions are largely independent of the choice of the prior.

2.2.5 Statistical power of the tests

Before analyzing the original Luria-Delbrück data, it is important to understand the statistical power of our approach in discriminating among the models. Anticipating that the L model will be easy to rule out (see *Results*), we focus on disambiguating just the D and the C models by investigating the relationship between N , the number of cultures, and the expected ratio $P(D)/P(C)$ for synthetic data that resembles that of the Luria-Delbrück Experiment 22 (see *Results*), which is fitted well either with the D model with $\theta_D = 2.0 \times 10^{-9}$, or with the C model with $\theta_D = 1.8 \times 10^{-9}$, and $\theta_L = 0.4 \times 10^{-9}$.

First, we note that “ruling out” the Lamarckian contribution would require having $P(D)/P(C) \gg 1$. Thus we created synthetic data with a fixed Darwinian mutation rate ($\theta_D = 2.0 \times 10^9$) and zero Lamarckian contribution and explored how strong the evidence in favor of purely Darwinian model would be at different N . For this, we varied N and investigated the ratio $P(D)/P(C)$, using a prior $P(\theta_{L/D})$ that is uniform between 0 and 4×10^{-9} , as we later use for the analysis of the actual experiments. Figure 2.2(a) illustrates the dependence. Notice that $P(D)/P(C) \sim 20$, which would correspond to rejection of the Combined model at about 95% confidence, only for $N > 10^3$. This suggests that the Luria-Delbrück experiment was not designed well to achieve this demarcation.

Next we analyzed the number of cultures that would be needed to demonstrate existence of the Lamarckian influence as a function of θ_L . We defined the demonstration as $P(D)/P(C) < 1/20$, and then explored the number of cultures required to reach this threshold for different Lamarckian rates, while keeping $\theta_D = 1.8 \times 10^{-9}$, cf. Fig. 2.2(b). As expected, when the Lamarckian contribution is higher, the number of cultures decreases. However, crucially, even at $\theta_L = \theta_D$, one would require $N > 200$ cultures to demonstrate the Lamarckian contribution conclusively. This is because at $\theta_L = \theta_D$, the expected number of Lamarckian mutations in a culture (half that of the Darwinian ones) is < 1 . This again suggests that the Luria-Delbrück experiment should have had larger numbers of cells to be designed optimally for discriminating among these different models.

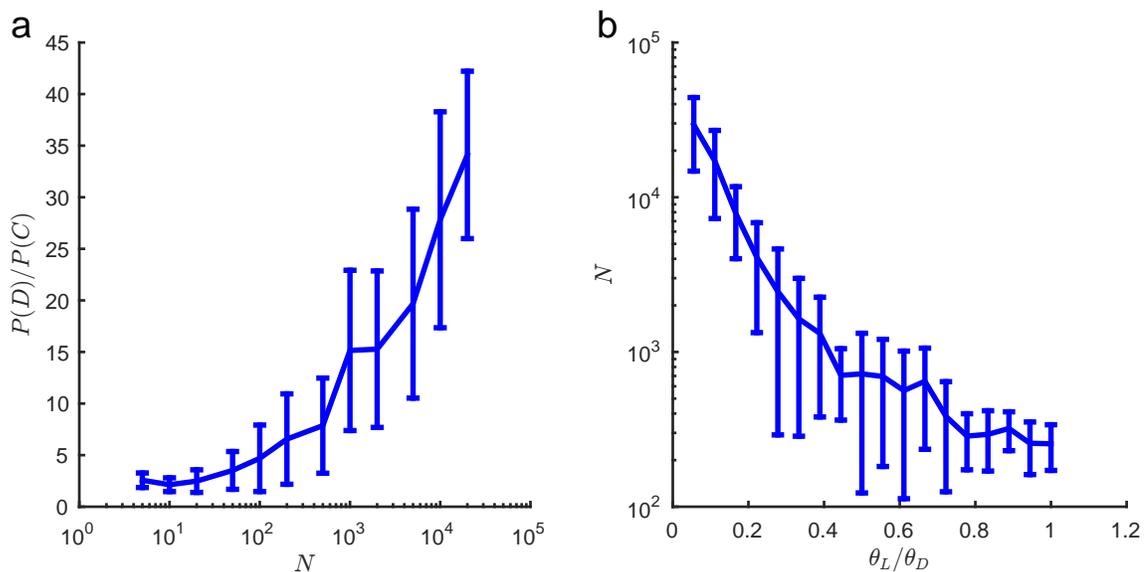


Figure 2.2: (a) For data with a known Darwinian mutation rate $\theta_D = 2.0 \times 10^{-9}$ and no Lamarckian mutations, we explore the power of the analysis to rule out the combined model. We see that the Combined model becomes improbable only at significantly higher values of N than those in the Luria-Delbrück experiment. (b) For data with a known Darwinian mutation rate $\theta_D = 1.8 \times 10^{-9}$ and a varying nonzero Lamarckian mutation rate, we found the minimum number of cultures N necessary to establish that there is a Lamarckian contribution (this is defined as $P(C)/P(D) \geq 20$).

2.3 Results

Luria and Delbrück’s paper provided data from multiple experiments, where in each experiment they grew a number of cultures, subjected them to the phage, and counted survivors. Most of the experiments have $O(10)$ cultures, which means that their statistical power for distinguishing different models is very low. We exclude these experiments from our analysis and focus only on experiments No. 22 and 23, which have $N = 100$ and $N = 87$ cultures, respectively. Our previous analysis suggests that even this is likely to be too few cultures for conclusive results, but these are the

numbers we have to work with. The experimental protocols differ in that Experiment 23 plated the entire culture subjected to the phage, while Experiment 22 plated only 1/4 of the culture. We analyze these experiments separately from each other.

2.3.1 Experiment 22

We evaluated the posterior probability of different parameter combinations, $P(\vec{\theta}_M|\{n_x\}, M)$, numerically, as described in *Methods*. The likelihood (posterior probability without the prior term) is illustrated in Fig. 2.3. Note that the peak of the likelihood is at $\theta_L^{22} \approx 4.0 \times 10^{-10} \neq 0$, $\theta_D^{22} \approx 1.8 \times 10^{-9}$, illustrating that the data suggests that the Combination model is better than either of the pure models in explaining the data, though the pure Darwinian model comes close. The fit of the maximum likelihood Combination model is shown in Fig. 2.4. The quality of the best fit $\log_{10} P(\{n_x\}|\theta_D^{22}, \theta_L^{22}) \approx 63.7$. This matches surprisingly well with the likelihood expected if the data was indeed generated by the model, $\log(P(\{n_x^{22}\}|\theta_D^{22}, \theta_L^{22})) = 62.1 \pm 5.1$. Thus the model fits the data perfectly despite numerous simplifying assumptions, suggesting no need to explore more complex physiological scenarios, such as asynchronous divisions, or different growth rates for mutated and non-mutated bacteria.

Next we evaluate the posterior probabilities of all three models by performing the Bayesian integral, Eq. (2.7). We use two different priors for θ_L and θ_D to verify if our conclusions are prior-independent: uniform between 0 and 4×10^{-9} and uniform

in the logarithmic space between 1×10^{-10} and 4×10^{-9} . For the uniform prior,

$$\frac{P(D)}{P(C)} \approx \frac{2.8}{1}, \quad \frac{P(L)}{P(C)} \approx 10^{-106}. \quad (2.9)$$

In other words, the purely Lamarckian model is ruled out by an enormous margin, as suggested in the original publication. However, in agreement with our estimate of the statistical power of the analysis, the ratio of posterior probabilities of the Darwinian and the Combination models is only 2.8, and this ratio is 2.0 for the logarithmic prior, which is way over the usual 5% significance threshold for ruling out a hypothesis. In other words,

The Darwinian and the Combination models of evolution have nearly the same posterior probabilities after controlling for different number of parameters in the models. Thus contribution of Lamarckian mechanisms to evolution in the Luria-Delbrück Experiment 22 cannot be ruled out.

2.3.2 Experiment 23

We performed similar analysis for Experiment 23 and evaluated the posterior likelihood, Fig. 2.5, for each combination of parameters. Here, however, the posterior is several orders of magnitude smaller than for Experiment 22. This is because the experimental data has a tail that is *heavier* than typical realizations of even the Darwinian model would predict. Indeed, Luria and Delbrück themselves noted this excessively heavy tail. However, as they were only choosing whether the Darwinian or the Lamarckian model fits better, this led further credence to the claim that the

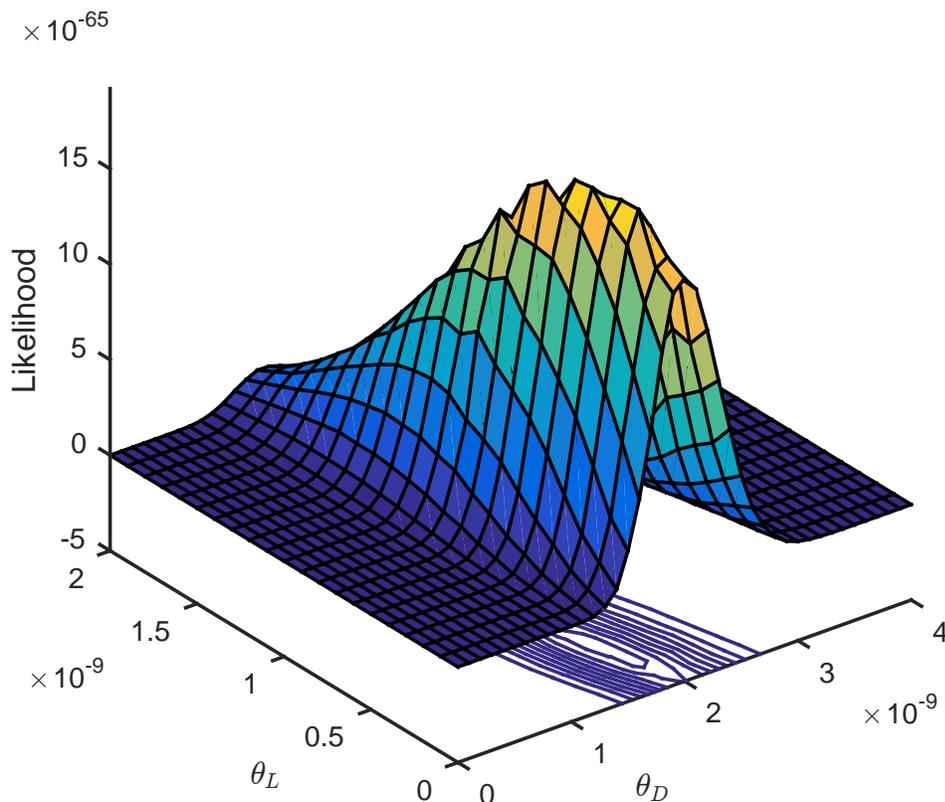


Figure 2.3: Posterior likelihood of the Darwinian, θ_D , and the Lamarckian, θ_L , mutation parameters evaluated for the Luria-Delbrück Experiment 22. Notice that the likelihood peaks away from $\theta_L = 0$.

Lamarckian model could not describe the data. Now we are able to quantify this: for Experiment 23, at the maximum likelihood parameters ($\theta_L^{23} = 0, \theta_D^{23} = 4.4 \times 10^{-9}$), the quality of fit is $\log_{10} P(\{n_x\}|\theta_D^{23}, \theta_L^{23}) \approx -90.0$. In contrast, for data generated from the model, we get $\log_{10} P(\text{data}|\theta_D, \theta_L) = -76.9 \pm 3.1$. In fact, by generating 10^5 data sets using these parameter combination, we estimate that the probability of generating data from this model that is as unlikely as the experimental data is $p < 10^{-4}$. Thus the tail of the distribution of the number of mutants in Experiment 23 is so heavy that it cannot be fit well by either of the hypotheses considered. In-

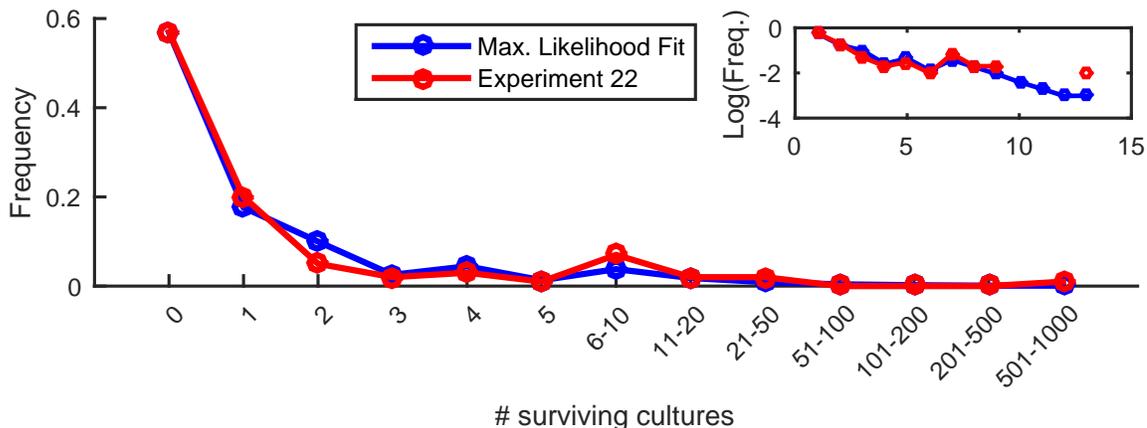


Figure 2.4: Luria-Delbrück experimental data (red) for Experiment 22 and the maximum likelihood fit of the Combination model with $\theta_L^{22} = 4.0 \times 10^{-10}$ and $\theta_D^{22} = 1.8 \times 10^{-9}$.

stead, it is likely that some other dynamics are at play here, such as some form of contamination, or additional non-Darwinian processes. In other words,

Luria-Delbrück Experiment 23 cannot be explained by any of the proposed hypotheses (the Lamarckian, the Darwinian, or the Combination one), and thus cannot be used to rule out one hypothesis over another.

2.4 Discussion

The classic 1943 experiment by Luria and Delbrück [1943] is credited with ruling out the Lamarckian model in favor of Darwinism for explaining acquisition of phage resistance in bacteria. However, while heralded as a textbook example of quantitative approaches to biology, the data in the paper was analyzed semi-quantitatively at best. We performed a *quantitative* analysis of the fits of three models of evolution (Lamarckian, Darwinian, and Combination) to these classic data, Experiments

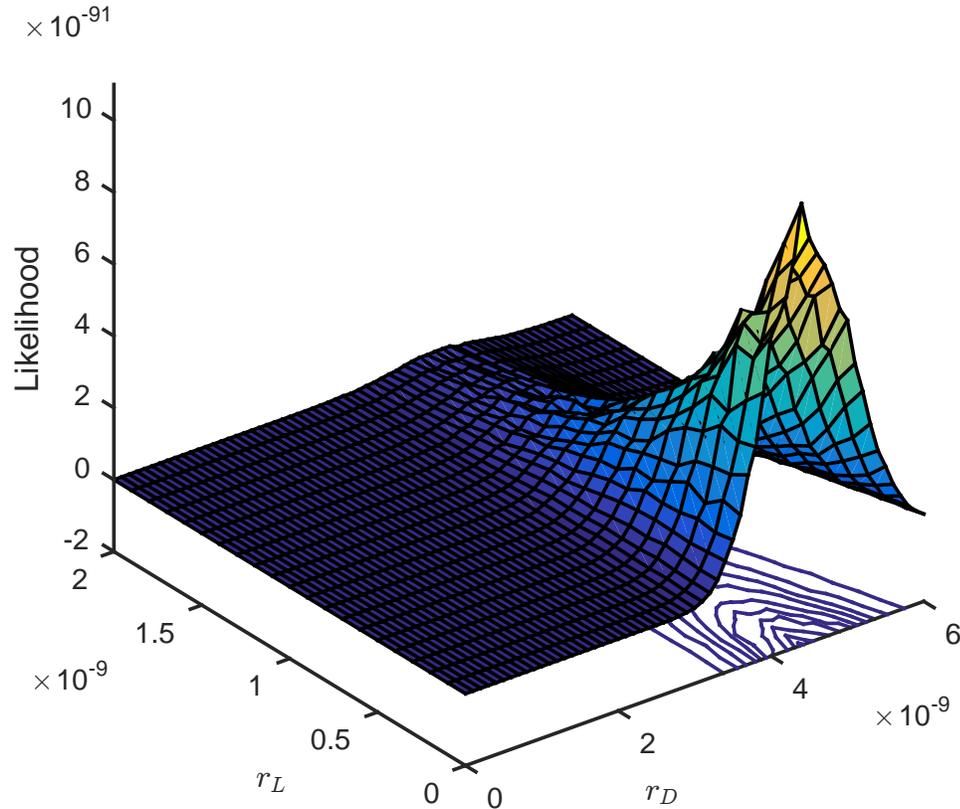


Figure 2.5: Posterior likelihood of the mutation parameters for Experiment 23. The maximum likelihood is at $\theta_L = 0$. However, neither of the three considered models is capable of fitting the data well (see text).

22 and 23. Our analysis was based on a very simplified model of the process: we started each colony with a Poisson-distributed (mean 135) wild-type bacteria and allowed them to replicate synchronously for exactly 21 times, with mutations occurring continuously (Darwinian model) or at the last generation (Lamarckian model). Additionally, we did not consider the possibility that multiple mutations might be needed to acquire resistance, or that growth rates of bacteria may be inhomogeneous. Nonetheless, the simple model fits Experiment 22 data perfectly, suggesting no need for more complex modeling scenarios.

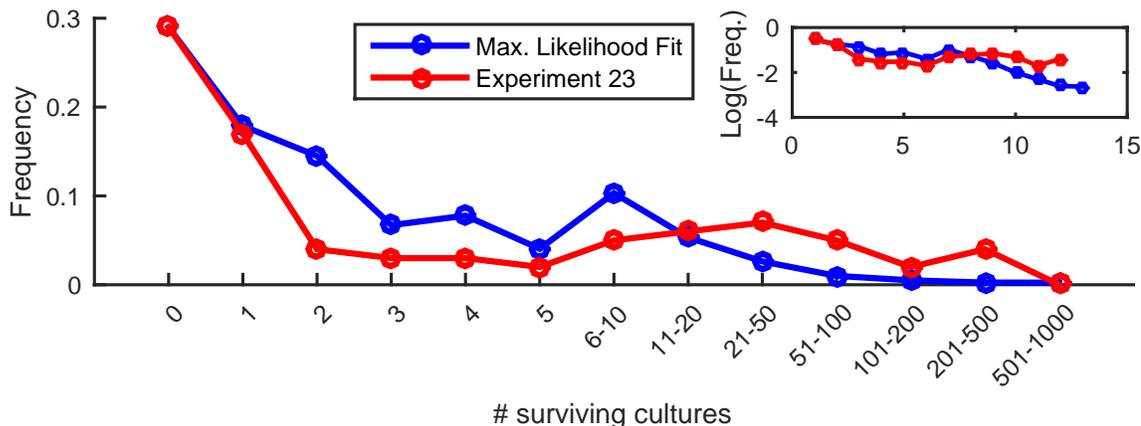


Figure 2.6: Luria-Delbrück experimental data (red) for Experiment 23 and the maximum likelihood fit of the Darwinian model (also the best fit Combination model) with $\theta_D^{23} = 4.4 \times 10^{-9}$. Here we can see that the tail in the experimental data is too heavy to be reproduced even by the Darwinian model.

For Experiment 22, by a ratio of $\approx 10^{-10^6}$, the Lamarckian model is a posteriori less likely than the Darwinian one, agreeing with the original Luria and Delbrück conclusion that the pure Darwinian evolution is a better explanation of the data than the pure Lamarckian evolution. However, the posterior odds of the pure Darwinian model are only 2 – 3 times higher than those for the Combination model (suitably penalized for model complexity), which has nonzero Darwinian *and* Lamarckian mutation rates. Even by liberal standards of modern day hypothesis testing, there is insufficient evidence to rule out the Combination model, and, therefore, contribution of Lamarckian processes to bacterial evolution in this experiment. This was in agreement with our analysis of the statistical power of the data: the number of cultures and the mean number of mutations per culture were too small to effectively discriminate between the D and the C model with a small Lamarckian rate. In contrast, for Experiment 23, neither of the three considered models could quantitatively explain the data, suggesting that additional processes must be in play beyond simple

Lamarckian and Darwinian mutations.

In summary, while subsequent experiments have certainly established the Darwinian nature of mutations in the T1-*E. coli* system, our analysis shows that the classic Luria-Delbrück 1943 data cannot be used to rule out Lamarckian contributions to bacterial evolution in favor of Darwinism, potentially necessitating changes to the exposition in many biology textbooks.

3 The expected time to cross extended fitness plateaus

3.1 Introduction

Most mutations in most natural populations are effectively neutral. Considered in isolation, these are irrelevant for adaptation. But the fitness effect of a mutation generally depends on the genetic background on which it occurs, a phenomenon known as epistasis. Thus, there are likely to be combinations of these neutral mutations that interact epistatically to have an effect on fitness. If this effect is positive for a given combination, then that combination forms a *complex adaptation*, separated from the wild type by a fitness plateau. How frequently do we expect populations to acquire such adaptations? On one hand, a given complex adaptation should typically be harder for a population to find than a simple adaptation requiring only a single beneficial mutation. On the other hand, if a genome of length L has $\mathcal{O}(L)$ possible neutral mutations, then there are $\mathcal{O}(L^K)$ genotypes that could potentially be a complex adaptation involving K mutations. So if even a modest fraction of these genotypes are indeed adaptive, the number of possible complex adaptations could far exceed the number of available beneficial mutations, and it could be that they are collectively a frequent form of adaptation [Fisher, 2007, Weissman et al., 2009, Trotter et al., 2014]. To evaluate their importance, we must know more about how rapidly populations explore fitness plateaus.

Populations can cross fitness plateaus via a sequence of neutral mutations fixing by drift until only one more mutation is needed for the (formerly complex) adapta-

tion. But this process is slow and inefficient; in a high-dimensional fitness plateau, the population will be much more likely to drift away from a complex adaptation than towards it. Large, asexual populations can cross plateaus and even fitness valleys much more rapidly (e.g., [van Nimwegen and Crutchfield, 2000, Komarova et al., 2003, Iwasa et al., 2004, Weinreich and Chao, 2005, Durrett and Schmidt, 2008, Weissman et al., 2009]). They can do this because many mutations will be present in the population at low frequency. If the population is sufficiently large, even these low-frequency mutations will be present in a large absolute number of individuals, some of which will happen to also carry additional mutations. Thus, genotypes that are multiple mutations away from the consensus genotype will already be present in the population and exposed to natural selection, allowing the population to effectively “see” several steps away in the fitness landscape, and “tunnel” directly to the adaptive genotypes [Jain and Krug, 2006].

Recombination changes these dynamics in two ways. First, by combining mutations that occur in different lineages, it accelerates the population’s exploration of the plateau [Christiansen et al., 1998]. On the other hand, recombination breaks up the beneficial combination once it is formed [Eshel and Feldman, 1970, Feldman, 1971, Karlin and McGregor, 1971], slowing adaptation [Takahata, 1982, Michalakis and Slatkin, 1996]. While the latter effect is fairly easy to understand quantitatively, the former depends on the spectrum of mutant lineages that coexist in the population and has only been fully understood in the simplest case of two-locus plateaus [Weissman et al., 2010]. Here we extend this analysis to the three-locus case, considering the full spectrum of possible population sizes, recombination rates, mutation rates, and selective advantages of the adaptive genotype. We also analyze the dynamics for

arbitrary-width plateaus when recombination is frequent relative to selection.

3.2 Model

We consider a haploid Wright-Fisher population of size N . The genome consists of K loci each of which has two possible alleles, 0 and 1; for much of the analysis, we will focus on the case $K = 3$. Initially, all individuals have the all-0 genotype. All genotypes have the same fitness except the all-1 genotype, which has a strong selective advantage $s \gg 1/N$; see Figure 3.1. Individuals mutate (in both directions) at a rate μ per locus per generation. Each generation, each offspring is produced clonally (with possible mutations) with probability $1 - r$; with probability r , it is the product of recombination between two parents. Recombinant offspring sample each locus independently with equal probability from their parents (again, with possible mutation). We will focus on finding the expected time \mathcal{T} until the all-1 genotype first makes up the majority of the population. For simplicity, we will also refer to the “rate” of crossing the plateau, defined as \mathcal{T}^{-1} , even though it is not a true rate, as the distribution of time to cross the plateau is not exponential in general. The definitions of the most important symbols are collected in Table 1. Exact simulations were done in Python (Figs. 3.3 and 3.4) and Mathematica (Figs. 3.5 and 3.6).

3.3 Results

There are two fundamentally different plateau-crossing dynamics, depending on the relative rates of selection and recombination [Eshel and Feldman, 1970, Feldman,

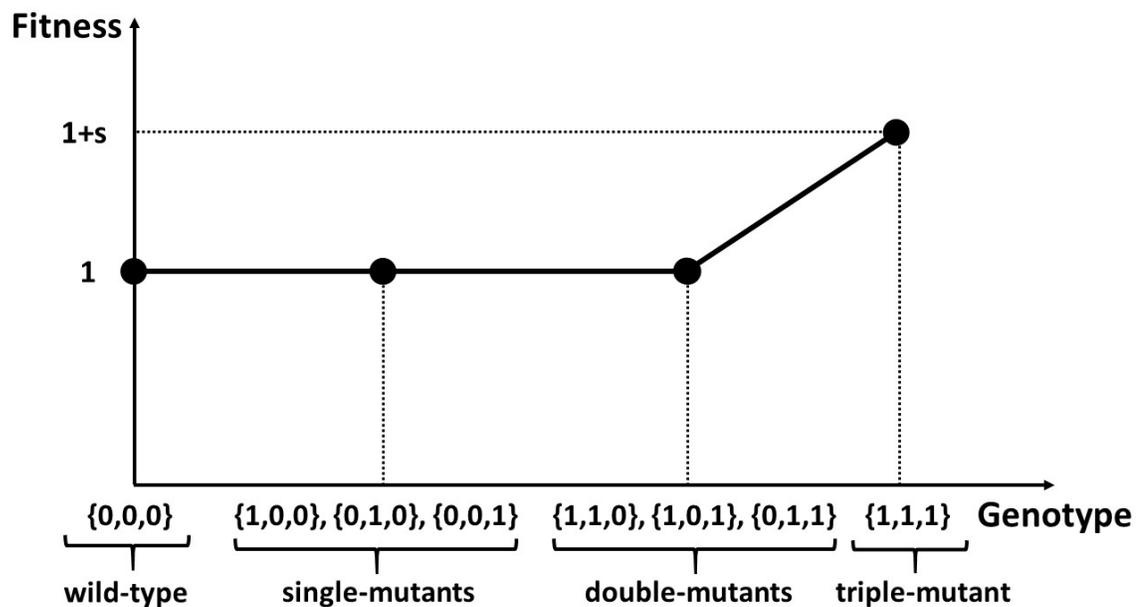


Figure 3.1: A visualization of the fitness landscape in the case of $K = 3$. The nodes represent the fitness of wild type, single-, double-, and triple-mutant genotypes. Wild-type alleles are denoted by 0 and mutants by 1. The $\{1, 1, 1\}$ genotype has a fitness $1 + s > 1$ and the rest have fitness 1.

Table 1: Symbol definitions

Symbol	Definition
N	Haploid population size
μ	Mutation rate per locus per generation
r	Recombination rate per generation
s	Selective advantage of the all-mutant genotype
\mathcal{T}	Expected time until the all-mutant genotype makes up the majority of the population
$R_i(t)$	The rate at which i -mutant individuals arise in the population at time t
$p_i(t)$	The probability that an i -mutant lineage arising at time t will be successful

Table 2: Rate of plateau-crossing (inverse of expected crossing time \mathcal{T}) for asymptotic regimes shown in Figure 3.2, organized by main divisions of parameter space.

Divisions	Name	\mathcal{T}^{-1} scaling	
$r < s$	$N\mu \gg 1$ Triple-mutants deterministic	$s/\ln(s/\mu)$	
	Double-mutants deterministic (asexual)	$(N\mu^3 s)^{1/3}$	
	Double-mutants deterministic (sexual)	$(N\mu^3 r s)^{1/4}$	
	Single-mutants deterministic (asexual)	$(N^2 \mu^5 s)^{1/4}$	
	Single-mutants deterministic (sexual)	$(N^2 \mu^5 r^3 s)^{1/7}$	
	$N\mu \ll 1$	Stochastic tunneling (asexual)	$N\mu^2 (s/\mu)^{1/4}$
		Stochastic tunneling (sexual)	$\sim N^{1.4} \mu^{1.9} r^{0.4} s^{0.1}$
		Semi-linkage-equilibrium tunneling	$N\mu^2 (N^3 \mu r s)^{1/4}$
		Tunneling after 1 mutation fixes (asexual)	μ
		Tunneling after 1 mutation fixes (sexual)	μ
$r \gg s$	$N\mu \gg 1$ Alleles deterministic	$(\mu^2 s)^{1/3}$	
	$N\mu \ll 1$ Linkage equilibrium (LE) tunneling	$N\mu^2 (N s)^{1/3}$	
	LE tunneling after 1 mutation fixes	μ	
$N \ll 1/\sqrt{\mu s}$	Sequential fixation	μ	

1971, Suzuki, 1997, Jain, 2009]. If recombination is weak relative to selection ($r \ll s$), the adaptive genotype is rarely broken up by recombination and can spread rapidly once formed even if the individual mutant alleles are very rare in the population. If, on the other hand, recombination is strong ($r \gg s$), the population is kept in quasi-linkage equilibrium, with the dynamics determined by the allele frequencies. Because the dynamics are so different, we consider these two regimes separately. Figure 3.2 and Table 2 summarize the different possible scaling behaviors of \mathcal{T} over all of parameter space.

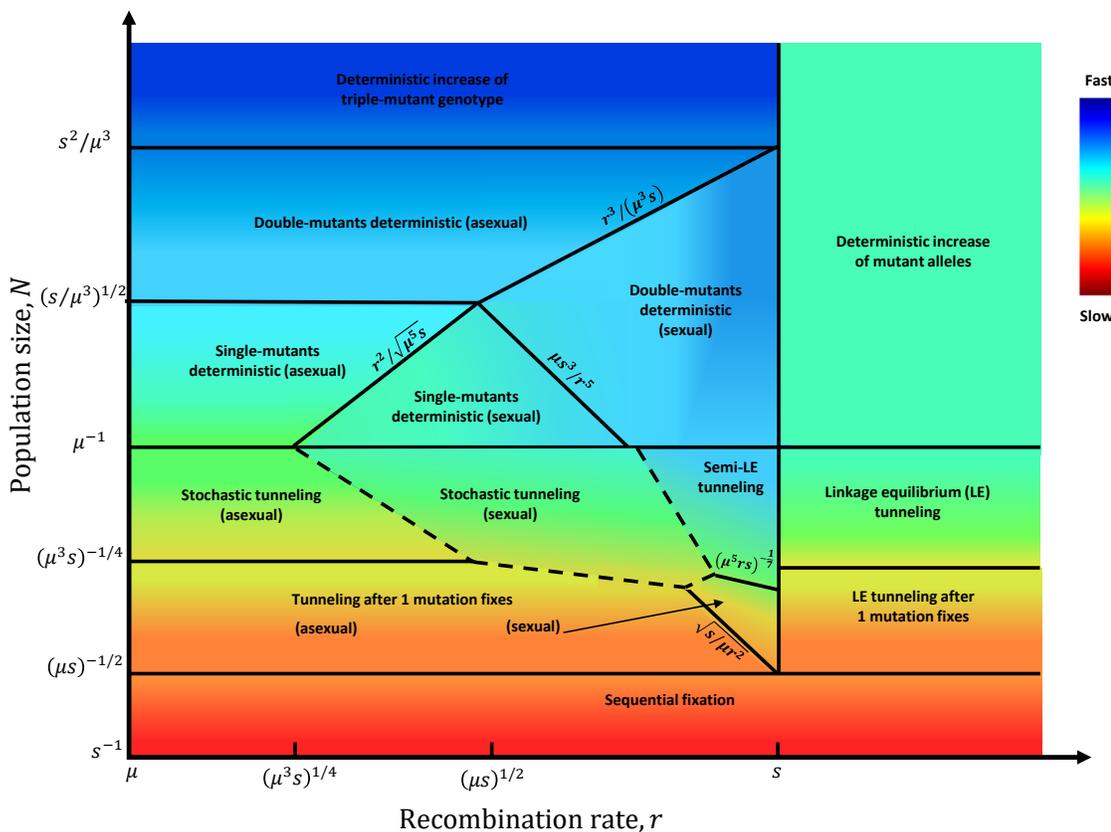


Figure 3.2: Schematic diagram of the asymptotic dynamics by which a population crosses a three-mutation ($K = 3$) fitness plateau to acquire a complex adaptation providing advantage s , as a function of recombination rate r and population size N . Color qualitatively represents the expected time \mathcal{T} for the population to cross; for quantitative expressions, see Table 2. For $r \ll s$, selection can drive the triple-mutant genotype to fixation while the other mutant genotypes remain rare, while for $r \gg s$ the population always remains close to linkage equilibrium; plateau crossing is fastest for intermediate recombination rates. The time to cross the plateau decreases as population size increases from the “sequential fixation” regime to the “deterministic” regimes. The “stochastic tunneling (sexual)” regime is a combination of several different regimes that can be practically indistinguishable, with boundaries that depend on the value of μ/s – see Appendix 3.5.

3.3.1 Rare recombination, $r \ll s$

For $r \ll s$, we must track genotype frequencies rather than just allele frequencies, so the complexity of the dynamics increases rapidly with the width of the plateau; we therefore focus on the simplest case that has not yet been fully characterized, $K = 3$. Even for this simplest case, there are many possible dynamical regimes (left half of Figure 3.2, Figure 3.3), depending on how difficult it is for the population to generate the adaptive genotype. For all but the largest population sizes, plateau-crossing becomes faster with increasing recombination rates, so the optimal rate is $r \lesssim s$. The equations in this section all apply to this regime as well, with s replaced by the average rate of increase of triple-mutants when rare, $\tilde{s} = s - r$.

If mutation and recombination are so frequent and the population is big enough that the triple-mutant genotype is generated effectively instantaneously, then the expected plateau-crossing time \mathcal{T} is just the time for a selective sweep, and depends primarily on s . For smaller μ , r , and N , most of \mathcal{T} is waiting for the successful triple-mutant to be produced and the strongest dependence is on μ . N and r are most important at intermediate levels of diversity, where producing triple-mutants is difficult but there are opportunities for simultaneous polymorphisms at multiple loci to recombine. Quantitatively, when the mutation supply is large ($N\mu \gg 1$), then the

expected plateau-crossing time is approximately:

$$\mathcal{T} \approx \begin{cases} \ln(s/\mu)/s & \text{if } (Nr\mu^3/s^3)^{1/4} \gg 1 \text{ or } N\mu(Ns)^{-2/3} \gg 1 \\ (N\mu^3rs)^{-1/4} & \text{if } r \gg (N\mu^3s)^{1/3}, (N^3\mu^5/(rs))^{1/4} \gg 1, (Nr\mu^3/s^3)^{1/4} \ll 1 \\ (N^2\mu^5r^3s)^{-1/7} & \text{if } r \gg (N^2\mu^5s)^{1/4} \text{ and } (N^3\mu^4r/s^2)^{1/7} \ll 1. \end{cases} \quad (3.1)$$

The first line of Equation 3.1 corresponds to the approximately deterministic dynamics of very large populations, which are insensitive to rare recombination (because the only substantial linkage disequilibrium is that generated by selection on the triple mutants after they are already on their way to fixation). In the second line, fluctuations in the number of triple-mutants are important, but single- and double-mutants can be treated deterministically (the “doubles deterministic (sexual)” regime). In the third line, fluctuations in the numbers of both triple- and double-mutants are important, but single-mutants can still be treated deterministically (the “single deterministic (sexual)” regime). If the recombination rate is lower than the thresholds in the second and third lines of Equation 3.1, the population is effectively asexual and \mathcal{T} follows the scaling behavior described in Weissman et al. [2009] and Equations 3.6 and 3.9 below.

If the mutation supply is low ($N\mu \ll 1$), then \mathcal{T} is approximately the expected waiting time for the first successful mutation. Since exploration of genotype space is more of a challenge for populations when mutations are rare, recombination has the potential to make more of a difference. When the recombination is very rare, the population is effectively asexual, with plateau-crossing rate $\mathcal{T}_{\text{asex}}^{-1} \approx N\mu^2(s/\mu)^{1/4}$ (Equation 3.10, see also Weissman et al. [2009]). As the recombination rate increases,

it becomes easier mutations to be successful, and plateau-crossing speeds up. There are eight different asymptotic scaling regimes for rare recombination as $N \rightarrow \infty$, depending on exactly how $\mu, r, s \rightarrow 0$, but for reasonable parameter values they are generally fairly similar (see Appendix 3.5), with the expected rate of plateau-crossing roughly given by $\mathcal{T}^{-1} \approx N^{1.4} \mu^{1.9} r^{0.4} s^{0.1}$. As recombination becomes more frequent (but still $r \ll s$), pairs of large single-mutant lineages are able to succeed by reaching linkage equilibrium with each other and then recombining with a smaller third lineage (“semi-linkage-equilibrium tunneling”), and the rate of crossing increases further to $\mathcal{T}^{-1} \approx (N\mu)^2 (\mu r s / N)^{1/4}$ (Equation 3.11). This is the regime where recombination speeds plateau-crossing the most; comparing Equations 3.10 and 3.11, we see that it increases the rate by a factor $\sim (N^3 \mu^2 r)^{1/4}$, which could exceed an order of magnitude if $Nr > 10^6$.

3.3.2 Frequent recombination, $r \gg s$

For frequently recombining populations ($r \gg s$), we find the expected time for plateau crossing across the full spectrum of possible plateau widths K , mutation rates μ , population sizes N , and selective coefficients s (Figure 3.4). These population will be in quasi-linkage equilibrium and selection will therefore act on alleles rather than genotypes. In this regime, the plateau-crossing time depends primarily on the mutation rate and is typically $\sim \mathcal{O}(1/\mu)$. When mutations are frequent ($N\mu \gg 1$), the population crosses the plateau nearly deterministically and solving the deterministic mutation-selection dynamics gives plateau-crossing time: $\mathcal{T} \approx (s/\mu)^{1/K} / \mu$.

When mutations are rare ($N\mu \ll 1$), stochasticity is important and the dynamics

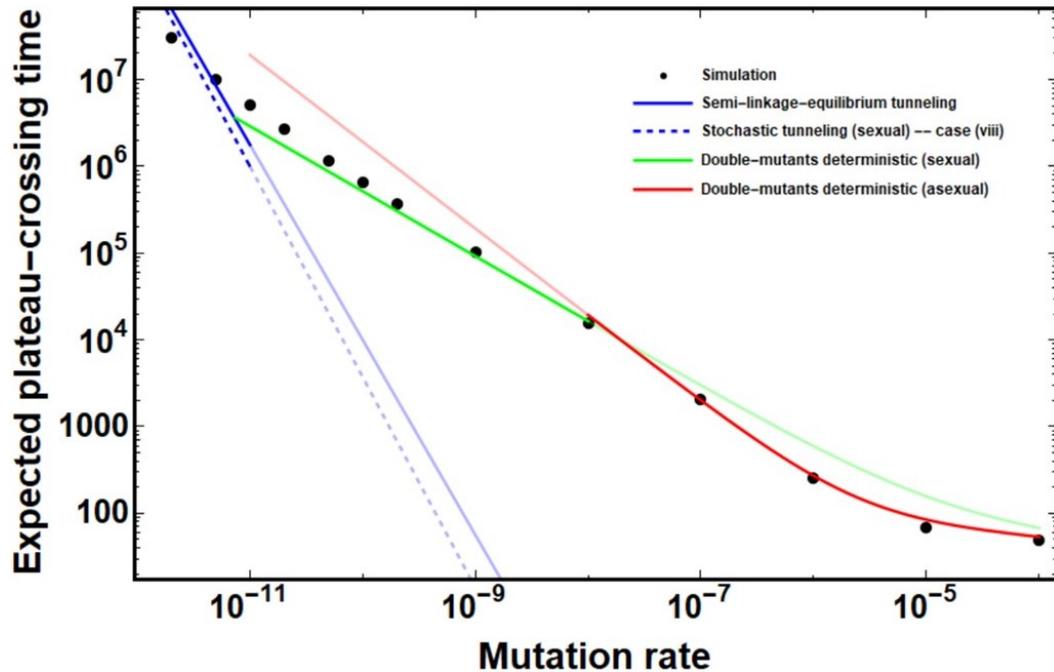


Figure 3.3: Expected time \mathcal{T} for a rarely recombining population to cross a three-mutation fitness plateau, as a function of the mutation rate μ . Points show simulation results, curves show analytical predictions. Note that in some regions, several analytical expressions give almost the same exact expected time. Typical plateau-crossing dynamics of the different asymptotic regimes are illustrated in Figures 3.5 and 3.6. The analytical solution for the double-mutants deterministic asexual and sexual paths (red and green), semi-linkage-equilibrium tunneling (solid blue) and single-mutant stochastic tunneling (blue dashed line) regimes is given by Equations 3.6, 3.11, and A.9, respectively. Parameter values: $N = 10^{11}$, $r = 10^{-3}$, $s = 1$. Error bars are smaller than the size of the points.

typically proceed in two stages: first, $K - m$ of the necessary mutations sequentially drift to fixation by chance; then, once the population is sufficiently close to the adaptive genotype, it relatively rapidly acquires the last m mutations together via stochastic tunneling. The typical value of m is the largest integer such that the probability of a new mutation triggering a tunneling event of m mutations is higher than the probability $1/N$ of fixation by drift:

$$m \approx \left\lfloor \frac{1}{2} + \sqrt{\frac{1}{4} - \frac{2 \ln(Ns)}{\ln(N\mu)}} \right\rfloor, \quad (3.2)$$

where $\lfloor \cdot \rfloor$ represents the floor function. Therefore, the plateau-crossing time is typically dominated by the time for $K - m$ mutations to drift to fixation, unless $m \geq K$, in which case the population tunnels directly. Summarizing these regimes, the expected time for a frequently recombining population to cross a fitness plateau is:

$$\mathcal{T} \approx \begin{cases} \frac{1}{\mu} \left(\frac{\mu}{s}\right)^{1/K} & \text{if } N\mu \gg 1 \\ N(Ns)^{-\frac{1}{K}} (N\mu)^{-\frac{K+1}{2}} C_K & \text{if } N\mu \ll 1 \text{ and } m \geq K \\ \frac{\ln(K/m)}{\mu} & \text{if } N\mu \ll 1 \text{ and } m < K, \end{cases} \quad (3.3)$$

where C_K in the second line (pure tunneling) is a combinatorial factor that depends only on K (see Equation 3.16).

Comparing the Equation 3.3 to the expected time for an asexual population to cross the plateau ($\mathcal{T} \approx \frac{1}{N\mu^2} (s/\mu)^{-2^{1-K}}$ for $N \ll 1/\mu$, $\mathcal{T} \approx \frac{K}{s} \ln(s/\mu)$ for $N \gg s^{K-1}/\mu^K$, with additional asymptotic regimes for intermediate population sizes [Weissman et al., 2009]), we see that frequent recombination tends to speed up adaptation in small populations (relative to asexuality), where the primary challenge is produc-

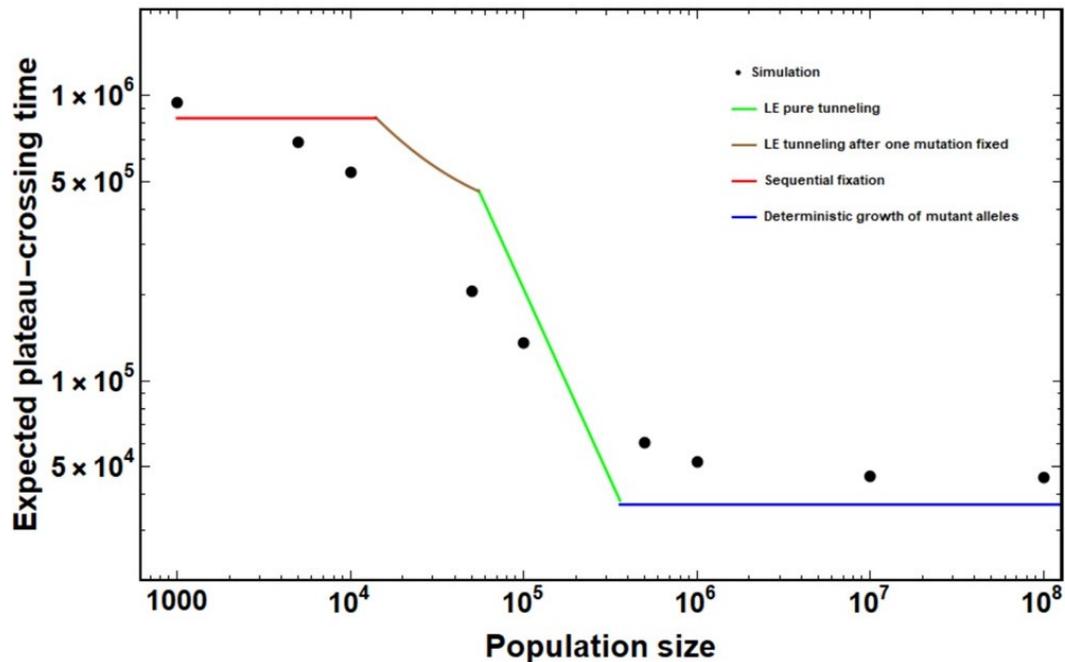


Figure 3.4: Expected time \mathcal{T} for a frequently recombining population to cross a three-mutation fitness plateau, as a function of the population size N . Points show simulation results, curves show analytical predictions Equation 3.13 (blue) and Equation 3.18 (green, brown, and red). Parameter values: $r = 0.5$, $\mu = 10^{-6}$, $s = 0.05$. The time to cross the plateau depends strongly on N for $N \lesssim 1/\mu$, and levels off for large or very small populations. Error bars are smaller than the size of the points.

ing the beneficial genotype, while slowing it down in large populations, where most of the time is spent fixing the genotype after it has been produced.

3.4 Analysis

3.4.1 Rare recombination ($r \ll s$)

In this section we will consider the plateau-crossing process in populations with rare recombination, starting with very large populations and progressively decreasing in size. As N decreases, the population's ability to efficiently explore genotype space (measured by N , μ , and r) becomes more important, and its ability to exploit its discoveries (s) less so. At the largest population sizes, \mathcal{T} is essentially determined by s . For all the lower population size regimes, there will be at least some genotypes that are only rarely produced, and \mathcal{T} will be approximately the waiting time for the production of the first successful lineage of a rare genotype.

3.4.1.1 Very large populations: deterministic dynamics

For extremely large population sizes, the number of single-, double-, and triple-mutant individuals are well approximated by their expected values after only a few generations. Triple mutants are produced almost instantaneously, and the plateau-crossing time is dominated by the time it takes them to sweep to fixation. This can easily be found by solving the deterministic equations for the dynamics of the genotype frequencies under mutation and selection, with recombination only reducing the effective selective advantage of the triple mutants:

$$\mathcal{T} \approx \frac{3}{s} \ln \left(\frac{s}{\mu} \right), \quad (3.4)$$

It is straightforward to generalize Equation 3.4 to arbitrary plateau widths K :

$$\mathcal{T} \approx \frac{K}{s} \ln \left(\frac{s}{\mu} \right).$$

For this deterministic approximation to be accurate, the production rate of triple mutants, $R_3(t)$, must be large at the time $t \sim 1/s$ when the first triple-mutant lineage reaches number $1/s$ and becomes established. Triple mutants are produced by double-mutant individuals that either acquire another mutation or recombine with single mutants, so R_3 is:

$$R_3(t) = \mu n_2(t) + \frac{r}{24N} n_1(t) n_2(t), \quad (3.5)$$

where n_i is the total number of i -mutant individuals, so $n_1(t) \approx 3N\mu t$ and $n_2(t) \approx 3N\mu^2 t^2$. (The factor of $1/24$ in Equation 3.5 is because to make a triple mutant, each double mutant can only successfully recombine with $1/3$ of the single mutants, and only $1/8$ of the offspring will inherit the correct alleles.) At $t \sim 1/s$, Equation 3.5 gives $R_3 \approx 3N\mu^3/s^2$. (The recombination term is smaller by a factor $\sim \mathcal{O}(r/s)$.) So to be in this regime, the population must have size $N \gg s^2/\mu^3$. Note that recombination is almost irrelevant in this regime: mutants are produced so frequently that there is no need for recombination to generate new combinations. (It does slightly slow down adaptation, as technically s should be replaced in the equations by $\tilde{s} = s - r$.)

3.4.1.2 Large populations: single- and double-mutants common, triple-mutants rare

Slightly smaller populations (with $N \ll s^2/\mu^3$) will still only occasionally be producing triple-mutants at the time they cross the plateau, so while the single- and double-mutant populations will have nearly deterministic dynamics, fluctuations in the number of triple-mutants will be important. Because triple-mutant lineages are rare, we can consider them in isolation, and \mathcal{T} will be the waiting time for the first successful triple-mutant. The probability that a successful triple-mutant lineage will have been produced by time t is $P_3(t) = 1 - \exp\left[-\int_0^t dt' s R_3(t')\right]$, using that the probability that a triple-mutant lineage is successful once it has been produced is $\sim s$ [Ewens, 2004]. Therefore, the expected waiting time \mathcal{T} is given by:

$$\begin{aligned} \mathcal{T} &= \int_0^\infty dt \exp\left[-s \int_0^t dt' R_3(t')\right] \\ &\approx \begin{cases} (N\mu^3 s)^{-1/3}, & r \ll (N\mu^3 s)^{1/3} \text{ (asexual path)} \\ (N\mu^3 r s)^{-1/4}, & r \gg (N\mu^3 s)^{1/3} \text{ (sexual path)}, \end{cases} \end{aligned} \quad (3.6)$$

where we are ignoring constants of $\mathcal{O}(1)$. In the first line, the population is effectively asexual, i.e., the successful triple-mutant is likely to arise via mutation from a double-mutant. In the second line, it is more likely to arise via recombination between a double-mutant and a single-mutant. The two expressions in Equation 3.6 are generally close, differing by a factor of only $\sim (r^3/(N\mu^3 s))^{1/2}$: recombination can provide a mild increase in speed, but as in the previous section, the population is so large that the triple-mutant genotype will rapidly be produced by mutation alone

anyway.

In deriving Equation 3.6, we ignored fluctuations in the number of double-mutants (as well as of single-mutants), so these must be negligible on time scales similar to \mathcal{T} for the expressions to be valid. There are two ways that this approximation can hold. First, if the number of double-mutants is in fact close to its expectation with high probability. This will be true if the production rate $R_2(t)$ of double-mutants is high, i.e., $R_2(\mathcal{T}) \gg 1$, so that the double-mutant population is composed of many lineages and fluctuations in the individual lineage sizes average out. R_2 is given by:

$$R_2(t) = 2\mu n_1(t) + \frac{r}{12N} n_1(t)^2 \quad (3.7)$$

Plugging in the first line of Equation 3.6 for t gives the requirement $R_2(\mathcal{T}) \sim \mu (N^2/s)^{1/3} \gg 1$

In the recombination-dominated regime in the second line of Equation 3.6, there is an additional way for the fluctuations to be negligible: recombination can cap their size by preventing them from greatly exceeding linkage equilibrium with the much larger and approximately deterministic wild-type and single-mutant populations. In this case, if $R_2 \ll 1$ the number of double-mutants will be fluctuating as lineages are sporadically produced and die out, but no one lineage will drift for a time much exceeding $\sim 1/r$ before being broken up by recombination. We will see what condition this puts on the parameters in the following section, but for now, note that this mechanism requires the single-mutants to be approximately deterministic, so at a minimum we require $R_1 \sim N\mu \gg 1$.

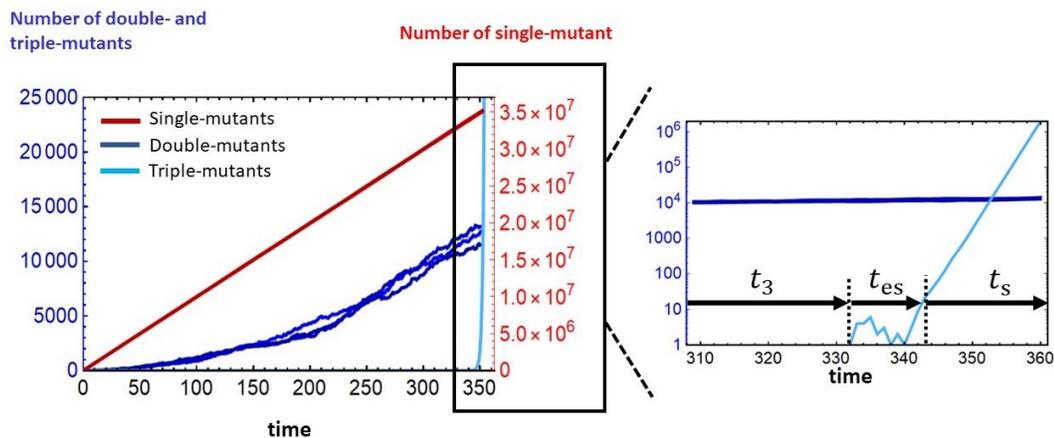


Figure 3.5: Typical simulation results of plateau-crossing dynamics for very large population sizes where \mathcal{T} is dominated by the waiting time for the arrival of the first successful triple-mutant individual that has been produced following a mutation event in the growing population of double-mutants. The inset is a magnified view of the last 50 generations before the adaptive genotype fixes in the population which demonstrates the establishment time and sweep time of the triple-mutants. The model parameters for this simulation are $N = 10^{11}$, $\mu = 10^{-6}$, $r = 10^{-3}$, and $s = 1$.

Finally, we must also check the conditions for our assumption that the triple-mutant lineages are rare enough to be considered in isolation. This is equivalent to $R_3(\mathcal{T}) \ll 1$ – the flip side to the parameter condition in the preceding section requiring that triple-mutants be approximately deterministic. Plugging in our expressions for R_3 and \mathcal{T} , we get $\mu(N/s^2)^{1/3} \ll 1$, which is indeed the reverse of the previous condition.

3.4.1.3 Moderately large populations: single-mutants common, double-mutants rare

For populations slightly smaller than those in the previous section, the mutation supply will still be high ($N\mu \gg 1$), so single-mutants will still be approximately deterministic, but double-mutant lineages will be rare ($R_2(\mathcal{T}) \ll 1$) and we must consider their fluctuations. Since they are rare, we can consider each lineage in isolation, and \mathcal{T} will be the waiting time for the first successful double-mutant to arise.

A double-mutant lineage can be successful by either mutating or recombining with single-mutants to produce a successful triple-mutant. Since the single-mutants are deterministic ($n_1(t) \approx 3N\mu t$), we can lump these two processes into a single time-dependent effective mutation rate, $\tilde{\mu} \equiv \mu + \frac{r}{24N}n_1(t) \approx \mu \left(1 + \frac{1}{8}rt\right)$. Since in this regime we expect the waiting time for the first successful double-mutant lineage to be long compared to the time for which that lineage must drift before producing the successful triple-mutant, we can further treat this effective mutation rate as being approximately constant over each lineage's lifetime. With this approximation, the problem is reduced to that considered in Weissman et al. [2009]: a lineage mutates at rate $\tilde{\mu}$ to a genotype with advantage s ; additionally, the double-mutant lineage has an effective selective disadvantage r due to being broken up by recombination with the wild type. The probability $p_2(t)$ that a double-mutant lineage arising at time t

will be successful is therefore Weissman et al. [2009]:

$$p_2(t) \approx \begin{cases} \sqrt{\mu s \left(1 + \frac{1}{8}rt\right)} & \text{if } \mu s \left(1 + \frac{1}{8}rt\right) \gg r^2 \\ \frac{\mu s}{r} \left(1 + \frac{1}{8}rt\right) & \text{if } \mu s \left(1 + \frac{1}{8}rt\right) \ll r^2. \end{cases} \quad (3.8)$$

In the first line of Equation 3.8, a lineage is most likely to succeed by drifting for long enough to produce many ($\sim 1/s$) triple-mutants. In the second line, recombination is too frequent and lineages are broken up before they can drift for that long. We therefore see that the condition for being able to ignore fluctuations in the double-mutant numbers as in the previous section is $\mu s \left(1 + \frac{1}{8}r\mathcal{T}\right) \ll r^2$. Since this case is covered by that section's analysis, we will now focus on the case $p_2(t) \approx \sqrt{\mu s \left(1 + \frac{1}{8}rt\right)}$ where fluctuations are key. Combining the success probability $p_2(t)$ with the production rate $R_2(t) \approx N\mu^2t \left(2 + \frac{3}{4}rt\right)$, we can find the expected waiting time \mathcal{T} for the first successful double-mutant (ignoring $\mathcal{O}(1)$ constants):

$$\begin{aligned} \mathcal{T} &\approx \int_0^\infty \exp\left[-\int_0^t dt' R_2(t') p_2(t')\right] \\ &\approx \begin{cases} (N^2\mu^5s)^{-1/4} & r \ll (N^2\mu^5s)^{1/4} \text{ (asexual path)} \\ (N^2\mu^5r^3s)^{-1/7} & r \gg (N^2\mu^5s)^{1/4} \text{ (sexual path)}. \end{cases} \end{aligned} \quad (3.9)$$

Both R_2 and p_2 switch from being mutation-dominated to recombination-dominated at time $t \sim 1/r$. In the first line of Equation 3.9, $\mathcal{T} \ll 1/r$ so the population is effectively asexual. In the second line, $\mathcal{T} \gg 1/r$ so the successful double-mutant is likely both to be produced by recombination and to produce the successful triple via recombination.

3.4.1.4 Moderately small populations: occasional triple polymorphisms

If the mutation supply is low ($N\mu \ll 1$), then the population will typically be monomorphic. The plateau-crossing time is dominated by the waiting time for a lucky single-mutant lineage that drifts long enough to either fix or encounter additional mutations that allow it to tunnel across the plateau. We will consider the latter process in this section. Call this mutation A , and let T_A be the time for which this mutation's lineage must drift to be likely to be successful; over this time, the lineage will typically reach a size $n_A \sim T_A$. The mutation will manage to drift this long with probability $p_1 \sim 1/T_A$, so the expected plateau-crossing time is $\mathcal{T} = \frac{1}{3N\mu p_1} = \frac{T_A}{3N\mu}$. Note that if $T_A > N$ (or, equivalently, $\mathcal{T} > 1/\mu$), the lineage is more likely to fix than tunnel. We now find expressions for the necessary drift time T_A .

First, we will review the asexual process. Weissman et al. [2009] showed that $T_A \sim (\mu^3 s)^{-1/4}$ (ignoring combinatoric factors) is long enough for the lineage to be likely to acquire two additional mutations (which we will call B and C) and be successful. The expected time to cross the plateau is thus:

$$\mathcal{T} \sim \frac{1}{N\mu^2} \left(\frac{\mu}{s}\right)^{1/4}. \quad (3.10)$$

Comparing T_A to N , we see that the population will only tunnel if $N(\mu^3 s)^{-1/4} > 1$. This result therefore applies only to populations within a fairly narrow band of sizes, with the lower limit of validity only a factor of $(\mu s)^{1/4}$ smaller than the upper limit – less than three orders of magnitude for realistic parameters.

Recombination can speed up tunneling (i.e., reduce the necessary T_A) by allowing

the original A lineage to acquire B and C from the $\sim N\mu T_A$ independent mutant lineages that arise while it is drifting. Let B be the mutation carried by the largest such lineage; it will typically drift for $T_B \sim N\mu T_A$ generations, reaching a size $n_B \sim N\mu T_A$. If $T_B \gg 1/r$, recombination will effectively reduce the linkage disequilibrium between A and B , i.e., there will be an average of $n_{AB} \sim n_A n_B / N \approx \mu T_A^2$ AB individuals over most of the T_B generations for which both mutations are drifting. During this time, there will be $\sim N\mu T_B$ C lineages produced by mutation, the largest of which will therefore typically drift for $T_C \sim N\mu T_B \approx (N\mu)^2 T_A$ generations to size $n_C \sim T_C$. AB and C individuals will therefore coexist for $\sim T_C$ generations, during which they will generate $\sim \frac{r}{N} n_{AB} n_C T_C \approx N^3 \mu^5 r T_A^4$ triple-mutant recombinants. Each of these has a probability $\approx s$ of being successful, so we see that for our original A lineage to be likely to be successful, its drift time T_A must satisfy $N^3 \mu^5 r T_A^4 \sim 1/s$. Solving for T_A gives $T_A \sim (N^3 \mu^5 r s)^{-1/4}$, corresponding to an expected plateau-crossing time of:

$$\mathcal{T} \sim \frac{1}{(N\mu)^2} \left(\frac{N}{\mu r s} \right)^{1/4}. \quad (3.11)$$

We refer to this as “semi-linkage-equilibrium tunneling”, since the two most frequent mutations are in linkage equilibrium with each other while drifting, but the third mutation may not be, and the triple-mutant will produce large linkage disequilibria once it starts to sweep.

The derivation of Equation 3.11 assumed that A and B were close to being in linkage equilibrium with each other, i.e., $rT_B \gg 1$. Substituting in $T_B \sim N\mu T_A$, this is equivalent to a condition that $T_A \gg 1/(N\mu r)$. However, it may be the case that the A and B lineages can produce enough recombinants to be successful before they

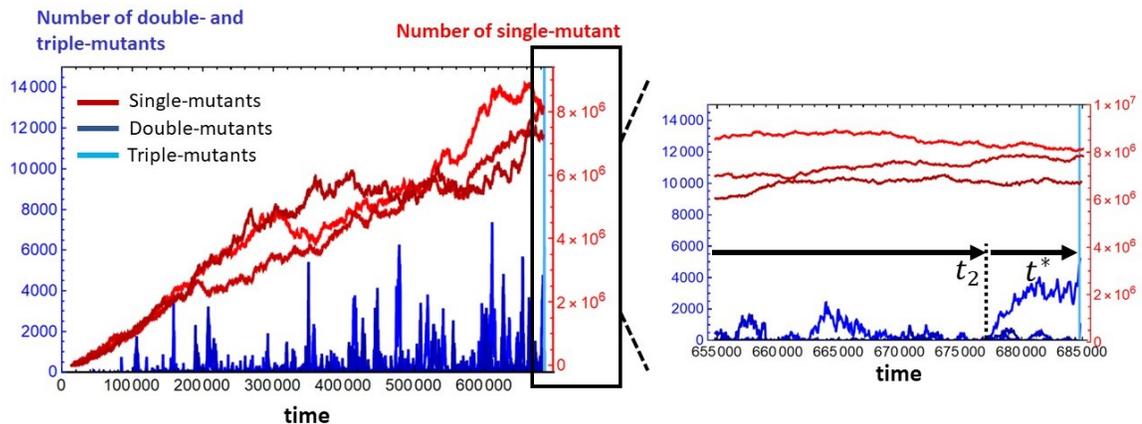


Figure 3.6: Typical simulation results of plateau-crossing dynamics for moderately large population sizes where \mathcal{T}_2 is dominated by the waiting time for the arrival of the first successful double-mutant lineage t_2 (indicated on the inset) that has been produced following a recombination event between two single-mutants. The inset is a magnified view of the last few thousands of generations before the adaptive genotype fixes in the population which demonstrates that while the successful double-mutant lineage is drifting for t^* generations, the population of single-mutants has, to a very good approximation, remained constant. The model parameters for this simulation are $N = 10^{11}$, $\mu = 5 \times 10^{-10}$, $r = 10^{-3}$, and $s = 1$.

approach linkage equilibrium. This is true for small values of r , where the time to approach linkage equilibrium becomes very long. In this situation, the analysis here overestimates how large T_A must be, and therefore overestimates the time required to cross the plateau. The correct analysis of this regime is even more involved, and we leave it for Appendix 3.5. We also ignored the possibility that the AB individuals might produce a triple mutant directly via mutation, but it is straightforward to check that this is rare in the relevant parameter range: as long as $N \gg 1/\sqrt{\mu r}$, acquiring the third mutation via recombination is more likely.

3.4.1.5 Small populations: single-mutants drift to fixation

For smaller populations, the most likely way for a single-mutant to be successful is for it to drift to fixation, which occurs with probability $1/N$. The expected waiting time is therefore $\mathcal{T} = 1/(3\mu)$. Once the single-mutant has fixed, the population only needs two additional mutations, so Weissman et al. [2010]’s two-locus analysis applies. The average time to tunnel will necessarily be small compared to the time for the first mutant to drift to fixation, so it can be neglected in \mathcal{T} . The exception is for very small populations, $N \ll 1/\sqrt{\mu s}$, where the *second* mutation is also more likely to drift to fixation than to tunnel [Weissman et al., 2010]. In this case, the total waiting time is $\mathcal{T} \approx \frac{1}{3\mu} + \frac{1}{2\mu} = \frac{5}{6\mu}$. (The final fixation of the third mutation is relatively rapid as long as $Ns \gg 1$.)

3.4.2 Frequent recombination ($r \gg s$)

If recombination is frequent ($r \gg s$), selection will be too weak to generate linkage disequilibrium, and the population will stay close to linkage equilibrium (LE). We can therefore simply track allele frequencies, rather than genotype frequencies. For this much easier problem, we can consider plateaus of arbitrary width K .

3.4.2.1 Large populations ($N\mu \gg 1$): deterministic dynamics

When the mutation supply is large, $N\mu \gg 1$, the mutant allele frequency trajectories are nearly deterministic, and therefore almost the same as each other, i.e., a single

variable $x(t)$ can describe the frequency of all the mutant alleles. When the mutations are rare ($x \ll 1$), their increase according to:

$$\dot{x} \approx \mu + sx^K, \quad (3.12)$$

where x^K is the frequency of the beneficial genotype. Solving Equation 3.12 for t such that $x(t) \approx 1$ gives the time to cross the plateau:

$$\mathcal{T} \approx \frac{1}{\mu} \left(\frac{\mu}{s} \right)^{1/K}. \quad (3.13)$$

We can understand this as the time it takes for mutation to drive the mutations to the frequency $x \sim (\mu/s)^{1/K}$ at which selection takes over, after which fixation is rapid.

3.4.2.2 Small populations ($N\mu \ll 1$): sequential fixation + stochastic tunneling of mutant alleles

When the mutational supply of the population is small ($N\mu \ll 1$), most loci will usually be monomorphic, with occasional drifting mutant lineages. To cross the plateau, the population needs some combination of mutations drifting to fixation, and others producing the beneficial genotype and tunneling together. We can think of the tunneling dynamics as allowing the population to “see” the adaptive genotype once the dominant genotype is within m mutations of it, for some m .

We must first find how the maximum tunneling range m depends on N , μ , and s . A population can cross the plateau via a rare mutant lineage that grows to a

large size over an extend period of time. Suppose that such a lineage persists for $\sim T_1$ generations, typically growing to size $\sim T_1$. There will be $\sim (m-1)N\mu T_1$ mutations at other loci during this time, the largest of which will typically persist for $T_2 \sim (m-1)N\mu T_1$ while the original allele is still drifting. During T_2 , the longest-drifting mutation at a third locus will typically persist for $T_3 \sim (m-2)N\mu T_2 = (m-1)(m-2)(N\mu)^2 T_1$, and so on, with the the m^{th} mutation persisting for $T_m \sim (K-1)!(N\mu)^{m-1} T_1$. The frequency x_m of the m -mutant genotype will peak at:

$$\begin{aligned} x_m &\sim \prod_{k=1}^m (T_k/N) \\ &\sim \left((N\mu)^{\frac{m-1}{2}} \frac{T_1}{N} \right)^m \frac{(m-1)!^{m-1}}{G(m+1)}, \end{aligned}$$

where G is the double gamma function. For the mutations to establish, this peak frequency must exceed $\sim 1/Ns$. Solving this condition for T_1 gives the time scale over which the first mutation must drift to be likely to be successful:

$$T_1 \sim N(Ns)^{-\frac{1}{m}} (N\mu)^{-\frac{m-1}{2}} e^{-\frac{m}{4}} (m-1)^{\frac{m-4}{2}}, \quad (3.14)$$

where the final combinatorial factors are approximations valid for large m , and negligible for small m . For the initial mutation to be more likely to tunnel than to fix, T_1 must be small compared to N . Solving $T_1 \sim N$ for m therefore gives the maximum tunneling range:

$$m \approx \left\lfloor \frac{1}{2} \left(1 + \sqrt{1 + \frac{8 \ln(Ns)}{|\ln(N\mu)|}} \right) \right\rfloor, \quad (3.15)$$

where $\lfloor \cdot \rfloor$ is the floor function.

If $m \geq K$, then a wild-type population can tunnel directly to the beneficial genotype. The probability that a mutation successfully tunnels is $\sim 1/T_1$, so the expected waiting time is:

$$\begin{aligned} \mathcal{T} &\approx \frac{T_1}{KN\mu} \\ &\approx N(Ns)^{-\frac{1}{K}}(N\mu)^{-\frac{K+1}{2}}e^{-\frac{K}{4}}(K-1)^{\frac{K-6}{2}}, \end{aligned} \quad (3.16)$$

where we have substituted Equation 3.14 with $m = K$ for T_1 . If $1 < m < K$, the total plateau-crossing time is dominated by the time it takes for the population to fix $K - m$ mutations via drift so that it can get close enough to the adaptive genotype to tunnel the rest of the way. (If $m = 1$, then the population cannot tunnel and must fix $K - 1$ mutations by drift, at which point the K^{th} mutation becomes beneficial.) The k^{th} mutation fixes after an expected waiting time of $1/(K - k + 1)\mu$, so the total expected waiting time for $K - m$ mutations to fix is

$$\begin{aligned} \mathcal{T} &\approx \sum_{k=1}^{K-m} \frac{1}{(K - k + 1)\mu} \\ &\approx \frac{\ln(K/m)}{\mu} \text{ for } K - m \gg 1. \end{aligned} \quad (3.17)$$

3.4.2.3 Three-mutation plateaus

Plugging in $K = 3$ to the above analysis, the expected time to cross the plateau is (with $\mathcal{O}(1)$ combinatorial factors included for clarity):

$$\mathcal{T} \approx \begin{cases} \frac{1}{\mu}(\mu/s)^{1/3} & \text{for } N\mu \gg 1 \text{ (deterministic)} \\ (3N\mu^2)^{-1} (2Ns)^{-1/3} & \text{for } (\mu^3s)^{-1/4} \ll N \ll 1/\mu \text{ (pure tunneling)} \\ 1/(3\mu) + \frac{1}{N}(\mu^3s)^{-1/4} & \text{for } (\mu s)^{-1/2} \ll N \ll (\mu^3s)^{-1/4} \text{ (tun. after 1 mut. fixes)} \\ 5/(6\mu) & \text{for } N \ll (\mu s)^{-1/2} \text{ (sequential fixation)}. \end{cases} \quad (3.18)$$

The second term in the third line is the $K = 2$ tunneling time [Weissman et al., 2010].

3.5 Discussion

We have shown that even moderately large populations can acquire complex adaptations requiring three individually-useless mutations substantially faster than would be expected if mutations had to fix sequentially by drift. In other words, natural selection can at least somewhat effectively promote mutations that not only provide no direct selective benefit, but also do not directly increase evolvability, i.e., do not change the distribution of mutational effects. Recombination helps most at intermediate population sizes, where there can be simultaneous polymorphisms at multiple loci but triple mutants are rare. In this range, the rate of plateau-crossing is maximized when recombination is just somewhat rarer than selection.

Across regimes, the rate of crossing the three-mutant fitness plateau scales sub-cubically in the mutation rate, i.e., complexity is not strongly suppressing the rate of adaptation, suggesting that even more complex adaptations could also potentially be acquired. However, analyzing even the three-mutation case for $r \lesssim s$ involved a proliferation of different dynamical regimes, so simply extending our analysis to wider plateaus is likely to be impractical. The asexual case [Weissman et al., 2009] and the case $r \gg s$ analyzed above are simpler but plateau-crossing is fastest for $r \lesssim s$, meaning that these easier limiting cases may be missing essential dynamics.

How practically important could adaptive paths across three-mutation plateaus be? Could we hope to observe experimental populations following them? Viruses often have large populations and high mutation rates; if we consider an RNA virus with a mutation rate of $\sim 10^{-4}$ per base per replication and a potential adaptation providing a $\sim 10\%$ fitness advantage, a population size of $N > 10^9$ – fewer than might be present in a single infected host – would be enough for the population to deterministically acquire the triple-mutant genotype. On the other hand, if we consider a yeast population in which the relevant mutations have target sizes of ~ 300 base pairs, for a mutation rate of $\sim 10^{-7}$ [Lynch et al., 2008], it would be difficult to maintain a large enough experimental population for long enough to reliably acquire the adaptation via any of the paths we have described here.

The major limitation in seriously applying any of our analysis to real populations is that we have considered the necessary loci in isolation. As mentioned in the Introduction, a major part of our motivation in considering the possibility of complex adaptation is that a combinatorial argument suggests that there are potentially very

many of them available. But if there are in fact very many possible complex adaptations, then the first one that actually fixes in the population is likely to be one that happened unusually quickly, potentially by different dynamics than those considered here [Weissman et al., 2010]. Thus at a minimum, we would need to consider the entire distribution of plateau-crossing times rather than just its mean. More precisely, we would need to describe the left tail of the distribution. This may in fact simplify the analysis – there may be only a few ways for a lineage to get a lot of mutations quickly, regardless of the population parameters [Weissman et al., 2010] – and thus provide a way forward to analyzing wider plateaus.

The fact that the population is likely to be adapting at more than just K loci does not only mean that we need to think about the left tail of crossing-time distribution; it also means that we need to think about how adaptation elsewhere in the genome may affect evolution at the focal loci. If there is substantial fitness variance due to the rest of the genome and limited recombination, the dynamics of the mutant lineages will be completely different due to hitchhiking [Neher and Shraiman, 2011]. In addition, the complex adaptation may be lost due to clonal interference once it is produced, reducing its fixation probability. The fixation probability in this case is likely to require a careful calculation in its own right, as the background fitness is likely to systematically differ from the mean because of the required conditioning on long-lasting lineages carrying the intermediate mutations.

Perhaps even more importantly than clonal interference is the potential epistatic interference. When we consider just the K focal loci, substitutions at other loci should turn the fitness landscape into a constantly shifting metaphoric “seascape”

[Mustonen and Lässig, 2009]. In the most extreme case, other mutations may fix that permanently disrupt the potential complex adaptation, forcing the population onto another path. We have no understanding of how this should affect the probability of complex adaptation in anything beyond the simplest possible case of a single beneficial mutation blocking a two-mutation complex adaptation in an asexual population [Ochs and Desai, 2015]. We can already see that is likely to substantially change the interpretation of our results by looking at Table 2 and our results for generic K with $r \gg s$ and $N\mu \ll 1$. The regimes in which the population only tunnels through the last mutations while initially fixing the others via drift appear to have roughly the same rate of plateau-crossing as the sequential fixation regime in which all mutations but the very last must drift to fixation. But this is because our model assumes that all populations reach the adaptive genotype eventually. In a more realistic model in which populations can get diverted and miss potential adaptations entirely, being able to tunnel through m mutations greatly increases the zones of attraction of adaptive genotypes in the fitness landscape, and could make a large difference in the probability of finding them.

In addition to epistatic interactions with other loci, the plateau could shift because of environmental changes [Masel, 2006, Kim, 2007]. It is difficult to say which process is likely to be more important. We currently do not even know whether changes in the selective coefficient of single mutations are driven more often by environmental changes or changes in the rest of the genome, let alone what drives changes in selection on the rest of the genome. More generally, the basic difficulty in analyzing more complex, realistic fitness landscapes is that we have no idea what they should look like. Even mapping out the local fitness landscape of a single gene

requires a heroic experimental effort (e.g., [Bank et al., 2016]) – and then we only know it in a limited number of artificial environments. Our best hope may be to try to develop a theory that can reduce the full complexity of landscapes to a reasonable number of parameters describing their features that are most relevant for adaptation, but it is an open question whether such a theory exists.

3.A. Appendix

3.A.1 Small populations with rare recombination

Here we focus on populations with low mutation supply, $N\mu \ll 1$, and rare recombination, $r \ll s$. In particular, we focus on those that fall in between the asexual and semi-linkage equilibrium cases discussed above, for which recombination is frequent enough to speed plateau-crossing but too rare to bring even the largest mutant lineages into linkage equilibrium with each other. As above, the expected plateau-crossing time is dominated by the waiting time for the production of the first successful single-mutant lineage A which drifts for time T_A , with the other possible mutations labeled B and C . All genotypes that drift for a time T_X reach a typical size $n_X \sim T_X$, so we will not need to distinguish between drift times and lineage sizes in the following. We will exploit our freedom in labeling the B and C mutations to always label the double mutant AB if it has the A allele, so the AC genotype will not appear in our analysis. Throughout, we will ignore $\mathcal{O}(1)$ numerical factors arising from combinatorics and integration, none of which change the results significantly. We can identify eight possible asymptotic scenarios, depending on the relative sizes

of the different relevant lineages (Figure A.1):

(i) $T_A \gg T_B \gg T_C \gg T_{AB}$: While three independent mutant lineages are drifting, the larger two recombine, and then that recombinant recombines with the third lineage to produce the successful triple-mutant. Typical sizes are $T_B \sim N\mu T_A$, $T_C \sim N\mu T_B$, and $T_{AB} \sim \frac{r}{N} T_A T_B T_C$ (because we only consider the largest AB lineage that arises while C is drifting). The number of ABC individuals produced by recombination between C and AB during the $\sim T_{AB}$ generations that they coexist is $\approx \frac{r}{N} T_C T_{AB}^2$; we need this quantity to be $\sim 1/s$ for success to be likely:

$$\begin{aligned} 1 &\sim \frac{rs}{N} T_C T_{AB}^2 \\ &\sim \frac{rs}{N} ((N\mu)^2 T_A) \left(\frac{r}{N} (N\mu T_A)^3 \right)^2 \\ &\approx N^5 \mu^8 r^3 s T_A^7. \end{aligned}$$

Solving for T_A and the other drift times gives:

$$\begin{cases} T_A \sim (N^5 \mu^8 r^3 s)^{-1/7} \\ T_B \sim (\mu r^3 s / N^2)^{-1/7} \\ T_C \sim (r^3 s / (N^9 \mu^6))^{-1/7} \\ T_{AB} \sim (N \mu^3 r^2 s^3)^{-1/7}. \end{cases} \quad (\text{A.1})$$

(ii) $T_A \gg T_C \gg T_B \gg T_{AB}$: While three independent mutant lineages are drifting, the largest recombines with the smallest, and then that recombinant recombines with the middle lineage to produce the successful triple-mutant. Typical sizes are

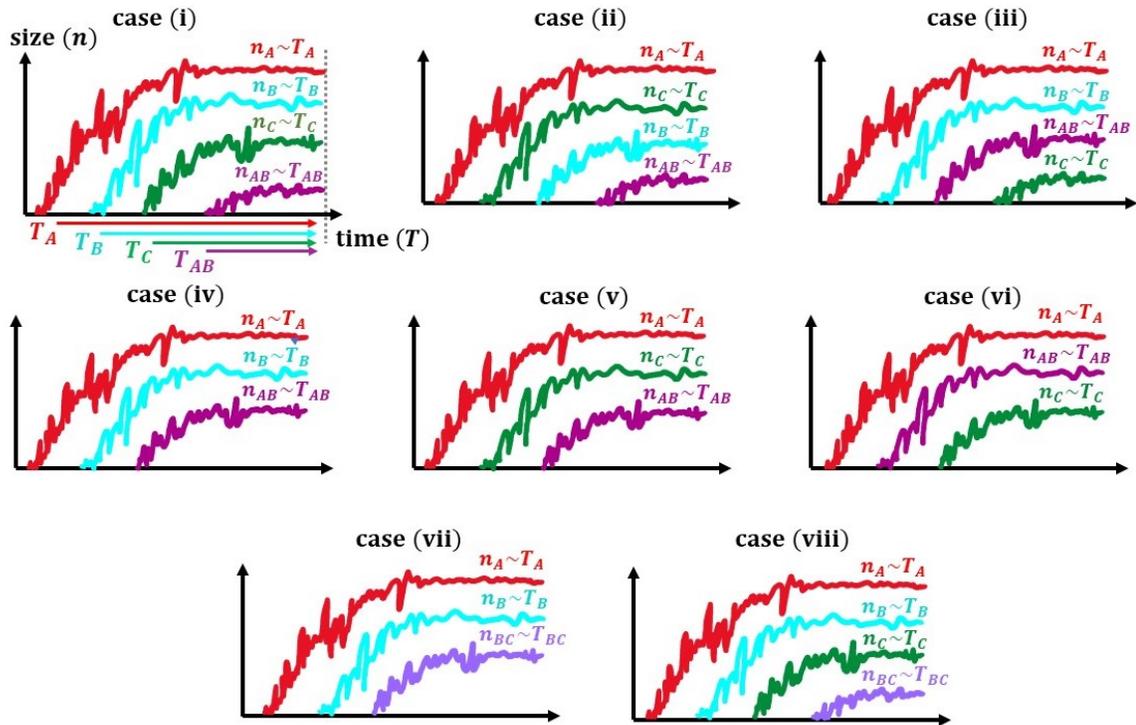


Figure A.1: A schematic plot of 8 different ways in which the plateau-crossing can occur given that all the mutant alleles are not frequently produced in the population. In case (i), the two bigger lineages, A and B , recombine with each other to produce the AB lineage which in turns recombines with the C lineage to produce the first successful ABC lineage after T_A generations. Cases (ii) and (iii) also correspond to the same dynamics as case (i) except for that the second and third biggest mutant lineages could be different. Case (iv) occurs when lineages A and B recombine to produce an AB lineage which later mutates to produce the beneficial mutant. In both cases (v) and (vi), the A lineage mutates to produce the AB lineage which then recombines with the C lineage to produce the successful ABC lineage. In case (vii), the B lineage mutates to produce a BC lineage which later recombines with A . Case (viii) occurs when the B and C lineages recombine to produce a BC lineage which then recombines with the A lineage.

$T_C \sim N\mu T_A$, $T_B \sim N\mu T_C$, and $T_{AB} \sim \frac{r}{N} T_A T_B^2$. To get $\sim 1/s$ triple-mutants, we need:

$$\begin{aligned} 1 &\sim \frac{rs}{N} T_C T_{AB}^2 \\ &\sim \frac{rs}{N} (N\mu T_A) \left(\frac{r}{N} (N\mu)^4 T_A^3 \right)^2 \\ &\approx N^6 \mu^9 r^3 s T_A^7. \end{aligned}$$

Solving for T_A and the other drift times gives:

$$\begin{cases} T_A \sim (N^6 \mu^9 r^3 s)^{-1/7} \\ T_C \sim (\mu^2 r^3 s / N)^{-1/7} \\ T_B \sim (r^3 s / (N^8 \mu^5))^{-1/7} \\ T_{AB} \sim (N^2 \mu^4 r^2 s^3)^{-1/7}. \end{cases} \quad (\text{A.2})$$

(iii) $T_A \gg T_B \gg T_{AB} \gg T_C$: Two single-mutant lineages recombine. While that recombinant double-mutant is drifting, a third single-mutant lineage arises and recombines with it to produce a successful triple-mutant. Typical sizes are $T_B \sim N\mu T_A$, $T_{AB} \sim \frac{r}{N} T_A T_B^2$, and $T_C \sim N\mu T_{AB}$. To get $\sim 1/s$ triple-mutants, we need:

$$\begin{aligned} 1 &\sim \frac{rs}{N} T_{AB} T_C^2 \\ &\sim \frac{rs}{N} (N\mu)^2 \left(\frac{r}{N} (N\mu)^2 T_A^3 \right)^3 \\ &\sim N^4 \mu^8 r^4 s T_A^9. \end{aligned}$$

Solving for T_A and the other drift times gives:

$$\begin{cases} T_A \sim (N^4 \mu^8 r^4 s)^{-1/9} \\ T_B \sim (r^4 s / (N^5 \mu))^{-1/9} \\ T_{AB} \sim (N \mu^2 r s)^{-1/3} \\ T_C \sim (r s / (N^2 \mu))^{-1/3}. \end{cases} \quad (\text{A.3})$$

(iv) $T_A \gg T_B \gg T_{AB}$: Two single-mutant lineages recombine, and that recombinant lineage then mutates and succeeds. Typical sizes are $T_B \sim N \mu T_A$ and $T_{AB} \sim \frac{r}{N} T_A T_B^2$. The AB lineage will produce $\sim \mu T_{AB}^2$ mutants while it is drifting; setting this equal to $\sim 1/s$ gives:

$$\begin{aligned} 1 &\sim \mu s T_{AB}^2 \\ &\sim \mu s \left(\frac{r}{N} (N \mu)^2 T_A^3 \right)^2 \\ &\sim N^2 \mu^5 r^2 s T_A^6. \end{aligned}$$

Solving for T_A and the other drift times gives:

$$\begin{cases} T_A \sim (N^2 \mu^5 r^2 s)^{-1/6} \\ T_B \sim (r^2 s / (N^4 \mu))^{-1/6} \\ T_{AB} \sim (\mu s)^{-1/2}. \end{cases} \quad (\text{A.4})$$

(v) $T_A \gg T_C \gg T_{AB}$: While two single-mutant lineages are drifting, the mutates

at the third locus. This double-mutant then recombines with the other single-mutant lineage. Typical sizes are $T_C \sim N\mu T_A$ and $T_{AB} \sim \mu T_A T_C$ (because we only consider the largest-double mutant lineage that arises while C is drifting). To get $\sim 1/s$ triple-mutants, we need:

$$\begin{aligned} 1 &\sim \frac{rs}{N} T_C T_{AB}^2 \\ &\sim \frac{rs}{N} N\mu T_A (N\mu^2 T_A^2)^2 \\ &\sim N^2 \mu^5 rs T_A^5. \end{aligned}$$

Solving for T_A and the other drift times gives:

$$\begin{cases} T_A \sim \frac{1}{\mu} (N^2 rs)^{-1/5} \\ T_C \sim (N^3 / (rs))^{1/5} \\ T_{AB} \sim (N / (rs)^2)^{1/5}. \end{cases} \quad (\text{A.5})$$

(vi) $T_A \gg T_{AB} \gg T_C$: A single-mutant lineage mutates. While the resulting double-mutant lineage drifts, a new lineage with a mutation at the third locus arises and successfully recombines with it. Typical sizes are $T_{AB} \sim \mu T_A^2$ and $T_C \sim N\mu T_{AB}$. To get $\sim 1/s$ triple-mutants, we need:

$$\begin{aligned} 1 &\sim \frac{rs}{N} T_{AB} T_C^2 \\ &\sim \frac{rs}{N} \mu T_A^2 (N\mu^2 T_A^2)^2 \\ &\sim N\mu^5 rs T_A^6. \end{aligned}$$

Solving for T_A and the other drift times gives:

$$\begin{cases} T_A \sim (N\mu^5rs)^{-1/6} \\ T_{AB} \sim (N\mu^2rs)^{-1/3} \\ T_C \sim (N^2\mu/(rs))^{1/3}. \end{cases} \quad (\text{A.6})$$

(vii) $T_A \gg T_B \gg T_{BC}$: While two single-mutant lineages are drifting, the smaller one acquires an additional mutation at the third locus. This double-mutant lineage then successfully recombines with the larger single-mutant lineage. Typical sizes are $T_B \sim N\mu T_A$ and $T_{BC} \sim \mu T_B^2$. To get $\sim 1/s$ triple-mutants, we need:

$$\begin{aligned} 1 &\sim \frac{rs}{N} T_A T_{BC}^2 \\ &\sim \frac{rs}{N} N^4 \mu^6 T_A^5 \end{aligned}$$

Solving for T_A and the other drift times gives:

$$\begin{cases} T_A \sim (N^3\mu^6rs)^{-1/5} \\ T_B \sim (N^2/(\mu rs))^{1/5} \\ T_{BC} \sim (N^4\mu^3/(rs)^2)^{1/5}. \end{cases} \quad (\text{A.7})$$

(viii) $T_A \gg T_B \gg T_C \gg T_{BC}$: While three single-mutant lineages are drifting, the smaller two recombine. The recombinant then successfully recombines with the

largest single-mutant lineage. Typical sizes are $T_B \sim N\mu T_A$, $T_C \sim N\mu T_B$, and $T_{BC} \sim \frac{r}{N} T_B T_C^2$. To get $\sim 1/s$ triple-mutants, we need:

$$\begin{aligned} 1 &\sim \frac{rs}{N} T_A T_{BC}^2 \\ &\sim \frac{rs}{N} T_A (N^4 \mu^5 r T_A^3)^2 \\ &\sim N^7 \mu^{10} r^3 s T_A^7. \end{aligned}$$

Solving for T_A and the other drift times gives:

$$\left\{ \begin{array}{l} T_A \sim \frac{1}{N\mu} (\mu^3 r^3 s)^{-1/7} \\ T_B \sim (\mu^3 r^3 s)^{-1/7} \\ T_C \sim N\mu (\mu^3 r^3 s)^{-1/7} \\ T_{BC} \sim N (\mu^5 / (r^2 s^3))^{1/7}. \end{array} \right. \quad (\text{A.8})$$

For all of these cases, the expected plateau-crossing time is $\mathcal{T} \sim T_A / (N\mu)$. All require that the double-mutant drift times T_{AB} or T_{BC} be small compared to $1/r$, so that the lineage is not broken up by recombination. We collect the predicted rates

and conditions here:

$$\mathcal{T}^{-1} \sim \begin{cases} N\mu (N^5\mu^8r^3s)^{1/7}, & N \gg r^5/(\mu s)^3 \text{ (i)} \\ N\mu (N^6\mu^9r^3s)^{1/7}, & N \gg \sqrt{r^5/(\mu^4s^3)} \text{ (ii)} \\ N\mu (N^4\mu^8r^4s)^{1/9}, & N \gg r^2/(\mu^2s) \text{ (iii)} \\ N\mu (N^2\mu^5r^2s)^{1/6}, & r \ll \sqrt{\mu s} \text{ (iv)} \\ N\mu^2 (N^2rs)^{1/5}, & N \ll s^2/r^3 \text{ (v)} \\ N\mu (N\mu^5rs)^{1/6}, & N \gg r^2/(\mu^2s) \text{ (vi)} \\ N\mu (N^3\mu^6rs)^{1/5}, & N \ll (s^2/(\mu r)^3)^{1/4} \text{ (vii)} \\ (N\mu)^2 (\mu^3r^3s)^{1/7}, & N \ll (s^3/(\mu r)^5)^{1/7} \text{ (viii)}. \end{cases} \quad (\text{A.9})$$

For parameter values where multiple cases apply, the predicted \mathcal{T} value is the one corresponding to the case with the smallest T_A – the rates for the different cases do not add, since all are dependent on the same initial dynamic of an unusually long-lived single-mutant. If even the smallest T_A is greater than N , single-mutants are more likely to fix than tunnel. For most reasonable parameter values, multiple different cases give similar values in Equation A.9, i.e., populations are not in the true asymptotic regimes corresponding to one case or another. However, since they all roughly agree, the predicted value for \mathcal{T} is still accurate.

References

- W. P. Angerer. An explicit representation of the Luria–Delbrück distribution. *Journal of Mathematical Biology*, 42(2):145–174, 2001.
- P. Armitage. The statistical theory of bacterial populations subject to mutation. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–40, 1952.
- C. Bank, S. Matuszewski, R. T. Hietpas, and J. D. Jensen. On the (un)predictability of a large intragenic fitness landscape. *Proceedings of the National Academy of Sciences*, 113(49):14085–14090, 2016.
- R. Barrangou, C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero, and P. Horvath. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 315(5819):1709–1712, 2007.
- M. S. Bartlett. *An introduction to stochastic processes: with special reference to methods and applications*. Cambridge University Press, Cambridge, UK, 1978.
- F. B. Christiansen, S. P. Otto, A. Bergman, and M. W. Feldman. Waiting with and without recombination: the time to production of a double mutant. *Theoretical Population Biology*, 53(3):199–215, 1998.
- R. Durrett and D. Schmidt. Waiting for two mutations: With applications to regulatory sequence evolution and the limits of Darwinian evolution. *Genetics*, 180(3):1501–1509, 2008.
- I. Eshel and M. W. Feldman. On the evolutionary effect of recombination. *Theoretical Population Biology*, 1(1):88–100, 1970.
- W. J. Ewens. *Mathematical Population Genetics 1: Theoretical Introduction*. Springer-Verlag, New York, 2nd edition, 2004.
- M. W. Feldman. Equilibrium studies of two locus haploid populations with recombination. *Theoretical Population Biology*, 2(3):299–318, 1971.
- D. S. Fisher. Course 11 evolutionary dynamics. *Les Houches*, 85:395–446, 2007.
- D. S. Fisher, M. Lässig, and B. I. Shraiman. Evolutionary dynamics and statistical physics. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(01):N01001, 2013.

- M. Ghafari and D. B. Weissman. The expected time to cross extended fitness plateaus. *bioRxiv*, (343053), 2018.
- K. Hantke and V. Braun. Functional interaction of the *tonA/tonB* receptor system in *Escherichia coli*. *Journal of Bacteriology*, 135(1):190–197, 1978.
- W. Hayes. *The genetics of bacteria and their viruses. Studies in basic genetics and molecular biology*. Blackwell Scientific Publications, Oxford, UK, 1964.
- C. M. Holmes, M. Ghafari, A. Abbas, V. Saravanan, and I. Nemenman. Luria–Delbrück, revisited: the classic experiment does not rule out Lamarckian evolution. *Physical Biology*, 14(5):055004, 2017.
- Y. Iwasa, F. Michor, and M. A. Nowak. Stochastic tunnels in evolutionary dynamics. *Genetics*, 166(3):1571–1579, 2004.
- E. Jablonka and M. J. Lamb. The changing concept of epigenetics. *Annals of the New York Academy of Sciences*, 981(1):82–96, 2002.
- E. Jablonka and G. Raz. Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *The Quarterly Review of Biology*, 84(2):131–176, 2009.
- G. Jaeger and S. Sarkar. On the distribution of bacterial mutants: the effects of differential fitness of mutants and non-mutants. *Genetica*, 96(3):217–223, 1995.
- K. Jain. Time to fixation in the presence of recombination. *Theoretical Population Biology*, 77(1):23–31, 2009.
- K. Jain and J. Krug. Deterministic and stochastic regimes of asexual evolution on rugged fitness landscapes. *Genetics*, 175(3):1275–1288, 2006.
- M. Jones, S. Thomas, and A. Rogers. Luria–Delbrück fluctuation experiments: design and analysis. *Genetics*, 136(3):1209–1216, 1994.
- S. Karlin and J. McGregor. On mutation selection balance for two-locus haploid and diploid populations. *Theoretical Population Biology*, 2(1):60–70, 1971.
- Y. Kim. Rate of adaptive peak shifts with partial genetic robustness. *Evolution*, 61(8):1847–1856, 2007.
- A. L. Koch. Mutation and growth rates from Luria–Delbrück fluctuation tests. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 95(2-3):129–143, 1982.

- N. L. Komarova, A. Sengupta, and M. A. Nowak. Mutation-selection networks of cancer initiation: tumor suppressor genes and chromosomal instability. *Journal of Theoretical Biology*, 223(4):433–450, 2003.
- E. V. Koonin and Y. I. Wolf. Is evolution Darwinian or/and Lamarckian? *Biology Direct*, 4(1):42, 2009.
- E. V. Koonin and Y. I. Wolf. Just how Lamarckian is CRISPR-Cas immunity: the continuum of evolvability mechanisms. *Biology Direct*, 11(1):9, 2016.
- D. Lea and C. A. Coulson. The distribution of the numbers of mutants in bacterial populations. *Journal of Genetics*, 49(3):264–285, 1949.
- J. Lederberg and E. M. Lederberg. Replica plating and indirect selection of bacterial mutants. *Journal of Bacteriology*, 63(3):399, 1952.
- S. E. Luria and M. Delbrück. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics*, 28(6):491, 1943.
- M. Lynch, W. Sung, K. Morris, N. Coffey, C. R. Landry, E. B. Dopman, W. J. Dickinson, K. Okamoto, S. Kulkarni, D. L. Hartl, and W. K. Thomas. A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proceedings of the National Academy of Sciences*, 105(27):9272–9277, 2008.
- W. Ma, G. v. H. Sandri, and S. Sarkar. Analysis of the Luria–Delbrück distribution using discrete convolution powers. *Journal of Applied Probability*, 29(02):255–267, 1992.
- D. J. MacKay. Bayesian interpolation. *Neural computation*, 4(3):415–447, 1992.
- D. J. Mackay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, Cambridge, UK, 2003.
- B. Mandelbrot. A population birth-and-mutation process, I: explicit distributions for the number of mutants in an old culture of bacteria. *Journal of Applied Probability*, 11(03):437–444, 1974.
- J. Masel. Cryptic genetic variation is enriched for potential adaptations. *Genetics*, 172(3):1985–1991, 2006.
- Y. Michalakis and M. Slatkin. Interaction of selection and recombination in the fixation of negative-epistatic genes. *Genetical Research*, 67:257–269, 1996.

- S. Montgomery-Smith, A. Le, G. Smith, S. Billstein, H. Oveys, D. Pisechko, and A. Yates. Estimation of mutation rates from fluctuation experiments via probability generating functions. *arXiv*, (1608.04175), 2016.
- V. Mustonen and M. Lässig. From fitness landscapes to seascapes: non-equilibrium dynamics of selection and adaptation. *Trends in Genetics*, 25(3):111–119, 2009.
- R. A. Neher and B. I. Shraiman. Genetic draft and quasi-neutrality in large facultatively sexual populations. *Genetics*, 188(4):975–996, 2011.
- P. Nelson. *Physical Models of Living Systems*. Freeman, San Francisco, CA, 2015.
- H. B. Newcombe. Origin of bacterial variants. *Nature*, 164(4160):150–150, 1949.
- I. E. Ochs and M. M. Desai. The competition between simple and complex evolutionary trajectories in asexual populations. *BMC Evolutionary Biology*, 15(1):813, 2015.
- S. Sarkar. Haldane’s solution of the Luria–Delbrück distribution. *Genetics*, 127(2):257, 1991.
- S. Sarkar, W. Ma, and G. v. H. Sandri. On fluctuation analysis: a new, simple and efficient method for computing the expected number of mutants. *Genetica*, 85(2):173–179, 1992.
- F. Stewart. Fluctuation analysis: the effect of plating efficiency. *Genetica*, 84(1):51–55, 1991.
- F. M. Stewart, D. M. Gordon, and B. R. Levin. Fluctuation analysis: the probability distribution of the number of mutants under different conditions. *Genetics*, 124(1):175–185, 1990.
- H. Suzuki. The optimum recombination rate that realizes the fastest evolution of a novel functional combination of many genes. *Theoretical Population Biology*, 51(3):185–200, 1997.
- N. Takahata. Sexual recombination under the joint effects of mutation, selection, and random sampling drift. *Theoretical Population Biology*, 22(2):258–277, 1982.
- A. L. Taylor. Bacteriophage-induced mutation in *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 50(6):1043–1051, 1963.
- M. V. Trotter, D. B. Weissman, G. I. Peterson, K. M. Peck, and J. Masel. Cryptic genetic variation can make “irreducible complexity” a common mode of adaptation in sexual populations. *Evolution*, 68(12):3357–3367, 2014.

- E. van Nimwegen and J. Crutchfield. Metastable evolutionary dynamics: Crossing fitness barriers or escaping via neutral paths? *Bulletin of Mathematical Biology*, 62(5):799–848, 2000.
- D. M. Weinreich and L. Chao. Rapid evolutionary escape by large populations from local fitness peaks is likely in nature. *Evolution*, 59(6):1175–1182, 2005.
- D. B. Weissman, M. M. Desai, D. S. Fisher, and M. W. Feldman. The rate at which asexual populations cross fitness valleys. *Theoretical Population Biology*, 75:286–300, 2009.
- D. B. Weissman, M. W. Feldman, and D. S. Fisher. The rate of fitness-valley crossing in sexual populations. *Genetics*, 186:1389–1410, 2010.
- B. Ycart. Fluctuation analysis: can estimates be trusted? *PloS one*, 8(12):e80958, 2013.
- Q. Zheng. Progress of a half century in the study of the Luria–Delbrück distribution. *Mathematical Biosciences*, 162(1):1–32, 1999.
- Q. Zheng. On Haldane’s formulation of Luria and Delbrück’s mutation model. *Mathematical Biosciences*, 209(2):500–513, 2007.