bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fitness-valley crossing in subdivided asexual populations

Michael R. McLaren

Department of Biology, Stanford University

mmclaren@alumni.stanford.edu

Abstract

Adaptations may require multiple mutations that are beneficial only in combination. To adapt, a lineage must acquire mutations that are individually neutral or deleterious before gaining the beneficial combination, thereby crossing a plateau or valley, respectively, in the mapping from genotype to fitness. Spatial population structure can facilitate plateau and valley crossing by allowing neutral and deleterious lineages to survive longer and produce more beneficial mutants. Here, we analyze adaptation across a two-mutation plateau or valley in an asexual population that is subdivided into discrete subpopulations, or demes, connected by migration. We describe how subdivision alters the dynamics of adaptation from those in an equally sized unstructured population and give a complete quantitative description of these dynamics for the island migration model. Subdivision can significantly decrease the waiting time for the adaptation if demes and migration rates are small enough that single-mutant lineages fix in one or more demes before producing the beneficial double mutant. But, the potential decrease is small in very large populations and may also be limited by the slow spread of the beneficial mutant in extremely subdivided populations. Subdivision has a smaller effect on the probability that the population adapts very quickly than on the mean time to adapt, which has important consequences in some applications, such as the development of cancer. Our results provide a general and intuitive framework for predicting the effects of spatial structure in other models and in natural populations.

Keywords: complex adaptation, fitness valleys, sign epistasis, population subdivision, spatial structure

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Contents

1 Introduction **3**

2 Model **7**

3 Summary of results **9**

4 Fate of a single-mutant lineage **13**

4.1 Qualitative effects of subdivision 13

4.2 Unstructured populations 18

4.3 Limit as $m \to 0$ and the isolated-demes regime 23

4.4 Locally neutral single mutants ($\delta \ll \alpha/N$) 26

4.5 Locally deleterious single mutants ($\delta \gg \alpha/N$) 33

4.6 Sweep time 38

5 Average waiting time for the population to adapt **39**

6 Distribution of the waiting time **44**

7 Discussion **47**

References **56**

Tables **60**

Figures **61**

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 Introduction

Adaptations that give rise to new biological functions are often *complex*, emerging from the interactions of multiple mutations. Examples include the evolution of aerobic citrate utilization in experimentally evolving populations of *Escherichia coli* (Blount *et al.* 2012) and the evolution in vertebrates of a ligand-activated transcription factor to bind a new target ligand (Ortlund *et al.* 2007) or new target DNA sequence (McKeown *et al.* 2014). The acquisition of drug resistance in pathogenic bacteria (Wang *et al.* 2002; Weinreich *et al.* 2006) and immune system escape in influenza (Gong *et al.* 2013) and HIV (Silva *et al.* 2010) can also be complex adaptations.

The mutations involved in complex adaptations have effects on function and fitness that depend on which other mutations are present. As a result, some or all of the sequences of mutations that lead from the ancestral to the adapted genotype may have steps that are neutral or deleterious. Such sequences are said to cross a plateau or valley, respectively, in the map from genotype to fitness. Determining how a species can adapt via such a sequence of mutations (in both cases generically referred to as *valley crossing*) is crucial to understanding how and when complex adaptations evolve.

Valley crossing also has implications for the evolution of human pathogens and in the development of cancer. Persistence of drug resistance in bacteria and viruses has been explained by the observation that, in the absence of drugs, drug-resistant and drug-sensitive strains are separated by a fitness valley, making reversion to drug sensitivity difficult (Levin *et al.* 2000; Trindade *et al.* 2009). The evolution of Human Immunodeficiency Virus-1 (HIV-1) to use an alternate host co-receptor, seen in roughly one-half of HIV-1 patients, may also require crossing a valley (Regoes and Bonhoeffer 2005; Silva *et al.* 2010; Silva and Wyatt 2014). Finally, cancer usually results from a sequence of mutations (often called *hits*) accumulating within an individual cell that give the cell's lineage an increased growth rate. For example, loss of function (LOF) mutations may be needed in both copies of a particular tumor-suppressor gene (TSG) before a lineage begins growing into a tumor (Knudson 2001;

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Frank 2007). If both LOF mutations are needed for cells to have a competitive advantage over cells with no mutations, then inactivating the TSG requires crossing a valley or plateau.

In asexual populations, valley crossing requires a clonal lineage to sequentially acquire all of the necessary mutations. Large asexual populations typically cross plateaus and valleys by a process in which a lineage acquires the mutations while still rare in the population. This process also occurs in sexual populations if the mutations are closely linked (Weissman *et al.* 2010), as is likely if the mutations occur in the same gene. Komarova *et al.* (2003) used the term *tunneling* to distinguish this process from the typical mode of adaptation in small populations by which the mutations fix sequentially. They were also the first to analyze tunneling where two mutations are required and the single-mutant genotype is either neutral or deleterious. Gillespie (1984) first analyzed tunneling for strongly deleterious intermediate mutants and Kimura (1985) analyzed tunneling as a form of neutral evolution, where the single mutant is deleterious and the double mutant is neutral. Later studies analyzed asexual tunneling under new regimes or provided additional details (e.g., Iwasa *et al.* 2004,Weinreich and Chao 2005, Durrett and Schmidt 2008, and Proulx 2011; see Weissman *et al.* (2009) for the most complete account of valley crossing and tunneling in asexual populations). Genomic evidence suggests that tunneling occurred repeatedly in the evolution of mitochondrial tRNA genes (Meer *et al.* 2010). Theory, observations, and experiments suggest that tunneling occurs during co-receptor switch in HIV (Regoes and Bonhoeffer 2005; Silva and Wyatt 2014). The large size of stem cell populations also suggests that tunneling will often be involved in the acquisition of multiple hits leading to some cancers. Even in cases where intermediate genotypes offer a small fitness advantage, tunneling may be a more likely mode of adaptation that sequential fixation of individual mutations (Weissman *et al.* 2009).

Tunneling studies typically assume that the population is completely unstructured, with all individuals competing equally with one another; however, natural populations are spatially structured to various degrees, such that individuals disperse and compete within a region smaller than the population's total range. Spatially structured populations can be

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

subdivided into distinct subpopulations, or *demes*; distributed continuously in space; or fall somewhere in between these extremes. Early on, Sewall Wright suggested that spatial structure should facilitate valley crossing (Wright 1932). Wright was mainly concerned with sexual species having high rates of recombination between loci and, in these species, spatial structure reduces the rate at which sex breaks up beneficial combinations of mutations. This idea forms the backbone of Wright's famous shifting balance theory of evolution (Wright 1970; Charlesworth and Charlesworth 2010). However, the extensive literature on the shifting balance theory says little about how spatial structure might affect valley crossing in asexual species.

We aim to understand how spatial structure affects valley crossing in asexual populations for the simplest scenario, where the adaptation requires two mutations. Spatial structure can have qualitatively different effects on evolutionary dynamics depending on assumptions about how competition and demography interact with space. We set out to describe the consequences of perhaps the most basic aspect of spatial structure, that individuals tend to compete and reproduce within a local region that is small compared to the population's total geographic range. This basic feature of space can affect asexual tunneling in the following way. If migration is sufficiently limited, mutant lineages that segregate in the population will tend to be unevenly spatially distributed, and mutant individuals will tend to have a greater than average proportion of mutants nearby. As a result, mutants tend to compete more with other mutants than they would if the population were unstructured. This phenomenon, which we call *positive competitive assortment*, reduces the ability for selection and drift to purge deleterious and neutral mutant lineages from the population. Single-mutant lineages are therefore more likely to survive long enough to successfully tunnel across the valley by producing a double mutant than in unstructured populations.

Komarova (2006) showed that spatial structure can accelerate tunneling across valleys and plateaus by this mechanism. Structure was modeled in the form of a population where each each individual occupies a unique site on a one-dimensional lattice, competes with its

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

nearest neighbors, disperses to adjacent sites, and does not migrate. In this model, single-mutant lineages form contiguous colonies that change slowly in size since genetic change occurs only at the colony's two edges. As a result, single-mutant lineages survive longer and produce more double mutants, which can lead to faster tunneling on average. In contrast, Takahasi (2007) found no effect of population subdivision on plateau crossing in the absence of recombination; however, their mutation model differs from that of Komarova (2006) by allowing each mutation to occur only once in the history of the population.

The results of Komarova (2006) show that structure can in principle accelerate valley crossing; however, questions remain about when this effect will be significant, which we address in the present study. First, how do the effects of spatial structure on valley crossing vary with the type of structure (e.g., subdivided versus continuous habitats) and the degree of structure (e.g., migration rates)? Second, do limits exist for the extent to which spatial structure can accelerate valley crossing? In other words, are there situations where increasing the degree of structure does not decrease the waiting time for the population to adapt?

For certain applications, the rare times when the valley is crossed very quickly are of primary importance; examples include the development of cancer within a human lifetime (Weissman *et al.* 2010) or if crossing the valley before a mutually exclusive beneficial mutation occurs could lead the population down a divergent long-term evolutionary trajectory. In unstructured populations, the dynamics associated with quickly crossing the valley differ in important ways from those associated with crossing the valley in a roughly average amount of time (Weissman *et al.* 2010). Therefore, we also ask how spatial structure affects the probability that the population crosses the valley within a period of time that is much shorter than the average.

To address these questions, we analyzed how population structure affects the dynamics of adaptation across a two-mutation valley or plateau in an asexual population that is subdivided into discrete subpopulations, or *demes*, with occasional migration between demes. Subdivision may have particularly large effects on tunneling, since a rare lineage that becomes

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

fixed within a deme only competes with itself and not with other genotypes. Subdivision is also relevant for the evolution of pathogens and cancer development, as human tissues create subdivided habitats. The population of stem cells from which colon cancers arise, for example, are subdivided into $\sim 10^7$ crypts (Tomasetti and Vogelstein 2015), each with an independently evolving subpopulation of ~ 5 stem cells that are susceptible to developing into cancer (Baker *et al.* 2014; Vermeulen and Snippert 2014), and populations of cells carrying HIV are subdivided into pulps within the spleens of infected patients (Frost *et al.* 2001).

We first describe the general mechanism by which subdivision affects valley crossing and then give a complete description of valley-crossing dynamics for the finite island model of subdivision, in which migration between any pair of demes occurs equally frequently. We consider the effects of varying the degree of spatial structure in the island model by varying the size of demes or the migration rates between them. Though the island model is unrealistic for many species, many of our results are general and provide fundamental insights into how subdivision, and spatial structure more generally, affect valley crossing. Section 7 considers how the results, methods, and intuition we present can be extended to models where migration occurs primarily between nearby demes.

2 Model

We consider the process by which a subdivided asexual population obtains an adaptation that requires two mutations before any fitness benefit is gained. The mutations leading to such adaptations can be of a variety of types. For example, two point mutations might be required in the same or different genes. Alternatively, for the case of inactivation of a tumor-suppressor gene in a clonally reproducing diploid cell, a nonsense or frameshift mutation may inactivate one copy of the gene and be followed by gene conversion to inactivate the second copy. We analyze how subdivision changes how the population adapts by a particular sequence of two mutations while ignoring the possibility of other mutations. This assumption simplifies our

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

analysis by allowing us to consider just three types of individuals—wild-type individuals that carry zero mutations and individuals carrying either one or two mutations. Wild types, single mutants, and double mutants have relative fitnesses of 1, $1-\delta$, and $1+s$, respectively, with $\delta \leq 0$ and $s > 0$ (Figure 1). Wild types mutate to single mutants at rate μ_0 and single mutants mutate to double mutants at rate μ_1; for simplicity, we ignore back mutations. In general, a population can adapt by multiple mutational pathways; valley-crossing dynamics can then be found from our results using the method described in Weissman *et al.* (2009).

The population consists of N_{T} individuals subdivided into L demes, with each individual competing and reproducing locally, that is, within its own deme. The number of individuals within a deme is constant and independent of the average fitness of the deme. Migration exchanges individuals between demes but does not alter deme size or the total frequency of genotypes within the population; thus, in our model selection is *soft* and migration is *conservative* (Nagylaki 1980; Whitlock 2004; Charlesworth and Charlesworth 2010). With these assumptions, subdivision affects tunneling by the same general mechanism as in the lattice model of Komarova (2006). We describe how subdivision qualitatively affects valley crossing under a general model where demes may differ in size and allowing for any pattern of migration, so long as each deme can be reached from any other in a finite time. Our quantitative analysis assumes the finite island model, for which each deme has N individuals that each migrate at rate m per generation to a different, randomly chosen deme.

The rate of genetic drift is controlled by a parameter, α, which we call the *drift coefficient*, defined as one-half of the variance in the number of descendants left by an individual after one generation. This definition is equivalent to $\alpha \equiv N/2N_{\mathrm{e}}$, where N_{e} is the variance effective population size of a deme of size N. Larger values of α correspond to smaller effective population sizes and faster stochastic changes in genotype frequencies with a deme for a given N. Our inclusion of the parameter α serves two purposes. First, in many cases the effects of subdivision on the evolution of a mutant lineage can be understood as decreasing the effective drift and selection coefficients. Second, our model can represent a variety of

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

reproduction models by appropriately choosing α; for example, the Moran reproduction model corresponds to $\alpha = 1$ and the Wright-Fisher model to $\alpha = 1/2$. Table 1 summarizes our model parameters.

We make several additional assumptions in order to facilitate our analysis and discussion. First, we assume that $s \gg \alpha/N_{\mathrm{T}}$ so that the double mutant has a significant selective advantage over wild type. Second, we assume that $\mu_1 \ll \max\{\delta, s\}$, which, for reasons described below, is required for spatial structure to accelerate valley crossing. Third, we assume that the dynamics of genotype frequencies within demes can be described using the standard diffusion approximation from population genetics (Ewens 2004). In practice, this assumption requires that the census and effective sizes of demes are large ($N \gg 1$ and $N/\alpha \gg 1$), selection has a small effect on an individual's reproductive success ($\delta \ll \alpha$ and $s \ll \alpha$), and that third and higher moments of the distribution of the number of descendants per individual per generation are negligible (i.e., the distribution is characterized only by its mean and variance). Finally, unless otherwise stated we assume that the population is large enough that valley crossing occurs by tunneling—that is, single mutants remain rare in the total population until the double mutant begins to sweep.

We verified our predictions using computer simulations of a Wright-Fisher population. Additional details for our simulations are given in Supplementary Appendix A. Source code for simulations and figures is available at https://github.com/mmclaren42/valley-crossing-subdivided.

3 Summary of results

Our goal is to find the average and probability distribution of the time $\mathcal{T}$ when, starting from an entirely wild-type population, a double mutant first arises whose lineage is destined to sweep through the population, as well as the average time for the double mutant to become fixed within the population.

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

For the population to adapt across the valley, a single-mutant lineage must survive long enough to produce a double-mutant lineage that is destined to survive and sweep to fixation. We refer to the single and double mutants that give rise to such lineages, and the lineages themselves, as *successful*. Figure 2 illustrates the process of adaptation in a population that is not too large or too heavily subdivided. We let $\mathcal{T}_0$ be the waiting time for the first successful single mutant; the *drift time*, $\mathcal{T}_1$, be the time that the first successful single-mutant lineage drifts before producing a successful double mutant; and the *sweep time*, $\mathcal{T}_{\mathrm{sw}}$, be the time for the successful double-mutant lineage to sweep to fixation. This and other notation is reviewed in Table 2.

In Figure 2, the first successful single-mutant lineage produces the first successful double mutant, which in turn becomes the most recent common ancestor of the population once the double-mutant genotype is fixed. Therefore, the time when the first successful double mutant occurs is $\mathcal{T} = \mathcal{T}_0 + \mathcal{T}_1$ and the time for the population to become fixed for the double mutant is $\mathcal{T}_0 + \mathcal{T}_1 + \mathcal{T}_{\mathrm{sw}}$. We call this scenario *stochastic tunneling* to distinguish it from other tunneling scenarios that can occur in very large or very heavily subdivided populations. A sufficient condition for stochastic tunneling is that the average time $\langle \mathcal{T}_0 \rangle$ is much larger than both $\langle \mathcal{T}_1 \rangle$ and $\langle \mathcal{T}_{\mathrm{sw}} \rangle$. In this case, the total time for the population to adapt is typically dominated by $\mathcal{T}_0$. Let p_1 be the probability that a single mutant is successful. Successful single mutants are produced as a Poisson process with rate $N_{\mathrm{T}} \mu_0 p_1$, and so the average waiting time for the first successful single mutant is

$$\langle \mathcal{T}_0 \rangle \approx \frac{1}{N_{\mathrm{T}} \mu_0 p_1}. \tag{1}$$

Subdivision accelerates stochastic tunneling by increasing p_1 above its corresponding value in unstructured populations, $p_{1,\mathrm{wm}}$ (the subscript "wm" stands for "well-mixed"). Increases in p_1 are primarily attributable to the effects of competitive assortment, which reduces the rates of selection and drift experienced by single-mutant lineages and so allows them to

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

survive longer and produce more double mutants.

The quantitative effect subdivision has on a given single-mutant lineage depends roughly on the average degree of competitive assortment experienced by the lineage before producing a successful double mutant. We measure competitive assortment with a statistic F that quantifies how much more likely a pair of individuals from the same deme are to share a genotype than a pair drawn at random from the total population. The statistic F is equivalent to a version of Wright's fixation index F_{ST} defined in Whitlock (2002); it varies between zero and one, with $F = 0$ when a lineage has equal frequency across all demes and $F = 1$ when a lineage is either fixed or absent in each deme. Increasing the degree of subdivision by decreasing the size of demes or the migration rate increases the typical values of F experienced by single-mutant lineages and thus increases p_1. Large increases in p_1, of an order of magnitude or more, only occur if successful single-mutant lineages typically fix in one or more demes before producing a successful double mutant and so have average values of F close to 1.

Successful single-mutant lineages will only tend to fix within a deme if both N and m are small. Demes must have substantially fewer than $N_\times$ individuals, where $N_\times$ is the size of an unstructured population in which valley crossing is equally likely to occur by tunneling or sequential fixation. If $N \gg N_\times$, then $p_1 \approx p_{1,\mathrm{wm}}$, since, even at low migration rates, successful single-mutant lineages remain at low frequency within individual demes and always have small assortments. If $N \ll N_\times$ and m is sufficiently small, then successful lineages typically fix within their initial deme and $p_1 \gg p_{1,\mathrm{wm}}$.

When $N \ll N_\times$, the probability p_1 increases over a wide range of decreasing migration rates, to a maximum $\lim_{m\to 0} p_1 = \psi_{0\to 1}\theta$, where $\psi_{0\to 1}$ is the probability that a single-mutant lineage fixes in a wild-type deme and θ is the probability that a double-mutant lineage that has fixed within a deme goes on to fix in the total population. A sufficient condition for $p_1 \approx \psi_{0\to 1}\theta$ is that migration is rare enough that a single-mutant deme always transitions into a double-mutant deme before being displaced by a wild-type migrant.

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Variation in p_1 for higher migration rates depends critically on whether selection against single mutants is strong enough to inhibit their fixing within individual demes. If $\delta \ll \alpha/N$, then single mutants are *locally neutral*—they have a neutral probability $\approx 1/N$ of fixing within a deme when migration is rare. (Such mutants are still deleterious with respect to fixation in the total population if $\delta \gg \alpha/N_{\mathrm{T}}$.) If $\delta \gg \alpha/N$, then single mutants are *locally deleterious*, and are much less likely than a neutral mutation to fix within a deme due to selection.

If single mutants are locally neutral and the number of demes is large, the effects of subdivision can be found by substituting effective drift and selection parameters into results for unstructured populations. These effective parameters are $\alpha_{\mathrm{T}} \equiv (1-\hat{F})\alpha$ and $\delta_{\mathrm{T}} \equiv (1-\hat{F})\delta$, where $\hat{F} \approx 1/(1+Nm/\alpha)$ is the equilibrium assortment in the infinite neutral island model. Subdivision has a large effect for small migration rates $m \ll \alpha/N$, for which the effective parameters are greatly reduced below their unstructured values.

If single mutants are locally deleterious, successful single-mutant lineages will either, depending on the migration rate, drift to $\sim \alpha/\delta \ll N$ copies before producing a successful double mutant, and so never fix within a deme, or drift to $\approx N$ copies and fix within their initial deme. The latter dominate for migration rates $m \ll \eta\alpha/N$, where η is a constant that is $\ll 1$ if $\delta \gg \alpha/N$.

Finding the extent to which subdivision accelerates valley crossing requires that we also consider the drift and sweep times. We find that subdivision always increases the average drift and sweep times, by factors at least as large as p_1. Extreme subdivision therefore makes it much more likely for tunneling to be limited by the drift and sweep times. Moreover, these increases in the drift and sweep times can greatly limit the degree to which subdivision accelerates tunneling. The drift time limits how much subdivision can decrease the waiting time for the successful double mutant. We find that subdivision can decrease the average waiting time for a successful double mutant to a minimum of $\sqrt{\pi\alpha/2N_{\mathrm{T}}\mu_0\mu_1 s}$. This minimum value is already obtained by an unstructured population if $N_{\mathrm{T}}\mu_0/\alpha \gg \max\{1, \delta^2/\mu_1 s\}$; in

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

such cases, subdivision has no effect on the average (or on the distribution) of $\mathcal{T}$. The sweep time is proportional to α/Nm for small migration rates $m \ll \alpha/N$, and so will dominate the waiting time for the double mutant to fix in heavily subdivided populations.

Increases in the drift time also limit the degree to which subdivision can increase the probability that the valley is crossed unusually quickly. The probability that a successful double mutant appears by time t can only be increased by subdivision to a maximum of $1 - \exp[-N_{\mathrm{T}}\mu_0\mu_1 s t^2/2\alpha]$. For t less than $\min\{1/\sqrt{\mu_1 s}, 1/\delta\}$, the average drift time in an unstructured population, this maximum occurs even without subdivision and subdivision has no effect on the probability, even if it greatly decreases the average, $\langle\mathcal{T}\rangle$.

4 Fate of a single-mutant lineage

This section describes how subdivision affects the fate of a single-mutant lineage—in particular, the probability that the lineage is successful and the drift and sweep times. Sections 5 and 6 use these results to find the average and distribution of the waiting time for the population to adapt, including cases where the time is not dominated by the waiting time for a successful single mutant.

4.1 Qualitative effects of subdivision

Before presenting our quantitative results for the island model, we first describe the general mechanism by which subdivision increases the probability that a single-mutant lineage is successful as well as the drift and sweep times.

Subdivision and lineage dynamics: To see how subdivision affects the dynamics of the total copy number of a mutant lineage, we consider a mutant lineage with selection coefficient f segregating in an otherwise wild-type population. We let $n(t)$ be the number of mutant individuals at time t, n_i be the number of mutants in the i-th deme, and N_i be

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the size of the i-th deme, so that $N_{\mathrm{T}} = \sum_i N_i$ and $n = \sum_i n_i$. It is often useful to consider frequencies instead of numbers, and so we also let $x \equiv n/N_{\mathrm{T}}$ be the frequency of mutants in the total population and $x_i \equiv n_i/N_i$ be the frequency of mutants in the i-th deme, so that $x = \sum_i (N_i/N_{\mathrm{T}})x_i$. Our goal is to understand how competitive assortment in subdivided populations influences how $n(t)$ (or $x(t)$) changes over time.

We quantify the competitive assortment at time t with a measure F describing the increase in probability that a pair of competing individuals in the population share a genotype over the probability for a random pair of individuals. First, we consider sampling two individuals with replacement from the total population, and let J_T be the probability that the two individuals share a genotype, or are *identical by state*. Therefore,

$$J_T = x^2 + (1-x)^2 \tag{2}$$

where x is the current frequency of the mutation. Next, we consider sampling two individuals by sampling the first from the total population and the second from the set of individuals that compete with the first. For the subdivision models we consider, this set includes all individuals within the deme of the first individual, including the first individual itself. By conditioning on which deme the first individual is chosen from, we can see that

$$J_S = \sum_{i=1}^{L} \left(\frac{N_i}{N_{\mathrm{T}}}\right) \left[x_i^2 + (1-x_i)^2\right]. \tag{3}$$

It follows by Jenson's inequality that if mutants are evenly spatially distributed—that is, if $x_i = x$ for all i—then $J_S = J_T$, but if mutants are unevenly distributed, then $J_S > J_T$. (This fact is equivalent to the well-known Wahlund effect by which subdivision leads to departures from Hardy-Weinberg proportions in randomly mating diploid species (Charlesworth and Charlesworth 2010).) Intuitively, if mutants are unevenly distributed, they tend to be found in demes with a greater than average frequency of mutants. We define the competitive

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

assortment F as the difference between J_S and J_T normalized so that $0 \leq F \leq 1$, or

$$F \equiv \frac{J_S - J_T}{1 - J_T}. \tag{4}$$

By definition, $F = 0$ when mutants are evenly distributed and $F = 1$ when mutants are either fixed or absent in every deme.

We briefly mention some similarities and differences between the definition (4) and related quantities used in other evolutionary studies. The quantity F is equal to G_{ST} as defined in Nei (1973) for a biallelic locus if demes are weighted by their size and is also equal to F_{ST} as defined in Whitlock (2002) and Whitlock (2003). We emphasize, however, that our definition differs from other quantities also called F_{ST} and more general "relatedness" or "inbreeding" coefficients (*sensu* Rousset (2002)) in important ways. These quantities as typically defined are deterministic and as such are effectively parameters of the model (Rousset 2002). In contrast, here F is a summary statistic of the current state of the population. Some, but not all, of the theoretical knowledge relating to these other quantities carries over when considering the values of F that are experienced by mutant lineages. Our reason for the definition of F used here is that it provides a direct way of understanding, with minimal assumptions, how competitive assortment affects the evolution of $x(t)$.

The change in x over a short period of time due to selection and drift is characterized by the mean and variance of the change in x over one generation. The mean, M_x, describes the action of selection, or the deterministic change in x, while the variance, V_x, describes the action of drift, or the stochastic change in x. Given the current state of the population, the mean and variance can be expressed in terms of x and F (Supplementary Appendix B) as

$$M_x = (1 - F) f x (1 - x) \tag{5}$$

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and

$$V_x = (1 - F)\frac{2\alpha}{N_{\mathrm{T}}}x(1 - x). \tag{6}$$

When $F = 0$, Equations (5) and (6) reduce to the familiar expressions for the mean and variance in an unstructured population with mutant frequency x. In general, however, both M_x and V_x are reduced by a factor $1 - F$ below their respective unstructured values. This reduction occurs because the fraction of competitive interactions in the population that take place between individuals with different genotypes is reduced by a factor $1 - F$ relative to in an unstructured population.

Even without saying anything about how F evolves, we can use Equations (5) and (6) to make some general conclusions about how subdivision affects the long term evolution of the total frequency of the lineage. Starting from a initial frequency $x(0)$, whether $x(t)$ first reaches a frequency A before a frequency B (with $A < x(0) < B$) is determined by M_x/V_x in the interval between A and B, and is therefore independent of subdivision (Supplementary Appendix B). This is a more general statement of Muruyama's well-known result (Maruyama 1970; Maruyama 1974) that the probability that the lineage reaches fixation is independent of subdivision. If we consider the trajectories of x that reach A from $x(0)$ (or B from $x(0)$), the frequency will tend to change more slowly in subdivided than in unstructured populations at every intermediate frequency. Compared to an equivalent trajectory in an unstructured population, the time taken to reach A (or B) will be longer by a factor of $1/(1 - \bar{F})$, where $\bar{F}$ is the average value of F over the trajectory (Supplementary Appendix B).

Given our above observations, we can define the probability that a lineage with selection coefficient f initially present in n_0 copies reaches k copies before going extinct (with $0 \leq n_0 \leq k \leq N_{\mathrm{T}}$), which we denote $P_f(n_0, k)$, independently of population structure. This probability can be approximated using standard diffusion methods for unstructured

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

populations (reviewed in Ewens (2004)) and is

$$P_f(n_0, k) \approx \frac{1 - e^{-n_0 f/\alpha}}{1 - e^{-kf/\alpha}}. \tag{7}$$

The specific case $n_0 = 1$ and $k = N_\mathrm{T}$ corresponds to the familiar diffusion approximation for the probability of fixation of a mutant lineage when initially present in a single copy.

Fate of a single-mutant lineage: Subdivision can potentially affect the probability that the single-mutant lineage is successful both through the number of double mutants it produces and through the probability that a double-mutant lineage is successful.

A double-mutant lineage is successful if it is destined for fixation instead of going extinct. The probability that a double-mutant lineage is successful, which we denote p_2, is not entirely independent of subdivision due to the presence of single mutants in the population. If the population were all wild-type, then $p_2 = P_s(1, N_\mathrm{T})$ independently of subdivision. For $s \gg \alpha/N_\mathrm{T}$, this probability is $\approx s/\alpha$, which is usefully interpreted as the reciprocal of the number α/s above which the double-mutant lineage is *established*, or very likely to survive drift and sweep through the population (Desai and Fisher 2007). Since single mutants are rare in the total population, in unstructured populations the probability p_2 remains $\approx s/\alpha$. In subdivided populations, however, single mutants may be frequent in one or more demes and double mutants in primarily single-mutant demes will have a significant extra advantage if $\delta \gtrsim s$. In addition, new double-mutant lineages in subdivided populations usually occur in demes with greater than average frequencies of single mutants—if x and F are the frequency and assortment of single mutants, then the expected frequency of single mutants in the deme where a double mutation occurs is $F + (1 - F)x$. As a result, subdivision may increase p_2, although we will see that the effect this has on p_1 is usually fairly small.

The number of double mutants produced by a single-mutant lineage is proportional to the total number of single mutant descendants of the lineage, or its *weight* (Weissman *et al.* 2009). The weight W is equal to $\int_0^\infty n(t)\, dt$, where $n(t)$ is the number of single mutants at

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

time t, and a single-mutant lineage with weight W on average produces $W\mu_1$ double mutants. Subdivision increases the weights of single-mutant lineages, since a lineage is equally likely to reach a maximum of k copies in subdivided and unstructured populations, but those in subdivided populations drift for longer in the population. The increase in weight for a given lineage that can be attributable to subdivision grows with the assortment experienced by the lineage, particularly during times when $n(t)$ is larger.

Overall, subdivision increases the probability p_1 that a single-mutant lineage is successful by increasing W without decreasing p_2. The magnitude of the increase in p_1 depends on the assortments experienced by single-mutant lineages; significant increases can only occur if successful lineages tend to have assortments $F \sim 1$. Since F is bounded from above the largest frequency of mutants within any individual deme, large assortments require that the lineage has become frequent in at least one deme. Increases in p_1 by an order of magnitude or more require $F \approx 1$—in other words, successful lineages must typically fix in one or more demes.

Competitive assortment also causes the drift and sweep times to increase in subdivided populations. The drift time increases because single-mutant lineages with high assortments can survive longer and are also less likely to grow very large and produce many double mutants very quickly. The sweep time increases because double-mutant lineages can also have high assortments and so take longer to spread through the population.

4.2 Unstructured populations

This section provides the results for unstructured populations needed to describe tunneling in subdivided populations. Previous studies of tunneling in unstructured populations have assumed either the Wright-Fisher model of reproduction, for which $\alpha = 1/2$, or the Moran model of reproduction, for which $\alpha = 1$. Supplementary Appendix C uses an analytical approach similar to Weissman *et al.* (2009) and Weissman *et al.* (2010) to account for general variation in α. Here, we summarize the main results and explain the intuition using heuristic

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

arguments also adapted from these earlier studies.

It will be useful later on to distinguish between two ways the drift coefficient, α, influences the fate of a single-mutant lineage. First, the drift coefficient determines how quickly the number of single mutants changes by drift. Conditional that a lineage, initially present in a single copy, reaches a number $k \lesssim \alpha/\delta$ before going extinct, it will typically do so in $\sim k/\alpha$ generations. The lineage will typically then go extinct over an additional $\sim k/\alpha$ generations, thereby accumulating a weight $\sim k^2/\alpha$ over its lifetime. A single-mutant lineage has a probability $\approx 1/k$ of reaching a number $k \ll \alpha/\delta$, but is much less likely to drift to a number $k \gg \alpha/\delta$; we can understand this observation by noting that the time required to drift to $k \gg \alpha/\delta$ is much greater than the time $\sim 1/\delta$ for selection to significantly reduce the lineage number. In order to be able to translate between the time a lineage has survived and the number of the lineage, we also note that a lineage that has survived for $t \lesssim 1/\delta$ generations will typically have grown to a number $\sim \alpha t$. These observations follow by extending those of Fisher (2007) and Weissman *et al.* (2009), given for $\alpha = 1$, to general α.

The second effect of the drift coefficient occurs in determining the probability p_2 that a double mutant produced by the single-mutant lineage is successful. Recall that the probability p_2 that a double-mutant lineage is successful in an unstructured population is $\approx s/\alpha$, the reciprocal of the number at which the lineage becomes established. Intuitively, a double-mutant lineage that has drifted to a number $\gg \alpha/s$ is unlikely to drift to extinction within $\sim 1/s$ generations, by which time the lineage will have grown substantially from selection. To distinguish the effect of α on p_2 from its effect on the rate of stochastic changes in single-mutant lineages, we initially treat the probability p_2 as a new parameter of our model. Later, we substitute $p_2 \approx s/\alpha$ to find results in terms of our original parameters. Meanwhile, our assumption that $\mu_1 \ll \max\{\delta, s\}$ is taken to be equivalent to $\mu_1 \ll \max\{\delta, \alpha p_2\}$.

Tunneling dynamics qualitatively differ depending on the strength of selection against

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

single mutants. The probability that a single mutant is successful is

$$p_1 \approx \begin{cases} \sqrt{\frac{\mu_1 p_2}{\alpha}} & \text{for } \delta \ll \sqrt{\alpha \mu_1 p_2}, \\ \frac{\mu_1 p_2}{\delta} & \text{for } \delta \gg \sqrt{\alpha \mu_1 p_2}. \end{cases} \tag{8}$$

For $\delta \ll \sqrt{\alpha \mu_1 p_2}$, single mutants are (effectively) neutral for tunneling—the probability p_1 does not depend on the strength of selection δ. In contrast, if $\delta \gg \sqrt{\alpha \mu_1 p_2}$, then mutants are deleterious for tunneling and p_1 is inversely proportional to δ. The average drift time for neutral and deleterious single mutants is

$$\langle \mathcal{T}_1 \rangle \approx \begin{cases} \frac{\ln 2}{\sqrt{\alpha \mu_1 p_2}} & \text{for } \delta \ll \sqrt{\alpha \mu_1 p_2}, \\ 1/\delta & \text{for } \delta \gg \sqrt{\alpha \mu_1 p_2}. \end{cases} \tag{9}$$

The full distribution of $\mathcal{T}_1$ is found in Supplementary Appendix C. The critical feature of this distribution is that a successful lineage is very unlikely to drift for much fewer or much more than $\langle \mathcal{T}_1 \rangle$ generations. Instead, successful lineages typically drift for $\sim \langle \mathcal{T}_1 \rangle$ generations and reach a number $\sim \alpha \langle \mathcal{T}_1 \rangle$, before producing a successful double mutant.

We can understand the expressions for p_1 and $\langle \mathcal{T}_1 \rangle$ in (8) and (9) in terms of the behavior of typical successful lineages when single mutants are either neutral or deleterious. Conditional on a single-mutant lineage having weight W, it produces at least one successful double mutant with probability $1 - e^{-W \mu_1 p_2}$. Lineages typically become successful in one of two ways (Weissman *et al.* 2009; Weissman *et al.* 2010): either they have a large weight $W \gtrsim 1/\mu_1 p_2$ and thus are nearly assured to produce at least one successful double mutant, or they have a small weight $W \ll 1/\mu_1 p_2$, yet manage to produce exactly one successful double mutant. Which strategy is used depends on whether or not selection is strong enough to limit the single-mutant lineages to small numbers and thus to small weights.

Successful neutral lineages follow the first strategy of having a large weight: they drift

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

to numbers $\sim \sqrt{\alpha/\mu_1 p_2}$ over $\sim 1/\sqrt{\alpha\mu_1 p_2}$ generations, reaching a weight $\sim 1/\mu_1 p_2$. The probability p_1 in (8) corresponds to the probability of drifting to such a number, while $\langle \mathcal{T}_1 \rangle$ in (9) roughly corresponds to the time required to do so. These lineages strike a sweet spot for producing a successful double mutant; lineages that only reach smaller numbers have much smaller weights, while lineages that reach larger numbers have larger weights but are not significantly more likely to produce a successful double mutant. The condition $\delta \ll \sqrt{\alpha\mu_1 p_2}$ ensures that the lineages can drift to a number $\sim \sqrt{\alpha/\mu_1 p_2}$ neutrally (equivalently, it ensures that the time $1/\sqrt{\alpha\mu_1 p_2}$ required to do so is smaller than the time $\sim 1/\delta$ for selection to purge the lineage (Weissman *et al.* 2009)).

If $\delta \gg \sqrt{\alpha\mu_1 p_2}$, selection limits single-mutant lineages to much smaller numbers and weights, forcing successful deleterious lineages to follow the second strategy. Since almost all deleterious lineages have weights $\ll 1/\mu_1 p_2$, we can approximate the probability that a lineage produces a successful double mutant as $\langle W \rangle \mu_1 p_2$, where $\langle W \rangle$ is the average weight of a deleterious lineage and is approximately $1/\delta$. The dominant contribution to $\langle W \rangle$, and thus to p_1, comes from lineages that drift to numbers $\sim \alpha/\delta$ over periods of $\sim 1/\delta$ generations. Although lineages that drift to larger numbers tend to have larger weights, these lineages are so unlikely that they make a negligible contribution p_1. These observations explain the expressions $p_1 \approx \mu_1 p_2/\delta$ in (8) and $\langle \mathcal{T}_1 \rangle \approx 1/\delta$ in (9).

Substituting $p_2 = s/\alpha$ into (8) and (9) gives expressions for p_1 and $\langle \mathcal{T}_1 \rangle$ in terms of our basic model parameters. These expressions are

$$p_{1,\mathrm{wm}} \approx \begin{cases} \frac{\sqrt{\mu_1 s}}{\alpha} & \text{for } \delta \ll \sqrt{\mu_1 s}, \\ \frac{\mu_1 s}{\alpha\delta} & \text{for } \delta \gg \sqrt{\mu_1 s}, \end{cases} \tag{10}$$

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and

$$\langle \mathcal{T}_{1,\mathrm{wm}} \rangle \approx \begin{cases} \frac{\ln 2}{\sqrt{\mu_1 s}} & \text{for } \delta \ll \sqrt{\mu_1 s}, \\ 1/\delta & \text{for } \delta \gg \sqrt{\mu_1 s}. \end{cases} \tag{11}$$

We have used the notation $p_{1,\mathrm{wm}}$ in (10) and $\langle \mathcal{T}_{1,\mathrm{wm}} \rangle$ in (11) to enable comparison between these expressions and our results for p_1 and $\langle \mathcal{T}_1 \rangle$ in subdivided populations (the subscript "wm" stands for "well-mixed").

We conclude our discussion of tunneling in unstructured populations by returning to our assumption that $\mu_1 \ll \max\{\delta, s\}$ (or that $\mu_1 \ll \max\{\delta, \alpha p_2\}$). Equations (10) and (11) show that this assumption is equivalent both to $p_1 \ll s/\alpha$ and to $\langle \mathcal{T}_1 \rangle \ll 1/\mu_1$. If instead $\mu_1 \gg \max\{\delta, s\}$, then $p_1 \approx s/\alpha$ and $\langle \mathcal{T}_1 \rangle \approx 1/\mu_1$; in other words, $p_2 \approx s/\alpha$ forms an upper bound for p_1 and $1/\mu_1$ an upper bound for $\langle \mathcal{T}_1 \rangle$. By assumption, these upper bounds are not met if the population is unstructured, but we will see that they can be reached if the population is sufficiently subdivided.

Sequential fixation in small populations: In sufficiently small populations, successful lineages are likely to fix in the population before producing a successful double mutant and valley crossing will occur by sequential fixation. More generally, we can compare the probabilities that a single mutant is successful by the tunneling and sequential pathways to determine which is more likely. Since we do not allow back mutation, a single-mutant lineage that reaches fixation always produces a successful double mutant. Therefore, the probability that a single-mutant lineage is successful by sequential fixation is simply the probability that the lineage drifts to fixation, equal to $P_{-\delta}(1, N_\mathrm{T})$ or approximately

$$P_{-\delta}(1, N_\mathrm{T}) \approx \begin{cases} \frac{1}{N_\mathrm{T}} & \text{for } N_\mathrm{T} \ll \frac{\alpha}{\delta} \\ \frac{\delta}{\alpha e^{N_\mathrm{T}\delta/\alpha}} & \text{for } N_\mathrm{T} \gg \frac{\alpha}{\delta}. \end{cases} \tag{12}$$

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

After previous authors (Weinreich and Chao 2005; Weissman *et al.* 2009), we define the threshold population size $N_\times$ as the size where the probability p_1 of success by tunneling equals $P_{-\delta}(1, N_\mathrm{T})$. If $N_\mathrm{T} \gg N_\times$, then $p_1 \gg P_{-\delta}(1, N_\mathrm{T})$ and tunneling is much more likely than sequential fixation, while the opposite is true if $N_\mathrm{T} \ll N_\times$. The threshold size if single mutants are neutral or deleterious for tunneling, found by assuming the approximation (10) for p_1, is

$$N_\times \approx \begin{cases} \frac{\alpha}{\sqrt{\mu_1 s}} & \text{for } \delta \ll \sqrt{\mu_1 s} \\ \frac{\alpha}{\delta} \ln\left(\frac{\delta^2}{\mu_1 s}\right) & \text{for } \delta \gg \sqrt{\mu_1 s}. \end{cases} \tag{13}$$

Equation (13) is valid if $N_\times \gg \alpha/s$, so that $p_2 \approx s/\alpha$ for $N_\mathrm{T} \sim N_\times$. This requirement is always met if $\delta \lesssim s$, but in general may not be; see Supplementary Appendix D for the general solution for $N_\times$. For neutral single mutants, $N_\times$ equals the typical copy number $\alpha/\sqrt{\mu_1 s}$ reached by successful lineages in the tunneling pathway. For deleterious single mutants, $N_\times$ is smaller than $\alpha/\sqrt{\mu_1 s}$, but is still substantially larger than the typical number α/δ reached by successful tunneling lineages. When $N_\mathrm{T} \approx N_\times$ in the deleterious case, the fact that selection greatly limits the probability that a lineage drifts to N_T copies is exactly offset by the assurance of producing a successful double mutant if it does so.

4.3 Limit as $m \to 0$ and the isolated-demes regime

Demes must be small enough for subdivision to increase the probability that a single mutant is successful. Otherwise, successful single-mutant lineages will produce a successful double mutant without ever reaching high frequency in a deme, even under extremely restricted migration. The only significant effect of subdivision in such cases is possibly increasing the sweep time.

Here, we determine how small demes must be for subdivision to potentially increase p_1 by a large amount by considering the fate of single-mutant lineages in the limit $m \to 0$—the

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

best case scenario for subdivision to increase p_1. Valley-crossing dynamics are easily studied in the limit, since a single-mutant lineage must either go extinct or produce a double mutant that fixes in the initial deme before migration in or out of that deme can occur. Once fixed within the deme, a double-mutant lineage has a probability $P_s(N, N_\mathrm{T}) \equiv \theta$ of then fixing in the total population, approximately equal to

$$\theta \approx 1 - e^{-Ns/\alpha}. \tag{14}$$

As in unstructured populations, we consider two ways a single-mutant lineage might produce a successful double mutant—by tunneling within the deme, or by first reaching fixation in the deme. The probability that a single mutant is successful by tunneling within the deme is approximately $p_{1,\mathrm{wm}}$, the probability of tunneling in an unstructured population of size N_T, while the probability the mutant is successful by first fixing in the deme is $P_{-\delta}(1, N)\theta$. The two are approximately equal when $N = N_\times$, and for demes with much more than $N_\times$ individuals, tunneling is much more likely, while for demes with much fewer than $N_\times$ individuals, fixation is much more likely.

At least a fraction of demes must have sizes $N \ll N_\times$ for subdivision to have a large effect on single-mutant lineages at low migration rates. Since we are interested in determining when subdivision can significantly accelerate valley crossing, we assume $N \ll N_\times$ for the remainder of our analysis. We will also assume that $N_\mathrm{T} \gg N_\times$, so that tunneling is the dominant mode of valley crossing at high migration rates. This assumption simplifies our analysis by ensuring that single-mutant lineages remain rare in the total population across all migration rates; we describe these lineages as tunneling with respect to the total population even if they fix in one or more demes. It is also possible for subdivision to accelerate valley crossing when $N_\mathrm{T} \ll N_\times$. Although we do not explicitly consider this situation, it can be analyzed straightforwardly by combining our results for the rate of tunneling in subdivided populations to that of sequential fixation, which is unaffected by subdivision.

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Over the next two sections, we will determine how p_1 and $\mathcal{T}_1$ vary as a function of migration rate, m, while the size, N, of demes is fixed so that $N \ll N_\times \ll N_\mathrm{T}$, and find it useful to keep in mind the above dynamics as occurring for sufficiently low migration rates. These dynamics are characterized by new single-mutant lineages having a probability $\psi_{0\to1}$ of fixing within their initial deme and, once fixed, being guaranteed to produce a double-mutant deme before being replaced by a wild-type migrant or colonizing additional demes. Since migration does not play a role in determining the fate of these lineages, we refer to the range of parameters where such dynamics apply the *isolated-demes regime*.

In the isolated-demes regime, the probability that a single mutant is successful is approximately

$$p_1 \approx \psi_{0\to1}\theta, \tag{15}$$

and the drift time is approximately exponentially distributed with mean

$$\langle \mathcal{T}_1 \rangle \approx \frac{1}{N\mu_1\psi_{1\to2}}, \tag{16}$$

which follows since the single-mutant deme produces double mutants at rate $N\mu_1$ that each have a probability $\psi_{1\to2}$ of fixing in the deme. These expressions for p_1 and $\langle \mathcal{T}_1 \rangle$ are decreasing in N and are significantly greater than $p_{1,\mathrm{wm}}$ and $\langle \mathcal{T}_{1,\mathrm{wm}} \rangle$ if and only if $N \ll N_\times$. Both (15) and (16) become insensitive to N at small numbers—in particular, $p_1 \approx s/\alpha$ and $\langle \mathcal{T}_1 \rangle \approx \mu_1$ for $N \ll \min\{\alpha/\delta, \alpha/s\}$. These values for p_1 and $\langle \mathcal{T}_1 \rangle$ are the maximum that occur as the degree of subdivision is increased by decreasing N and m.

We conclude our discussion of the isolated-demes regime by determining the range of migration rates for which it applies. To do so, we use the fact that when m and N are sufficiently small, we can approximate fixation or loss of a lineage that has entered a deme by migration or mutation as being practically instantaneously. Therefore, we approximate a single-mutant lineage that has fixed in its initial deme by a *deme birth-death (DBD) process*,

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in which single-mutant demes can give birth by colonizing a new deme, die by being replaced by a wild-type migrant, and mutate into a double-mutant deme. To simplify notation for the rates of these events, we let $\psi_{i\to j}$ denote the probability that an individual with j mutations fixes within a deme currently occupied by individuals with i mutations. For example, a single-mutant lineage fixes in a wild-type deme with probability $\psi_{0\to 1} = P_{-\delta}(1, N)$ and a wild-type lineage fixes in a single-mutant deme with probability $\psi_{1\to 0} = P_{\delta}(1, N)$. So long as single mutants are rare in the total population, each single-mutant deme gives birth at rate $B \equiv Nm\psi_{0\to 1}$, dies at rate $D \equiv Nm\psi_{1\to 0}$, and mutates at rate $U_1 \equiv N\mu_1\psi_{1\to 2}$ into a double-mutant deme. The isolated-demes regime corresponds to migration rates where $U_1 \gg \max\{B, D\}$, ensuring that a single-mutant deme always mutates before dying or giving birth; since $D \geq B$, this condition is equivalent to $U_1 \ll D$. We define m^* as the migration rate where $U_1 = D$, given by

$$m^* = \frac{\mu_1\psi_{1\to 2}}{\psi_{1\to 0}}. \tag{17}$$

The isolated-demes regime corresponds to $N \ll N_\times$ and $m \ll m^*$.

4.4 Locally neutral single mutants ($\delta \ll \alpha/N$)

How single-mutant lineages are impacted by low migration rates critically depends on whether or not these lineages can fix neutrally within a deme, making valley-crossing dynamics most easily understood by considering locally neutral and locally deleterious single mutants separately. This section presents results for p_1 and $\mathcal{T}_1$ when single mutants are locally neutral and the next presents results when single mutants are locally deleterious. The intermediate scenario that occurs when $\delta \sim \alpha/N$ is analyzed in Supplementary Appendix F.

Summary of methods: Analyzing valley crossing by modeling the frequency dynamics within each deme is intractable; therefore, we must develop approaches for approximating lineage dynamics. We develop two primary methods for predicting valley crossing when sin-

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

gle mutants are locally neutral by extending previous methods for approximating frequency dynamics in the island model. The $\hat{F}$ method, outlined below and detailed in Supplementary Appendix E, predicts p_1 and $\mathcal{T}_1$ when single-mutant lineages are locally neutral and typically drift to $\gg N$ copies before producing a successful double mutant. The DBD method, described in Section 4.3 and Supplementary Appendix C, predicts p_1 and $\mathcal{T}_1$ when single-mutant lineages typically fix in their initial deme before producing a successful double mutant. For $\alpha/\delta \ll N \ll \alpha/\sqrt{\mu_1 s}$, the two methods are valid for distinct but overlapping ranges of migration rates, together predicting dynamics across the full range of m.

Before describing how these methods are used, we make some useful observations about successful single-mutant lineages. By assumption, the deme size N is smaller than the number $\sim \alpha/\sqrt{\mu_1 s}$ reached by successful neutral lineages and the number $\sim \alpha/\delta$ reached by successful deleterious lineages in unstructured populations. Successful single-mutant lineages therefore drift to numbers $\gg N$ before producing a successful double mutant at high migration rates. They only drift to $\approx N$ copies at migration rates $m \ll m^*$; however, lineages that reach fewer than N copies always have a negligible probability of being successful. Consequently, so long as $m \ll \alpha/N$, we can assume that successful single-mutant lineages fix in their initial deme. In addition, we expect a wide range of migration rates $m^* \ll m \ll \alpha/N$ over which successful lineages fix in their initial deme, but drift to $\gg N$ copies, fixing in multiple demes, before producing a successful double mutant. (Note that this situation contrasts with that for locally deleterious single mutants, which drift to numbers $\sim \alpha/\delta \ll N$ at high migration rates.)

The $\hat{F}$ method reduces the effects of subdivision to a constant reduction in the rates of selection and drift experienced by single-mutant lineages. In particular, we model the single-mutant lineage in the subdivided population by one in an unstructured population of equal total size, but with effective parameters $\alpha_{\mathrm{T}} \equiv (1-\hat{F})\alpha$ and $\delta_{\mathrm{T}} \equiv (1-\hat{F})\delta$, where $\hat{F}$ is a constant that approximates the long-run average assortment of locally neutrally lineages. Meanwhile, since the establishment probability of the double mutant is approximately un-

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

affected by subdivision for $\delta \ll \alpha/N$, we keep $p_2 = s/\alpha$. In this way, predictions for p_1 and $\mathcal{T}_1$ follow directly from the theory for unstructured populations described in Section 4.2 and Supplementary Appendix C.

Over sufficiently long periods, the average assortment of locally neutral lineages tends towards a value that is independent of both the lineage frequency in the total population and the strength of selection against the lineage. This value is approximately equal to the equilibrium assortment at a polymorphic neutral locus with no mutation in the infinite island model, consisting of an infinite number of demes of N individuals. In the neutral infinite island model, frequency at the locus remains constant in the population as a whole; however, the frequency within each deme changes due to drift and migration, causing the assortment to deterministically reach an equilibrium value, $\hat{F}$. The equilibrium value equals the probability that a pair of alleles sampled from the same deme, looking backward in time, coalesce to a common ancestor before either migrates to a different deme. Intuitively, a pair that coalesce must be identical, while a pair that migrates resembles a sample from the total population at a short time in the past, but where the population has the same total frequency as the present; thus, $J_S = \hat{F} + (1 - \hat{F})J_T$. The equilibrium $\hat{F}$ is approximated by Wright's well-known formula (Wright 1951)

$$\hat{F} \approx \frac{1}{1 + Nm/\alpha}, \tag{18}$$

which follows immediately from the observation a pair of lineages coalesce within a deme at rate $2\alpha/N$ and migrate at rate $2m$. Thus, large assortments of $F \sim 1$ in the neutral, infinite island model correspond to migration rates $m \lesssim \alpha/N$.

Unlike in the infinite model, the assortment in the finite model is stochastic; however, if $L \gg 1$, $\delta \ll \alpha/N$, and the lineage is large enough, then the average assortment of the lineage approximates $\hat{F}$. Recall that F is bounded by maximum local frequency of the lineage; thus, mutant lineages with numbers $n < N\hat{F}$ necessarily have assortments smaller

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

than $\hat{F}$. However, the assortments of locally neutral lineages with numbers $n > N\hat{F}$ tend to fluctuate around a value $\approx \hat{F}$, such that averaged over a increasing number of generations the assortment becomes closer and closer to $\hat{F}$ (Figure 3 and Supplementary Appendix E). Weak local selection does not affect the assortment of these lineages with $n > N\hat{F}$ because selection takes $1/\delta$ generations to affect genealogies within demes, while assortment dynamics are governed by the faster processes of coalescence and migration within demes (Roze and Rousset 2003).

On the basis of similar observations, previous authors suggested approximating the dynamics of frequency in the total population by taking $F = \hat{F}$ (Cherry and Wakeley 2003; Roze and Rousset 2003; Wakeley 2003; Whitlock 2003; Wakeley and Takahashi 2004); however, conditions for when the stochastic fluctuations in F can be ignored for predicting lineage dynamics were not completely described. Several authors (Cherry and Wakeley 2003; Roze and Rousset 2003; Wakeley 2003) argued the constant F approximation is valid when L is sufficiently large that fluctuations in F are negligible. However, small fluctuations in F can require that the lineage is spread over an extremely large number of demes. Lineages that begin as a single copy necessarily occupy a relatively small number of demes and have significant variation in F even for large L (e.g., Figure 3 and Supplementary Figure S4). In Supplementary Appendix E, we find that a constant F approximation accurately predicts lineage dynamics if $n \gg N$, but may not for smaller n. The condition $n \gg N$ ensure that even large fluctuations in F typically average out over the time periods required for significant changes in n by selection or drift (Supplementary Appendix E). A single-mutant lineage that drifts to a maximum number $k \gg N$ typically spends a majority of its life-time (and accumulates an even larger majority of its weight) at numbers $\gg N$. Thus, if successful lineages typically drift to $\gg N$ copies, we can use the constant F approximation to model tunneling dynamics.

A second method for approximating dynamics at low migration rates is suggested by the fact that for $m \ll \alpha/N$, fixation or loss of a migrant within a demes occurs approximately

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

independently of further migration. If successful single-mutant lineages are likely to reach at least N copies and fix in their initial deme, then we can approximate lineage dynamics by treating fixation and loss within a deme as instantaneous and model the number of single-mutant demes by the DBD process described in Section 4.3. We assume that lineages that reach fewer than N copies make a negligible contribution to p_1, a valid assumption since $N \ll \min\{\alpha/\delta, \alpha/\sqrt{\mu_1 s}\}$. Predictions for valley-crossing dynamics under the DBD approximation are derived in Supplementary Appendix C. These predictions agree with those for the isolated-demes regime for $m < m^*$. They also agree with predictions from the $\hat{F}$ method at migration rates $m^* \ll m \ll \alpha/N$—at these migration rates, successful lineages always fix in their initial deme, but tend to reach numbers $\gg N$, fixing in multiple demes, before producing a successful double mutant. Consequently, we use the $\hat{F}$ method and isolated-demes approximation primarily for understanding the main quantitative and qualitative effects of subdivision, and use the DBD method results mainly for verification and for combining with the $\hat{F}$ method to make numerical predictions across full range of migration rates. Simulations show that numerical predictions for p_1 and $\langle \mathcal{T}_1 \rangle$ found by these methods are very accurate across all migration rates for $N\delta/\alpha$ as large as 0.2 (Figure 5 and Supplementary Figure S1).

Results: The relationships of p_1 and $\mathcal{T}_1$ with m depend on whether double mutants are locally beneficial or neutral ($s \gg \alpha/N$ or $s \ll \alpha/N$), and we first present results when they are locally beneficial. When $\delta \ll \alpha/N$ and $s \gg \alpha/N$, tunneling dynamics fall into one of five distinct parameter regimes. These regimes can be conveniently visualized in a phase diagram as a function of migration rate m and the single mutant selection coefficient δ (Figure 4). The five regimes occupy the region $\delta < \alpha/N$ in Figure 4; the two additional regimes with $\delta > \alpha/N$ are described in the next section. The probability that a single mutant is successful

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in these five regimes is approximately

$$p_1 \approx \begin{cases} \frac{\mu_1 s}{\alpha\delta} & \text{for } m \gg \frac{\alpha}{N} \text{ and } \delta \gg \sqrt{\mu_1 s} \\ \frac{\sqrt{\mu_1 s}}{\alpha} & \text{for } m \gg \frac{\alpha}{N} \text{ and } \delta \ll \sqrt{\mu_1 s} \\ \frac{\mu_1 s}{Nm\delta} & \text{for } \frac{\alpha\mu_1 s}{N\delta^2} \ll m \ll \frac{\alpha}{N} \\ \sqrt{\frac{\mu_1 s}{Nm\alpha}} & \text{for } \frac{N\mu_1 s}{\alpha} \ll m \ll \min\left\{\frac{\alpha}{N}, \frac{\alpha\mu_1 s}{N\delta^2}\right\} \\ \frac{1}{N} & \text{for } m \ll \frac{N\mu_1 s}{\alpha}, \end{cases} \tag{19}$$

and the average drift time is approximately

$$\langle \mathcal{T}_1 \rangle \approx \begin{cases} \frac{1}{\delta} & \text{for } m \gg \frac{\alpha}{N} \text{ and } \delta \gg \sqrt{\mu_1 s} \\ \frac{\ln(2)}{\sqrt{\mu_1 s}} & \text{for } m \gg \frac{\alpha}{N} \text{ and } \delta \ll \sqrt{\mu_1 s} \\ \frac{\alpha}{Nm\delta} & \text{for } \frac{\alpha\mu_1 s}{N\delta^2} \ll m \ll \frac{\alpha}{N} \\ \frac{\ln(2)\sqrt{\alpha}}{\sqrt{Nm\mu_1 s}} & \text{for } \frac{N\mu_1 s}{\alpha} \ll m \ll \min\left\{\frac{\alpha}{N}, \frac{\alpha\mu_1 s}{N\delta^2}\right\} \\ \frac{\alpha}{N\mu_1 s} & \text{for } m \ll \frac{N\mu_1 s}{\alpha}. \end{cases} \tag{20}$$

Complete analytical predictions are given in Supplementary Appendix C. Figure 5 shows our predictions for p_1 and $\langle \mathcal{T}_1 \rangle$ as a function of m, along with estimates from simulations, in a case where $\delta \gg \sqrt{\mu_1 s}$, so that single mutants are deleterious for tunneling at high migration rates, and Supplementary Figure S1 shows a case where $\delta \ll \sqrt{\mu_1 s}$. In these figures, the range of migration rates corresponding to each regime is color coded according to Figure 4.

These results show that p_1 and $\langle \mathcal{T}_1 \rangle$ show little variation with m at very high and very low migration rates, and increase significantly with decreasing m over a range of intermediate migration rates. For $m \gg \alpha/N$, migration is sufficiently frequent that the population is effectively unstructured for tunneling. The condition $m \gg \alpha/N$ corresponds to $\hat{F} \ll 1$; since

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

single-mutant lineages have small assortments, subdivision has almost no effect on tunneling dynamics. The two regimes with $m \gg \alpha/N$ in (19) and (20) simply correspond to whether single mutants are deleterious or neutral for tunneling in an unstructured population. Low migration rates $m \ll N\mu_1 s/\alpha$ correspond to the isolated-demes regime described in Section 4.3. For $\alpha/\delta \ll N \ll \alpha/s$, the various fixation probabilities described in Section 4.3 are approximately $\psi_{0\to 1} \approx \psi_{1\to 0} 1/N$ and $\theta \approx 1$, so that $m^* \approx N\mu_1 s/\alpha$ and $p_1 \approx 1/N$ and $\langle \mathcal{T}_1 \rangle \approx \alpha/N\mu_1 s$ for $m \ll m^*$.

Over the range of intermediate migration rates $\alpha/N\mu_1 s \ll m \ll \alpha/N$, both p_1 and $\langle \mathcal{T}_1 \rangle$ increase with a decreasing migration rate from their unstructured values to their isolated-demes values. The mathematical relationship of p_1 and $\langle \mathcal{T}_1 \rangle$ on m is easily understood using the $\hat{F}$ method, substituting the effective drift and selection coefficients α_{T} and δ_{T} into Equation (8) for p_1 and (9) for $\langle \mathcal{T}_1 \rangle$. Noting that $1 - \hat{F} \approx Nm/\alpha$ for $m \ll \alpha/N$, we see that α_{T} and δ_{T} are each reduced by a factor Nm/α as compared to their unstructured values. Accordingly, for single mutants that are deleterious for tunneling, p_1 and $\langle \mathcal{T}_1 \rangle$ are increased by a factor α/Nm relative to their unstructured values, while for single mutants that are neutral for tunneling, p_1 and $\langle \mathcal{T}_1 \rangle$ are increased by a factor of $\sqrt{\alpha/Nm}$. Whether single mutants are neutral or deleterious for tunneling is determined by whether δ_{T} is smaller or larger than $\sqrt{\alpha_{\mathrm{T}}\mu_1 p_2}$. Single mutants with $\delta \ll \sqrt{\mu_1 s}$ are neutral across all migration rates. Single mutants with $\delta \gg \sqrt{\mu_1 s}$ are deleterious for $m \gg \alpha\mu_1 s/N\delta^2$ and neutral for $m \ll \alpha\mu_1 s/N\delta^2$; intuitively, for migration rates below $\alpha\mu_1 s/N\delta^2$, subdivision has reduced the number a lineage must reach to have a weight $\sim 1/\mu_1 p_2$ to less than α/δ. Figure 5 shows p_1 and $\langle \mathcal{T}_1 \rangle$ in this second case where $\delta \gg \sqrt{\mu_1 s}$, so that, as m decreases, single mutants transition from being deleterious to neutral for tunneling.

When $s \ll \alpha/N$, we must modify the above results as follows. Equations (19) and (20) continue to predict p_1 and $\langle \mathcal{T}_1 \rangle$ for migration rates $m \gg \mu_1 \min\{\alpha/N\delta, \alpha/Ns\}$, but not for lower migration rates. For rates $m \ll \mu_1 \min\{\alpha/N\delta, \alpha/Ns\}$, we instead find that $p_1 \approx s/\alpha$ and $\langle \mathcal{T}_1 \rangle \approx 1/\mu_1$; thus, both p_1 and $\langle \mathcal{T}_1 \rangle$ are approximately independent of further

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

decreases in m. The isolated-demes regime corresponds to $m \ll \mu_1$. Therefore, when $s \ll \alpha/N$, p_1 and $\langle \mathcal{T}_1 \rangle$ reach their maximum values at migration rates outside of the isolated-demes regime and where single-mutant lineages still reach $\gg N$ copies before becoming successful. This observation contrasts with what we observed when $s \gg \alpha/N$, where p_1 and $\langle \mathcal{T}_1 \rangle$ increased with decreasing m until m^*. Intuitively, this difference is because for $s \ll \alpha/N$, both p_1 and $\langle \mathcal{T}_1 \rangle$ reach their theoretical upper bounds of p_2 and $1/\mu_1$, respectively, once $m \ll \mu_1 \min\{\alpha/N\delta, \alpha/Ns\}$, and so further decreases in α_{T} and δ_{T} have no effect.

As Figure 5 makes apparent, when single mutants are locally neutral, subdivision increases p_1 and $\langle \mathcal{T}_1 \rangle$ by similar proportions. Our analytical results show that if $\delta \ll \alpha/N$, then $p_1/\langle \mathcal{T}_1 \rangle \sim \mu_1 s/\alpha$ across all regimes, and the same relationship between p_1 and $\langle \mathcal{T}_1 \rangle$ applies in unstructured populations. Therefore, increasing the degree of subdivision by decreasing either N or m has a near-constant trade-off between decreasing $\langle \mathcal{T}_0 \rangle$ and increasing $\langle \mathcal{T}_1 \rangle$ once N is smaller than α/δ. We consider the consequences of this trade-off for the mean and distribution of the valley-crossing time in Sections 5 and 6.

4.5 Locally deleterious single mutants ($\delta \gg \alpha/N$)

When single mutants are locally deleterious, at migration rates $m \ll \alpha/N$, valley crossing may still be dominated by successful lineages that never reach high frequency within a deme. Our assumption that $\alpha/\delta \ll N \ll \alpha/\sqrt{\mu_1 s}$ implies $\delta \gg \sqrt{\mu_1 s}$, so that single mutants are deleterious for tunneling in unstructured populations. Therefore, successful lineages drift to $\sim \alpha/\delta$ copies at high migration rates, not nearly enough to fix within a deme in the absence of migration. And yet, our assumption that $N \ll N_\times$ ensures that successful lineages reach N copies and fix within a deme for sufficiently low migration rates. For lineages to fix within their deme (and for subdivision to increase p_1), the benefit of fixing in a deme—potentially gaining a much larger weight than lineages that reach only $\sim \alpha/\delta$ copies—must outweigh the cost of a low probability of drifting to N copies.

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Summary of methods: To solve valley crossing dynamics when single mutants are locally deleterious, we separately consider the contributions of lineages that drift to numbers $\lesssim \alpha/\delta$ and those that drift to at least N copies. Lineages that reach intermediate numbers k such that $\alpha/\delta \ll k \ll N$ do not contribute to valley crossing for the same reason as in large unstructured populations—compared to lineages reaching $\sim \alpha/\delta$ copies, their lower probability of occurrence greatly outweighs their slightly higher weights. Only lineages that reach N or more copies and fix within a deme have a sufficient weight advantage to potentially offset their low probability. For brevity, we refer to single-mutant lineages that reach a maximum number $\lesssim \alpha/\delta$ as *type A* and those that reach N or more copies as *type B*. Note that type-B lineages only have a possibility of influencing valley crossing at migration rates low enough that they will fix in their initial deme. We define $p_1^{(\mathrm{A})}$ as the contribution to p_1 from type-A lineages; i.e., the probability that a new lineage is type A, multiplied by the probability that a type-A lineage is successful, and similarly define $p_1^{(\mathrm{B})}$. The probability that a single-mutant is successful follows from the separate contributions as $p_1 \approx p_1^{(\mathrm{A})} + p_1^{(\mathrm{B})}$. In addition, we let $\langle \mathcal{T}_1^{(\mathrm{A})} \rangle$ be the average drift time of successful type As and $\langle \mathcal{T}_1^{(\mathrm{B})} \rangle$ be that of type Bs. The overall average drift time $\langle \mathcal{T}_1 \rangle$ is given by the average of $\langle \mathcal{T}_1^{(\mathrm{A})} \rangle$ and $\langle \mathcal{T}_1^{(\mathrm{B})} \rangle$ weighted by $p_1^{(\mathrm{A})}$ and $p_1^{(\mathrm{B})}$, respectively.

Type-A lineages have small assortments and so act similarly to deleterious lineages in unstructured populations. Successful type-A lineages have a maximum assortment of $\sim \alpha/N\delta$, set by the maximum frequency a lineage present in $\sim \alpha/\delta$ copies can obtain in any one deme. At migration rates $m \ll \delta$, we expect successful lineages to remain within their initial deme and thus have $F \sim \alpha/N\delta$. This small positive assortment leads to an increase in p_1 on the order of $\alpha/N\delta$ percent (Supplementary Appendix F analytically confirms this prediction). To a good approximation, we can ignore this small effect of structure in determining the fate of a type-A lineage. Since the vast majority of single-mutant lineages are type A, the approximate contribution to p_1 and $\langle \mathcal{T}_1 \rangle$ from type As are simply given by p_1 and $\langle \mathcal{T}_1 \rangle$ for deleterious tunneling in an unstructured population; i.e., $p_1^{(\mathrm{A})} \approx \mu_1 s/\alpha\delta$ and

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$\langle \mathcal{T}_1^{(\mathrm{A})} \rangle \approx 1/\delta$.

Since significant contributions from type-B lineages are limited to low migration rates ensuring fixation in the initial deme, we approximate their dynamics using the DBD model described in Section 4.3 and Supplementary Appendix C. The contribution to p_1 from type Bs equals the probability $\psi_{0\to 1}$ that a lineage fixes in its deme, times the probability that a single-mutant deme is successful via the DBD process, producing a successful double-mutant deme before going extinct. The dynamics of the DBD process simplify, however, for $\delta \gg \alpha/N$, since the deme birth rate is much smaller than the deme death rate D, allowing us to ignore the possibility that the lineage spreads beyond its initial deme before being driven extinct by an immigrating wild-type lineage (Supplementary Appendix C). The probability that a single-mutant deme is successful is thus approximately $U_1/(D+U_1)$, the probability that the deme mutates before going extinct, times the probability θ that the double-mutant deme is successful. The contribution to p_1 from type-B mutants is therefore

$$p_1^{(\mathrm{B})} \approx \frac{\psi_{0\to 1} U_1 \theta}{D+U_1}. \tag{21}$$

The drift time of successful type-B lineages is approximately exponentially distributed with average

$$\langle \mathcal{T}_1^{(\mathrm{B})} \rangle \approx \frac{1}{D+U_1}. \tag{22}$$

We recall that the migration rate m^* determining the boundary of the isolated-demes regime is given by migration rate where $D = U_1$. For migration rates $m \ll m^*$, where $D \ll U_1$, the expressions (21) and (22) for $p_1^{(\mathrm{B})}$ and $\mathcal{T}_1^{(\mathrm{B})}$, respectively, reduce to the values of p_1 and $\langle \mathcal{T}_1 \rangle$ in the isolated-demes regime (i.e., $p_1^{(\mathrm{B})} \approx \psi_{0\to 1}\theta$ and $\langle \mathcal{T}_1^{(\mathrm{B})} \rangle \approx 1/(N\mu_1\psi_{1\to 2})$. For migration rates $m \gg m^*$, for which $D \gg U_1$), these expressions can be approximately written as

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$p_1^{(\mathrm{B})} \approx \eta\mu_1 s/Nm\delta$ and $\langle \mathcal{T}_1^{(\mathrm{B})} \rangle \approx \alpha/Nm\delta$, where

$$\eta \equiv \frac{\alpha N \psi_{0\to 1}\psi_{1\to 2}\theta}{s} \tag{23}$$

is a constant that is $\ll 1$ for $\delta \gg \alpha/N$ and ≈ 1 for $\delta \ll \alpha/N$.

Results: Combining these predictions for the two types of lineages completely describes the fate of single-mutant lineages across the full range of migration rates. The relative size of s and α/N determines the maximum values of p_1 and $\langle \mathcal{T}_1 \rangle$ as m decreases by determining θ, but, unlike when $\delta \ll \alpha/\delta$, does not qualitatively change the relationships of p_1 and $\mathcal{T}_1$ with m. Figure 6 shows predictions for p_1 and $\langle \mathcal{T}_1 \rangle$ and the contribution from the two types in a case where $N\delta/\alpha = 8$ (large enough for a clear separation type-A and B lineages), and $Ns/\alpha = 10$ (large enough that $\theta \approx 1$). As expected, p_1 and $\langle \mathcal{T}_1 \rangle$ are driven by type-A lineages at high migration rates and by type-B lineages at low migration rates. The two types contribute equally to p_1 (i.e., $p_1^{(\mathrm{A})} = p_1^{(\mathrm{B})}$) at a migration rate $m \approx \eta\alpha/N$, so that for $p_1 \approx p_1^{(\mathrm{A})}$ at migration rates $m \gg \eta\alpha/N$ and $p_1 \approx p_1^{(\mathrm{B})}$ at rates $m \ll \eta\alpha/N$. The migration rate must therefore be significantly lower for subdivision to increase p_1 than when $\delta \ll \alpha/N$. Due to the much longer drift times of type Bs at rates $m \ll \alpha/N$, type Bs dominate $\langle \mathcal{T}_1 \rangle$ over a wider range, $m \approx \sqrt{\eta}\alpha/N$, than they do p_1. In other words, for rates $\eta\alpha/N \ll m \ll \sqrt{\eta}\alpha/N$, the average drift time approximates $\langle \mathcal{T}_1^{(\mathrm{B})} \rangle$ and is much larger than $\langle \mathcal{T}_1^{(\mathrm{A})} \rangle$, even though the majority of successful mutants are type A.

There are three distinct regimes for the probability that a single mutant is successful, mapped in the region $\delta > \alpha/N$ in Figure 4 and colored in Figure 6. In these regimes, the

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

probability that a single mutant is successful is approximately

$$p_1 \approx \begin{cases} \frac{\mu_1 s}{\alpha\delta} & \text{for } m \gg \frac{\eta\alpha}{N} \\ \frac{\eta\mu_1 s}{Nm\delta} & \text{for } \frac{\mu_1(\delta+s)}{\delta} \ll m \ll \frac{\eta\alpha}{N} \\ \frac{\delta\theta}{\alpha e^{N\delta/\alpha}} & \text{for } m \ll \frac{\mu_1(\delta+s)}{\delta}. \end{cases} \tag{24}$$

The regime $m \gg \eta\alpha/N$ corresponds to migration rates where type-A mutants make the dominant contribution to p_1 and so tunneling dynamics are similar to deleterious tunneling in an unstructured population. In the other two regimes, type Bs dominate p_1 such that $p_1 \approx p_1^{(\mathrm{B})}$ and is significantly increased over its unstructured value. The regime $m \ll \mu_1(\delta+s)/\delta$ corresponds to the isolated-demes regime; here, the fraction $\psi_{0\to 1} \approx \delta/(\alpha e^{N\delta/\alpha})$ of mutants that are type B are successful with probability θ. In the intermediate regime $\mu_1(\delta+s)/\delta \ll m \ll \eta\alpha/N$, type-B mutants dominate but only have a probability U_1/D of producing a double-mutant deme, causing $p_1 \propto 1/m$. The drift time has four regimes—the three regimes for p_1 plus the additional regime $\sqrt{\eta}\alpha/N$ described above—in which the average drift time is approximately

$$\langle \mathcal{T}_1 \rangle \approx \begin{cases} \frac{1}{\delta} & \text{for } m \gg \frac{\sqrt{\eta}\alpha}{N} \\ \left(\frac{\eta\alpha}{Nm}\right)\left(\frac{\alpha}{Nm\delta}\right) & \text{for } \frac{\eta\alpha}{N} \ll m \ll \frac{\sqrt{\eta}\alpha}{N} \\ \frac{\alpha}{Nm\delta} & \text{for } \frac{\mu_1(\delta+s)}{\delta} \ll m \ll \frac{\eta\alpha}{N} \\ \frac{\alpha}{N\mu_1(\delta+s)} & \text{for } m \ll \frac{\mu_1(\delta+s)}{\delta}. \end{cases} \tag{25}$$

Subdivision tends to increase $\langle \mathcal{T}_1 \rangle$ much more than p_1 at migration rates low enough for subdivision to have a significant effect. The larger increase can be traced to the fact that $p_1/\langle \mathcal{T}_1 \rangle$ equals $\mu_1 s/\alpha$ for type-A lineages but equals $\eta\mu_1 s/\alpha$ for type-B lineages. Therefore, an even steeper trade-off between $\langle \mathcal{T}_0 \rangle$ and $\langle \mathcal{T}_1 \rangle$ occurs for locally deleterious single mutants

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

than for locally neutral.

We conclude this section with an observation about the distribution of drift times when single mutants are locally deleterious. The distribution of $\mathcal{T}_1$ has two humps, centered at $\langle \mathcal{T}_1^{(\mathrm{A})} \rangle \approx 1/\delta$ and $\langle \mathcal{T}_1^{(\mathrm{B})} \rangle \approx \alpha/Nm\delta$, for $m \ll \alpha/N$. These humps can be seen most clearly when $m \approx \eta\alpha/N$, so that successful lineages are equally likely to be of either type (Figure 7). For $m \ll \eta\alpha/N$, the hump from successful type As forms a tiny fraction of the overall distribution. But, if valley crossing is limited by the drift time, the shorter drift times of type-A lineages may allow them to dominate valley crossing. We return to this point when considering the mean and distribution of $\mathcal{T}$ in the next two sections.

4.6 Sweep time

After a single-mutant lineage produces a successful double mutant, it takes an additional $\mathcal{T}_{\mathrm{sw}}$ generations for the double-mutant lineage to become fixed in the population. Supplementary Appendix G reviews sweep dynamics in unstructured populations and develops new approximations for sweeps in the island model. In unstructured populations, the average sweep time is approximately

$$\langle \mathcal{T}_{\mathrm{sw}} \rangle \approx \frac{2\ln(N_{\mathrm{T}} s/\alpha)}{s}. \tag{26}$$

The dominant factor in this expression for $\langle \mathcal{T}_{\mathrm{sw}} \rangle$ is $1/s$, the timescale over which selection significantly increases the number of double mutants. In island populations at migration rates $m \ll \alpha/N$, the average sweep time is approximately

$$\langle \mathcal{T}_{\mathrm{sw}} \rangle \approx \begin{cases} \frac{2\alpha \ln(N_{\mathrm{T}} s/\alpha)}{Nms} & \text{for } s \ll \alpha/N \\ \\ \frac{2\alpha \ln(L)}{Nms} & \text{for } s \gg \alpha/N. \end{cases} \tag{27}$$

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In (27), the dominant factor is α/Nms, which is sets the timescale for selection to significantly increase the number of double mutants when $m \ll \alpha/N$. See Supplementary Appendix G for more detailed expressions and interpretations further showing how subdivision influences the sweep. Equation (27) shows that heavy subdivision can lead to large sweep times even in modestly sized populations. Comparing (27) to (26) and ignoring logarithmic terms shows that subdivision increases $\langle \mathcal{T}_{\text{sw}} \rangle$ by a factor $\sim 1/(1-\hat{F})$, which is always similar to or larger than the factor of increase in p_1. In Section 5, we consider how the long sweep times in heavily subdivided populations can limit the ability for large amounts of subdivision to accelerate adaptation.

5 Average waiting time for the population to adapt

This section uses the results from Section 4 to describe the average waiting time $\mathcal{T}$ for a successful double mutant to arise and start to spread and the average waiting time for complete fixation of the double mutant. If these waiting times are typically dominated by $\mathcal{T}_0$, then increases in p_1 due to subdivision directly translate into decreases in the average time for the population to adapt. However, in very large or heavily subdivided populations, the drift and sweep times can form a substantial fraction of the adaptation time. Since subdivision increases the drift and sweep times, in such cases the decreases in the time to adapt will be substantially less than the increase in p_1. We first present our results for $\langle \mathcal{T} \rangle$ for subdivided populations with $\delta \ll \alpha/N$ (together with unstructured populations), then for subdivided populations with $\delta \gg \alpha/N$, and finally consider the overall time for the double mutant to fix, accounting for the sweep.

Unstructured populations and subdivided populations with $\delta \ll \alpha/N$: The average waiting time for a successful double mutant in unstructured populations and in subdivided populations for which single mutants are locally neutral are both approximately given

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

by (Supplementary Appendix C)

$$\langle \mathcal{T} \rangle \approx \begin{cases} \langle \mathcal{T}_0 \rangle & \text{for } \langle \mathcal{T}_0 \rangle \gg \langle \mathcal{T}_1 \rangle \\ \sqrt{\frac{\pi\alpha}{2N_{\mathrm{T}}\mu_0\mu_1 s}} & \text{for } \langle \mathcal{T}_0 \rangle \ll \langle \mathcal{T}_1 \rangle. \end{cases} \tag{28}$$

The effects of subdivision in (28) are encompassed solely by the determination of $\langle \mathcal{T}_0 \rangle$ and $\langle \mathcal{T}_1 \rangle$.

If $\langle \mathcal{T}_0 \rangle \gg \langle \mathcal{T}_1 \rangle$, then the first successful single-mutant lineage that occurs in the population will typically produce a successful double mutant before another successful single mutant can arise. In this case, we will typically have that $\mathcal{T} = \mathcal{T}_0 + \mathcal{T}_1$ and so $\langle \mathcal{T} \rangle \approx \langle \mathcal{T}_0 \rangle + \langle \mathcal{T}_1 \rangle \approx \langle \mathcal{T}_0 \rangle$. If $\langle \mathcal{T}_0 \rangle \ll \langle \mathcal{T}_1 \rangle$, then the approximation $\langle \mathcal{T}_0 \rangle$ significantly underestimates $\langle \mathcal{T} \rangle$ by ignoring the drift time. But, since multiple successful single-mutant lineages are likely to appear before the first successful double mutant, the approximation $\langle \mathcal{T}_0 \rangle + \langle \mathcal{T}_1 \rangle$ significantly overestimates $\langle \mathcal{T} \rangle$—if many successful single-mutant lineages segregate simultaneously, the lineage that is first to produce a successful double mutant must do so much more quickly than usual.

Weissman *et al.* (2009) called the regime in unstructured populations with $\langle \mathcal{T}_0 \rangle \ll \langle \mathcal{T}_1 \rangle$ the *neutral semi-deterministic tunneling (NSD)* regime. In the NSD regime, the total number of single mutants in the population grows as $\approx N_{\mathrm{T}}\mu_0 t$ until the time $\mathcal{T}$ when a successful double mutant is produced. Our results from Section 4.2 show that the condition $\langle \mathcal{T}_0 \rangle \ll \langle \mathcal{T}_1 \rangle$ corresponds to $N_{\mathrm{T}}\mu_0/\alpha \gg \max\{1, \delta^2/\mu_1 s\}$ in unstructured populations. The condition $N_{\mathrm{T}}\mu_0/\alpha \gg 1$ ensures that the number of single mutants only shows small fluctuations about its expectation, while $N_{\mathrm{T}}\mu_0/\alpha \gg \delta^2/\mu_1 s$ ensures that a successful double mutant is produced within $1/\delta$ generations, before negative selection has an effect. In the NSD regime, the role of drift only affects the probability $p_2 \approx s/\alpha$ that double mutants establish. From the fact that successful double mutants are produced at rate $\approx N_{\mathrm{T}}\mu_0\mu_1 st/\alpha$ it is easily shown that the average waiting time for a successful double mutant is $\langle \mathcal{T} \rangle_{\mathrm{nsd}} \equiv \sqrt{\pi\alpha/2N_{\mathrm{T}}\mu_0\mu_1 s}$.

In Supplementary Appendix C, we show if $\delta \ll \alpha/N$ and $\langle \mathcal{T}_0 \rangle \ll \langle \mathcal{T}_1 \rangle$, then NSD dynamics

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

also occur in subdivided populations and $\langle \mathcal{T} \rangle \approx \langle \mathcal{T} \rangle_{\text{nsd}}$. By decreasing $\langle \mathcal{T}_0 \rangle$ and increasing $\langle \mathcal{T}_1 \rangle$, subdivision greatly expands the situations where NSD dynamics occur. This finding makes intuitive sense given our discussion in Section 4.1 of how subdivision reduces the rates of drift and selection acting on the number of single mutants in the population.

Since the approximation for $\langle \mathcal{T} \rangle$ in (28) is roughly given by the maximum of $\langle \mathcal{T}_0 \rangle$ and $\langle \mathcal{T} \rangle_{\text{nsd}}$, the time $\langle \mathcal{T} \rangle_{\text{nsd}}$ sets a lower bound on $\langle \mathcal{T} \rangle$ regardless of the degree of subdivision in the population. One can imagine increasing the degree of subdivision by decreasing N and/or m while keeping the total population size fixed, causing $\langle \mathcal{T}_0 \rangle$ to decrease and $\langle \mathcal{T}_1 \rangle$ to increase. If the population is small enough that $N_{\text{T}}\mu_0/\alpha \ll \mu_1/s$, eventually p_1 will reach its theoretical maximum of $\approx s/\alpha$ and $\langle \mathcal{T} \rangle$ will reach $\approx \alpha/N_{\text{T}}\mu_0 s$ while the drift time is still negligible. However, if $N_{\text{T}}\mu_0/\alpha \gg \mu_1/s$, then $\langle \mathcal{T} \rangle$ can only decrease to a minimum of $\langle \mathcal{T} \rangle_{\text{nsd}}$. This minimum roughly occurs at the point where $\langle \mathcal{T}_0 \rangle \sim \langle \mathcal{T}_1 \rangle \sim \langle \mathcal{T} \rangle_{\text{nsd}}$, which can be seen as minimizing $\langle \mathcal{T}_0 \rangle + \langle \mathcal{T}_1 \rangle$ subject to the constraint $\langle \mathcal{T}_0 \rangle \langle \mathcal{T}_1 \rangle \sim \alpha/N_{\text{T}}\mu_0\mu_1 s$ that arises from the trade-off between p_1 and $\langle \mathcal{T}_1 \rangle$. Figure 8 shows $\langle \mathcal{T} \rangle$ in this scenario as the migration rate is decreased and N is held fixed. The average $\langle \mathcal{T} \rangle$ reaches $\approx \langle \mathcal{T} \rangle_{\text{nsd}}$ and stops decreasing at a migration rate that is still much larger than the migration rate m^* at which p_1 stops increasing.

We conclude our discussion of $\langle \mathcal{T} \rangle$ when $\delta \ll \alpha/N$ by observing that if $N_{\text{T}}\mu_0/\alpha \gg \max\{1, \delta^2/\mu_1 s\}$, then NSD dynamics already occur in the absence of population structure, and no degree of subdivision can effect a decrease in $\langle \mathcal{T} \rangle$.

Subdivided populations with $\delta \gg \alpha/N$: When single mutants are locally deleterious, we find $\langle \mathcal{T} \rangle$ by separately considering the average waiting times for a successful double mutant to first be produced by a type-A or a type-B lineage. If $\mathcal{T}^{(\text{A})}$ and $\mathcal{T}^{(\text{B})}$ are the waiting times for a successful double mutant to be produced by a type-A and by a type-B lineage, respectively, then $\mathcal{T} = \min\{\mathcal{T}^{(\text{A})}, \mathcal{T}^{(\text{B})}\}$. The averages and distributions of $\mathcal{T}^{(\text{A})}$ and $\mathcal{T}^{(\text{B})}$ are derived in Supplementary Appendix C. For most parameter choices, one of $\langle \mathcal{T}^{(\text{A})} \rangle$

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

or $\langle \mathcal{T}^{(\mathrm{B})} \rangle$ will be significantly smaller than the other, and the type with the smaller average almost always produces the first successful double mutant. In this case, the average waiting time $\langle \mathcal{T} \rangle$ is approximately given by the minimum of $\langle \mathcal{T}^{(\mathrm{A})} \rangle$ and $\langle \mathcal{T}^{(\mathrm{B})} \rangle$.

If the total waiting time is dominated by the waiting time for the first successful single mutant, then the condition for $\langle \mathcal{T}^{(\mathrm{B})} \rangle \ll \langle \mathcal{T}^{(\mathrm{A})} \rangle$ is simply that $m \ll \eta\alpha/N$. But, due to the much longer drift times of type Bs, this condition is in general not sufficient. We let $\mathcal{T}_0^{(\mathrm{A})}$ be the waiting time for the first successful type-A single mutant and $\mathcal{T}_0^{(\mathrm{B})}$ be the waiting time for the first successful type-B single mutant. Both $\mathcal{T}_0^{(\mathrm{A})}$ and $\mathcal{T}_0^{(\mathrm{B})}$ are exponentially distributed with rates $N_{\mathrm{T}}\mu_0 p_1^{(\mathrm{A})}$ and $N_{\mathrm{T}}\mu_0 p_1^{(\mathrm{B})}$, respectively, so that successful type Bs occur at a faster rate when $m < \eta\alpha/N$. Since type-A lineages act like deleterious lineages in an unstructured population, the average waiting time for a successful double mutant from a type-A lineage is roughly given by the maximum of $\langle \mathcal{T}_0^{(\mathrm{A})} \rangle$ and $\langle \mathcal{T} \rangle_{\mathrm{nsd}}$ (see Equation (28)). The average waiting time for a successful double mutant from a type-B lineage is approximately (Supplementary Appendix C)

$$\langle \mathcal{T}^{B} \rangle \approx \begin{cases} \langle \mathcal{T}_0^{(\mathrm{B})} \rangle & \text{for } \langle \mathcal{T}_0^{(\mathrm{B})} \rangle \gg \langle \mathcal{T}_1^{(\mathrm{B})} \rangle \\ \sqrt{\frac{\pi\alpha}{2N_{\mathrm{T}}\mu_0\mu_1 s\eta}}, & \text{for } \langle \mathcal{T}_0^{(\mathrm{B})} \rangle \ll \langle \mathcal{T}_1^{(\mathrm{B})} \rangle. \end{cases} \tag{29}$$

The average waiting time in (29) is approximately given by the maximum of $\langle \mathcal{T}_0^{(\mathrm{B})} \rangle$ and $\langle \mathcal{T} \rangle_{\mathrm{nsd}}/\sqrt{\eta}$.

Comparing these results for $\langle \mathcal{T}^{(\mathrm{A})} \rangle$ and $\langle \mathcal{T}^{(\mathrm{B})} \rangle$ shows that $\langle \mathcal{T}^{(\mathrm{B})} \rangle \ll \langle \mathcal{T}^{(\mathrm{A})} \rangle$ if and only if $N_{\mathrm{T}}\mu_0/\alpha \ll \eta\delta^2/\mu_1 s$ and $m \ll \eta\alpha/N$. Since $\langle \mathcal{T}^{(\mathrm{A})} \rangle$ approximately equals $\langle \mathcal{T} \rangle$ in an unstructured population, these conditions are also the conditions for subdivision to significantly decrease $\langle \mathcal{T} \rangle$ when $\delta \gg \alpha/N$. The requirement on $N_{\mathrm{T}}\mu_0/\alpha$ can be understood by noting that if $N_{\mathrm{T}}\mu_0/\alpha \gg \eta\alpha/N$, then $\langle \mathcal{T} \rangle_{\mathrm{nsd}}/\sqrt{\eta}$ is much larger than $\langle \mathcal{T}_0^{(\mathrm{A})} \rangle$. In this case, tunneling by type-B lineages remains slower than by type-As even at migration rates $m \ll \eta\alpha/N$ due to the much longer drift times of type-B lineages.

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Conditions for a significant reduction in $\langle \mathcal{T} \rangle$: These results show that for subdivision to significantly decrease the average waiting time for a successful double mutant, not only must demes and migration rates be small enough, but the scaled rate of new single mutants entering the population, $N_{\mathrm{T}}\mu_0/\alpha$, must not be too large. Explicit conditions for a significant decrease in $\langle \mathcal{T} \rangle$ across all regimes for the island model are that $N \ll N_{\times}$ and $m \ll \eta\alpha/N$, which ensure that $p_1 \gg p_{1,\mathrm{wm}}$, and that $N_{\mathrm{T}}\mu_0/\alpha \ll \min\{1, \eta\delta^2/\mu_1 s\}$, which ensures that longer drift times do not completely offset larger values of p_1 in the subdivided population.

Total time to fixation: For approximating the total time until complete fixation of the double mutant, there are two scenarios to consider. The first scenario is that all or a vast majority of the final population traces its ancestry back to the first successful double mutant; in this scenario, the sweep is said to be *hard* (Messer and Petrov 2013). If a hard sweep is likely, the average waiting time until complete fixation of the double mutant is approximately $\langle \mathcal{T} \rangle + \langle \mathcal{T}_{\mathrm{sw}} \rangle$. The second scenario is that multiple successful double-mutant lineages arise and make significant contributions to the final population; in this scenario, the sweep is said to be *soft* (Messer and Petrov 2013). If a soft sweep is likely, the approximation $\langle \mathcal{T} \rangle + \langle \mathcal{T}_{\mathrm{sw}} \rangle$ may substantially overestimate the total time to fixation since the contributions from later successful double mutants tend to shorten the time between the arrival of the first successful double mutant and fixation. A necessary (but insufficient) condition for soft sweeps is that $\langle \mathcal{T}_{\mathrm{sw}} \rangle \gg \langle \mathcal{T} \rangle$ (Supplementary Appendix G). Supplementary Appendix G provides more details about the fixation process under both the hard and soft scenarios.

By decreasing $\langle \mathcal{T} \rangle$ and increasing $\langle \mathcal{T}_{\mathrm{sw}} \rangle$, subdivision makes the sweep time more likely to limit the total time to fixation of the double mutant. The fastest rate of adaptation typically occurs at intermediate levels of subdivision where $\langle \mathcal{T} \rangle \sim \langle \mathcal{T}_{\mathrm{sw}} \rangle$. An example is shown in Figure 8, where both the average time to fixation is shown as a function of migration rate in addition to $\langle \mathcal{T} \rangle$. The time to fixation begins increasing with decreasing m below the point where $\langle \mathcal{T}_{\mathrm{sw}} \rangle \approx \langle \mathcal{T} \rangle$ (coincidentally, this happens to be near the migration rate where

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$\langle \mathcal{T}_0 \rangle \sim \langle \mathcal{T}_1 \rangle$). This figure also illustrates the overestimation of the time to fixation by the hard-sweep approximation at low migration rates.

6 Distribution of the waiting time

In biological applications such as the development of cancer or drug resistance, rare instances where a population adapts atypically quickly may be of utmost importance. For these applications, it is critical to consider not only the average but also the distribution of $\mathcal{T}$, or the probability $\Pr[\mathcal{T} < t]$ that a successful double mutant has occurred by a specific time t. Supplementary Appendix C finds the probability distribution of $\mathcal{T}$ for unstructured populations and subdivided populations with locally neutral or locally deleterious single mutants. This section summarizes these results and uses them to find the conditions in which subdivision significantly increases $\Pr[\mathcal{T} < t]$ as well as the maximum values of $\Pr[\mathcal{T} < t]$ that can be obtained by subdividing a population.

Unstructured populations and subdivided populations with $\delta \ll \alpha/N$: As with the mean, the distribution of $\mathcal{T}$ can be described similarly for unstructured populations and subdivided populations when single mutants are locally neutral. When $\langle \mathcal{T}_0 \rangle \gg \langle \mathcal{T}_1 \rangle$, so that $\langle \mathcal{T} \rangle \approx 1/N_{\mathrm{T}}\mu_0 p_1$, the probability that a successful double mutant has occurred by t is approximately

$$\Pr[\mathcal{T} < t] \approx \begin{cases} \frac{N_{\mathrm{T}}\mu_0\mu_1 s t^2}{2\alpha} & \text{for } t \ll \langle \mathcal{T}_1 \rangle \\ 1 - e^{-N_{\mathrm{T}}\mu_0 p_1 t} & \text{for } t \gg \langle \mathcal{T}_1 \rangle. \end{cases} \tag{30}$$

At times $t \gg \langle \mathcal{T}_1 \rangle$, the distribution of $\mathcal{T}$ is approximately exponential with mean $\langle \mathcal{T}_0 \rangle$, as at these times, $\mathcal{T}$ is dominated by the waiting time for the first successful single mutant (i.e., the drift time is negligible). However, the probability that a successful double mutant appears by a time $t \ll \langle \mathcal{T}_1 \rangle$ is much less than under the exponential approximation, since

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

tunneling at these times requires a single-mutant lineage that is extra lucky—not must the lineage produce a successful double mutant, it must do so much more quickly than a typical successful lineage.

The expression for the probability in (30) at times $t \ll \langle \mathcal{T}_1 \rangle$ can be understood using a simple argument. Let $n_1(t)$ be the total number of single mutants in the population at time t and $W_1(t) \equiv \int_0^t n_1(t)\, dt$ be the total weight from all single-mutant lineages up to t. Early on, the total weight from single mutants tends to be small. If t is small enough that we can safely assume that $W_1(t) \ll 1/\mu_1 p_2$, then $\Pr[\mathcal{T} < t]$ is approximately $\langle W_1(t) \rangle \mu_1 p_2$. Initially, the average number of single mutants grows simply as $\langle n_1(t) \rangle \approx N_{\mathrm{T}} \mu_0 t$, independently of population structure. Thus, early on, the average total weight grows as $N_{\mathrm{T}} \mu_0 t^2/2$ and the probability of at least one successful double mutant as $N_{\mathrm{T}} \mu_0 \mu_1 p_2 t^2/2$. This argument fails for $t \gtrsim \langle \mathcal{T}_1 \rangle$, at which point enough time has passed for selection and mutation to double mutants have reduced the growth of $\langle n_1(t) \rangle$ or for W_1 to become $\gtrsim 1/\mu_1 p_2$ (Supplementary Appendix C). Weissman *et al.* (2010) gives an alternate informal derivation for the approximation for $\Pr[\mathcal{T} < t]$ for $t \ll \langle \mathcal{T}_1 \rangle$ in unstructured populations based on the likely dynamics of lineages that become successful within t generations.

Equation (30) shows that even if subdivision greatly reduces $\langle \mathcal{T} \rangle$ relative to an unstructured population, it only significantly increases $\Pr[\mathcal{T} < t]$ for times much longer than the average drift time in the unstructured population (Figure 9). The range of times in which $\Pr[\mathcal{T} < t]$ is unaffected by subdivision is substantial if single mutants are close to neutral (as in the figure) and so have long drift times even when the population is effectively unstructured. This effect can be seen as another consequence of the trade-off between p_1 and the drift time. Subdivision increases p_1 by causing lineages to survive longer than they otherwise would, but also causes these lineages to grow more slowly in number and weight.

For unstructured populations and subdivided populations with $\delta \ll \alpha/N$ that are in the NSD regime ($\langle \mathcal{T}_0 \rangle \gg \langle \mathcal{T}_1 \rangle$ and $\langle \mathcal{T} \rangle \approx \langle \mathcal{T} \rangle_{\mathrm{nsd}}$), the probability of a successful double mutant

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

by t is

$$\Pr[\mathcal{T} < t] \approx 1 - e^{-N_{\mathrm{T}}\mu_0\mu_1 s t^2/2\alpha}. \tag{31}$$

Equation (31) forms an upper bound on the probability of adaptation by time t regardless of the level of subdivision, which is approximately met for $t \ll \langle \mathcal{T}_1 \rangle$. As the degree of subdivision is increased while N_{T} is held fixed, the average drift time increases and along with it the range of times for which $\Pr[\mathcal{T} < t]$ is given by (31), as seen in Figure 9.

Subdivided populations with $\delta \gg \alpha/N$: When single mutants are locally deleterious, we determine $\Pr[\mathcal{T} < t]$ by separately considering the probabilities that a successful double mutant has been produced by a type-A lineage or by a type-B lineage before t. In general, due to their longer drift times, type-B lineages are less likely to contribute to $\Pr[\mathcal{T} < t]$ at very early times than at times $\sim \langle \mathcal{T} \rangle$. If $\langle \mathcal{T}^{(\mathrm{B})} \rangle \gg \langle \mathcal{T}^{(\mathrm{A})} \rangle$, then type-A lineages form the dominate contribution to $\Pr[\mathcal{T} < t]$ for all t and subdivision has no effect on the distribution. Therefore, we assume that $N_{\mathrm{T}}\mu_0/\alpha \ll \eta\delta^2/\mu_1 s$ and $m \ll \alpha/N$, so that $\langle \mathcal{T}^{(\mathrm{B})} \rangle \ll \langle \mathcal{T}^{(\mathrm{A})} \rangle$ and adaptation typically occurs by type-B lineages. If $\langle \mathcal{T}_0^{(\mathrm{B})} \rangle \gg \langle \mathcal{T}_1^{(\mathrm{B})} \rangle$, then

$$\Pr[\mathcal{T} < t] \approx \begin{cases} \frac{N_{\mathrm{T}}\mu_0\mu_1 s t^2}{2\alpha} & \text{for } t \ll \frac{1}{\delta} \\ \frac{N_{\mathrm{T}}\mu_0\mu_1 s t}{\alpha\delta} & \text{for } \frac{1}{\delta} \ll t \ll \frac{2}{\eta\delta} \\ \frac{N_{\mathrm{T}}\mu_0\mu_1 s \eta t^2}{2\alpha} & \text{for } \frac{2}{\eta\delta} \ll t \ll \langle \mathcal{T}_1^{(\mathrm{B})} \rangle \\ 1 - e^{-N_{\mathrm{T}}\mu_0 p_1^{(\mathrm{B})} t} & \text{for } t \gg \langle \mathcal{T}_1^{(\mathrm{B})} \rangle, \end{cases} \tag{32}$$

where $p_1^{(\mathrm{B})}$ is given by (24) and $\langle \mathcal{T}_1^{(\mathrm{B})} \rangle$ by (25). For $t \ll 2/\eta\delta$, the probability of adaptation is determined solely from type-A lineages. The two regimes for $t \ll 2/\eta\delta$ correspond to whether t is shorter or longer than the average drift time $\approx 1/\delta$ of a type-A lineage. For $t \ll 2/\eta\delta$, the probability of adaptation is determined solely by type-B lineages. Again, the

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

two regimes here correspond to t shorter or longer than $\langle \mathcal{T}_1^{(\mathrm{B})} \rangle$. If instead $\langle \mathcal{T}_0^{(\mathrm{B})} \rangle \ll \langle \mathcal{T}_1^{(\mathrm{B})} \rangle$, then $\Pr[\mathcal{T} < t]$ is given by (32) for $t \gg 2/\eta\delta$ and by $\Pr[\mathcal{T} < t] \approx 1 - e^{-N_{\mathrm{T}}\mu_0\mu_1 s\eta t^2/2\alpha}$ for $t \gg 2/\eta\delta$. In both cases, type-B lineages only make the dominant contribution to $\Pr[\mathcal{T} < t]$ at times $t \gg 2/\eta\delta$.

These results show that if $\delta \gg \alpha/N$, the range of times over which subdivision has hardly any effect on $\Pr[\mathcal{T} < t]$ extends beyond the average drift time in an unstructured population, $1/\delta$, to the much longer time $2/\eta\delta$. For $\Pr[\mathcal{T} < t]$ to be significantly increased by subdivision when $\delta \gg \alpha/N$, it is necessary and sufficient both that $\langle \mathcal{T} \rangle$ is significantly decreased and that $t \gg 2/\eta\delta$.

7 Discussion

We have described the general ways in which population subdivision affects valley crossing in asexual species and given a complete analysis of valley-crossing dynamics for the island subdivision model. The competitive assortment caused by subdivision allows single-mutant lineages to have larger weights, with the result that single mutants have a greater probability p_1 of being successful. Decreasing deme size or migration rate increases the level of assortment and thus p_1, though significant increases in p_1 occur only if demes and migration rates are small enough that successful single-mutant lineages fix in one or more demes before producing a successful double mutant. Subdivision also increases the drift and sweep times, which can greatly limit the extent to which subdivision will decrease the average time for the population to adapt or increase the probability that the population has adapted by a certain time.

In addition to the study of Komarova (2006), a number of additional studies on valley crossing in spatially structured asexual populations were published during the preparation of this manuscript. Komarova (2006) analyzed the cases of one neutral or deleterious intermediate in a population structured as a one-dimensional integer lattice, where competition

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

occurs between individuals on neighboring sites on the lattice. Durrett and Moseley (2015) extended these results to include neutral intermediates on integer lattices of $d \geq 1$ dimensions. Komarova *et al.* (2014) used simulations to consider one or more neutral, deleterious, or beneficial intermediates in a two-dimensional lattice, allowing for competition within a fixed radius. These lattice-population studies sought the waiting time for the first beneficial mutant to occur. Bitbol and Schwab (2014) considered the average time to fixation of the double mutant from a wild-type population in subdivision model very similar to ours, but only analyzed the limits of extremely frequent and extremely infrequent migration, while using simulations to consider variation in the migration rate. Spatial structure affects valley crossing by the same fundamental mechanism in all these studies (except in the second model of Komarova *et al.* (2014), for which spatial structure leads to lower total population sizes). Accordingly, all find that structure can accelerate valley crossing by increasing p_1, although our results significantly extend these works; we compare our results with these works below. The seemingly contradictory finding of Takahasi (2007) that subdivision does not affect the crossing of a plateau in the absence of recombination is a consequence their unrealistic assumption that each mutation can occur only once in the population.

Ability for subdivision to increase p_1: In order for subdivision to significantly increase the probability that a single mutant is successful, we find the intuitive condition that successful single-mutant lineages must typically fix within one or more demes before producing a successful double mutant. This condition requires that both demes and migration rates are sufficiently small. Demes must contain fewer than $N_\times$ individuals to ensure that successful lineages do not simply tunnel within demes at low migration rates; this requirement was also found by Bitbol and Schwab (2014). In addition, we showed that for the island migration model, the migration rate m must be less than a threshold migration rate $\eta\alpha/N$ for p_1 to be significantly increased. If single mutants are locally neutral, the migration threshold simplifies to α/N, which corresponds to Wright's classic result that large values of F_{ST} at neutral

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

loci occur in the island model when $m \lesssim 1/N_{\mathrm{e}}$ (Wright 1931). If single mutants are locally deleterious, the migration threshold is significantly less than α/N, reflecting the role of selection in limiting preventing locally deleterious lineages from obtaining large assortments by fixing within a deme (Whitlock 2002).

If $N \ll N_{\times}$, then p_1 increases with decreasing migration rates over a wide range. If double mutants are sufficiently beneficial (such that $s \gg \alpha/N$), then p_1 increases until the maximal increase is approximately met when $m \ll m^*$, corresponding to the isolated-demes regime in which single-mutant lineages that fix within a deme always produce a double-mutant deme. Bitbol and Schwab (2014) analyzed the isolated-demes regime and also estimated the migration boundary m^*. However, for $\delta \ll \alpha/N$, the result in their Equation (10) (see also their Equation 17) overestimates m^* by a factor of $\ln(L)$ compared to the approximation $m^* \approx N\mu_1 s/\alpha$ following from our (17). They estimate m^* by comparing the time for a single-mutant lineage that has fixed in a deme to produce a successful double mutant, versus the average time for the lineage to go extinct. However, the latter time is not relevant for single mutants that are neutral for tunneling, which is always the case for m near m^* when $\delta \ll \alpha/N$. As a result, the approximation for m^* in Bitbol and Schwab (2014) overestimates the range of migration rates where p_1 takes its limiting value as $m \to 0$ when $\delta \ll \alpha/N$.

Consequences of the drift and sweep times: The effects that subdivision has on the drift and sweep times can greatly limit the ability for subdivision to accelerate valley crossing. Increases in the drift time caused by subdivision have two important consequences. First, subdivision can only decrease $\langle \mathcal{T} \rangle$ to a minimum of $\sqrt{\pi\alpha/2N_{\mathrm{T}}\mu_0\mu_1 s}$, equal to the average waiting time for a successful double mutant under neutral semi-deterministic tunneling dynamics. If $N_{\mathrm{T}}\mu_0/\alpha \gg \max\{1, \delta^2/\mu_1 s\}$, then this minimum is met even in the absence of subdivision and subdivision has no effect on either the mean or the distribution of $\mathcal{T}$. Komarova *et al.* (2014) observe a similar condition in their simulations for when structure does not affect $\mathcal{T}$ in two-dimensional lattice populations. Natural populations are frequently large

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

enough that we might expect that $N_{\mathrm{T}}\mu_0/\alpha \gg 1$. This can often be expected for bacterial and viral populations, and can also occur in cancer-susceptible stem-cell populations (see discussion below). In these cases, significant decreases in $\mathcal{T}$ due to subdivision will only occur if single mutants are sufficiently deleterious.

The second important consequence of the drift time that subdivision can only increase $\Pr[\mathcal{T} < t]$ to a maximum of $1 - e^{-N_{\mathrm{T}}\mu_0\mu_1 s t^2/2\alpha}$. As a result, subdivision has no affect on $\Pr[\mathcal{T} < t]$ at times earlier than the average drift time in an unstructured population, even if it greatly reduces $\langle\mathcal{T}\rangle$.

These results for the minimum of $\langle\mathcal{T}\rangle$ and maximum of $\Pr[\mathcal{T} < t]$ are both consequences of the fundamental trade-off between how spatial structure affects p_1 and $\langle\mathcal{T}_1\rangle$. As a result, we expect them to apply in other types of spatial structure such as lattice population models. Other studies of tunneling have either only considered $\langle\mathcal{T}\rangle$ or assumed the exponential approximation approximation for $\Pr[\mathcal{T} < t]$ and so have failed to observe that the effects of subdivision on the average are in general not representative of the effects on the probability of adapting very quickly. This has critical implications for how the results of these studies should be interpreted when considering the effects of subdivision on $\Pr[\mathcal{T} < t]$ in applications to cancer development, which we discuss in further detail below.

Extreme subdivision also greatly increases the sweep time, which may more than offset any decreases in the time until the occurrence of a successful double mutant. Bitbol and Schwab (2014) also emphasized this point and found a similar expression for the average sweep time at low migration rates when $s \gg \alpha/N$.

Other subdivision models: For many species, subdivision models where demes occupy geographic locations and individuals migrate preferentially to nearby demes are more realistic than the island model. Our results may also be used to predict valley-crossing dynamics in such cases, provided that demes have equal sizes and all individuals migrate at a fixed rate of m per generation. Our results for p_1 and $\mathcal{T}_1$ in the isolated-demes regime and for locally

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

deleterious single mutants are approximately independent of the migration pattern between demes. Therefore, significant differences in our predictions are limited to regimes where $\delta \ll \alpha/N$ and $m \gg m^*$, the regime in the island model described by the $\hat{F}$ method. In general, single-mutant lineages that drift to $\gg N$ copies will experience larger assortments in models where migration is limited to nearby demes than in the island model. Thus, locally neutral single mutants will tend to have larger values of p_1 outside of the isolated-demes regime, and may show significant increases in p_1 at migration rates higher than α/N. Further study is required to determine how variation in deme size (e.g., if only a fraction of demes have fewer than $N_\times$ individuals) or variation in migration rates in different demes affect p_1. We expect our results for the mean and distribution of $\mathcal{T}$ in terms of $\langle \mathcal{T}_0 \rangle$ and $\langle \mathcal{T}_1 \rangle$ to continue to hold in these other subdivision models.

Beneficial single mutants: When two mutations lead to a strongly beneficial complex adaptation, the first mutation may often have a smaller but still significant fitness advantage. We consider the effects of subdivision on the waiting time for such an adaptation in Supplementary Appendix H. Suppose that single mutants have a fitness $1 + s_1$ and double mutants have fitness $1 + s$ with $0 < s_1 < s$. If $s_1 \ll \min\{\alpha/N_\mathrm{T}, \sqrt{\mu_1 s}\}$, then single mutants are effectively neutral for the purposes of adaptation in both unstructured (Weissman *et al.* 2009) and subdivided populations, and subdivision can accelerate adaptation by increasing p_1. On the other hand, if $s_1 \gg \max\{\alpha/N_\mathrm{T}, \sqrt{\mu_1 s}\}$, then in an unstructured population a successful single-mutant lineage will begin to sweep in the population before producing a successful double mutant. In such cases, subdivision can significantly increase the waiting time for a successful double mutant by slowing the sweep of the single mutant. Komarova *et al.* (2014) also observed this effect in simulations of two-dimensional lattice populations.

As with neutral and deleterious single mutants, extreme subdivision moves the mean and distribution of $\mathcal{T}$ for beneficial single mutants towards the values predicted under neutral semi-deterministic tunneling. We consider the potential implications of this finding for cancer

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

incidence rates below.

Effects of subdivision in nature: To determine whether subdivision affects tunneling dynamics in nature, we must first ask whether natural populations are sufficiently subdivided that successful single-mutant lineages are likely to fix within a deme. Empirical measures of F_{ST} and related statistics using polymorphism data can provide some insight into the levels of competitive assortment in natural populations. Values of F_{ST} near 1 have been found in some species, such as the nematode *C. elegans* and in plant species with limited pollen dispersal (Charlesworth and Charlesworth 2010). Frost *et al.* (2001) found values of F_{ST} ranging between 0.08 and 0.59 for subpopulations of HIV in pulps within the spleens of infected patients. However, reported values of F_{ST} provide only limited insight into levels of competitive assortment, since they typically consider assortment over spatial scales much larger than the spatial scale at which individuals compete with one another. For example, studies commonly report estimates of F_{ST} between large geographic regions that most likely each consist of many smaller demes. Estimates of F_{ST} between regions likely underestimate competitive assortment, since rare mutant lineages are more likely to fix within a small deme than within a large geographic region.

Frequent local extinctions of demes followed by recolonization from a small number of founder individuals may be an important driver of large assortments in natural populations. For example, Frost *et al.* (2001) inferred that such dynamics drive the high values of F_{ST} in HIV populations within the spleen. Strong genetic hitchhiking also occurs in HIV (Zanini and Neher 2013; Zanini *et al.* 2015) and may act similarly by allowing a single-mutant lineage to quickly hitchhike to fixation on a linked beneficial mutation. Founder effects and hitchhiking may be particularly important in allowing locally deleterious lineages to reach fixation within a deme, but introduce additional complications and so further study of how these dynamics interact with subdivision to affect tunneling is needed.

Finally, we note that extreme subdivision occurs in two examples of tunneling in nature.

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The mitochondrial population of a mammalian species is divided into subpopulations corresponding to separate germlines, each with a small effective population size of $N_e \lesssim 200$ (Howell *et al.* 1992; Jenuth *et al.* 1996) and no migration between them. These mitochondrial populations regularly cross valleys involving changes in the two complementary nucleotides that bind together to form the stems of tRNAs (Meer *et al.* 2010). In such cases, the first mutation alone is deleterious while the two mutations together are neutral. Meer *et al.* (2010) present genomic evidence that the deleterious single mutants remain at low frequency in the species' population of mitochondria. Given the low germline effective population size and typical mitochondrial mutation rates, it is likely that the single-mutant lineage fixes within the germline before gaining the second mutation. Subdivision may increase the rate of tunneling by minimizing selection against single-mutant lineages within germlines. However, fixation within the germline is also thought to increase selection acting on mitochondrial mutations at the level of competition between hosts (Bergstrom and Pritchard 1998; Neiman and Taylor 2009), and so it is possible that subdivision may in fact reduces the rate of tunneling. Human tissues in which tunneling can give rise to cancer are also often extremely subdivided, and we now consider the implications of our results to this process.

Applications to the development of cancer: Our results provide insight into how the structure of human tissue might influence the development of cancer, which we illustrate by considering the waiting time for a stem cell to gain a loss of function (LOF) mutation in both copies of a tumor suppressor gene (TSG). Details for our parameter estimates and calculations are given in Supplementary Appendix I. First hits to TSGs in healthy stem cells have been estimated to occur at rates of $\sim 10^{-6}$ to $\sim 10^{-7}$ per generation, while second hits are thought occur somewhat faster as loss-of-heterozygosity mutations may occur that replace the functional with the non-functional copy. We suppose that the second hit has a rate of $\mu_1 = 10^{-6}$ per generation. Cellular reproduction within somatic tissues is similar to that in the Moran model, so that $\alpha \sim 1$. If the first hit is neutral, then the average

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

drift time if the population were unstructured is $\log(2)/\sqrt{\mu_1} \approx 700$ generations if double mutants always lead to cancer and longer if only a fraction do. The number of generations experienced by stem cells in cancer-susceptible tissues over an 80 year human lifespan ranges from a high of ~ 6000 in the colon to $\lesssim 80$ for many tissues (Supplementary Appendix I). Thus, if single mutants are typically neutral then the average drift time without structure may be comparable to or longer than a typical human lifespan for many cancers, greatly limiting the ability for structure to increase the probability of acquiring two hits. Komarova (2006), Komarova *et al.* (2014), and Durrett and Moseley (2015) claim that spatial structure should increase the probability of cancer when the first hit is neutral, but ignore the drift time in their calculations and only account for the time until a successful single mutant appears.

In the colon, where a relatively large number of stem-cell generations occur, the susceptible stem-cell population is very large, at $N_{\mathrm{T}} \sim 8\times10^7$ (Supplementary Appendix I). Colon cancer often begins by two hits to the APC gene, for which we estimate that $N_{\mathrm{T}}\mu_0/\alpha \sim 50$. Thus, if the first hit is neutral, we predict neutral semi-deterministic tunneling regardless of the level of subdivision. However, Vermeulen *et al.* (2013) recently showed that some single LOF mutants carry a large fitness advantage over wild-types within a crypt. This advantage is large enough that single mutants would likely sweep before producing a successful double mutant if the population were unstructured, but instead are confined to a single crypt due to the extreme subdivision present in the colon (Supplementary Appendix I). In reality, different LOF mutations are likely to have different fitness effects and cannot necessarily be lumped together into one mutation rate. However, if a large fraction of LOF mutations have a similar selective advantage, then subdivision in the colon likely greatly limits the appearance of stem cells with two LOF mutations.

Conclusion: Population subdivision likely has important but complex effects on valley crossing in nature. Our results can be used to predict the effects of a broad range of different

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

types of spatial structure on valley-crossing dynamics when recombination between mutations can be ignored. Our assumptions about population dynamics will frequently be invalid in nature due to processes we have ignored, such as extinction-recolonization dynamics, genetic hitchhiking, and selection at levels of population structure higher than demes. These require careful further consideration before we fully understand the effects of subdivision on tunneling in nature. However, we hope that our results may provide a foundation for studying these processes and for incorporating the effects of subdivision and spatial structure more generally in future theoretical and empirical investigations of valley crossing.

Acknowledgements

I am grateful to I. Arbisser, B. J. Callahan, N. Creanza, M. D. Edge, M. W. Feldman, and D. B. Weissman for invaluable discussions and comments on the manuscript. I was supported during this research by the Stanford Department of Biology, the Stanford Genome Training Program (SGTP; NIH/NHGRI) and the Stanford Center for Computational, Evolutionary, and Human Genomics' Fellows program. Computing resources were provided by NIH grant GM28016.

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

Baker, A. M., B. Cereser, S. Melton, A. G. Fletcher, M. Rodriguez-Justo, *et al.*, 2014 Quantification of crypt and stem cell evolution in the normal and neoplastic human colon. Cell Reports 8.4: 940–947.

Bergstrom, C. T., and J. Pritchard, 1998 Germline Bottlenecks and the Evolutionary Maintenance of Mitochondrial Genomes. Genetics 149.4: 2135–2146.

Bitbol, A.-F., and D. J. Schwab, 2014 Quantifying the role of population subdivision in evolution on rugged fitness landscapes. PLoS Computational Biology 10.8: e1003778.

Blount, Z. D., J. E. Barrick, C. J. Davidson, and R. E. Lenski, 2012 Genomic analysis of a key innovation in an experimental Escherichia coli population. Nature 489.7417: 513–518.

Charlesworth, B., and D. Charlesworth, 2010 *Elements of Evolutionary Genetics*. Greenwood Village: Roberts and Company Publishers.

Cherry, J. L., and J. Wakeley, 2003 A diffusion approximation for selection and drift in a subdivided population. Genetics 163.1: 421–8.

Desai, M. M., and D. S. Fisher, 2007 Beneficial mutation selection balance and the effect of linkage on positive selection. Genetics 176.3: 1759–1798.

Durrett, R., and S. Moseley, 2015 Spatial Moran Models I. Stochastic Tunneling in the Neutral Case. The Annals of Applied Probability 25.1: 104–115.

Durrett, R., and D. Schmidt, 2008 Waiting for two mutations: With applications to regulatory sequence evolution and the limits of Darwinian evolution. Genetics 180.3: 1501–1509.

Ewens, W. J., 2004 *Mathematical Population Genetics 1: Theoretical Introduction*. New York: Springer-Verlag.

Fisher, D. S., 2007 Course 11 Evolutionary dynamics. *Complex Systems: Lecture Notes of the Les Houches Summer School 2006*. Ed. by J.-P. Bouchaud, M. Mézard, and J. Dalibard. Vol. 85. Elsevier: pp. 395–446.

Frank, S. A., 2007 *Dynamics of cancer: Incidence, inheritance, and evolution*. Princeton and Oxford: Princeton University Press.

Frost, S. D., M. J. Dumaurier, S. Wain-Hobson, and A. J. Brown, 2001 Genetic drift and within-host metapopulation dynamics of HIV-1 infection. Proceedings of the National Academy of Sciences of the United States of America 98.12: 6975–6980.

Gillespie, J. H., 1984 Molecular evolution over the mutational landscape. Evolution 38.5: 1116–1129.

Gong, L. I., M. A. Suchard, and J. D. Bloom, 2013 Stability-mediated epistasis constrains the evolution of an influenza protein. eLife 2: e00631.

Howell, N., S. Halvorson, I. Kubacka, D. a. McCullough, L. a. Bindoff, *et al.*, 1992 Mitochondrial gene segregation in mammals: is the bottleneck always narrow? Human genetics 90.1-2: 117–120.

Iwasa, Y., F. Michor, and M. A. Nowak, 2004 Stochastic tunnels in evolutionary dynamics. Genetics 166.3: 1571–1579.

Jenuth, J. P., A. C. Peterson, K. Fu, and E. A. Shoubridge, 1996 Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA. Nature Genetics 14.3: 353–356.

Kimura, M., 1985 The role of compensatory neutral mutations in molecular evolution. Journal of Genetics 64.1: 7–19.

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Knudson, A. G., 2001 Two genetic hits (more or less) to cancer. Nature reviews Cancer 1.2: 157–162.

Komarova, N. L., 2006 Spatial stochastic models for cancer initiation and progression. Bulletin of mathematical biology 68.7: 1573–1599.

Komarova, N. L., A. Sengupta, and M. A. Nowak, 2003 Mutation-selection networks of cancer initiation: tumor suppressor genes and chromosomal instability. Journal of Theoretical Biology 223.4: 433–450.

Komarova, N. L., L. Shahriyari, and D. Wodarz, 2014 Complex role of space in the crossing of fitness valleys by asexual populations. Journal of The Royal Society Interface 11: 20140014.

Levin, B. R., V. Perrot, and N. Walker, 2000 Compensatory Mutations, Antibiotic Resistance and the Population Genetics of Adaptive Evolution in Bacteria. Genetics 154.3: 985–997.

Maruyama, T., 1970 On the fixation probability of mutant genes in a subdivided population. Genetical Research, Cambridge 15: 221–225.

Maruyama, T., 1974 A Markov process of gene frequency change in a geographically structured population. Genetics 76.2: 367–377.

McKeown, A. N., J. T. Bridgham, D. W. Anderson, M. N. Murphy, E. A. Ortlund, *et al.*, 2014 Evolution of DNA specificity in a transcription factor family produced a new gene regulatory module. Cell 159.1: 58–68.

Meer, M. V., A. S. Kondrashov, Y. Artzy-Randrup, and F. A. Kondrashov, 2010 Compensatory evolution in mitochondrial tRNAs navigates valleys of low fitness. Nature 464.7286: 279–282.

Messer, P. W., and D. A. Petrov, 2013 Population genomics of rapid adaptation by soft selective sweeps. Trends in Ecology and Evolution 28.11: 659–669.

Nagylaki, T., 1980 The strong-migration limit in geographically structured populations. Journal of Mathematical Biology 9.2: 101–114.

Nei, M., 1973 Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences of the United States of America 70.12: 3321–3323.

Neiman, M., and D. R. Taylor, 2009 The causes of mutation accumulation in mitochondrial genomes. Proceedings of the Royal Society B: Biological Sciences 276.1660: 1201–1209.

Ortlund, E. A., J. T. Bridgham, M. R. Redinbo, and J. W. Thornton, 2007 Crystal structure of an ancient protein: Evolution by conformational epistasis. Science 317.5844: 1544–1548.

Proulx, S. R., 2011 The rate of multi-step evolution in Moran and Wright-Fisher populations. Theoretical Population Biology 80.3: 197–207.

Regoes, R. R., and S. Bonhoeffer, 2005 The HIV coreceptor switch: A population dynamical perspective. Trends in Microbiology 13.6: 269–277.

Rousset, F., 2002 Inbreeding and relatedness coefficients: what do they measure? Heredity 88.5: 371–380.

Roze, D., and F. Rousset, 2003 Selection and drift in subdivided populations: A straightforward method for deriving diffusion approximations and applications involving dominance, selfing and local extinctions. Genetics 165.4: 2153–2166.

Silva, J. da, and S. K. Wyatt, 2014 Fitness valleys constrain HIV-1's adaptation to its secondary chemokine coreceptor. Journal of Evolutionary Biology 27: 604–615.

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Silva, J. da, M. Coetzer, R. Nedellec, C. Pastore, and D. E. Mosier, 2010 Fitness epistasis and constraints on adaptation in a human immunodeficiency virus type 1 protein region. Genetics 185.1: 293–303.

Takahasi, K. R., 2007 Evolution of coadaptation in a subdivided population. Genetics 176.1: 501–511.

Tomasetti, C., and B. Vogelstein, 2015 Variation in cancer risk among tissues can be explained by the number of stem cell divisions. Science 347.6217: 78–81.

Trindade, S., A. Sousa, K. B. Xavier, F. Dionisio, M. G. Ferreira, *et al.*, 2009 Positive epistasis drives the acquisition of multidrug resistance. PLoS Genetics 5.7: e1000578.

Vermeulen, L., E. Morrissey, M. van der Heijden, A. M. Nicholson, A. Sottoriva, *et al.*, 2013 Defining stem cell dynamics in models of intestinal tumor initiation. Science 342.6161: 995–998.

Vermeulen, L., and H. J. Snippert, 2014 Stem cell dynamics in homeostasis and cancer of the intestine. Nature reviews Cancer 14.7: 468–480.

Wakeley, J., 2003 Polymorphism and divergence for island-model species. Genetics 163.1: 411–420.

Wakeley, J., and T. Takahashi, 2004 The many-demes limit for selection and drift in a subdivided population. Theoretical Population Biology 66.2: 83–91.

Wang, X., G. Minasov, and B. K. Shoichet, 2002 Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. Journal of Molecular Biology 320.1: 85–95.

Weinreich, D. M., and L. Chao, 2005 Rapid evolutionary escape by large populations from local fitness peaks is likely in nature. Evolution 59.6: 1175–82.

Weinreich, D. M., N. F. Delaney, M. A. Depristo, and D. L. Hartl, 2006 Darwinian evolution can follow only very few mutational paths to fitter proteins. Science 312.5770: 111–114.

Weissman, D. B., M. M. Desai, D. S. Fisher, and M. W. Feldman, 2009 The rate at which asexual populations cross fitness valleys. Theoretical Population Biology 75.4: 286–300.

Weissman, D. B., M. W. Feldman, and D. S. Fisher, 2010 The rate of fitness-valley crossing in sexual populations. Genetics 186.4: 1389–1410.

Whitlock, M. C., 2002 Selection, load and inbreeding depression in a large metapopulation. Genetics 160.3: 1191–1202.

Whitlock, M. C., 2003 Fixation probability and time in subdivided populations. Genetics 164.2: 767–779.

Whitlock, M. C., 2004 Selection and drift in metapopulations. *Ecology, Genetics and Evolution of Metapopulations*. Ed. by I. Hanski, and O. E. Gaggiotti. Elsevier: pp. 153–173.

Wright, S., 1970 Random drift and the shifting balance theory of evolution. *Mathematical Topics in Population Genetics*. Ed. by K.-i. Kojima. Berlin, Heidelberg: Springer-Verlag: pp. 1–31.

Wright, S., 1931 Evolution in Mendelian populations. Genetics 16.2: 97–159.

Wright, S., 1932 The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proceedings of the Sixth International Congress on Genetics*. Vol. 1: pp. 356–366.

Wright, S., 1951 The genetical structure of populations. Annals of Eugenics 15.4: 323–354.

Zanini, F., J. Brodin, L. Thebo, C. Lanz, G. Bratt, *et al.*, 2015 Population genomics of intrapatient HIV-1 evolution. eLife 4: e11282.

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Zanini, F., and R. A. Neher, 2013 Quantifying selection against synonymous mutations in HIV-1 env evolution. Journal of Virology 87.21: 11843–11850.

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Tables

Table 1: Model parameters.

L	Number of demes
N_{T}	Census population size of the total population
N	Census population size of a deme in the island model
m	Migration rate in the island model
α	Drift parameter
μ_0	Mutation rate of wild type to single mutant
μ_1	Mutation rate of single to double mutant
δ	Selective disadvantage of single mutant
s	Selective advantage of double mutant

Table 2: Notation.

p_1	Probability that a single mutant is successful
p_2	Probability that a double mutant is successful
$\mathcal{T}_0$	Waiting time for the first successful single mutant
$\mathcal{T}_1$	Drift time of the first successful single mutant
$\mathcal{T}$	Waiting time for the first successful double mutant
$\mathcal{T}_{\mathrm{sw}}$	Time for the first successful double mutant to sweep to fixation
$p_{1,\mathrm{wm}}$	Value of p_1 in an unstructured population of N_{T} individuals
$\langle \mathcal{T}_{1,\mathrm{wm}} \rangle$	Value of $\langle \mathcal{T}_1 \rangle$ in an unstructured population of N_{T} individuals
F	Measure of genetic assortment
$P_f(i,k)$	Probability that a mutant lineage with fitness $1+f$ present in i copies reaches k copies before extinction
$\psi_{i\to j}$	Probability that a lineage with j mutations fixes in a deme of individuals with i mutations
θ	Probability that a double mutant is successful after fixing in one deme
$N_\times$	Threshold deme size
m^*	Migration boundary for the isolated-demes regime
η	Compound parameter defined in Equation (23)
$\langle \mathcal{T} \rangle_{\mathrm{nsd}}$	$\langle \mathcal{T} \rangle$ under neutral semi-deterministic tunneling

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figures

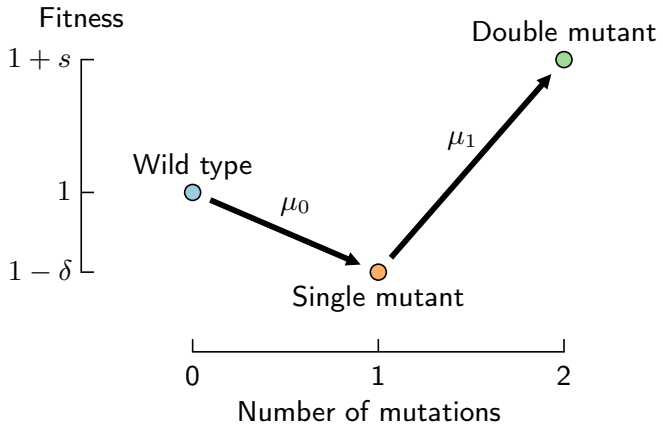

Figure 1: Genotype-to-fitness mapping for a two-mutation fitness valley ($\delta > 0$) or plateau ($\delta = 0$).

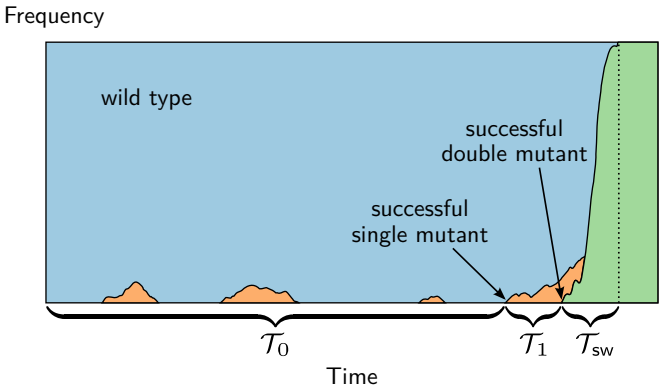

Figure 2: When successful single mutants are infrequent, the time for the population to tunnel across the valley is typically dominated by the time $\mathcal{T}_0$ until the first successful single mutant is produced.

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

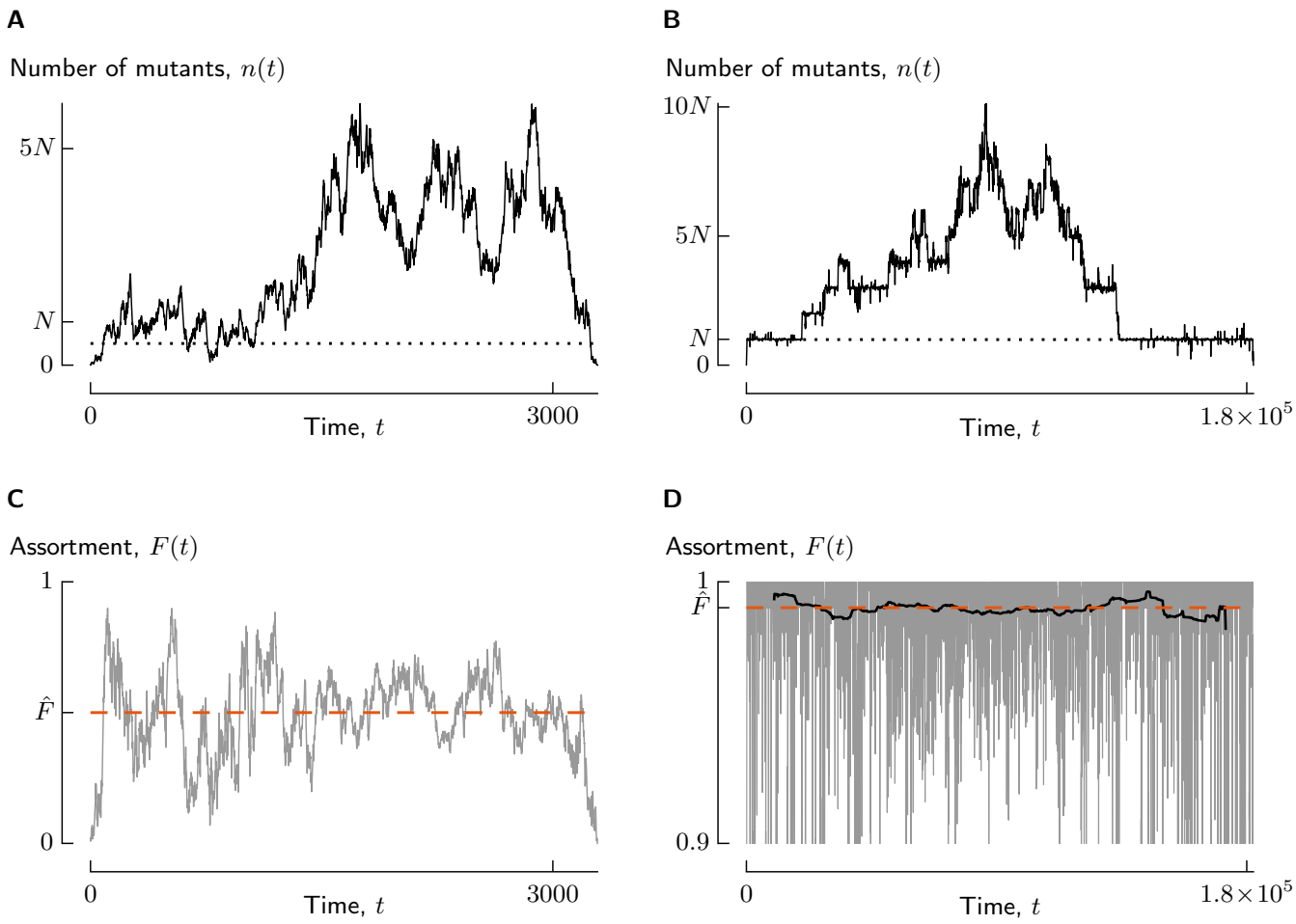

Figure 3: The assortment of a locally neutral lineage fluctuates around $\approx \hat{F}$ while the lineage is present in more than $N\hat{F}$ copies. Shown are the trajectories of n and F for a locally neutral lineage ($N\delta/\alpha = 0.2$) at each of two migration rates, corresponding to $Nm/\alpha = 1$ (Panels A and C) and $Nm/\alpha = 10^{-2}$ (Panels B and D). Each lineage began as a single copy and reached at least $\alpha/\delta = 5N$ copies before going extinct. In Panels A and B, the solid line shows $n(t)$ while the dotted line marks the line $n = N\hat{F}$. In Panels C and D, the grey line shows F and the dashed orange line indicates $\hat{F}$, while the black line in Panel D shows the centered moving average of F over a window of $1/m$ generations. Note that assortments below 0.9 are not shown in Panel D. Parameters are $L = 100$, $N = 100$, $\alpha = 0.5$, and $\delta = 10^{-3}$, with $m = 5\times10^{-3}$ for A and C and $m = 5\times10^{-5}$ for B and D.

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

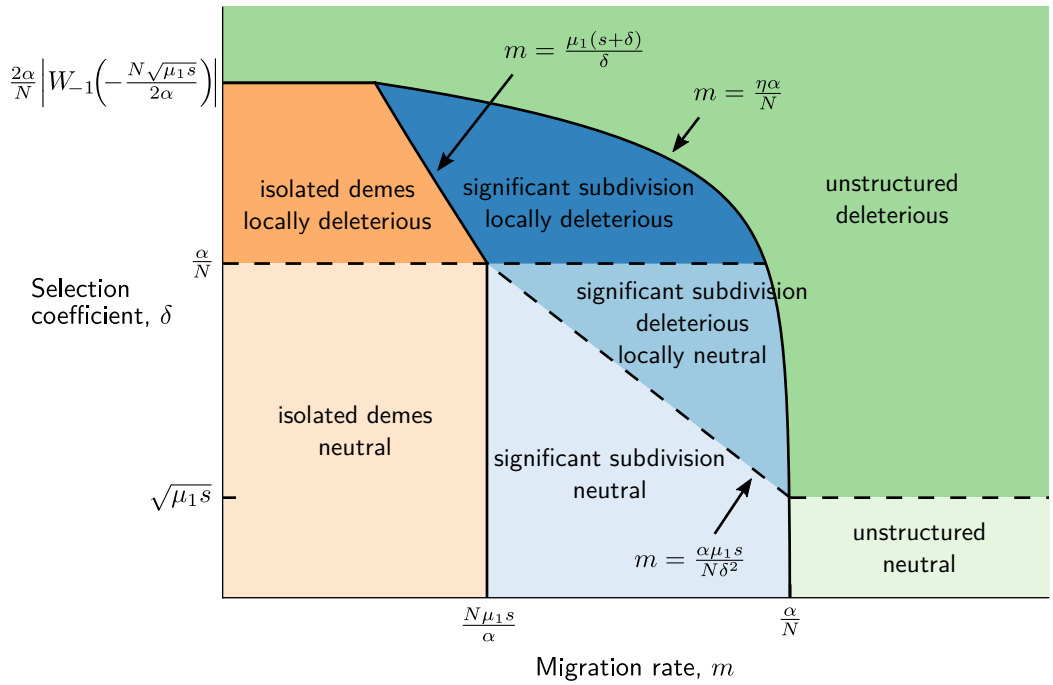

Figure 4: The dynamics of successful single mutants separate into seven regimes depending on the migration rate, m, and the selection coefficient of the single mutant, δ, when demes are intermediately sized with $\alpha/s \ll N \ll \alpha/\sqrt{\mu_1 s}$, so that effects of subdivision are possible and double mutants always spread after fixing within a deme. Both axes are on a logarithmic scale, and W_{-1} is the negative real branch of the product-log function. To make the figure, we fixed $N = 10^3$, $\alpha = 0.5$, $\mu_1 = 10^{-8}$, and $s = 0.05$ and varied m and δ.

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

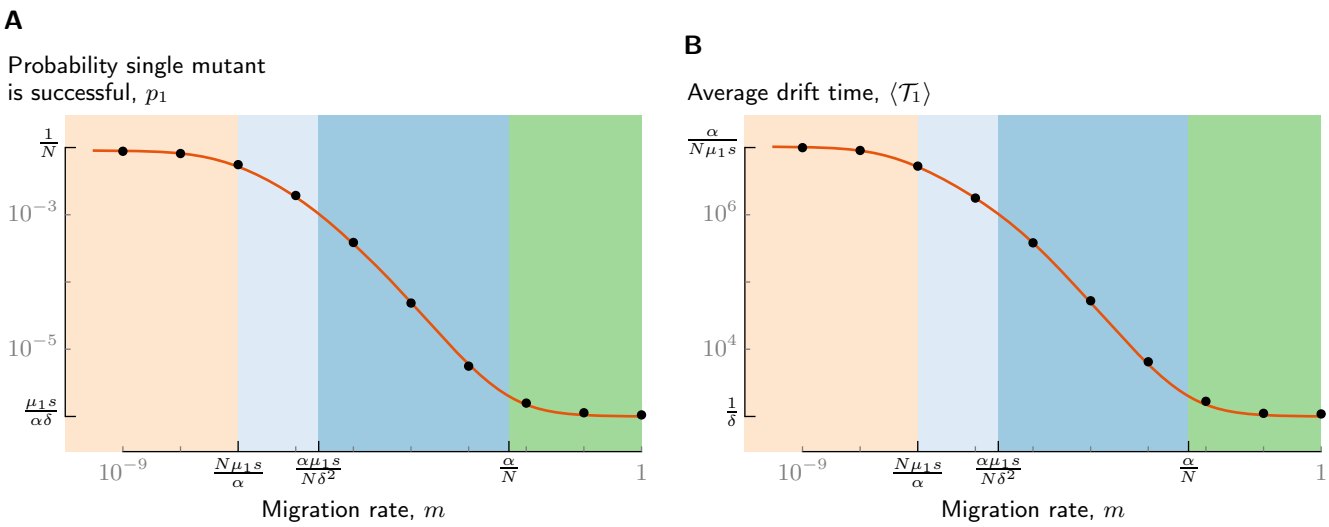

Figure 5: When single mutants are locally neutral ($\delta \ll \alpha/N$), our analytical predictions using the $\hat{F}$ and DBD methods accurately predict similar increases in the probability that a single mutant is successful, p_1, and the average drift time, $\langle\mathcal{T}_1\rangle$, as the migration rate, m, decreases. Here, single mutants are deleterious for tunneling at high migration rates ($\delta \gg \sqrt{\mu_1 s}$) and double mutants are locally beneficial ($s \gg \alpha/N$), so that there are four regimes for p_1 and $\langle\mathcal{T}_1\rangle$ depending on m. In each panel, the orange line shows the predictions for p_1 and $\langle\mathcal{T}_1\rangle$ by smoothly joining the $\hat{F}$ and DBD approximations. The dots show estimates from simulations (95% confidence intervals are smaller than the dots). Background colors correspond to the regimes labeled in Figure 4. Both axes have a log scale, with orders of magnitude delineated by small tick marks. Large tick marks indicate the regime boundaries on the x-axis and the limiting values of p_1 and $\langle\mathcal{T}_1\rangle$ on the y-axis as given in Equations (19) and (20). Parameters are $L = 100$, $N = 100$, $\alpha = 0.5$, $\mu_1 = 10^{-8}$, $\delta = 10^{-3}$, and $s = 0.05$, so that $N\delta/\alpha = 0.2$, $Ns/\alpha = 10$, and $\delta^2/\mu_1 s = 2\times10^3$.

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

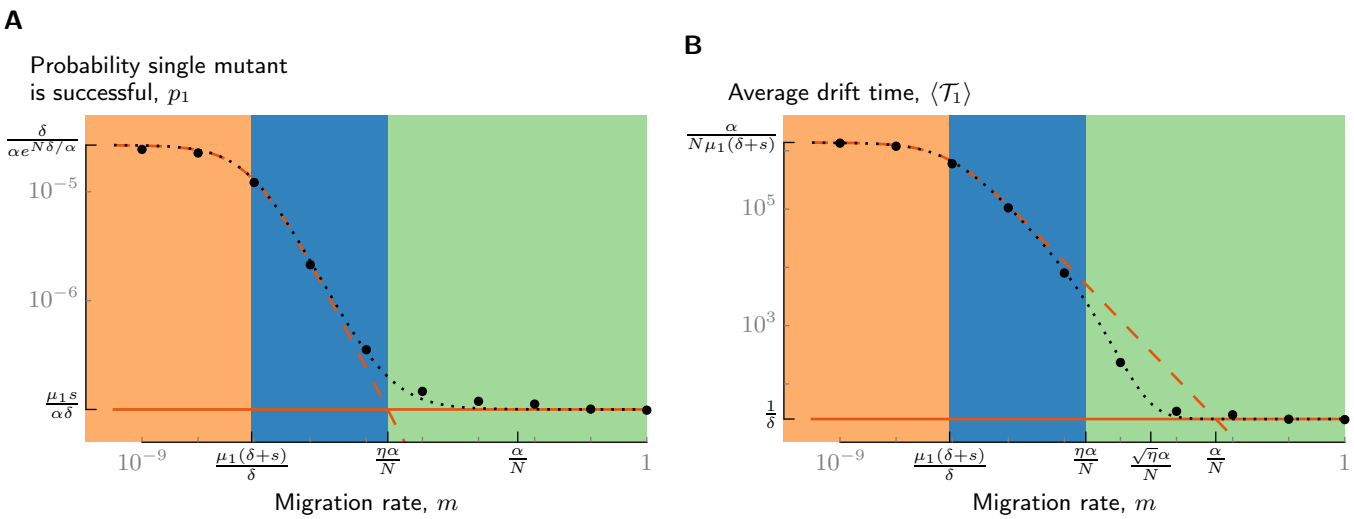

Figure 6: When single mutants are locally deleterious ($\delta \gg \alpha/N$), increases in the probability that a single mutant is successful, p_1, and the average drift time, $\langle\mathcal{T}_1\rangle$, at low migration rates are due to the contributions of type-B mutants. Solid orange lines show predictions for type-A lineages ($p_1^{(A)}$ or $\langle\mathcal{T}_A^{(1)}\rangle$) dashed orange lines show predictions for type-B lineages ($p_1^{(B)}$ or $\langle\mathcal{T}_B^{(1)}\rangle$) and the dotted black line shows predictions for p_1 and $\langle\mathcal{T}_1\rangle$ accounting for both types. Large black dots show estimates from simulations (95% confidence intervals are smaller than the dots). Background colors correspond to the regimes labeled in Figure 4. Both axes have a log scale, with orders of magnitude delineated by small tick marks. Large tick marks indicate the regime boundaries on the x-axis and the limiting values of p_1 and $\langle\mathcal{T}_1\rangle$ on the y-axis as given in Equations (24) and (25). Parameters are $L = 50$, $N = 100$, $\alpha = 0.5$, $\mu_1 = 4\times10^{-8}$, $\delta = 0.04$, and $s = 0.05$, so that $N\delta/\alpha = 8$ and $Ns/\alpha = 10$.

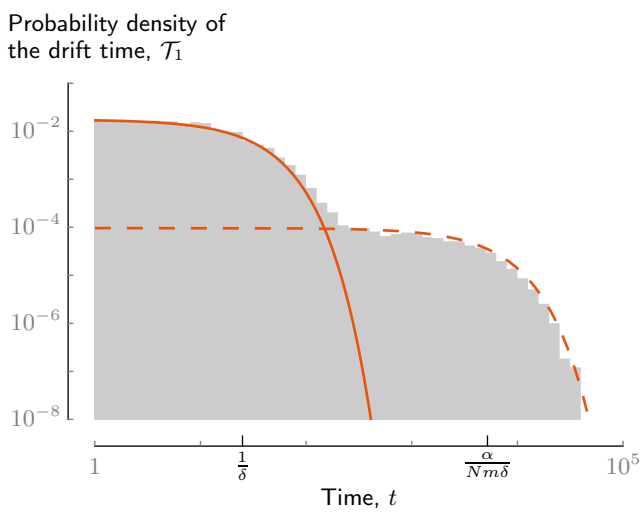

Figure 7: When single mutants are locally deleterious ($\delta \gg \alpha/N$), the distinct behavior of type-A and type-B successful single mutants can be seen in the two distinct humps in the distribution of drift times when $m \sim \eta\alpha/N$. The hump at times $\sim 1/\delta$ is from successful type-A mutants, while the hump at $\sim \alpha/Nm\delta$ is from successful type-B mutants. The solid and dashed orange lines show predictions for the distributions for type-A and type-B mutants, respectively, while the grey area is a histogram of simulation results. Parameters are as in Figure 6, but with $m = 2.4\times10^{-5} \approx \eta\alpha/N$, so that a successful mutant is equally likely to be type A or type B.

bioRxiv preprint doi: https://doi.org/10.1101/079624; this version posted October 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

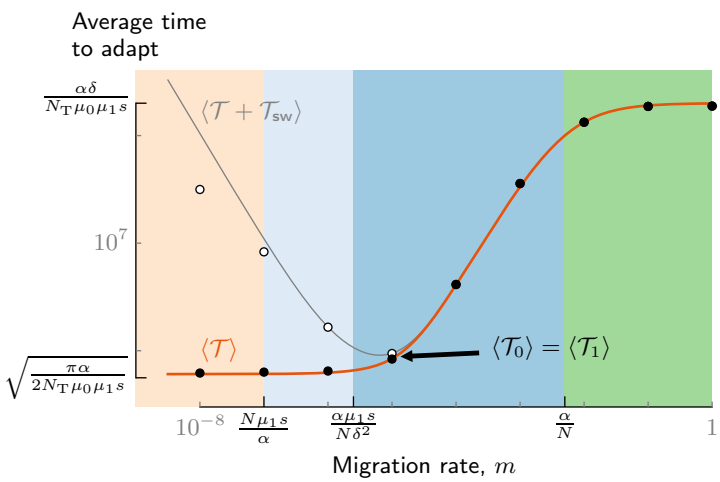

Figure 8: The drift and sweep times limit how much extreme subdivision decreases the average time for the population to adapt. Shown is the average waiting time for a successful double mutant, $\langle\mathcal{T}\rangle$, (prediction: orange line; simulation results: black dots) and the average waiting time for the double mutant to fix in the total population (prediction: grey line; simulation results: white dots). Single mutants are locally neutral ($N\delta/\alpha = 0.2$); parameters, x-axis, and background colors are the same as in Figure 5, while $\mu_0 = 5\times10^{-7}$, so that $N_{\mathrm{T}}\mu_0/\alpha = 10^{-2}$. For these parameters, the migration rate where $\langle\mathcal{T}_0\rangle = \langle\mathcal{T}_1\rangle$ is coincidentally near that where $\langle\mathcal{T}_{\mathsf{sw}}\rangle = \langle\mathcal{T}\rangle$.

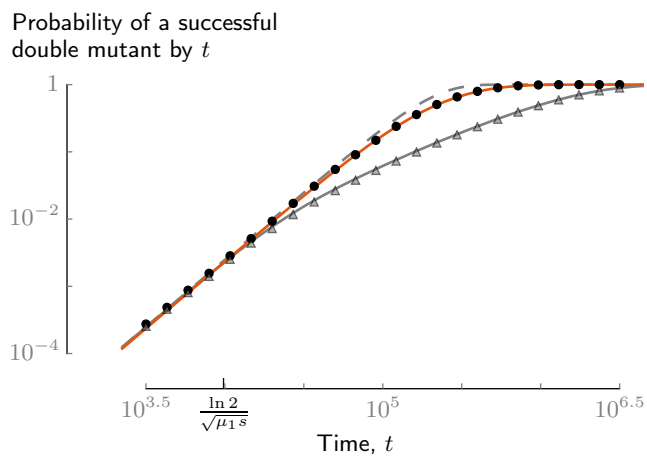

Figure 9: The probability $\Pr[\mathcal{T} < t]$ that a successful double mutant arises by time t is only increased by subdivision for times later than the average drift time in an unstructured population (here, approximately $\ln(2)/\sqrt{\mu_1 s}$). The grey line shows $\Pr[\mathcal{T} < t]$ for an unstructured population, while grey triangles show estimates from simulations. The orange line shows $\Pr[\mathcal{T} < t]$ for a subdivided population of equal total size in which $\langle\mathcal{T}\rangle$ is approximately five times smaller, while black dots show estimates from simulations. The dashed grey line shows the prediction for $\Pr[\mathcal{T} < t]$ under neutral semi-deterministic dynamics. Here, single mutants are neutral for tunneling in the unstructured and subdivided populations. Parameters are $N_{\mathrm{T}} = 5\times10^5$, $\mu_0 = 10^{-8}$, $\mu_1 = 10^{-7}$, $\alpha = 0.5$, $\delta = 0$, and $s = 0.05$; the subdivided population is split into $L = 10^3$ demes of $N = 500$ individuals each, with migration rate $m = 10^{-5}$.