BMC
Evolutionary Biology

The competition between simple and complex evolutionary trajectories in asexual populations

Ian E Ochs and Michael M Desai*

Abstract

Background: On rugged fitness landscapes where sign epistasis is common, adaptation can often involve either individually beneficial "uphill" mutations or more complex mutational trajectories involving fitness valleys or plateaus. The dynamics of the evolutionary process determine the probability that evolution will take any specific path among a variety of competing possible trajectories. Understanding this evolutionary choice is essential if we are to understand the outcomes and predictability of adaptation on rugged landscapes.

Results: We present a simple model to analyze the probability that evolution will eschew immediately uphill paths in favor of crossing fitness valleys or plateaus that lead to higher fitness but less accessible genotypes. We calculate how this probability depends on the population size, mutation rates, and relevant selection pressures, and compare our analytical results to Wright-Fisher simulations.

Conclusion: We find that the probability of valley crossing depends nonmonotonically on population size: intermediate size populations are most likely to follow a "greedy" strategy of acquiring immediately beneficial mutations even if they lead to evolutionary dead ends, while larger and smaller populations are more likely to cross fitness valleys to reach distant advantageous genotypes. We explicitly identify the boundaries between these different regimes in terms of the relevant evolutionary parameters. Above a certain threshold population size, we show that the probability that the population finds the more distant peak depends only on a single simple combination of the relevant parameters.

Keywords: Epistasis, Rugged fitness landscape, Fitness valley

Background

In an adapting population, evolution often has the potential to follow many distinct mutational trajectories. In order to predict how the population will adapt, we must understand how evolution chooses among these possibilities. Many experimental and theoretical studies have analyzed this question, focusing primarily on the simple case where epistasis is absent, so that each mutation has some fixed fitness effect [1-6]. This work can explain the probability that a given mutation will fix as a population adapts, as a function of its fitness effect, the population size, mutation rate, distribution of fitness effects of other mutations, and other parameters of the evolutionary process.

However, the fitness effect of a mutation often depends on the genetic background in which it occurs. A particularly interesting form of this phenomenon, *sign epistasis*, occurs when several mutations are individually neutral or deleterious but their combination is beneficial [7]. Sign epistasis has been observed repeatedly in experiments [8-13], and plays a central role in the evolution of complex phenotypes that involve multiple interacting components. When sign epistasis is present, adaptation can involve passing through genotypes of lower fitness — i.e. a population may have to cross a fitness valley or plateau. Thus the fate of a mutation depends not only on its fitness, but also on its adaptive potential [14].

Several recent theoretical studies have analyzed the evolutionary dynamics of fitness valley crossing [15-20]. This

*Correspondence: mmdesai@fas.harvard.edu
Department of Organismic and Evolutionary Biology, Department of Physics, and FAS Center for Systems Biology, Harvard University, 02138 Cambridge, MA, USA

work has focused on calculating the rate at which adapting populations cross a valley or plateau, in the absence of any other possible mutational trajectories. However, individually beneficial mutations may often compete with more complex evolutionary trajectories. We must then ask how likely evolution is to eschew the immediately uphill paths, and instead cross valleys or plateaus to reach better but less accessible genotypes. In other words, when the fitness landscape is rugged, we wish to understand whether evolution will take the more "farsighted" path to reach distant advantageous genotypes, rather than a "greedy" trajectory that fixes immediately beneficial mutations regardless of whether these may lead to evolutionary dead ends.

In this article, we analyze this evolutionary choice between immediately beneficial mutations and more complex mutational trajectories that ultimately lead to higher fitness. We calculate the probability that an adapting population will follow each type of competing trajectory, as a function of the population size, mutation rates, and selection pressures. We focus on asexual populations, where the only way for a population to acquire a complex adaptation is for a single lineage to acquire each mutation in turn. Our analysis is similar in spirit to earlier work which also considered the tradeoff between short-term and long-term fitness advantages [21-24]. However, these earlier studies dealt with competition between different strictly uphill or neutral paths, and considered the case where the less beneficial initial mutation led to better long-term evolutionary opportunities. In contrast, our analysis describes the competition between uphill mutations and more complex trajectories. While these two cases can be qualitatively similar in very small populations, they lead to very different dynamics in larger populations where the sign of the effect of the intermediate mutation can play a crucial role.

Our results show that population size has a crucial impact on how "farsighted" evolution can be. This dependence is not monotonic: evolution at intermediate population sizes is most "greedy", while both larger and smaller populations are more likely to eschew uphill paths in favor of complex trajectories. In large populations, our results show that a single parameter reliably predicts the extent of this evolutionary "foresight" across a wide range of parameters. Finally, we describe how our analysis can be generalized to predict how evolution will choose among even more complex trajectories, such as broad fitness valleys with multiple intermediate genotypes, and we discuss evolution in genotype spaces with many possible evolutionary paths.

Methods

We are interested in how a population makes an evolutionary choice when confronted with multiple possible mutational trajectories. Specifically, we focus on

the extent to which adaptation proceeds by crossing fitness valleys rather than acquiring immediately beneficial (uphill) mutations. Of course, the relative frequency of valley crossing will depend on the number of available fitness valleys, their depth, and the fitness advantage of the multiple-mutants, as well as the distribution of fitness effects (DFE) of the uphill mutants. Our goal is to understand how the prevalence of valley crossing depends on these factors.

Model

Throughout most of this article, we consider the simplest context in which we can address this question: the choice between a single uphill path and a single fitness valley. Specifically, we consider a haploid asexual population of constant size N which can either acquire an uphill mutation (u) that confers an immediate fitness advantage s_u, or alternatively acquire a deleterious fitness valley intermediate (i) with fitness deficit δ_i on which background a double-mutant (v) with fitness $s_v > s_u$ can arise. This scenario is illustrated in Figure 1. We also consider the case of a fitness plateau, where $\delta_i = 0$.

Because we are interested in the evolutionary choice between competing mutational trajectories, we assume that these two trajectories are mutually exclusive (i.e. the mixed genotypes ui and uv are strongly deleterious), so that only one genotype (either u or v) can eventually fix in the population. As a measure of evolutionary foresight, we analyze the probability that the double-mutant v fixes as a function of the relevant mutation rates, selection coefficients, and population size. In some situations, we could imagine that after either genotype u or v fixes, another set of competing potential trajectories become available. In this case, our analysis predicts the long-term relative ratio of fixed uphill versus valley-crossing mutations. In the Discussion, we consider how this model can be extended to the situation where there are many different competing uphill paths and valleys, and to broader fitness valleys involving multiple intermediate genotypes.

Simulations

In addition, we compare our analytical predictions for valley crossing probability to Wright-Fisher simulations. Each simulated population was evolved until either the uphill genotype or valley-crossing genotype fixed. Valley crossing probabilities were then inferred from the number of trials in which the valley-crossing genotype fixed, out of 1000 trials per parameter set.

Results

In the absence of the uphill genotype, fitness valley crossing can be modeled as a homogeneous Poisson process with rates as calculated by [17]. In small populations, the primary role of the uphill genotype is to introduce an
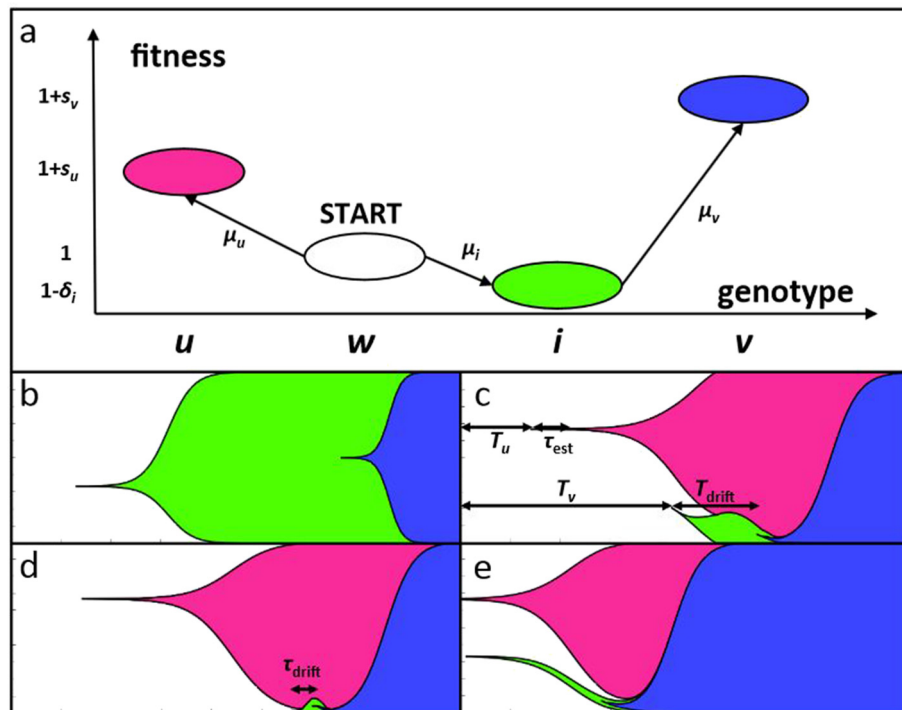
Figure 1 Model and characteristic trajectories. (a) The model to study fitness valley crossing prevalence. The population starts as wild type (*w*), and then acquires uphill mutations (*u*) at rate μ_u that confer an immediate fitness advantage s_u, and acquires deleterious fitness valley intermediates (*i*) at rate μ_i with fitness deficit δ_i on which background double-mutants (*v*) with fitness $s_v > s_u$ arise at rate μ_v. **(b)-(e)** The four main forms of fitness valley crossing. **(b)** Small populations are characterized by low genetic diversity and strong genetic drift, leading sequential fixation of intermediates to dominate the dynamics. **(c)** For larger populations, genetic diversity is maintained longer, and double mutants will tend to arise on transient single-mutant backgrounds, in a process known as stochastic tunneling. **(d)** If the drift time is small compared to the maximal rate of change in background fitness, we can approximate the drift time of the intermediate by its expectation, dramatically simplifying the mathematical analysis. **(e)** For very large populations, we can treat single-mutants deterministically, in a process dubbed semi-deterministic tunneling.

effective *time limit* on this process: once an uphill mutation destined to survive drift first occurs, it very quickly fixes, leading to the extinction of the wild-type. The probability of valley-crossing can thus be calculated as the probability that the intermediate *i* fixes before the uphill genotype *u*. An example of this is shown in Figure 1b.

In larger populations, the dynamics are more complex, as illustrated in Figure 1c. Rather than leading to a single cutoff time for valley-crossing to occur, the single-mutant occurs and gradually increases in frequency. This leads to a decline in the size of the wild-type background on which intermediate and valley-crossing mutants can arise, and a corresponding increase in the mean fitness of the population (Figure 1c). These effects gradually reduce the rate at which intermediates are produced, and make these intermediates effectively more deleterious relative to the mean fitness. These factors reduce the rate of the valley-crossing process. Thus valley-crossing becomes an *inhomogenous* Poisson process, with a rate that depends on the random appearance time T_u of the uphill mutant.
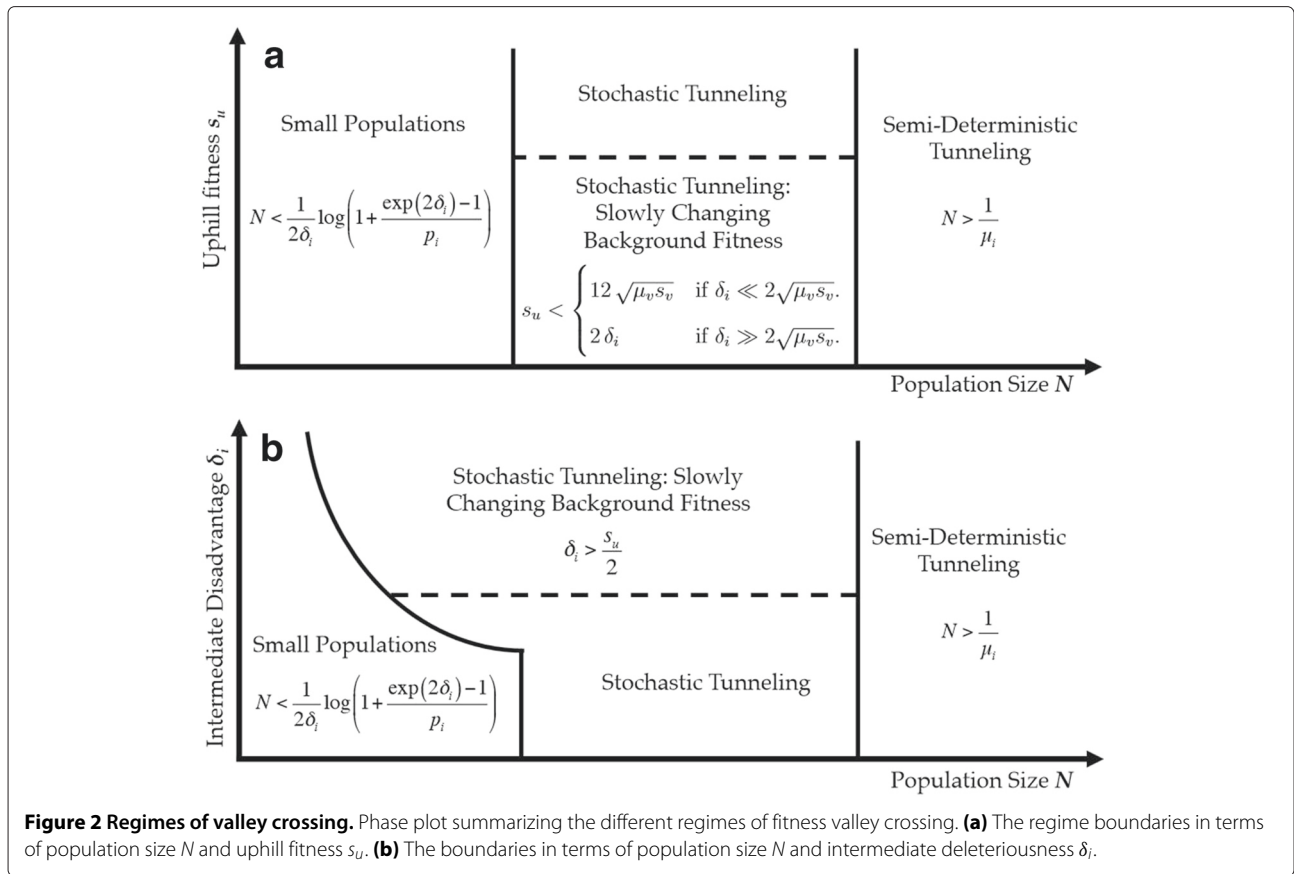
In general, these effects of interference and tunneling are complex. However, the analysis becomes simpler in

two specific regimes. When the expected drift time of the intermediate genotype is short, we can neglect the changing background fitness due to the uphill mutant during this drift time (Figure 1d). Alternatively, for very large populations ($N\mu > 1$), the Poisson process approximation breaks down and both uphill and intermediate mutations can be treated deterministically (Figure 1e), and only the valley-crossing genotype must be treated stochastically.

These various regimes are illustrated in Figure 2. We now analyze each in turn, assuming weak selection ($s_j < 1$) for all genotypes throughout. Taken together, this provides a complete picture of the probability that evolution will eschew the immediately uphill path in favor of the more complex adaptation.

Small populations

When the population size is small enough that the probability of stochastic tunneling is very low, the population is generally clonal or nearly clonal, and moves in Markovian jumps between neighboring genotypes. The transition between genotypes *i* and *j* occurs at rate

Figure 2 Regimes of valley crossing. Phase plot summarizing the different regimes of fitness valley crossing. **(a)** The regime boundaries in terms of population size N and uphill fitness s_u. **(b)** The boundaries in terms of population size N and intermediate deleteriousness δ_i.

$$r_{ij} = N\mu_{ij}\pi_{ij}, \qquad (1)$$

where π_{ij} is the probability that a single j mutant will give rise to a lineage that fixes, given by the standard formula,

$$\pi_{ij} = \frac{1 - e^{-2(s_j - s_i)}}{1 - e^{-2N(s_j - s_i)}}. \qquad (2)$$

We refer to this as the sequential fixation regime. Because we are considering neutral and weakly deleterious intermediates, we account for the possibility of back-mutation to the wild type if the intermediate fixes. Therefore, the process can be modeled as an absorbing states Markov chain, where the wild type and intermediates act as transient states, and the uphill and double-mutant genotypes act as absorbing states. From elementary Markov chain theory, we find

$$P_{\text{cross}} = \frac{r_{wi}r_{iv}}{r_{wi}r_{iv} + r_{wu}(r_{iw} + r_{iv})} = \left[1 + \left(\frac{\pi_{wu}}{\pi_{wi}}\right)\left(\frac{\mu_u}{\mu_i}\right)\left(1 + \frac{\mu_i\pi_{iw}}{\mu_v\pi_{iv}}\right)\right]^{-1}. \qquad (3)$$

As the population size increases, $\pi_{wu} \rightarrow 2s_u$ and $\pi_{wi} \rightarrow 0$, so $\frac{\pi_{wu}}{\pi_{wi}} \rightarrow \infty$, and $P_{\text{cross}} \rightarrow 0$. Thus we find that within the sequential fixation regime, larger population sizes are less likely to cross fitness valleys.

Stochastic tunneling

For large populations, the probability that deleterious intermediates will fix declines drastically, and successful double mutants will instead arise on the unfixed single-mutant background in a process known as *stochastic tunneling* [16]. This transition occurs when

$$N > \frac{1}{2\delta_i} \log\left[1 + \frac{\exp(2\delta_i) - 1}{p_v}\right], \qquad (4)$$

where p_v is the probability that the intermediate lineage survives drift long enough to give rise to an ultimately successful double mutant lineage (we will explicitly calculate this probability below). We can then model the appearance of an intermediate mutant lineage destined to give rise to a double mutant lineage as a Poisson process. The rate λ_v at which these lineages appear is given by the rate at which intermediate mutations arise times the probability of success of the lineage, integrated over the drift time t_d after appearance of the single-mutant intermediate:

$$\lambda_v = N_{\text{wt}}\mu_i \int_0^\infty \frac{\partial p_i(t_d)}{\partial t_d} dt_d. \qquad (5)$$

Here N_{wt} is the wild-type population size, and $p_i(t_d)$ is the cumulative probability that a single-mutant lineage

will give rise to a successful double-mutant lineage by time t_d after it appears. This probability is given by [17]:

$$p_i(t_d) = 2 \left(\frac{(a_+ - 1)(1 - a_-)(1 - \exp[-(1 - \delta_{i,\text{eff}})(a_+ - a_-)t_d])}{a_+ - 1 + (1 - a_-)(1 - \exp[-(1 - \delta_{i,\text{eff}})(a_+ - a_-)t_d])} \right),$$
(6)

where

$$a_{\pm} = \frac{2 - \delta_{i,\text{eff}} - \mu_v s_{v,\text{eff}} \pm \sqrt{(\delta_{i,\text{eff}} + \mu_v s_{v,\text{eff}})^2 + 4\mu_v s_{v,\text{eff}}}}{2(1 - \delta_{i,\text{eff}})},$$
(7)

and $\delta_{i,\text{eff}}$ and $s_{v,\text{eff}}$ are the fitnesses of the intermediate and valley-crossing genotypes relative to the (time-dependent) mean population fitness. These effective fitnesses and N_{wt} depend on the background at time t_d in a way we must now consider. The background in turn is determined by the frequency $f_u(T_v + t_d)$ of the uphill genotype at time t_d after the appearance of the first single-mutant intermediate lineage destined for success at time T_v (Figure 1c). Thus the uphill genotype frequency sets the fitness background on which valley-crossing probabilities are determined.

To calculate the relevant effective parameters, we note that the appearance of uphill lineages destined for success can be modeled as a Poisson process. Moreover, because the valley-crossing genotypes make up a tiny fraction of the population (unless the double-mutant has already established), we can treat the genetic background on which these uphill lineages appear as essentially fixed. Therefore, the first uphill lineage destined to survive genetic drift will appear at time T_u, distributed exponentially with rate

$$\lambda_u = N\mu_u \pi_{wu} \approx N\mu_u(2s_u).$$
(8)

Once a successful uphill lineage appears, we assume it establishes in time $\tau_{\text{est}} = \gamma_e/(2s_u)$, where $\gamma_e \approx .577$ is the Euler-Mascheroni constant [25], and then sweeps deterministically according to

$$f_u(\hat{t}) = \frac{1 - \exp\left[-(\mu_u + s_u)\hat{t}\right]}{1 + (s_u/\mu_u)\exp\left[-(\mu_u + s_u)\hat{t}\right]},$$
(9)

where $\hat{t} \equiv t - T_u - \tau_{\text{est}}$ is the time after establishment of the uphill mutant.

Conditioning on the appearance time T_u, we can thus work out our effective parameters

$$N_{\text{wt}}(t \mid T_u) = N(1 - f_u(t - T_u - \tau_{\text{est}})) \quad (10)$$

$$\delta_{i,\text{eff}}(t + t_d \mid T_u) = -\delta_i + f_u(t + t_d - T_u - \tau_{\text{est}})s_u \quad (11)$$

$$s_{v,\text{eff}}(t + t_d \mid T_u) = s_v - f_u(t + t_d - T_u - \tau_{\text{est}})s_u. \quad (12)$$

These effective parameters encompass the two main effects of the sweeping uphill mutation on the valley crossing probability: the first represents the declining wild-type background on which new mutations can arise, and the remaining two represent the decreasing relative fitness of the valley-crossing lineage.

We are interested in the probability that a double-mutant lineage destined for success appears before the uphill genotype fixes. Integrating over all possible appearance times T_u, this is given by:

$$P_{\text{cross}} = \int_0^\infty dt_u \left(\lambda_u e^{-\lambda_u t_u}\right) \times \quad (13)$$
$$\left[1 - \exp\left[-\int_0^\infty dt \left(N_{\text{wt}}(t, t_u)\mu_i \int_0^\infty dt_d \frac{\partial p_i(t + t_d, t_u)}{\partial t_d}\right)\right]\right].$$

This integral is a complete solution for the probability of valley-crossing in the stochastic tunneling regime, provided that the population size is small enough that the Poisson process approximation above holds. Although it does not have a simple closed-form solution, we can easily evaluate the integral numerically. Alternatively, there is a simple and relevant parameter regime in which background fitnesses change slowly. We now consider this case, and show that it allows us to evaluate our expression for the valley-crossing probability explicitly. In a later section below, we turn to the alternative case where the Poisson process approximation breaks down, and we can instead treat all single-mutants deterministically; the valley-crossing probability also simplifies considerably in this very large population regime.

Slowly changing background fitness

One of the main complications of Equation (13) is the integral over possible drift times t_d, which reflects the increasing effective deleteriousness of the intermediate genotype as the uphill genotype sweeps to fixation and increases the mean fitness of the background population. However, when s_u is small or δ_i is large, this background fitness changes slowly compared to the intermediate drift time. In this case, we can treat the background during intermediate drift as effectively constant (Figure 1d). This eliminates the need for an integral over t_d, since the time-dependent probability of success of a single-mutant at time t is fully determined by $f_u(t + t_d) \approx f_u(t)$.

We can further simplify the analysis if we treat the probability of crossing the valley as a function of two probabilities: the probability $P_{v,1}$ that the first successful valley-crossing lineage appears before the first successful uphill lineage establishes, and the probability $P_{v,2}$ that a successful valley-crossing lineage appears after the uphill mutant establishes:

$$P_{\text{cross}} = P_{v,1} + (1 - P_{v,1}) P_{v,2}. \quad (14)$$

The calculation of $P_{v,1}$ takes place on a purely wild-type background, so we can use

$$\lambda_v = N\mu_i p_v, \quad (15)$$

where p_v, the probability that the intermediate lineage survives drift long enough to give rise to an ultimately successful double mutant lineage, is given by [17]:

$$p_v = -\delta_i + \sqrt{\delta_i^2 + 4\mu_v s_v} \approx \begin{cases} 2\sqrt{\mu_v s_v} & \text{if } \delta_i \ll 2\sqrt{\mu_v s_v} \\ 2\mu_v s_v/\delta_i & \text{if } \delta_i \gg 2\sqrt{\mu_v s_v}. \end{cases} \tag{16}$$

Meanwhile, λ_u is unchanged from the original analysis. $P_{v,1}$ is determined by a race between these two exponential random variables. Using basic properties of the exponential, the probability that the double-mutant appears before the uphill genotype establishes is therefore

$$P_{v,1} = \begin{cases} \left(\frac{\lambda_v}{\lambda_v + \lambda_u}\right) e^{-\lambda_u \Delta\tau} & \text{if } \Delta\tau > 0 \\ 1 - \left(\frac{\lambda_u}{\lambda_v + \lambda_u}\right) e^{-\lambda_v \Delta\tau} & \text{if } \Delta\tau < 0, \end{cases} \tag{17}$$

where $\Delta\tau$ represents the difference between the mean drift time τ_{drift} of the valley-crossing lineage and the mean establishment time τ_{est} of the uphill lineage,

$$\Delta\tau = \tau_{\text{drift}} - \tau_{\text{est}}. \tag{18}$$

Here τ_{est} is as given above, and from [17], we can approximate the drift time as

$$\tau_{\text{drift}} = \begin{cases} \log 2/\sqrt{\mu_v s_v} & \text{if } \delta_i \ll 2\sqrt{\mu_v s_v} \\ 1/\delta_i & \text{if } \delta_i \gg 2\sqrt{\mu_v s_v}. \end{cases} \tag{19}$$

If the successful uphill lineage establishes, the first successful valley-crossing lineage still has a chance to appear and outcompete it, albeit on a declining wild-type background. Thus for $P_{v,2}$ we get a similar integral as in the original analysis. However, since we are approximating p_i as constant, the rate of successful lineage generation simplifies to

$$\hat{\lambda}_v(t) = \mu_i N_{\text{wt}} \, p_i(\delta_{i,\text{eff}}, s_{v,\text{eff}})$$
$$= \mu_i N(1-f_u) \left(-(\delta_i + s_u f_u) + \sqrt{(\delta_i + s_u f_u)^2 + 4\mu_v(s_v - s_u f_u)}\right). \tag{20}$$

Integrating and assuming mutation rates are small compared to selection pressures, we find:

$$\int_0^\infty dt \, \hat{\lambda}_v(t) \approx \left(\frac{\log N s_u}{s_u}\right) N \mu_i p_i = \left(\frac{\log N s_u}{s_u}\right) \lambda_v. \tag{21}$$

Combining these results, we find

$$P_{\text{cross}} = P_{v,1} + (1 - P_{v,1})\left(1 - e^{-\left(\frac{\log N s_u}{s_u}\right)\lambda_v}\right)$$
$$= 1 - (1 - P_{v,1}) \, e^{-\left(\frac{\log N s_u}{s_u}\right)\lambda_v}. \tag{22}$$

We expect this result to be valid provided that δ_i is effectively constant over the expected drift time. This will hold when

$$s_u \ll \begin{cases} 2\sqrt{\frac{2}{\log 2}}\sqrt{\mu_v s_v} & \text{if } \delta_i \ll 2\sqrt{\mu_v s_v} \\ 2\delta_i & \text{if } \delta_i \gg 2\sqrt{\mu_v s_v} \end{cases}. \tag{23}$$

Semi-deterministic tunneling

We now consider the case where $N\mu_i > 1$ and $N\mu_u > 1$, and hence the Poisson process approximation used to derive Equation (13) breaks down. Fortunately, in this regime the number of single-mutant intermediates and uphill mutants in the population are well approximated by their deterministic expectation (Figure 1e). Thus the only random variable is the appearance time of the first successful double-mutant lineage, which occurs with rate

$$\lambda_v(t) = N f_i \mu_v \pi_{iv} \approx 2N f_i \mu_v(s_v - s_u f_u). \tag{24}$$

Because intermediates never make up a large portion of the population, f_u is unaffected by f_i, and hence is still given by Equation (9). We can then approximate the frequency f_i of the single-mutant intermediates using mutation-selection balance with a declining wild-type population:

$$f_i = f_i^*(1 - f_u), \tag{25}$$

where f_i^* gives the independent deterministic dynamics (mutation-selection balance) of single mutants on the wild-type background, and $(1 - f_u)$ is the size of the wild-type background. It is useful to transform this into an integral in the frequency domain:

$$\int_0^\infty \lambda_v(t)dt = \int_0^1 \lambda_v(f_u)\left(\frac{\partial f_u}{\partial t}\right)^{-1} df_u. \tag{26}$$

We note from equation 9 that

$$\frac{\partial f_u}{\partial t} = \mu_u(1 - f_u) + s_u f_u(1 - f_u), \tag{27}$$

$$t(f_u) = \frac{1}{\mu_u + s_u} \log\left(\frac{1 + (s_u/\mu_u)f_u}{1 - f_u}\right). \tag{28}$$

We further approximate

$$f_i^* = \begin{cases} \mu_i t & \text{if } \delta_i \ll 2\sqrt{\mu_v s_v} \\ \frac{\mu_i}{\delta_i}(1 - \exp(-\delta_i t)) & \text{if } \delta_i \gg 2\sqrt{\mu_v s_v}. \end{cases} \tag{29}$$

Combining these expressions and assuming $\mu_u \ll s_u$, we find

$$\int_0^\infty \lambda_v(t)dt = 2N\mu_i \mu_v s_v \gamma / s_u, \tag{30}$$

where

$$\gamma = \begin{cases} s_u^{-1}\left(\frac{1}{2}\log(s_u/\mu_u)^2 - (s_u/s_v)\log(s_u/\mu_u) + \pi^2/6\right) & \text{if } \delta_i \ll 2\sqrt{\mu_u s_v} \\ \delta_i^{-1}\left(\log(s_u/\mu_u) - (s_u/\delta_i)(1 - (s_u/\mu_u)^{-\delta_i/s_u})\right) & \text{if } \delta_i \gg 2\sqrt{\mu_u s_v}. \end{cases} \tag{31}$$

This gives

$$P_{\text{cross}} = 1 - \exp\left[-\int \lambda_v(t)dt\right] = 1$$
$$- \exp\left[\frac{-2N\mu_i\mu_v s_v \gamma}{s_u}\right] = 1 - \exp\left[-2N\Gamma\right],$$
(32)

where we have defined the useful quantity

$$\Gamma \equiv \frac{\mu_i\mu_v s_v \gamma}{s_u} .$$
(33)

Thus we see that in very large populations, the probability of valley-crossing depends in a simple way on the single composite parameter Γ. The form of this composite parameter depends crucially on whether the fitness cost of the intermediate genotype is large or small, as defined in Equation (31).
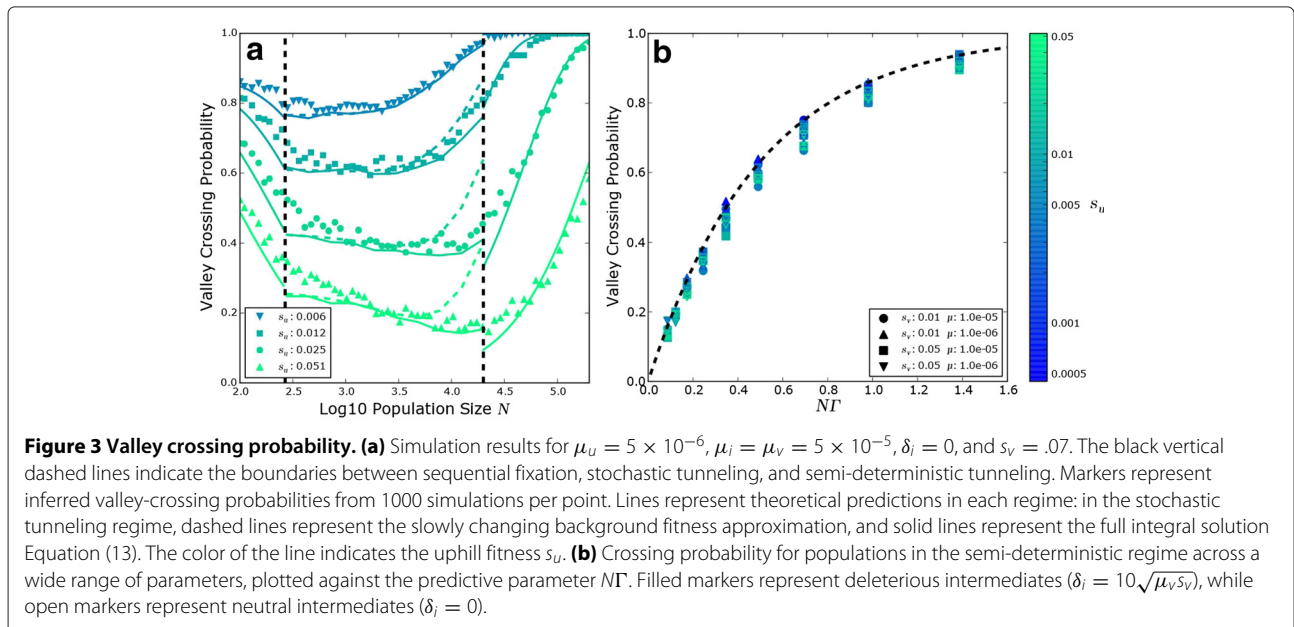
Discussion

Our results have shown that the size of a population strongly influences how "farsighted" it can be. In small populations, genetic drift is strong relative to selection, so the evolutionary dynamics proceeds by sequential fixation. Since fixation of a deleterious intermediate becomes less likely in larger populations, this means that increasing the population size initially decreases the relative influence of fitness-valley crossing. However, as the population size increases further, beneficial mutations take longer to fix, maintaining diversity in the population and allowing double mutants to stochastically tunnel on the declining wild-type background [15-17]. Together, these effects lead to a non-monotonic relationship between population size

and the probability that evolution will favor the complex adaptation over the directly uphill path, as illustrated in Figure 3a. This nonmonotonic dependence on population size is similar in spirit to the results of earlier work analyzing evolution on epistatic landscapes in the absence of fitness valleys [21-23].

It is interesting to note that P_{cross} does not immediately begin to rise with the onset of tunneling. Instead, the dependence is more complex, as a consequence of the tradeoff between increasing mutation rates and fixation times. Nevertheless, for populations in which the transition to valley-crossing behavior occurs in the semi-deterministic regime, we can derive a simple expression for the threshold size at which the population will tend to cross valleys with probability P_{cross}. A straightforward inversion of (32) gives

$$N = \frac{-\log\left[1 - P_{\text{cross}}\right]}{2\Gamma},$$
(34)

valid for large population sizes in the semi-deterministic regime. Thus in this regime the threshold size above which a population exhibits a given degree of foresight (i.e. has a particular P_{cross}) depends only on Γ. To illustrate this, in Figure 3b we show P_{cross} as a function of $N\Gamma$ for a variety of simulations across different values of $\mu_u, \mu_i, \mu_v, s_u, \delta_i$ and s_v. It is clear that even across a wide parameter range, $N\Gamma$ is a reliable predictor of valley crossing probability in the semi-deterministic limit.

Figure 3 Valley crossing probability. (a) Simulation results for $\mu_u = 5 \times 10^{-6}$, $\mu_i = \mu_v = 5 \times 10^{-5}$, $\delta_i = 0$, and $s_v = .07$. The black vertical dashed lines indicate the boundaries between sequential fixation, stochastic tunneling, and semi-deterministic tunneling. Markers represent inferred valley-crossing probabilities from 1000 simulations per point. Lines represent theoretical predictions in each regime: in the stochastic tunneling regime, dashed lines represent the slowly changing background fitness approximation, and solid lines represent the full integral solution Equation (13). The color of the line indicates the uphill fitness s_u. **(b)** Crossing probability for populations in the semi-deterministic regime across a wide range of parameters, plotted against the predictive parameter $N\Gamma$. Filled markers represent deleterious intermediates ($\delta_i = 10\sqrt{\mu_v s_v}$), while open markers represent neutral intermediates ($\delta_i = 0$).

Multiple intermediates and evolutionary predictability

Recently, Szendro *et al.* [26] simulated evolution across a wide range of population sizes on an experimentally-derived epistatic fitness landscape, finding that "evolutionary entropy" (i.e. unpredictability over all possible outcomes, given by $S = -\sum p_j \log p_j$) varied nonmonotonically with population size. Specifically, these authors found that entropy initially decreased with N above a characteristic population size $N \propto 1/\mu$, before increasing again above a second characteristic size $N \propto 1/\mu^2$; they argued that these points were related to the supply rate of single and double mutants respectively. Our analysis is consistent with these results. For example, the increase in entropy at $N \propto 1/\mu^2$ found by [26] corresponds to our result that valley-crossing begins to significantly influence evolution when $N \sim 1/\Gamma$: this is approximately proportional to $1/\mu^2$, albeit with an additional log dependence on μ from the γ factor that would be harder to observe experimentally.

We find related behavior if we extend our analysis to valleys with more intermediates. As a simple example, we consider a fitness landscape where we add deleterious intermediates u_0 and v_0 (each occurring at rate μ_0 and leading to fitness cost δ_i) to the uphill and valley-crossing branches respectively, so that we now have competition between a single-intermediate valley and a two-intermediate valley. In a large enough population, mutation-selection balance will ensure that

$$N_{u_0} = N_{v_0} = \frac{N\mu_0}{\delta_i}. \qquad (35)$$

If we assume that these sub-populations are large enough that double-mutants behave deterministically, then we find the crossing probability obeys

$$-\log\left[1 - P_{\text{cross}}\right] = 2\left(\frac{\mu_0}{\delta_i}\right)N\Gamma = 2\left(\frac{s_v\gamma}{s_u\delta_i}\right)N\mu_0\mu_i\mu_v. \qquad (36)$$

Thus our analysis of valleys with multiple intermediates suggests that the $N\mu^2$ entropy peak is not unique: as the population grows larger, there should be entropy peaks corresponding to foresight across valleys with increasing numbers of intermediates. The emergence of such second peaks has been observed in simulations [23], and our model offers a quantitative outline of where such peaks should occur given the relevant evolutionary parameters. In the example above, for instance, there would be an entropy peak approximately proportional to $N\mu^3$, and in general, we could expect entropy peaks at points approximately proportional to $N\mu^n$ for $(n - 1)$-intermediate valleys. In practice, however, the semideterministic approximation will break down for any sizable number of intermediates, unless the population size is unrealistically large.

Many paths

Throughout this paper, we have assumed the presence of a single uphill mutation and fitness valley. We now consider how our analysis can be extended to predict how evolution chooses among many such possible mutational trajectories.

In small populations that are in the sequential fixation regime, we simply add additional transient transition matrix elements representing different mutations, with the uphill mutations transitioning to the uphill absorbing state, and similarly for the valley-crossing mutations. When stochastic tunneling is important, we must instead add the rates of single valley-crossing mutants to get a total rate of

$$\Lambda_v = \sum_v \lambda_v, \qquad (37)$$

and similarly find the total rate of uphill mutants that are destined to survive drift,

$$\Lambda_u = \sum_u \lambda_u = \sum_u N\mu_u\pi(s_u) = \int NU_b\pi(s_u)\rho(s_u)ds_u, \qquad (38)$$

where in the last equality we have replaced a discrete collection of uphill mutations with a continuous distribution of uphill fitness effects $\rho(s)$, and a discrete collection of mutation rates μ_u with a total beneficial mutation rate U_b. This is valid as long as $\Lambda_u \ll 1$. Once an uphill mutation destined to survive drift occurs, the probability that it has fitness s is given by the ratio between its partial rate and the total rate; formally, the probability density is given by:

$$f(s) = N\mu_u\pi(s)\rho(s)/\Lambda_u. \qquad (39)$$

Using these expressions, we can integrate our results from the analysis over all possible trajectories. However, we note that if there are a large number of weakly beneficial mutations, it is possible the first successful lineage to appear will be outcompeted by a stronger uphill mutation that arises later but fixes first. Our analysis applies provided that we consider only uphill mutations that reach a significant portion k of the population before a new, more fit uphill mutant is expected to be produced: i.e.

$$\Lambda_u\tau_k = \left(\int_{s_{\text{cutoff}}}^{\infty} N\mu_s\pi(s)\rho(s)ds\right)(\tau_k) < 1, \qquad (40)$$

where τ_k is the expected time for a single-mutant destined for success to make up frequency k of the population. This is consistent with our intuition that as the population size grows larger, we increasingly expect the mutations of largest effect to dominate the dynamics.

Conclusions

Using a simple three-locus fitness landscape model (Figure 1a), we identified several regimes of valley-crossing with qualitatively different behavior (Figure 2). By examining the behavior in each of these regimes in turn, we found that the probability of valley-crossing has a complex, non-monotonic dependence on population size (Figure 3a), and identified a parameter Γ that reliably predicts the population size at which valley crossing becomes preferred (Figure 3b). Finally, we showed how these results can be extended to fitness valleys with more intermediates, and to fitness landscapes with many possible evolutionary trajectories, as is the case in most naturally occurring populations.

Abbreviation

DFE: Distribution of fitness effects.

References

1. Perfeito L, Fernandes L, Mota C, Gordo I. Adaptive mutations in bacteria: High rate and small effects. Science. 2007;317(5839):813–5.
2. Tenaillon O, Rodríguez-Verdugo A, Gaut RL, McDonald P, Bennett AF, Long AD, et al. The molecular diversity of adaptive convergence. Science. 2012;335(6067):457–61.
3. Lang GI, Rice DP, Hickman MJ, Sodergren E, Weinstock GM, Botstein D, Desai MM. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. Nature. 2013;500(7464):571–4.
4. Gerrish P, Lenski R. The fate of competing beneficial mutations in an asexual population. Genetica. 1998;102:127–44.
5. Schiffels S, Szollosi GJ, Mustonen V, Lassig M. Emergent neutrality in adaptive asexual evolution. Genetics. 2011;189(4):1361–75.
6. Good BH, Rouzine IM, Balick DJ, Hallatschek O, Desai MM. Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. Proc Nat Acad Sci. 2012;109:4950–5.
7. Weinreich DM, Watson RA, Chao L. Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. Evolution. 2005;59(6):1165–74.
8. Weinreich DM, Delaney NF, DePristo MA, Hartl DL. Darwinian evolution can follow only very few mutational paths to fitter proteins. Science. 2006;312(5770):111–4.
9. Poelwijk FJ, Kiviet DJ, Weinreich DM, Tans SJ. Empirical fitness landscapes reveal accessible evolutionary paths. Nature. 2007;445:383–6.
10. Kvitek DJ, Sherlock G. Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. PLoS Genet. 2011;7(4):1002056.
11. Silva RF, Mendonça SC, Carvalho LM, Reis AM, Gordo I, Trindade S, et al. Pervasive sign epistasis between conjugative plasmids and drug-resistance chromosomal mutations. PLoS Genet. 2011;7(7):1002181.
12. Dawid A, Kiviet DJ, Kogenaru M, de Vos M, Tans SJ. Multiple peaks and reciprocal sign epistasis in an empirically determined genotype-phenotype landscape. Chaos: Interdisciplinary J Nonlinear Sci. 2010;20(2):026105.
13. Salverda ML, Dellus E, Gorter FA, Debets AJ, Van Der Oost J, Hoekstra RF, et al. Initial mutations direct alternative pathways of protein evolution. PLoS Genet. 2011;7(3):1001321.
14. Woods RJ, Barrick JE, Cooper TF, Shrestha U, Kauth MR, Lenski RE. Second-order selection for evolvability in a large escherichia coli population. Science. 2011;331(6023):1433–6.
15. Weinreich DM, Chao L. Rapid evolutionary escape by large populations from local fitness peaks is likely in nature. Evolution. 2005;59:1175–82.
16. Iwasa Y, Michor F, Nowak MA. Stochastic tunnels in evolutionary dynamics. Genetics. 2004;166(3):1571–9.
17. Weissman DB, Desai MM, Fisher DS, Feldman MW. The rate at which asexual populations cross fitness valleys. Theor Population Biol. 2009;75(4):286–300.
18. Gokhale CS, Iwasa Y, Nowak MA, Traulsen A. The pace of evolution across fitness valleys. J Theor Biol. 2009;259(3):613–20.
19. Altland A, Fischer A, Krug J, Szendro IG. Rare events in population genetics: stochastic tunneling in a two-locus model with recombination. Phys Rev Lett. 2011;106(8):088101.
20. Weissman DB, Feldman MW, Fisher DS. The rate of fitness-valley crossing in sexual populations. Genetics. 2010;186(4):1389–410.
21. Rozen DE, Habets MG, Handel A, de Visser JAG. Heterogeneous adaptive trajectories of small populations on complex fitness landscapes. PLoS One. 2008;3(3):1715.
22. Handel A, Rozen DE. The impact of population size on the evolution of asexual microbes on smooth versus rugged fitness landscapes. BMC Evolutionary Biol. 2009;9(1):236.
23. Jain K, Krug J, Park S-C. Evolutionary advantage of small populations on complex fitness landscapes. Evolution. 2011;65(7):1945–55.
24. Van Nimwegen E, Crutchfield JP. Metastable evolutionary dynamics: crossing fitness barriers or escaping via neutral paths?. Bull Math Biol. 2000;62(5):799–848.
25. Desai MM, Fisher DS. Beneficial mutation–selection balance and the effect of linkage on positive selection. Genetics. 2007;176(3):1759–8.
26. Szendro IG, Franke J, de Visser JAG, Krug J. Predictability of evolution depends nonmonotonically on population size. Proc Nat Acad Sci. 2013;110(2):571–6.