The Rate of Fitness-Valley Crossing in Sexual Populations

Daniel B. Weissman,*,[1] **Marcus W. Feldman**[†] **and Daniel S. Fisher**[‡]

Institute of Science and Technology Austria, A-3400 Klosterneuburg, Austria and [†]*Department of Biology and*
[‡]*Department of Applied Physics, Stanford University, Stanford, California 94305*

ABSTRACT

Biological traits result in part from interactions between different genetic loci. This can lead to sign epistasis, in which a beneficial adaptation involves a combination of individually deleterious or neutral mutations; in this case, a population must cross a "fitness valley" to adapt. Recombination can assist this process by combining mutations from different individuals or retard it by breaking up the adaptive combination. Here, we analyze the simplest fitness valley, in which an adaptation requires one mutation at each of two loci to provide a fitness benefit. We present a theoretical analysis of the effect of recombination on the valley-crossing process across the full spectrum of possible parameter regimes. We find that low recombination rates can speed up valley crossing relative to the asexual case, while higher recombination rates slow down valley crossing, with the transition between the two regimes occurring when the recombination rate between the loci is approximately equal to the selective advantage provided by the adaptation. In large populations, if the recombination rate is high and selection against single mutants is substantial, the time to cross the valley grows exponentially with population size, effectively meaning that the population cannot acquire the adaptation. Recombination at the optimal (low) rate can reduce the valley-crossing time by up to several orders of magnitude relative to that in an asexual population.

MOST phenotypes depend on interactions between many different genes. For any given trait, there are likely to be many possible adaptations that involve combinations of mutations at different loci. Examples include signaling pathways that require coevolution of receptors and ligands (GOH *et al.* 2000), pathogens that may require several mutations to escape their hosts' immune response or to become resistant to drugs (LEVIN *et al.* 2000), and bacteria that require multiple mutations to evolve the ability to metabolize a new nutrient (BLOUNT *et al.* 2008). While such complex adaptations may be acquired through successive individually beneficial mutations (BRIDGHAM *et al.* 2006), it is likely that some adaptations require a combination of mutations that are each individually neutral or deleterious in the absence of the other mutations. Using the fitness landscape metaphor, populations must cross a "fitness valley" to acquire such adaptations. Valley crossing provides a means for populations to escape local "fitness peaks" or "plateaus" where no single mutation is individually beneficial. It may also be a common mode of adaptation even when individually beneficial mutations are available, especially in large populations (WEISSMAN *et al.* 2009), as is typical of microbes or Drosophila.

Understanding the dynamics of crossing fitness valleys is particularly important for understanding pathogen evolution. An adaptation that requires crossing a fitness valley should generally be more difficult for a population to acquire than an adaptation that can be reached by going straight "uphill." For a method of combating a pathogen to be effective, it must be difficult for the pathogen to adapt to it; this suggests that treatments should typically confront the pathogen with a fitness valley. Similarly, for an attenuated strain of a virus to be useful for a vaccine, it must be difficult for it to revert to virulence, so one would like the attenuated genotype to be separated from pathogenic genotypes by fitness valleys. But as we will see, in many parameter regimes fitness valleys can be crossed rapidly, so it is important to understand how wide and deep a valley must be for a particular large population to be highly unlikely to cross it.

Intuitively, recombination affects the valley-crossing process in two opposing ways. First, recombination can speed up valley crossing by bringing together separate neutral or deleterious mutations present in different individuals, perhaps producing a beneficial combination. Second, recombination can slow down valley crossing by breaking up beneficial combinations of mutations once they have been formed. In this article, we provide a quantitative description of the dynamics by which a sexual population crosses a fitness valley, focusing on the effect of recombination on the expected time this takes. We find that over a broad region of parameter space, a small amount of recombination speeds up valley crossing, while a large amount slows it down, often drastically, with the scale for the recombi-

[1]*Corresponding author:* Institute of Science and Technology Austria, Am Campus 1, A-3400 Klosterneuburg, Austria.   E-mail: dbw@ist.ac.at

nation rate at which the behavior changes set by the selective advantage obtained by crossing the valley.

In general, the dynamics of complex adaptation will depend on the population size, the mutation rates, the recombination rates between the different loci, the selective advantage provided by the adaptation, and the selective disadvantages of each of the intermediate genotypes. The population size will generally be very large and the mutation rates very small compared to unity, but both vary over many orders of magnitude between different populations and species. Recombination rates and selection coefficients typically range over many orders of magnitude among different sets of loci within one organism. Given this wide parameter space, the potentially subtle interactions between very large and very small parameters, and the difficulty of accurately measuring these underlying parameters in real populations, arriving at a detailed theoretical picture of all possible dynamics by simulations is neither practical nor particularly useful. Here we focus on the simplest possible model for understanding valley crossing with recombination: a population with a single possible adaptation requiring one mutation at each of two different loci, with both mutations individually neutral or deleterious. Even in this simple case, a complete description is quite complicated, and exact mathematical expressions for dynamics cannot, generally, be obtained. However, given the inaccuracies involved in applying our simple mathematical model to any biological system, the increased precision given by exact expressions is of limited usefulness. On the other hand, given the large range in possible parameter values mentioned above, approximate expressions giving the dependence of quantities of interest on the parameters can be very informative. We focus on developing an intuitive understanding of the major features of the dynamics, emphasizing aspects that are insensitive to the details of the model and that can serve as a basis for understanding more complicated situations. Detailed calculations are provided in APPENDIX B.

Our analysis of crossing simple fitness valleys provides a basis for understanding more complicated fitness valleys, as WEISSMAN *et al.* (2009) did for asexual populations. But it also serves as one step toward understanding how recombination and other population parameters act in general fitness landscapes to determine the rate and mode of adaptation. It is not known (even in simple models) under what circumstances adaptation proceeds primarily via the fixation of alleles that provide a selective advantage in a wide range of genetic backgrounds or when, instead, adaptation primarily involves formation and selection of particularly advantageous combinations of alleles. NEHER and SHRAIMAN (2009) recently analyzed the interplay between selection on alleles and selection on complicated combinations of alleles in populations with high diversity initially, but no new mutations. Their results

share with ours the feature that which type of adaptation dominates is controlled by the relative magnitudes of the selective advantage of a combination of alleles and the recombination rate among the loci involved.

The problem of the acquisition of an adaptation requiring individually deleterious mutations at two loci was analyzed in the limit of infinite population size by HALDANE (1931), CROW and KIMURA (1965), ESHEL and FELDMAN (1970), and KARLIN and MCGREGOR (1971), who found that a sufficiently high recombination rate made adaptation impossible. MICHALAKIS and SLATKIN (1996) considered several combinations of population parameters for finite populations and found that recombination generally slows valley crossing, while noting that recombination can have a positive effect if selection against single mutants is much weaker than selection for double mutants. CHRISTIANSEN *et al.* (1998) quantified the increase in the production of double mutants caused by recombination in the case of neutral single mutants, while for deleterious single mutants and weak recombination HADANY (2003) calculated the expected time for the population to cross the valley. In the limit of extremely weak recombination, the population becomes effectively asexual, a case that has been analyzed by KOMAROVA *et al.* (2003), IWASA *et al.* (2004), WEINREICH and CHAO (2005), DURRETT and SCHMIDT (2008), WEISSMAN *et al.* (2009), and LYNCH and ABEGG (2010). Simulations have usually found that recombination reduces the rate of valley crossing (*e.g.*, TAKAHATA 1982; KIM 2007), although some have found a slight increase for small amounts of recombination (TAKAHASI and TAJIMA 2005; WEINREICH and CHAO 2005; DE VISSER *et al.* 2009). In this article, we provide a quantitative framework tying together these previous results and describe the dynamics in new regions of parameter space. We also describe the parameter range for which the general analysis of the rate of transitions between local fitness peaks in BARTON and ROUHANI (1987) is valid and apply this analysis to our case. While we are primarily concerned with the rate of adaptive evolution, the problem of purely compensatory evolution (where the single mutants are deleterious but the double mutant has the same fitness as wild type) is closely related and has been analyzed by, *e.g.*, KIMURA (1985), STEPHAN (1996), and CARTER and WAGNER (2002). We do not consider the effect of environmental change on valley crossing; this has been investigated by, *e.g.*, MASEL (2006) and KIM (2007). While this article was being revised, another article (LYNCH 2010) examining similar questions was published; we compare those results with ours in the DISCUSSION.

MODEL

We consider a haploid population of N individuals with two diallelic loci with recombination rate r between them. Genotype ab is initially fixed in the pop-
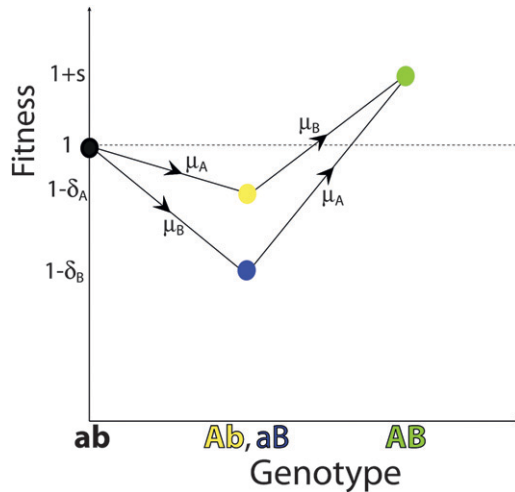
FIGURE 1.—The fitnesses of the four genotypes. The wild-type ab has fitness 1, and the double-mutant AB has fitness $1 + s > 1$. The single mutants Ab and aB are no more fit than wild type, with fitnesses $1 - \delta_A \leq 1$ and $1 - \delta_B \leq 1$, respectively. Allele a mutates to A at rate μ_A, and b mutates to B at rate μ_B.

ulation, with allele a mutating to A at rate μ_A and b mutating to B at rate μ_B. (We neglect back mutation, as it will have only a minor effect on the dynamics in the interesting regions of parameter space.) The four possible genotypes have fitnesses

$$w_{ab} = 1$$
$$w_{Ab} = 1 - \delta_A$$
$$w_{aB} = 1 - \delta_B$$
$$w_{AB} = 1 + s,$$

with $\delta_A, \delta_B \geq 0$ so that the mutant alleles are individually neutral or deleterious, and $1/N \ll s \ll 1$, so that the double mutant has a substantial but not enormous selective advantage over the wild type (Figure 1). For finite N, the fittest genotype AB will eventually dominate the population. We wish to find the distribution of the time $\mathcal{T}$ that it takes for this to occur, in particular the expected time $\tau \equiv E[\mathcal{T}]$. Except in very large populations, the time that it takes for AB to sweep to fixation in the population will be a negligible component of τ, which will be dominated by the waiting time before the selective sweep begins (WEISSMAN *et al.* 2009) We therefore ignore the sweep time and refer to the beginning of the sweep (when the double mutants establish and begin to increase in number exponentially) as the point at which the population has crossed the valley.

SUMMARY OF RESULTS

We find that the time τ for the population to acquire the adaptation is usually minimized at an optimum level of recombination $r \lesssim s$, which balances the two effects of recombination on valley crossing: bringing together
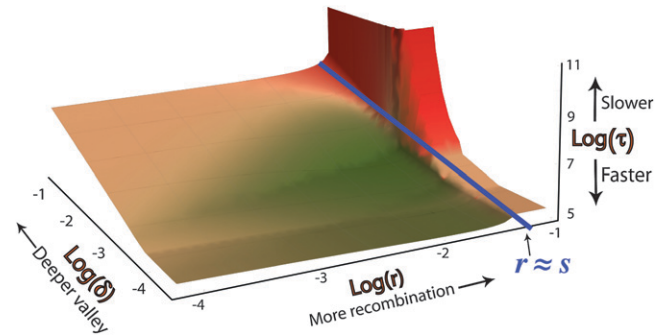

FIGURE 2.—Typical slice of parameter space, showing the expected time to cross the fitness valley, τ, as a function of the recombination rate r and the selective disadvantage of the single mutants δ. The other parameters are held constant at $N = 10^5$, $\mu = 5 \times 10^{-7}$, and $s = 0.05$. Color indicates the effect of recombination on the expected time: recombination significantly reduces τ in the green region [*i.e.*, $\log(\tau(r)/\tau_{r=0}) < 0$], significantly increases τ in the red region [$\log(\tau(r)/\tau_{r=0}) > 0$], and has only a minor effect on τ elsewhere. We see that small recombination rates $r < s$ can reduce τ, with this effect strongest at intermediate δ, while frequent recombination $r \gg s$ can drastically increase τ for large δ. The rapid transition between these regimes occurs at $r \approx s$. Note that increasing δ always increases τ, but can increase or decrease the ratio $\tau/\tau_{r=0}$. Also note that recombination tends to reduce τ for small values of δ and increase τ for large values of δ, increasing the dependence of τ on δ.

single mutants and breaking up double mutants. For higher recombination rates, $r \gg s$, recombination effectively breaks up the double-mutant combination; thus double mutants can increase in frequency only along with single mutants, slowing down the former's spread (Figure 2). In this case, if single mutants have a significant fitness disadvantage, then these mutants must drift against selection to the frequency at which they can begin to take over the population. The probability of this happening decreases exponentially as N increases, so τ increases exponentially: this means that large populations effectively cannot acquire adaptations that require individually deleterious mutations at unlinked genes. For less deleterious single mutants, mutation and drift are strong enough to allow the mutants to reach the necessary frequency relatively easily. In this regime, the two effects of recombination approximately cancel, and τ is close to its asexual value $\tau_{r=0}$, which decreases with increasing N—the opposite dependence.

For $r \ll s$, recombination is rare enough that it has only a small effect on the increase in double mutants once they are present in the population. If r is not too small, however, it can effectively bring together A and B alleles from single mutants and reduce the valley-crossing time. In the regime where this is true, τ decreases roughly as a power of r and is minimized at an intermediate recombination rate $r \lesssim s$, just below the value at which recombination starts effectively breaking up double mutants (Figures 3 and 4). Interestingly,
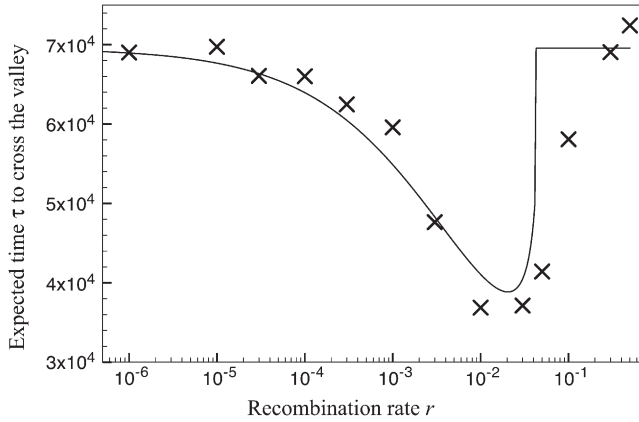
FIGURE 3.—Analytical approximation and simulation results for τ, the expected time in generations to cross the valley, as a function of the recombination rate r, for $N = 10^5$, $\mu = 5 \times 10^{-7}$, $\delta = 10^{-5}$, and $s = 0.05$, so that single mutants are effectively neutral. τ decreases with increasing r until it reaches a minimum at $r \approx s/2$, after which it quickly increases back to roughly its asexual value. The overall effect of recombination is weak in this regime, with τ varying only by a factor of $\approx 2 = \mathcal{O}((N^2 \mu s)^{1/6})$ over seven orders of magnitude of recombination rates.

the factor by which recombination reduces τ, $\tau_{r=0}/\tau(r)$, is also maximized at intermediate values of the other parameters, N, μ_A, μ_B, δ_A, and δ_B. In this region of parameter space, recombination can reduce τ by up to several orders of magnitude—a large effect, although small compared to the delay in valley crossing that can be caused by too frequent recombination $r \gg s$: see Figure 5. While frequent recombination can potentially slow down valley crossing far more than occasional recombination can speed it up, recombination may still generally increase the frequency of valley crossing in populations, because the deep valleys for which recombination has a negative effect are already far less likely to be crossed, even by asexual populations, than the shallow valleys where it has a positive effect (Figure 2).

While most of our analysis focuses on typical valley-crossing dynamics, we also briefly consider the probability that a population crosses a fitness valley unusually rapidly—a problem that is likely to be relevant to subdivided populations, in which the first subpopulation to cross the valley may then colonize the rest of the population. We find that the most likely dynamics by which populations cross fitness valleys change significantly as a function of time, with several different dynamics dominating at successively later times before the long-term dynamics dominate. The naive inference from the long-term rate of valley crossing can dramatically overestimate the early-time probability. However, even the early-time probability is typically maximized at an intermediate rate of recombination $r \lesssim s$.

Quantitative summary: Here we provide a list of some of the expressions for τ derived below and in APPENDIX B for readers who wish to skip the calculations but want a more quantitative summary than the one given above.
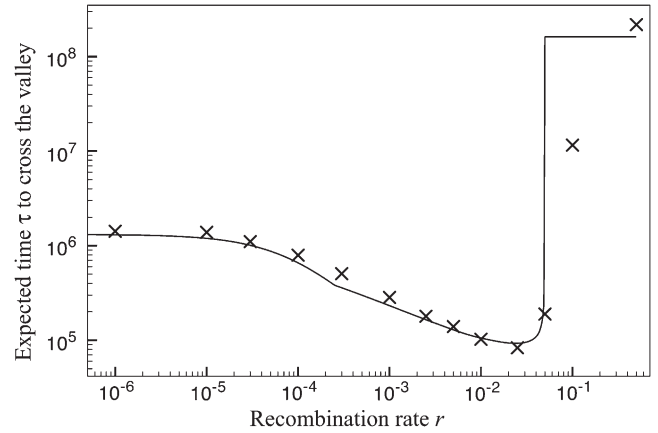

FIGURE 4.—Analytical approximation and simulation results for τ, the expected time in generations to cross the valley, as a function of the recombination rate r, for $N = 10^8$, $\mu = 2 \times 10^{-9}$, $\delta = 5 \times 10^{-5}$, and $s = 0.05$, so that selection against single mutants is strong enough to affect the dynamics, but not so strong to prevent the two single-mutant genotypes from occasionally arising together and recombining to produce double mutants. τ decreases with increasing r until it reaches a minimum at $r \approx s/2 \sim 20 = \mathcal{O}(\sqrt{N\delta})$ times lower than its asexual value, after which it quickly increases to a maximum $\sim 200 = \mathcal{O}(e^{N\delta^2/s})$ times higher than its asexual value.

We omit all numerical factors and leave most of the interpretation of the results for other sections of this article. For simplicity, we assume here and in the heuristic arguments below that the two loci are equivalent; i.e., $\delta_A = \delta_B \equiv \delta$, $\mu_A = \mu_B \equiv \mu$ (see APPENDIX C for the effect of relaxing this assumption).

First, in the asexual case $r = 0$ it has been shown that the rate of valley crossing, τ^{-1}, is given approximately by

$$\tau_{r=0}^{-1} \approx \begin{cases} N\sqrt{\mu^3 s} & \text{for } \delta \ll \sqrt{\mu s},\, N\mu \ll 1 \\ \mu\sqrt{Ns} & \text{for } \delta \ll \mu\sqrt{Ns},\, N\mu \gg 1 \\ \dfrac{N\mu^2 s}{\delta} & \text{for } \delta \gg \sqrt{(1 + N\mu)\mu s}. \end{cases} \quad (1)$$

Equation 1 is invalid for very small and very large N; see WEISSMAN *et al.* (2009) for a detailed discussion.

For frequent recombination, $r \gg s$, and single mutants no more than mildly deleterious, $\delta \ll \max\{\sqrt{\mu s},\ \sqrt{s/N}\}$, τ is given approximately by

$$\tau_{r>s} \approx \tau_{r=0}, \quad (2)$$

where $\tau_{r=0}$ is given by Equation 1. (See APPENDIX B for the derivation of Equation 2.) For substantially deleterious single mutants, $\max\{\sqrt{\mu s},\ \sqrt{s/N}\} \ll \delta \ll s$, τ is given approximately by

$$\log \tau_{r>s} \approx \frac{N\delta^2}{s}. \quad (3)$$

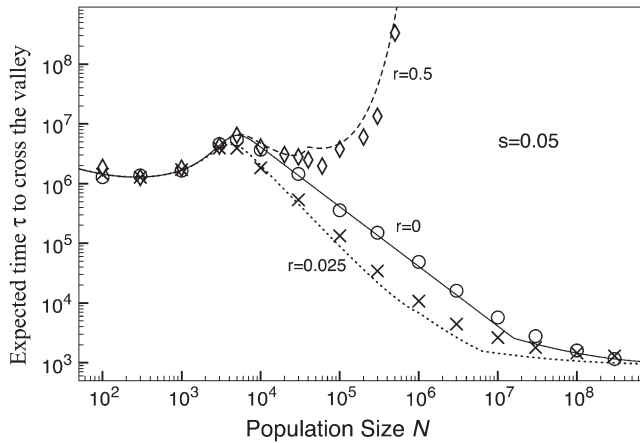(See the heuristic analysis below and BARTON and ROUHANI 1987.)

FIGURE 5.—Analytical approximations and simulation results for τ, the expected time in generations to cross the valley, as a function of the population size N, for $r = 0.5$ (frequent recombination), $r = 0.025$ (occasional recombination), and $r = 0$ (asexual), with the other parameters held constant at $\mu = 5 \times 10^{-7}$, $\delta = 10^{-3}$, and $s = 0.05$. Lines show analytical approximations (dashed for $r = 0.5$, dotted for $r = 0.025$, and solid for $r = 0$). Diamonds, crosses, and circles show simulation results for $r = 0.5$, $r = 0.025$, and $r = 0$, respectively. Note that when τ depends on the recombination rate, it is always lowest for occasional recombination and highest for frequent recombination, with the value for asexuality lying in between.

For infrequent recombination, $r \ll s$, and $N\mu \ll 1$ (single-mutant lineages arising infrequently), the rate of valley crossing is given approximately by

$$
\tau_{r<s}^{-1} \approx
\begin{cases}
\tau_{r=0}^{-1} & \text{for } r \ll \max\left\{2\delta, \min\left\{\frac{1}{N}\sqrt{\frac{s}{\mu}}, \frac{s}{N\delta}\right\}\right\} \\[2mm]
(N^4\mu^5 rs)^{1/3} & \text{for } \max\{\sqrt{\mu s}, \delta\} \ll (N\mu^2 rs)^{1/3} \\[2mm]
\mu^2\sqrt{\frac{N^3 rs}{\delta}} & \text{for } \max\left\{\frac{s}{Nr}, (N\mu^2 rs)^{1/3}\right\} \ll \delta \ll \left(\frac{rs}{N}\right)^{1/3} \\[2mm]
\frac{N\mu^2 rs}{\delta^2} & \text{for } \left(\frac{rs}{N}\right)^{1/3} \ll \delta \ll \frac{r}{2}.
\end{cases}
\tag{4}
$$

(See the heuristic analysis below and APPENDIX B.) Equation 4 remains valid for intermediate recombination rates $r \lesssim s$ if we replace s with $\tilde{s} \equiv s - r$. Note that τ generally decreases with increasing $r\tilde{s}$, meaning that a recombination rate $r \approx s/2$ maximizes the rate of valley crossing.

For $r \ll s$ and $N\mu \gg 1$ (single-mutant lineages arising frequently), τ is given by Equation B27 in APPENDIX B.

The boundaries of the parameter regimes in Equations 1–4 are somewhat complicated; see Figure B1 in APPENDIX B for an illustration of a typical slice of parameter space. Note that when $r \gg s$ and single mutants are substantially deleterious, τ grows exponentially with increasing N—in this parameter regime, the population relies on drift to cross the fitness valley. However, in all other parameter regimes, τ decreases with increasing N.

We now turn to the derivation and explanation of these results.

HEURISTIC ANALYSIS

For much of parameter space, the valley-crossing dynamics can be analyzed at least semideterministically, with the numbers of single mutants n_{Ab} and n_{aB} approximated by their expected values and only the double mutants treated stochastically (see, e.g., CHRISTIANSEN et al. 1998). However, in other parameter regimes, the stochastic fluctuations in the numbers of single mutants are crucial to the dynamics; these include the parameter regimes where recombination has the largest effect. Here we give a heuristic analysis of the effect of recombination on the time for the population to cross the fitness valley in two of these regimes where the semideterministic approximation fails. First, we consider the regime where recombination slows down valley crossing the most: $r \gg s$ and $\delta^2/s \gg \max\{\mu, 1/N\}$, where the mutants must drift far above mutation–selection balance before they begin to be favored by selection. Second, we consider the regime $r \ll s$ with $N\mu \ll 1$, where recombination can speed up valley crossing. Together, these two regimes display all the interesting qualitative behavior of the system.

Our goal in this section is to provide an intuitive understanding of the dynamics underlying the results summarized above, focusing only on the most important features and the regimes in which stochastic fluctuations in the number of single mutants are important. More careful derivations are unfortunately quite long and technical and are reserved for APPENDIX B. While we explicitly analyze only a limited subsection of parameter space in this section, this should provide the tools for similar heuristic analyses of the rest of parameter space. Indeed, we hope that these kinds of heuristic arguments can be used broadly for understanding evolutionary processes that depend on rare mutant lineages.

Frequent recombination, $r \gg s$: First we consider a high rate of recombination, $r \gg s, \delta$. In this case, recombination is frequent enough to keep the population close to linkage equilibrium, meaning that the double mutants will not begin to spread deterministically in the population until the mean fitness of all mutants is significantly greater than that of the wild type. When mutants are rare, this condition on the mean fitness of the mutants can be expressed in terms of the allele frequencies x_A, x_B as

$$
(s + 2\delta)x_A x_B - \delta(x_A + x_B) > \frac{1}{N}.
\tag{5}
$$

The left-hand side of this inequality is the mean selective advantage of a mutant individual over wild type assuming linkage equilibrium, while the right is the strength of drift; when the inequality is satisfied, selection is likely to overcome drift and drive the mutant alleles to fixation.

For the double mutants to spread deterministically while the single mutants are still at low frequencies, condition (5) must have a solution with x_A, $x_B \ll 1$. This

is possible only if $\delta \gg s$; we focus on this case here. When selection against single mutants is sufficiently strong compared to mutation and drift (specifically, when $\delta^2/s \gg \max\{\mu, 1/N\}$—see APPENDIX B for the derivation of this condition), the rate at which the single mutants reach frequencies satisfying the condition (5) is strongly reduced. In the limit of infinite population size, this rate goes to zero: there are two deterministic equilibria, one with ab as the dominant genotype and the other with AB as the dominant genotype, and the population will remain at the first equilibrium (KARLIN and MCGREGOR 1971). The boundary between the two basins of attraction is given by the curve for which the left-hand side of this inequality (5) is zero, with an unstable fixed point at $x_A = x_B \approx \delta/s$.

If the population is large ($N \gg \delta^2/s$) but finite, the analysis of BARTON and ROUHANI (1987) can be applied to find the rate of valley crossing. Because this analysis is quite general and involved, we provide a brief heuristic summary before quoting the result. For finite N, there will be an equilibrium probability distribution for the allele frequencies, $\Psi(x_A, x_B)$. Assuming linkage equilibrium and using a diffusion approximation, $\Psi(x_A, x_B)$ is given by

$$\Psi(x_A, x_B) \sim \exp[-N(\delta(x_A + x_B) - sx_Ax_B)] \quad (6)$$

for $1/N \ll x_A, x_B \ll 1$, with a prefactor that is changing slowly compared to the exponent near the saddle point (KIMURA 1964); see Figure 6. The infinite-population deterministic equilibria correspond to peaks in Ψ, and the unstable fixed point at $x_A = x_B \approx \delta/s$ corresponds to a saddle point. The lower peak at $x_A, x_B \approx 0$, the saddle point, and the higher peak at $x_A, x_B \approx 1$ are joined by a ridge in Ψ running along $x_A = x_B$. For $\delta \ll s$, the most likely way for the population to cross from the lower peak to the higher peak is to drift along the ridge to the saddle, after which selection will drive the mutant alleles to fixation. Since Ψ falls off rapidly away from the ridge, we can treat the population as being effectively one-dimensional, with a single two-allele locus subject to disruptive frequency-dependent selection. The largest factor in τ comes from the tiny probability of reaching the saddle point from the initial peak. Since the probability of being near the saddle point is lower than the probability of being near the wild-type-dominated peak by a factor of $\sim \exp(-N\delta^2/s)$, the time to cross the valley is exponentially large,

$$\log \tau \approx \frac{N\delta^2}{s}, \quad (7)$$

with additional smaller terms that depend on the mutation rate, the rate of drift, higher-order terms in the location of the saddle point, and linkage disequilibrium. Using the calculations leading to Equation 17 in BARTON and ROUHANI (1987), we can find the more accurate expression
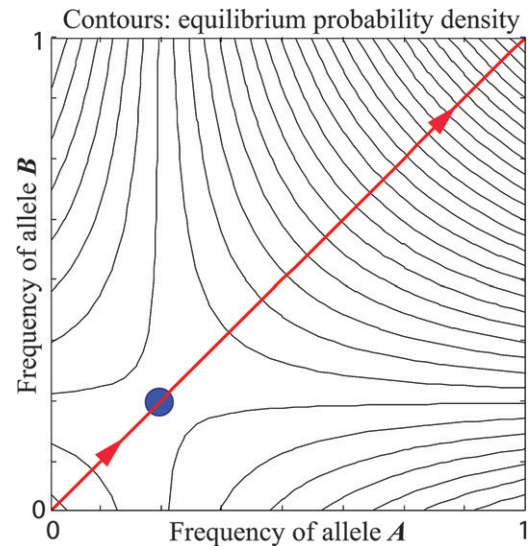

FIGURE 6.—Schematic contour plot of equilibrium allele frequency probability distribution for a large population with deleterious single mutants and frequent recombination (Equation 6). The red curve shows the most likely path through frequency space from the initial wild-type-dominated state to the final mutant-dominated state. This path passes through the saddle point, shown in blue, separating the two deterministic equilibria. The probability of being near the saddle point at $x_A \approx \delta_B/s$, $x_B \approx \delta_A/s$ is lower than the probability of being near the initial peak by a factor $\sim \exp(-N\delta_A\delta_B/s)$, producing a factor in the expected time τ to cross the valley of $\sim \exp(N\delta_A\delta_B/s)$.

$$\tau \approx \frac{\Gamma(N\mu)^2}{\delta}\left(\frac{N\delta^2}{s+2\delta}\right)^{1-2N\mu}\exp\left[\frac{N\delta^2}{s+2\delta}\right], \quad (8)$$

where Γ is the gamma function. For large populations, Equation 7 is an enormously long time; they essentially cannot cross the valley.

Equations 7 and 8 are valid for $r \gg s \gg \delta \gg \max\{\sqrt{\mu s}, \sqrt{s/N}\}$. The case of infrequent recombination, in which the first inequality is violated, is discussed below; the case of shallow valleys, in which the third inequality is violated, is discussed in APPENDIX B. As mentioned above, for deeper valleys, $\delta \gg s$ (but $r \gg \delta$), the condition (5) is satisfied only for x_A, x_B close to 1—before the single mutants can begin to increase deterministically in frequency, at least one must drift close to fixation. The expected time for this to happen is very long, with $\log \tau \approx N\delta$ (KIMURA 1962). Note that this matches smoothly with Equation 7 at the boundary of the two approximations, $\delta \approx s$. Finally, for very deep valleys, $\delta \gg r$, selection against single mutants is strong enough to keep the population in linkage disequilibrium. In this case, we expect that the most likely way for the population to cross the valley is for an AB lineage to drift close to fixation while undergoing unusually little recombination. The expected time for this to happen is extremely long, with $\log \tau \approx Nr$.

Infrequent recombination, $r \ll s$: Now we consider infrequent recombination, $r \ll s$. We focus on calculating the rate of valley crossing in populations large

enough that they are likely to cross via "stochastic tunneling" (KOMAROVA *et al.* 2003), in which double mutants take over the population without single mutants ever reaching high frequencies. For a double-mutant lineage to take over the population, it must avoid going extinct due to drift or being broken up by recombination. This occurs with probability $\approx s - r \equiv \tilde{s}$ (HADANY 2003). We refer to a mutant individual as *successful* if its descendants include double mutants that eventually take over the population; we wish to find the waiting time for the first such successful single mutant.

If the population is large enough that new mutants are continually being produced, $N\mu \gg 1$, the dynamics of the single mutants can be treated approximately deterministically; this case is discussed in APPENDIX B. Here we focus on low population mutation rates, $N\mu \ll 1$. In this case, the total waiting time $\mathcal{T}$ is generally dominated by the waiting time for the production of the first successful single mutant, with the first successful double mutant being produced shortly afterward. Since single-mutant lineages are produced at rate $2N\mu$, $\mathcal{T}$ will have an approximately exponential distribution with expected time $\tau \approx (2N\mu p)^{-1}$, where p is the probability that a single-mutant lineage will be successful (KOMAROVA 2006; WEISSMAN *et al.* 2009). To find τ, it thus suffices to find p.

Lineage dynamics: A single-mutant lineage must be lucky to be successful, but this luck can come in several different places: for instance, the lineage could luckily drift to such an unusually large size that it becomes likely to produce so many double mutants that one will be successful, or it could drift to only a small size and produce one double mutant that luckily is successful. To distinguish between these cases, we can decompose p into the probability prob(T) that a lineage drifts for T generations before going extinct and the probability $P|_T$ that it will be successful given that it does so, integrated over all possible drift times T:

$$p \approx \int P|_T \, \text{prob}(T) \, dT. \tag{9}$$

To evaluate Equation 9, we need to understand the typical dynamics of a single-mutant lineage. Here we summarize the aspects of the dynamics that are important for calculating p; slightly more detailed expressions are given in APPENDIX A. Let the number of individuals in the lineage at time t be given by $n(t)$, with $n(0) = 1$. For $t \ll \min\{N, 1/\delta\}$, the lineage is effectively neutral and unlikely to reach a high frequency. For the rest of this section, we assume that N is sufficiently large that all single mutants remain at low frequency. In this case, for $t \ll 1/\delta$ (and $\gg 1$), the probability that the lineage has avoided extinction by time t is approximately proportional to t^{-1}, with the constant of proportionality depending on the strength of drift. This can be seen by modeling the dynamics of the lineage by a nearly critical branching process (as in APPENDIX A for continuous time or in ATHREYA and NEY 1972, pp. 19ff, for discrete time) or by using a diffusion approximation (EWENS 2004, p. 162). We use the symbol "$\sim$" to denote approximate proportionality up to a model-dependent constant. (In APPENDIX B, we use a continuous-time Moran model chosen such that this constant is 1 for most expressions of interest.)

Since the expected number of individuals in the lineage $E[n(t)] = \exp(-\delta t)$ remains ≈ 1 until $t = \mathcal{O}(1/\delta)$, the typical size of the lucky lineages that avoid extinction for a time $t \gg 1$ is $E[n(t)|n(t) > 0] = E[n(t)]/\text{Prob}(n(t) > 0) \sim t$. At longer times, $t \gtrsim 1/\delta$, the probability that the lineage has avoided extinction decreases like $\exp(-\delta t)$, as does $E[n(t)]$: selection makes it very unlikely that lineages will persist for much longer than $\sim 1/\delta$ generations, and the few that do last for a long time typically do not grow past a characteristic size of $n \sim 1/\delta$. Thus, the distribution of lineage drift times T (and sizes n) has a long tail until it is cut off by selection at T, $n \sim 1/\delta$. This suggests for small δ, rare large lineages may play an important role in the population dynamics. Note that these basic features of the lineage dynamics are the most important for determining the rate of valley crossing; we expect our results to be insensitive to changes in the model details.

While T is broadly distributed over $t < 1/\delta$, the distributions of the other random variables that play a role in valley-crossing dynamics are for the most part relatively narrow. In particular, besides the distribution of T for $t > 1/\delta$, the distributions of $n(t)$, the number of mutants produced by a lineage with a given trajectory, the number of mutants produced in a given time interval, and the number of recombinations between two lineages with given trajectories all have exponential tails. Thus, as a good first approximation, we can generally treat the drift times of the mutant lineages as the only random variables, with all other variables having fixed values given the drift times, and treat even the drift times as being limited to values $< 1/\delta$. The exponentially unlikely valley-crossing dynamics that violate this approximation produce only small rates of valley crossing, as seen in the case of frequent recombination and deep valleys discussed above.

Given these lineage dynamics, we can now find an approximate expression for $P|_T$, the probability that a lineage that lasts for T generations will be successful. We can write $P|_T$ as the sum of two terms: $P_{as}|_T$, the probability of success via the asexual path, in which a single-mutant individual acquires an additional mutation, and $P_r|_T$, the probability of success via the sexual path, in which an Ab individual recombines with an aB individual. In the next section, we derive approximate expressions for $P_{as}|_T$ and $P_r|_T$, assuming for simplicity that single mutants are neutral, $\delta = 0$. (We discuss the effect of larger δ later.) Recall that in this case we expect rare large lineages to be important.

Neutral single mutants—asexual path across the valley: First consider the probability $P_{as}|_T$ of a single-mutant lineage being successful via the asexual path given that it drifts for at least T generations. For a detailed discussion of this case, see WEISSMAN *et al.* (2009); here we provide just a brief summary to introduce the heuristic arguments that we use in the more complicated sexual path below. To be successful, the single-mutant lineage must produce a successful double mutant. Each double mutant has a probability $\tilde{s} \approx s - r$ of being successful, so if the single-mutant lineage gives rise to n_2 double mutants, it will be successful with probability $P_{as}|_T \approx n_2 \tilde{s}$, until this saturates at 1 for $n_2 \approx 1/\tilde{s}$—lineages that give rise to $\gtrsim 1/\tilde{s}$ double mutants are likely to be successful.

Thus, to find $P_{as}|_T$, we need to find the typical number n_2 of double mutants that the lineage will produce. Each individual in the lineage has a probability μ per generation of producing a double mutant, so the total number produced over T generations will typically be $n_2 \approx \mu \int n(t) dt$. To evaluate this integral, we must use the fact, mentioned above, that the size of the lineage will typically grow like $n(t) \sim t$; doing so, we find that the typical number of double mutants produced in T generations is $n_2 \sim T^2 \mu$. Multiplying by $\tilde{s}$ to get the total probability of success gives an approximate expression for $P_{as}|_T$:

$$P_{as}|_T \sim \begin{cases} T^2 \mu \tilde{s} & \text{for } T \ll \frac{1}{\sqrt{\mu \tilde{s}}} \\ 1 & \text{for } T \gg \frac{1}{\sqrt{\mu \tilde{s}}}. \end{cases} \quad (10)$$

Since $P_{as}|_T$ grows like T^2 and the probability of lasting for T generations only falls off like T^{-1}, the asexual valley-crossing dynamics are dominated by the rare lineages that grow to a large size over a long time; the vast majority of lineages that reach only much smaller sizes make little contribution to the total rate of valley crossing. In other words, the luck needed to cross the valley is likely to be concentrated in a single very large lineage rather than spread out over multiple moderately sized lineages. (When single mutants are strongly deleterious, the opposite is true; see WEISSMAN *et al.* 2009.)

Neutral single-mutants—sexual path across the valley: We now turn to finding $P_r|_T$, the probability that a lineage that drifts for at least T generations is successful via the sexual path. To do so, we must consider the lineages with complementary genotypes that will coexist with the focal lineage. During the T generations for which the focal lineage (of genotype, say, Ab) drifts, $\approx N\mu T$ aB lineages typically arise. The probability of being successful therefore grows like $p_r|_T \approx N\mu T p'_T$, where p'_T denotes the probability that a single aB lineage will be successful by recombining with the focal lineage.

As in Equation 9, we can decompose the probability p'_T that an aB lineage will be successful into the probability that it avoids extinction for T' generations

and the probability $P_r|_{T,T'}$ that it is successful given that it does so: $p'_T \approx \int P_r|_{T,T'} \text{prob}(T') dT'$. The total probability that the focal Ab lineage will be successful therefore grows like

$$P_r|_T \approx N\mu T \int P_r|_{T,T'} \text{prob}(T') dT' \quad (11)$$

until saturating at 1 for sufficiently large T.

We already know the distribution of the lineage sizes T', so to evaluate Equation 11 it remains to find $P_r|_{T,T'}$. The number of AB individuals produced by the pair of the lineages via recombination is typically $\approx (r/N) \int n_{Ab}(t) n_{aB}(t) dt$. We know that $n_{Ab} \sim T$ and $n_{aB} \sim T'$; assuming $T' \le T$, the two lineages coexist for T' generations. (It is easy to check that $T' > T$ makes only a small contribution to Equation 11.) Evaluating the integral, we find that the two lineages typically recombine to produce $\sim (r/N) TT'^2$ AB individuals. Thus $P_r|_{T,T'}$ grows like $P_r|_{T,T'} \sim (r\tilde{s}/N) TT'^2$ until it approaches 1 at $T' \sim \sqrt{N/Tr\tilde{s}}$. Using this to evaluate Equation 11, we find that the probability of success is

$$P_r|_T \sim \begin{cases} T^3 \mu r \tilde{s} & \text{for } T \ll \left(\frac{N}{r\tilde{s}}\right)^{1/3} \\ T^{3/2} \mu \sqrt{Nr\tilde{s}} & \text{for } \left(\frac{N}{r\tilde{s}}\right)^{1/3} \ll T \ll (N\mu^2 r\tilde{s})^{-1/3} \\ 1 & \text{for } T \gg (N\mu^2 r\tilde{s})^{-1/3}. \end{cases}$$

$$(12)$$

For small Ab lineages ($T \ll (N/r\tilde{s})^{1/3}$), the most likely path to being successful requires an aB lineage to arise at roughly the same time and grow to roughly the same size. Larger Ab lineages, however, are most likely to be successful by recombining with an aB lineage much smaller than themselves, because $P_r|_{T,T'}$ saturates at 1 for $T' \ll T$. Because $P_r|_T$ grows faster than linearly with T before saturating, the sexual valley-crossing dynamics, like the asexual dynamics, are dominated by the rare single-mutant lineages that grow so large that they are then likely to be successful [$T \sim (N\mu^2 r\tilde{s})^{-1/3}$]. Note that this means that in the most likely sexual valley-crossing path, the two single-mutant lineages are likely to be of very different sizes, even when the two loci are symmetric. (See Figure 7 for a typical valley-crossing trajectory.)

Neutral single mutants—advantage of recombination: We can now compare Equations 10 and 12 to find whether the population is more likely to cross the valley via the asexual or the sexual path. Since both the asexual and the sexual valley-crossing dynamics are dominated by rare large single-mutant lineages, this is equivalent to asking whether the size at which a lineage is likely to be successful is smaller for the asexual or the sexual path. Success via an additional mutation becomes likely at $T \sim 1/\sqrt{\mu \tilde{s}}$, while success via recombination becomes likely at $T \sim 1/(N\mu^2 r\tilde{s})^{1/3}$.
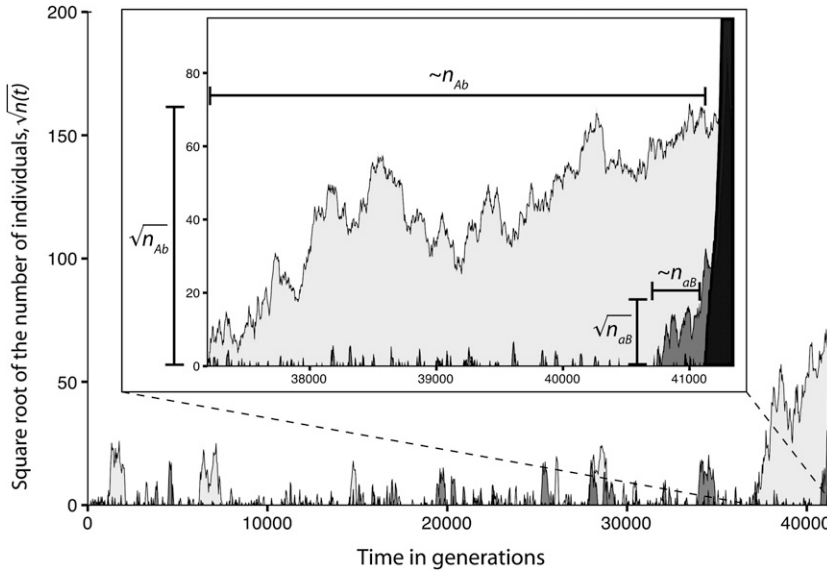
FIGURE 7.—Typical simulated population dynamics in the "stochastic tunneling" regime, with neutral single mutants, $N\mu \ll 1$, and $\sqrt{s/\mu}/N \ll r \ll s$. Light shading shows $\sqrt{n_{Ab}(t)}$, dark shading shows $\sqrt{n_{aB}(t)}$, and solid shading shows $\sqrt{n_{AB}(t)}$. [Square roots are shown to make low-frequency lineages more visible and because this is the natural scale for measuring the variance of the frequency of rare mutants (CAVALLI-SFORZA and EDWARDS, 1967).] The inset is a magnified view of the last few thousand generations before AB takes over the population, starting from the birth of the first successful single mutant. We see that the time to cross the valley is dominated by the waiting time for an unusually large single-mutant lineage that drifts to size n_{Ab} over $\sim n_{Ab}$ generations. While this lineage drifts, many small aB and AB lineages arise and go extinct. The largest aB lineage drifts to a size $n_{aB} \sim N\mu n_{Ab}$ over $\sim n_{aB}$ generations, recombining many times with the Ab lineage and producing the successful AB lineage.

Therefore, the asexual path dominates for $1/\sqrt{\mu\tilde{s}} \ll 1/(N\mu^2 r\tilde{s})^{1/3}$ (i.e., $Nr\sqrt{\mu/\tilde{s}} \ll 1$), while for $Nr\sqrt{\mu/\tilde{s}} \gg 1$, the sexual path dominates. In these two extreme cases, Equation 9 for the total probability p that a lineage is successful simplifies to

$$p \sim \begin{cases} \sqrt{\mu\tilde{s}} & \text{for } Nr\sqrt{\mu/\tilde{s}} \ll 1 \text{ (effectively asexual)} \\ (N\mu^2 r\tilde{s})^{1/3} & \text{for } Nr\sqrt{\mu/\tilde{s}} \gg 1 \text{ (recombination dominates).} \end{cases}$$
$$(13)$$

The characteristic time τ for valley crossing is therefore given by

$$\tau \sim \begin{cases} (2N\sqrt{\mu^3\tilde{s}})^{-1} & \text{for } Nr\sqrt{\mu/\tilde{s}} \ll 1 \text{ (effectively asexual)} \\ \frac{1}{2}(N^4\mu^5 r\tilde{s})^{-1/3} & \text{for } Nr\sqrt{\mu/\tilde{s}} \gg 1 \text{ (recombination dominates),} \end{cases}$$
$$(14)$$

corresponding to the first two lines of Equation 4.

Comparing the two expressions for τ in Equation 14, we see that an intermediate level of recombination, $(1/N)\sqrt{s/\mu} \ll r \ll s$, reduces the time to cross the valley by a factor of $\sim (N^2\mu r^2/s)^{1/6}$. However, because of the small exponent, and the fact that $N\mu, r/s \ll 1$ by assumption, this is generally not a big effect (see Figure 3). Moderate recombination thus typically provides a small increase in the rate of crossing very shallow valleys. (This is also true for $N\mu \gtrsim 1$; see Figure 5 and APPENDIX B.)

Deleterious single mutants: Intuitively, small amounts of recombination reduce the time for valley crossing by reducing the size to which a single-mutant lineage must drift to be likely to be successful. Because the lineage size distribution has a long tail for neutral single mutants, this is only a small effect. However, if single mutants are deleterious, $\delta > 0$, then the long tail is cut off at $T \sim 1/\delta$, and recombination can have a larger effect. In particular, if $\sqrt{\mu\tilde{s}} \ll \delta \ll (N\mu^2 r\tilde{s})^{1/3}$, then selection against single mutants has little effect on the rate of valley crossing via recombination, but slows down the asexual route significantly.

For even larger values of δ, both the asexual and the sexual paths will be affected. This case can be approximately analyzed by simply recalculating the expressions for $P_{as}|_T$ and $P_r|_T$ from the previous section using the approximation $\text{prob}(T > 1/\delta) \approx 0$. The results are somewhat complicated and are given in Equation 4 and in Equations B21 and B23 in APPENDIX B, where they are derived by a more careful calculation. The most important point to note is that for $\delta \ll \max\{\sqrt{\mu\tilde{s}}, (N\mu^2 r\tilde{s})^{1/3}\}$, selection against single mutants does not significantly affect the rate of valley crossing (i.e., τ is given by Equation 14). Thus, in large populations, many mutations that are strongly selected against ($N\delta \gg 1$) may still be effectively neutral for the purposes of valley crossing.

It is also interesting to note that while small values of δ affect the asexual path across the valley more strongly than the sexual path, the opposite is true for large values of δ. This is because the sexual path requires two single-mutant lineages; if δ is large enough to affect both lineages, then selection will impede the sexual path more than the asexual path (which requires only one single-mutant lineage). In fact, for $\delta > s/2$, $P_r|_T$ becomes negligible compared to $P_{as}|_T$, and any amount of recombination slows down valley crossing by reducing $\tilde{s}$. At the intermediate values of δ where recombination provides the greatest benefit, it reduces τ by a factor of $\sqrt{N\delta r/s}$: for $r \approx s/2$, this factor is $\sim \sqrt{N\delta}$, meaning that recombination can increase the rate of valley crossing by as much as several orders of magnitude (see Figures 2 and 4).

Distribution of valley-crossing times: We have focused on the expectation τ of the time for the

population to cross the valley. The full distribution of the valley-crossing time $\mathcal{T}$ is close to exponential in the regimes where it is typically dominated by the waiting time for the first successful single mutant (because mutations are rare, $N\mu \gg 1$, or because each mutant is very unlikely to be successful, $s \ll r, N\delta^2$), so it is given by

$$P(\mathcal{T} < t) \approx 1 - \exp\left(-\frac{t}{\tau}\right). \tag{15}$$

This is valid, however, only for t large compared to the typical time that a successful single-mutant lineage drifts before the double mutants take over. This drift time makes only a small contribution to τ, but it reduces the probability that $\mathcal{T}$ will be very small: even if the first successful single mutant is produced extremely early, it will still usually take time to generate the first successful double mutant and for the double mutants to spread. It is worth understanding this effect, because the left tail of the τ-distribution can be important for some processes. For example, in a metapopulation made up of many demes (say, for a pathogen or attenuated vaccine strain colonizing a host population), it may be that the lucky deme that crosses the valley first will colonize the other demes; the typical time for this to occur depends on the left tail of $P(\mathcal{T})$ rather than τ. The asexual case $r = 0$ is also relevant for understanding the incidence of cancers that typically occur only after a series of mutations (see popular accounts by Wodarz and Komarova 2005, Nowak 2006, and Frank 2007): the rare unlucky individuals that develop any given form of cancer are likely to have had cells that acquired multiple mutations anomalously fast.

As an example, we now consider the case of infrequent recombination ($r \ll s$) and rare mutation ($N\mu \ll 1$) described above. We saw that the successful single-mutant lineage typically drifted for a time $T \sim (N\mu^2 rs)^{-1/3}$, so Equation 15 is invalid for $t < (N\mu^2 rs)^{-1/3}$. We can, however, perform a similar heuristic calculation to derive $P(\mathcal{T})$ in this regime. We focus on calculating the probability that the successful double-mutant lineage has been produced by time t. We expect that this probability will be dominated by the probability that a lineage will be produced very early (with probability $\sim N\mu$) and will avoid extinction for $T \sim t$ generations and grow to a size $n \sim t$, which occurs with probability $\sim t^{-1}$. The lineage will then be successful with probability $P|_t \approx P_{as}|_t + P_r|_t$ (where $P_{as}|_t$ and $P_r|_t$ are given by Equations 10 and 12 with $T = t$), so the rate of valley crossing at time t will be $\sim N\mu t^{-1} P|_t$, and the cumulative probability of having produced a successful double mutant by time t will grow like $P(t) \sim \int_0^t N\mu t'^{-1} P|_{t'} dt'$. (Here we are ignoring numerical factors of order 1, as we will throughout the rest of this section.)

Plugging in the values from Equations 10 and 12 to get $P|_t$, we can find the complete $\mathcal{T}$ distribution. In the

regime $Nr\sqrt{\mu/\hat{s}} \gg 1$ where the sexual path dominates at long times, we have

$$P(\mathcal{T} < t)$$
$$\sim \begin{cases} N\mu^2 \tilde{s}t^2 & \text{for } t \ll \frac{1}{r} \text{ (asexual route dominates)} \\ N\mu^2 r\tilde{s}t^3 & \text{for } \frac{1}{r} \ll t \ll \left(\frac{N}{r\tilde{s}}\right)^{1/3} \\ \mu^2\sqrt{N^3 r\tilde{s}}t^3 & \text{for } \left(\frac{N}{r\tilde{s}}\right)^{1/3} \ll t \ll (N\mu^2 r\hat{s})^{-1/3} \\ 1 - \exp[-(N^4\mu^5 r\hat{s})^{1/3}t] & \text{for } (N\mu^2 r\hat{s})^{-1/3} \ll t \end{cases} \tag{16}$$

(see Figure 8). We see that for very short times, the population is more likely to cross the valley by the asexual route, with the sexual route becoming more likely at a time $t = \mathcal{O}(1/r)$ when it becomes likely that individuals will have undergone recombination.

Note that it is straightforward to use this approach in combination with the analysis of Weissman *et al.* (2009) to find the complete $\mathcal{T}$ distribution for an asexual population crossing a fitness valley of arbitrary width K. Doing so, one finds that for shallow valleys $P(\mathcal{T} < t)$ generally grows like t^K for small t and then like t^{K-1}, $t^{K-2}, \ldots, t$ before exponentially approaching 1. This suggests that we should expect cancer incidence curves to generically show multiple different power-law behaviors as a function of age. Note the contrast between this result and the prediction of, *e.g.*, Frank (2007) or Schweinsberg (2008) that the incidence should grow like a single power of t at early times. (This result also differs from those of Schweinsberg 2008 in that we find the same early time dependence for populations that have different long-time exponential behavior.) Our results show that the use of time-dependent incidence data to infer the number of mutations responsible for a cancer is seriously problematic. This is especially so in the absence of much information about relevant mutation rates, numbers of potentially oncogenic mutations, and fitness effects of intermediaries.

SIMULATIONS AND NUMERICAL RESULTS

In addition to the analytical work described above, we also performed stochastic simulations, similar to those used in Weissman *et al.* (2009) (see appendix d for simulation methods). The results of these simulations are generally in good agreement with our analytical results predictions, although there are significant differences in the intermediate parameter regimes, particularly for r just slightly larger than s, where the approximations made in the analysis are invalid; see Figures 3, 4, and 5.

In Figures 3 and 4, we show our analytical predictions and simulation results for the expected valley-crossing time τ as a function of the recombination rate r, with the lowest recombination rate $r = 10^{-6}$ corresponding to, *e.g.*, a distance of $\sim$100 bp in Drosophila (Hilliker *et al.* 1994; Andolfatto and Wall 2003) and the highest recombination rate $r = 0.5$ corresponding to
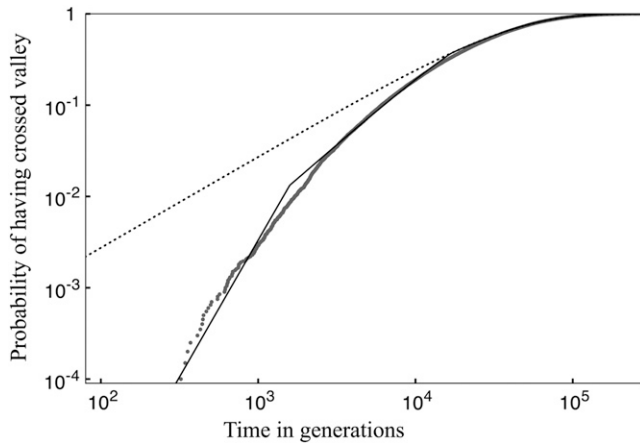
FIGURE 8.—Analytical approximations and simulation results for the full distribution of $\mathcal{T}$, the time in generations to cross the valley, with parameters $N = 10^5$, $\mu = 5 \times 10^{-7}$, $r = 0.01$, $\delta = 10^{-5}$, and $s = 0.05$. The dotted curve shows the exponential approximation Equation 15, the solid curve shows the more accurate approximation Equation 16 (with numerical factors included), and the circles show the results of simulations. For these parameters, the expected time to cross the valley is $\tau \approx 4 \times 10^4$ generations. We see that for times short compared to τ, Equation 15 overestimates the probability of having crossed the valley, while Equation 16 is accurate for all times.

unlinked genes. In Figure 3, the other parameters are chosen such that $\sqrt{\mu s} \gg \delta = 1/N \gg \mu$, so that single mutants are effectively neutral (even though they would normally be described as being weakly deleterious) and we are in the stochastic tunneling regime, described above for $r \ll s$. We see that a small amount of recombination indeed reduces τ compared to its asexual value, but only slightly: at the optimal recombination rate, $r \approx s/2$, τ is reduced only by a factor of $\approx (N^2 \mu s/2)^{1/6} \approx 2$. We also see that at large values of the recombination, τ returns to approximately its asexual value.

In Figure 4, the fixed parameters are chosen such that $\delta^2 > \sqrt{\mu s}$, s/N and $N\mu < 1$, so that for $r < s$ we are in the stochastic tunneling regime with deleterious single mutants, and for $r \gg s$ we are in the exponentially suppressed regime discussed above. We see that recombination has a much stronger effect on the valley-crossing time for deleterious single mutants than it does when single mutants are neutral, with τ reduced by a factor of $\approx 20 \sim \sqrt{N\delta}$ from its asexual value at the optimum recombination rate and increased by a factor of $\approx 200 \sim e^{N\delta^2/s}$ at the maximum recombination rate. Because the population parameters are chosen to be closer to the boundaries of our analytical approximations, the agreement between the analysis and the simulations is not as close in Figure 4 as it is in Figure 3, although it is still quite good.

In Figure 5 we show our analytical predictions and simulation results for τ as a function of the population size N. For small population sizes, the two loci are essentially never simultaneously polymorphic, so recombination has no effect. In this small-population

regime, the most likely way for the population to cross the valley is for one single mutant to drift to fixation, so τ increases as population size increases and selection against single mutants becomes more effective. At intermediate population sizes, the single mutants can produce a successful double-mutant lineage while still at low frequency in the population, and valley crossing becomes easier, with τ decreasing as N increases, and mutants are produced more frequently. A small amount of recombination increases the rate at which double mutants are produced, decreasing τ. Frequent recombination both increases the rate of production of double mutants and reduces the probability that a double-mutant lineage will be successful; these two effects cancel and τ is approximately the same as for an asexual population. At larger population sizes, τ continues to decrease with N when recombination is rare or absent, but rapidly increases with N for frequent recombination. At very large population sizes, double mutants are essentially produced instantly, and the population crosses the valley instantly as long as recombination is not too frequent. The agreement between the analytical predictions and the simulation results is very close for most of parameter space. The biggest differences occur at the boundaries of the analytical approximations, particularly at $N\mu \approx 1$ for occasional recombination and $N \approx s/\delta^2$ for frequent recombination.

DISCUSSION

We have provided a complete and intuitive description of the rate at which populations acquire adaptations requiring two mutations at different loci; these span a very broad range of parameter space and multiple regimes. Several subregimes of parameter space have previously been examined in detail. In particular, the case of strongly deleterious single mutants has been described for $r \ll s$ by MICHALAKIS and SLATKIN (1996) and HADANY (2003) and for $r \gg s$ as a special case by the results of BARTON and ROUHANI (1987), while for the case of neutral single mutants and $r \ll s$, $N\mu \gtrsim 1$ is examined in great detail in CHRISTIANSEN et al. (1998). That $r \approx s$ is the critical value of the recombination rate, above which recombination effectively breaks up double mutants, was described by CROW and KIMURA (1965) and further investigated by ESHEL and FELDMAN (1970), KARLIN and MCGREGOR (1971), WEINREICH and CHAO (2005), and JAIN (2009), among others. The present work serves to tie together these previous results, providing the boundaries of the regions of parameter space that they describe. In addition, we explore new regions of parameter space, including the case of mildly deleterious single mutants, the region in which recombination can speed up valley crossing the most. We also discuss the early-time dynamics most relevant for understanding cancer rates and many

aspects of pathogen evolution; these have not previously been analyzed at any depth for the most part, and our results contradict previous findings (SCHWEINSBERG 2008) in the one part of parameter space to have been considered in detail.

As mentioned earlier, LYNCH (2010) recently addressed some of the same questions. Here we compare his results to ours for the regions of parameter space where the two overlap. For the interplay between the asexual and sexual processes, LYNCH's (2010) Equation 1 for τ is incorrect. It relies on the assumption that for each single-mutant lineage, success via the asexual path occurs independently from success via the sexual path. But it follows from our heuristic arguments that the converse is true, since for neutral single mutants both paths depend primarily on the same random variable T, the time that the lineage drifts (see our Equations 9, 10, and 12). (And since drifting to fixation is also primarily dependent on T, earlier asexual versions of this equation such as Equation 3b in LYNCH and ABEGG 2010 are also incorrect.) For neutral single mutants and $r \ll s$, Equation 2 in LYNCH (2010) is similar to the second line of our Equation 13 and is found using a method that appears to be similar to our heuristic argument. However, the approximations given for the two parts of the process are valid in opposite limits and thus inconsistent: an anomalously long-lived first mutant dominates when $N\mu \ll 1$, but the second mutant can be treated only deterministically, as assumed, for $N\mu \gg 1$. For $r \gg s$ and large population sizes, $N\mu \gg 1$, LYNCH (2010) claims $\tau \gg \tau_{r=0}$, in contrast with our result $\tau \approx \tau_{r=0}$, although it is not clear in this case how he has defined τ or the boundaries of the regime. For deleterious single mutants and $r \ll s$, Equation 5 in LYNCH (2010) is clearly incorrect, since it adds Equation 4b, which is valid only for strongly deleterious single mutants, and Equation 4a, which overcounts the asexual path across the valley that is already included in Equation 4b. The (fairly poor) agreement with simulations is obtained only by the use of a fitting parameter with no biological meaning. For $r \gg s$, the regime where recombination slows down valley crossing the most, LYNCH (2010) gives an expression for τ (Equations 6 and 7) that differs somewhat from our Equation 8 because it is taken from results for a different kind of fitness valley from the one that both he and we are considering.

It was previously established that large asexual populations can rapidly cross shallow valleys, suggesting that complex adaptation may be common in such populations (see, e.g., WEISSMAN et al. 2009 and LYNCH and ABEGG 2010). Here we find that crossing shallow valleys is made even easier by moderate amounts of recombination (and is not affected by large amounts of recombination), suggesting that this kind of complex adaptation may be even more common in large sexual populations, especially among different sites within single genes for obligately sexual populations. On the other hand, we have found that sexual populations are less likely than asexual populations to cross deep valleys. Thus, recombination increases the existing bias in favor of crossing shallow rather than deep valleys toward large rather than small peaks.

Previous studies such as STEPHAN and KIRBY (1993) have found that the rate of compensatory double substitutions at distant loci is suppressed in natural populations, evidence of the decrease in the rate of valley crossing for $r \gg s$ predicted by our work as well as many previous results. Our results further suggest that for very closely linked loci, adaptive double substitutions should actually increase as a function of the genetic distance between the loci. However, this signal is likely to be harder to detect in data from natural populations, because it is difficult to identify adaptive double substitutions. (PISKOL and STEPHAN 2008 do observe that the rate of double substitutions at a certain set of interacting sites does increase with genetic distance at short distances, but this is unlikely to be due to recombination, as the double substitutions in question are expected to be only compensatory, not adaptive.)

This article has focused on the specific case of a population with a single possible adaptation, requiring two specific individually neutral or deleterious mutations. Real populations, however, are likely to have many possible adaptations, some requiring only a single mutation and others requiring many, with each adaptation potentially reachable by multiple different mutational paths. Understanding what determines the rate and manner of adaptation in these more realistic situations remains a very difficult problem, but the work in this article should be helpful. In particular, for populations with possible complex adaptations requiring mutations at several loci, we expect that it will still be true that a small amount of recombination among the necessary loci will make the adaptation more likely to be acquired, while a large amount will make it unlikely, with the scale separating small and large recombination rates being set by the advantage provided by the adaptation. Such an effect has indeed been observed in simulations (SUZUKI 1997). More generally, we should expect to see a transition from selection acting on some combinations of mutations at tightly linked loci to selection acting roughly independently on different haplotype blocks at longer genetic distances. NEHER and SHRAIMAN (2009) find such effects in simulations and analyses of evolution on certain types of (random) fitness landscapes. Because the genetic length scale at which this transition occurs depends on the strength of selection, we expect that complex adaptations that provide large selective advantages may involve longer regions of chromosomes than adaptations providing smaller advantages.

LITERATURE CITED

ANDOLFATTO, P., and J. D. WALL, 2003 Linkage disequilibrium patterns across a recombination gradient in african *Drosophila melanogaster*. Genetics **165:** 1289–1305.

ATHREYA, K. B., and P. E. NEY, 1972 *Branching Processes*, Ed. 1. Springer-Verlag, New York.

BARTON, N. H., 1995 Linkage and the limits to natural selection. Genetics **140:** 821–841.

BARTON, N. H., and S. ROUHANI, 1987 The frequency of shifts between alternative equilibria. J. Theor. Biol. **125:** 397–418.

BLOUNT, Z. D., C. Z. BORLAND and R. E. LENSKI, 2008 Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **105:** 7899–7906.

BRIDGHAM, J. T., S. M. CARROLL and J. W. THORNTON, 2006 Evolution of hormone-receptor complexity by molecular exploitation. Science **312:** 97–101.

CARTER, A. J. R., and G. P. WAGNER, 2002 Evolution of functionally conserved enhancers can be accelerated in large populations: a population-genetic model. Proc. R. Soc. B **269:** 953–960.

CAVALLI-SFORZA, L., and A. W. F. EDWARDS, 1967 Phylogenetic analysis: models and estimation procedures. Am. J. Hum. Genet. **19:** 233–257.

CHRISTIANSEN, F. B., S. P. OTTO, A. BERGMAN and M. W. FELDMAN, 1998 Waiting with and without recombination: the time to production of a double mutant. Theor. Popul. Biol. **53:** 199–215.

CROW, J. F., and M. KIMURA, 1965 Evolution in sexual and asexual populations. Am. Nat. **99:** 439–450.

DE VISSER, J. A. G. M., S.-C. PARK and J. KRUG, 2009 Exploring the effect of sex on empirical fitness landscapes. Am. Nat. **174:** S15–S30.

DURRETT, R., and D. SCHMIDT, 2008 Waiting for two mutations: with applications to regulatory sequence evolution and the limits of Darwinian evolution. Genetics **180:** 1501–1509.

ESHEL, I., and M. W. FELDMAN, 1970 On the evolutionary effect of recombination. Theor. Popul. Biol. **1:** 88–100.

EWENS, W. J., 2004 *Mathematical Population Genetics*, Ed. 2. Springer-Verlag, New York.

FISHER, D. S., 2007 Evolutionary dynamics, pp. 395–446 in *Complex Systems* (Lecture Notes of the Les Houches Summer School, Vol. LXXXV), edited by J. P. BOUCHAUD, M. MEZARD and J. DALIBARD. Springer-Verlag, Amsterdam.

FRANK, S. A., 2007 *Dynamics of Cancer: Incidence, Inheritance, and Evolution*. Princeton University Press, Princeton, NJ.

GOH, C.-S., A. A. BOGAN, M. JOACHIMIAK, D. WALTHER and F. E. COHEN, 2000 Co-evolution of proteins with their interaction partners. J. Mol. Biol. **299:** 283–293.

HADANY, L., 2003 Adaptive peak shifts in a heterogenous environment. Theor. Popul. Biol. **63:** 41–51.

HALDANE, J. B. S., 1931 A mathematical theory of natural selection, part VIII: metastable populations. Math. Proc. Camb. Philos. Soc. **27:** 137–142.

HILLIKER, A. J., G. HARAUZ, A. G. REAUME, M. GRAY, S. H. CLARK *et al.*, 1994 Meiotic gene conversion tract length distribution within the rosy locus of *Drosophila melanogaster*. Genetics **137:** 1019–1026.

IWASA, Y., F. MICHOR and M. A. NOWAK, 2004 Stochastic tunnels in evolutionary dynamics. Genetics **166:** 1571–1579.

JAIN, K., 2009 Time to fixation in the presence of recombination. Theor. Popul. Biol. **77:** 23–31.

KARLIN, S., and J. MCGREGOR, 1971 On mutation selection balance for two-locus haploid and diploid populations. Theor. Popul. Biol. **2:** 60–70.

KENDALL, D. G., 1948 On the generalized "birth-and-death" process. Ann. Math. Stat. **19:** 1–15.

KIM, Y., 2007 Rate of adaptive peak shifts with partial genetic robustness. Evolution **61:** 1847–1856.

KIMURA, M., 1962 On the probability of fixation of mutant genes in a population. Genetics **47:** 713–719.

KIMURA, M., 1964 Diffusion models in population genetics. J. Appl. Probab. **1:** 177–232.

KIMURA, M., 1985 The role of compensatory neutral mutations in molecular evolution. J. Genet. **64:** 7–19.

KOMAROVA, N. L., 2006 Spatial stochastic models for cancer initiation and progression. Bull. Math. Biol. **68:** 1573–1599.

KOMAROVA, N. L., A. SENGUPTA and M. A. NOWAK, 2003 Mutation-selection networks of cancer initiation: tumor suppressor genes and chromosomal instability. J. Theor. Biol. **223:** 433–450.

LEVIN, B. R., V. PERROT and N. WALKER, 2000 Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in bacteria. Genetics **154:** 985–997.

LYNCH, M., 2010 Scaling expectations for the time to establishment of complex adaptations. Proc. Natl. Acad. Sci. USA **107:** 16577–16582.

LYNCH, M., and A. ABEGG, 2010 The rate of establishment of complex adaptations. Mol. Biol. Evol. **27:** 1404–1414.

MASEL, J., 2006 Cryptic genetic variation is enriched for potential adaptations. Genetics **172:** 1985–1991.

MICHALAKIS, Y., and M. SLATKIN, 1996 Interaction of selection and recombination in the fixation of negative-epistatic genes. Genet. Res. **67:** 257–269.

NEHER, R. A., and B. I. SHRAIMAN, 2009 Competition between recombination and epistasis can cause a transition from allele to genotype selection. Proc. Natl. Acad. Sci. USA **106:** 6866–6871.

NOWAK, M. A., 2006 *Evolutionary Dynamics: Exploring the Equations of Life*, Ed. 1. Belknap Press, Cambridge, MA.

PISKOL, R., and W. STEPHAN, 2008 Analyzing the evolution of RNA secondary structures in vertebrate introns using Kimura's model of compensatory fitness interactions. Mol. Biol. Evol. **25:** 2483–2492.

SCHWEINSBERG, J., 2008 The waiting time for m mutations. Electron. J. Probab. **13:** 1442–1478.

STEPHAN, W., 1996 The rate of compensatory evolution. Genetics **144:** 419–426.

STEPHAN, W., and D. A. KIRBY, 1993 RNA folding in Drosophila shows a distance effect for compensatory fitness interactions. Genetics **135:** 97–103.

SUZUKI, H., 1997 The optimum recombination rate that realizes the fastest evolution of a novel functional combination of many genes. Theor. Popul. Biol. **51:** 185–200.

TAKAHASI, K. R., and F. TAJIMA, 2005 Evolution of coadaptation in a two-locus epistatic system. Evolution **59:** 2324–2332.

TAKAHATA, N., 1982 Sexual recombination under the joint effects of mutation, selection, and random sampling drift. Theor. Popul. Biol. **22:** 258–277.

WEINREICH, D. M., and L. CHAO, 2005 Rapid evolutionary escape by large populations from local fitness peaks is likely in nature. Evolution **59:** 1175–1182.

WEISSMAN, D. B., M. M. DESAI, D. S. FISHER and M. W. FELDMAN, 2009 The rate at which asexual populations cross fitness valleys. Theor. Popul. Biol. **75:** 286–300.

WODARZ, D., and N. L. KOMAROVA, 2005 *Computational Biology of Cancer: Lecture Notes and Mathematical Modeling*, Ed. 1. World Scientific, Hackensack, NJ.

APPENDIX A: BRANCHING PROCESS DYNAMICS

When mutants are sufficiently rare in the population that they do not influence each other, the dynamics of a mutant lineage can be described by a continuous-time branching process in which individuals reproduce via binary fission at rate $1 - \delta$ (with $\delta \ll 1$) and die at rate 1. Here we list several basic facts about the dynamics of such lineages that serve as the basis for the analysis in the rest of the article. Let $n_{Ab}(t)$ be the number of individuals at time t, with $n_{Ab}(0) = 1$, and let T be the extinction time, $T \equiv \inf\{t: n_{Ab}(t) = 0\}$. The probability that the lineage persists for at least t generations is

$$P(T > t) = \left(1 + \frac{e^{\delta t} - 1}{\delta}\right)^{-1} \tag{A1}$$

$$\approx \begin{cases} 1/t & \text{for } 1 \ll t \ll 1/\delta \\ \delta e^{-\delta t} & \text{for } t \gg 1/\delta \end{cases} \tag{A2}$$

(Kendall 1948), and given that it does so, it has size $n_{Ab}(t) = n$ with probability

$$p_t(n_{Ab} = n \mid n_{Ab} > 0) = \frac{\delta}{1 - (1 - \delta)e^{-\delta t}} \left(\frac{(1 - \delta)(1 - e^{-\delta t})}{1 - (1 - \delta)e^{-\delta t}}\right)^{n-1}$$

$$\approx \begin{cases} \frac{1}{t}(1 + 1/t)^{-n} & \text{for } t \ll \frac{1}{\delta} \\ \delta(1 - \delta)^{n-1} & \text{for } t \gg \frac{1}{\delta}. \end{cases} \tag{A3}$$

Note that the distribution of extinction times has a long tail before falling off exponentially for $t \gtrsim 1/\delta$. Note also that the number of individuals at a given time has an approximately geometric distribution, falling off exponentially for $n \gtrsim \min\{t, 1/\delta\}$. The exact expressions above are also valid for $\delta < 0$, corresponding to a lineage with a selective advantage. Thus we can see from Equation A1 that (ignoring recombination) an AB lineage has a probability $\approx s$ of succeeding and that those AB lineages that do go extinct typically do so in $t \lesssim 1/s$ generations.

If we instead start with $n_{Ab}(t = 0) = n_0$ individuals, then the expected number of individuals at time t will be

$$E[n_{Ab}(t)] = n_0 e^{-\delta t},$$

and the variance will be

$$\text{Var}[n_{Ab}(t)] = (2 - \delta)n_0 e^{-\delta t}\frac{1 - e^{-\delta t}}{\delta}$$

$$\approx \begin{cases} 2n_0 t & \text{for } t \ll \frac{1}{\delta} \\ 2n_0 e^{-\delta t}/\delta & \text{for } t \gg \frac{1}{\delta}. \end{cases}$$

Note that the number of individuals is likely to remain $n_{Ab}(t) \sim n_0$ until $t \approx \min\{n_0, 1/\delta\}$ and that the lineage is likely to be much smaller or extinct for $t > 1/\delta$, regardless of n_0.

APPENDIX B: GENERATING FUNCTION CALCULATIONS

Why generating functions: We want to estimate the rate of stochastic tunneling, in which all mutants are rare until the double mutants begin to take over the population. Let $p_t(n_{Ab}, n_{aB}, n_{AB})$ be the probability that at time t there are n_{Ab} Ab individuals, n_{aB} aB individuals, and n_{AB} AB individuals in the population. It is useful to consider the generating function $\mathcal{Z}$ of p_t defined by

$$\mathcal{Z}(\phi_{Ab}, \phi_{aB}, \phi_{AB}, t) \equiv E[\exp(-\phi_{Ab}n_{Ab} - \phi_{aB}n_{aB} - \phi_{AB}n_{AB})]. \tag{B1}$$

In particular, to calculate the probability $P(t)$ that the double mutants dominate the population at time t (*i.e.*, that the population has crossed the valley by this time), it is useful to consider the generating function for the distribution of n_{AB}, $\mathcal{Z}(0, 0, h, t) = E[\exp(-hn_{AB})]$. In typical valley-crossing dynamics, n_{AB} will remain small for a time, until it reaches a threshold number at which selection begins to effectively favor the double mutants, after which they quickly sweep to near fixation. In large populations, this threshold number will be $\ll N$, so we can choose $h \gg 1/N$ such that n_{AB} will typically remain $\ll 1/h$ until the double mutants take over the population. For such h, the random variable $\exp(-hn_{AB})$ remains ≈ 1 until the population crosses the valley, at which point it rapidly drops to ≈ 0 (because n_{AB} approaches $N \gg 1/h$). Thus $\exp(-hn_{AB})$ is an approximate indicator variable. Therefore, its expected value $\mathcal{Z}(0, 0, h, t)$ approximately gives the probability of *not* having crossed the valley by time t, and we can write

$$P(t) \approx 1 - \mathcal{Z}(0, 0, h, t). \tag{B2}$$

We see that if we find $\mathcal{Z}$, we will have found the distribution of valley-crossing times $\mathcal{T}$.

Partial differential equation for $\mathcal{Z}$**:** To calculate $\mathcal{Z}$, we use a continuous-time Moran model for the population. All individuals die at an equal rate, which sets the unit of time, and are chosen to replace dead individuals with a probability proportional to their fitness. Mutation and recombination occur independently from each other and from reproduction. (It is trivial to adjust the parameters to allow for mutations or recombinations occurring at birth.) When mutants are rare, so that mutants essentially do not compete with each other, p_t evolves according to the approximate differential equation

$$
\begin{aligned}
\frac{\partial p_t(n_{Ab}, n_{aB}, n_{AB})}{\partial t} &\approx [(1 - \delta_A)(n_{Ab} - 1) + N\mu_A]p_t(n_{Ab} - 1, n_{aB}, n_{AB}) + (n_{Ab} + 1)p_t(n_{Ab} + 1, n_{aB}, n_{AB}) \\
&+ [(1 - \delta_B)(n_{aB} - 1) + N\mu_B]p_t(n_{Ab}, n_{aB} - 1, n_{AB}) + (n_{aB} + 1)p_t(n_{Ab}, n_{aB} + 1, n_{AB}) \\
&+ [(1 + s)(n_{AB} - 1) + \mu_A n_{Ab} + \mu_B n_{aB}]p_t(n_{Ab}, n_{aB}, n_{AB} - 1) \\
&+ (n_{AB} + 1)p_t(n_{Ab}, n_{aB}, n_{AB} + 1) + r(n_{AB} + 1)p_t(n_{Ab} - 1, n_{aB} - 1, n_{AB} + 1) \\
&+ \frac{r}{N}(n_{Ab} + 1)(n_{aB} + 1)p_t(n_{Ab} + 1, n_{aB} + 1, n_{AB} - 1) \\
&- \Big[N(\mu_A + \mu_B) + (2 - \delta_A)n_{Ab} + (2 - \delta_B)n_{aB} + \frac{r}{N}n_{Ab}n_{aB} \\
&+ (2 + s + r)n_{AB}\Big]p_t(n_{Ab}, n_{aB}, n_{AB}).
\end{aligned}
\tag{B3}
$$

It is straightforward to use Equations B1 and B3 to write down a partial differential equation for $\mathcal{Z}$. For $\phi_{Ab}, \phi_{aB}, \phi_{AB} \ll 1$, this is approximately

$$
\begin{aligned}
\frac{\partial \mathcal{Z}}{\partial t} &\approx -N(\mu_A \phi_{Ab} + \mu_B \phi_{aB})\mathcal{Z} + \frac{r}{N}[\phi_{Ab} + \phi_{aB} - \phi_{AB}]\frac{\partial \mathcal{Z}}{\partial \phi_{Ab}\partial \phi_{aB}} \\
&+ [\mu_A \phi_{AB} - \delta_B \phi_{aB} - \phi_{aB}^2]\frac{\partial \mathcal{Z}}{\partial \phi_{aB}} + [\mu_B \phi_{AB} - \delta_A \phi_{Ab} - \phi_{Ab}^2]\frac{\partial \mathcal{Z}}{\partial \phi_{Ab}} \\
&+ [(s - r)\phi_{AB} + r(\phi_{Ab} + \phi_{aB}) - \phi_{AB}^2]\frac{\partial \mathcal{Z}}{\partial \phi_{AB}}.
\end{aligned}
\tag{B4}
$$

Note that if it were not for the recombination between Ab and aB individuals, which gives rise to the second-order term $\partial \mathcal{Z}/\partial \phi_{Ab}\partial \phi_{aB}$, Equation B4 would describe the evolution of the generating function of a familiar multitype branching process. For such processes, the problem of finding the probability of fixation of mutant alleles has been extensively studied (see, *e.g.*, BARTON 1995 and the references therein). Unfortunately, the addition of recombination makes our problem significantly more complicated; Equation B4 cannot be solved exactly.

Small populations ($N\mu \ll 1$): To simplify the problem, we must make some approximations. For $N\mu \gg 1$, n_{Ab} and n_{aB} are usually well approximated by their expected values (FISHER 2007), greatly simplifying the analysis; see below for a discussion of this case. Here we focus on $N\mu \ll 1$. In this case, single-mutant lineages arise infrequently, so for an Ab lineage and an aB lineage to coexist and recombine to produce an AB lineage, one of two rare events must occur: either the first lineage to arise must persist for an unusually long time or the two lineages must arise unusually close together in time. Recalling the heuristic description of lineage dynamics given in the main text (with the distribution of lineage lifetimes T having a long tail like $\sim 1/t$ until it begins to fall off exponentially at $t \sim 1/\delta$), we expect that the former possibility will be more common for sufficiently small δ (neutral or mildly deleterious single mutants) while the latter will be relatively more common for larger δ (strongly deleterious single mutants).

We focus on the first possibility, that the first single-mutant lineage persists for an unusually long time. In this case, there is an asymmetry between the two successful single-mutant lineages, since one arises well before the other. Here we find p_{Ab}, the probability that an Ab lineage will be successful by drifting for a long time and then producing a successful AB individual either directly by mutation or by recombining with a later aB lineage. p_{aB}, the probability that an aB lineage will be successful by drifting for a long time, will be the same (with $\delta_A \leftrightarrow \delta_B$, $\mu_A \leftrightarrow \mu_B$), and when this is the dominant mode of valley crossing, $P(t)$ at long times will be given by

$$
P(t) \approx 1 - \exp\big[-Nt(\mu_A p_{Ab} + \mu_B p_{aB})\big].
\tag{B5}
$$

Probability that an Ab lineage will be successful: Now we turn to finding p_{Ab}. By assumption, the Ab lineage arises and drifts for many generations without interacting with aB individuals, so its dynamics can be modeled by a simple single-type branching process. The approximate dynamics of such a process are discussed in the main text, while the exact distributions of the extinction time and the number of individuals are given in APPENDIX A. Because the Ab

lineage is likely to be large by the time the aB lineage arises (at least for δ_A not too large), we make the approximation that the Ab lineage's dynamics are relatively unaffected by the aB and AB lineages until these lineages become large and the mutants are beginning to sweep to fixation.

The Ab lineage will, however, affect the other mutant lineages. To find p_{Ab}, we consider the dynamics of $n_{aB}(t)$ and $n_{AB}(t)$ conditioned on a particular trajectory $\{n_{Ab}(t)\}$ and then take the expectation over all possible trajectories. Let $Z(\phi_{aB}, \phi_{AB}, t \,|\, \{n_{Ab}\})$ be the generating function of this conditioned process, defined as in Equation B1 by $Z(\phi_{aB}, \phi_{AB}, t \,|\, \{n_{Ab}\}) \equiv E[\exp(-\phi_{aB}n_{aB} - \phi_{AB}n_{AB}) \,|\, \{n_{Ab}\}]$. Then, as in Equation B2, the probability that the Ab lineage with trajectory $\{n_{Ab}\}$ has produced a successful AB lineage t generations after it has arisen is

$$P \,|_{n_{Ab}}(t) \approx 1 - Z(0, h, t \,|\, \{n_{Ab}\}), \tag{B6}$$

and the total probability that the Ab lineage is successful is

$$p_{Ab} \approx 1 - \lim_{t \to \infty} E[Z(0, h, t \,|\, \{n_{Ab}\})]. \tag{B7}$$

We can thus find p_{Ab} by finding Z.

Generating function for the aB and AB lineages: To find Z, note that it satisfies the partial differential equation

$$\frac{\partial Z}{\partial t} = -(N\phi_{aB} + n_{Ab}\phi_{AB})\mu_B Z + \left[(s - r)\phi_{AB} + r\phi_{aB} - \phi_{AB}^2 \right] \frac{\partial Z}{\partial \phi_{AB}}$$
$$+ \left[(rx_A + \mu_A)\phi_{AB} - (rx_A + \delta_B)\phi_{aB} - \phi_{aB}^2 \right\} \frac{\partial Z}{\partial \phi_{aB}}, \tag{B8}$$

where $x_A \equiv n_{Ab}/N$, with $x_A \ll 1$ by assumption. (Equation B8 can be derived either directly from the master equation for the conditioned process or by simply substituting $\phi_{Ab} \to 0$, $\partial \mathcal{Z}/\partial \phi_{Ab} \to -n_{Ab}\mathcal{Z}$, $\mathcal{Z} \to Z$ in Equation B4.)

Equation B8 describes the evolution of a two-type branching process with time-dependent transition rates, and we can solve it using the method of characteristics (KENDALL 1948). Taking the generating function variables to be functions of time, $\phi_i = \phi_i(t)$, the total time derivative of Z is given by

$$\frac{dZ}{dt} = -(N\phi_{aB} + n_{Ab}\phi_{AB})\mu_B Z + \left[(s - r)\phi_{AB} + r\phi_{aB} - \phi_{AB}^2 + \frac{d\phi_{AB}}{dt} \right] \frac{\partial Z}{\partial \phi_{AB}}$$
$$+ \left[(rx_A + \mu_A)\phi_{AB} - (rx_A + \delta_B)\phi_{aB} - \phi_{aB}^2 + \frac{d\phi_{aB}}{dt} \right] \frac{\partial Z}{\partial \phi_{aB}}. \tag{B9}$$

Note that if we choose $\phi_{aB}(t)$ and $\phi_{AB}(t)$ to cancel the second and third terms on the right-hand side of Equation B9, we are left with a simple ordinary differential equation for Z, with the solution

$$Z(t) = \exp\left[-\mu_B \int_0^t (N\phi_{aB}(t') + n_{Ab}\phi_{AB}(t'))\,dt' \right]. \tag{B10}$$

Note that Equation B10 implies that the probability that a lineage with trajectory $\{n_{Ab}\}$ is successful after t generations is

$$P \,|_{n_{Ab}}(t) \approx 1 - \exp\left[-\mu_B \int_0^t (N\phi_{aB}(t') + n_{Ab}\phi_{AB}(t'))\,dt' \right], \tag{B11}$$

and the total probability of success is

$$p_{Ab} \approx 1 - E\left[\exp\left[-\mu_B \int_0^t (N\phi_{aB}(t') + n_{Ab}\phi_{AB}(t'))\,dt' \right] \right]. \tag{B12}$$

From Equations 11 and 12, we see that ϕ_{aB} and ϕ_{AB} can be naturally interpreted as the conditioned probability of success for an aB lineage and an AB lineage, respectively. (The first term in the integrand corresponds to the aB lineages that arise while the Ab lineage drifts, while the second term corresponds to the AB lineages produced directly by mutation from Ab.) See BARTON (1995) for similar expressions obtained in a slightly different way.

Characteristic equations for probabilities of success: For Equation B11 to be valid, ϕ_{aB} and ϕ_{AB} must satisfy the (Riccati) differential equations

$$-\frac{d\phi_{aB}}{dt} = -(rx_A + \delta_B)\phi_{aB} + (rx_A + \mu_A)\phi_{AB} - \phi_{aB}^2 \tag{B13}$$

$$-\frac{d\phi_{AB}}{dt} = r\phi_{aB} + (s - r)\phi_{AB} - \phi_{AB}^2. \tag{B14}$$

We are interested in the parameter regime where aB individuals are much more likely to produce AB individuals by recombining with Ab than by mutating ($rx_A \gg \mu_A$), so we neglect the term proportional to μ_A in Equation B13 for the rest of this article. (Note that Equation B11 does include the possibility that the focal Ab lineage produces AB individuals via mutation.)

We can solve Equations B13 and B14 approximately for the trajectory of ϕ_{aB} and ϕ_{AB}, moving backward in time. Let u be a backward-time variable, with $u = 0$ corresponding to the time t at which $\phi_{aB} = 0$, $\phi_{AB} = h$, and $u = t$ corresponding to the time at which the Ab lineage arises. For small u (*i.e.*, at late times), ϕ_{aB} and ϕ_{AB} are very small and we can ignore the quadratic terms in Equations B13 and B14. The eigenvalues of these linearized equations are

$$\lambda_{\pm} = \frac{1}{2}\left(s - r(1 + x_A) - \delta_B \pm \sqrt{(s - r(1 + x_A) - \delta_B)^2 + 4x_A rs - 4\delta_B(r - s)} \right). \tag{B15}$$

After an initial transient of the order of $(\lambda_+ - \lambda_-)^{-1} = \mathcal{O}(\min\{1/s, 1/r, 1/\delta_B\})$ generations, the larger eigenvalue λ_+ dominates and ϕ_{aB} and ϕ_{AB} increase as $\exp(\int_0^u \lambda_+ du')$ until the quadratic terms in Equations B13 and B14 become important. Note that λ_+ is a function of x_A and therefore changes over time. However, we will see that the dynamics of ϕ_{aB} and ϕ_{AB} are fast compared to the changes in x_A for many of the most important trajectories $\{n_{Ab}\}$. For such trajectories, ϕ_{Ab} and ϕ_{AB} quickly approach the fixed point $(\phi_{aB}^*, \phi_{AB}^*)$ of Equations B13 and B14, which then changes slowly as x_A changes.

Fixed point and heuristic interpretation: Setting the left-hand sides of Equations B13 and B14 to 0, we find that the first coordinate ϕ_{aB}^* of the constant-x_A fixed point satisfies

$$0 = \phi_{aB}^*\left[rx_A s + (s - r)\delta_B + \left(s - r - rx_A\left(1 + \frac{\delta_B}{rx_A}\right)^2 \right)\phi_{aB}^* - 2\left(1 + \frac{\delta_B}{rx_A}\right)\phi_{aB}^{*2} - \frac{1}{rx_A}\phi_{aB}^{*3} \right]. \tag{B16}$$

Equation B16 can be solved exactly using the cubic formula, but the solution is not illuminating. Given that it is already an approximation, it is more useful to look at the simple approximate solutions in different limits, along with the corresponding approximate value of ϕ_{AB}^*:

$$(\phi_{aB}^*, \phi_{AB}^*) \approx \begin{cases} (\sqrt{rx_A s}, s) & \text{for } r \ll s, \delta_B^2 \ll rx_A s \\ \left(\frac{rx_A s}{\delta_B}, s\right) & \text{for } r \ll s, \delta_B^2 \gg rx_A s \\ (x_A s - \delta, x_A s - \delta) & \text{for } r \gg s, \delta_B < x_A s \\ (0, 0) & \text{for } r \gg s, \delta_B > x_A s. \end{cases} \tag{B17}$$

Recalling that ϕ_{aB} and ϕ_{AB} can be interpreted as the conditioned probabilities of success of aB and AB individuals, respectively, we can understand some aspects of Equation B17 intuitively. For $r \ll s$, we have $\phi_{AB}^* \approx s$ (the asexual probability of success for AB), because in this case recombination is unlikely to break up the beneficial combination. For small δ_B, $\phi_{aB}^* \approx \sqrt{rx_A s}$ is the probability of drifting over $\sim 1/\sqrt{rx_A s}$ to a size $\sim \sqrt{rx_A s}$, at which point the lineage is likely to produce $\sim 1/s$ AB individuals via recombination and be successful. (Note that this agrees with the heuristic argument given in the main text.) For larger δ_B, $\phi_{aB}^* \approx rx_A s/\delta_B$ is the probability of being successful by drifting for $\sim 1/\delta_B$ generations to a size $\sim 1/\delta_B$ and producing rx_A/δ_B^2 AB individuals.

For $r \gg s$ we have $\phi_{AB}^* \approx \phi_{aB}^*$, indicating that linkage is not important (*i.e.*, given $n_{Ab} \gg 1$, a B allele has roughly the same probably of success on an a background as on an A background). The probability of success $x_A s - \delta$ is just the mean selective advantage of the B allele when rare. Note that the B allele has lower mean fitness than the b allele for $x_A < \delta/s$, so for these values of x_A the mutants will not succeed and $\phi_{aB}^* \approx \phi_{AB}^* \approx 0$. (In fact, there is still a small probability $\sim e^{-N\delta}$ that a B allele drifts against selection to fixation, and a small probability $\sim e^{-N(r-s)}$ that an AB lineage avoids recombination while sweeping to fixation, but we neglect them here.)

Timescale for aB and AB dynamics: As mentioned above, ϕ_{aB} and ϕ_{AB} will approach the values given in Equation B17 only when their dynamics are fast compared to the rate at which x_A changes. When ϕ_{aB} and ϕ_{AB} are far from the fixed point, they change at a rate λ_+. To find the rate of change as they approach the fixed point, we can linearize Equations B13 and B14 about the fixed point and find the eigenvalues $\lambda'_{\pm}$ of this system. The rate of approach will be set by λ'_+ (where $\lambda'_- < \lambda'_+ < 0$). Skipping the straightforward but tedious calculations, we find that in the asymptotic regimes of Equation B17, the dynamics occur at rates λ_+ and λ'_+ given by

$$(\lambda_+, |\lambda'_+|) \approx \begin{cases} (s, 2\sqrt{rx_A s}) & \text{for } r \ll s, \delta_B^2 \ll rx_A s \\ (s, \min\{s, \delta_B\}) & \text{for } r \ll s, \delta_B^2 \gg rx_A s \\ (x_A s - \delta_B, x_A s - \delta_B) & \text{for } r \gg s, \delta_B < x_A s \end{cases} \tag{B18}$$

(Note that we have ignored the trivial fixed point in the last line of Equation B17, which does not contribute to the probability of success.)

As with Equation B17, we can understand some aspects of Equation B18 intuitively. From the first two lines of Equation B17, we saw that for $r \ll s$ we could roughly neglect the effect of recombination on the dynamics of an AB lineage once it is formed. As mentioned in APPENDIX A, we thus expect an AB lineage to drift for $\sim 1/s$ generations before either becoming successful or going extinct; we see that this timescale corresponds exactly to the time $\sim \lambda_+^{-1}$ that it takes for ϕ_{AB} to begin to approach its fixed point. For small δ_B, we argued that $\phi_{aB}^* \approx \sqrt{rx_A s}$ corresponded to the probability of an aB lineage drifting for $\sim 1/\sqrt{rx_A s}$ generations; we now see that this is indeed the timescale $|\lambda'_+|^{-1}$ in the first line of Equation B18. For larger values of δ, the dominant process is for the aB lineage to drift for $\sim 1/\delta_B$ generations; this, together with the $1/s$ timescale associated with the AB lineage, sets the scale $|\lambda'_+|^{-1}$ in the second line of Equation B18. Finally, for $r \gg s$, the B allele has mean selective advantage $x_A s - \delta$ and thus typically drifts for $\sim 1/(x_A s - \delta)$ generations before beginning to increase deterministically or going extinct.

Recall that we wish to compare Equation B18 to the typical rates of change of x_A. As discussed in APPENDIX A, a trajectory that reaches $n_{Ab}(t) = n$ will typically remain $\mathcal{O}(n)$ for at least $\sim \min\{n, 1/\delta_A\}$ generations. So if the values in Equation B18 are large compared to $\max\{1/(Nx_A), \delta_A\}$ for the trajectories making the largest contributions to p_{Ab} in Equation B12, we can make the approximation that ϕ_{aB} and ϕ_{AB} are given by their fixed-point values for almost all the time that the Ab lineage is drifting. Doing so, Equation B12 becomes

$$p_{Ab} \approx 1 - E\left[\exp\left(-\mu_B \int_0^\infty \left[N\phi_{aB}^*(n_{Ab}(t)) + n_{Ab}(t)\phi_{AB}^*(n_{Ab}(t))\right] dt\right)\right]. \tag{B19}$$

We can now plug the values from Equation B17 into Equation B19 and evaluate the expectation to find p_{Ab}, checking that the dominant trajectories are indeed changing slowly compared to λ_+ and λ'_+.

Evaluating the probability that a single-mutant lineage will be successful: Unfortunately, there is no simple general expression for Equation B19; instead, we consider the different asymptotic forms, as in Equations B17 and B18. To reduce the number of different parameter regimes that we must consider, we now assume that the two loci are symmetric, with $\mu_A = \mu_B = \mu$ and $\delta_A = \delta_B = \delta$. (The generalization to the asymmetric case is straightforward—see APPENDIX C.) We then have $p_{aB} = p_{Ab} \equiv p$, and Equation B5 for the probability of having crossed the valley at long times becomes

$$P(t) \approx 1 - \exp\left[\frac{-t}{\tau}\right],$$

where τ is the expected time to cross the valley, and $\tau \approx (2N\mu p)^{-1}$, with p given by Equation B19.

We now proceed to evaluate Equation B16 and find expressions for the approximate rate of valley crossing, τ^{-1}. Even with the simplifying assumption of symmetric loci, there are a number of different limiting parameter regimes in which τ^{-1} takes different forms. (See Figure B1 for the different parameter regimes.) We skip the tedious details and provide only the results of the calculations. The crucial fact that we use is the one mentioned in the main text and described in APPENDIX A: that lineages are exponentially unlikely to drift for a time $t \gg 1/\delta$ and that given that a lineage drifts for a time t, it is exponentially unlikely to have a size $n \gg \min\{t, 1/\delta\}$ for much of that time. Because lineage trajectories that last longer than this or reach larger sizes are so unlikely, and because their contributions to Equation B19 typically grow no faster than $\sim nt^2$ at the most, the trajectories that make the dominant contribution to p for most of parameter space are the ones that last no longer than $\sim 1/\delta$ generations and reach typical sizes given their lifetime. On the other hand, because the lineage size distribution has a long tail until it reaches these boundaries, large (but not exponentially suppressed) lineages are generally important.

Asexual regime: First note that for very small r, the expressions for ϕ_{aB}^* in the first and second lines of Equation B17 go to 0, and the integrand in Equation B19 is dominated by $N\phi_{AB}^*$:

$$p \approx 1 - E\left[\exp\left(-\mu s \int_0^\infty n_{Ab}(t) dt\right)\right]. \tag{B20}$$

In this case, double mutants are more likely to be produced by mutation of single mutants than by recombination between single mutants, and the population is effectively asexual. We considered this case in WEISSMAN *et al.* (2009). The rate of valley crossing will be given by the asexual rate, $\tau_{r=0}^{-1}$, namely

$$\tau_{r=0}^{-1} \approx \begin{cases} 2N\sqrt{\mu^3 s} & \text{for } \delta \ll \sqrt{\mu s} \text{ (effectively neutral single mutants)} \\ \frac{2N\mu^2 s}{\delta} & \text{for } \delta \gg \sqrt{\mu s} \text{ (deleterious single mutants).} \end{cases} \tag{B21}$$

Equation B21 holds in moderately large populations, with $N \gg \min\{1/\sqrt{\mu s}, 1/\delta\}$; for smaller populations, the dominant valley-crossing process is for one mutant genotype to drift to fixation before the double mutant is produced. (In this small-population regime, because the two loci are essentially never simultaneously polymorphic, recombination has no effect; see Figure 5.)

Infrequent recombination ($r \ll s$): Now we consider larger r (but still $\ll s$). First, we verify that our approximation $\phi_{aB} \approx \phi_{aB}^*$, $\phi_{AB} \approx \phi_{AB}^*$ is correct. We give the typical lifetime T for the trajectories that dominate p, along with the corresponding timescale $\lambda^{-1} \sim \lambda_+^{-1} + \lambda_+'^{-1}$ for ϕ_{aB} and ϕ_{AB} to approach their fixed point value:

$$(T, \lambda^{-1}) \approx \begin{cases} \left((N\mu^2 rs)^{-1/3}, \left(\frac{N^2\mu}{rs}\right)^{1/3} \right) & \text{for } \max\{\sqrt{\mu s}, \delta\} \ll (N\mu^2 rs)^{1/3} \\ \left(\frac{1}{\delta}, \sqrt{\frac{\delta}{Nrs}} \right) & \text{for } \max\{\frac{s}{Nr}, (N\mu^2 rs)^{1/3}\} \ll \delta \ll \left(\frac{rs}{N}\right)^{1/3} \\ \left(\frac{1}{\delta}, \frac{1}{\delta} + \frac{1}{s} \right) & \text{for } \left(\frac{rs}{N}\right)^{1/3} \ll \delta \end{cases} \tag{B22}$$

In the first two regimes, $\lambda^{-1} \ll T$ and our guess is consistent, but for sufficiently deleterious single mutants it fails. In this last case, selection limits both n_{Ab} and n_{aB} to $\sim 1/\delta$, so valley crossing typically occurs via two similarly sized lineages and the analysis of this APPENDIX fails. Fortunately, in this case selection limits the size of fluctuations in the numbers of single mutants, and n_{Ab} and n_{aB} can be approximated by their expectations. This actually allows the rate of valley crossing to be calculated far more easily. We quote only the result here (see Equation B23 below); for a detailed analysis, see HADANY (2003). (It is interesting to note that even in this case where our approximations fail, they overestimate the rate of valley crossing only by a factor of 2.)

Evaluating Equation B19 in the regimes where $\lambda^{-1} \ll T$, and combining this with the large-δ result and Equation B21, we finally have the rate of valley crossing for $r \ll s$:

$$\tau_{r<s}^{-1} \approx \begin{cases} \tau_{r=0}^{-1} & \text{for } r \ll \max\left\{ 2\delta, \min\left\{ \frac{1}{N}\sqrt{\frac{s}{\mu}}, \frac{s}{N\delta} \right\} \right\} \text{ (effectively asexual)} \\ 2(N^4\mu^5 rs)^{1/3} & \text{for } \max\left\{ \sqrt{\mu s}, \delta \right\} \ll (N\mu^2 rs)^{1/3}\} \text{ (effectively neutral single mutants)} \\ 2\mu^2 \sqrt{\frac{\pi N^3 rs}{\delta}} & \text{for } \max\left\{ \frac{s}{Nr}, (N\mu^2 rs)^{1/3} \right\} \ll \delta \ll \left(\frac{rs}{N}\right)^{1/3} \text{ (slightly deleterious single mutants)} \\ \frac{N\mu^2 rs}{\delta^2} & \text{for } \left(\frac{rs}{N}\right)^{1/3} \ll \delta \ll \frac{r}{2} \text{ (moderately deleterious single mutants)} \end{cases} \tag{B23}$$

(see Figure B1). In the first line, we see that for sufficiently small r, the population is effectively asexual, and the rate of valley crossing is given by Equation B21. For recombination to effectively speed up valley crossing, there must be a significant probability that Ab and aB individuals are simultaneously present (N and μ not too small, δ not too large) and the asexual valley-crossing process must not be too easy (s not too large). Given these conditions, recombination will be the dominant means of producing double mutants, and the population will fall into one of the three remaining parameter regimes in Equation B23, depending on the size of δ. In the "effectively neutral" parameter regime, selection is sufficiently weak that the most likely way for the population to cross the valley is for one single-mutant genotype, $e.g.$, Ab, to drift to such a high number $n_{Ab} \approx (N\mu^2 rs)^{-1/3}$ that success becomes likely ($P|_{n_{Ab}} \approx 1$); all the luck is concentrated in this initial large lineage. In the "slightly deleterious" regime, selection makes it unlikely that a lineage will reach this size. Instead, the most likely way for the population to cross the valley is for one lucky lineage to drift to size $n_{Ab} \sim 1/\delta$, during which time another lucky lineage drifts to size $n_{aB} \sim \sqrt{N\delta/rs}$, at which point success becomes likely. At even higher values of δ, in the "moderately deleterious" regime, both single-mutant lineages are limited to size $\sim 1/\delta$, and the double mutants that they generate must get lucky to escape drift and take over the population.

Frequent recombination ($r \gg s$): Now we turn to the case $r \gg s$. Note that since $\phi_{aB}^* \approx \phi_{AB}^*$ in this case and $n_{Ab} \ll N$ by assumption, the integrand in Equation B19 is dominated by the first term, and p is approximately

$$p \approx 1 - E\left[\exp\left(-\mu \int_0^\infty (n_{Ab}(t)s - \delta)\theta\left[n_{Ab}(t) - \frac{N\delta}{s}\right]dt\right)\right],\tag{B24}$$

where θ is the Heaviside step function. Note that for $\delta = 0$, this is just Equation B20, the probability of success of a single-mutant lineage in an *asexual* population. If $\delta \ll \sqrt{s/N}$, then the extra terms in Equation B24 do not affect the main contributions of the trajectories that dominate the asexual case, and $\tau_{r>s} \approx \tau_{r=0}$. (It is easy to check that in these cases the condition $Ns \gg 1$ implies that the timescale λ^{-1} in Equation B18 is small compared to the timescale for the Ab lineage dynamics, so our approximations are consistent.)

On the other hand, if we have $\delta \gg \sqrt{s/N}$, then only the exponentially unlikely trajectories with $n_{Ab} \ll 1/\delta$ contribute to p. In this case, because the integrand is positive only for $n_{Ab} > N\delta/s$, we can see from Equation A3 in appendix a that $p = \mathcal{O}(\exp(-N\delta^2/s))$, and thus $\log \tau \approx N\delta^2/s$. Note that this agrees with the result of Barton and Rouhani (1987) quoted in the main text, although it was obtained via a very different method. Note also that p is tiny compared to its values in other parameter regimes; this is why we can generally ignore unlikely trajectories that last much longer than $\sim 1/\delta$ or reach disproportionate sizes—their contribution to p is negligible as long as the nonsuppressed lineages can be successful.

Putting these results together, we find that the rate of valley crossing for $r \gg s$ is given approximately by

$$\tau_{r>s}^{-1} \approx \begin{cases} \tau_{r=0}^{-1} & \text{for } \delta \ll \sqrt{\frac{s}{N}} \\ \sim e^{-N\delta^2/s} & \text{for } \delta \gg \sqrt{\frac{s}{N}}. \end{cases}\tag{B25}$$

See Figure B1 for a typical slice of parameter space. We see that as long as the single mutants are not too deleterious, the two effects of recombination on valley crossing (bringing together mutant alleles and breaking up the beneficial combination) roughly cancel. In contrast, for deeper valleys, frequent recombination enormously decreases the rate of valley crossing. Note that the arguments of this appendix are not sufficient to determine the prefactor in $\tau_{r>s}$ given in Equation 8; this requires the detailed calculation in Barton and Rouhani (1987).

Assumptions and consistency checks: It is worth reviewing the assumptions required in the lengthy and admittedly somewhat opaque derivation above. The most important assumption is that successful Ab lineages typically do not follow exponentially unlikely trajectories. This is because the probability of success for a lineage typically grows much slower than exponentially with the lineage size. It is straightforward to obtain rough estimates of the contributions of the exponentially unlikely trajectories to the overall probability of success and confirm that they are small. We have done so for the one regime ($r \gg s$, $\delta \gg \sqrt{s/N}$) where the probability of success increases faster than exponentially in the lineage size and where the exponentially unlikely lineages are therefore important. Note that this assumption is only for the initial, large Ab lineage; we do not make a similar assumption for the aB and AB lineages. Rather, we start with the full distribution of possible trajectories and find that the exponentially unlikely trajectories are unimportant, verifying this part of the heuristic argument in the main text.

We have made the additional assumption that the Ab lineage is relatively unaffected by the aB and AB lineages until the B allele is already increasing deterministically. As a check on this assumption, one can check that for $r \ll s$ the aB lineage dynamics are relatively unaffected by the much larger Ab lineage. While the two lineages must obviously interact to produce recombinants, the total number of individuals affected ($\lesssim 1/s$) is typically small compared not just to n_{Ab}, but also to n_{aB}. For $r \gg s$, one can check that the B-allele dynamics are only slightly affected by the more common A allele until $x_A x_B \approx 1/s$ (at which point the mutants begin to deterministically increase in frequency), and thus the B allele has only a slight effect on the A dynamics before this point.

The remaining assumptions (*e.g.*, that the aB and AB dynamics are fast compared to the Ab dynamics) are justified within the derivation, and the (numerous) remaining approximations have effects that are small in the asymptotic parameter regimes and straightforward to estimate.

Larger population sizes: Equations B23 and B25 are valid only for $N\mu \ll 1$. For $N\mu \gg 1$, there will typically be many single-mutant lineages in the population at one time, and the single-mutant frequencies initially increase approximately deterministically. For $r \ll s$, their numbers are given by (assuming symmetric loci)

$$n_{Ab}(t) \approx n_{aB}(t) \approx \frac{N\mu}{\delta}(1 - e^{-\delta t}).\tag{B26}$$

As long as N is not so large that the successful double mutant is produced instantly ($N\mu^2/s \ll 1$), τ is dominated by the waiting time for the first successful double mutant to be produced. Double mutants are produced by recombination at a rate $(r/N)n_{Ab}(t)n_{aB}(t)$ and by mutation at a rate $\mu(n_{Ab}(t) + n_{aB}(t))$, so the expected waiting time for the first successful one is

$$\tau \approx \int_0^\infty dt \exp\left[-s \int_0^t \left(\frac{r}{N} n_{Ab}(t') n_{aB}(t') + \mu(n_{Ab}(t') + n_{aB}(t'))\right) dt'\right]$$

$$\approx \begin{cases} (\mu\sqrt{Ns})^{-1} & \text{for } r, \quad \delta \ll \mu\sqrt{Ns} \text{ (effectively asexual, neutral single mutants)} \\ \frac{\delta}{2N\mu^2 s} & \text{for } \delta \gg r, \mu\sqrt{Ns} \text{ (effectively asexual, deleterious single mutants)} \\ (N\mu^2 rs)^{-1/3} & \text{for } r \gg \mu\sqrt{Ns}, \delta \ll (N\mu^2 rs)^{1/3} \text{ (effectively neutral single mutants)} \\ \frac{\delta^2}{N\mu^2 rs} & \text{for } r \gg \delta \gg (N\mu^2 rs)^{1/3} \text{ (deleterious single mutants)} \end{cases} \quad \text{(B27)}$$

For small r or large δ, the population is effectively asexual; this regime is discussed in WEISSMAN *et al.* (2009). For a more detailed discussion of the the third case, where recombination dominates and single mutants are effectively neutral, see CHRISTIANSEN *et al.* (1998). The fourth case is discussed in HADANY (2003). Note that for deleterious single mutants, τ has the same value as it does for deleterious single mutants with $N\mu \ll 1$ in Equations B21 and B23. In the neutral cases, on the other hand, not only is the expected time τ different, but also the rate of valley-crossing increases with time, so that the distribution of the valley-crossing time is no longer exponential.

For $r \gg s$, we can consider the population to have crossed the valley when the mean fitness of all the mutants is significantly greater than that of the wild type; once this is true, selection will rapidly drive the mutants to high frequencies. For $\delta \ll \sqrt{\mu s}$, single mutants are effectively neutral, and selection will begin to favor the mutants when $n_{AB} \approx 1/s$. This typically occurs at time $\tau \approx (\mu\sqrt{Ns})^{-1}$, approximately the same as for an asexual population. (Note that $\tau \gg 1/r$, so the population will be close to linkage equilibrium at $t \approx \tau$.) For $\delta \gg \sqrt{\mu s}$, the population cannot cross the valley until the single mutants drift far higher than their frequencies at mutation–selection balance, so the semideterministic approximation breaks down; this regime is discussed in the main text and in BARTON and ROUHANI (1987). Note that in this case, while there are many single-mutant lineages in the population at one time, only rare lineages drift to the frequencies needed to cross the valley, so the exponential dependence of τ on $N\delta^2/s$ is the same as for $N\mu \ll 1$.

APPENDIX C: ASYMMETRIC LOCI

For most of this article, we have assumed that the two loci are equivalent to reduce the number of parameters. Relaxing this assumption, *i.e.*, allowing $\mu_A \neq \mu_B$ and $\delta_A \neq \delta_B$, does not change the qualitative conclusions. It is straightforward to repeat the heuristic arguments in the main text or the calculations in APPENDIX B for any particular choice of parameters, although it is tedious to do so for all possible parameter combinations. Two points are worth briefly noting. First, asymmetry between the loci generally reduces the advantage of $r \ll s$ and increases the disadvantage of $r \gg s$ relative to asexual populations. This is because the asexual valley-crossing dynamics will be dominated by the more favorable single-mutant genotype, while the sexual dynamics will necessarily involve both genotypes. Second, for $r \gg s$ and very asymmetric loci, it is possible to be in the regime where valley crossing is exponentially suppressed even if one of the single mutants is nearly neutral. For instance, suppose δ_A is very small. As long as δ_B is large enough that the product satisfies $\delta_A\delta_B \gg \max\{\mu_A s, \mu_B s, s/N\}$, the population will be in the regime discussed in the main text, where to cross the valley it must cross a saddle point in the equilibrium allele frequency probability distribution. The saddle point in this case is at $x_A \approx \delta_B/s$, $x_B \approx \delta_A/s$, with Ab typically drifting to a much higher frequency than aB. The expected valley-crossing time increases exponentially with the population size, with

$$\log(\tau) \sim N\delta_A\delta_B/s.$$

APPENDIX D: SIMULATION METHOD

We evolved populations using time steps of $dt = 10^{-2}$ generations, during each of which the state of the population was updated via the following procedure:

1. The mean fitness $\bar{w}$ of the population was calculated, and the number n_{ij} of individuals with genotype ij was changed by random numbers of births $b_{ij} \sim \text{Binomial}(n_{ij}, (1 + w_{ij} - \bar{w})dt)$, deaths $d_{ij} \sim \text{Binomial}(n_{ij}, dt)$, and mutations $m_{ij} \sim \delta_{iA}\text{Binomial}(n_{aj}, \mu_A dt) + \delta_{jB}\text{Binomial}(n_{ib}, \mu_B dt)$ (where δ_{kl} is the Kronecker delta): $n^*_{i,j} = n_{ij} + b_{ij} - d_{ij} + m_{ij}$.

2. If the total number of individuals $N^* = \sum n_{ij}^*$ was different from N, the number of individuals of each genotype was multiplied by N/N^* and rounded to the nearest integer, and then the number of individuals with the most common genotype was adjusted to bring the total back to N.

3. A random number of pairs of individuals with complementary genotypes ($ab - AB$ or $Ab - aB$) underwent recombination. For each pair of genotypes, the number of recombinations was distributed as Binomial(n_{ij}, $rn_{i'j'}$ dt/N), with $n_{ij} \leq n_{i'j'}$.

Each run of the simulation proceeded until the population was fixed for AB or 2×10^7 generations had passed. Note that the simulation results for τ include the time it takes for AB to sweep through the population. This has a negligible effect, except for the smallest population sizes shown in Figure 5; in this case, we also included the sweep time in the curves showing the analytical predictions.
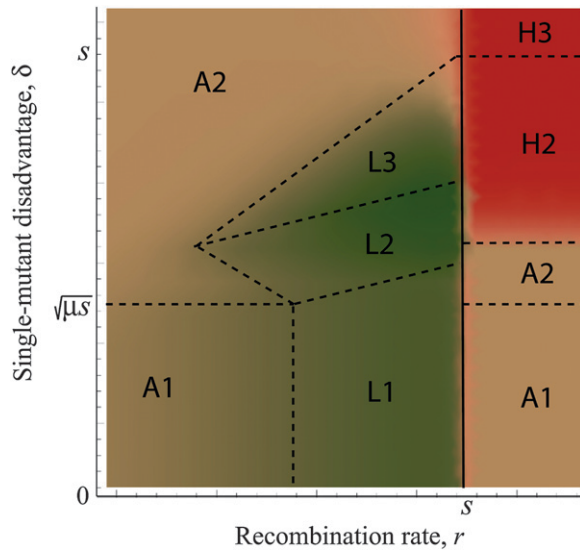

FIGURE B1.—Typical slice of parameter space for $1/\sqrt{\mu s} \ll N \ll 1/\mu$, showing regions where recombination has a strong effect on τ, the expected time for the population to cross the fitness valley. Recombination reduces τ in the green region, increases τ in the red region, and has only a minor effect on τ elsewhere. In the regions labeled A1 and A2, τ is given by the first and second lines of Equation B21 (the first and third lines of Equation 1 in the main text), respectively. In regions L1, L2, and L3, τ is given by the second, third, and fourth lines of Equation B23 (Equation 4 in the main text), respectively. In region H2, τ is given by Equation 8. In region H3, τ is approximately the expected time for a deleterious single mutant to drift to fixation, with $\log \tau \approx N\delta$. In regions A1 and L1, single mutants are effectively neutral.