

Fitness-valley crossing with generalized parent-offspring transmission

M.M. Osmond^{a,*}, S.P. Otto^a

^a*Department of Zoology, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada*

Abstract

Simple and ubiquitous gene interactions create rugged fitness landscapes composed of coadapted gene complexes separated by “valleys” of low fitness. Crossing such fitness valleys allows a population to escape suboptimal local fitness peaks to become better adapted. This is the premise of Sewall Wright’s shifting balance process. Here we generalize the theory of fitness-valley crossing in the two-locus, bi-allelic case by allowing bias in parent-offspring transmission. This generalization extends the existing mathematical framework to genetic systems with segregation distortion and uniparental inheritance. Our results are also flexible enough to provide insight into shifts between alternate stable states in cultural systems with “transmission valleys”. Using a semi-deterministic analysis and a stochastic diffusion approximation, we focus on the limiting step in valley crossing: the first appearance of the genotype on the new fitness peak whose lineage will eventually fix. We then apply our results to specific cases of segregation distortion, uniparental inheritance, and cultural transmission. Segregation distortion favouring mutant alleles facilitates crossing most when recombination and mutation are rare, i.e., scenarios where crossing is otherwise unlikely. Interactions with more mutable genes (e.g., uniparental inherited cytoplasmic elements) substantially reduce crossing times. Despite component traits being passed on poorly in the previous cultural background, small advantages in the transmission of a new combination of cultural traits can greatly facilitate a cultural transition. While peak shifts are unlikely under many of the common assumptions of population genetic theory, relaxing some of these assumptions can promote fitness-valley crossing.

Keywords: cultural evolution, cytonuclear inheritance, meiotic drive, peak shift, population genetics, valley crossing

1. Introduction

Epistasis and underdominance create rugged fitness landscapes on which adaptation may require a population to acquire multiple, individually-deleterious mutations that are collectively advantageous. Using the adaptive landscape metaphor, we say the population faces a fitness “valley” (Wright, 1932). Such valleys

*Corresponding author

Email addresses: mmosmond@zoology.ubc.ca (M.M. Osmond), otto@zoology.ubc.ca (S.P. Otto)

11 appear to be common in nature (Weinreich et al., 2005; Szendro et al., 2013, but see Carneiro and Hartl,
12 2010) and affect, among other things, speciation by reproductive isolation, the evolution of sex, the evolu-
13 tions of populations, and the predictability of evolution (Szendro et al., 2013). Here we are interested in
14 the speed and likelihood of fitness-valley crossing, which we determine by examining the first appearance
15 of an individual with the collectively advantageous set of mutations whose lineage will eventually spread to
16 fixation.

17 Believing epistasis to be ubiquitous, Sewall Wright (1931; 1932) formulated his “shifting balance theory”,
18 which describes evolution as a series of fitness-valley crossings. In phase one of the shifting balance process,
19 small, partially-isolated subpopulations (demes) descend into fitness valleys by genetic drift. The new
20 mutations are selected against when rare, as they will tend to occur alone as single deleterious alleles.
21 Eventually drift may allow the deleterious mutations to reach appreciable frequencies in at least one deme.
22 Once multiple synergistically-acting mutations arise together, they begin to be locally favoured by selection.
23 In phase two, these favoured combinations of mutations sweep to fixation, and those demes ascend the new
24 “fitness peak”. Finally, in phase three, the demes that reach the new fitness peak send out migrants whose
25 genes invade and fix in the remaining demes, eventually “pulling” the entire population up to the new fitness
26 peak. Our focus here is in the first appearance of a genotype on the new fitness peak whose lineage will
27 eventually fix, considering a single isolated deme. This is typically the longest stage of phases one and two
28 (Stephan, 1996) and hence is likely the limiting step in fitness-valley crossing.

29 Fitness-valley crossing has been investigated in a large number of theoretical studies. In the context of
30 multiple loci with reciprocal sign epistasis, the first appearance of the genotype with the best combination
31 of alleles has been the focus of a few studies (Phillips, 1996; Christiansen et al., 1998; Hadany, 2003; Hadany
32 et al., 2004; Weissman et al., 2009, 2010). Many authors have gone on to examine the remainder of phases
33 one and two (Crow and Kimura, 1965; Eshel and Feldman, 1970; Karlin and McGregor, 1971; Kimura,
34 1985; Barton and Rouhani, 1987; Kimura, 1990; Phillips, 1996; Michalakis and Slatkin, 1996; Stephan,
35 1996; Weinreich and Chao, 2005; Weissman et al., 2009, 2010), as well as phase three (Kimura, 1990; Crow
36 et al., 1990; Barton, 1992; Kondrashov, 1992; Phillips, 1993; Gavrillets, 1996; Hadany, 2003; Hadany et al.,
37 2004). Similar attention has been given to situations with a single underdominant locus (Slatkin, 1981;
38 Gillespie, 1984; Barton and Rouhani, 1993; Peck et al., 1998) or a quantitative trait (Lande, 1985a; Barton
39 and Rouhani, 1987; Rouhani and Barton, 1987a,b; Charlesworth and Rouhani, 1988; Barton and Rouhani,
40 1993). The theoretical and empirical support for Wright’s shifting balance process has been summarized and
41 debated (Coyne et al., 1997; Wade and Goodnight, 1998; Coyne et al., 2000; Whitlock and Phillips, 2000;
42 Coyne et al., 2000; Goodnight and Wade, 2000; Goodnight, 2013), and the general consensus appears to be
43 that, unless the valley is shallow (weakly deleterious intermediates), crossing a fitness valley is unlikely.

44 Despite the abundance of literature on fitness-valley crossing, the above studies all assume perfect
45 Mendelian inheritance. The question therefore remains: how robust are our ideas of fitness-valley crossing

46 to deviations from Mendelian inheritance? Specifically, how does transmission bias (e.g., meiotic drive or
47 uniparental inheritance) affect the speed and likelihood of valley crossing? Departing from strict Mendelian
48 inheritance also allows us to consider the idea of valley crossing in cultures, considering the spread of memes
49 (Dawkins, 1976) rather than genes. This simultaneously adds a level of complexity to current mathematical
50 models of cultural transmission, which typically consider only one cultural trait at a time (e.g., Cavalli-Sforza
51 and Feldman, 1981; but see, e.g., Ihara and Feldman, 2004; Creanza et al., 2012).

52 Transmission bias in the form of segregation distortion is likely to have a large effect on valley cross-
53 ing, as distortion represents a second level of selection (Sandler and Novitski, 1957; Hartl, 1970). Insight
54 into how segregation distortion affects valley crossing comes from models of underdominant chromosomal
55 rearrangements (mathematically equivalent to models with *one* diploid biallelic locus), which often find
56 meiotic drive to be a mechanism allowing fixation of a new mutant homokaryotype (Bengtsson and Bod-
57 mer, 1976; Hedrick, 1981; Walsh, 1982). Populations that have fixed alternate homokaryotypes produce
58 heterokaryotype hybrids, which have low viability and/or fertility; thus gene flow between these populations
59 is reduced. Segregation distortion is therefore thought to be a mechanism that promotes rapid speciation
60 (stasipatric speciation; White, 1978). Although the role of underdominance in chromosomal speciation has
61 recently been questioned (revisited in Rieseberg, 2001; Hoffmann and Rieseberg, 2008; Faria and Navarro,
62 2010; Kirkpatrick, 2010), it is hypothesized to be relevant in annual plants (Hoffmann and Rieseberg, 2008)
63 and appears to play a dominant role in maintaining reproductive isolation in sunflowers (Lai et al., 2005)
64 and monkey flowers (Stathos and Fishman, 2014).

65 Another common form of transmission bias is sex specific, with the extreme case being uniparental
66 inheritance. In genetic transmission, strict uniparental inheritance is common for organelle genomes, such
67 as the mitochondria, which is typically inherited from the mother. Uniparental inheritance will tend to
68 imply further asymmetries. For instance, the mutation rate of mitochondrial genes is estimated to be two
69 orders of magnitude larger than the mutation rate of nuclear genes in many animals (e.g., Linnane et al.,
70 1989). Higher mutation rates will likely facilitate crossing. That said, higher mutation rates in only one gene
71 may have limited effect because the production of double mutants by recombination will be constrained by
72 the availability of the rarer single mutant. Previous models of fitness-valley crossing have tended to ignore
73 asymmetries (but see Appendix C of Weissman et al., 2010).

74 Transmission bias is an integral characteristic of cultural transmission, where it is referred to as “cultural
75 selection” (Cavalli-Sforza and Feldman, 1981; Boyd and Richerson, 1985). However, to the best of our
76 knowledge, no attempts have been made to examine the evolution of cultural traits (memes) in the presence
77 of a “fitness” valley. Boyd (2001) reviews the genetic theory of the shifting balance, and notes that it could
78 be applied to culture, but no explicit cultural models were presented. Meanwhile, instances such as the
79 so-called “demographic transition” in 19th century western Europe, where societies transitioned from less
80 educated, large families to more educated, small families (Borgerhoff Mulder, 1998), suggest that alternate

81 combinations of cultural traits (e.g., ‘value of education’ and ‘family-size preference’) can be stable and that
82 peak shifts may occur in cultural evolution. In fact, alternate stable cultural states may be pervasive (Boyd
83 and Richerson, 2010), as alluded to by the common saying that people are “stuck in their ways.” Paradigm
84 shifts in the history of science (Kuhn, 1962) may provide further examples (Fog, 1999). Cultural peak shifts
85 can also be relatively trivial; for instance, changing the unit of time from seconds, minutes, and hours to
86 a decimal system is only advantageous if we also change units that are based on seconds, such as the joule
87 and volt (Fog, 1999).

88 Here we focus on a population genetic model with two bi-allelic loci under haploid selection in a randomly-
89 mating, finite population. This model can easily be reduced to a single-locus model with two alleles and
90 diploid selection, which is formally equivalent to a model of chromosomal rearrangements (e.g., a chromosome
91 has an inversion or not). Interpreting genes as memes produces a model of vertically-transmitted cultural
92 evolution. Our model incorporates both transmission bias and asymmetries in mutation and initial numbers
93 of single mutants. We first give a rough semi-deterministic sketch to develop some intuition, then follow
94 with a stochastic analysis using a diffusion approximation. Our analysis corresponds to the stochastic
95 simultaneous fixation regime of Weinreich and Chao (2005), and the neutral stochastic tunnelling and
96 deleterious tunnelling regimes of Weissman et al. (2010), where the appearance of the new, favourable,
97 and eventually successful “double mutant” occurs before the fixation of the neutral or deleterious “single
98 mutants”. Finally, we apply our results to the specific cases of segregation distortion, uniparental inheritance,
99 and cultural transmission.

100 We derive the expected time until the appearance of a double mutant whose lineage will fix when single
101 mutants are continuously generated by mutation from residents (the stochastic model assumes neutral
102 single mutants). We also use the stochastic model to derive the probability that a double mutant appears
103 and fixes given an initial stock of deleterious single mutants that is not replenished by mutation. Given
104 typical per-locus mutation rates, valley crossing is generally found to be a slow and unlikely outcome under
105 fair Mendelian transmission, even when single mutants are selectively neutral. Segregation distortion, in
106 favour of wild-type or mutant alleles, affects crossing most when recombination and mutation are rare,
107 the scenarios where crossing is otherwise unlikely. Cytonuclear inheritance allows increased mutational
108 asymmetries between the two loci; higher mutation rates lead to more single mutants and hence faster
109 valley crossing, but, when holding the average mutation rate constant, asymmetries hinder crossing by
110 reducing the probability that the single mutants recombine to produce double mutants. Finally, we show
111 that, when new cultural ideas or practices are not too poorly transmitted when arising individually within the
112 previous cultural background, a transmission advantage of the new combination greatly facilitates cultural
113 transitions.

114 2. Model and Results

115 Consider two loci, **A** and **B**, with x_{ij} the current frequency of genotype A_iB_j , where $i, j \in \{1, 2, \dots, p\}$ are
 116 the alleles carried by the individual. When an A_iB_j individual mates with an A_kB_l individual, they produce
 117 an A_mB_n offspring with probability $b_{ij}^{kl}(mn)$. Summing over all possible offspring types, $\sum_{m,n=1}^p b_{ij}^{kl}(mn) =$
 118 1. We can specify that the bottom index (here ij) denotes the genotype of the mother, while the top index
 119 (here kl) denotes the genotype of the father. As a consequence, transmission biases according to parental
 120 sex [$b_{ij}^{kl}(mn) \neq b_{kl}^{ij}(mn)$] are allowed. When considering sex-biased transmission we assume the frequencies
 121 x_{ij} are the same in females and males (i.e., no sex linkage and no sex-based differences in selection), which
 122 is automatically the case in hermaphrodites.

123 Random mating and offspring production is followed by haploid viability selection, which occurs immedi-
 124 ately before censusing. The population size, N , is constant and discrete, and generations are non-overlapping.
 125 Then the expected frequency of A_mB_n in the next generation, x'_{mn} , solves

$$Vx'_{mn} = w_{mn} \sum_{i,j,k,l=1}^p x_{ij} x_{kl} b_{ij}^{kl}(mn), \quad (1)$$

126 where $w_{mn} \geq 0$ is the relative viability of A_mB_n and V is the sum of the right hand side of Equation (1)
 127 over all genotypes, which keeps the frequencies summed to one.

128 Denote the probability that a mating between an A_iB_j mother and an A_kB_l father produces an A_mB_n
 129 offspring that survives one round of viability selection by $b_{ij}^{kl}(mn)^* = w_{mn}b_{ij}^{kl}(mn)$, where the asterisk
 130 indicates “after selection”. And let the average probability that a mating produces A_mB_n , regardless of
 131 which parent was which, be $\bar{b}_{ij}^{kl}(mn)^* = \frac{1}{2}w_{mn}[b_{ij}^{kl}(mn) + b_{kl}^{ij}(mn)]$. Then (as we will see below) selection on
 132 A_iB_j in a population of “residents” (A_1B_1) is described by $s_{ij} = 2\bar{b}_{11}^{ij}(ij)^* - 1$. Letting $w_{ij} = 1 + d_{ij} > 0$
 133 describe viability selection and $2\bar{b}_{11}^{ij}(ij) = 1 + k_{ij}$ describe transmission bias ($-1 \leq k_{ij} \leq 1$), then $s_{ij} =$
 134 $(1 + d_{ij})(1 + k_{ij}) - 1$. Here we define the relative fitness of genotype A_iB_j as $1 + s_{ij}$, which is determined by
 135 both viability and transmission. Thus defined, fitness is a measure of the “transmissibility” of a genotype as
 136 it includes several processes (e.g., viability, meiotic drive, recombination, mutation) that affect the number
 137 of offspring of a given genotype produced by a parent of that genotype. We will see that it is transmissibility
 138 that determines the dynamics of valley crossing.

139 Without mutation or recombination, fair transmission implies $k_{ij} = 0$, or $\bar{b}_{11}^{ij}(ij) = 1/2 \forall i \neq 1, j \neq 1$.
 140 In words, with fair transmission we expect half of all offspring from matings between A_1B_1 and A_iB_j to be
 141 of parental type A_iB_j . Sex-based inheritance is expected to arise in the form of $b_{11}^{ij}(ij) = 1 - b_{ij}^{11}(ij)$ [e.g.,
 142 maternal inheritance implies $b_{11}^{ij}(ij) = 1$ and $b_{ij}^{11}(ij) = 0$], which does not directly impose selection as $k_{ij} = 0$.
 143 Segregation distortion can, however, impose selection. For example, ignoring mutations, if the A_2 allele is
 144 more likely to be transmitted than the A_1 allele (in a B_1 background) we would have $\bar{b}_{11}^{21}(21) > 1/2$, giving
 145 $k_{21} > 0$. Interpreting genes as memes, transmission bias k_{ij} determines the strength of “cultural selection”

146 (*sensu* Cavalli-Sforza and Feldman, 1981) on meme combination $A_i B_j$. Previous work on multi-locus peak
 147 shifts has assumed that bias does not influence selection ($k_{ij} = 0$) and that maternal and paternal types are
 148 equally transmitted [$b_{11}^{ij}(ij) = b_{ij}^{11}(ij) = 1/2 \forall i \neq 1, j \neq 1$].

149 Here we focus on bi-allelic loci ($p = 2$). We are specifically interested in the case where, in a popula-
 150 tion composed entirely of residents, “single mutants” ($A_2 B_1$ and $A_1 B_2$) are selected against while “double
 151 mutants” ($A_2 B_2$) are selectively favoured: $s_{21}, s_{12} < 0 < s_{22}$.

152 Given that the population is composed primarily of residents, with no double mutants as of yet, the
 153 population faces a fitness valley. The valley can be created by differences in viability alone, or it can be
 154 created by differences in transmission, or both. Here we focus on the limiting step in the peak-shift process,
 155 the probability and expected time until a double mutant arises whose lineage will eventually fix. Following
 156 the lead of Christiansen et al. (1998), we begin by developing a rough semi-deterministic analysis to gain
 157 intuition. A stochastic analysis follows. Table 1 provides a summary of notation and a supplementary
 158 *Mathematica* file gives a more detailed derivation of the results.

159 TABLE 1 HERE

160 2.1. Semi-deterministic analysis

161 2.1.1. Single mutant dynamics

162 Selection against single mutants keeps their frequencies (x_{21} and x_{12}) small. Let these frequencies be
 163 proportional to some small number $\epsilon \ll 1$. Let the probability that an offspring inherits an allele that
 164 neither parent possesses [i.e., mutation; e.g., $b_{11}^{11}(21)$] be of the same small order ϵ . Then, for large N , the
 165 frequencies of the single mutants in the next generation are

$$x'_{ij} \approx \frac{w_{ij}}{V} \left[b_{11}^{11}(ij) + 2\bar{b}_{11}^{ij}(ij)x_{ij} \right] + O(\epsilon^2), \quad (2)$$

166 where $i \neq j$ and $O(\epsilon^2)$ captures terms of order ϵ^2 and smaller.

167 We will write $\mu_{ij}^{kl}(mn)^* = \bar{b}_{ij}^{kl}(mn)^*$ when $m \notin \{i, k\}$ or $n \notin \{j, l\}$ to highlight the fact that a mutation
 168 has occurred. Then, ignoring $O(\epsilon^2)$, the frequencies of single mutants, which are initially absent [$x_{ij}(0) = 0$],
 169 in generation t are

$$x_{ij}(t) \approx \begin{cases} \mu_{11}^{11}(ij)^* \left[(1 + s_{ij})^t - 1 \right] s_{ij}^{-1} & : s_{ij} \neq 0 \\ \mu_{11}^{11}(ij)^* t & : s_{ij} = 0 \end{cases} \quad (3)$$

170 Viability and transmission are thus coupled together (in s_{ij}) throughout our results, and it is primarily the
 171 total amount of selection on $A_i B_j$ in a population of residents (s_{ij}) that determines the dynamics. [As a
 172 technical aside, this is not true in the first generation that mutants appear, via $\mu_{ij}^{kl}(mn)^*$, but this is simply
 173 because of the order of the life cycle chosen, where these mutants experience viability selection, but not
 174 transmission biases, when they first occur.]

175 Equation (3) assumes the normalizing factor V remains near 1 over the t generations, which is the

176 case when single mutants are rare, as is generally true when single mutants are selected against, $s_{ij} <$
 177 $0 \forall i \neq j$. When $s_{ij} < 0$ and there has been a sufficiently long period of selection, $t > -1/s_{ij}$, the
 178 single mutant frequencies approach mutation-selection balance $x_{ij}(t) \approx -\mu_{11}^{11}(ij)^*/s_{ij}$. This assumes the
 179 probability of mutation, $\mu_{11}^{11}(ij)^*$, is small relative to the strength of selection, s_{ij} . We next derive a semi-
 180 deterministic solution for the crossing time, T , given mutation-selection balance is reached. In Appendix
 181 [A](#) we follow [Christiansen et al. \(1998\)](#) to derive the semi-deterministic crossing time when crossing occurs
 182 before mutation-selection balance is reached; this occurs when $-s_{ij}T \ll 1$, which can only be the case if
 183 the valley is shallow, $-s_{ij} \ll 1$.

184 2.1.2. Waiting time for first successful double mutant

185 We now turn to calculating the waiting time until a double mutant that is able to establish first arises.
 186 Assume the probability two residents mate to produce a double mutant (i.e., a double mutation), $b_{11}^{11}(22)$, is
 187 very rare, on the order of ϵ^2 . Then the expected frequency of double mutants in the next generation before
 188 selection, assuming single mutant are rare and there are currently no double mutants $x_{22} = 0$, is

$$x'_{22} = \left[\mu_{11}^{11}(22) + 2\mu_{11}^{21}(22)x_{21} + 2\mu_{11}^{12}(22)x_{12} + 2r_{21}^{12}(22)x_{21}x_{12} \right] + O(\epsilon^3), \quad (4)$$

189 where we write $r_{21}^{12}(22) = \bar{b}_{21}^{12}(22)$ to highlight the fact that a double mutant has effectively been produced
 190 by recombination. The expected frequency of double mutants (Equation 4) is measured before viability
 191 selection to avoid artificially adjusting the double mutant frequency by its viability difference before it
 192 appears.

193 In a truly deterministic model ($N \rightarrow \infty$) double mutants are present at frequency x'_{22} after a single
 194 bout of reproduction. However, assuming no double mutants have yet appeared, we can use $x_{22}(t)$ as a
 195 rough approximation for the probability of a double mutant first arising in generation t ([Christiansen et al.](#),
 196 [1998](#)). Summing t from 0 to t' gives the cumulative probability of observing a double mutant in any of the t'
 197 generations. The generation T' at which the cumulative probability reaches $1/N$ can be used as an estimate
 198 of the time we expect to wait until the first double mutant has arisen ([Christiansen et al., 1998](#)).

199 Here we are more interested in the waiting time until the first *successful* double mutant appears (i.e.,
 200 one whose lineage will eventually fix). We therefore want to multiply the probability that a double mutant
 201 appears at time t , $x_{22}(t)$, by the probability it will fix before taking the sum over t . Using Kimura's ([1962](#))
 202 approximation, the probability a double mutant fixes is

$$u_{22} = \frac{1 - e^{-2s_{22}}}{1 - e^{-2Ns_{22}}}. \quad (5)$$

203 With a weak double mutant advantage, $0 < s_{22} \ll 1$, in a large population, $Ns_{22} \gg 1$, Equation (5)
 204 simplifies to the familiar $2s_{22}$ ([Haldane, 1927](#)).

205 The selection coefficient s_{22} can be calculated from the number of double mutant offspring a newly

206 arisen double mutant is expected to leave in the next generation, given that the mean number of offspring
207 per individual is one, such that the population size is constant. This expectation, $1 + s_{22}$, is the probability
208 of mating with a given type, multiplied by the probability of producing a double mutant offspring, multiplied
209 by the probability of surviving to the next generation, summed over all possible matings

$$1 + s_{22} = \sum_{i,j=1}^2 2\bar{b}_{ij}^{22}(22)^* x_{ij}, \quad (6)$$

210 where $x_{22} = 0$ in the remaining population (i.e., the double mutant does not mate with itself). Without
211 transmission bias, mutation, or recombination, $\bar{b}_{ij}^{22}(22) = 1/2 \forall i, j$ and Equation (6) reduces to the familiar
212 $s_{22} = w_{22} - 1$. Here we allow bias, mutation, and recombination, and assume single mutants are sufficiently
213 rare, giving $s_{22} \approx 2\bar{b}_{11}^{22}(22)^* - 1$. This implies that selection on the double mutant (including transmission) is
214 constant over time and that fixation depends only on its dynamics in a population composed almost entirely
215 of residents. With recombination and otherwise fair transmission we have $\bar{b}_{11}^{22}(22) = (1 - r)/2$, where r
216 is the probability of recombination between a double mutant and a resident. Writing $w_{22} = 1 + s$ and
217 assuming both s and r are small, recovers the well-known first-order approximation $s_{22} = s - r$ (Crow and
218 Kimura, 1965). This expression highlights the fact that recombination can reduce the probability of fixation
219 by breaking up favourable gene combinations (Crow and Kimura, 1965).

220 When selection is strong and mutation is rare, relative to the strength of genetic drift, the time to fixation
221 is dominated by the time to the arrival of a successful mutant (Gillespie, 1984; Weinreich and Chao, 2005;
222 Weissman et al., 2010). The waiting time until the first successful double mutant, which we derive below,
223 therefore well approximates the fixation time of a double mutant within a population when double mutants
224 are advantageous but rarely produced, $x'_{22} \ll 1/N < s_{22}$.

225 *Crossing time given mutation-selection balance.* When enough time has passed ($t > -1/s_{ij}$) the single-
226 mutant frequencies approach mutation-selection balance (MSB), $x_{ij}(t) \approx -\mu_{11}^{11}(ij)^*/s_{ij}$. Using these fre-
227 quencies in Equation (4) gives the expected frequency of double mutants in the next generation, which does
228 not change until a double mutant arises, i.e., $x_{22}(t) = x'_{22} \forall t$. Summing $u_{22}x'_{22}$ over T_{MSB} generations,
229 setting equal to $1/N$, and solving for T_{MSB} gives an estimate of the number of generations we expect to
230 wait for a successful double mutant to arise when beginning from mutation-selection balance

$$\begin{aligned}
 T_{MSB} \approx & \frac{1}{u_{22}N} \left[\left(1 + \frac{\mu_{11}^{11}(21)^*}{s_{21}} + \frac{\mu_{11}^{11}(12)^*}{s_{12}} \right)^2 \mu_{11}^{11}(22) \right. \\
 & - \left(1 + \frac{\mu_{11}^{11}(21)^*}{s_{21}} + \frac{\mu_{11}^{11}(12)^*}{s_{12}} \right) \left(\frac{\mu_{11}^{11}(21)^*}{s_{21}} \right) 2\mu_{11}^{21}(22) \\
 & - \left(1 + \frac{\mu_{11}^{11}(21)^*}{s_{21}} + \frac{\mu_{11}^{11}(12)^*}{s_{12}} \right) \left(\frac{\mu_{11}^{11}(12)^*}{s_{12}} \right) 2\mu_{11}^{12}(22) \\
 & + \left(\frac{\mu_{11}^{11}(21)^*}{s_{21}} \right)^2 \mu_{21}^{21}(22)^* + \left(\frac{\mu_{11}^{11}(12)^*}{s_{12}} \right)^2 \mu_{12}^{12}(22) \\
 & \left. + \left(\frac{\mu_{11}^{11}(21)^*}{s_{21}} \right) \left(\frac{\mu_{11}^{11}(12)^*}{s_{12}} \right) 2r_{21}^{12}(22) \right]^{-1} - 1. \tag{7}
 \end{aligned}$$

231 In our numerical examples, we will track the waiting time until a successful double mutant arises in a
 232 population that has recently established and is fixed for the resident type (e.g., following a bottleneck or
 233 a founder event). This time can be approximated by the time that it takes to reach mutation-selection
 234 balance, T_0 , and the establishment time once there

$$T \approx T_0 + T_{MSB}. \tag{8}$$

235 Here we use $T_0 = \max\{\frac{1}{-s_{21}}, \frac{1}{-s_{12}}\}$. As the deleterious single mutants approach neutrality ($s_{ij} \rightarrow 0^- \forall i \neq j$)
 236 the waiting time *from* mutation-selection balance, T_{MSB} , decreases (because there are more single mutants
 237 segregating), but the waiting time *to* mutation-selection balance, T_0 , increases dramatically because it takes
 238 longer to produce the higher segregating frequencies of single mutants. As $-s_{ij}$ becomes small enough such
 239 that $T < -1/s_{ij}$ the approximation breaks down and we must use the non-equilibrium solution derived in
 240 Appendix A.

241 With symmetric Mendelian assumptions, weak selection on single mutants ($\delta = 1 - w_{ij} \forall i \neq j$), rare
 242 mutation (μ), and infrequent recombination [such that $u_f \approx 2(s - r) \approx 2s$], the rate of production of
 243 successful double mutants from mutation selection balance (Equation 7) is

$$T_{MSB}^{-1} \approx \frac{2sN\mu^2r}{\delta^2}, \tag{9}$$

244 aligning with equation 4 in Weissman et al. (2010, see supplementary *Mathematica* file). This result preforms
 245 well when $T_{MSB}^{-1} < \delta$, or, equivalently, when $\sqrt[3]{2sN\mu^2r} < \delta$.

246 2.2. Stochastic analysis

247 2.2.1. Markov process

248 Fitness-valley crossing is naturally a stochastic process. We thus now consider the Wright-Fisher model,
 249 where the next generation is formed by choosing N offspring, with replacement, from a multinomial distri-
 250 bution with frequency parameters x'_{ij} (Equation 1). Let the number of A_2B_1 and A_1B_2 single mutants in
 251 generation t be i_t and j_t , respectively. Given that there are currently no double mutants, we have $N - i_t - j_t$

252 resident individuals and we let $X(t) = (i_t, j_t)$ describe the state of the system in generation t . Let the
 253 expected frequencies in the $t + 1$ generation, conditional on $X(t) = (i, j)$, be $x'_{kl}(i, j) = x'_{kl}$, with $x_{22} = 0$.
 254 The transition probabilities to states without double mutants are then

$$P_{ij}^{kl} = P\{X(t+1) = (k, l) \mid X(t) = (i, j)\} = \binom{N}{k, l, N-k-l} (x'_{21})^k (x'_{12})^l (x'_{11})^{N-k-l}. \quad (10)$$

255 Note that summing over all $k, l \in \{0, 1, \dots, N\}$ gives $(1 - x'_{22})^N$, the probability that no double mutant is
 256 sampled. Equation (10) describes a sub-stochastic transition matrix for the Markov process.

257 Next, let H be the state with any positive number of double mutants. We then have the transition
 258 probabilities $P_{ij}^H = 1 - (1 - x'_{22})^N \approx Nx'_{22}$, where the approximation assumes a small expected frequency
 259 of double mutants in the next generation, $x'_{22} \ll 1$. To calculate the waiting time until the first *successful*
 260 double mutant, we replace P_{ij}^H with $\tilde{P}_{ij}^H = P_{ij}^H u_{22} \approx Nx'_{22}u_{22}$, ignoring the segregation of double mutants
 261 when lost. H is now the state with any positive number of successful double mutants. Dividing each
 262 x'_{ij} in Equation (10) by the probability a double mutant does not arise $(1 - x'_{22})$ and multiplying by the
 263 probability a double mutant does not arise and fix $(1 - x'_{22}u_{22})$ ensures the columns sum to one. To complete
 264 the transition matrix we make H an absorbing state: $P_H^H = 1$ and $P_H^{ij} = 0$.

265 We can describe this process, in part, by the moments for the change in number of single mutants,
 266 conditional on the process not being killed by a successful double mutant. The n^{th} moment for the change
 267 in the number of A_2B_1 individuals, $\Delta i = i_{t+1} - i_t$, is

$$E[(\Delta i)^n \mid i_t = i] = \sum_{k=0}^N (k - i)^n \binom{N}{k} \left(\frac{x'_{21}}{1 - x'_{22}u_{22}} \right)^k \left(\frac{x'_{12} + x'_{11}}{1 - x'_{22}u_{22}} \right)^{N-k}. \quad (11)$$

268 Similar equations can be computed for the change in the number of A_1B_2 individuals, $\Delta j = j_{t+1} - j_t$.

269 To make analytic progress we use the moment equations to approximate the Markov chain with a diffusion
 270 process (Karlin and Taylor, 1981, Ch. 15). We do so by taking the large population limit ($N \rightarrow \infty$) while
 271 finding the appropriate scalings to ensure finite drift and diffusion terms (Appendix B).

272 2.2.2. Crossing time with neutral single mutants

273 The diffusion process yields a partial differential equation describing the expected time until a successful
 274 double mutant arises given that we begin with $N^\beta y$ individuals of type A_2B_1 and $N^\beta z$ individuals of type
 275 A_1B_2 (Christiansen et al., 1998)

$$\frac{1}{2}\sigma_Y^2(y) \frac{\partial^2 \tilde{T}(y, z)}{\partial y^2} + \frac{1}{2}\sigma_Z^2(z) \frac{\partial^2 \tilde{T}(y, z)}{\partial z^2} + \mu_Y(y) \frac{\partial \tilde{T}(y, z)}{\partial y} + \mu_Z(z) \frac{\partial \tilde{T}(y, z)}{\partial z} - \kappa(y, z) \tilde{T}(y, z) = -1, \quad (12)$$

276 where $\tilde{T}(y, z)$ refers to time scaled in units of N^β generations (parameters defined in Table 1 and Appendix
 277 B). In Appendix C we solve Equation (12) under the two scenarios explored in Christiansen et al. (1998):
 278 with and without recombination from neutral single mutants to double mutants when the population begins

279 with only residents, but here generalized to allow unequal mutation rates and sex-biased transmission. While
 280 the neutrality assumption precludes the existence of a fitness valley, it provides a minimum for the expected
 281 time to observe a successful double mutant. Previous studies have suggested that fitness valleys will only
 282 be crossed if single mutants are nearly neutral (e.g., [Walsh, 1982](#)).

283 2.2.3. Probability of crossing from standing variation

284 The diffusion process can also be used to describe the production of successful double mutants from an
 285 initial stock of single mutants (i.e., evolution from standing variation). Specifically, assuming that residents
 286 don't mutate [$b_{11}^{11}(12) = b_{11}^{11}(21) = b_{11}^{11}(22) = 0$] the process has two absorbing states, fixation of A_1B_1 and
 287 fixation of A_2B_2 (a successful double mutant appears and the process is killed). The probability of fixation
 288 of residents is the solution, $u(y, z)$, of ([Karlin and Taylor, 1981](#))

$$\frac{1}{2}\sigma_Y^2(y)\frac{\partial^2 u(y, z)}{\partial y^2} + \frac{1}{2}\sigma_Z^2(z)\frac{\partial^2 u(y, z)}{\partial z^2} + \mu_Y(y)\frac{\partial u(y, z)}{\partial y} + \mu_Z(z)\frac{\partial u(y, z)}{\partial z} - \kappa(y, z)u(y, z) = 0, \quad (13)$$

289 with terms defined in Appendix B. The probability that a successful double mutant arises is therefore
 290 $1 - u(y, z)$. [Karlin and Tavaré \(1981\)](#) used a similar equation to find the probability of detecting a lethal
 291 homozygote in the one locus, diploid case with Mendelian transmission.

292 *Deleterious single mutants without recombination.* With no recombination from single mutants to double
 293 mutants [$r_{21}^{12}(22) = 0$] we have scaling parameter $\beta = 1/2$. Then, with equal selection on single mutants and
 294 some mutational symmetry between the two loci (see supplementary *Mathematica* file), the single mutants
 295 are equivalent and we can concern ourselves with only their sum $\xi = y + z$. Equation (13) then collapses to

$$\frac{1}{2}\xi\frac{d^2 u(\xi)}{d\xi^2} + S_m\xi\frac{du(\xi)}{d\xi} - u_{22}w_{22}\left[B_{11}^m(22) + B_m^{11}(22)\right]\xi u(\xi) = 0, \quad (14)$$

296 where $S_m = s_{21}N^\beta = s_{12}N^\beta$ is scaled selection on single mutants and $B_{11}^m(22) + B_m^{11}(22) = [b_{11}^{21}(22) =$
 297 $b_{21}^{11}(22)]N^\beta = [b_{11}^{12}(22) = b_{12}^{11}(22)]N^\beta$ is the scaled mutation probability from single mutants to double
 298 mutants.

299 The boundary conditions are $u(0) = 1$ and $u(\infty) = 0$. Solving the boundary value problem gives the
 300 probability of a double mutant appearing when starting with $n_0 = i_0 + j_0$ single mutants

$$1 - u(n_0) = 1 - \exp\left[n_0\left[-s_m - \sqrt{s_m^2 + 2u_{22}2\mu_{11}^m(22)^*}\right]\right], \quad (15)$$

301 where $s_m = s_{21} = s_{12}$ is the total strength of selection on each single mutant type. Setting $n_0 = 1$ gives the
 302 probability a newly arisen single mutant will begin a lineage which eventually produces a successful double
 303 mutant.

304 Interestingly, Equation (15) does not depend strongly on population size, N . Without recombination
 305 double mutants are primarily produced by mutations from single mutants, which are rare and hence always

mate with one of the large number of residents. In other words, the production of A_2 and B_2 alleles does not rely on the number of residents but only on the dynamics of the rare single mutants.

Deleterious single mutants with recombination. Finally, we examine the probability of a successful double mutant appearing when there is recombination between deleterious single mutants, $r_{21}^{12}(22) > 0$. With sufficiently strong selection against single mutants the single mutant frequencies scale as $c_n y \approx z$ when we begin with initial frequencies $c_n y(0) = z(0)$ and both single mutants are under the same selection pressure, $S_{21} = S_{12}$. Then, without mutation from residents to single mutants, Equation (13) collapses to

$$\frac{1}{2(1+c_n)} \xi \frac{d^2 u(\xi)}{d\xi^2} + S_{21} \xi \frac{du(\xi)}{d\xi} - u_{22} c_n 2r_{21}^{12}(22)^* \xi^2 u(\xi) = 0, \quad (16)$$

where $\xi = c_n y = z$.

With boundary conditions $u(0) = 1$ and $u(\infty) = 0$ the probability of valley crossing is

$$1 - u(i_0, N) = 1 - \exp \left[- (1 + c_n) i_0 s_{21} \right] \frac{A_i \left[\frac{N(1+c_n)^2 (s_{21})^2 + i_0 2u_{22} c_n (1+c_n) 2r_{21}^{12}(22)^*}{N^{1/3} [2u_{22} c_n (1+c_n) 2r_{21}^{12}(22)^*]^{2/3}} \right]}{A_i \left[\frac{N(1+c_n)^2 (s_{21})^2}{N^{1/3} [2u_{22} c_n (1+c_n) 2r_{21}^{12}(22)^*]^{2/3}} \right]}, \quad (17)$$

where A_i is the Airy function. Equation (17) extends the one-locus diploid result with Mendelian transmission (equation 28 in [Karlin and Tavaré, 1981](#)) by allowing unequal single mutant frequencies ($c_n \neq 1$) while also incorporating transmission bias, recombination, and double mutant fitness. Equation (17) well approximates the Mendelian simulation results of [Michalakis and Slatkin \(1996\)](#), see supplementary *Mathematica* file).

When $(s_{21})^2$ and i_0 are small, we have the first order approximation

$$1 - u(i_0, N) = i_0 \left[(1 + c_n) s_{21} + \frac{3^{1/3} \Gamma[2/3]}{\Gamma[1/3]} \left(\frac{2u_{22} c_n (1 + c_n) 2r_{21}^{12}(22)^*}{N} \right)^{1/3} \right], \quad (18)$$

which is valid only when the term in the large square brackets is positive. Equation (18) can be used to show that when holding the initial number of single mutants, $(1 + c_n) i_0$, constant, the probability the double mutant fixes is maximized when there are equal numbers of single mutants, $c_n = 1$. This is because recombination is most efficient in creating double mutants when single mutants are equally frequent.

2.3. Three scenarios

We next apply our results to three different scenarios: segregation distortion, cytonuclear inheritance, and cultural transmission.

2.3.1. Segregation distortion

One form of segregation distortion, found in the heterothallic fungi *Neurospora intermedia*, is autosomal killing ([Burt and Trivers, 2006](#)). In heterozygotes, the presence of a “killer” allele results in the death of a proportion of the spores that contain the wild-type (“susceptible”) allele, leading to a $(1 + k)/2$ frequency of

332 the killing allele at fertilization, $0 < k \leq 1$. Letting A_2 and B_2 represent the killing alleles, and, for the sake
333 of exploration, assuming that cells with one killing allele are functionally equivalent to cells with two, the
334 transmission probabilities are shown in Table 2. The Mendelian case is given by $k = 0$. When $-1 \leq k < 0$
335 the allele identities are reversed: A_1 and B_1 are killers and A_2 and B_2 are susceptibles.

336 INSERT TABLE 2 HERE

337 Since segregation distortion imposes selection on single mutants, we can only investigate the effect of seg-
338 regation distortion on valley crossing with the semi-deterministic crossing time estimates allowing selection
339 on single mutants (Equations 8 and A3) and with the crossing probability estimates from standing variation
340 (Equations 15 and 17).

341 Figure 1 shows the crossing time as a function of the probability of recombination, and how segregation
342 distortion affects this time. Simulations (X 's) well match the numerical solution (Equation A1; dots) and
343 the MSB approximation (Equation 8; solid curves in top panel) over the range of parameters tested. When
344 valley crossing occurs before reaching MSB (bottom panel) a transition occurs between when mutation
345 drives crossing (dashed line; Equation A2) and when recombination does (solid curves; Equation A3), here
346 approximately $r \approx 10^{-4}$. The largest effect of segregation distortion occurs when the crossing time is long,
347 the scenario in which single-mutants must persist the longest before a successful double mutant appears.
348 In addition, observe that as the probability of recombination, r , increases above a critical value such that
349 $s_{22} < 0$, the double mutant is broken apart faster than its selective advantage and valley crossing takes
350 longer [equations B21 and B25 in Weissman et al., 2010 approximate the crossing times with no segregation
351 distortion ($k = 0$) when $s_{22} < 0$; see also Lynch, 2010; Altland et al., 2011].

352 INSERT FIGURE 1 HERE

353 Figure 2 shows the probability of crossing from standing variation. Again, segregation distortion has a
354 large effect when mutation (top panel) and recombination (bottom panel) are rare, the conditions under
355 which single mutants must persist the longest before a successful double mutant is formed. When crossing
356 occurs by recombination our analytical approximation (Equation 17) overestimates the probability of crossing
357 (bottom panel), especially when the initial number of single mutants is small and therefore subject to strong
358 stochasticity (results not shown). This occurs because the assumption that the ratio of single mutant
359 frequencies in these simulations remains roughly $c_n = 1$ is violated, reducing the probability that double
360 mutants are formed by recombination.

361 INSERT FIGURE 2 HERE

362 2.3.2. Cytonuclear inheritance

363 We next explore how fitness-valley crossing is affected by uniparental inheritance of one of the traits. This
364 might occur if, for example, there was reciprocal sign epistasis between cytoplasmic and nuclear loci. Without
365 loss of generality we assume that the **B** trait is always inherited from the mother. For simplicity we assume

366 individuals are hermaphroditic. Here we can use only those results that allow recombination (Equations
367 8, A3, C5, and 17), as cytoplasmic and nuclear elements are expected to be inherited independently (i.e.,
368 $r_{21}^{12}(22) + r_{12}^{21}(22) = 1/2$).

369 One likely implication of uniparental transmission is asymmetric mutation probabilities. For instance,
370 in animals the mitochondrial mutation rate is two orders of magnitude larger than typical nuclear rates
371 (Linnane et al., 1989). Let μ be the mutation probability in the biparentally inherited **A** trait and ν be
372 the mutation probability in the uniparentally inherited **B** trait, with $c_\mu = \nu/\mu$ the ratio of uniparental to
373 biparental mutation probabilities. The transmission probabilities are shown in Table 3.

374 INSERT TABLE 3 HERE

375 The top panel of Figure 3 shows the crossing time as a function of the mutation probability in the **B**
376 locus, ν . Increasing ν increases the rate at which single and double mutants are created, aiding valley
377 crossing. The bottom panel of Figure 3 shows the crossing time as a function of the ratio of the mutation
378 probabilities at the two loci, c_μ , while holding the average mutation probability, $(\mu + \nu)/2 = \mu(1 + c_\mu)/2$,
379 constant. When holding the average mutation probability constant the time to fixation is minimized when
380 $\nu = \mu$ because single mutant types are then equally frequent, increasing the chances they mate with one
381 another to produce a double mutant by recombination. As c_μ departs from one, the mutation rate at one
382 of the loci becomes small, causing those single mutants to become rare. The highly stochastic nature of the
383 rare single mutant frequencies causes our semi-deterministic (Equation A3) and stochastic (Equation C5)
384 approximations to underestimate the crossing time and, instead, the single mutants first reach mutation-
385 selection balance (Equation 8; dashed gray curve).

386 INSERT FIGURE 3 HERE

387 Given that crossing occurs by recombination from standing variation (Equation 17), asymmetric mutation
388 rates have little effect given a particular starting population (i_0, j_0) . However, standing variation will also
389 tend to vary in proportion to mutation rates, implying that uniparental inheritance will cause differences in
390 the initial numbers of the two single mutants, which can have a large effect. Let c_μ now also determine the
391 ratio of the initial numbers of single mutants, $c_\mu = c_n = j_0/i_0$. Figure 4 shows the probability of crossing
392 from standing variation as a function of c_μ . When we hold i_0 and μ constant and increase j_0 and ν (grey
393 curve), the probability of crossing increases with c_μ as there are then more single mutants segregating. When
394 we instead hold the total initial number of single mutants $(i_0 + j_0)$ and the average mutation probability
395 $[(\mu + \nu)/2]$ constant (black curve), the probability of crossing is maximized at $c_\mu = 1$ because the single
396 mutants are then equally frequent and hence more likely to mate with one another and produce a double
397 mutant through recombination.

398 INSERT FIGURE 4 HERE

399 *2.3.3. Cultural inheritance*

400 Finally, we remove Darwinian selection, such that transmission bias alone determines the dynamics, and
401 interpret the model in a cultural context. For the sake of exposition we consider only one simplified case of
402 cultural transmission. Let trait combinations with only one new trait (A_2B_1 and A_1B_2) be inherited relative
403 to the previous combination (A_1B_1) with probability q . Let the new combination of cultural traits (A_2B_2)
404 be inherited relative to the previous combination with probability p . We are most interested in the case of a
405 “transmission valley”, where the previous combination of traits is transmitted more effectively than mixed
406 combinations of new and old ($q < 1/2$), but the all-new combination spreads even more effectively than
407 the previous combination ($p > 1/2$). We assume that parental trait combinations can be broken up with
408 probability r and mutation occurs with probability μ . The transmission probabilities are shown in Table 4.

409 INSERT TABLE 4 HERE

410 Figure 5 shows that the crossing time is substantially faster when the new combination of traits has a
411 stronger transmission advantage (T decreases with p ; compare thick curve with thin). Nevertheless, even
412 combinations that are transmitted very effectively (thick curve) spread very slowly when their component
413 traits are passed on poorly in the previous cultural background ($q \ll 1/2$). In particular, the crossing time
414 increases most quickly as q decreases from $1/2$, demonstrating that slight biases in the transmission of the
415 new traits when arising within the previous cultural background have a strong influence on the spread of
416 new combinations of cultural traits, effectively preventing establishment if $q \ll 1/2$.

417 INSERT FIGURE 5 HERE

418 Figure 6 shows the probability of crossing from standing variation. In this case, with such a large
419 mutation rate, crossing can be more likely by mutation (Equation 15) than by recombination (Equation 17).
420 Recombination has the added effect of breaking up the new combination of traits, reducing the probability
421 of crossing. With a lower mutation rate crossing is most likely with moderate amounts of recombination
422 (e.g., Figure 1). Figure 6 again shows that the transmission advantage of the new combination of traits
423 (p ; compare thick lines to thin) and slight biases in the transmission of new traits in the previous cultural
424 background ($q \approx 1/2$) greatly influence the probability that a new combination of cultural traits successfully
425 spreads.

426 INSERT FIGURE 6 HERE

427 **3. Discussion**

428 Our results support the general consensus that, given reasonable population sizes and per locus per
429 generation mutation rates, crossing a particular fitness valley by genetic drift is typically a slow and unlikely
430 event (Crow and Kimura, 1965; Bengtsson and Bodmer, 1976; Lande, 1979; Hedrick, 1981; Walsh, 1982;
431 Lande, 1985b; Michalakis and Slatkin, 1996; Phillips, 1996; Coyne et al., 1997). For example, with a per locus
432 per generation mutation probability of $\mu = 10^{-8}$, a double mutant viability of $w_{22} = 1.01$, recombination

433 between the two loci with probability $r = 0.01$, and a population size of $N = 10^4$, the waiting time for a
434 successful double mutant, in the best case scenario where single mutants are selectively neutral, is on the
435 order of 10^7 generations (Equation C5). As this is the typical age for living animal genera (Van Valen, 1973;
436 Lande, 1979), we should not expect to see this fitness valley forded. Of course, with many potential fitness
437 valleys across the genome, the chance that one of them is forded can become substantial.

438 By broadening previous treatments to allow for non-Mendelian inheritance, we have shown that a small
439 amount of segregation distortion can greatly impact the chances of fitness-valley crossing. Of course, segre-
440 gation distortion has a large impact because it provides a second level of selection (Sandler and Novitski,
441 1957), often acting like gametic selection (but see Hartl, 1970, 1977). When the A_2 and B_2 alleles are more
442 likely to be passed down than the A_1 and B_1 alleles, respectively, in matings between single mutants and
443 residents, the depth of the valley is effectively reduced and hence crossing is much more likely. For example,
444 when single mutants have a relative viability of $w_m = 0.95$, the mutation rate is $\mu = 10^{-8}$, double mutants
445 are weakly favoured ($w_{22} = 1.01$), and we begin with one single mutant ($n_0 = 1$) in a population of size
446 $N = 10^4$, in the absence of recombination [$r_{21}^{12}(22) = 0$] and segregation distortion ($k_{ij} = 0$), the probability
447 of crossing is on the order of 10^{-9} (Equation 15). With a 5% distortion in favour of A_2 and B_2 alleles
448 ($k_{21} = k_{12} = 0.05$) the single mutants are effectively neutral and the probability increases seven orders of
449 magnitude to 10^{-2} . And with a 10% distortion the single mutants are selectively favoured and the double
450 mutant fixes with probability 0.25.

451 Segregation distortion, in the form of meiotic drive, has often been implicated as a force that could help
452 fix underdominant chromosomal rearrangements (Sandler and Novitski, 1957; Bengtsson and Bodmer, 1976;
453 Hedrick, 1981; Walsh, 1982; Faria and Navarro, 2010). Chromosomal rearrangements, such as translocations
454 and inversions, are often fixed in alternate forms in closely related species (White, 1978; Coyne, 1989;
455 Faria and Navarro, 2010). Because heterokaryotypes typically have severely reduced fertility (Sandler and
456 Novitski, 1957; Lande, 1979), such rearrangements are thought to promote rapid speciation (stasipatric
457 speciation; White, 1978, but see Faria and Navarro, 2010; Kirkpatrick, 2010). The trouble is explaining
458 how such rearrangements originally increase in frequency when they are so strongly selected against when
459 rare (Navarro and Barton, 2003; Kirkpatrick, 2010). Meiotic drive provides one possible answer. Our
460 results can be used to investigate valley crossing with chromosomal rearrangements by assuming **A** and
461 **B** are homologous chromosomes, with A_2 and B_2 being the novel chromosomes, and A_2B_1 and A_1B_2
462 interchangeable. For example, with free recombination [$r_{21}^{12}(22) = 1/4$, $r_{11}^{22}(22) = 1/4$], a 5% viability
463 reduction in heterokaryotypes ($w_m = 0.95$), no meiotic drive [$\bar{b}_{11}^m(m) = 1/2$], a very beneficial mutant
464 homokaryotype ($w_{22} = 2.5$), and a spontaneous chromosome mutation rate of $\mu = 10^{-3}$ (Lande, 1979),
465 when starting with one copy of each mutant chromosome ($i_0 = 1$, $c_n = 1$) in a population of size $N = 10^4$,
466 Equation (17) gives a 0.4% chance of fixing the mutant homokaryotype. When the mutated chromosome has
467 a 70% chance of being passed down in matings with residents, a relatively weak amount of drive (Sandler

468 [and Novitski, 1957](#)), the chance of crossing increases two orders of magnitude, to nearly 75%.

469 Here we have shown that, for a given number of single mutants, the chance of crossing a valley by
470 recombination is best when the two single mutant types are at equal frequencies. This is an important
471 factor when the mutation rates in **A** and **B** are highly asymmetric. One instance where this asymmetry is
472 likely is when one locus (say **B**) is in the mitochondrial genome, and is passed down maternally, while the
473 other (say **A**) is in the nuclear genome, and is passed down biparentally. Mutation rates in the mitochondria
474 can be orders of magnitude higher than in the nucleus ([Linnane et al., 1989](#)). With $r_{21}^{12}(22) = r/2 = 1/4$,
475 $N = 10^4$, $w_{22} = 2$, $b_{11}^{11}(21) = \mu = 10^{-6}$, and neutral single mutants, when the mutation rates in **A** and **B**
476 are equal [$b_{11}^{11}(12)/b_{11}^{11}(21) = c_\mu = 1$] the crossing time is 40,000 generations. When the mutation rate in **B**
477 is two orders of magnitude larger ($c_\mu = 100$) the waiting time is reduced to 2,500 generations. But when the
478 average mutation rate $(1 + c_\mu)b_{11}^{11}(21)/2$ is held constant, the asymmetrical mutation rates instead hinder
479 crossing, increasing the crossing time to nearly 120,000 generations.

480 By expanding a mathematical model of fitness-valley crossing beyond symmetrical Mendelian inheritance
481 we gain insight into transitions between alternate stable states in non-genetic systems, such as culture.
482 As mentioned in the introduction, culture may often exhibit alternate stable states; here valley crossing
483 corresponds to a shift between alternate combinations of cultural ideas or practices (e.g., the demographic
484 transition; [Borgerhoff Mulder, 1998](#)). The valley is a “transmission valley”, created by new cultural traits
485 that are transmitted effectively in concert but poorly when arising individually within the previous cultural
486 background. In this case our simplified example above demonstrates that, given that the component pieces
487 are not passed on too poorly in the previous cultural background, the probability that a new set of practices
488 or ideas becomes pervasive in society is greatly improved by its transmission advantage over the previous
489 set. Valley crossing might also be relevant in the context of gene-culture coevolution, where one trait is
490 cultural and the other genetic. For instance, the ability to absorb lactose as an adult is largely genetically
491 determined and is positively correlated with the cultural practice of dairy farming, reaching frequencies over
492 90% in cultures with dairy farming but typically remaining less than 20% in cultures without ([Feldman
493 and Laland, 1996](#)). If, as seems reasonable, the ability to absorb lactose as an adult has a cost in the
494 absence of dairy farming and the cultural practice of dairy farming has a cost when adults are unable to
495 absorb lactose, then the transition from non-pastoralist non-absorbers to pastoralist absorbers may represent
496 another example of fitness-valley crossing outside the purely genetic arena. We have used our generalized
497 model to begin to explore cultural transitions, but it should be emphasized that we neglect oblique and
498 horizontal transmission, common features of cultural evolution ([Cavalli-Sforza and Feldman, 1981](#)) and
499 likely components of the demographic transition ([Ihara and Feldman, 2004](#)). Generalizing models of fitness-
500 valley crossing further to include oblique and horizontal transmission would improve insight into cultural
501 transitions.

502 We have incorporated transmission bias in a model of multi-locus fitness-valley crossing. This allows us

503 to investigate fitness-valley crossing in new scenarios, such as in genetic systems with segregation distortion
504 and/or uniparental inheritance. Segregation distortion acts as a second level of selection and therefore can
505 greatly help or hinder fitness-valley crossing, especially when crossing is otherwise unlikely. Uniparental
506 inheritance will often imply asymmetric mutation rates, which in turn lead to unequal frequencies of single
507 mutants, and therefore, all else being equal, a lower probability of fitness-valley crossing by recombination.
508 However, uniparental-inherited cytoplasmic elements tend to have increased mutation rates, which helps
509 crossing. Generalizing transmission also allows us to begin to extend the theory of valley crossing to non-
510 genetic systems, such as culture. Despite component traits being passed on poorly in the previous cultural
511 background, we find that small advantages in the transmission of the new set of cultural traits will greatly
512 facilitate a cultural transition. While crossing a deep fitness valley is difficult under Mendelian inheritance,
513 it can be easier when Mendel is left behind.

514 4. Acknowledgements

515 We thank Eva Kisdi, Stefan Geritz, Helene Weigang, and the labs of S.P. Otto, Michael Whitlock,
516 Michael Doebeli, and Christoph Hauert for helpful discussions. We thank Mark Kirkpatrick for suggesting
517 the simplicity of considering crossing times from mutation-selection balance. Funding provided by Natural
518 Sciences and Engineering Research Council (Canada) Discovery (SPO) and CGS-D (MMO) grants.

519 5. References

- 520 Altland, A., Fischer, A., Krug, J., and Szendro, I. G. (2011). Rare events in population genetics: Stochastic tunneling in a
521 two-locus model with recombination. *Physical Review Letters*, 106(8):088101.
- 522 Barton, N. H. (1992). On the spread of new gene combinations in the third phase of Wright's shifting-balance. *Evolution*,
523 46(2):551–557.
- 524 Barton, N. H. and Rouhani, S. (1987). The frequency of shifts between alternative equilibria. *Journal of Theoretical Biology*,
525 125(4):397–418.
- 526 Barton, N. H. and Rouhani, S. (1993). Adaptation and the shifting balance. *Genetical Research*, 61:57–74.
- 527 Bengtsson, B. and Bodmer, W. (1976). On the increase of chromosome mutations under random mating. *Theoretical Population*
528 *Biology*, 9:260–281.
- 529 Borgerhoff Mulder, M. (1998). The demographic transition: are we any closer to an evolutionary explanation? *Trends in*
530 *Ecology and Evolution*, 13(7):266–270.
- 531 Boyd, R. (2001). Equilibrium selection and the evolution of norms. In Nicita, A. and Pagano, U., editors, *The evolution of*
532 *economic diversity*. Routledge, New York, NY.
- 533 Boyd, R. and Richerson, P. J. (1985). *Culture and the evolutionary process*. The University of Chicago Press, Chicago, IL.
- 534 Boyd, R. and Richerson, P. J. (2010). Transmission coupling mechanisms: cultural group selection. *Philosophical Transactions*
535 *of the Royal Society B: Biological Sciences*, 365(1559):3787–3795.
- 536 Burt, A. and Trivers, R. (2006). *Genes in conflict*. Belknap Press, Cambridge, Massachusetts.

- 537 Carneiro, M. and Hartl, D. L. (2010). Adaptive landscapes and protein evolution. *Proceedings of the National Academy of*
538 *Sciences of the United States of America*, 107:1747–1751.
- 539 Cavalli-Sforza, L. L. and Feldman, M. W. (1981). *Cultural transmission and evolution: a quantitative approach*. Princeton
540 University Press, Princeton, NJ.
- 541 Charlesworth, B. and Rouhani, S. (1988). The probability of peak shifts in a founder population. II. An additive polygenic
542 trait. *Evolution*, 42(6):1129–1145.
- 543 Christiansen, F. B., Otto, S., Bergman, A., and Feldman, M. W. (1998). Waiting with and without recombination: the time
544 to production of a double mutant. *Theoretical Population Biology*, 53(3):199–215.
- 545 Coyne, J. A. (1989). A test of the role of meiotic drive in fixing a pericentric inversion. *Genetics*, 123:241–243.
- 546 Coyne, J. A., Barton, N. H., and Turelli, M. (1997). A critique of Sewall Wright’s shifting balance theory of evolution. *Evolution*,
547 51(3):643–671.
- 548 Coyne, J. A., Barton, N. H., and Turelli, M. (2000). Is Wright’s shifting balance process important in evolution? *Evolution*,
549 54(1):306–317.
- 550 Creanza, N., Fogarty, L., and Feldman, M. W. (2012). Models of cultural niche construction with selection and assortative
551 mating. *PLoS ONE*, 7(8):e42744.
- 552 Crow, J., Engels, W., and Denniston, C. (1990). Phase three of Wright’s shifting-balance theory. *Evolution*, 44(2):233–247.
- 553 Crow, J. and Kimura, M. (1965). Evolution in sexual and asexual populations. *American Naturalist*, 99(909):439–450.
- 554 Dawkins, R. (1976). *The selfish gene*. Oxford University Press, Oxford, UK.
- 555 Eshel, I. and Feldman, M. W. (1970). On the evolutionary effect of recombination. *Theoretical Population Biology*, 1(1):88–100.
- 556 Faria, R. and Navarro, A. (2010). Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends in*
557 *Ecology and Evolution*, 25(11):660–669.
- 558 Feldman, M. W. and Laland, K. N. (1996). Gene-culture coevolutionary theory. *Trends in Ecology and Evolution*, 5347(96):453–
559 457.
- 560 Fog, A. (1999). *Cultural selection*. Kluwer Academic Publishers, Dordrecht, Netherlands.
- 561 Gavrillets, S. (1996). On phase three of the shifting-balance theory. *Evolution*, 50(3):1034–1041.
- 562 Gillespie, J. (1984). Molecular evolution over the mutational landscape. *Evolution*, 38(5):1116–1129.
- 563 Goodnight, C. and Wade, M. (2000). The ongoing synthesis: a reply to Coyne, Barton, and Turelli. *Evolution*, 54(1):317–324.
- 564 Goodnight, C. J. (2013). Wrights shifting balance theory and factors affecting the probability of peak shifts. In Svensson, E.
565 and Calsbeek, R., editors, *The adaptive landscape in evolutionary biology*, pages 74–86. Oxford University Press, Oxford,
566 UK.
- 567 Hadany, L. (2003). Adaptive peak shifts in a heterogenous environment. *Theoretical Population Biology*, 63(1):41–51.
- 568 Hadany, L., Eshel, I., and Motro, U. (2004). No place like home: competition, dispersal and complex adaptation. *Journal of*
569 *Evolutionary Biology*, 17(6):1328–1336.
- 570 Haldane, J. B. S. (1927). A mathematical theory of natural and artificial selection, part V: selection and mutation. *Mathematical*
571 *Proceedings of the Cambridge Philosophical Society*, 23:838–844.
- 572 Hartl, D. L. (1970). Population consequences of non-Mendelian segregation among multiple alleles. *Evolution*, 24(2):415–423.
- 573 Hartl, D. L. (1977). Applications of meiotic drive in animal breeding and population control. In Pollak, E., Kempthorne, O.,
574 and Bailey, Jr., T. B., editors, *Proceedings of the International Conference on Quantitative Genetics, 1976*. The Iowa State
575 University Press, Ames, IA.

- 576 Hedrick, P. (1981). The establishment of chromosomal variants. *Evolution*, 35(2):322–332.
- 577 Hoffmann, A. A. and Rieseberg, L. H. (2008). Revisiting the impact of inversions in evolution: from population genetic markers
578 to drivers of adaptive shifts and speciation? *Annual Review of Ecology, Evolution, and Systematics*, 39:21–42.
- 579 Ihara, Y. and Feldman, M. W. (2004). Cultural niche construction and the evolution of small family size. *Theoretical Population*
580 *Biology*, 65(1):105–111.
- 581 Karlin, S. and McGregor, J. (1971). On mutation selection balance for two-locus haploid and diploid populations. *Theoretical*
582 *Population Biology*, 70:60–70.
- 583 Karlin, S. and Tavaré, S. (1981). The detection of particular genotypes in finite populations. I. Natural selection effects.
584 *Theoretical Population Biology*, 19(2):187–214.
- 585 Karlin, S. and Taylor, H. M. (1981). *A second course in stochastic processes*. Academic Press, New York, NY.
- 586 Kimura, M. (1962). On the probability of fixation of mutant genes in a population. *Genetics*, 47:713–719.
- 587 Kimura, M. (1985). The role of compensatory neutral mutations in molecular evolution. *Journal of Genetics*, 64(1):7–19.
- 588 Kimura, M. (1990). Some models of neutral evolution, compensatory evolution, and the shifting balance process. *Theoretical*
589 *Population Biology*, 37:150–158.
- 590 Kirkpatrick, M. (2010). How and why chromosome inversions evolve. *PLoS Biology*, 8(9):e1000501.
- 591 Kondrashov, A. (1992). The third phase of Wright’s shifting-balance: a simple analysis of the extreme case. *Evolution*,
592 46(6):1972–1975.
- 593 Kuhn, T. S. (1962). *The structure of scientific revolutions*. University of Chicago Press, Chicago, IL.
- 594 Lai, Z., Nakazato, T., Salmaso, M., Burke, J. M., Tang, S., Knapp, S. J., and Rieseberg, L. H. (2005). Extensive chromosomal
595 repatterning and the evolution of sterility barriers in hybrid sunflower species. *Genetics*, 171(1):291–303.
- 596 Lande, R. (1979). Effective deme sizes during long-term evolution estimated from rates of chromosomal rearrangement. *Evo-*
597 *lution*, 33(1):234–251.
- 598 Lande, R. (1985a). Expected time for random genetic drift of a population between stable phenotypic states. *Proceedings of*
599 *the National Academy of Sciences of the United States of America*, 82(22):7641–7645.
- 600 Lande, R. (1985b). The fixation of chromosomal rearrangements in a subdivided population with local extinction and colo-
601 nization. *Heredity*, 54:323–332.
- 602 Linnane, A., Ozawa, T., Marzuki, S., and Tanaka, M. (1989). Mitochondrial DNA mutations as an important contributor to
603 ageing and degenerative diseases. *The Lancet*, 333:642–645.
- 604 Lynch, M. (2010). Scaling expectations for the time to establishment of complex adaptations. *Proceedings of the National*
605 *Academy of Sciences of the United States of America*, 107(38):16577–16582.
- 606 Michalakis, Y. and Slatkin, M. (1996). Interaction of selection and recombination in the fixation of negative-epistatic genes.
607 *Genetical Research*, 67(3):257–269.
- 608 Navarro, A. and Barton, N. H. (2003). Chromosomal speciation and molecular divergence-accelerated evolution in rearranged
609 chromosomes. *Science*, 300(5617):321–324.
- 610 Peck, S., Ellner, S., and Gould, F. (1998). A spatially explicit stochastic model demonstrates the feasibility of Wright’s shifting
611 balance theory. *Evolution*, 52(6):1834–1839.
- 612 Phillips, P. (1993). Peak shifts and polymorphism during phase three of Wright’s shifting-balance process. *Evolution*,
613 47(6):1733–1743.
- 614 Phillips, P. (1996). Waiting for a compensatory mutation: phase zero of the shifting-balance process. *Genetical Research*,

- 615 67:271–283.
- 616 Rieseberg, L. H. (2001). Chromosomal rearrangements and speciation. *Trends in Ecology and Evolution*, 16(7):351–358.
- 617 Rouhani, S. and Barton, N. H. (1987a). Speciation and the “shifting balance” in a continuous population. *Theoretical Population*
618 *Biology*, 31:465–492.
- 619 Rouhani, S. and Barton, N. H. (1987b). The probability of peak shifts in a founder population. *Journal of Theoretical Biology*,
620 126:51–62.
- 621 Sandler, L. and Novitski, E. (1957). Meiotic drive as an evolutionary force. *American Naturalist*, 91(857):105–110.
- 622 Slatkin, M. (1981). Fixation probabilities and fixation times in a subdivided population. *Evolution*, 35(3):477–488.
- 623 Stathos, A. and Fishman, L. (2014). Chromosomal rearrangements directly cause underdominant F1 pollen sterility in *Mimulus*
624 *lewisii* – *Mimulus cardinalis* hybrids. *Evolution*, 68:3109–3119.
- 625 Stephan, W. (1996). The rate of compensatory evolution. *Genetics*, 144(1):419–426.
- 626 Szendro, I. G., Schenk, M. F., Franke, J., Krug, J., and de Visser, J. A. G. M. (2013). Quantitative analyses of empirical fitness
627 landscapes. *Journal of Statistical Mechanics: Theory and Experiment*, P01005.
- 628 Van Valen, L. (1973). A new evolutionary law. *Evolutionary Theory*, 1:1–30.
- 629 Wade, M. and Goodnight, C. (1998). Perspective: the theories of Fisher and Wright in the context of metapopulations: when
630 nature does many small experiments. *Evolution*, 52(6):1537–1553.
- 631 Walsh, J. (1982). Rate of accumulation of reproductive isolation by chromosome rearrangements. *American Naturalist*,
632 120(4):510–532.
- 633 Weinreich, D. and Chao, L. (2005). Rapid evolutionary escape by large populations from local fitness peaks is likely in nature.
634 *Evolution*, 59(6):1175–1182.
- 635 Weinreich, D. M., Watson, R. A., and Chao, L. (2005). Perspective: sign epistasis and genetic constraint on evolutionary
636 trajectories. *Evolution*, 59:1165–1174.
- 637 Weissman, D. B., Desai, M. M., Fisher, D. S., and Feldman, M. (2009). The rate at which asexual populations cross fitness
638 valleys. *Theoretical Population Biology*, 75(4):286–300.
- 639 Weissman, D. B., Feldman, M., and Fisher, D. S. (2010). The rate of fitness-valley crossing in sexual populations. *Genetics*,
640 186(4):1389–1410.
- 641 White, M. J. D. (1978). *Modes of speciation*. W.H. Freeman and Company, San Francisco, CA.
- 642 Whitlock, M. and Phillips, P. (2000). The exquisite corpse: a shifting view of the shifting balance. *Trends in Ecology and*
643 *Evolution*, 5347:347–348.
- 644 Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, 16(2):97–159.
- 645 Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proceedings of The Sixth*
646 *International Congress of Genetics*, 1:356–366.

647 A. Dynamic single mutants

Here we calculate the waiting time for a successful double mutant to arise, starting from a population that is composed entirely of residents. While single mutants are far from mutation-selection balance, $|s_{ij}t| < 1 \forall i \neq j$, we can write x_{22} as a function of t by replacing x_{ij} in Equation (4) with the appropriate x_{ij} in

Equation (3). With rare mutation, rare single mutants, and weak selection on single mutants, the expected number of generations until a successful double mutant appears, T , solves

$$\begin{aligned}
 \frac{1}{u_{22}N} &= \mu_{11}^{11}(22) \\
 &+ T \left[\mu_{11}^{11}(22) [1 - \mu_{11}^{11}(21)^* - \mu_{11}^{11}(12)^*] + \mu_{11}^{11}(21)^* \mu_{11}^{21}(22) + \mu_{11}^{11}(12)^* \mu_{11}^{12}(22) \right. \\
 &+ \left. \frac{1}{3} \mu_{11}^{11}(21)^* s_{21} \mu_{11}^{21}(22) + \frac{1}{3} \mu_{11}^{11}(12)^* s_{12} \mu_{11}^{12}(22) + \frac{1}{3} \mu_{11}^{11}(21)^* \mu_{11}^{11}(12)^* r_{21}^{12}(22) \right] \\
 &+ T^2 \left[\mu_{11}^{11}(21)^* \mu_{11}^{21}(22) + \mu_{11}^{11}(12)^* \mu_{11}^{12}(22) + \mu_{11}^{11}(21)^* \mu_{11}^{11}(12)^* r_{21}^{12}(22) \right. \\
 &- \left. \mu_{11}^{11}(22) [\mu_{11}^{11}(21)^* + \mu_{11}^{11}(12)^*] \right] \\
 &+ T^3 \left[\frac{2}{3} \mu_{11}^{11}(21)^* \mu_{11}^{11}(12)^* r_{21}^{12}(22) + \frac{1}{3} \mu_{11}^{11}(21)^* s_{21} \mu_{11}^{21}(22) + \frac{1}{3} \mu_{11}^{11}(12)^* s_{12} \mu_{11}^{12}(22) \right] \\
 &+ O(\epsilon^5). \tag{A1}
 \end{aligned}$$

648 The $O(\epsilon^5)$ terms disappear and the equation is exact when single mutants are neutral, $s_{ij} = 0$. Otherwise,
 649 with selection against single mutants, the higher order terms can only be ignored as long as the crossing
 650 time, T , is much smaller than the inverse of the selection coefficients, s_{21} and s_{12} . Because Equation (A1)
 651 is a cubic in T , its solution is cumbersome (see supplementary *Mathematica* file). Here we examine two
 652 scenarios which give more interpretable approximations for T .

653 Without selection on single mutants ($s_{21} = s_{12} = 0$) and without recombination from single mutants to
 654 double mutants [$r_{21}^{12}(22) = 0$] the T^3 term in Equation (A1) vanishes. In addition, if the crossing time T is
 655 long, the dominant term is the one proportional to T^2 . Solving for T from this term alone gives

$$T \approx \left[u_{22}N [\mu_{11}^{11}(21)^* \mu_{11}^{21}(22) + \mu_{11}^{11}(12)^* \mu_{11}^{12}(22)] \right]^{-1/2}, \tag{A2}$$

656 where we have ignored double mutants arising instantaneously [$\mu_{11}^{11}(22) = 0$]. Equation (A2) shows that the
 657 crossing time without selection on or recombination among single mutants is roughly proportional to $N^{-1/2}$
 658 generations. The crossing time decreases with N because increasing N increases the per generation input of
 659 mutations. Holding mutation input $\theta_{ij}^{kl}(mn) = N\mu_{ij}^{kl}(mn)$ constant, the crossing time becomes proportional
 660 to $N^{1/2}$. When the single-mutation transmission probabilities are equal [$\mu_{11}^{11}(21)^* = \mu_{11}^{11}(12)^* = 2\mu_{11}^{21}(22) =$
 661 $2\mu_{11}^{12}(22) = \mu$] and we calculate the first appearance of any double mutant (successful or not; $u_{22} = 1$),
 662 the expected time until the first double mutant appears simplifies to the neutral genetic case without
 663 recombination, $T \approx 1/\sqrt{\mu^2 N}$ (equation 8 in [Christiansen et al., 1998](#)). Equation (A2) clarifies the role of

664 the various, *potentially different*, mutation probabilities $\mu_{11}^{ij}(kl)$ on the time until the first double mutant,
 665 while also allowing us to ignore double mutants that are lost.

666 When there is recombination between single mutants to produce double mutants [$r_{21}^{12}(22) > 0$] and the
 667 crossing time, T , is smaller than the inverse of the selection coefficients, s_{12} and s_{21} , the dominant term
 668 in Equation (A1) is proportional to T^3 . This term is positive when recombination is frequent relative to
 669 selection against single mutants. Again, if the time T is long we can use this term alone to approximate T ,
 670 which gives

$$T \approx \frac{3^{1/3}}{\left[u_{22}N [\mu_{11}^{11}(21)^* \mu_{11}^{11}(12)^* r_{21}^{12}(22) + s_{21} \mu_{11}^{11}(21)^* \mu_{11}^{21}(22) + s_{12} \mu_{11}^{11}(12)^* \mu_{11}^{12}(22)] \right]^{1/3}}, \quad (\text{A3})$$

671 where we have once again ignored the instantaneous production of double mutants. Notice that, for a
 672 given mutation input $\theta_{ij}^{kl}(mn)$, when there is recombination between single mutants, the crossing time is
 673 roughly proportional to $N^{1/3}$ generations (rather than $N^{1/2}$ generations without recombination), implying
 674 that recombination between single mutants tends to shorten the expected time until the first (successful or
 675 unsuccessful) double mutant arises. However, because recombination can also occur between residents and
 676 double mutants (reducing u_{22}) Equation (A3) shows that the waiting time until the first *successful* double
 677 mutant is minimized at intermediate levels of recombination.

678 Equation (A3) reduces to $T \approx 1/\sqrt[3]{Nr\mu^2/3}$ (equation 9 in [Christiansen et al., 1998](#)) when we ignore
 679 the weak selection against single mutants ($s_{ij} = 0$), there is equal mutation probability at each locus
 680 [$\mu_{11}^{11}(21)^* = \mu_{11}^{11}(12)^* = \mu$], and we wait until the first double mutant appears, successful or not ($u_{22} = 1$).
 681 Once again our analysis clarifies the role of the various, potentially different, mutation probabilities $\mu_{ij}^{kl}(mn)$
 682 on the waiting time until the first successful double mutant. Equation (A3) also allows (weak) selection on
 683 single mutants and incorporates transmission bias, which we explore more fully in the main text.

684 Figure A1 compares the approximations derived here (Equations A2 and A3) with that derived in the
 685 text assuming mutation-selection balance is reached before crossing (Equation 8). The approximations given
 686 by Equations (A2) and (A3) break down as the depth of the valley ($\delta = -s_{21} = s_{12}$) increases such that the
 687 crossing time becomes long, $T > 1/\delta$.

688 B. Diffusion approximation

689 Here we take the large population limit ($N \rightarrow \infty$), scale time such that one unit of time in the scaled
 690 diffusion process ($\tau \in \mathbb{Z}_{\geq 0}$) is N^α generations in the unscaled Markov process ($\Delta t = \tau N^\alpha$) and define new

691 frequency parameters $Y(\tau) = i_\tau/N^\beta$ and $Z(\tau) = j_\tau/N^\beta$, with $0 < \alpha, \beta < 1$.

692 We are concerned with three quantities for each variable ΔY and ΔZ . The first is the infinitesimal mean

$$\mu_Y(y) = \lim_{N \rightarrow \infty} E[\Delta Y | Y(\tau) = y = i/N^\beta] = \lim_{N \rightarrow \infty} \frac{N^\alpha}{N^\beta} E[\Delta i | i_t = i]. \quad (\text{B1})$$

693 The second quantity is the infinitesimal variance

$$\sigma_Y^2(y) = \lim_{N \rightarrow \infty} E[(\Delta Y)^2 | Y(\tau) = y = i/N^\beta] = \lim_{N \rightarrow \infty} \frac{N^\alpha}{N^{2\beta}} E[(\Delta i)^2 | i_t = i]. \quad (\text{B2})$$

694 And the third quantity of interest is a higher ($n > 2$) infinitesimal moment

$$\lim_{N \rightarrow \infty} E[(\Delta Y)^n | Y(\tau) = y = i/N^\beta] = \lim_{N \rightarrow \infty} \frac{N^\alpha}{N^{n\beta}} E[(\Delta i)^n | i_t = i]. \quad (\text{B3})$$

695 We can similarly calculate $\mu_Z(z)$, $\sigma_Z^2(z)$, and a higher moment in ΔZ .

696 The final quantity of interest is the scaled “killing rate”

$$\kappa(y, z) = \lim_{N \rightarrow \infty} N^\alpha \tilde{P}_{ij}^H \approx \lim_{N \rightarrow \infty} N^\alpha N x'_{22} u_{22}, \quad (\text{B4})$$

697 where the approximation assumes $x'_{22} u_{22} \ll 1$.

698 For the Markov chain to converge to a diffusion process as $N \rightarrow \infty$ we require: 1) $\mu_Y(y)$ and $\mu_Z(z)$ to be
 699 finite; 2) $\sigma_Y^2(y)$, $\sigma_Z^2(z)$, and $\kappa(y, z)$ to be positive and finite; and 3) some higher moment (in both ΔY and
 700 ΔZ) to be equal to zero (Karlin and Taylor, 1981). We first take a hint from the genetic case (Christiansen
 701 et al., 1998) and scale transmission probabilities as

$$b_{ij}^{kl}(mn) = \begin{cases} B_{ij}^{kl}(mn) & : m \in \{i, k\}, n \in \{j, l\} \\ \frac{B_{ij}^{kl}(mn)}{N^2} + O(1/N^3) & : m \notin \{i, k\}, n \notin \{j, l\} \\ \frac{B_{ij}^{kl}(mn)}{N} + O(1/N^2) & : \text{otherwise} \end{cases} \quad (\text{B5})$$

702 In the genetic case this can be interpreted as making the probability of mutation proportional to the inverse
 703 of population size $\mu = B/N$. Then, as $N \rightarrow \infty$ mutation probability decreases ($\mu \rightarrow 0$), such that mutation
 704 input $B = N\mu$ is constant. This prevents the process from taking large jumps in frequency space, which
 705 violate the diffusion process (Karlin and Taylor, 1981).

706 In order for the transmission parameters to satisfy the logical constraint $\sum_{m,n=1}^2 b_{ij}^{kl}(mn) = 1$ the diffusion

707 also requires, as $N \rightarrow \infty$, that

$$B_{ij}^{ij}(ij) = 1 + O(1/N^\beta) \quad (\text{B6})$$

708 and

$$B_{ij}^{kl}(ij) + B_{ij}^{kl}(kl) = 1 + O(1/N^\beta) \quad (\text{B7})$$

709 when either $\{i \neq k, j = l\}$ or $\{i = k, j \neq l\}$. In words, the sum total mutation probability for parents $A_i B_j$
710 and $A_k B_l$ must be relatively small, at most on the order of $1/N^\beta$.

711 Finally, our approximation requires weak selection, relative to $w_{11} = 1$. In particular, total selection on
712 single mutants must be weak, on the order of $1/N^\beta$,

$$w_{ij}[b_{11}^{ij}(11) + b_{ij}^{11}(ij)] = 1 + S_{ij}/N^\beta + O(1/N^{2\beta}) \quad (\text{B8})$$

713 for $i \neq j$, where S_{ij} is the scaled selection strength. And selection on double mutants must also be weak,
714 such that

$$s_{22} = S_{22}/N + O(1/N^2). \quad (\text{B9})$$

715 With the above assumptions (Equations B5-B9) the Markov chain converges to a diffusion process as
716 $N \rightarrow \infty$ when

$$\alpha = \beta = \begin{cases} 1/2 & : r_{12}^{21}(22) \leq O(1/N^{1/2}) \\ 1/3 & : \text{otherwise} \end{cases} \quad (\text{B10})$$

717 This scaling implies that if recombination between single mutants to make double mutants $r_{12}^{21}(22)$ is less
718 likely that mutation (which is on the order of $N^{-1/2}$; Equation B6), then the time until the process is killed
719 scales with $N^{1/2}$. Meanwhile, if recombination is more likely than mutation the killing time scales with
720 $N^{1/3}$. These results align with our semi-deterministic analysis (Equations A2 and A3).

721 When $\alpha = \beta$ the infinitesimal variances are $\sigma_Y^2(y) = y$ and $\sigma_Z^2(z) = z$. The infinitesimal means and the
722 killing term depend on the probability of recombination. When recombination is rare the single mutants are
723 expected to reach higher frequencies and therefore have a greater influence on the dynamics. To simplify,
724 when recombination is rare [$r_{21}^{12}(22) \leq O(1/N^{1/2})$] we assume weak transmission bias for residents mating
725 with single mutants [$b_{11}^{ij}(11) + b_{ij}^{11}(11) = 1 + O(1/N^\beta)$] and for single mutants mating with each other
726 [$b_{ij}^{kl}(ij) + b_{kl}^{ij}(ij) = 1 + O(1/N^\beta)$]. We further assume weak viability selection on single mutants, $w_{ij} =$

727 $1 - O(1/N^\beta)$, regardless of recombination. The infinitesimal mean is then always

$$\mu_Y(y) = B_{11}^{11}(21) - y S_{21} \quad (\text{B11})$$

728 and similarly for $\mu_Z(z)$. The first term, $B_{11}^{11}(21) \approx b_{11}^{11}(21)N$, describes mutation to single mutants in
 729 resident-resident matings and the second term, with $S_{21} \approx s_{21}N^\beta$, describes the removal of single mutants
 730 by selection (both through transmission bias when mating with the resident and survival).

731 The killing terms are

$$\kappa(y, z) = \begin{cases} u_{22}w_{22} \left[y[B_{11}^{21}(22) + B_{21}^{11}(22)] + z[B_{11}^{12}(22) + B_{12}^{11}(22)] + yz R_{21}^{12}(22) \right] & : r_{21}^{12}(22) \leq O(1/N^{1/2}) \\ u_{22} yz r_{21}^{12}(22)^* & : \text{otherwise} \end{cases} \quad (\text{B12})$$

732 where $R_{21}^{12}(22) \approx r_{21}^{12}(22)N^{1/2}$ describes a (low) probability of recombination. The first line shows that
 733 the process can be killed by mutations in single mutants that mate with residents [$B_{11}^{ij}(22) + B_{ij}^{11}(22) \approx$
 734 $(b_{11}^{ij}(22) + b_{ij}^{11}(22))N$] or by rare recombination between single mutants to produce double mutants $R_{21}^{12}(22)$.
 735 When recombination is more likely than $N^{-1/2}$ the process is essentially always killed by recombination
 736 $r_{21}^{12}(22)$ (second line in Equation B12).

737 C. Stochastic crossing times

738 *Neutral single mutants without recombination.* With no chance of recombination from single mutants to
 739 double mutants [$r_{21}^{12}(22) = 0$] we have scaling parameter $\beta = 1/2$. Then, without selection on single mutants
 740 ($s_{21} = s_{12} = 0$) and with some mutational symmetry between the two loci [$\bar{b}_{11}^{21}(22) = \bar{b}_{11}^{12}(22) = \bar{b}_{11}^m(22)$], the
 741 single mutants are equivalent and we can concern ourselves with only their sum, $\xi = y + z$. Letting m be
 742 either single mutant type ($m = 21$ or 12), Equation (12) reduces to

$$\frac{1}{2}\xi \frac{d^2\tilde{T}(\xi)}{d\xi^2} + [B_{11}^{11}(21) + B_{11}^{11}(12)] \frac{d\tilde{T}(\xi)}{d\xi} - u_{22}\xi [B_{11}^m(22) + B_m^{11}(22)] w_{22} \tilde{T}(\xi) = -1 \quad (\text{C1})$$

743 where $B_{ij}^{kl}(mn) = b_{ij}^{kl}(mn)N$.

744 When there are an infinite number of single mutants a successful double mutant is produced immedi-
 745 ately, giving one boundary condition $\lim_{\xi \rightarrow \infty} \tilde{T}(\xi) = 0$. The second boundary condition is $d\tilde{T}(0)/d\xi =$
 746 $-[B_{11}^{11}(21) + B_{11}^{11}(12)]^{-1}$, which can be derived directly from Equation (C1) by setting $\xi = 0$ (see appendix

747 A in [Christiansen et al., 1998](#), for a more complete derivation).

748 The solution to the boundary value problem, evaluated at $\xi = 0$, corresponding to the expected number
 749 of generations until a successful double mutant arises when beginning with only residents, $T = N^{1/2}\tilde{T}(0)$,
 750 is then

$$T = \frac{N^{1/2}\Gamma[1/2]\Gamma[B_{11}^{11}(21) + B_{11}^{11}(12)]}{\Gamma[1 + B_{11}^{11}(21) + B_{11}^{11}(12)]\sqrt{u_{22}2[B_{11}^m(22) + B_m^{11}(22)]w_{22}}}, \quad (\text{C2})$$

751 where $\Gamma[\cdot]$ is the gamma function. Setting mutation probabilities equal [$B_{11}^{11}(21) = B_{11}^{11}(12) = 2B_{11}^m(22) =$
 752 $2B_m^{11}(22) = \theta = N\mu]$ reduces Equation (C2) to the neutral genetic case (equation 27 in [Christiansen](#)
 753 [et al., 1998](#)) divided by $\sqrt{u_{22}w_{22}}$ because we census after selection and consider double mutant fixation.
 754 By separating the various mutational terms our analysis clarifies that, while the crossing time is inversely
 755 proportional to the mutation probability from residents to single mutants, it is inversely proportional to the
 756 *square root* of mutation probabilities from single mutants to double mutants. The crossing time is therefore
 757 increased much more by a reduction in mutations from residents to single mutants than it is by a reduction
 758 in mutations from single mutants to double mutants.

759 When mutations from residents to single mutants [$b_{11}^{11}(21)$ and $b_{11}^{11}(12)$] are rare, an approximation for
 760 the crossing time, in terms of our unscaled parameters, is

$$T \approx \frac{1}{N[b_{11}^{11}(21) + b_{11}^{11}(12)]\sqrt{u_{22}4\mu_{11}^m(22)^*}}. \quad (\text{C3})$$

761 Increasing the mutational supply of single mutants [$N(b_{11}^{11}(21) + b_{11}^{11}(12))$] or the probability of mutation
 762 from single mutants to successful double mutants [$u_{22}\mu_{11}^m(m)^*$] decreases the amount of time we expect to
 763 wait before a successful double mutant arises. Holding mutation input, θ , constant, Equation (C3) shows
 764 that the crossing time without recombination is roughly proportional to $N^{1/2}$ generations, aligning with the
 765 semi-deterministic analysis (Equation A2) and indicating that, for a given mutational input, genetic drift
 766 increases the speed at which fitness valleys are crossed.

767 *Neutral single mutants with recombination.* With recombination the scaling parameter is $\beta = 1/3$. We
 768 can reduce and solve Equation (12) with recombination when the frequencies of single mutants remain
 769 proportional to one another, such that we need follow only $c_\mu y = z = \xi$, where c_μ is a constant. This
 770 requires mutation input [$Nb_{11}^{11}(21)$, $Nb_{11}^{11}(12)$] to be large enough to make the dynamics of y and z relatively
 771 deterministic. We further assume no selection on single mutants ($s_{21} = s_{12} = 0$). We then have $c_\mu y = z$
 772 for all time, t , when the ratio of mutation probabilities is c_μ [i.e., $c_\mu b_{11}^{11}(21) = b_{11}^{11}(12)$] and we begin with

773 $c_\mu y(0) = z(0)$. Equation (12) then collapses to

$$\frac{\xi}{2(1+c_\mu)} \frac{d^2 \tilde{T}(\xi)}{d\xi^2} + B_{11}^{11}(21) \frac{d\tilde{T}(\xi)}{d\xi} - u_{22} c_\mu r_{21}^{12}(22)^* \xi^2 \tilde{T}(\xi) = -1. \quad (C4)$$

774 The boundary conditions are $\lim_{\xi \rightarrow \infty} \tilde{T}(\xi) = 0$ and $d\tilde{T}(0)/d\xi = -B_{11}^{11}(21)^{-1}$. The solution to the
775 boundary-value problem, evaluated at $\xi = 0$, in units of generations, $T = N^{1/3} \tilde{T}(0)$, is

$$T = \frac{2^{5/3} \pi}{3^{11/6} \Gamma[2/3]} \frac{N^{1/3} (1+c_\mu) \Gamma[2(1+c_\mu) B_{11}^{11}(21)/3]}{\Gamma[2(1+(1+c_\mu) B_{11}^{11}(21))/3] \sqrt[3]{u_{22} c_\mu (1+c_\mu) r_{21}^{12}(22)^*}}. \quad (C5)$$

776 Letting $c_\mu = 1$, $B_{11}^{11}(21) = \theta = N\mu$, and $r_{21}^{12}(22) = r/2$ reduces Equation (C5) to the neutral genetic
777 case (equation 30 in [Christiansen et al., 1998](#)) divided by $\sqrt[3]{u_{22} w_{22}}$ because we census after selection and
778 consider double mutant fixation. Our result extends the insight of [Christiansen et al. \(1998\)](#) by allowing
779 the frequencies of single mutants to differ, $c_\mu \neq 1$. Holding average mutation input $[(1+c_\mu) N b_{11}^{11}(21)/2]$
780 constant, Equation (C5) shows that the crossing time is minimized when there are equal numbers of the two
781 single mutants ($c_\mu = 1$) and increases as the asymmetry grows. This occurs because recombination is most
782 effective in creating double mutants when the single mutants are equally frequent.

783 Converting the full solution back in terms of our unscaled parameters and letting the mutation probability
784 $b_{11}^{11}(21)$ be small, we have the approximation

$$T \approx \frac{2^{2/3} \pi}{3^{5/6} \Gamma[2/3]^2} \frac{1}{N^{2/3} b_{11}^{11}(21) \sqrt[3]{u_{22} c_\mu (1+c_\mu) r_{21}^{12}(22)^*}}. \quad (C6)$$

785 Holding mutation input $[N b_{11}^{11}(21)]$ constant, Equation (C6) shows that the crossing time is roughly propor-
786 tional to $N^{1/3}$ generations, aligning with the semi-deterministic analysis (Equation A3).

787 D. Stochastic simulations

788 We performed stochastic simulations to verify our analytical and numerical results. Simulation code is
789 supplied in the supplementary *Mathematica* file. Briefly, we performed random multinomial sampling of
790 genotypes with frequency parameters given by Equation (1) and transition probabilities defined in Tables
791 2-4. Crossing time simulations ended on double mutant fixation and the generation in which this occurred
792 was recorded as the crossing time. Crossing time was averaged over all trials (10^3 trials in Figure 1, 10^2
793 trials in Figure 3 and 5). Crossing probability simulations ended on resident or double mutant fixation and

794 the genotype which fixed in each trial was recorded. The crossing probability was calculated as the fraction
795 of trials in which the double mutant fixed (10^3 trials in Figure 2 and 4, 10^5 trials in Figure 6).

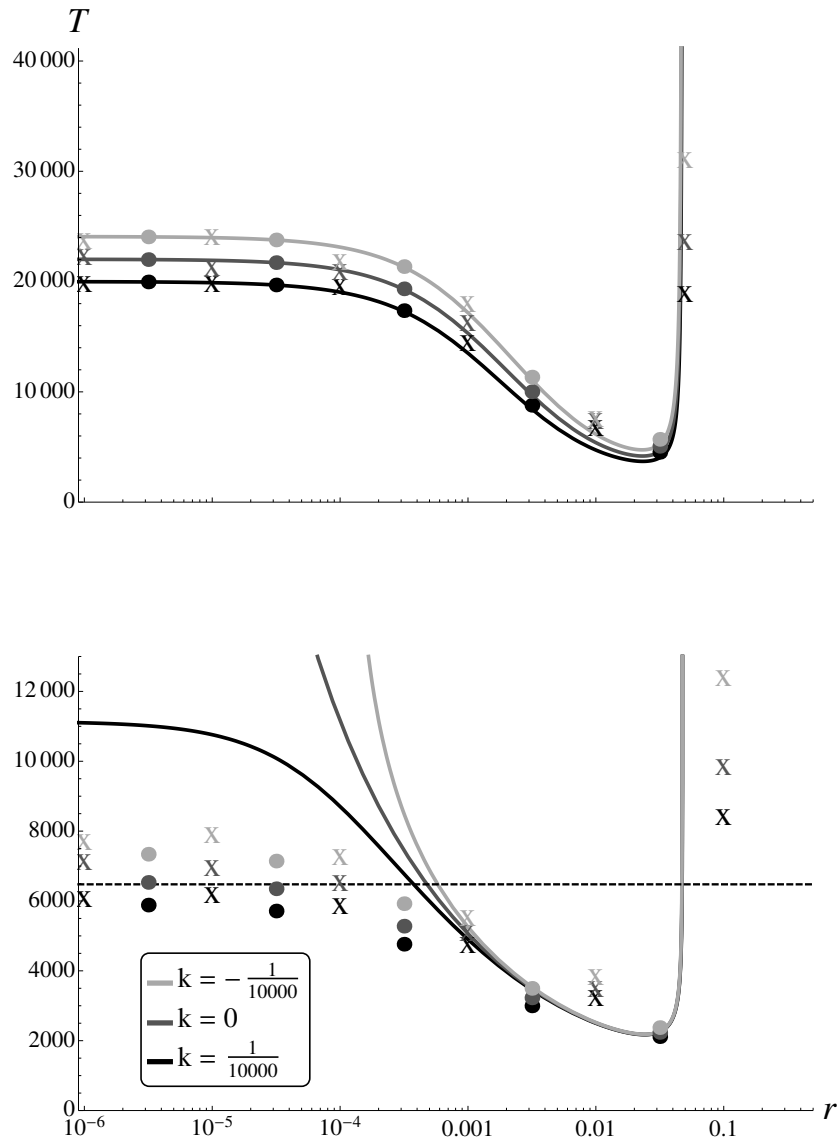


Figure 1: Expected number of generations until a double mutant begins to fix, T , as a function of the probability of recombination, r . The dots show the full semi-deterministic solution (numerical solution to Equation A1, including higher order terms, allowing both recombination and mutation to generate double mutants). The solid curves show the semi-deterministic results when (**top**) mutation-selection balance is first reached (Equation 8) and (**bottom**) mutation-selection balance is not reached and crossing can occur by recombination (Equation A3). The dashed line gives the crossing time when crossing occurs by mutation only, before mutation selection balance is reached, and single mutants are selectively neutral (Equation A2). The X's are mean simulation results (Appendix D). The grayscale corresponds to (*dark*) distortion favouring single mutants, $k = 10^{-4}$; (*medium*) the Mendelian case, $k = 0$; and (*light*) distortion favouring wild-type, $k = -10^{-4}$. Parameters: $\mu = 5 \times 10^{-7}$, $N = 10^6$, $w_{22} = 1.05$, and (**top**) $w_{21} = w_{12} = 1 - 10^{-3}$ and (**bottom**) $w_{21} = w_{12} = 1 - 10^{-5}$.

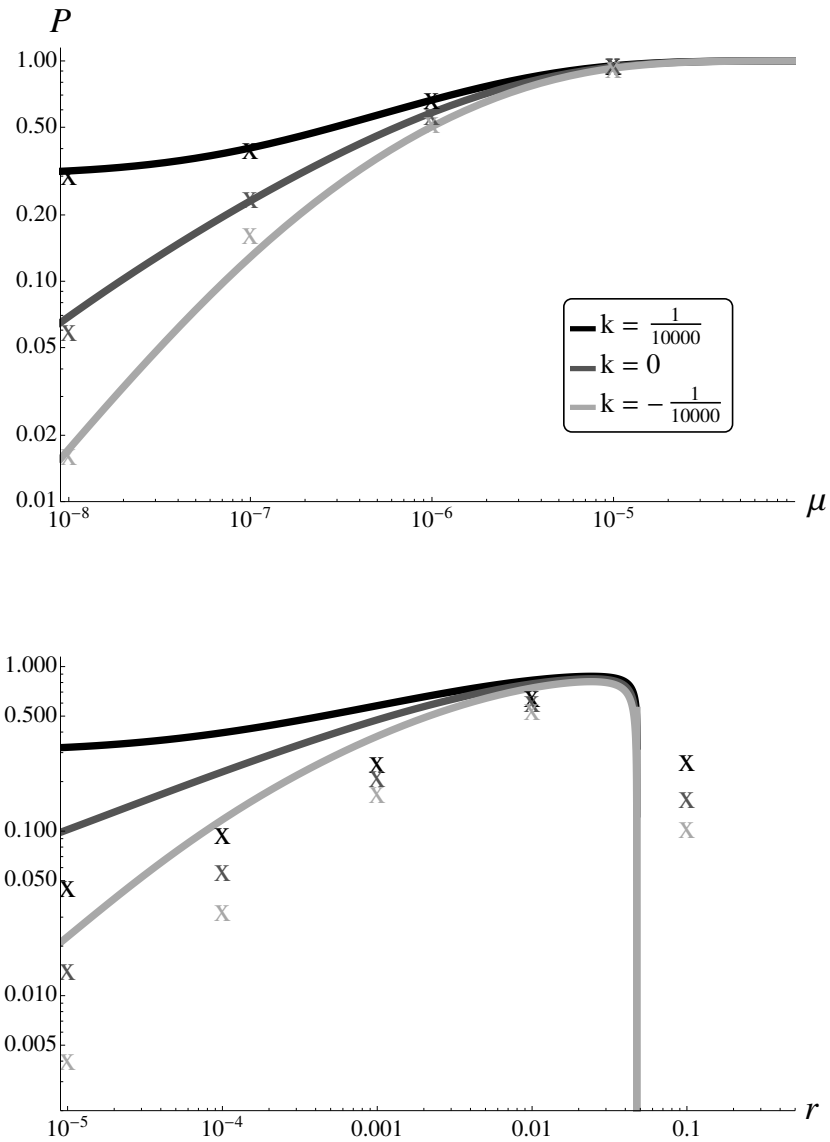


Figure 2: The probability, $P = 1 - u$, of crossing the valley given an initial stock of single mutants (with no further mutations from resident-resident matings) as a function of the rate at which single mutants produce double mutants (**top**) without recombination (Equation 15) and (**bottom**) by recombination only (Equation 17). The X 's are simulation results (Appendix D). Parameters and grayscale as in Figure 1 (bottom) with $i_0 = j_0 = 1000$.

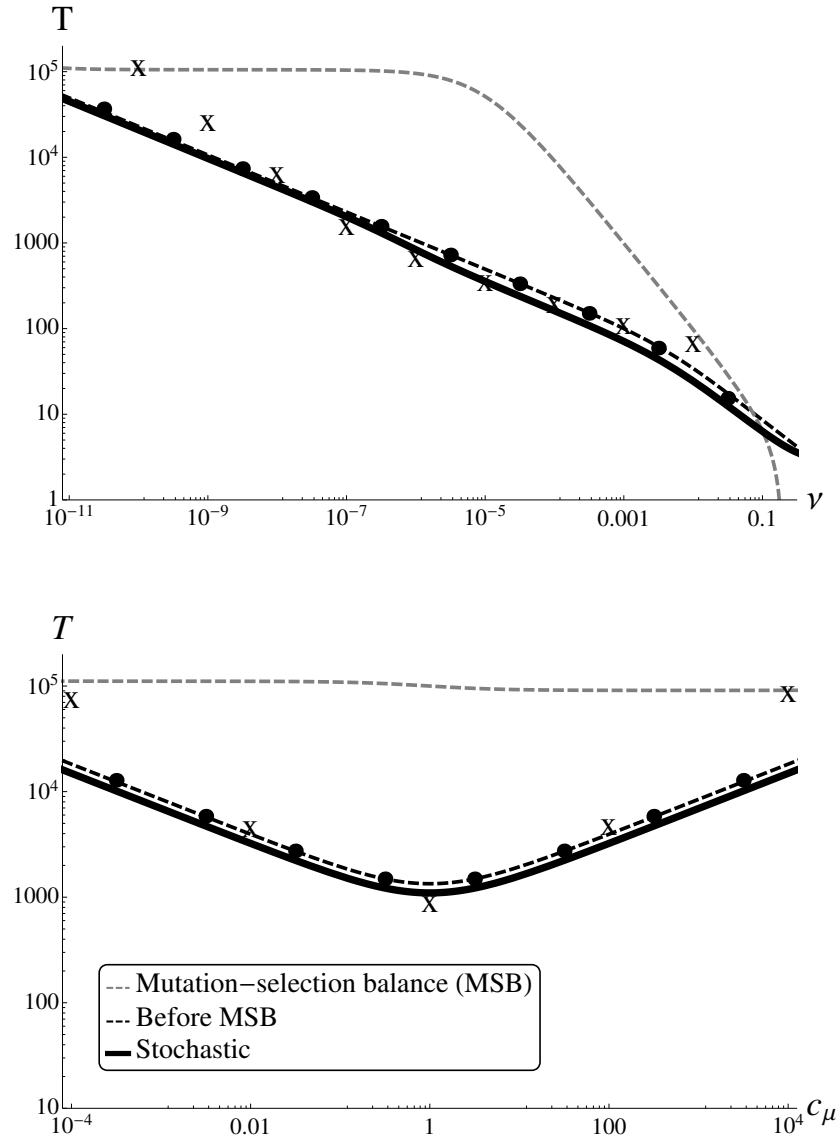


Figure 3: Expected number of generations until a double mutant begins to fix, T , as a function of (**top**) the mutation probability in locus **B**, ν , and (**bottom**) the relative mutability of the two loci, $c_\mu = \nu/\mu$. The top panel holds the mutation probability in locus **A** ($\mu = 5 \times 10^{-7}$) constant while the bottom panel holds the average mutation probability [$(\mu + \nu)/2 = \mu(1 + c_\mu)/2 = 5 \times 10^{-7}$] constant. The solid curves show the stochastic crossing time by recombination with neutral single mutants (Equation C5). The dashed curves show the semi-deterministic results when crossing occurs (*black*) before (Equation A3) and (*gray*) after (Equation 8) mutation-selection balance is first reached. The dots show the full semi-deterministic solution (numerical solution to Equation A1, including higher order terms, allowing both mutation and recombination to generate double mutants). The X's are mean simulation results (Appendix D). Parameters as in Figure 1 (bottom), except $w_{22} = 2.01$, which ensures $s_{22} \geq 0 \forall c_\mu$.

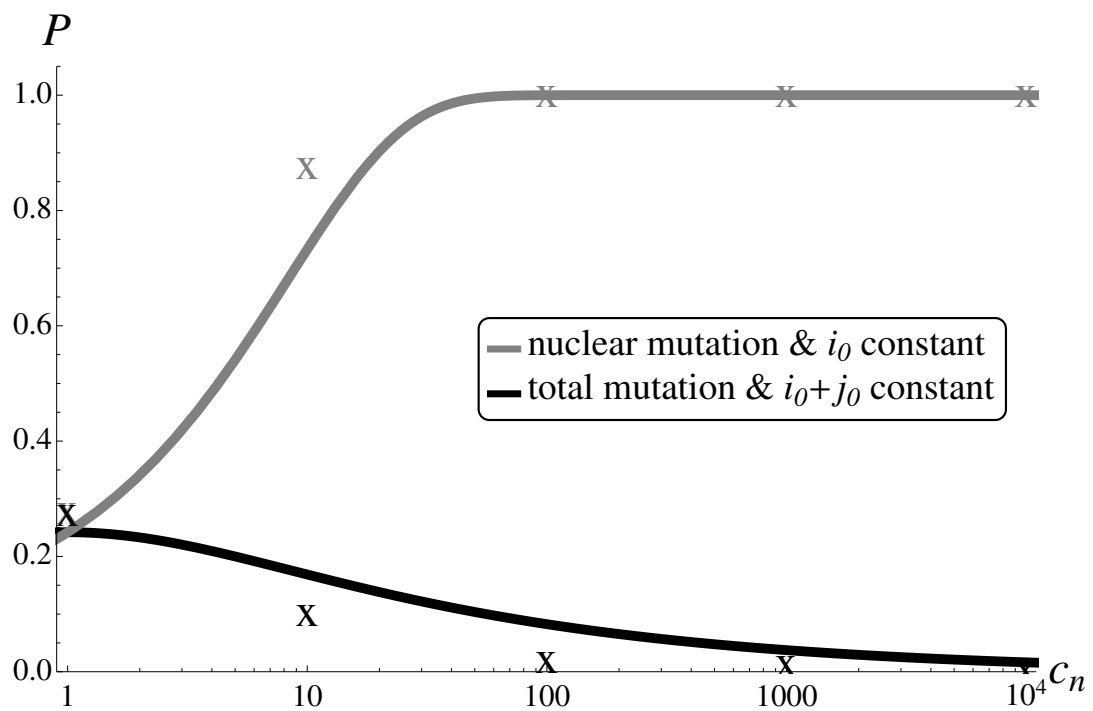


Figure 4: The probability, $P = 1 - u$, of crossing the valley given an initial stock of single mutants (with no further mutations from resident-resident matings) as a function of the ratio of the initial numbers of single mutants and mutation probabilities, $c_\mu = \nu/\mu = c_n = j_0/i_0$ (Equation 17). The *gray* curve holds the initial number of A_2B_1 ($i_0 = 100$) and the mutation probability in the **A** locus ($\mu = 5 \times 10^{-7}$) constant and varies the initial number of A_1B_2 (j_0) and the mutation probability in the **B** locus (ν). The *black* curve holds the initial number of single mutants ($n_0 = i_0 + j_0 = 200$) and average mutation probability [$(\mu + \nu)/2 = 5 \times 10^{-7}$] constant. The *X*'s are simulation results (Appendix D). Other parameters as in Figure 3.

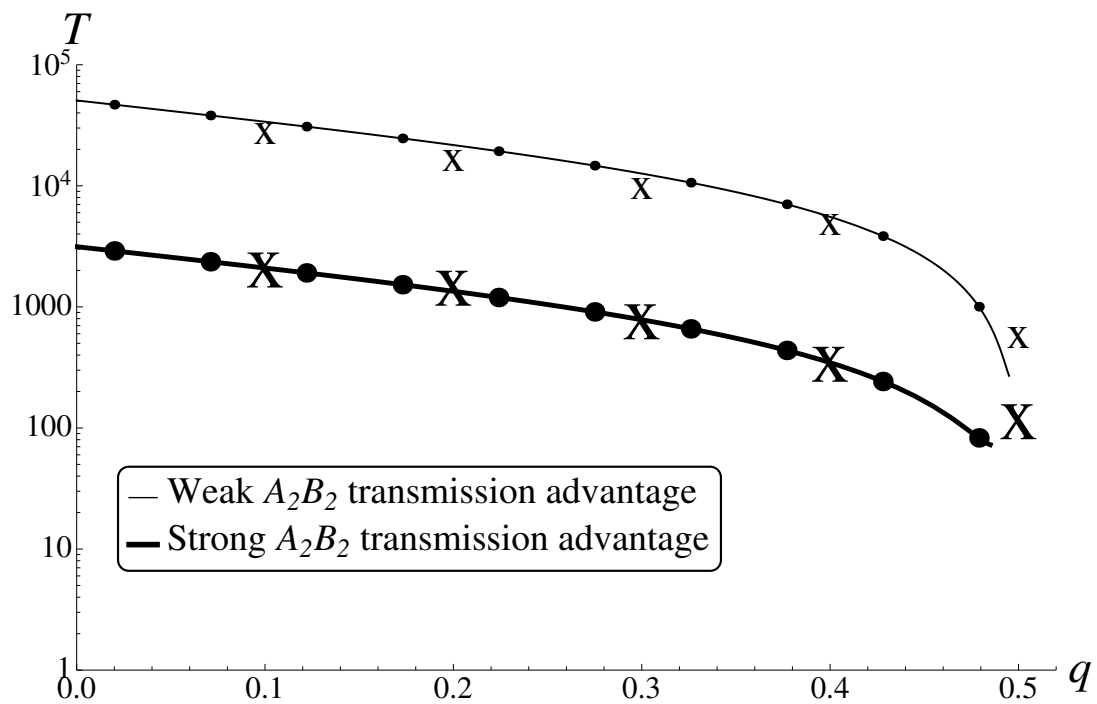


Figure 5: Expected number of generations until the new combination of cultural traits (A_2B_2) begins to fix, T , as a function of the probability of inheritance, q , of the new traits singly (A_2B_1 , A_1B_2) over the previous combination (A_1B_1). The curves show the estimate given mutation-selection balance is first reached (which assumes A_2B_1 and A_1B_2 are disfavoured, $q < 0.5$; Equation 8). The dots show the full semi-deterministic solution (numerical solution to Equation A1, including higher order terms, allowing both recombination and mutation to generate double mutants). The X 's are mean simulation results (Appendix D). The transmission advantage for the new combination of cultural traits is either weak (thin curves, small dots: $p = 0.51$) or strong (thick curves, large dots: $p = 0.6$). Parameters: $N = 10^3$, $\mu = 10^{-3}$, $r = 0.01$, and $w_{21} = w_{12} = w_{22} = 1$.

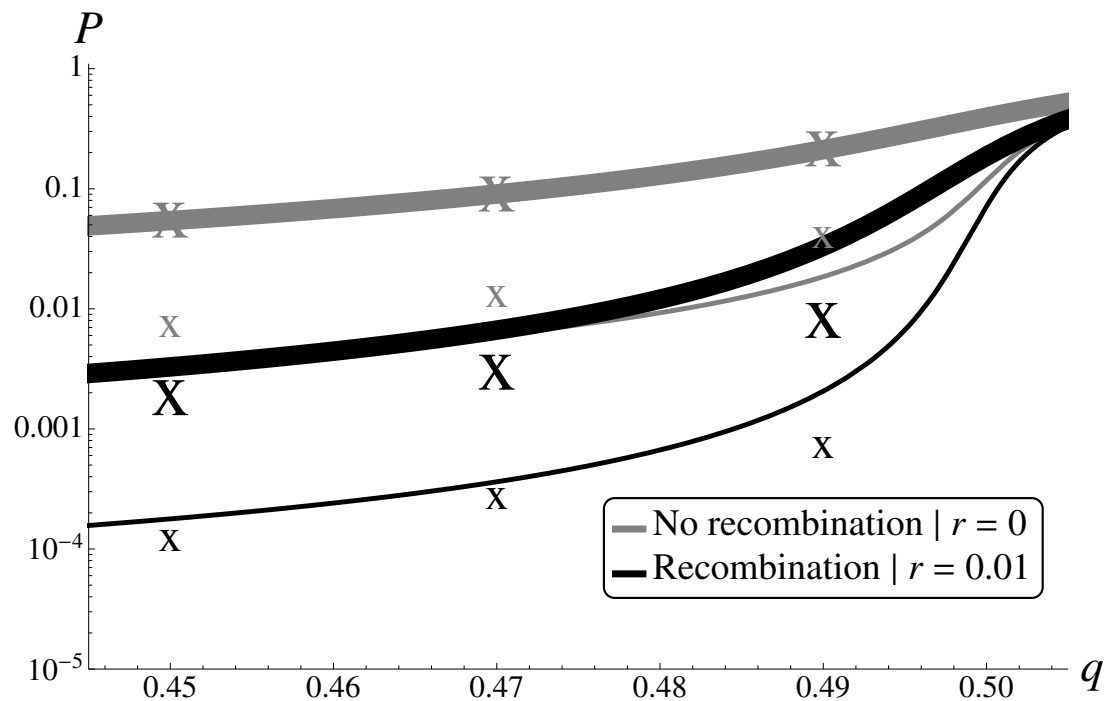


Figure 6: The probability, $P = 1 - u$, that the new combination of cultural traits (A_2B_2) fixes given an initial number of A_2B_1 and A_1B_2 (with no further mutations from resident-resident matings) as a function of the probability of inheritance, q , of the new traits singly (A_2B_1 , A_1B_2) over the previous combination (A_1B_1). The grey curves show the probability of crossing in the absence of recombination ($r = 0$; Equation 15), with a strong (thick curves: $p = 0.6$) or weak (thin curves: $p = 0.51$) transmission advantage for the new combination of cultural traits. The black curves show the probability of crossing by recombination only ($r = 0.01$; Equation 17). The X's are simulation results (Appendix D). With such a large mutation probability, crossing can be more likely without recombination, which has the added effect of breaking apart the new combination. Parameters as in Figure 5 with $i_0 + j_0 = 20$ and $c = 1$.

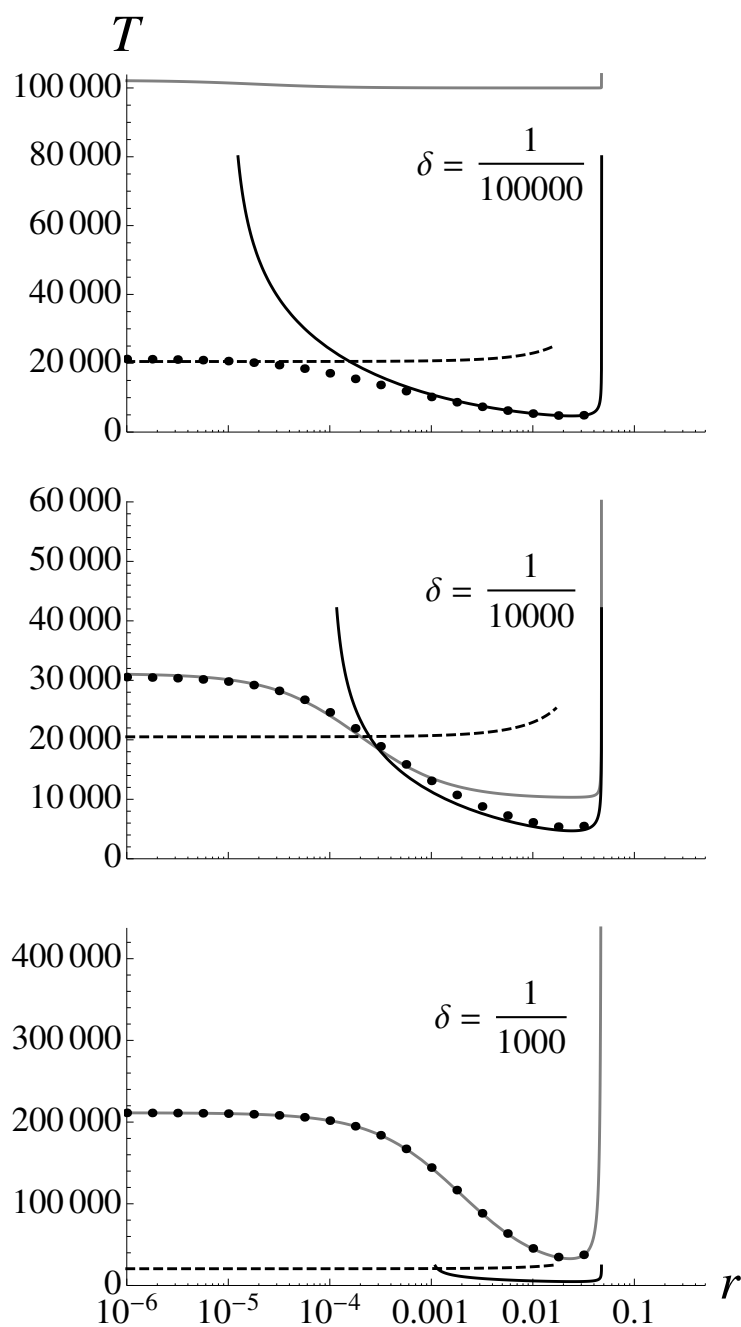


Figure A1: Expected number of generations until a double mutant begins to fix, T , as a function of recombination, r , given (gray) mutation-selection balance is first reached (Equation 8) or mutation-selection balance is not reached and (black, solid) crossing can occur by recombination (Equation A3) or (black, dashed) crossing occurs by mutation only and $-s_{21} = -s_{12} = \delta = 0$ (Equation A2). The dots show the full semi-deterministic solution (numerical solution to Equation A1, including higher order terms, allowing both recombination and mutation to generate double mutants). The mutation-selection balance estimate (gray) performs better than the dynamic estimates (black) when $\delta T > 1$, and vice-versa. Parameters: symmetrical, Mendelian inheritance with $N = 10^5$, $s_{22} = 0.05$, and $\mu = 5 \times 10^{-7}$ (see supplementary *Mathematica* file).

Table 1: Parameters used throughout text

| Symbol | Description |
|---------------------------|--|
| x_{ij} | frequency of $A_i B_j$ in the current generation |
| x'_{ij} | expected frequency of $A_i B_j$ in the next generation |
| w_{ij} | viability of $A_i B_j$ relative to viability of $A_1 B_1$ |
| V | normalizing factor |
| N | number of individuals in the population |
| t | time, in units of generations |
| $b_{ij}^{kl}(mn)$ | probability $A_i B_j$ mother and $A_k B_l$ father produce $A_m B_n$ offspring |
| $\bar{b}_{ij}^{kl}(mn)^*$ | average probability of surviving $A_m B_n$ offspring from $A_i B_j \times A_k B_l$ mating, $\frac{1}{2} w_{mn} [b_{ij}^{kl}(mn) + b_{kl}^{ij}(mn)]$ |
| $\mu_{ij}^{kl}(mn)^*$ | probability of surviving <i>mutant</i> offspring, $\bar{b}_{ij}^{kl}(mn)^*$, $m \notin \{i, k\}$, $n \notin \{j, l\}$ |
| $r_{ij}^{kl}(mn)^*$ | probability of surviving <i>recombinant</i> offspring, $\bar{b}_{ij}^{kl}(mn)^*$, $m \in \{i, k\}$, $n \in \{j, l\}$, $mn \notin \{ij, kl\}$ |
| $\bar{b}_{ij}^{kl}(mn)$ | average transmission probability before selection, $\bar{b}_{ij}^{kl}(mn)^* / w_{mn}$ [similarly for $\mu_{ij}^{kl}(mn)$ and $r_{ij}^{kl}(mn)$] |
| s_{ij} | selection on $A_i B_j$ in a resident population, $2\bar{b}_{11}^{ij}(ij)^* - 1$ |
| T | generations until first successful double mutant arises |
| u_{22} | probability that a double mutant begins a lineage that will fix |
| i_t | number of $A_2 B_1$ individuals in generation t (similarly for $A_1 B_2$, j_t) |
| $X(t)$ | numbers of single mutants in generation t assuming no double mutants, (i_t, j_t) |
| Δi | change in number of $A_2 B_1$ individuals, $i_{t+1} - i_t$ (similarly for $A_1 B_2$, $\Delta j = j_{t+1} - j_t$) |
| α, β | scaling parameters in diffusion process |
| τ | scaled unit of time, t/N^α |
| $Y(\tau)$ | scaled frequency of $A_2 B_1$, i_τ/N^β (similarly for $A_1 B_2$, $Z(\tau) = j_\tau/N^\beta$) |
| $\mu_Y(y)$ | first moment of $\Delta Y = Y(\tau + 1) - Y(\tau)$ given $Y(\tau) = y = i/N^\beta$ (similarly for Z) |
| $\sigma_Y^2(y)$ | second moment of ΔY given $Y(\tau) = y$ (similarly for Z) |
| $\kappa(y, z)$ | rate diffusion killed by successful double mutants given $Y(0) = y, Z(0) = z$ |
| $B_{ij}^{kl}(mn)$ | scaled transmission probability, $b_{ij}^{kl}(mn)N^\beta$ |
| $R_{21}^{12}(22)$ | scaled (rare) recombination from single mutants to double mutants, $r_{21}^{12}(22)N^{1/2}$ |
| S_{ij} | scaled selection on $A_i B_j$ in population of residents, $s_{ij}N^\beta$ |
| $\tilde{T}(y, z)$ | scaled time until first successful double mutant given $Y(0) = y, Z(0) = z, TN^\alpha$ |
| m | index for single mutant types when equivalent (e.g., $s_m = s_{21} = s_{12}$) |
| c_μ | mutation rate at locus B relative to locus A , ν/μ |
| c_n | initial number of $A_1 B_2$ individuals, relative to $A_2 B_1$, j_0/i_0 |
| ξ | scaled frequency of single mutants ($y + z$ or $c_i y = z$, depending on assumptions) |
| $u(y, z)$ | probability no successful double mutant appears given $Y(0) = y, Z(0) = z$ |
| n_0 | initial number of single mutants, $i_0 + j_0$ |

Table 2: Transmission probabilities, $k_i^{kl}(mn)$, with segregation distortion (autosomal killing). Recombination occurs with probability r , followed by autosomal killing of strength $0 \leq k \leq 1$, and mutation with probability μ . When $k > 0$ the killing alleles are the mutant alleles (A_2 and B_2) and when $k < 0$ the killing alleles are the resident alleles (A_1 and B_1).

| Parents | | Offspring | | | |
|----------|----------|-------------------------------|--|--|---|
| Mother | Father | A_1B_1 | A_2B_1 | A_1B_2 | A_2B_2 |
| A_1B_1 | A_1B_1 | $(1-\mu)^2$ | $\mu(1-\mu)$ | $\mu(1-\mu)$ | μ^2 |
| A_1B_1 | A_2B_1 | $\frac{1-k}{2}(1-\mu)^2$ | $\frac{1+k}{2}\mu(1-\mu)$ | $\frac{1-k}{2}\mu(1-\mu)$ | $\frac{1+k}{2}\mu + \frac{1-k}{2}\mu^2$ |
| A_1B_1 | A_1B_2 | $\frac{1-k}{2}(1-\mu)^2$ | $\frac{1+k}{2}\mu(1-\mu)$ | $\frac{1+k}{2}\mu(1-\mu)$ | $\frac{1+k}{2}\mu + \frac{1-k}{2}\mu^2$ |
| A_1B_1 | A_2B_2 | $(1-r)\frac{1-k}{2}(1-\mu)^2$ | $(1-r)\frac{1-k}{2}\mu(1-\mu) + \frac{r}{2}$ | $(1-r)\frac{1-k}{2}\mu(1-\mu) + \frac{r}{2}$ | $(1-r)\frac{1+k}{2}\mu + (1-r)\frac{1-k}{2}\mu^2$ |
| A_2B_1 | A_1B_1 | $\frac{1-k}{2}(1-\mu)^2$ | $\frac{1+k}{2}\mu(1-\mu)$ | $\frac{1+k}{2}\mu(1-\mu)$ | $\frac{1+k}{2}\mu + \frac{1-k}{2}\mu^2$ |
| A_2B_1 | A_2B_1 | 0 | $1-\mu$ | 0 | μ |
| A_2B_1 | A_1B_2 | $r\frac{1-k}{2}$ | $\frac{1-r}{2}(1-\mu)$ | $\frac{1-r}{2}(1-\mu)$ | $r\frac{1+k}{2}$ |
| A_2B_1 | A_2B_2 | 0 | $\frac{1-k}{2}$ | 0 | $\frac{1+k}{2}$ |
| A_1B_2 | A_1B_1 | $\frac{1-k}{2}(1-\mu)^2$ | $\frac{1+k}{2}\mu(1-\mu)$ | $\frac{1-k}{2}\mu(1-\mu)$ | $\frac{1+k}{2}\mu + \frac{1-k}{2}\mu^2$ |
| A_1B_2 | A_2B_1 | $r\frac{1-k}{2}$ | $\frac{1-r}{2}(1-\mu)$ | $\frac{1-r}{2}(1-\mu)$ | $r\frac{1+k}{2}$ |
| A_1B_2 | A_1B_2 | 0 | 0 | $1-\mu$ | μ |
| A_1B_2 | A_2B_2 | 0 | 0 | $\frac{1-k}{2}$ | $\frac{1+k}{2}$ |
| A_2B_2 | A_1B_1 | $(1-r)\frac{1-k}{2}(1-\mu)^2$ | $(1-r)\frac{1-k}{2}\mu(1-\mu) + \frac{r}{2}$ | $(1-r)\frac{1-k}{2}\mu(1-\mu) + \frac{r}{2}$ | $(1-r)\frac{1+k}{2}\mu + (1-r)\frac{1-k}{2}\mu^2$ |
| A_2B_2 | A_2B_1 | 0 | $\frac{1-k}{2}$ | 0 | $\frac{1+k}{2}$ |
| A_2B_2 | A_1B_2 | 0 | 0 | $\frac{1-k}{2}$ | $\frac{1+k}{2}$ |
| A_2B_2 | A_2B_2 | 1 | 0 | 0 | 0 |

Table 3: Transmission probabilities, $b_{ij}^{kl}(mn)$, with cytonuclear inheritance. The **A** locus is biparentally inherited with μ the mutation probability from A_1 to A_2 . The **B** locus is uniparentally inherited with ν the mutation probability from B_1 to B_2 .

| Parents | | Offspring | | | |
|----------|----------|----------------------------|----------------------------|----------------------|----------------------|
| Mother | Father | A_1B_1 | A_2B_1 | A_1B_2 | A_2B_2 |
| A_1B_1 | A_1B_1 | $(1 - \mu)(1 - \nu)$ | $\mu(1 - \nu)$ | $(1 - \mu)\nu$ | $\mu\nu$ |
| A_1B_1 | A_2B_1 | $\frac{1-\mu}{2}(1 - \nu)$ | $\frac{1+\mu}{2}(1 - \nu)$ | $\frac{1-\mu}{2}\nu$ | $\frac{1+\mu}{2}\nu$ |
| A_1B_1 | A_1B_2 | $(1 - \mu)(1 - \nu)$ | $\mu(1 - \nu)$ | $(1 - \mu)\nu$ | $\mu\nu$ |
| A_1B_1 | A_2B_2 | $\frac{1-\mu}{2}(1 - \nu)$ | $\frac{1+\mu}{2}(1 - \nu)$ | $\frac{1-\mu}{2}\nu$ | $\frac{1+\mu}{2}\nu$ |
| A_2B_1 | A_1B_1 | $\frac{1-\mu}{2}(1 - \nu)$ | $\frac{1+\mu}{2}(1 - \nu)$ | $\frac{1-\mu}{2}\nu$ | $\frac{1+\mu}{2}\nu$ |
| A_2B_1 | A_2B_1 | 0 | $1 - \nu$ | 0 | ν |
| A_2B_1 | A_1B_2 | $\frac{1-\mu}{2}(1 - \nu)$ | $\frac{1+\mu}{2}(1 - \nu)$ | $\frac{1-\mu}{2}\nu$ | $\frac{1+\mu}{2}\nu$ |
| A_2B_1 | A_2B_2 | 0 | $1 - \nu$ | 0 | ν |
| A_1B_2 | A_1B_1 | 0 | 0 | $1 - \mu$ | μ |
| A_1B_2 | A_2B_1 | 0 | 0 | $\frac{1-\mu}{2}$ | $\frac{1+\mu}{2}$ |
| A_1B_2 | A_1B_2 | 0 | 0 | $1 - \mu$ | μ |
| A_1B_2 | A_2B_2 | 0 | 0 | $\frac{1-\mu}{2}$ | $\frac{1+\mu}{2}$ |
| A_2B_2 | A_1B_1 | 0 | 0 | $\frac{1-\mu}{2}$ | $\frac{1+\mu}{2}$ |
| A_2B_2 | A_2B_1 | 0 | 0 | 0 | 1 |
| A_2B_2 | A_1B_2 | 0 | 0 | $\frac{1-\mu}{2}$ | $\frac{1+\mu}{2}$ |
| A_2B_2 | A_2B_2 | 0 | 0 | 0 | 1 |

Table 4: Transmission probabilities, $b_{ij}^k(m,m)$, with cultural inheritance. Parental trait combinations are broken up with probability r , followed by biased transmission (A_2B_1 and A_1B_2 are passed down over A_1B_1 with probability q , A_2B_2 is passed down over A_1B_1 with probability p), and mutation with probability μ .

| Parents | | Offspring | | | |
|----------|----------|-----------------------|---|---|--|
| Mother | Father | A_1B_1 | A_2B_1 | A_1B_2 | A_2B_2 |
| A_1B_1 | A_1B_1 | $(1-\mu)^2$ | $\mu(1-\mu)$ | $\mu(1-\mu)$ | μ^2 |
| A_1B_1 | A_2B_1 | $(1-q)(1-\mu)^2$ | $(1-q)\mu(1-\mu) + q(1-\mu)$ | $(1-q)\mu(1-\mu)$ | $(1-q)\mu^2 + q\mu$ |
| A_1B_1 | A_1B_2 | $(1-q)(1-\mu)^2$ | $(1-q)\mu(1-\mu)$ | $(1-q)\mu(1-\mu) + q(1-\mu)$ | $(1-q)\mu^2 + q\mu$ |
| A_1B_1 | A_2B_2 | $(1-r)(1-p)(1-\mu)^2$ | $(1-r)(1-p)\mu(1-\mu) + \frac{r}{2}(1-\mu)$ | $(1-r)(1-p)\mu(1-\mu) + \frac{r}{2}(1-\mu)$ | $(1-r)[(1-p)\mu^2 + p] + r\mu$ |
| A_2B_1 | A_1B_1 | $(1-q)(1-\mu)^2$ | $(1-q)\mu(1-\mu) + q(1-\mu)$ | $(1-q)\mu(1-\mu)$ | $(1-q)\mu^2 + q\mu$ |
| A_2B_1 | A_2B_1 | 0 | $1-\mu$ | 0 | μ |
| A_2B_1 | A_1B_2 | $r(1-p)(1-\mu)^2$ | $\frac{1-r}{2}(1-\mu) + r(1-p)\mu(1-\mu)$ | $\frac{1-r}{2}(1-\mu) + r(1-p)\mu(1-\mu)$ | $(1-r)\mu + r[p + (1-p)\mu^2]$ |
| A_2B_1 | A_2B_2 | 0 | $(\frac{1}{2}-p+q)(1-\mu)$ | 0 | $(\frac{1}{2}-p+q)\mu + \frac{1}{2}-p+q$ |
| A_1B_2 | A_1B_1 | $(1-q)(1-\mu)^2$ | $(1-q)\mu(1-\mu)$ | $(1-q)\mu(1-\mu) + q(1-\mu)$ | $(\frac{1}{2}-p+q)\mu + \frac{1}{2}-p+q$ |
| A_1B_2 | A_2B_1 | $r(1-p)(1-\mu)^2$ | $\frac{1-r}{2}(1-\mu) + r(1-p)\mu(1-\mu)$ | $\frac{1-r}{2}(1-\mu) + r(1-p)\mu(1-\mu)$ | $(1-r)\mu + r[p + (1-p)\mu^2]$ |
| A_1B_2 | A_1B_2 | 0 | 0 | $1-\mu$ | μ |
| A_1B_2 | A_2B_2 | 0 | 0 | $(\frac{1}{2}-p+q)(1-\mu)$ | $(\frac{1}{2}-p+q)\mu + \frac{1}{2}-p+q$ |
| A_2B_2 | A_1B_1 | $(1-r)(1-p)(1-\mu)^2$ | $(1-r)(1-p)\mu(1-\mu) + \frac{r}{2}(1-\mu)$ | $(1-r)(1-p)\mu(1-\mu) + \frac{r}{2}(1-\mu)$ | $(1-r)[(1-p)\mu^2 + p] + r\mu$ |
| A_2B_2 | A_2B_1 | 0 | $(\frac{1}{2}-p+q)(1-\mu)$ | 0 | $(\frac{1}{2}-p+q)\mu + \frac{1}{2}-p+q$ |
| A_2B_2 | A_1B_2 | 0 | 0 | $(\frac{1}{2}-p+q)(1-\mu)$ | $(\frac{1}{2}-p+q)\mu + \frac{1}{2}-p+q$ |
| A_2B_2 | A_2B_2 | 0 | 0 | 0 | 1 |