SUPPLEMENTARY TEXT

Contents

1 Fisher's Geometric Model Simulations

We model adaptive walks in sexual diploid populations with Wright-Fisher simulations using Fisher's geometric model (FGM) as in Sellis *et al.* (2011). FGM is a phenotypic model, where the phenotype is represented as a point in n-dimensional coordinate space. We assume phenotypes are additive, such that the phenotype of a diploid individual is the midpoint of the phenotypes of the constituent alleles (Sellis *et al.* 2011). Note that this does not assume that alleles are additive in fitness space. Mutations in this system are vectors that modify the phenotype of the allele.

Since FGM is a phenotypic model, the underlying genetic basis for the mutations is not explicitly defined. For the purposes of this work, we will assume that there is a large number of completely linked loci underlying the phenotype, resulting in an infinite alleles model with no recombination. Each phenotypic mutation vector could be thought of as representing a mutation at a unique locus in this underlying genotype space, analogous to different functional mutations in a single gene.

1.1 FGM models

We explore three different parameterizations of FGM. In all parameter regimes, the population initially contains a single allele with a distance of 2 units from the optimum. The first two regimes use a symmetrical Gaussian fitness landscape with a single phenotypic optimum at the origin of the form

$$w(x) = e^{-x^2/2} \tag{1}$$

in two and 25 phenotypic dimensions, respectively, where x is the distance of the

individual's phenotype to the optimum. The mutation rate is set to $\mu = 5 * 10^{-6}$, which results in one mutation every 20 generations on average with a diploid population of size $N = 5000$. Mutations are vectors in phenotype space that get added to the phenotypic position of the underlying allele to generate a new allele. The angle of the mutation vector is drawn from a spherically uniform distribution, while the magnitude of the mutation vector is drawn from an exponential distribution. For the two dimensional regime, the mean of the mutational magnitude is 0.5, while for the 25 dimensional regime, the mean is set to 5. The mutational magnitudes were chosen to generate sufficient numbers of adaptive walks both with and without overdominant mutations.

We are forced to use a larger mutational magnitude for the 25 dimensional regime, as the magnitude of the component of the mutation vector in the direction of the phenotypic optimum becomes smaller when the number of dimensions increases. Once this component becomes too small, most mutations are nearly neutral in fitness effect and are mostly lost from the population. Thus, without increasing the mutational magnitude to compensate for this effect, it would be impossible for us to generate adaptive trajectories of sufficient length for our analysis. Despite this correction, we still had trouble sampling a sufficient number of trajectories without overdominant mutations for our statistics, as the fraction of adaptive mutations that are overdominant increases with increasing dimensionality under phenotypic additivity (Sellis *et al.* (2011), Figure S2), so we had to run additional simulations until we obtained another 200 simulations with at least 5 adaptive mutations that did not contain any balanced states.

The third parametrization of FGM is another two-dimensional landscape with a different Gaussian fitness function. The Gaussian fitness surface was first used by Lande (1976) under an assumption that the population is initially close enough to the optimum to be adapting under a concave-down surface (fitness increases at a slower rate as the phenotype

approaches the optimum). In our original simulations, the initial population is in a concave-up portion of the fitness surface, so we conducted further simulations with the above fitness function to ensure that the shape of the local fitness surface did not have a qualitative impact on our results. These additional simulations in two dimensions are conducted with:

$$w(x) = e^{-x^2/18} \tag{2}$$

This parametrization is chosen such that the initial population, at distance 2 from the optimum, is in the concave-down portion of the Gaussian curve and thus close to the optimum. The remainder of the parameters (μ, N, and mutational magnitude) are identical to the previous two-dimensional regime.

1.2 Implementation of simulations

The simulations use the code modified from Sellis *et al.* (2011) to allow for more than 2 dimensions. We perform 10,000 replicate simulations using a standard Wright-Fisher approach (Fisher 1930; Wright 1931) . Simulations are conducted for 10,000 generations, where each generation consists of mutating alleles and then propagating alleles to the next generation. For propagation, all possible offspring genotypes are computed and assigned a weight proportional to their frequency (assuming that all offspring genotypes are in Hardy-Weinberg equilibrium based on the allele frequencies of the previous generation) multiplied by the fitness of the diploid offspring genotype. We then conduct multinomial sampling over these weighted offspring genotypes to determine the frequencies of the alleles

in the next generation (i.e., viability selection on the offspring). Complete source code is available at `https://github.com/sunthedeep/Fisher-Geometric-Model`.

To conduct backward predictability inference, we identify the most frequent allele in each simulated population at the end of 10,000 generations of evolution and study the mutations present on that allele. We limit our analysis to studying the first five mutations of each adaptive walk and ignore simulations with fewer than five mutations in order to control for the length of the adaptive walk when studying predictability. We partition our five-mutation adaptive walks into those that do and those that do not contain overdominant mutations to study the impact of balanced states on predictability.

1.3 Partitioning walks

Throughout all of our analysis, we have separated walks with and without overdominant mutations. The methodology for this separation is as follows. For each FGM simulation, we have identified the most frequent allele at the end of the simulation, and isolated the first five mutations to occur on this allele. We first determine the time t_5 at which the allele containing these first five mutations exceeded 5% frequency in the population. All time-points after t_5 are no longer considered for analysis. At each generation $t <= t_5$, we isolate all alleles in the population at $>= 1\%$ frequency. For every subset of these alleles, we compute their equilibrium frequencies and mean fitness using the method of Kimura (1956) (ST1.4). If a set of alleles generates a stable polymorphic state at equilibrium, we infer that there is an overdominant mutation present among those alleles. An FGM simulation is determined to contain an overdominant mutation if, for any generation $t <= t_5$, the subset of alleles with the highest mean fitness at generation t is a stable polymorphism at equilibrium. For simplicity, we removed simulations that contained stable polymorphisms with $>= 3$ alleles for $>= 50$ generations so that we only need to consider 2 allele balanced states for the remainder of this work.

1.4 Computing whether a set of alleles generates a stable equilibrium

The method of Kimura (1956) generates a square fitness matrix A of size n, where n is the number of alleles present at the locus at non-zero frequency. The value of $A_{i,j}$ (row i, column j of matrix A) is the fitness of the genotype containing alleles i and j. The system of alleles is stable if: 1) by replacing each column of A with 1's and computing the determinant of the resulting matrix, the sign of the determinant is always positive and 2) the matrix T, where $t_{i,j} = A_{i,j} - A_{i,n} - A_{j,n} + A_{n,n}$, must be negative definite. The frequencies of each allele and mean fitness can also be computed from these matrices. For a further discussion of computing the stability of a balanced system, please see Kojima (1959); Mandel (1959); Kingman (1961).

1.5 Identification of hidden alleles

We identify hidden alleles (Figure 4b) by comparing the set of alleles present in the equilibrium state of each generation throughout a given FGM simulation from ST1.3 to the set of 5 alleles along the direct mutational trajectory from the ancestral allele to the 5-mutant allele (i.e. the 0, 1, 2, 3, 4 and 5- mutant alleles). Any allele that is present in the equilibrium states, but not part of the mutational trajectory of the 5-mutant allele, is deemed a hidden allele.

2 Backward Predictability Inference

Backward predictability inference seeks to reconstruct the order in which a set of mutations arose in an adaptive trajectory. By estimating the likelihood of every possible order, we can try to predict the "true" adaptive trajectory as the inferred trajectory with the highest probability. We can also study the probability distribution of all of the possible adaptive trajectories to understand how predictable the system is overall.

Experimental studies conducting backward predictability inference assumed a strong selection / weak mutation (SSWM) model of evolution (Weinreich *et al.* 2006), which we also use in our study. With the SSWM assumption, the population is assumed to reach equilibrium after the successful invasion of each mutation before the next mutation is introduced. This is appropriate as our per-generation mutation rate of 0.05 is much smaller than one. For simplicity of analysis. we also assume that all balanced states contain exactly two alleles, as we have excluded all simulations that generated balanced states with 3 or more alleles.

We utilize two variants of the backward predictability inference method. The first, which we call the fixation assumption method (FA method), assumes that every mutation that successfully invades the population reaches fixation, and is comparable to the method of Weinreich *et al.* (2006). The second, which we call the polymorphism assumption method (PA method), allows for the presence of stable two-allele polymorphic states.

We conduct backward predictability inference within the framework of FGM. We explicitly model the phenotypes of the alleles and mutation vectors, and use the same fitness functions as in the FGM simulations to compute fitness.

2.1 Computing the likelihood of a particular order of mutations

We begin with an overview of the backward predictability inference method used by Weinreich *et al.* (2006) and then continue on to a description of our implementation of the FA and PA methods.

Weinreich et al (2006) inference method Weinreich et al (2006) describe the probability of the ancestral allele (A^{wt}) evolving into the derived allele containing all 5 mutations available (A^{der}) going through a particular order of mutations (M_i) with

intermediate alleles a, b, c and d. This can be computed as

$$\begin{aligned}\Pr(M_i) &= \Pr(A^{wt} \to a \to b \to c \to d \to A^{der}) \\ &= \Pr(A^{wt} \to a) * \Pr(a \to b) * \Pr(b \to c) * \Pr(c \to d) * \Pr(d \to A^{der}) \quad (3)\end{aligned}$$

because "along any particular trajectory the choice of each next fixation is statistically independent of all previous fixations. Here, the $\Pr(i \to j)$ are the conditioned fixation probabilities of a particular single mutant neighbour j of an allele i given by

$$\Pr(i \to j) = \frac{\Pi_{i \to j}}{\sum\limits_{k \epsilon N_i} \Pi_{i \to k}} \quad (4)$$

where $\Pi_{i \to j}$ is the unconditioned fixation probability of allele j from allele i, and N_i is the set of all mutational neighbours of allele i." (modified from Weinreich *et al.* (2006) Supplementary Methods). In essence, (Weinreich *et al.* 2006) compute the probability of a particular order of mutations as the product of the probabilities of each mutation in that order successfully fixing in the population in succession.

The FA and PA methods Our methods for backward predictability inference are necessarily more complicated.

First, since we are using a diploid model, new mutations occur as heterozygotes and thus must invade the population as heterozygotes. Therefore, we cannot compute the fixation probability, but must compute the probability of an allele successfully invading the population from low frequency and reaching its equilibrium frequency. Secondly, in the presence of a balanced polymorphism in the PA method, new mutations can occur on multiple available backgrounds. This allows for the generation of hidden alleles when

conducting the PA method. This also implies that it may take more than 5 mutations in a mutation order to generate the allele with all 5 mutations. Finally, a new mutation in the PA method that successfully invades can either fix or balance with any of the alleles already present in the population, whereas a successful invasion by new mutation in the FA method can only result in fixation. As these properties make it challenging to describe the FA and PA methods using closed form analytic equations as in WEINREICH *et al.* (2006), we will describe the recursive algorithm we use to implement the FA and PA methods using pseudocode. Every call to the algorithm keeps requires a population state (set of alleles and their frequencies), a set of alleles observed during the recursion and the probability of the mutation order so far. Using global variables outside of the algorithm, we keep track of $\Phi(M_i)$, the unconditioned probability of every possible mutation order M_i. All $\Phi(M_i)$ are initialized to 0.

computeBackwardInference($S_{existing}$, $A^{existing}$, $P_{existing}$)

$S_{existing} \leftarrow$ the population state = a set of alleles and their frequencies
$A^{existing} \leftarrow$ the set of alleles observed so far in this mutation order
$P_{existing} \leftarrow$ the unconditioned probability of this order of mutations so far
if $A^{der} \epsilon S_{existing}$ **then**
    We need to first determine the order M_i in which the mutations were introduced into A^{der} and add $P_{existing}$ to the unconditioned probability for this order of mutations ($\Phi(M_i)$)
    return // We are done since we have successfully generated A^{der}
else
    $\rho_{total} = 0$
    for all new alleles A^n that can be generated by a single mutation on the alleles in $S_{existing}$, excluding those where $A^n \in A_{existing}$ **do**
        for all pairs of alleles A^i, A^j in the set of alleles including A^n and every allele in $S_{existing}$ **do**
            Compute the frequency of A^i and A^j and the mean fitness of the population at equilibrium assuming these are the only two alleles in the population
        S_{new} =the pair of alleles and their frequencies with the highest mean fitness computed in the preceding for loop excluding all alleles at frequency 0.
        if $A^n \notin S_{new}$ **then**
            A^n cannot invade $S_{existing}$ and can thus be ignored
        else
            compute $P^i_{A^n}$ = the probability of invasion of A^n into $S_{existing}$ through 10,000 forward Wright-Fisher simulations
            The unconditioned probability of A^n succeeding in this population $\rho_n = P^i_{A^n} *$ the frequency of the allele in $S_{existing}$ that was mutated to generate A^n
            $\rho_{total} += \rho_n$
    for all new alleles A^n with $\rho_n > 0$ **do**
        S_{new} and ρ_n defined as above for A^n
        if Using the FA method **then**
            $S_{new} = A^n$ at frequency 1 (fixation)
        $A_{new} = A_{existing} \cup A^n$
        $P_{new} = P_{existing} * \frac{\rho_n}{\rho_{total}}$
        computeBackwardInference(S_{new}, A_{new}, P_{new}) // recursive call

The initial call to this algorithm has $S_{existing}$ be the ancestral population used in the FGM simulations i.e. a population monomorphic for an allele two units from the optimum, $A_{existing}$ as the set containing A^{wt} and $P_{existing} = 1$. Once we have computed the unconditioned probability for every M_i ($\Phi(M_i)$), we then use this information to compute the conditioned probability for each mutation order.

$$\Pr(M_i) = \frac{\Phi(M_i)}{\sum_j \Phi(M_j)} \tag{5}$$

Note that we track mutation orders by the order in which the mutations were introduced on allele A^{der}, which is always five mutations long, not the order in which the mutations were introduced in the population which is >= 5 mutations with the PA method but always exactly 5 mutations with the FA method.

In both the FA and PA methods, we compute the invasion probability of a new mutation $P^i_{A^n}$ using 10,000 forward Wright-Fisher simulations. In these simulations, we set N = 5,000 diploid individuals as in our FGM simulations, with no new mutations allowed.

The probability of a new allele successfully invading and reaching the deterministically inferred stable equilibrium is then the fraction of Wright-Fisher simulations where A^n reaches 90% of its expected equilibrium frequency in S_{new}. These simulations are entirely separate from the FGM simulations used to generate the adaptive walks used throughout the rest of this work.

We are forced to utilize empirical estimations through simulations and not the classical analytic solutions to compute $P^i_{A^n}$ (Haldane 1927; Kimura 1962) as many of the observed mutations have a selective advantage exceeding 100%, violating the assumptions of the analytic solutions that the mutations are weakly beneficial. Our simulations suggest

that the analytic solutions significantly overestimate the invasion probability under these conditions (data not shown).

2.2 Testing the accuracy of the backward predictability inference methods

We first test the accuracy of the FA and PA methods by running additional forward FGM simulations. For each set of 5 mutations observed in one of the original FGM simulations, we conduct 1000 additional FGM simulations where the only available mutations are these five. We terminate the simulation either at 10,000 generations or when A^{der} reaches 5% frequency in the population, whichever occurs first, and remove from consideration all simulations where A^{der} did not occur by 10,000 generations. We estimate $Pr(M_i)$ as the fraction of the additional FGM simulations that have mutation order M_i.

2.3 Quantifying backward predictability

To quantitatively study the results of backward predictability inference across simulations, we define the effective number of paths statistic as

$$\frac{1}{\sum_i Pr(M_i)^2} \tag{6}$$

The effective number of trajectories is defined to be 0 when there are no viable trajectories, e.g. $\sum_i Pr(M_i)^2 = 0$. This is similar to the effective number of alleles in a population (Kimura and Crow 1964), the predictability metric of Roy (2009) and the entropy metric of Palmer *et al.* (2013).

When a single trajectory dominates the probability density, the effective number of

trajectories is close to 1, indicating high backward predictability. On the other hand, if every trajectory has equal probability, $Pr(M_i) = \frac{1}{n!}$ since we know that there must be $n!$ possible mutation orders for a system of n mutations. In this situation, the effective number of paths = $n!$ = total number of possible mutation orders, indicating low backward predictability. This provides a single metric of the diversity of mutational orders that are possible while accounting for their relative likelihoods and summarizes the backward predictability of the adaptive walk. We also use the mean path divergence metric, which uses the Hamming distance between every pair of viable mutation orders scaled by the inferred probability of the mutation orders to quantify backward predictability (Lobkovsky *et al.* 2011, 2013) and get consistent results.

2.4 Phenotypic similarity of inferred trajectories from backward predictability inference

The effective number of paths metric captures the backward predictability of a set of mutations in terms of the likelihood of observing a given order of mutations. However, we can also consider the deviation of each of these possible trajectories from the true trajectory observed in the FGM simulations in phenotype space. In particular, we want to know if the presence of an overdominant mutation in a simulation influences the phenotypic similarity of incorrect mutation orders to the true mutation order.

We compute a maximum distance metric between the phenotypic states of each inferred trajectory and the phenotypic states of the true trajectory that were identified in ST1.3, excluding any inferred trajectory that matches the mutation order of the true trajectory. We first compute the average phenotype in each trajectory in all generations $t < t_5$. We use the alleles and their frequencies of equilibrium states computed in every generation $t < t_5$ in section 1.3, and compute the average phenotype of the generation as the midpoint of the phenotypes of every genotype weighted by their frequencies at Hardy-Weinberg

equilibrium. We compute the largest distance from any phenotype in the alternative mutation order to the phenotypes of the true adaptive trajectory. We then average all of these phenotypic distances for all inferred trajectories to compute the average phenotypic distance statistic. We also conduct randomized trials where we randomly select a viable inferred trajectory as the true trajectory and redo the analysis.

2.5 Maximum distance of observed trajectory from the optimal trajectory as a metric of forward predictability

For every forward FGM simulation, we compute the average phenotype of the population at every generation (the the midpoint of the phenotypes of every genotype present in the population weighted by their frequencies). We then compute the minimum distance of each these average phenotypes from the optimal trajectory (the line segment connecting the ancestral phenotype and optimal phenotype) for each generation $t < t_5$, and then take the maximum of these values as the maximal distance of the observed trajectory from the optimal trajectory for that particular simulation in a manner similar to the phenotypic similarity metric used in section 2.4.

2.6 Example of backward predictability inference

As a concrete example of how backward predictability inference is conducted, we will use a system of two mutations, m_1 and m_2. Let A^{wt} be the ancestral allele, A^1 be the allele containing mutation m_1, A^2 be the allele with m_2 and $A^{1,2} = A^{der}$ be the the derived allele containing both available mutations m_1 and m_2.

There are $2! = 2$ different orders of mutations that can generate allele A^{der}. In the mutation order under consideration, m_1 occurs first, then m_2:

$M_1 = A^{wt} \xrightarrow{m_1} A^1 \xrightarrow{m_2} A^{der}$

the remaining mutation order is:

$M_2 = A^{wt} \xrightarrow{m_2} A^2 \xrightarrow{m_1} A^{der}$

Let us first consider the FA method, using example data in Table S1, the results of which are shown in Figure S11. Using our recursive procedure, we start with $S_{existing} = A^{wt}$ at fixation. There are two possible mutations in this population, where m_1 occurs on A^{wt} to get A^1 and where m_2 occurs on A^{wt} to get A^2. We compute the unconditioned probability of allele A^1 successfully being generated and invading the population, ρ_{A^1} on $S_{existing}$ and similarly for A^2. These are 0.3 and 0.5, respectively. We then compute the conditioned probabilities of success for these alleles. For A^1, this is $\frac{0.3}{0.3+0.5} = 0.375$, while it is $\frac{0.5}{0.3+0.5} = 0.625$ for A^2 (Table S1). In essence, a single successful mutation in $S_{existing}$ (A^{wt} fixed in the population) will generate an S_{new} of A^1 fixed in the population 0.375 of the time, and an S_{new} of A^2 fixed in the population 0.625 of the time. From here, we can then do the recursive call for the next step of the inference procedure for each of these new population states. These two recursive calls will be: 1) $S_{new} = A^1$ fixed in the population, with $A_{new} = A^{wt}, A^1$ and $P_{new} = 1 * 0.375 = 0.375$ and 2) $S_{new} = A^2$ fixed in the population, with $A_{new} = A^{wt}, A^2$ and $P_{new} = 1 * 0.625 = 0.625$

Let us now consider the first of these recursive calls, when A^1 is the first successful allele to invade the population and $P_{existing}$ for this call is 0.375. In this case, there is only one available mutation, m_2, which will generate the fully adapted allele A^{der} with an unconditioned probability of 0.6 but a conditioned probability of 1. We now have $S_{new} = A^{der}$ fixed in the population, with a P_{new} of $0.375 * 1 = 0.375$. We then call the recursive condition again with this new S_{new}, where we find that the termination condition of having A^{der} in $S_{existing}$ has been reached. Therefore, we are done, and the unconditioned

probability of the mutation order used to get A^{der} this time, namely $A^{wt} \xrightarrow{m_1} A^1 \xrightarrow{m_2} A^{der}$ is 0.375.

A similar procedure with the other initial recursive call, where m_2 was the first mutation, finds that the unconditioned probability of the mutation order $A^{wt} \xrightarrow{m_2} A^2 \xrightarrow{m_1} A^{der}$ is 0.625. Therefore, we find two viable orders of mutations, one with conditioned probability $\frac{0.375}{0.375+0.625} = 0.375$ and one with probability $\frac{0.625}{0.375+0.625} = 0.625$. Note that with the FA method, the conditioned probability for a mutation order always equals its unconditioned probability since the number of mutations introduced into the population is always equal to the number of mutations in the mutation order. This is not the case in the PA method, as we will see below.

Table S1

$S_{existing}$	Mutation	New Allele A^n	Invasion Prob $P^i_{A^n}$	ρ_n	$P_{existing}$	P_{new}	S_{new}	Mutation Order for Adapted Allele A^{der}
A^{wt}	m_1 on A^{wt}	A^1	0.3	0.3	1	0.375	A^1, freq = 1	
A^{wt}	m_2 on A^{wt}	A^2	0.5	0.5	1	0.625	A^2, freq = 1	
A^1	m_2 on A^1	A^{der}	0.6	0.6	0.375	0.375	A^{der}, freq = 1	$A^{wt} \xrightarrow{m_1} A^1 \xrightarrow{m_2} A^{der}$
A^2	m_1 on A^2	A^{der}	0.2	0.2	0.625	0.625	A^{der}, freq = 1	$A^{wt} \xrightarrow{m_2} A^2 \xrightarrow{m_1} A^{der}$

We then turn to the PA method, using a different pair of mutations with example data in Table S2 and the results in Figure S12. In this case, let us suppose that the A^1 can successfully invade the ancestral population consisting of A^{wt} to result in a balanced state consisting of A^1 and A^{wt} at intermediate frequencies. Meanwhile, A^2 can also successfully invade the ancestral population, but it fixes, resulting in S_{new} consisting of A^2 at frequency 1. Given their relative invasion probabilities and the fact that A^{wt} was initially fixed, we find that the conditioned probability of A^1 invading A^{wt} and resulting in a balanced state $= 1$ (freq of A^{wt}) * 0.2 (invasion probability of A^1) / $(1 * 0.2 + 1 * 0.35) = 0.36$, while the probability of A^2 being the next mutation in A^{wt} is 0.64.

For the next recursion step, let us consider the $S_{existing}$ of A^1 and A^{wt} at intermediate frequencies. There are two possible mutations in this scenario, in which mutation m_2 can occur on either A^1 or A^{wt} to generate alleles A^{der} and A^2, respectively. The successful invasion of A^2 results in a S_{new} containing a balanced state consisting of both A^1 and A^2. The new allele A^{der} can also successfully invade the population and results in a stable polymorphism as well. Mutation m_1 is not allowed to occur on A^{wt}, since that would regenerate allele A^1 which has already been observed in this trajectory so far. The conditioned probability of A^2 succeeding in this population is $\frac{0.7*0.14}{0.7*0.14+0.3*0.6} = 0.35$, while the conditioned probability of A^{der} succeeding is 0.65. The running probability of these two mutation orders after two mutations have been introduced in the population are $0.36 * 0.35 = 0.126$ and $0.36 * 0.65 = 0.234$, respectively.

Now let us consider the mutations on the $S_{existing}$ where both A^1 and A^2 exist as a balanced polymorphism. In this situation, there are two possible mutations, where m_1 can arise on A^2 to give the adapted allele A^{der}, and m_2 can arise on A^1 to also give the adapted allele A^{der}. Even though this is the same allele being generated by the two mutations, the initial frequency of A^1 and A^2 are different, giving rise to different unconditioned

probabilities of their occurrence. The m_1 mutation has a conditioned probability of $\frac{0.8*0.4}{0.8*0.4+0.2*0.4} = 0.8$, while the m_2 mutation has a conditioned probability of 0.2. The running probability after each of these mutations are $0.36 * 0.35 * 0.8 = 0.1008$ and $0.36 * 0.35 * 0.2 = 0.0252$, respectively.

The final possible trajectory, where m_2 occurred first on $S_{existing} = A^{wt}$ and resulted in the fixation of A^2 has only one possible mutation. This is mutation m_1 on A^2 resulting in the allele A^{der}. Supposing that $A^1, 2$ is deleterious in this situation, it cannot invade and therefore has 0 probability of occurring. We then terminate this recursion as there are no valid beneficial mutations available to this population.

Finally, we now need to compute the conditioned likelihoods of each mutation order. We managed to successfully get A^{der} in 3 different ways when considering the mutations introduced into the population, but only 2 different ways when considering the mutations introduced onto the allele that generated A^{der}. The unconditioned probabilities of these two different mutation orders are: $0.234 + 0.0252 = 0.2592$ for mutation order $M_1 = A^{wt} \xrightarrow{m_1} A^1 \xrightarrow{m_2} A^{der}$ and 0.1008 for mutation order $M_2 = A^{wt} \xrightarrow{m_2} A^2 \xrightarrow{m_1} A^{der}$. The conditioned probabilities for these two mutation orders are thus $\frac{0.2592}{0.2592+0.1008} = 0.72$ and $\frac{0.1008}{0.2592+0.1008} = 0.28$, respectively.

Table S2

$S_{existing}$	Mutation	A^n	$P^i_{A^n}$	ρ_n	$P_{existing}$	P_{new}	S_{new}	Mutation Order of $A^{der} \in S_{new}$
A^{wt}	m_1	A^1	0.2	0.2	1	0.36	A^1 freq = 0.3, A^{wt} freq = 0.7	
A^{wt}	m_2	A^2	0.35	0.35	1	0.64	A^2 freq = 1	
A^1 freq = 0.3, A^{wt} freq = 0.7	m_2 on A^{wt}	A^2	0.14	0.098	0.36	0.126	A^1 freq = 0.2, A^2 freq = 0.8	
A^1 freq = 0.3, A^{wt} freq = 0.7	m_2 on A^1	A^{der}	0.6	0.18	0.36	0.234	A^{der} freq = 0.8, A^{wt} freq = 0.2	$A^{wt} \xrightarrow{m_1} A^1 \xrightarrow{m_2} A^{der}$
A^1 freq = 0.2, A^2 freq = 0.8	m_1 on A^2	A^{der}	0.4	0.32	0.126	0.1008	A^{der} freq = 1	$A^{wt} \xrightarrow{m_2} A^2 \xrightarrow{m_1} A^{der}$
A^1 freq = 0.2, A^2 freq = 0.8	m_2 on A^1	A^{der}	0.4	0.08	0.126	0.0252	A^{der} freq = 1	$A^{wt} \xrightarrow{m_1} A^1 \xrightarrow{m_2} A^{der}$
A^2 freq = 1	m_1 on A^2	A^{der}	0	0	0.64	0	A^2 freq = 1	

3 Supplementary Figures

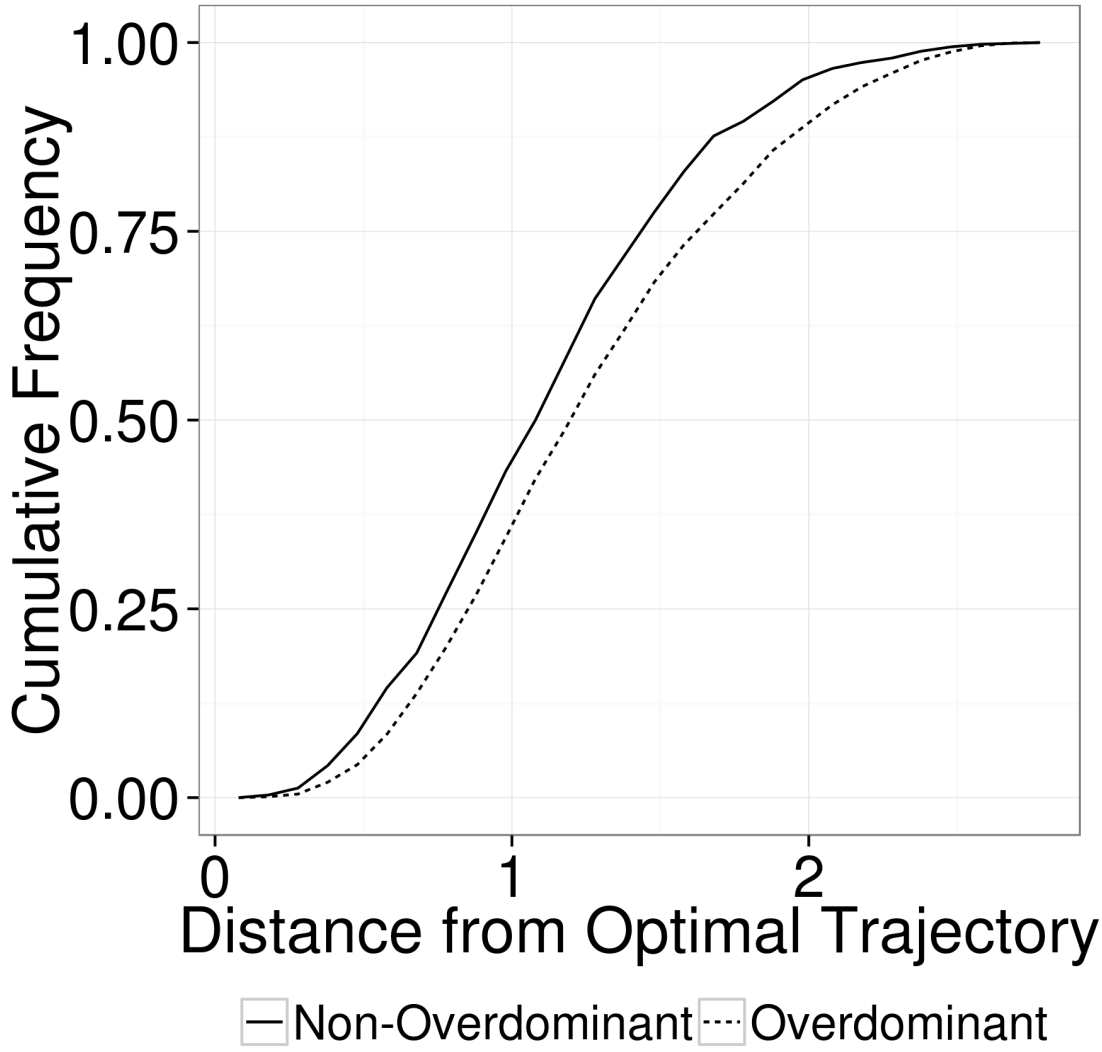


Figure S1. Maximum phenotypic deviation of the initial simulations in 2 dimensions with and without overdominant mutations from the optimal trajectory. Simulations without overdominant mutations are significantly closer to the optimal trajectory than those with overdominant mutations (Kolmogorov-Smirnov $p = 10^{-7}$).

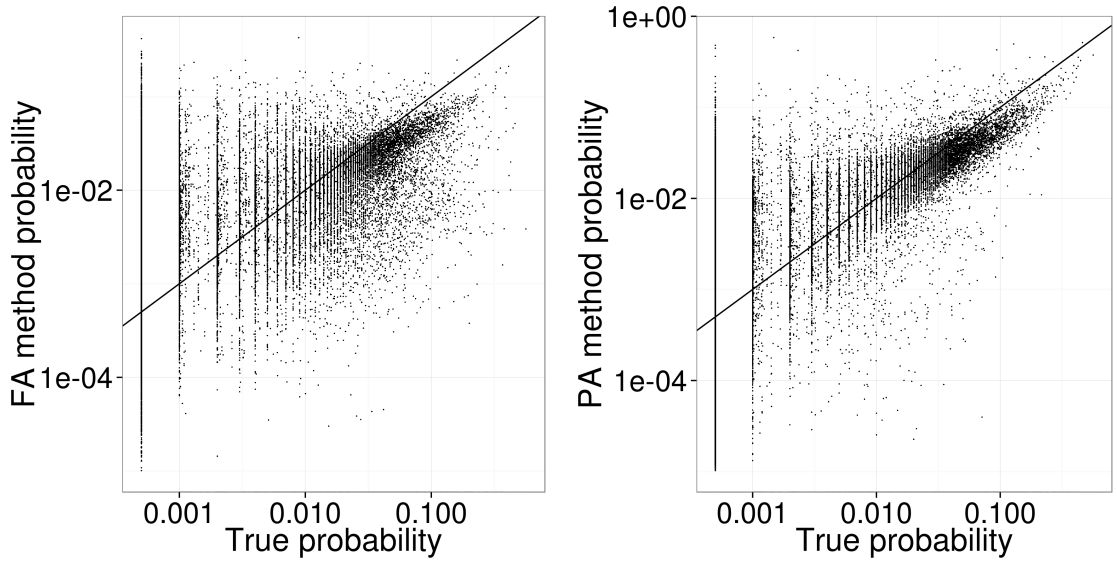


Figure S2. Similar to Figure 1 but for simulations at 25 dimensions. The inferred probabilities from both the FA method ($r^2 = 0.36$, $p < 10^{-10}$) and PA method ($r^2 = 0.65$, $p < 10^{-10}$) are significantly correlated with the true probabilities. The PA method is again significantly better correlated than the FA method ($p < 10^{-10}$).

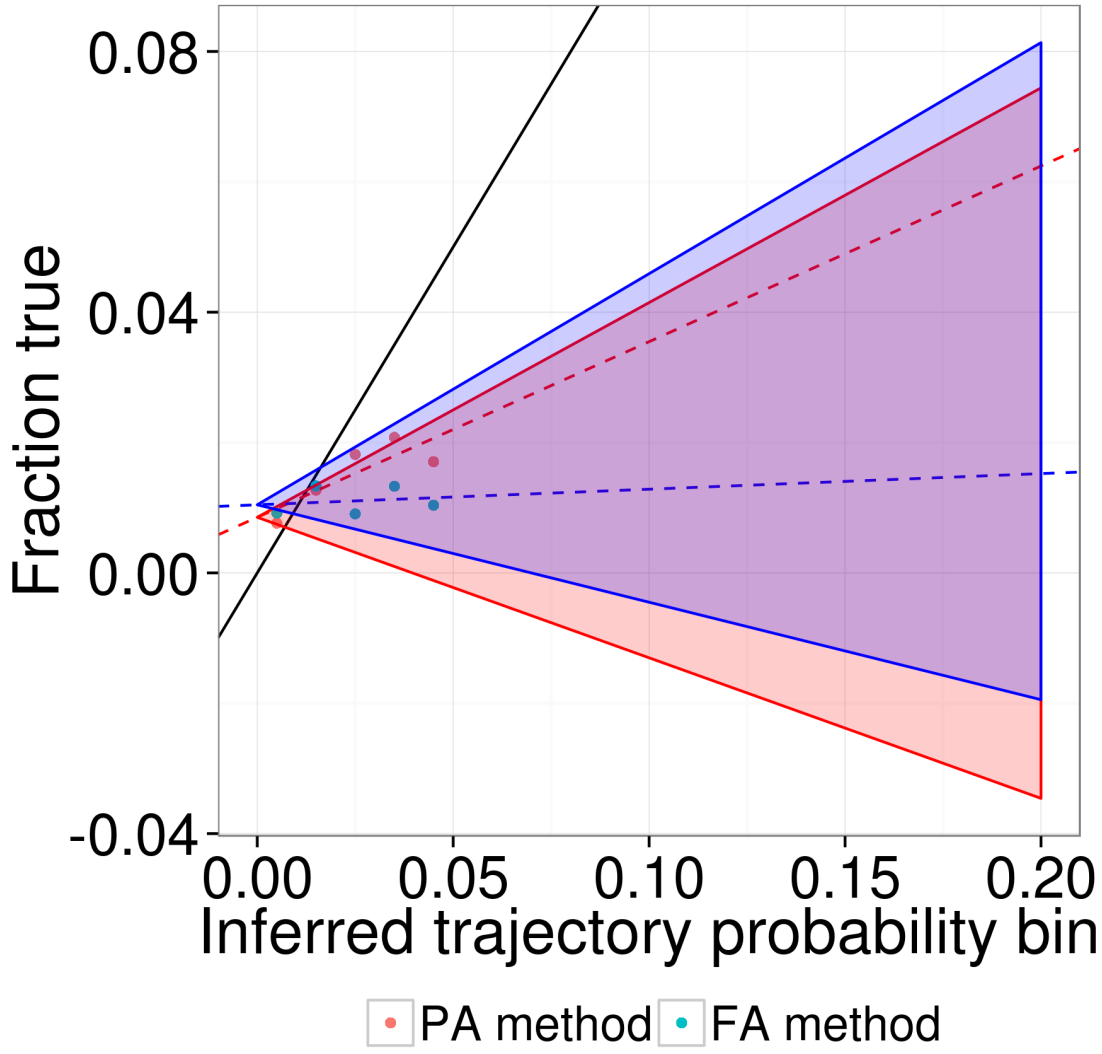
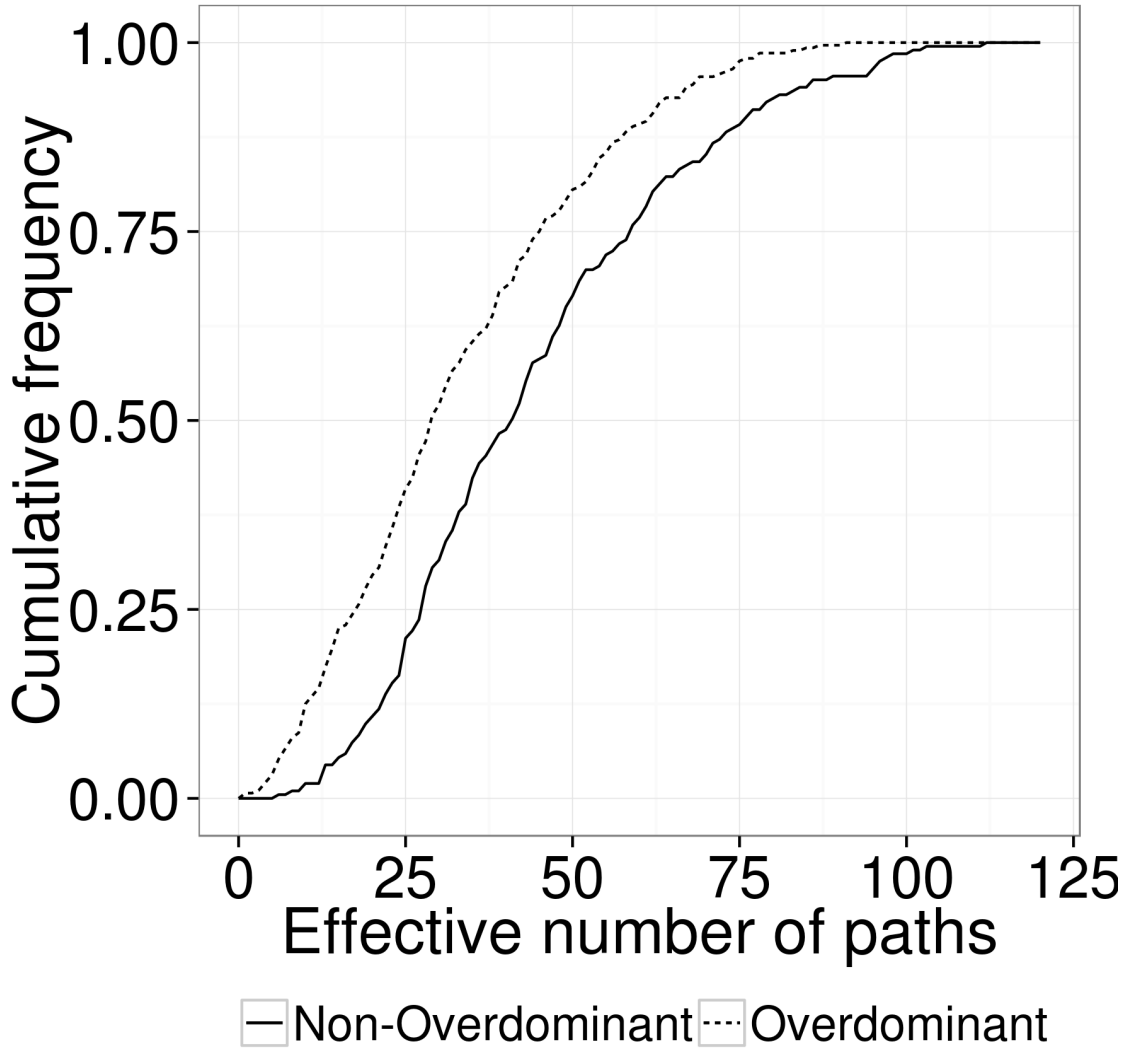
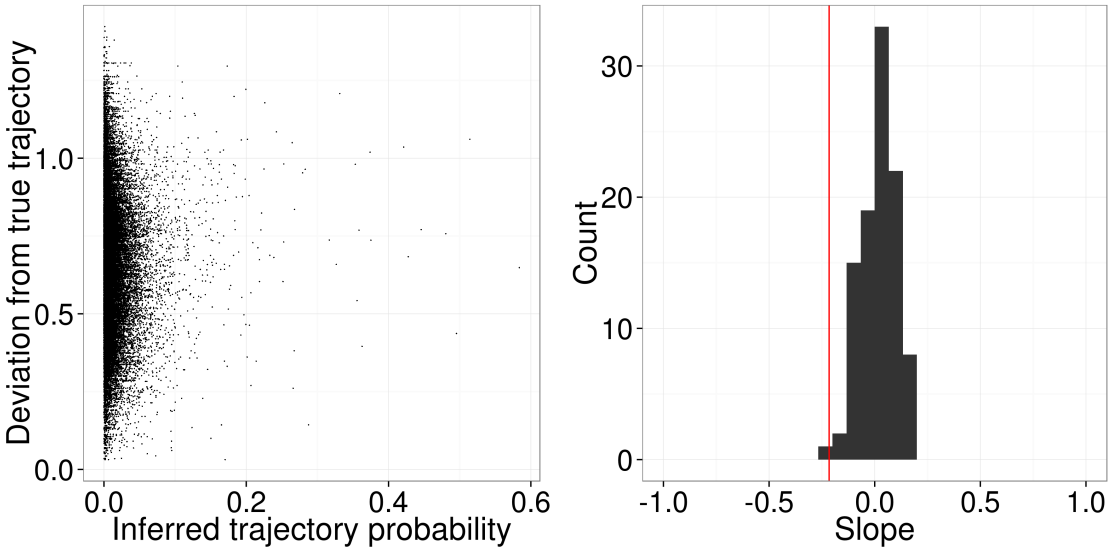


Figure S3. Similar to Figure 2 but for simulations in 25 dimensions. The slope for the FA method is not significantly different from the randomized trials (slope = 0.024, empirical $p = 0.733$), whereas the slope for the PA method is significantly better than all of the randomized trials (slope = 0.269, empirical $p = 0.009$).



370

371 **Figure S4.** Similar to Figure 3 but for 25 dimension simulations (Kolmogorov-Smirnov
372 $p < 10^{-10}$).



373

374 **Figure S5.** Similar to Figure 4 but for 25 dimension simulations (slope $= -0.22$, empirical

375 $p < 0.01$)

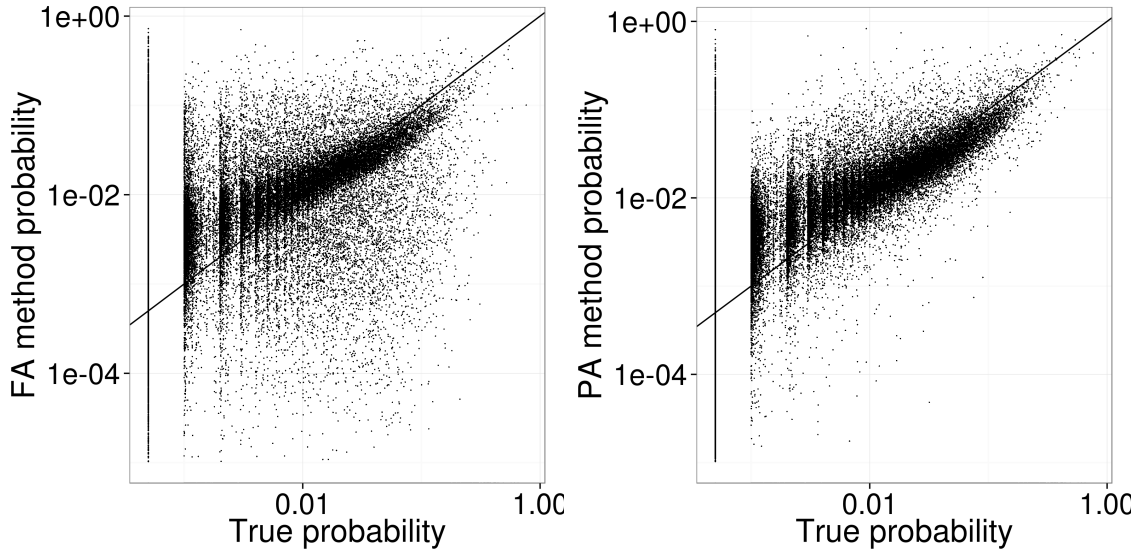


Figure S6. Similar to Figure 1 but for simulations in two dimensions close to the optimum. The inferred probabilities from both the FA method ($r^2 = 0.27$, $p < 10^{-10}$) and PA method ($r^2 = 0.52$, $p < 10^{-10}$) are significantly correlated with the true probabilities. The PA method is again significantly better correlated than the FA method ($p < 10^{-10}$).

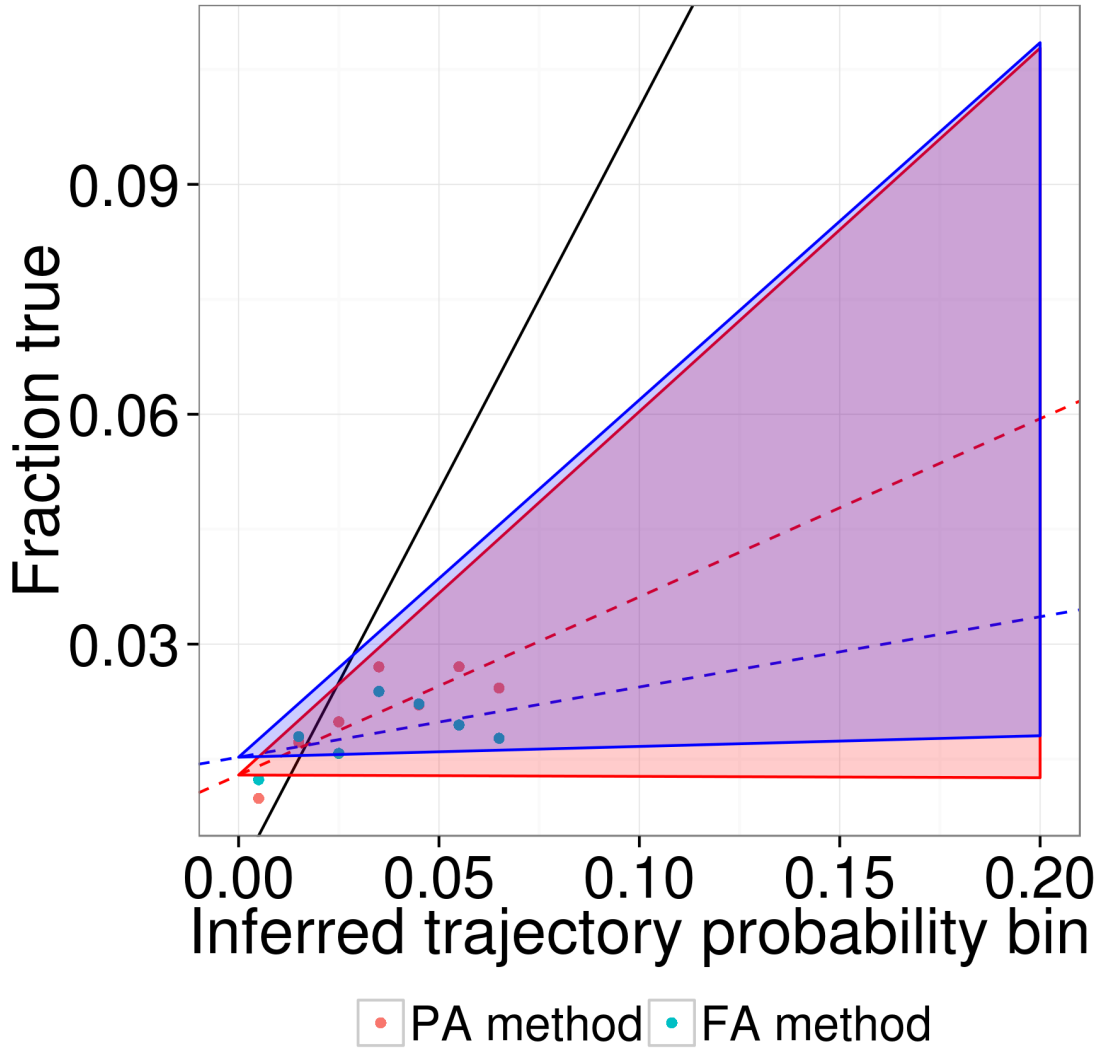
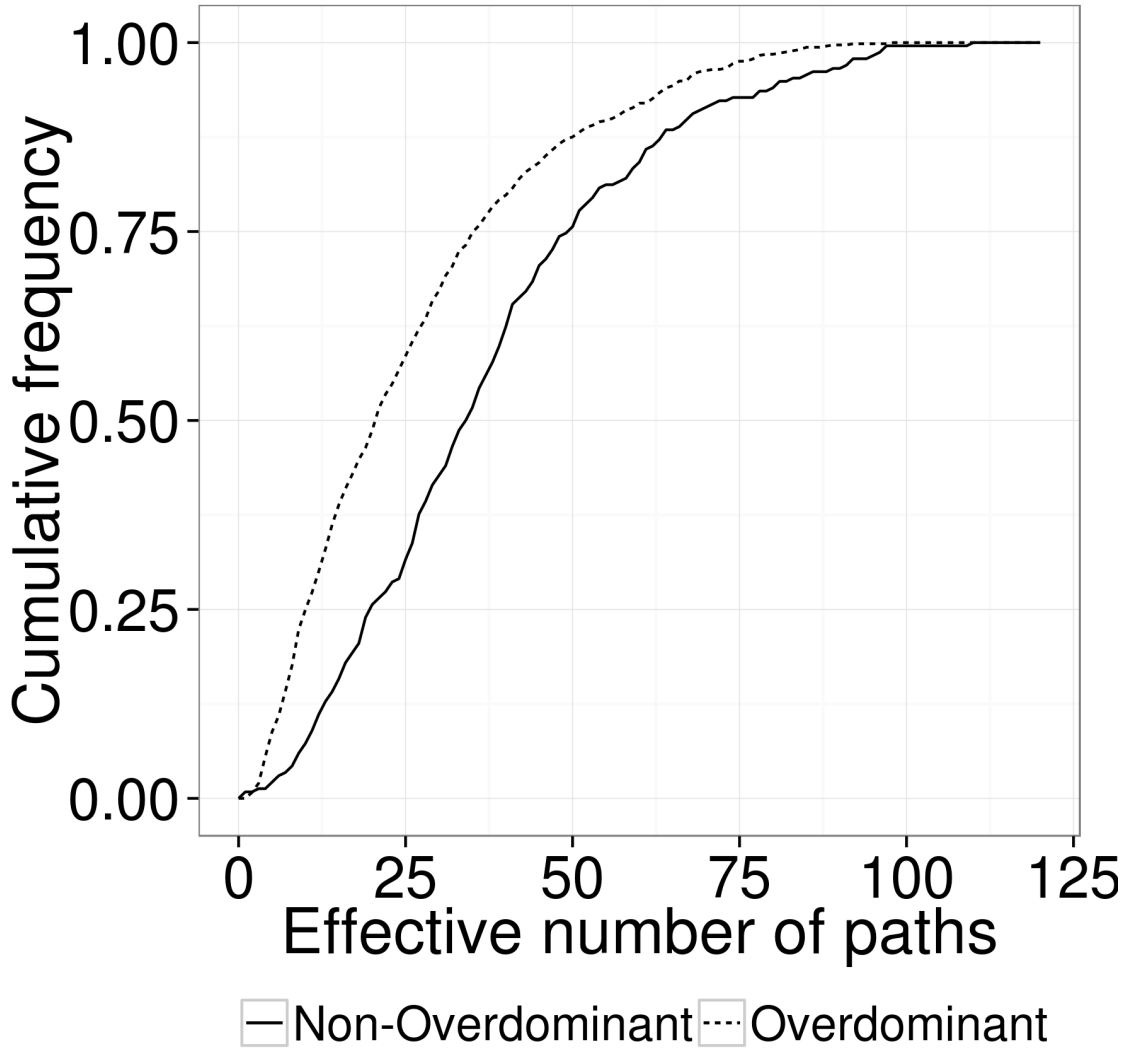


Figure S7. Similar to Figure 2 but for simulations in two dimensions close to the optimum. The slopes for both the FA method (slope = 0.092, empirical $p = 0.968$) and the PA method (slope = 0.232, empirical $p = 0.419$) are not significantly larger than the randomized trials.



386

387 **Figure S8.** Similar to Figure 3 but for simulations in two dimensions close to the

388 optimum (Kolmogorov-Smirnov $p = 10^{-10}$)

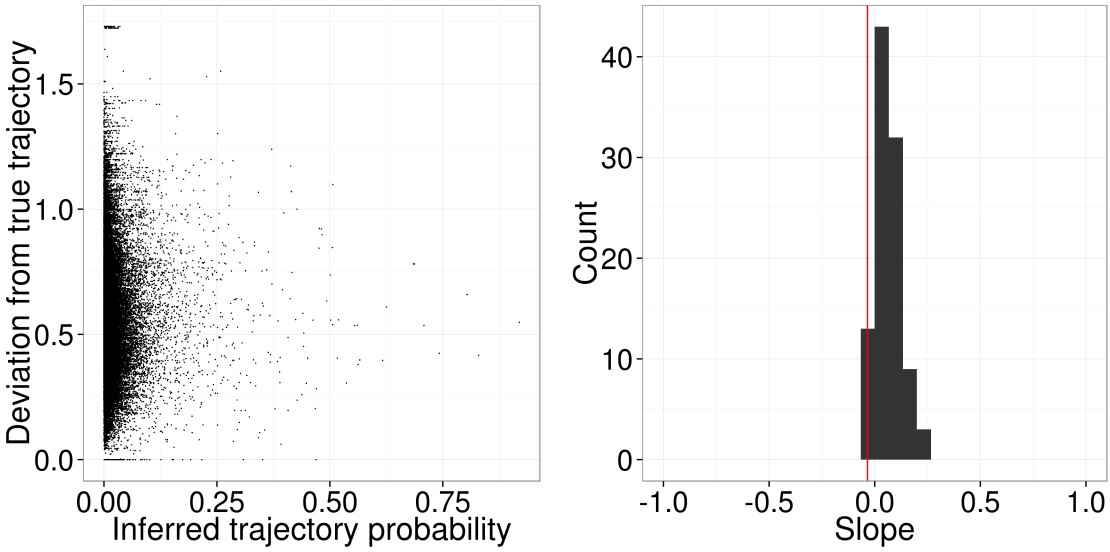


Figure S9. Similar to Figure 4 but for simulations in two dimensions close to the optimum (slope $= -0.03$, empirical $p = 0.02$)

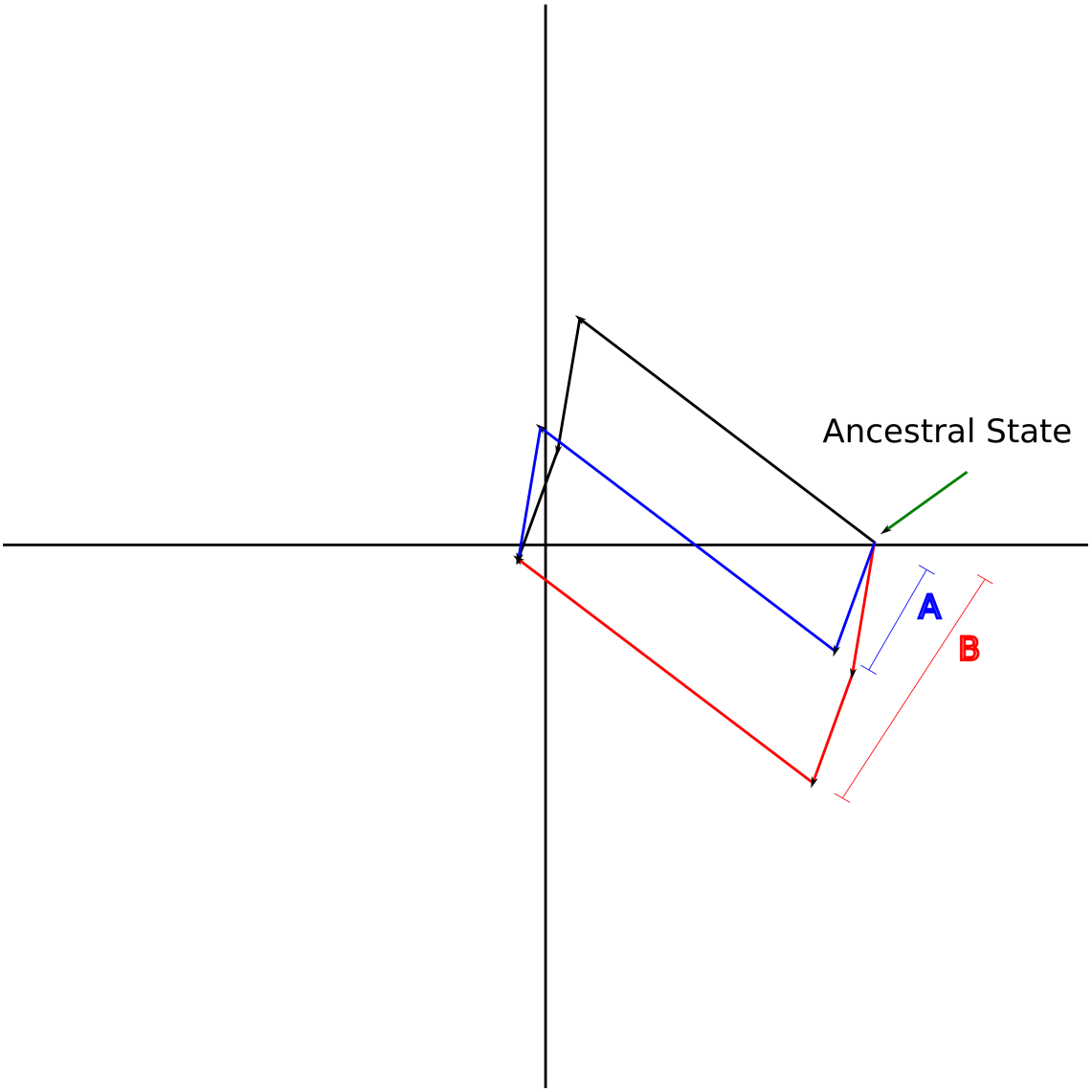


Figure S10. Cartoon of maximum phenotypic deviation calculation for a single adaptive trajectory. Suppose the black lines correspond to the true adaptive trajectory of the population, where three total mutations have occurred in succession (three black arrows). Consider two alternative orders of these mutations, in blue and red, both of which clearly deviate from the true adaptive trajectory in phenotype space. The distance **A** represents the maximal phenotypic deviation of the blue trajectory from the true adaptive trajectory, while the distance **B** represents the same thing for the red adaptive trajectory. This is essentially the largest distance from any point in the alternative mutation order to the true adaptive trajectory.

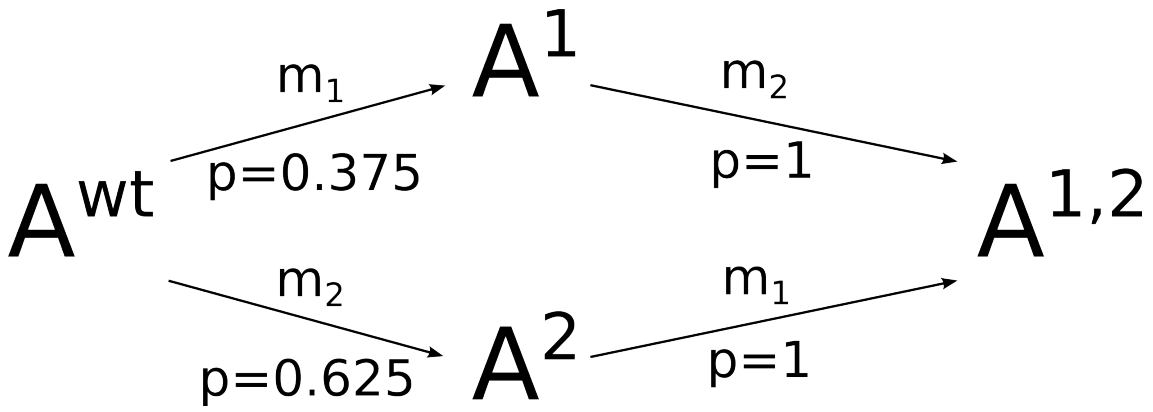


Figure S11. Along with table S1, representation of the FA method example in section S2.6. Arrows represent transitions after the introduction of an available mutation into the population, with the mutation above the arrow and the conditioned probability of the mutation successfully being generated and invading the population below the arrow.

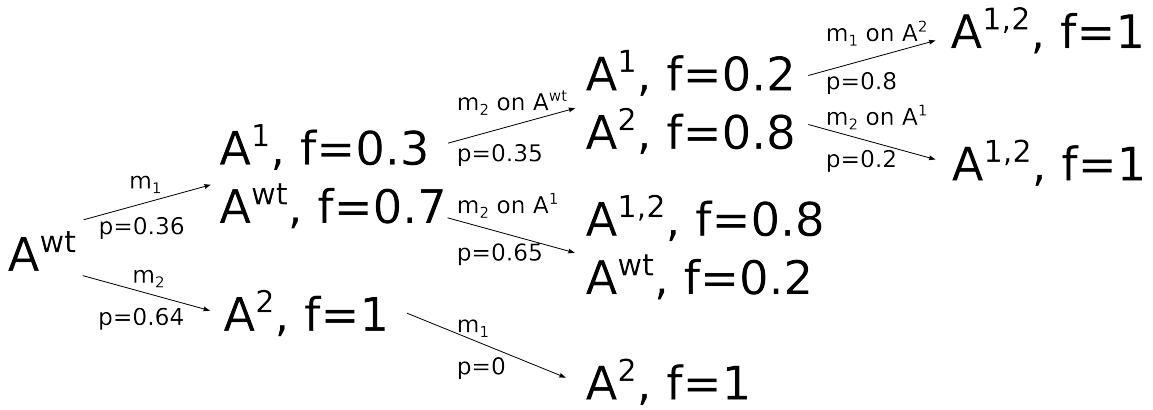


Figure S12. Along with table S2, representation of the PA method example in section S2.6. Note that this example uses a different set of mutations than the example for the FA method. Arrows represent transitions after the introduction of an available mutation into the population, with the mutation above the arrow and the conditioned probability of the mutation successfully being generated and invading the population below the arrow. Successful mutations that result in a balanced polymorphism are represented by the presence of multiple alleles each at some frequency (f).

Literature Cited

Fisher, R., 1930 *The genetical theory of natural selection*. Oxford at the Clarendon Press, Oxford, 1st edition.

Haldane, J. B. S., 1927 A Mathematical Theory of Natural and Artificial Selection, Part V: Selection and Mutation. Mathematical Proceedings of the Cambridge Philosophical Society **23**: 838–844.

Kimura, M., 1956 Rules for testing stability of a selective polymorphism. Proceedings of the National Academy of Sciences of . . . **1966**: 336–340.

Kimura, M., 1962 On the probability of fixation of mutant genes in a population. Genetics .

Kimura, M., and J. F. Crow, 1964 The number of alleles that can be maintained in a finite population. Genetics **49**: 725–738.

Kingman, J. F. C., 1961 A mathematical problem in population genetics. Mathematical Proceedings of the Cambridge Philosophical Society **57**: 574.

Kojima, K., 1959 Stable equilibria for the optimum model. Proceedings of the National Academy of Sciences of . . . **45**: 989–993.

Lande, R., 1976 Natural selection and random genetic drift in phenotypic evolution. Evolution1 **30**: 314–334.

Lobkovsky, A. E., Y. I. Wolf, and E. V. Koonin, 2011 Predictability of evolutionary trajectories in fitness landscapes. PLoS Computational Biology **7**: e1002302.

Lobkovsky, A. E., Y. I. Wolf, and E. V. Koonin, 2013 Quantifying the similarity of monotonic trajectories in rough and smooth fitness landscapes. Molecular bioSystems .

Mandel, S. P. H., 1959 The stability of a multiple allelic system. Heredity **13**: 289–302.

Palmer, M., A. Moudgil, and M. W. Feldman, 2013 Long-term evolution is surprisingly predictable in lattice proteins. Journal of the Royal Society, Interface **10**.

Roy, S. W., 2009 Probing evolutionary repeatability: neutral and double changes and the predictability of evolutionary adaptation. PLoS ONE **4**: e4500.

Sellis, D., B. Callahan, D. A. Petrov, and P. W. Messer, 2011 Heterozygote advantage as a natural consequence of adaptation in diploids. Proceedings of the National Academy of Sciences of the United States of America **2011**: 1–6.

Weinreich, D. M., N. F. Delaney, M. a. Depristo, and D. L. Hartl, 2006 Darwinian evolution can follow only very few mutational paths to fitter proteins. Science **312**: 111–4.

Wright, S., 1931 Evolution in mendelian populations. Genetics **16**: 97–159.