bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

1 **Title:** Predictability of adaptive evolution under the successive fixation assumption

2 Sandeep Venkataram, Diamantis Sellis, Dmitri A. Petrov

3 *Department of Biology, Stanford University, Stanford, CA 94305*

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

Running Title: Predictability of Evolution

Key Words: heterozygote advantage, fitness landscape, selection, population genetics

Corresponding Author:

Dmitri A. Petrov

Stanford University Department of Biology

371 Serra Mall Stanford CA 94305

(650) 736 1169

dpetrov@stanford.edu

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

ABSTRACT

Several recent experimental studies assessed the likelihood of all possible evolutionary paths between ancestral and evolved sequences. All of these studies measured the fitness of the intermediate genotypes and assumed that the advantageous genotypes fix in the population before acquiring the next adaptive mutation along the path. Unfortunately, the successive fixation assumption used by these studies is typically invalid, given that natural selection often maintain alleles at intermediate frequency by a variety of mechanisms such as frequency-dependent selection, local adaptation, clonal interference, and fitness overdominance. Here we simulate adaptive walks using Fisher's geometric model in diploid populations where previous work has shown that adaptation commonly generates balanced polymorphisms through overdominant mutations. We use these simulations to show that the use of the successive fixation assumption in this simple model is largely justified if the goal is to separate viable and inviable paths from each other. However, the estimates of the relative likelihoods of the viable paths become unreliable. We also show that the presence of balanced states along the true path significantly affects the number and likelihood distribution of viable paths when compared to walks without balanced states. These simple simulations highlight the importance of considering the effect of polymorphisms during adaptation especially given the prevalence of functional polymorphisms in natural populations.

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

INTRODUCTION

Predicting evolution is one of the fundamental challenges of evolutionary biology (reviewed in de Visser and Krug (2014)). This question became particularly prominent with Gould's famous thought-experiment on "replaying the tape of life" (Gould 1990), which asks how different our evolutionary history would have been if we could rerun evolution from some point in the past. We call this "forward predictability", or the predictability of the future evolutionary trajectory of a population from a given starting condition. Perhaps the best known study of forward predictability is a set of 12 parallel *Escherichia coli* lineages that have been experimentally evolving for over 50,000 generations (Wiser *et al.* 2013). These experiments showed that replicate independent evolution experiments frequently acquired similar adaptive mutations and experienced similar gains in fitness over the course of evolution (Crozat *et al.* 2010; Wiser *et al.* 2013), suggesting that evolution can be forward predictable to a surprising degree.

Recently, a number of experiments have instead focused on inferring past evolutionary history (backward predictability) given knowledge of both the ancestral state of the population and a set of adaptive mutations. The purpose of this method, which we call backward predictability inference, is to infer the relative likelihood of the different possible orders in which these specific adaptive mutations could have occurred. One of the earliest studies of backward predictability inference was done by Weinreich *et al.* (2006) based on a combinatorially complete reverse genetic study design pioneered by Malcolm *et al.* (1990). Weinreich *et al.* (2006) reconstructed all 32 possible combinations of 5 mutations in the beta-lactamase gene in *E. coli*, which are known to confer resistance to the drug rifampicin. They then quantitatively assayed the drug resistance of these 32 genotypes as a proxy for fitness, and used this information to analyze all $5! = 120$ possible trajectories (orders in which the 5 mutations could occur) from the ancestral to the resistant five-mutation genotype. A mutational path was deemed viable if resistance monotonically

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

increased with every mutational step. By this definition, Weinreich *et al.* (2006) found that only 18 of the 120 possible paths were viable and even fewer were highly likely, suggesting that there is substantial constraint in evolutionary trajectories that renders evolution highly predictable.

A handful of other studies using backward predictability inference have also suggested that evolution may be constrained to follow a few evolutionary paths. Franke *et al.* (2011) conducted backward predictability inference in all subsets of two to six mutations in an empirical eight-locus system of phenotypic marker mutations in *Aspergillus niger* and observed that different subsets of mutations had very different predictabilities. For example, they observed both zero and nine viable paths (out of 24 possible) in different four-mutation subsets. In addition, Buenrostro *et al.* (2014) studied an empirical RNA-protein binding landscape and found that only a few mutation orders were viable in various four-mutation subsets of the landscape, suggesting that this system is fairly backward predictable. All of these studies also computed the likelihood of each viable trajectory and found that only a few of these viable trajectories had a high likelihood.

The previous examples conducted backward predictability inference by surveying mutations of known function, or simply by sampling a large fraction of the possible sequence space through synthesis of random sequences. Another method is to use comparative genomics to infer the ancestral state of a protein (Yang *et al.* (1995); Hall (2006), reviewed in Thornton (2004); Harms and Thornton (2010)) and attempt to reconstruct the adaptive trajectory between the ancestral and derived states. Bridgham *et al.* (2006) conducted such a study using ancestral protein reconstruction of the aldosterone binding protein MR. After identifying a few functional mutations that differentiated the ancestral protein from the derived one using comparative genomic methods, the authors tested various intermediate genotypes for protein function. Along

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

with a follow up study (Ortlund *et al.* 2007), the authors found that some intermediate states generated nonfunctional proteins, which they suggest constrained evolution to one of the few adaptive trajectories with non-deleterious intermediate states.

In contrast to these results, Khan *et al.* (2011) performed an analysis of five adaptive mutations from experimentally evolved bacterial lineages using identical methodology and found that a majority of the mutational orders were viable. While backward predictability inference shows great promise in predicting evolution, we still need to understand how much predictability can vary across experimental systems. We also lack a sufficient number of gold-standard data sets to determine the utility of backward predictability inference in identifying either inviable mutation orders or the true adaptive trajectory.

Polymorphism in adaptive walks: A common assumption in all of these studies is that each putatively adaptive mutation fixes successively in a monomorphic population (Gillespie 1983, 1984; Orr 2002; Bridgham *et al.* 2006; Weinreich *et al.* 2006; Ortlund *et al.* 2007; Khan *et al.* 2011; Franke *et al.* 2011). This is clearly a major simplification for natural populations, since all natural populations are polymorphic. Beyond neutral polymorphisms that can be generated by genetic drift, genetic draft and admixture, we know that adaptive mutations can be maintained in a polymorphic state through the actions of natural selection. This can occur through negative frequency-dependent selection (Levin *et al.* 1988; Iserbyt *et al.* 2013), spatial and temporal fluctuations in selection (Rainey and Travisano 1998; Kasumovic *et al.* 2008; Bergland *et al.* 2014; Saltz and Nuzhdin 2014) and heterozygote advantage (also called overdominance, Takahata and Nei (1990)).

Backward predictability inference using the successive fixation assumption (FA method) is far easier to implement in practice than when polymorphisms are considered (PA method),

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

which is why the FA method has been used for all previous experimental studies. For the FA method, one only needs to assay the fitness of the 2^n possible genotypes in a system of n mutations (32 for n=5), and then infer the likelihood of every possible trajectory based on these fitness estimates. This could be further approximated by measuring a parameter related to fitness, such as the maximal growth rate of the organism. However, since polymorphisms are possible in the PA method, the marginal fitness of an allele is no longer independent of the other alleles present in the population. Therefore, the PA method requires one to directly test the successful invasion and stabilization of each new mutation in each of the possible adaptive trajectories. However, despite the challenges in implementation, the PA method may be necessary in systems where functional polymorphisms are frequently generated and maintained.

Beyond influencing the implementation of backward predictability inference, it is unknown whether the presence of polymorphic states systematically changes the backward predictability of evolution. We therefore want to study adaptive trajectories with and without polymorphic states to answer this question.

Predictability in Fisher's Geometric Model: Overall, there seems to be no experimental consensus on whether evolution is backward predictable using the method of Weinreich *et al.* (2006), and there is no experimental work on the impact of polymorphisms on backward predictability. Furthermore,the accuracy of backward predictability inference remains unknown. First, we do not know whether inferred trajectories that are deemed to be inviable are truly impossible. Second, it is unclear whether the likelihood estimate of an inferred trajectory, given that it is viable, accurately represents the probability that the population adapted along that trajectory. Finally, it is unknown how much improvement there will be in the inference procedure when we relax the successive-fixation assumption of prior studies and allow for stable polymorphic states

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

in systems where stable polymorphisms are frequently generated.

Due to the challenges of isolating sufficient numbers of independent adaptive trajectories from experimental populations, we utilize a simulation-based approach to study the accuracy and the impact of polymorphisms on backward predictability inference. A simulation framework also gives us perfect information about the true adaptive trajectory of the population so that we can conduct extensive validation of backward predictability inference. We employ Fisher's geometric model (FGM, Fisher (1930); Orr (1999, 2005)) to conduct forward simulations of adapting populations. FGM is a simple phenotypic model where individuals have phenotypes defined as points in n-dimensional space. A simple fitness function is then used to map phenotypes into fitness. Sellis *et al.* (2011) showed that adaptive mutations in diploid FGM simulations are frequently overdominant (exhibit heterozygote advantage) if the mutations are sufficiently large in phenotype space, resulting in balanced polymorphisms. These overdominant mutations are temporarily stable, but they can be driven out of the population by subsequent adaptive mutations. The simple nature of the model allows for a straightforward analysis procedure to study predictability. In addition, any evidence that the successive fixation assumption fails in this model would imply that it could easily fail in more realistic biological systems.

In this work, we simulate a large number of independent adaptive trajectories using FGM in three different parameter regimes, which we use to study the accuracy of the FA and PA methods. We then study the impact of stable polymorphisms on backward predictability in all three parameter regimes.

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

RESULTS

We generated adaptive trajectories using Wright-Fisher simulations with Fisher's geometric model in three parameter regimes, the first of which is a two dimensional regime with a poorly adapted initial population that is far from the optimum (2D-far regime, Supplementary Text ST1.1). We simulated diploid populations with a population size N=5000 and mutation rate $\mu = 5 * 10^{-6}$ per generation, with 10,000 replicate simulations (ST1.2). The low per-generation mutation rate allowed us to use the strong-selection weak-mutation assumption for our analysis to consider each mutation by itself. For all of our statistical analyses, we considered only those mutations that are present on the most frequent allele at generation 10,000. We additionally limited our analysis to the first five mutations of each adaptive walk, and ignored simulations with fewer than five mutations in order to compare adaptive walks of equal lengths. This is comparable to many recent studies utilizing backward predictability inference (backward predictability inference) (Weinreich *et al.* 2006; Khan *et al.* 2011; Franke *et al.* 2011). For simplicity, simulations that generated balanced states containing three or more alleles were removed from the analysis ($n = 1099$). We first used the remaining simulations to test the accuracy of the FA and PA methods. We then partitioned the simulations into those that do ($n = 4210$) and do not ($n = 872$) contain overdominant mutations to study the impact of balanced polymorphisms on the backward predictability of evolution (ST1.3). We tested the generality of our observations in 25-dimensional space with a poorly adapted initial population (25D-far regime) as well as 2-dimensional space with a well-adapted initial population (2D-close regime). Unless otherwise specified, all of our results described in the main text are for the 2D-far regime.

Impact of overdominant mutations on forward predictability: Sellis *et al.* (2011) found that overdominant mutations in FGM allow populations to explore a much larger portion of phenotype space than populations without such mutations. Therefore,

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

overdominant mutations should increase the phenotypic divergence between independent adaptive walks and thus decrease the forward predictability of evolution. To validate this with our simulations, we computed the maximum phenotypic distance of each simulated adaptive walk from the line segment connecting the ancestral and optimal phenotypes (i.e. the optimal adaptive trajectory) for simulations with and without overdominant mutations (Figure S1). We found that overdominant mutations significantly increase the deviation of an adaptive trajectory from the optimal trajectory, thus decreasing forward predictability and supporting the work of Sellis *et al.* (2011).

Conducting backward predictability inference: We conducted backward predictability inference in a manner similar to Weinreich *et al.* (2006) (implementation details in ST2.1) both with and without the successive fixation assumption (the FA and PA methods, respectively). Our analysis conditions on the first five mutations that occurred during a given FGM simulation and computes the relative likelihood of each of the $5! = 120$ possible orders of generating the adapted allele containing all five mutations (e.g. see Weinreich *et al.* (2006) Figure 2). This was done with a recursive procedure that successively added the five mutations into the ancestral population in every possible order until the allele containing all five mutations was generated. We computed the unconditioned likelihood of a mutation order as the product of the probabilities of each mutation occurring on the appropriate genetic background and then successfully invading the population, and then normalized the unconditioned likelihood across all viable mutation orders to compute the relative likelihood. Please see ST2.6 for a detailed example of this algorithm.

Validating the backward predictability inference method: We validated the FA and PA methods by using our knowledge of the order in which these five mutations actually occurred in the FGM simulation (from now on, "the true adaptive trajectory"),

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

similar to the validation methodology of (Ogden and Rosenberg 2006; Li *et al.* 2008; Messer and Petrov 2013; Bertels *et al.* 2014) and others. We eliminated, on average, 46% of the possible trajectories as they were inferred to be inviable (inferred trajectory probability = 0) using the FA method and 44% using the PA method. 97% of inferred inviable trajectories were inviable in both the FA and PA methods, while the rest were inferred to be inviable with only one of the methods. In $\sim 0.7\%$ of simulations with both the FA and PA methods, the true adaptive trajectory was eliminated by the inference analysis as inviable. Therefore, just by eliminating inviable trajectories, both the FA and PA methods greatly improve our ability to determine the true adaptive trajectory with a very low false negative rate.

Our implementations of the FA and PA methods use a diploid model, where new mutations must successfully invade the population as heterozygotes. We consider a further simplification of the FA model to exactly match the work of Weinreich *et al.* (2006), where we force new mutations to invade as homozygotes. We find that this results in a much higher (15%) false negative rate where true adaptive trajectories are mistakenly classified as inviable, so we do not consider this method for future analysis.

We assessed the accuracy of the FA and PA methods in two different ways. First, we compared the inferred probabilities for the trajectories predicted to be viable (ST2.2) to empirical estimates of their probability. For every set of five mutations, we conducted 1000 further forward Wright-Fisher FGM simulations while restricting the available set of mutations to those five mutations. We estimated the true likelihood of every possible mutation order as the frequency with which that mutation order occurred in the 1000 additional simulations. We found that the likelihoods inferred by both the FA and PA methods were significantly correlated with the true likelihoods (Figure 1), but the PA method was significantly better correlated with the true likelihoods than the FA method

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

(ANOVA $p < 10^{-10}$). In addition, we observe that the FA method both overestimates the likelihood of some highly unlikely mutation orders and underestimates the likelihood of some mutation orders with higher empirical likelihood. The PA method, in contrast, only appears to have some problems overestimating highly unlikely mutation orders. Therefore, the PA method matches the empirical likelihood estimates better than the FA method.

Our second assessment of accuracy was to test how informative these probabilities are in identifying the true adaptive trajectory. For both the FA and PA methods, we sorted all viable trajectories across all FGM simulations by their inferred probability and binned them into 100 bins. If the FA and PA methods work perfectly, we could expect a 1:1 relationship between the average trajectory probability for each bin and the fraction of the trajectories in that bin that match the mutation order of the true adaptive trajectory. If the inference method is not useful at all, we expect that there is no relationship between the two. We found that the trajectory probabilities for each bin were positively correlated with the fraction of the trajectories in that bin that were in fact observed in the original FGM simulations with both the FA and PA methods. The relationship was again weaker with the FA method than the PA method (FA slope = 0.49, PA: slope = 0.57, Figure 2).

We now wanted to know whether the inferred trajectory probabilities were better able to identify the true adaptive trajectory than random chance. Note that a positive slope is expected even if the FA and PA methods do not provide any information beyond removal of inviable lineages. Indeed, simulations where there are only a few viable trajectories will have, on average, a higher probability for each viable trajectory than those with many viable trajectories, and each of these viable trajectories are also more likely to be chosen as the true trajectory in a randomization since there are fewer of them. We tested whether our observed slope is greater than expected through 1000 randomizations of the data, where for each randomization, one viable trajectory in each simulation was randomly

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

assigned to be the true adaptive trajectory and the slope was recomputed as before. We found that both the FA and PA methods perform better than randomized data (FA and PA empirical $p < 0.001$; Figure 2, dashed lines compared to shaded areas).

Simulations in high dimensionality: In our implementation of Fisher's Model, balanced states arise when mutations are overdominant. The presence of additional phenotypic dimensions, which seems realistically plausible from observed rates of pleiotropy (Dudley *et al.* 2005; Albert *et al.* 2008), increases the fraction of new mutations that generate overdominant alleles (Sellis *et al.* 2011). However, this concordantly decreases the fitness advantage of the average new beneficial mutation, decreasing the number of adaptive mutations that successfully invade the population over our 10,000 generation FGM simulations. To study the impact of high dimensional landscapes on backward predictability, we conducted simulations using 25 dimensions with a mean mutation size of 5 (25D-far regime). The increase in mean mutation size, relative to our original two dimensional simulations, was necessary to generate a sufficient number of walks containing at least five mutations within 10,000 generations. We again partitioned the simulations into those with ($n = 288$) and without ($n = 203$) overdominant mutations and analyzed the backward predictability of the trajectories as before. We found that the FA and PA methods remove 17% and 16% of the inferred trajectories as inviable, respectively. Both methods also have very low false negative rates in inferring true adaptive trajectories as inviable (FA method 0.6%, PA method 1.4%). Consistent with previous results, the PA method performs significantly better than the FA method in accurately inferring the likelihood of a trajectory (Figure S2). Unlike our previous results, the slope computed with the FA method when comparing the likelihood of a set of binned trajectories to the fraction of those trajectories that are true (Figure S3) is not significantly better than randomizations (empirical $p = 0.733$), suggesting that the FA method is not useful in this high dimensional space beyond removing inviable trajectories. The PA method, in contrast,

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

is still significantly better than random chance in this regime (empirical $p = 0.009$).

Simulations close to the optimum: The Gaussian fitness surface was originally used to provide an approximation for populations close to their fitness optimum (Lande 1976), while our initial two parameter regimes had the ancestral population far from the optimum. We thus ran additional simulations in two diemnsions where the population was initialized close to the optimum by modifying the fitness function (2D-close regime, ST1.1). We generated a large number of 5-step adaptive walks ($n = 649$ and 234, respectively for simulations with and without overdominant mutations) and tested the accuracy and utility of the FA and PA methods as before (Figures S6-S7). Both methods remove a substantial fraction of the inferred trajectories as inviable (FA $\sim 44\%$, PA 42%), with a low false negative rate (FA 1.2%, PA 1.0%). The PA method is better correlated to the empirical probability estimates than the FA method (Figure S6, ANOVA $p < 10^{-10}$. However, neither method does better than random chance at identifying the true adaptive trajectory (Figure S7).

Across all parameter regimes, both the PA and FA methods are able to accurately identify inviable trajectories with a low false positive rate. However, the PA method has significantly higher accuracy in inferring the probability of a trajectory than the FA method in all regimes when compared to empirical estimates of trajectory probabilities (Figures 1, S2, S6). In addition, the PA method does better than the FA method at identifying the true trajectory in all regimes through the binning method. Unlike the PA method, the FA method fails to do better than random chance at identifying the true adaptive trajectory in the 25D-far regime, while both fail to do better than random chance in the 2D-close regime. The FA method is thus only reliable for identifying inviable trajectories but not for inferring their likelihoods. Therefore, we use the PA method for the remainder of our analysis.

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

Impact of overdominant mutations on backward predictability inference: Since we have shown that the PA method can accurately infer the probability of an adaptive trajectory, we utilized this method to study the impact of overdominant mutations on the backward predictability of adaptive trajectories. In particular, we are interested in testing whether backward predictability as inferred by the PA method is consistent with the previous results that forward simulations with overdominant mutations are less forward predictable than simulations without overdominant mutations (Sellis *et al.* 2011). For all of these analyses, we identified simulations that do and do not contain overdominant mutations and analyzed them separately (ST1.3). As before, all of our results are for the 2D-far regime unless stated otherwise.

We first study backward predictability in terms of the likelihoods of the different mutation orders. We compute the effective number of paths for each set of five mutations by weighting the number of viable paths found by their likelihoods (ST2.3). We found that the presence of overdominant mutations increased backward predictability in a walk by 31%, on average, when compared to simulations that lack overdominant mutations (Figure 3, Kolmogorov-Smirnov $p < 10^{-10}$). A similar statistic (mean path divergence, Lobkovsky *et al.* (2011)) also found that overdominant mutations resulted in walks that were 11% more backward predictable (Kolmogorov-Smirnov $p < 10^{-10}$). We found similar results in our other two regimes as well (Figures S4, S8). In other words, compared to an adaptive trajectory without overdominant mutations, it is more probable that independent evolutions experiencing at least one overdominant mutation will use the same mutational order. These results are in contrast with the previous result that overdominant mutations make adaptation less forward predictable (Figure S1, Sellis *et al.* (2011)).

While the effective number of paths and mean path divergence statistics capture the backward predictability of the mutational order, we can also consider the similarity of the

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

inferred trajectories with the wrong mutation order to the true trajectory in phenotype space (ST2.4). Simulations with overdominant mutations are 7% less phenotypically similar than simulations without such mutations in the 25D-far regime (Kolmogorov-Smirnov p = 0.001) but there is no significant difference between the two types of walks observed in the 2D-far or 2D-close regimes. However, in all parameter regimes, simulations with overdominant mutations are less phenotypically similar to the true adaptive trajectory than simulations without overdominant mutations. This is in contrast to the results looking at the backward predictability of mutation order where simulations with overdominant mutations were more predictable in all regimes.

We then wanted to test whether there was any relationship between the likelihood of a trajectory and its phenotypic deviation from the true trajectory. We compared the phenotypic deviation of each inferred trajectory against the probability of the inferred trajectory, and found a significantly negative regression slope (Figure 4a, slope $= -0.50$, $p < 10^{-10}$). This suggests that trajectories with high inferred probability but different mutation orders have similar phenotypic paths to the optimal trajectory. We compared this slope to slopes from 100 randomizations, where we randomly selected a viable mutation order to be the true trajectory. For each randomization, we recomputed the phenotypic deviations and found that the observed slope was more extreme than the slope of every randomized trial (Figure 4b, empirical $p < 0.01$). The significant of the slope suggests that even if backward predictability inference generates incorrect mutation orders with high probability, the inferred trajectory may still be a good phenotypic approximation for the true evolutionary history of the population. This significant reduction in slope compared to randomized data also holds for the 25D-far and 2D-close regimes (Figures S5, S9).

Multiple adapted states: We now turn to the first of two qualitatively novel features we uncovered while using the PA method. Backward predictability inference using the PA

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

method generates different intermediate alleles by introducing mutations in different orders. It is thus possible that the adapted allele containing all of the available mutations may be stably balanced against some of these intermediate alleles, but not others. Therefore, by introducing mutations in different orders, it may be the case that the final adapted allele is maintained as a stable polymorphism with different intermediate alleles in different mutational orders, or not maintained in a balanced state at all (Figure 5a).

We found that backward predictability inference frequently generated at least two different population states containing the final adapted allele when the mutations are introduced in different orders, with a maximum of 14 different population states for a single set of five mutations. In addition, the presence of an overdominant mutation in the observed walk increased the frequency of multiple adapted states from 40% to 55%. We also observed that 88% of the walks that did not experience balanced states in the FGM simulations generated at least one balanced state in some viable mutation order while using the PA method. Multiple end states were also observed in 67% and 57% of all simulations in the 25D-far and 2D-close regimes, respectively, suggesting that this is a general feature of our model.

Hidden alleles: In the course of our validation analysis, we were struck by the presence of simulations where the true adaptive trajectory was inferred to be inviable. Although 9 of these 35 simulations appear to be impossible to reconstruct due to clonal interference, we found that the remaining 26 contain derived alleles that reached high frequency through a balanced state but were eventually lost, which we term "hidden" alleles (Figure 5b). We hypothesized that in these 26 simulations, hidden alleles were necessary for the true trajectory to be viable, such that attempts to reconstruct the order of mutations using the PA method without including the hidden alleles would fail to recover the true adaptive trajectory. To test this, we selected one of these 26 simulations at random to rerun the PA method while including all hidden alleles in the inference. We successfully recovered the

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

true adaptive trajectory that had been previously impossible to reconstruct (data not shown), suggesting that hidden alleles can lead to significant errors in inference.

In general we found that 44% of all of our simulations contain at least one hidden allele (ST1.5). The vast majority of them (99%) are in simulations with overdominant mutations, as expected, since hidden alleles should only be able to reach high frequency when there is a balanced state. Hidden alleles in simulations without overdominant mutations appear to be the result of clonal interference. Given the high frequency of hidden alleles in our simulations, we tested whether they significantly impact the accuracy of the PA method, as in Figure 2. The slope for simulations without hidden alleles is 0.69, while the slope for simulations that do have hidden alleles is 0.59. While removing hidden alleles does increase the slope by 17%, this difference is non-significant (t-test $p = 0.29$), suggesting that the presence of hidden alleles did not significantly influence the accuracy of the likelihood estimates for the inferred trajectories, except for the rare cases where the hidden alleles were necessary for the viability of the true adaptive trajectory. Hidden alleles were also observed in 33% of simulations in the 25D-far regime and 27% of simulations in the 2D-close regime, suggesting that this is a general feature of our system.

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

DISCUSSION

In this work, we sought to answer two major questions. First, we wanted to know how polymorphisms impact backward predictability inference, and more generally, whether it is justified to assess the backward predictability of a trajectory by assuming successive fixation of mutations. Secondly, we wanted to investigate how predictability changes when comparing simulations with and without balanced polymorphisms. We use our results to interpret the many experimental studies using backward predictability inference that have been conducted, particularly the large variation in the number of viable trajectories.

Lower accuracy of the FA method compared to the PA method: When considering only whether or not a trajectory is viable, backward predictability inference both with (FA) and without (PA) the successive fixation assumption appear to do equally well at eliminating inviable trajectories with a low false-negative rate. However, the PA method is significantly more accurate than the FA method in all regimes when compared to empirically estimated probabilities from additional forward FGM simulations (Figures 1, S2, S6). The likelihoods estimated by the PA method appear to be more accurate than those from the FA method at identifying the true adaptive trajectory in all regimes as well (Figures 2, S3, S7, red dashed line compared to blue dashed line).

The lower accuracy of the FA method in estimating the likelihood of a trajectory when compared to the PA method results from two main consequences of using a diploid model. First, forcing an overdominant allele to fix when using the FA method reduces mean fitness, because overdominant alleles maximize mean fitness as stable polymorphisms. In some cases, mean fitness can actually decrease relative to the ancestral population. Therefore, certain mutational steps that confer marginal fitness benefits in the PA method can confer large fitness benefits in the FA method, due to the artificially reduced mean fitness of the population from the fixation of a prior overdominant mutation. Second, when

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

the PA method generates a balanced state, the likelihood of the next mutation in a particular mutation order is proportional to the frequency of the appropriate genetic background in that balanced state. This is irrelevant for the FA method, since every mutation is assumed to reach fixation. Therefore, mutational events that should be unlikely are not accounted for in the FA method.

One possible way to use the FA method and avoid the reduction in mean fitness is to use a haploid-like model where we only consider the fitness of homozygotes. However, we find that this version of the FA method results in a 15% false negative rate in the 2D-far regime. In addition, we still observe the systematic overestimation of mutation orders with low likelihoods found in the regular FA method (data not shown), suggesting that this is not an appropriate method for backward predictability inference.

Differences between parameter regimes: The usefulness of the inferred probabilities in detecting the true adaptive trajectory seems to be a function of the fitness of the ancestral population. The PA method does better than random chance at identifying the true adaptive trajectory in the 2-dimensional regime with a poorly adapted initial population that is phenotypically far from the optimum (2D-far regime) and the 25 dimensional regime far from the optimum (25D-far regime), but not in the well-adapted 2D regime close to the optimum (2D-close regime). We suspect that when the population is initialized close to the optimum and thus already has high fitness, every adaptive mutation can only slightly improve fitness, so most viable mutation orders for an adaptive trajectory in this regime are equally likely, making it challenging to identify the true adaptive trajectory by its probability alone.

In addition to the poor accuracy of the FA method in the 2D-close regime, the FA method also does poorly in the 25D-far regime, suggesting that a high phenotypic dimensionality

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

limits the accuracy of backward predictability inference specifically under the fixation assumption. Prior work has shown that increasing phenotypic dimensionality in FGM increases the fraction of adaptive mutations that are overdominant (Sellis *et al.* 2011), which are not accounted for with the FA method. This is likely what causes the poor accuracy of the FA method in the 25D-far regime.

We now test whether the poor accuracy of both the PA and FA methods in the 2D-close regime is due to the similar likelihoods of all viable trajectories. We compute the percentile rank of the true adaptive trajectory among all viable inferred trajectories for each simulation, and then compute the median across all simulations. A percentile rank of 0 implies that the true adaptive trajectory has, on average, the lowest probability of all viable trajectories in that simulation, while a percentile rank of 100 means that it has the highest probability. A rank of 50 implies that the inferred probabilities are not useful in identifying the true adaptive trajectory. We find that the median percentile ranks for the PA method are 70 and 66 for the 2D-far and 25D-far regimes, respectively, but is only 60 for the 2D-close regime, suggesting that the probabilities are less informative in identifying the true adaptive trajectory in this regime than the other regimes due to the true adaptive trajectory having a similar likelihood to many other viable trajectories.

We observe additional differences between our regimes when studying backward predictability in phenotype space. We only observe a significant difference in phenotypic similarity between simulations with and without overdominant mutations in the 25D-far regime, suggesting that the high dimensional regime is enhancing the phenotypic differences between these types of adaptive walks. In addition, both the 2D-far and 25D-far regimes have significantly negative slopes when comparing the phenotypic similarity of an inferred trajectory to its likelihood as computed by the PA method (Figure 4, S5). However, the 2D-close regime has a slope close to zero (Figure S9), again suggesting that

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

backward predictability inference on well-adapted populations is less informative in identifying the true adaptive trajectory than in poorly-adapted populations.

Interpretation of experimental studies: We can now compare the results of the PA method in our different parameter regimes to the previous results from natural systems. Weinreich *et al.* (2006) found that only 15% of the possible trajectories were viable, while others found 0%, 38% and $> 50\%$ in various systems. In addition, Franke *et al.* (2011) found significant variability in the number of viable trajectories in different non-independent sets of mutations in the same system. In our simulations, we find that the average number of viable trajectories depends on the dimensionality of the model, with 56% of the trajectories viable in the 2D-far regime and 58% in the 2D-close regime, while 84% of the trajectories are viable in the 25D-far regime. The number of viable trajectories is also highly variable within a single parameter regime, as the standard deviation of the number of viable trajectories in the 2D-far regime is 40.3. Therefore, the variation in the number of viable trajectories observed between experimental systems could be caused either by differences in the adaptive landscape or the inherent variability between independent adaptive walks. Backward predictability inference needs to be conducted in a large number of independent adaptive trajectories in multiple systems to resolve these sources of variation.

We found that the presence of overdominant mutations also influences the results of backward predictability inference. Overdominant mutations significantly increase backward predictability in terms of the order of mutations, but consistently decrease backward predictability in terms of phenotype space. Our analysis also corroborates previous results which suggest that overdominant mutations decrease forward predictability in phenotype space (Sellis *et al.* (2011), Figure S1). These results suggest that predictability can not be modeled as a scalar value, as different types of analysis give different results even in a

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

simple system like FGM.

Our analysis also uncovered qualitatively novel behaviors, such as the presence of hidden alleles. In particular, we found a small fraction of simulations where lack of knowledge of hidden alleles made it impossible to accurately infer the true trajectory using backward predictability inference. Therefore, in natural systems, there may be an unknown subset of adaptive trajectories where extinct hidden alleles would make it impossible to infer the true trajectory. Functionally important hidden alleles are challenging to identify even in extant populations, due to the vast amounts of variation present in any natural population.

As we are using a simulation system to carry out this study, there are a number of simplifications that we have made that will likely affect the accuracy of backward predictability inference in more realistic systems. These include alternative mechanisms by which adaptive mutations can be maintained as stable polymorphisms, including negative frequency dependent selection (Levin *et al.* 1988; Iserbyt *et al.* 2013), and spatially or temporally variable selection (Rainey and Travisano 1998; Kasumovic *et al.* 2008; Saltz and Nuzhdin 2014). Natural populations can also generate polymorphisms through a number of other mechanisms, such as clonal interference (Desai and Fisher 2007; Herron and Doebeli 2013; Kvitek and Sherlock 2013; Lang *et al.* 2013), genetic drift, admixture and other demographic processes. While phenomena such as hidden mutations are likely universal to all of these processes, it is unclear whether the predictability of systems with stable polymorphic states depends on the mechanism that generates the polymorphisms.

The underlying genetic, phenotypic and fitness landscape models used in our simulations are also limited in a number of ways, and could be expanded by including the possibility of multiple adaptive optima within the model, multiple genetically unlinked loci that are

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

capable of adaptation, epistasis between multiple loci, the presence of standing genetic variation and genetic draft. Consideration of these processes will likely further complicate backward predictability inference. Finally, simulation systems have the advantage of having exact knowledge of the fitness landscape, so we can precisely measure the fitness of every genotype. This information must be estimated in natural systems, which may introduce significant noise during backward predictability inference.

Despite our use of a very simple model, we have shown that the successive fixation assumption has limited utility when attempting to predict evolution. Our simulations suggest that the FA method is only useful for identifying inviable mutation orders. In contrast, the PA method can both identify inviable mutation orders and accurately infer the likelihood of a mutation order. Since natural systems are substantially more complex than our model, we suggest that the successive fixation assumption should be used with great caution. Additional experiments are required to test whether the PA method is accurate in inferring the likelihood of a mutation order in biological systems, as well as additional theoretical studies to understand how alternative sources of polymorphisms and our other model assumptions affect the accuracy of the PA method, and more generally, the inference of the predictability of evolution.

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

Literature Cited

Albert, A. Y. K., S. Sawaya, T. H. Vines, A. K. Knecht, C. T. Miller, *et al.*, 2008 The genetics of adaptive shape shift in stickleback: pleiotropy and effect size. Evolution **62**: 76–85.

Bergland, A., E. L. Behrman, K. R. O'Brien, P. S. Schmidt, and D. A. Petrov, 2014 Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in Drosophila. PLoS Genetics .

Bertels, F., O. K. Silander, M. Pachkov, P. B. Rainey, and E. van Nimwegen, 2014 Automated reconstruction of whole-genome phylogenies from short-sequence reads. Molecular biology and evolution **31**: 1077–88.

Bridgham, J. T., S. M. Carroll, and J. W. Thornton, 2006 Evolution of hormone-receptor complexity by molecular exploitation. Science (New York, N.Y.) **312**: 97–101.

Buenrostro, J. D., C. L. Araya, L. M. Chircus, C. J. Layton, H. Y. Chang, *et al.*, 2014 Quantitative analysis of RNA-protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. Nature biotechnology **32**: 562–8.

Crozat, E., C. Winkworth, J. Gaffé, P. F. Hallin, M. a. Riley, *et al.*, 2010 Parallel genetic and phenotypic evolution of DNA superhelicity in experimental populations of escherichia coli. Molecular Biology and Evolution **27**: 2113–2128.

de Visser, J. A. G. M., and J. Krug, 2014 Empirical fitness landscapes and the predictability of evolution. Nature reviews. Genetics **15**: 480–90.

Desai, M. M., and D. S. Fisher, 2007 Beneficial mutation selection balance and the effect of linkage on positive selection. Genetics **176**: 1759–98.

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

558 Dudley, A. M., D. M. Janse, A. Tanay, R. Shamir, and G. M. Church, 2005 A
559 global view of pleiotropy and phenotypically derived gene function in yeast. Molecular
560 systems biology **1**: 2005.0001.

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

563 Franke, J., A. Klozer, J. A. G. M. de Visser, and J. Krug, 2011 Evolutionary
564 accessibility of mutational pathways. PLoS Computational Biology **7**.

565 Gillespie, J., 1983 A simple stochastic gene substitution model. Theoretical Population
566 Biology **23**: 202–215.

567 Gillespie, J., 1984 Molecular evolution over the mutational landscape. Evolution **38**:
568 1116–1129.

569 Gould, S. J., 1990 *Wonderful Life: The Burgess Shale and the Nature of History*. W. W.
570 Norton & Company.

571 Hall, B. G., 2006 Simple and accurate estimation of ancestral protein sequences.
572 Proceedings of the National Academy of Sciences of the United States of America **103**:
573 5431–6.

574 Harms, M. J., and J. W. Thornton, 2010 Analyzing protein structure and function
575 using ancestral gene reconstruction. Current Opinion in Structural Biology **20**: 360–366.

576 Herron, M. D., and M. Doebeli, 2013 Parallel evolutionary dynamics of adaptive
577 diversification in Escherichia coli. PLoS Biology **11**: e1001490.

578 Iserbyt, A., J. Bots, H. Van Gossum, and T. N. Sherratt, 2013 Negative
579 frequency-dependent selection or alternative reproductive tactics: maintenance of female
580 polymorphism in natural populations. BMC evolutionary biology **13**: 139.

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

581 Kasumovic, M. M., M. J. Bruce, M. C. B. Andrade, and M. E. Herberstein,
582 2008 Spatial and temporal demographic variation drives within-season fluctuations in
583 sexual selection. Evolution **62**: 2316–25.

584 Khan, A. I., D. M. Dinh, D. Schneider, R. E. Lenski, and T. F. Cooper, 2011
585 Negative epistasis between beneficial mutations in an evolving bacterial population.
586 Science **332**: 1193–6.

587 Kvitek, D. J., and G. Sherlock, 2013 Whole genome, whole population sequencing
588 reveals that loss of signaling networks is the major adaptive strategy in a constant
589 environment. PLoS Genetics **9**: e1003972.

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

592 Lang, G. I., D. P. Rice, M. J. Hickman, E. Sodergren, G. M. Weinstock, *et al.*,
593 2013 Pervasive genetic hitchhiking and clonal interference in forty evolving yeast
594 populations. Nature .

595 Levin, B. R., J. Antonovics, and H. Sharma, 1988 Frequency-dependent selection in
596 bacterial populations. Philosophical Transactions of the Royal Society B: Biological
597 Sciences **319**: 459–472.

598 Li, H., J. Ruan, and R. Durbin, 2008 Mapping short DNA sequencing reads and calling
599 variants using mapping quality scores. Genome research **18**: 1851–8.

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

602 Malcolm, B., K. Wilson, and B. Matthews, 1990 Ancestral lysozymes reconstructed,
603 neutrality tested, and thermostability linked to hydrocarbon packing. Nature **345**:
604 86–89.

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

Messer, P. W., and D. a. Petrov, 2013 Frequent adaptation and the McDonald-Kreitman test. Proceedings of the National Academy of Sciences **2013**.

Ogden, T. H., and M. S. Rosenberg, 2006 Multiple sequence alignment accuracy and phylogenetic inference. Systematic biology **55**: 314–28.

Orr, H. A., 1999 The evolutionary genetics of adaptation: a simulation study. Genetical Research **74**: 207–14.

Orr, H. A., 2002 The population genetics of adaptation: the adaptation of DNA sequences. Evolution **56**: 1317–30.

Orr, H. A., 2005 The genetic theory of adaptation: a brief history. Nature reviews. Genetics **6**: 119–27.

Ortlund, E. A., J. T. Bridgham, M. R. Redinbo, and J. W. Thornton, 2007 Crystal structure of an ancient protein: evolution by conformational epistasis. Science (New York, N.Y.) **317**: 1544–8.

Rainey, P., and M. Travisano, 1998 Adaptive radiation in a heterogeneous environment. Nature **394**: 69–72.

Saltz, J. B., and S. V. Nuzhdin, 2014 Genetic variation in niche construction: implications for development and evolutionary genetics. Trends in ecology & evolution **29**: 8–14.

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.

Takahata, N., and M. Nei, 1990 Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. Genetics **124**: 967–78.

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

629 Thornton, J. W., 2004 Resurrecting ancient genes: experimental analysis of extinct
630 molecules. Nature reviews. Genetics **5**: 366–75.

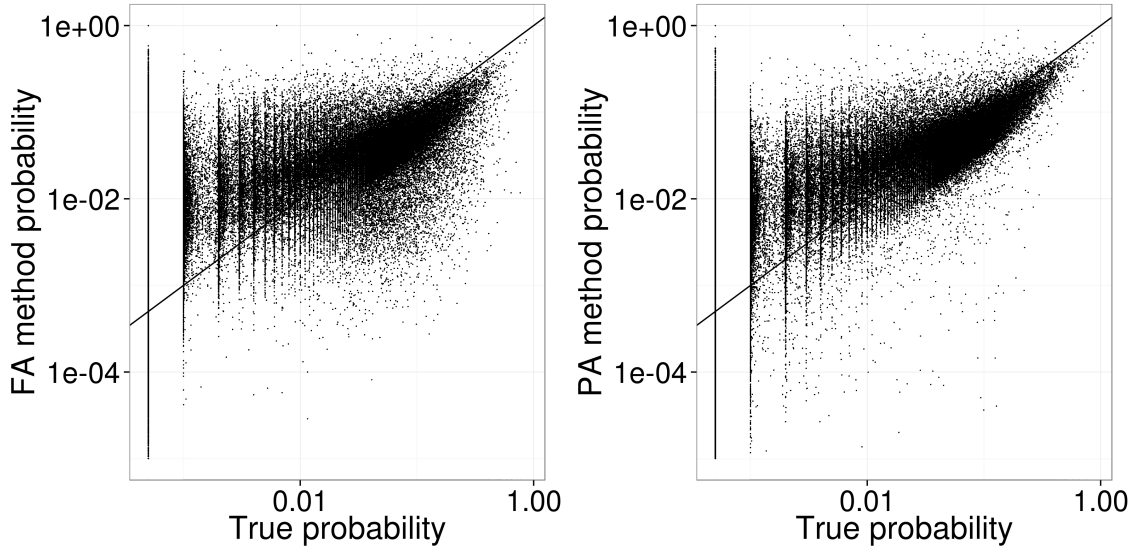
631 Weinreich, D. M., N. F. Delaney, M. a. Depristo, and D. L. Hartl, 2006
632 Darwinian evolution can follow only very few mutational paths to fitter proteins. Science
633 **312**: 111–4.

634 Wiser, M. J., N. Ribeck, and R. E. Lenski, 2013 Long-term dynamics of adaptation
635 in asexual populations. Science (New York, N.Y.) **1364**.

636 Yang, Z., S. Kumar, and M. Nei, 1995 A new method of inference of ancestral
637 nucleotide and amino acid sequences. Genetics **141**: 1641–1650.

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

638 # FIGURES



639

640 **Figure 1. Accuracy of inferred probabilities by the FA and PA methods.** We
641 used 1000 replicate forward simulations to estimate the true probability of every viable
642 path across all simulations to compare to the probabilities inferred by the **(A)** FA and **(B)**
643 PA methods. The straight lines show y=x. The probabilities by both methods are
644 significantly correlated with the true probability ($r^2 = 0.53$, $p < 10^{-10}$; $r^2 = 0.67$, $p < 10^{-10}$
645 for FA and PA methods, respectively). The PA method is significantly better correlated
646 than the FA method ($p < 10^{-10}$).

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

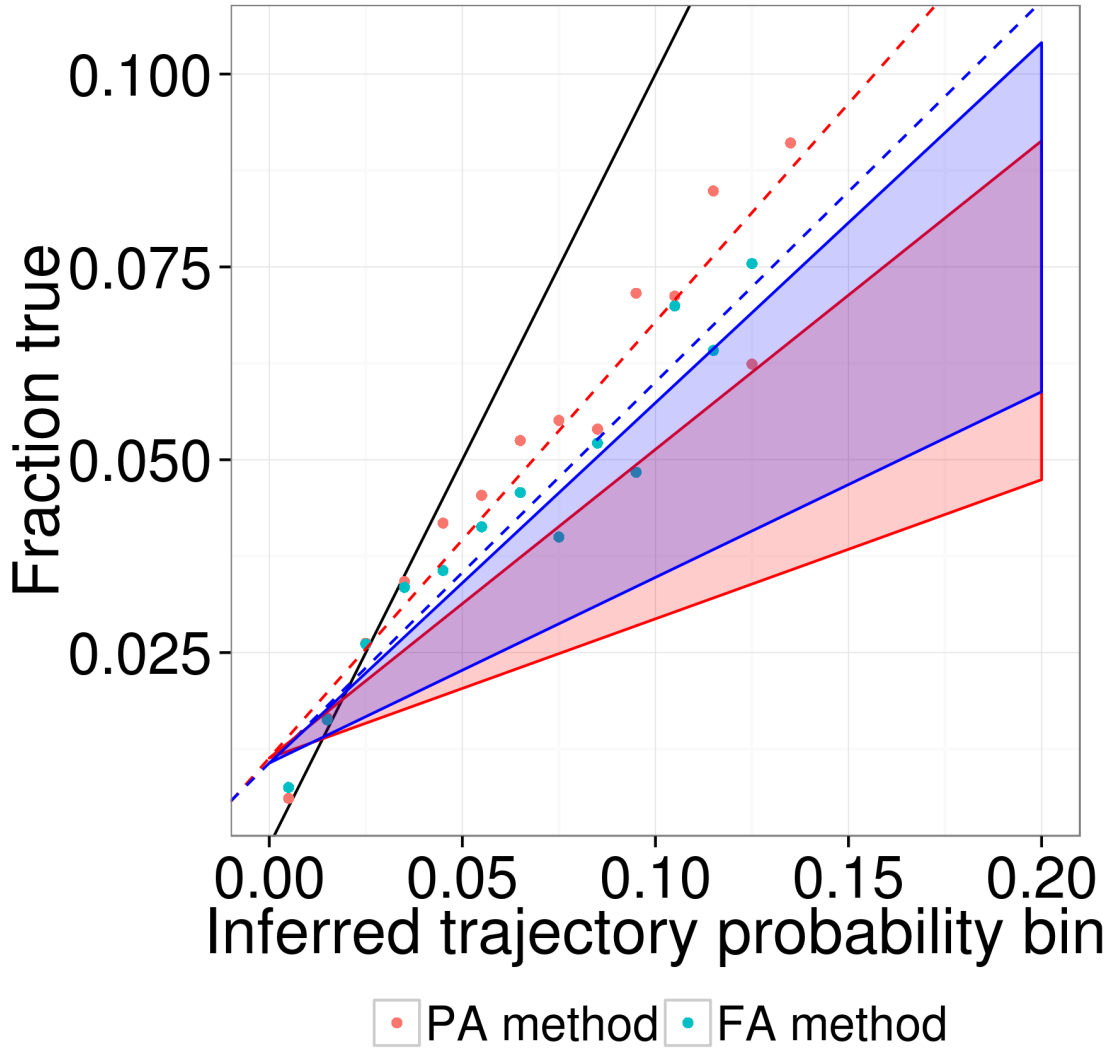


Figure 2. Utility of the inferred trajectory probabilities of the FA (blue) and PA (red) methods in determining the true adaptive trajectory. Dots are the probability of the inferred trajectories from the backward predictability inference binned into bins of width 1% against the fraction of the inferred trajectories in each bin that actually occurred in the underlying FGM simulation. Only bins containing at least 500 trajectories are shown, with the regression lines shown as dashed lines of the appropriate color, and the y=x line shown in solid black. Shaded areas of the appropriate colors show the distribution of regression lines when a random viable trajectory is selected as the true trajectory. Both the FA and PA methods do better than all randomizations.

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

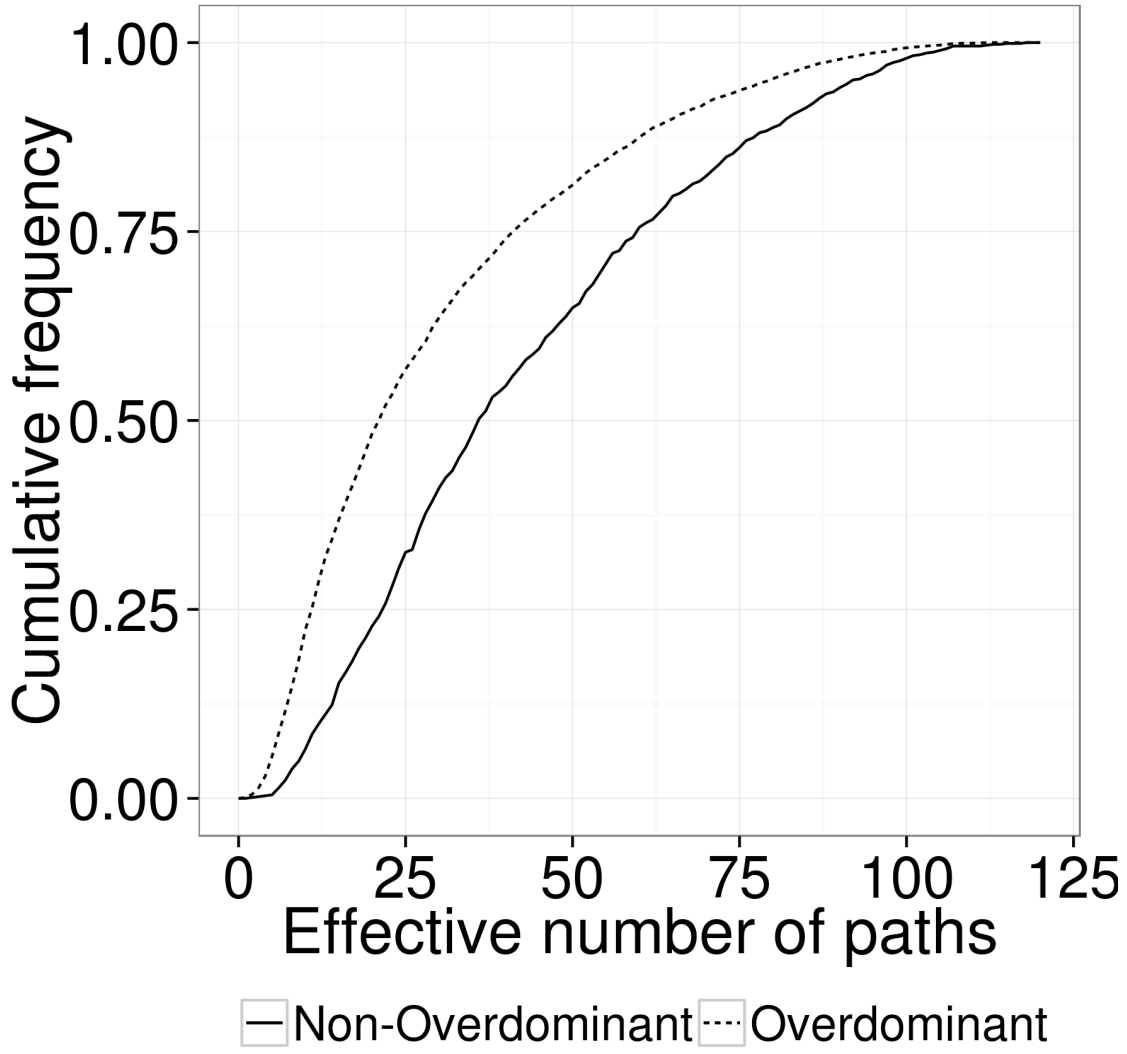


Figure 3. Cumulative distribution of the effective number of paths for adaptive walks with five mutations. This is a metric of backward predictability of evolution, which is computed as the inverse of the sum of the squared probabilities of the viable trajectories for that set of five mutations. A value of $5! = 120$ means that every possible mutation order is viable and equally likely, while low values indicate the presence of only a few mutation orders with high likelihoods. The effective number of paths in simulations without overdominant mutations is significantly greater than in simulations with such mutations, suggesting that overdominant mutations increase backward predictability (Kolmogorov-Smirnov $p < 10^{10}$).

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

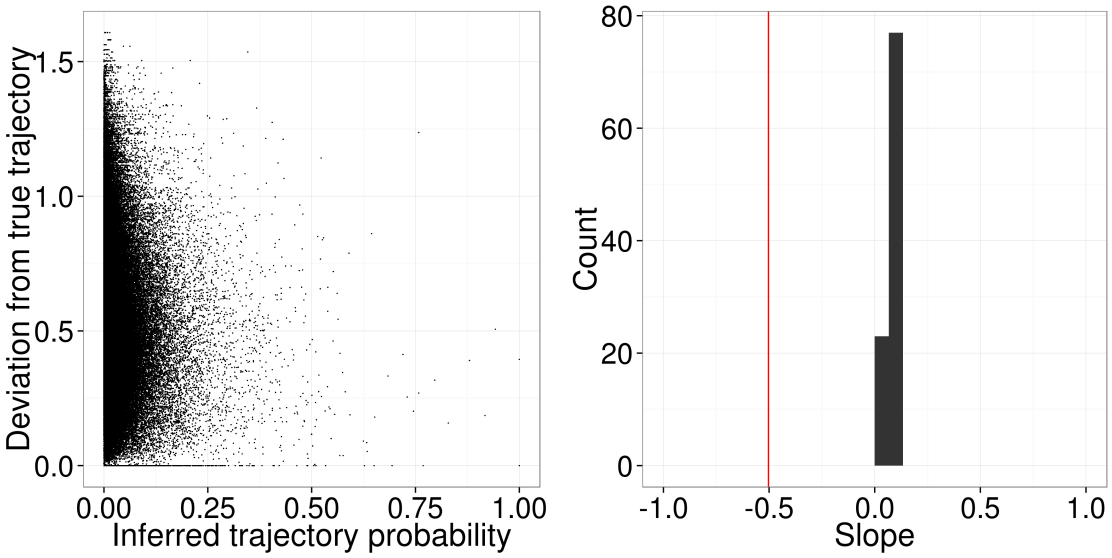
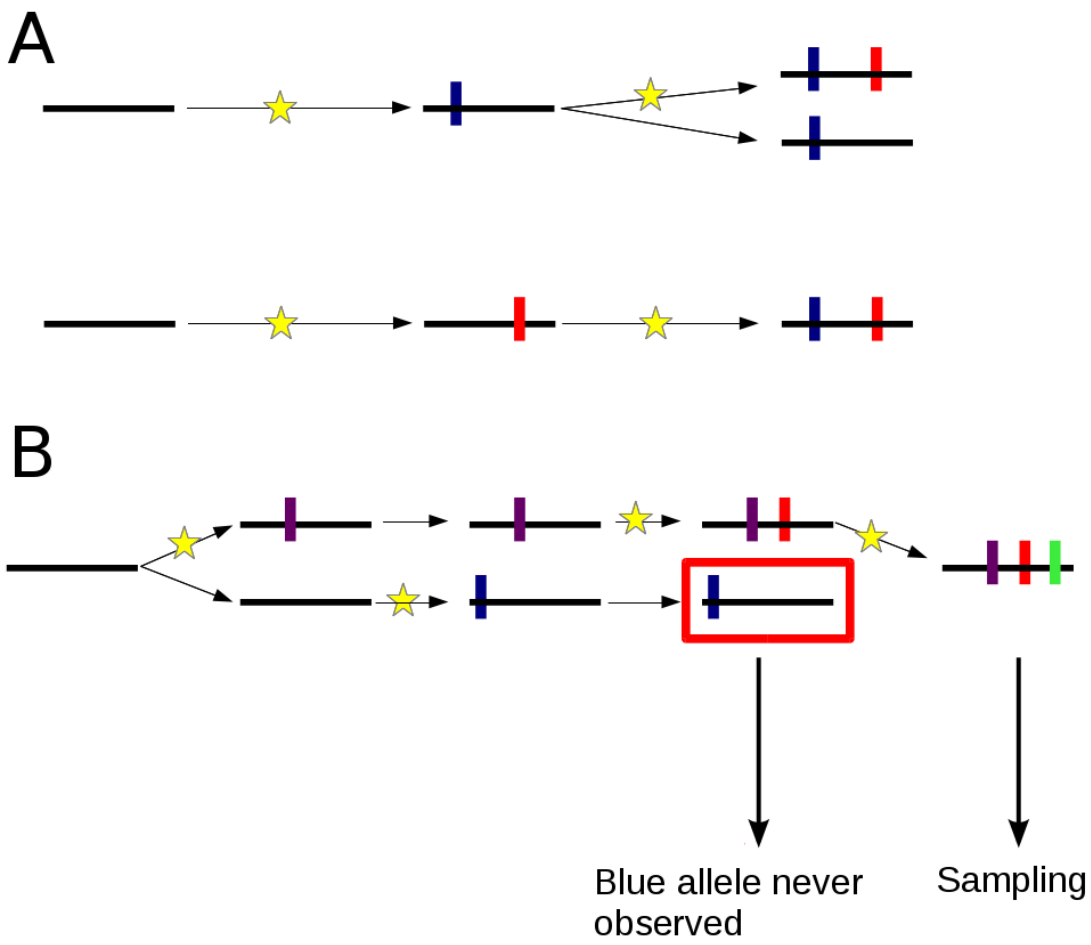


Figure 4. **(A)** Average phenotypic distance of inferred trajectories compared to the true trajectory. Linear regression slope $= -0.50$, t test of difference of slope from 0 $p < 10^{-10}$. **(B)** Comparison of empirical slope (vertical red line) to the slope distribution when randomly selecting the true trajectory and recomputing the deviation from this new "true" trajectory over 100 randomizations (empirical $p < 0.01$).

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.



672

673 **Figure 5. Potential impact of polymorphism on adaptive trajectories.** Black
674 horizontal lines represent alleles, while colored vertical bars represent mutated loci. Yellow
675 stars show the occurrence of mutations. Time increases from left to right, and each set of
676 arrows represents a transition from one population state to another through the process of
677 a mutation successfully invading the population and reaching equilibrium. **(A)** Multiple
678 Adapted States. A pair of mutations (red and blue) are introduced onto an ancestral
679 genotype in different orders for backward predictability inference. In the first scenario
680 (top), the blue mutation occurs first, with the second red mutation creating a balanced
681 polymorphic state with the allele containing only the blue mutation. In the second scenario
682 (bottom), the red mutation occurs first, and when the second blue mutation occurs, it fixes
683 in the population. Therefore, the same mutations can result in different adapted
684 population states when introduced in different orders. **(B)** Hidden Alleles. A four

bioRxiv preprint first posted online Dec. 2, 2013; doi: http://dx.doi.org/10.1101/001016. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

mutation system is depicted, where the blue mutation creates a derived allele that occurred on a polymorphic state that was stably maintained for some time but subsequently lost. As sampling occurs after the loss of this allele, we call the allele with the blue mutation a hidden allele, as it is hidden from sampling. In some cases, the blue allele may have been necessary for the purple, red and green mutations to occur in the order that they did, resulting in a true adaptive trajectory that is impossible to reconstruct without knowledge of the hidden allele.