Title: Ploidy and the Predictability of Evolution in Fisher's Geometric Model
Running Title: Ploidy and the Predictability of Evolution
First Author:
Sandeep Venkataram
Stanford University Department of Biology
Second Author:
Diamantis Sellis
Stanford University Department of Biology
Third Author:
Dmitri A Petrov (Corresponding author)
Stanford University Department of Biology
Key Words: epistasis, heterozygote advantage, fitness, landscape, selection

Abstract:

Predicting adaptive evolutionary trajectories is a primary goal of evolutionary biology. One can differentiate between forward and backward predictability, where forward predictability measures the likelihood of the same adaptive trajectory occurring in independent evolutions and backward predictability measures the likelihood of a particular adaptive path given the knowledge of starting and final states. Recent studies have attempted to measure both forward and backward predictability using experimental evolution in asexual haploid microorganisms. Similar experiments in diploid organisms have not been conducted. Here we simulate adaptive walks using Fisher's Geometric Model in haploids and diploids and find that adaptive walks in diploids are less forward- and more backward-predictable than adaptive walks in haploids. We argue that the difference is due to the ability of diploids in our simulations to generate transiently stable polymorphisms and to allow adaptive mutations of larger phenotypic effect. As stable polymorphisms can be generated in both haploid and diploid natural populations through a number of mechanisms, we argue that inferences based on experiments in which adaptive walks proceed through succession of monomorphic states might miss many of the key features of adaptation.

Introduction

Since Gould's famous thought-experiment (Gould, 1990) on "replaying the tape of life", scientists have been interested in the predictability of evolution. Gould wondered whether it is possible to forecast evolution, and determine the path or the final destination of the evolutionary process from a given starting population. It is also possible, however, to ask whether we can retrocast evolution, and reconstruct the true evolutionary trajectory given the final state and possibly the ancestral state. Forward predictability analysis tries to predict the future evolutionary trajectory or future adapted state of an evolving population, while backwards predictability analysis tries to determine the likelihood of the possible alternative adaptive trajectories that lead to the observed adapted state. This distinction between forward and backward predictability is rarely made (however see Szendro et al. (2013)) yet it is clear that these are distinct concepts and that high forward predictability does not necessarily imply high backward predictability and vice versa.

Forward predictability of evolution. Past experimental evolution studies have generally (but not always) focused on the forward predictability of evolution and suggest that evolution is indeed predictable to a surprising degree. For example, Ferea et al. (1999), Cooper et al. (2003) and Fong et al. (2005) evolved independent replicates of microbes and observed phenotypic changes in gene expression and growth rate in the evolved clones. A large study of 145 parallel long-term experimental evolutions with *E.coli* grown at elevated temperature showed that the same genes and pathways were repeatedly targeted for mutations in independent populations (Tenaillon et al., 2012) as did a study of 40 replicate S. cerevisiae batch culture evolutions(Lang et al., 2013). Herron and Doebeli (2013) evolved *E. coli* under multiple carbon sources and repeatedly observed the evolution of two distinct ecotypes with differential ability to grow on each carbon source. By sequencing independent replicate clones of both ecotypes they found the same genes, and sometimes the same exact mutations invading these replicate populations and differentiating the ecotypes.

Repeated evolution has been observed at both the genetic and morphological levels in natural systems as well. Kvitek et al. (2008) showed that highly divergent yeast strains isolated from oak trees had similar growth rates across a panel of diverse growth conditions. Studies of *Anolis* lizards in the Caribbean show repeated independent adaptive radiations into three phenotypically similar niches across the islands (Losos, 1998). In addition, a study of the adaptive radiation of cichlid fish in Lake Tanganyika showed convergent morphological evolution when the skeletal morphology of the various species was compared to their phylogeny (Muschick et al., 2012). Note, however, that practically all of these studies analyze whether the phenotypes of extant organisms are predictable (forward predictability of the adapted state), but they have limited power to assess the forward predictability of adaptive trajectories.

Backward predictability of evolution. Weinreich et al. (2006) conducted the first study to systematically probe the backward predictability of evolution, using a combinatorially complete reverse genetic study design pioneered by Malcolm et al. (1990). In their study, Weinreich et al. reconstructed every possible combination of five mutations in β-lactamase gene in *E. coli* known to lead to high levels of resistance to the drug rifampicin and assayed their levels of drug resistance, which they used as a proxy for fitness. From these data, they assayed the fitness changes involved in every step of each of the 120 possible paths involving these five mutations. A mutational path was deemed viable if fitness monotonically increased

with every step, and each viable path was assigned a probability based on the size of each fitness change. They found that only 18 of the 120 possible paths were viable, suggesting high backward predictability of evolution. In contrast, 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 orders were viable. Finally, Franke et al. (2011) studied backwards predictability in all subsets of two to six mutations in an empirical eight-locus system and found that the number of viable paths varied widely for a given subset size. For example, they observed both zero and nine viable paths (out of 24 possible) in different four locus subsets. The varying degrees of backward predictability found in these different systems does not yet allow us to draw general conclusions, and the laborious nature of the experiments makes it challenging to study more than a few mutations at a time.

The Weinreich et al method assays a particular type of backward predictability, where predictability is conditioned on the knowledge of the initial and final states and of all of the steps taken along the true path. Thus it only assays the likelihood of specific orders of steps. By definition, this metric cannot assay the likelihood of an alternative ancestral state or paths that involve mutations that are not present in the observed final state. In principle, one can first carry out a large number of forward evolutions starting from the same genotype and perform the Weinreich method for all of them. Such an experimental effort would illuminate the relationship between forward and backward predictabilities, but is unreasonably laborious and thus infeasible. We utilize simulated adaptive walks to circumvent these practical restrictions.

Here we employ simulations of adaptation in the context of Fisher's geometric model (FGM, Fisher (1930)) to study forward and backward predictability. FGM is a well-studied (e.g. Orr, 1999, 2005) phenotypic model with a single smooth fitness peak and a multidimensional phenotypic space (Figure 1a). We initialize our simulations with populations monomorphic for the same phenotype, and evolve them on a constant fitness landscape with a single smooth peak. Therefore, each independent adaptive walk tends to end up in the vicinity of the optimal phenotype. As we largely constrain both the beginning and the end of the walk, we focus on the predictability of evolutionary paths. We study both forward path predictability, by asking whether different paths tend to go through the same part of the phenotypic space, and backward path predictability, by using the method of Weinreich et al's.

We simulate adaptation in both haploids and diploids in a manner similar to that of Sellis et al. (2011). Their work showed that diploid FGM simulations generate a large number of overdominant adaptive mutations under certain parameter regimes, resulting in balanced polymorphisms, while haploid simulations always result in the classic successive fixation regime (also see Manna et al. (2011) for an analytical treatment of dominance in FGM). The comparison of haploid and diploid simulations allows us to investigate the impact of stable polymorphisms and overdominant mutations on the predictability of evolution. As reciprocal sign epistasis is necessary to generate the fitness valleys that increase predictability (Poelwijk et al., 2011), we also study the impact of various forms of sign epistasis on predictability. We conclude that the generation of stable polymorphic states has substantial effects on predictability of evolution, and discover that at least in this model, forward and backward predictability are inversely correlated.

Materials and Methods

Geometric Simulations

We model adaptive walks in haploid and diploid populations with Wright-Fisher simulations using Fisher's geometric model. The simulations use code modified from Sellis et al. (2011). We perform 10,000 replicate simulations with population size N=10,000 for haploids and N = 5,000 for diploids preserving the number of alleles in the population. Simulations are conducted for 10,000 generations in the geometric model. Each allele is represented as a point in coordinate space (PCS). The phenotype of haploids and homozygous diploids is simply the coordinates of the constituent allele. The phenotype of heterozygotes is the midpoint of the two points of the constituent alleles. This results in an assumption of phenotypic additivity of alleles in determining the phenotype of a genotype. This methodology is identical to the simulation methodology used by Sellis et al. (2011). The population initially contains a single allele with the PCS (2,0), and evolves on a fitness landscape with single phenotypic optimum at (0,0). Fitness is computed using a gaussian function :

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

where x is the distance of the phenotype to the optimum. The mutation rate is set to $\mu = 10^{-6}$ which results in one mutation every 100 generations on average for both haploids and diploids. Mutations are vectors in phenotype (coordinate) space which modify the PCS of the mutated allele by adding the mutation vector to the original PCS of that allele (Figure 1a). This results in the implicit assumption that mutations are additive in phenotype space. The magnitude of the mutation vector is drawn from an exponential distribution, with $\lambda = 1/2$. The angle of the mutation vector is drawn from a uniform distribution. Results using different values of λ (1/4, 1/2 or 2) and μ (10^{-6} to 10^{-4}), as well as results using 2, 3, 4 or 7 dimensions are described in the Supplementary Information.

In essence, FGM is an infinite allele single locus model where each allele has a phenotype that is a point in coordinate space. The model is additive in phenotype space, but nonadditive in fitness space. We utilize this model as it is very similar to asexually evolving systems such as the bacteria used in Weinreich et al. (2006), where all mutations that occur on a given background are completely linked to each other.

For the remainder of our analysis, 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.

Sign Epistasis and Conditional Mutations

We are first interested in the extent of sign epistasis between independent adaptive mutations (Figure 2a,b). We consider the extent of sign epistasis between initial adaptive mutations in 5,000 independent pairs of simulations for each parameter set in both haploids and diploids. Limiting our analysis to only the initial adaptive mutations allows us to compare mutations that occurred on the same ancestral allele. Sign epistasis, reciprocal sign epistasis and ancestral deleterious epistasis are defined as in Figure 2b. The level of epistasis is computed

as follows. We compute the equilibrium populations resulting from the invasion of each of the initial mutations on the ancestral population, as described below. We then compute the equilibrium probability (described below) of the double mutant allele invading each of these resulting equilibrium population states, as well as the equilibrium probability of the double mutant in the ancestral population. If the double mutant cannot invade the ancestor, it has ancestral deleterious epistasis, if it cannot invade either single mutant state, it has reciprocal sign epistasis, and if it cannot invade only one of the single mutant states, it has sign epistasis. Note that these double mutant alleles were never observed in any simulation, and we are artificially genereated these novel alleles to study epistasis between independent adaptive mutations.

To study the rate of conditional mutations (Figure 2c), we employ a similar procedure, but analyze the first two mutations of each simulated walk. If the second mutation of the simulated walk cannot invade the ancestral population, the mutation is said to be beneficial conditional on the presence of the initial mutation of the walk.

We note that computing the fitness advantage conferred by a mutation in diploids is more complicated than in haploids as mutations can be heterozygous or homozygous. Overdominant mutations, which maintain variation in diploid populations, add additional complications to the computations. New overdominant mutations have a high marginal fitness that depends on the fitness of all heterozygotes of the new allele with the established alleles and weighted by their frequency. For computing the fitness of a new mutation, we first computed the equilibrium probability of the mutation into the current population state when rare. If it cannot invade, we claim the mutation is deleterious. Assuming the mutation can invade, we compute the mean fitness and equilibrium allele frequencies that can be achieved by the new allele with each of the existing alleles in pairs. As our simulations suggest that equilibrium states with three or more alleles are very rare (0.1% of simulations), we ignore such possibilities with minimal impact on the qualitative results of our analysis.

Probability of a new mutation reaching an equilibrium frequency

The probability of new mutations reaching equlibrium from a single copy, which we call equilibrium probability, is empirically computed through 10,000 Wright-Fisher simulations for both haploids and diploids. If the marginal fitness of the new mutation is less than the mean fitness of the existing population, the mutation is deleterious and the equilibrium probability is zero. Otherwise, for haploids, we compute the probability as the likelihood of the allele reaching fixation from a single copy. In diploids, we first estimate the equilibrium allele frequency of the new mutation when the new allele is present with each of the alleles already in the ancestral population. This methodology requires the assumption that the equilibrium state consists of at most two alleles. Since we observe stable states consisting of 3 or more alleles less than ten times in 10,000 simulations, this approxiation seems reasonable. The equlibrium allele frequency of the new allele for each of these cases is

$$(h-1)/(2h-1)$$

(Gillespie, 2004), where h is the dominance coefficient of the new allele relative to the given extant allele. Given this equilibrium frequency, we can compute the mean population

fitness when the new allele is at equilibrium with that particular extant allele. This is repeated for all extant alleles, and the equilibrium state with the extant allele that grants the maximum equilibrium mean population fitness defines the new population state. The true equilibrium frequency of the new allele is thus the frequency of the allele in this new population state.

The equilibrium probability of the new diploid allele is computed as the probability of the allele reaching 90% of the equilibrium frequency from a single copy in the empirical Wright-Fisher simulations. This equilibrium frequency can be either fixation or some intermediate frequency due to heterozygote advantage. We are forced to utilize empirical estimations through simulations and not the classical analytic solutions for invasion and fixation probability 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.

Forward Predictability Analysis

We calculate forward predictability using two metrics. In both of these metrics, we only consider homozygous genotypes in diploids. Our first metric, maximum pariwise distance, considers pairs of adaptive walks with exactly five mutations. We compute the maximum of the phenotypic distances between the observed single mutant genotypes of the two adaptive walks, the double mutant genotypes, the triple mutant genotypes etc. Our second metric measures the maximal deviation from the optimal trajectory. For each adaptive walk (regardless of length), we compute the maximal phenotypic distance of any encountered (homozygous) genotype from the line segment connecting the ancestral phenotype and the optimum. This is thus the maximal phenotypic distance of the adaptive walk from the optimal adaptive trajectory.

Backward Predictability Analysis

We compute backward predictability using simulated adaptive walks of exactly five mutations. We calculate the probability of all possible mutational trajectories for the given set of mutations in a manner similar to Weinreich et al. (2006), but generalized to allow balanced states. The likelihood of a mutational trajectory is the product of the probabilities of each mutation being generated on the appropriate background and successfully reaching the equilibrium frequency in succession. The probability of a mutation being generated is proportional to the frequency of the appropriate ancestral background. For example, if the ancestral state is balanced with the 1-mutant state, the probability of generation of a 2-mutant state is proportional to the frequency of the 1-mutant state. The equilibrium probability of the mutation from a single copy is calculated empirically as described above. In the haploid model, where each successive mutation fixes in the population, there are $5! = 120$ possible orders of the five mutations observed in the simulated walk to generate the five-mutant allele observed at the end of the simulation. In the diploid model, each mutation can occur on any allele in the population where it has not already been introduced, leading to at most $(2 * 5 - 1)! = 362880$ possible mutation orders. The probabilities of all viable mutational orders are then rescaled to add up to 1, to give the probability of a trajectory conditional on reaching the final n-mutant state.

We define the effective number of trajectories as

$$1/\sum(probability\ of\ trajectory^2)$$

in a similar manner to the effective number of alleles in a population (Kimura and Crow, 1964), the entropy metric of Palmer et al. (2013), and the predictability metric of Roy (2009). Thus, when a single trajectory dominates the probability density, the effective number of trajectories is close to 1, even if there are many other trajectories with nonzero probability.

Results

We performed 10,000 replicate simulations of adaptation under FGM in haploids and diploids (haploid N = 10,000, diploid N = 5,000, mutation rate = 10^{-6}, two phenotypic dimensions and Gaussian fitness with a single peak at the origin, see materials and methods for details). Mutational magnitudes were drawn from an exponential distribution with parameter $\lambda = 1/2$, with the population initiated at 2 units from the optimum. For all of our analyses, we consider only those mutations that are on the path to the most frequent genotype at generation 10,000. Such mutations are typically the only ones available for analysis in a natural system. We begin with a study of epistasis, which is related to the ruggedness of the fitness landscape, and follow with a study of the predictability of evolution.

Figure 1b and 1c show the distribution of the points in coordinate space (PCS) for all alleles that have successfully invaded the population across all simulations. Note that in diploids we show the phenotype of the derived homozygotes. Black circles indicate the boundaries for α-dip, α-hap and γ for the ancestral state as described in Figure 1a. α-dip defines the phenotypic area for all mutations that are beneficial in diploids relative to the ancestral phenotype, while α-hap defines the set of beneficial mutations in haploids. γ defines the area where beneficial mutations would be replacing in diploids, as opposed to generating a balanced polymorphism. It is evident that simulated diploid populations explore more of the phenotypic space, which is consistent with the findings of Sellis et al. (2011). Intriguingly, this is not only because α-dip is larger than α-hap but also because many of the observed homozygous diploid phenotypes are located outside α-dip. This implies that these alleles (whether heterozygous or homozygous) are only beneficial due to the presence of other alleles in the population given that these alleles are deleterious by themselves. This also suggests that the ability of diploid populations to maintain variation during adaptation due to frequent overdominance substantially changes the process of adaptation. As we will demonstrate below, the presence of overdominant mutations in diploids changes patterns of sign epistasis, conditionality, and predictability compared to haploids.

Sign Epistasis

Sign epistasis between independent mutations has been empirically studied in a number of systems, but with limited sample size for each system(Kvitek and Sherlock, 2011; Rokyta et al., 2011; Lalić and Elena, 2013). Here, we isolated the first adaptive mutation along independent adaptive walks in 5,000 pairs of simulations (mutations A_1 and B_1 in Figure 2a). We computed the equilibrium population state resulting from each of the single mutant alleles invading the ancestral population individually. We then computed the probability of the synthetically generated double mutant allele invading each of these equilibrium single mutant states (purple lines in Figure 2a). If the double mutant could invade one but not the other of the single mutant equilibrium states, the system has sign epistasis. If the double mutant could invade neither of the single mutant states, there is reciprocal sign epistasis, and if it could not invade the ancestral population, the system has ancestral deleterious epistasis (Figure 2b).

The frequency of sign, reciprocal sign and ancestral deleterious epistasis in diploids and haploids is shown in Table 1. Full data for all parameter regimes is provided in Supplementary

Table 1. Haploids and diploids have similar levels of sign epistasis and reciprocal sign epistasis, but it appears that much of the sign epistasis in haploids is driven by ancestral deleterious epistasis, which is not the case in diploids.

We further recomputed the frequencies of epistasis in diploids conditioning on the number of overdominant mutations in each mutation pair (Figure 3). We found that pairs of mutations which include one or two overdominant mutations have a significantly higher prevalence of sign, reciprocal sign, and ancestral deleterious epistasis relative to pairs which include no overdominant mutations (Welch's t test $p<<10^{-10}$ in most cases, see Supplemental Table 5 for details). This is consistent with the idea that overdominant mutations are often mutations of large phenotypic effect, and combining two large effect mutations together should often lead to a maladaptive result in FGM.

Conditional Mutations

Of possibly greater interest are epistatic interactions between sequential mutations along adaptive walks. We examined the first two adaptive mutations that occurred in each of our simulations to analyze the ability of the second adaptive mutation to invade the ancestral population, and term those that are deleterious on the ancestral background as conditional (Figure 2a,c). In this system of two-mutation adaptive walks, there are two possible orders of mutations to generate the double mutant allele, only one of which was observed in the simulation. A conditional second mutation means that the order of the two mutations is completely backward predictable, as the order that was not observed has a deleterious intermediate. Conditionality is much more common in diploids compared to haploids (Table 2, χ^2 p $<10^{-10}$, df=1). We observed that conditional mutations were significantly more frequent when following an overdominant mutation compared to a non-overdominant mutation (Table 2, χ^2 p $<<10^{-10}$, df=1). The complete data for all parameter regimes provided in Supplementary Table 2 confirms this intuition. Our results suggest that all regimes where a large frequency of overdominant mutations are expected (Sellis et al., 2011) should also generate a large frequency of conditional mutations.

Predictability of Adaptive Walks

We first consider forward predictability of the phenotypic path, which we define as the tendency of independent adaptive walks to explore similar portions of the phenotypic space. We quantified this by measuring the distribution of maximal phenotypic distances between pairs of independent adaptive trajectories with exactly five mutations (see materials and methods). According to this measure, haploid walks are significantly more phenotypically forward path predictable than diploid walks (Figure 4, Welch's t test $p<<10^{-10}$). We also measure forward predictability as the maximal phenotypic deviation of each observed trajectory from the optimal trajectory- the line connecting the ancestral phenotype and the optimal phenotype. We again observe that haploid walks are significantly more forward predictable than diploid walks (Figure 5, Welch's t test $p<<10^{-10}$). These results are concordant with Figure 1, where haploids explored a much smaller phenotypic space than diploids over the course of evolution. The ability of diploids to explore a larger phenotypic space using overdominant mutations when evolving to the same phenotypic optimum as

haploids naturally leads to lower predictability of the phenotypic intermediates along the adaptive walk.

We then studied backward predictability *sensu* Weinreich et al. (2006). We limit our analysis to adaptive walks of exactly five mutations, which is comparable to many of the experimental studies of evolutionary predictability (Weinreich et al., 2006; Khan et al., 2011; Franke et al., 2011). We computed the probability of observing a particular path to the five-mutant state by successively introducing each of the five mutations on the ancestral background and assessing the probability of reaching equilibrium (see materials and methods). In diploids, this computation involves the probability of the next mutation occurring on either background in a potentially balanced system multiplied by the probability of the new state invading the existing population. Every available mutation was introduced onto every allele present in the population, with each allele having a particular mutation at most once. Mutational orders were deemed viable if the five-mutant state could be successfully reached by sequential introduction of mutations. The haploid computation does not consider balanced states, and is functionally identical to the method of Weinreich et al. (2006). We computed the effective number of adaptive trajectories for each observed trajectory, which weights the number of viable mutational orders by their probabilities.

The results of this analysis are shown in Figure 6. We found that haploid adaptive walks are substantially less backwards predictable than diploids as measured by the effective number of viable paths (Welch's t test $p<10^{-10}$). In other words, conditional on reaching a particular five-mutant state, it is more probable that independent diploid realizations will use the same mutational order. We also utilized the mean path divergence of Lobkovsky et al. (2011) and found that diploids were more backward predictable by this metric as well (Welch's t test $p = 10^{-10}$).

If our method is truly computing the likelihood of a particular adaptive trajectory, it should be the case that high probability trajectories are more likely to have been observed in the FGM simulations than low probability ones. We bin all trajectories with non-zero probability from the simulations with five mutations into 40 equal-sized bins, and compute the number of trajectories within each bin that was actually observed in the simulations. We find that the median bin probability is significantly positively correlated with the number of observed trajectories in both diploids ($p < 10^{-10}$) and haploids ($p < 10^{-7}$).

We found that 62% of diploid and 58% of haploid simulations had less than 25 viable paths. A larger proportion- 90% of simulations in diploids and 78% in haploids- have fewer than 25 effective paths. As there should be 120 possible paths for simulations containing five mutations, these statistics suggest that most mutational orders were inviable in most simulations, but even of those that were viable, there were a subset of mutational orders that were far more likely than the rest. As 29% of diploid simulations, but only 19% of haploid simulations had more than 25 viable orders but less than 25 effective paths, diploid sets of mutations were more likely to have a few high probability mutational orders compared to haploids. In contrast, only 1% of simulations had over 90 effective paths in both haploids and diploids, suggesting that these walks took relatively direct paths towards the optimum. The backward path predictability in our simulations appears to be highly variable and strongly depends on the particular set of mutations used to assess it.

We also observed that the same series of mutations, when introduced in different orders,

can give rise to different population states containing the final adapted allele. As different mutation orders generate intermediate alleles containing different sets of mutations, it is possible that the final five-mutant allele can balance against different intermediate alleles in different orders. We find that $> 65\%$ of diploid simulations have at least two different possible end population states containing the final adapted allele, with a median of 2 different population states and a maximum of 20 different population states for a single set of five mutations.

We note that backwards predictability analysis is conducted over genotypic landscapes with a limited number of possible states such as in Khan et al. (2011)'s Figure 1, while forward predictability is conducted over phenotypic landscapes (Figure 1a) which contains an infinite number of possible states in FGM. Therefore, it is possible to have no viable sequence of mutations to evolve one genotype to another in the restricted genotypic space.

Inaccessible Trajectories

In the process of computing backward predictability of diploid walks, it became clear that many of the adaptive states that are generated and are stably maintained during a diploid walk do not survive until the end of the evolution. We call these hidden mutations, as they are hidden from both our analysis and almost all modern experimental studies of adaptation in nature. Similar to our observations for transient mutations, lack of knowledge of hidden mutations may decrease the computed probability of the true adaptive path observed in the FGM simulations, and in extreme cases, even make the true path impossible to reconstruct. To study this phenomenon to a first approximation, we computed the excess number of mutations in the population that reached 5% frequency relative to the number of mutations present in the final adapted state. Such mutations are likely to have been beneficial to reach such high frequency, but died out before they fixed. Again, we limit our results to simulations with five mutations along the adaptive trajectory. We found that diploid simulations have many more extra mutations than haploids (58% vs 7.5%, Table 3, complete data in Supplementary Table 3) as predicted by the presence of balanced states in diploids but the lack of such states in haploids.

We quantify the frequency of simulations in which the observed mutational order is not viable using the backward predictability analysis *sensu* Weinreich et al. We found these apparently impossible evolutionary outcomes at a rare but appreciable frequency in diploids (7% of simulations), but never in haploids. This result strongly argues for the existence of hidden states in the diploid simulations that are necessary for the reconstruction of the true adaptive walk by our backwards predictability method.

Relationship Between Overdominance and Repeatability

We have so far analyzed epistasis and the predictability of evolution separately, and now explore the impact of epistasis on the predictability of evolution. As noted earlier, overdominant mutations are correlated with high levels of conditional mutations (Figure 4). As conditional mutations, by definition, are only beneficial on the background of the initial mutation and not the ancestor, such mutations limit the number of viable mutation orders and increase backward predictability. Therefore, overdominant mutations should increase backward predictability

by increasing the frequency of conditional mutations. Indeed, we found that the number of overdominant mutations in a simulated diploid adaptive walk is strongly negatively correlated with the effective number of paths computed for the same walk (Pearson correlation coefficient = -0.35, N=645, $p<<10^{-10}$). Therefore, a larger number of overdominant mutations (that frequently cause subsequent conditional mutations) in a walk decreases the effective number of paths, and thus increases backward predictability.

Discussion

In this study, we explored the predictability of evolution during adaptive walks using Fisher's geometric model. We distinguished between forward and backward predictability, where forward predictability measures the likelihood of the same or similar adaptive trajectory occurring in independent evolutions, while backward predictability measures the likelihood of a particular adaptive path given the ultimate adapted state. We knew from prior work that haploids and diploids behave very differently under the geometric model (Sellis et al., 2011), so we studied predictability in both haploids and diploids. We found haploids are more forward predictable than diploids, while diploids are more backwards predictable than haploids.

High backward predictability in diploids is consistent with high forward predictability in haploids. Indeed, an adaptive walk that goes directly to the optimum should have the lowest possible backward predictability as every order of mutations is adaptive. Given that adaptive walks in haploids tended to go more directly towards the optimum than diploids, it is not surprising that adaptive walks in haploids were less backward predictable than in diploids. An alternative metric of backward predictability called the mean path divergence (Lobkovsky et al., 2011) also supported our results (Supplementary Figure 6).

The primary difference between diploid and haploid adaptive walks in our model is the presence of overdominant mutations creating stable polymorphic states. In natural populations, stable polymorphisms can arise not only through overdominant mutations, but also through any mechanism of frequency dependent selection such as ecological niche construction (Rainey and Travisano, 1998), and spatially or temporally variable environments (Kasumovic et al., 2008), and can occur in haploids as well as diploids through these mechanisms. While the complications we describe are only possible in diploids in our simulations due to the simplistic nature of the geometric model, most of our results are pertinent to both haploid and diploid natural populations, as both can experience stable polymorphisms.

In our implementation of Fisher's Model, balanced states arise when mutations are overdominant. If most mutations are of small effect relative to the distance to the optimum, or if there are additional dimensions in the phenotype space, it becomes less likely that the mutation will be overdominant relative to the initial parameter set. The presence of additional dimensions essentially reduces the effect size of new mutations, and is therefore very similar in practice to reducing the mutational step size. In this case, diploids are at a disadvantage relative to haploids, as most mutations are now codominant, and the benefit of novel mutations is diluted in heterozygotes when rare. This can be seen when we vary the parameter values of our simulations (Supplementary Figures 1-6 and Supplementary Tables 1-4). When mutations have very small magnitudes in our simulations ($\lambda = 2$), haploids have higher frequencies of sign epistasis and fewer effective paths compared to diploids. In simulations with higher dimensions (3, 4 and 7), we again observe more sign epistasis in haploids compared to diploids. Therefore, the epistasis and backward predictability are increased in diploids compared to haploids when there is a high chance that mutations can be overdominant, otherwise the fitness benefit of new mutations is diluted to the point that haploids have more epistasis and backwards predictability than diploids.

The presence of balanced states can result in branching mutational lineages, resulting in

hidden mutations and transient population states that nevertheless significantly impact the future course of evolution. Depending on the equilibrium frequency of a balanced state, it can be more or less likely that a particular genotype will arise through a particular mutational order. Ignoring hidden mutations can significantly impact the likelihood of a particular adaptive path, to the point of making it inviable and different orders of mutations can generate different sets of heterozygous genotypes and different end population states, further complicating analysis of diploid adaptation.

Branching mutational paths drastically increase the number of possible adaptive paths. In haploids, as adaptation proceeds through sequential fixation, one only needs to consider the fitness of the 2^n possible genotypes relative to the ancestral background for an n-mutation system. This is the methodology used in the experimental studies of Weinreich et al. (2006); Khan et al. (2011); Franke et al. (2011). However, for diploids, each adaptive trajectory needs to be considered individually as different balanced states are possible in each one. Within each trajectory, every mutation along the trajectory needs to be introduced into the prior population at low frequency on every available allele, and tracked until the frequency of the new mutation has been stabilized. To our knowledge, such a study has never been conducted in any system.

Experimental Implications

In an experimental setting, high forward predictability means it is likely that the same set of mutations will be generated in independent adaptive walks, which make the probabilities generated through backward predictability analysis meaningful for predicting future events. This can occur by either a small mutational target size such as mutations that cause resistance to drugs, or a large mutational input into the population which makes rare but extremely beneficial mutations that dominate the adaptive process common (e.g. Kvitek and Sherlock, 2011; Pennings, 2012; Gerstein et al., 2012; Desai and Fisher, 2007). A study in FGM also suggests that a multi-locus FGM where each locus only influences a subset of the independent phenotypic dimensions (restricted pleiotropy) also promotes forward predictability, which the authors call parallel evolution (Chevin et al., 2010). In our studies, a large number of replicate simulations are required to get an accurate assessment of forward predictability, as there is a large variance between individual simulations. Experimentally determining forward predictability by studying many replicate evolutions either through fitness assays or high throughput methods such as DNA-seq and RNA-Seq seems laborious but feasible.

On the other hand, the possibility of hidden mutations makes accurate estimates of backwards predictability impossible. Since we do not have access to hidden mutations from the past, it is impossible to accurately compute the backwards predictability of the adaptive walk leading to natural modern day populations. Studying backwards predictability using forwards evolutions and constant sampling is equally infeasible. For example, in humans, there are dozens of mutations occuring in every genome in every generation. Even if we could sample every mutation in every human on the planet, almost all of these mutations will be lost, and there are far too many to determine the subset which are non-neutral. There is also the problem of combinatorially many adaptive walks possible for even a few mutations as described above. As acknowledged by Weinreich et al. (2006), sampling a

few high-fitness mutations and conducting backwards predictabiliy experiments may not generate a correct representation of the probability of any particular adaptive walk. Such studies ignore adaptation and potential epistatic interactions at sites throughout the rest of the genome. Additional phenomena such as fluctuating abiotic and biotic environments can further complicate accurate assessments of backwards predictability.

Finally, the impact of hidden states on evolutionary trajectories depends on the frequency at which stable polymorphic states are generated. Stable polymorphisms have been observed via niche construction, fluctuating selection pressures and frequency dependent selection (Levin et al., 1988; Takahata and Nei, 1990; Rainey and Travisano, 1998; Kasumovic et al., 2008). In addition, mutations of large phenotypic effect are relatively common in nature (Steiner et al., 2007; Kenny et al., 2012), suggesting that overdominant mutations should be relatively frequent as well. This is in addition to phenomena such as multiple mutation and clonal interference regimes, where large mutational inputs generate a constant source of polymorphisms in the population (Desai and Fisher, 2007; Lang et al., 2013). Therefore, it is likely impossible to estimate backward predictability of adaptation in any natural population. Additional factors in natural populations that can complicate predictability estimation are multipeaked and fluctuating fitness landscapes and stochastic tunneling through local fitness maxima.

Conclusion

In this study, we examined the interactions between epistasis, evolutionary predictability and ploidy using simulations conducted using Fisher's Geometric Model. We found that simulations of diploid populations, which frequently generate overdominant mutations leading to balanced polymorphisms, had similar levels of sign epistasis between independent mutations but far more conditional mutations than simulations of haploid population. We defined forward predictability as the phenotypic similarity of independent walks, and backward predictability as the probability that a population will evolve along a particular trajectory given the initial and final states defined by an observed set of adaptive mutations. We found that diploid simulations are less forward predictable, but more backward predictable than haploid simulations, demonstrating that different metrics of predictability are not interchangable.

The presence of balanced polymorphisms gives rise to complex evolutionary phenomena that are never observed in our haploid simulations. We observed instances of different mutation orders giving rise to different final adaptive population states, and transient alleles being necessary for the viability of the observed adaptive walk. In a natural system, these phenomena should be possible any time a balanced state is generated, such as in frequency dependent selection, niche construction or spatial or temporally variable selection. We suggest that predicting evolution is extraordinarily complex even for simple systems, and that acurrately computing backward predictability is extremely challenging for experimental studies.

References

Chevin, L.-M., G. Martin, and T. Lenormand (2010, November). Fisher's model and the genomics of adaptation: restricted pleiotropy, heterogenous mutation, and parallel evolution. *Evolution 64*(11), 3213–31.

Cooper, T. F., D. E. Rozen, and R. E. Lenski (2003, March). Parallel changes in gene expression after 20,000 generations of evolution in Escherichiacoli. *Proc Natl Acad Sci U S A 100*(3), 1072–7.

Desai, M. M. and D. S. Fisher (2007, July). Beneficial mutation selection balance and the effect of linkage on positive selection. *Genetics 176*(3), 1759–98.

Ferea, T., D. Botstein, P. O. Brown, and R. F. Rosenzweig (1999). Systematic changes in gene expression patterns following. *Proc Natl Acad Sci U S A 96*(August), 9721–9726.

Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*.

Fong, S. S., A. R. Joyce, and B. O. Palsson (2005, October). Parallel adaptive evolution cultures of Escherichia coli lead to convergent growth phenotypes with different gene expression states. *Genome Res 15*(10), 1365–72.

Franke, J., A. Klozer, J. A. G. M. de Visser, and J. Krug (2011). Evolutionary Accessibility of Mutational Pathways. *PLoS Comput Biol 7*(8).

Gerstein, A. C., D. S. Lo, and S. P. Otto (2012, June). Parallel Genetic Changes and Non-parallel Gene-environment Interactions Characterize the Evolution of Drug resistance in Yeast. *Genetics 192*(September), 241–252.

Gillespie, J. (2004). *Population Genetics: A Concise Guide* (2nd ed.). The Johns Hopkins University Press.

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

Herron, M. D. and M. Doebeli (2013, February). Parallel Evolutionary Dynamics of Adaptive Diversification in Escherichia coli. *PLoS Biol 11*(2), e1001490.

Kasumovic, M. M., M. J. Bruce, M. C. B. Andrade, and M. E. Herberstein (2008, September). Spatial and temporal demographic variation drives within-season fluctuations in sexual selection. *Evolution 62*(9), 2316–25.

Kenny, E. E., N. J. Timpson, M. Sikora, M.-C. Yee, A. Moreno-Estrada, C. Eng, S. Huntsman, E. G. Burchard, M. Stoneking, C. D. Bustamante, and S. Myles (2012, May). Melanesian blond hair is caused by an amino acid change in TYRP1. *Science 336*(6081), 554.

Khan, A. I., D. M. Dinh, D. Schneider, R. E. Lenski, and T. F. Cooper (2011, June). Negative epistasis between beneficial mutations in an evolving bacterial population. *Science 332*(6034), 1193–6.

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

Kvitek, D. J. and G. Sherlock (2011, April). Reciprocal Sign Epistasis between Frequently Experimentally Evolved Adaptive Mutations Causes a Rugged Fitness Landscape. *PLoS Genet 7*(4), e1002056.

Kvitek, D. J., J. L. Will, and A. P. Gasch (2008, October). Variations in stress sensitivity and genomic expression in diverse S. cerevisiae isolates. *PLoS Genet 4*(10), e1000223.

Lalić, J. and S. F. S. Elena (2013, February). Epistasis between mutations is host-dependent for an RNA virus. *Biol Lett 9*(1), 20120396.

Lang, G. I., D. P. Rice, M. J. Hickman, E. Sodergren, G. M. Weinstock, D. Botstein, and M. M. Desai (2013, July). Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature*.

Levin, B. R., J. Antonovics, and H. Sharma (1988, July). Frequency-Dependent Selection in Bacterial Populations [and Discussion]. *Philos Trans R Soc Lond B Biol Sci 319*(1196), 459–472.

Lobkovsky, A. E., Y. I. Wolf, and E. V. Koonin (2011, December). Predictability of evolutionary trajectories in fitness landscapes. *PLoS Comput Biol 7*(12), e1002302.

Losos, J. B. (1998, March). Contingency and Determinism in Replicated Adaptive Radiations of Island Lizards. *Science 279*(5359), 2115–2118.

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

Manna, F., G. Martin, and T. Lenormand (2011, November). Fitness landscapes: an alternative theory for the dominance of mutation. *Genetics 189*(3), 923–37.

Muschick, M., A. Indermaur, and W. Salzburger (2012, December). Convergent evolution within an adaptive radiation of cichlid fishes. *Curr Biol 22*(24), 2362–8.

Orr, H. a. (1999, December). The evolutionary genetics of adaptation: a simulation study. *Genet Res 74*(3), 207–14.

Orr, H. A. (2005, March). The genetic theory of adaptation: a brief history. *Nat Rev Genet 6*(2), 119–27.

Palmer, M., A. Moudgil, and M. W. Feldman (2013). Long-term evolution is surprisingly predictable in lattice proteins. *J Royal Soc Interface 10*(March).

Pennings, P. S. (2012, June). Standing Genetic Variation and the Evolution of Drug Resistance in HIV. *PLoS Comput Biol 8*(6), e1002527.

Poelwijk, F. J., S. Tnase-Nicola, D. J. Kiviet, and S. J. Tans (2011, March). Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. *J Theor Biol 272*(1), 141–4.

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

Rokyta, D. R., P. Joyce, S. B. Caudle, C. Miller, C. J. Beisel, and H. a. Wichman (2011, June). Epistasis between beneficial mutations and the phenotype-to-fitness Map for a ssDNA virus. *PLoS Genet 7*(6), e1002075.

Roy, S. W. (2009, January). Probing evolutionary repeatability: neutral and double changes and the predictability of evolutionary adaptation. *PLoS ONE 4*(2), e4500.

Sellis, D., B. Callahan, D. A. Petrov, and P. W. Messer (2011). Heterozygote advantage as a natural consequence of adaptation in diploids. *Proc Natl Acad Sci U S A 2011*, 1–6.

Steiner, C. C., J. N. Weber, and H. E. Hoekstra (2007, September). Adaptive variation in beach mice produced by two interacting pigmentation genes. *PLoS Biol 5*(9), e219.

Szendro, I. G., J. Franke, J. A. G. M. de Visser, and J. Krug (2013, January). Predictability of evolution depends nonmonotonically on population size. *Proc Natl Acad Sci U S A 110*(2), 571–6.

Takahata, N. and M. Nei (1990, April). Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics 124*(4), 967–78.

Tenaillon, O., a. Rodriguez-Verdugo, R. L. Gaut, P. McDonald, a. F. Bennett, a. D. Long, and B. S. Gaut (2012, January). The Molecular Diversity of Adaptive Convergence. *Science 335*(6067), 457–461.

Weinreich, D. M., N. F. Delaney, M. a. Depristo, and D. L. Hartl (2006, April). Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science 312*(5770), 111–4.

Figures

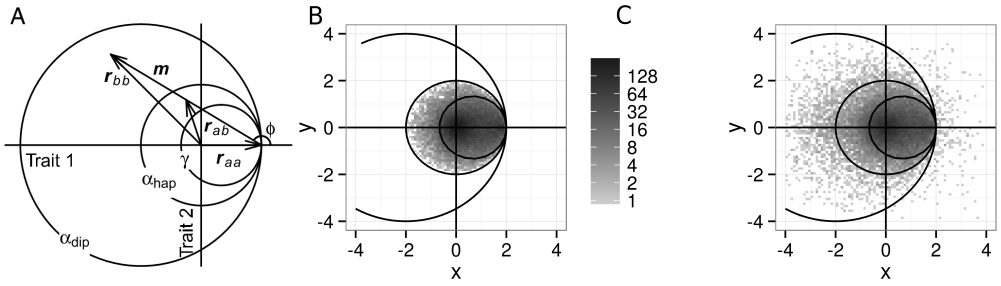

Figure 1. (A) Identical to Figure 2A Sellis et al 2011. Two orthogonal axes represent independent character traits. Fitness is determined by a symmetrical Gaussian function centered at the origin. Consider a population initially monomorphic for the wild-type allele $\vec{r_{aa}}$. A mutation m gives rise to a mutant phenotype vector $\vec{r_{bb}} = \vec{r_{aa}} + m$. The phenotype of the mutant heterozygote assuming phenotypic codominance ($h = 1/2$) is $\vec{r_{ab}} = \vec{r_{aa}} + m/2$. The different circles specify the areas in which mutations are adaptive in haploids (α-hap), adaptive in diploids (α-dip), and replacing in diploids (γ). (B). Phenotypes of all mutations along the adaptive walk of 10,000 haploid FGM simulations. (C) Homozygous phenotypes for diploid simulations. Red circles denote α-dip, α-hap and γ as described in (A).

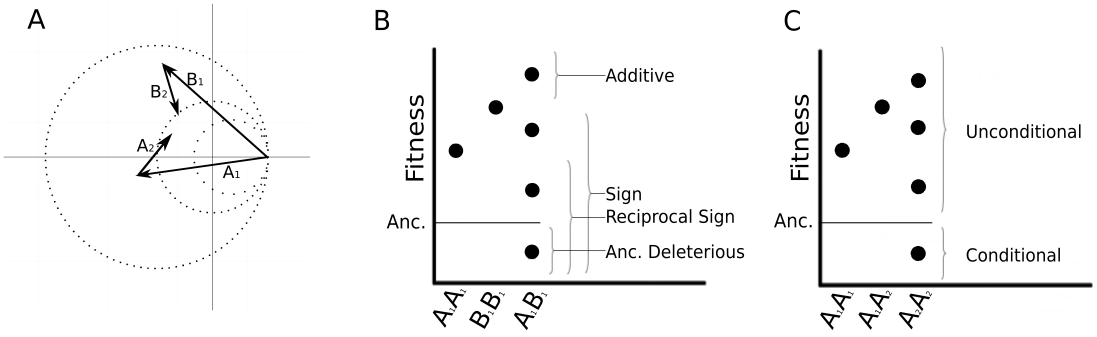

Figure 2. (A) Adaptive walks A, with mutations A_1 and A_2 and B, with mutations B_1 and B_2, occurred in independent diploid FGM simulations. Dotted circles denote α-dip, α-hap and γ as described in Figure 1A. Sign epistasis (B) between independent mutations occurs when the $A_1 + B_1$ double mutant is less fit than at least one of the two single mutants (A_1 or B_1 alone), with four possible fitnesses for the double mutant shown. Ancestral fitness is represented by the horizontal black line. Reciprocal sign epistasis occurs when the double mutant is less fit than both the A_1 and B_1 single mutants. Ancestral deleterious mutations occur when the double mutant is less fit than the ancestor. Conditional mutations (C) occur when latter mutations in the adaptive walk are deleterious on the ancestral background (A_2A_2 allele has the second mutation alone on the ancestral background) but beneficial conditional on the presence of the first mutation (A_1A_2 allele).

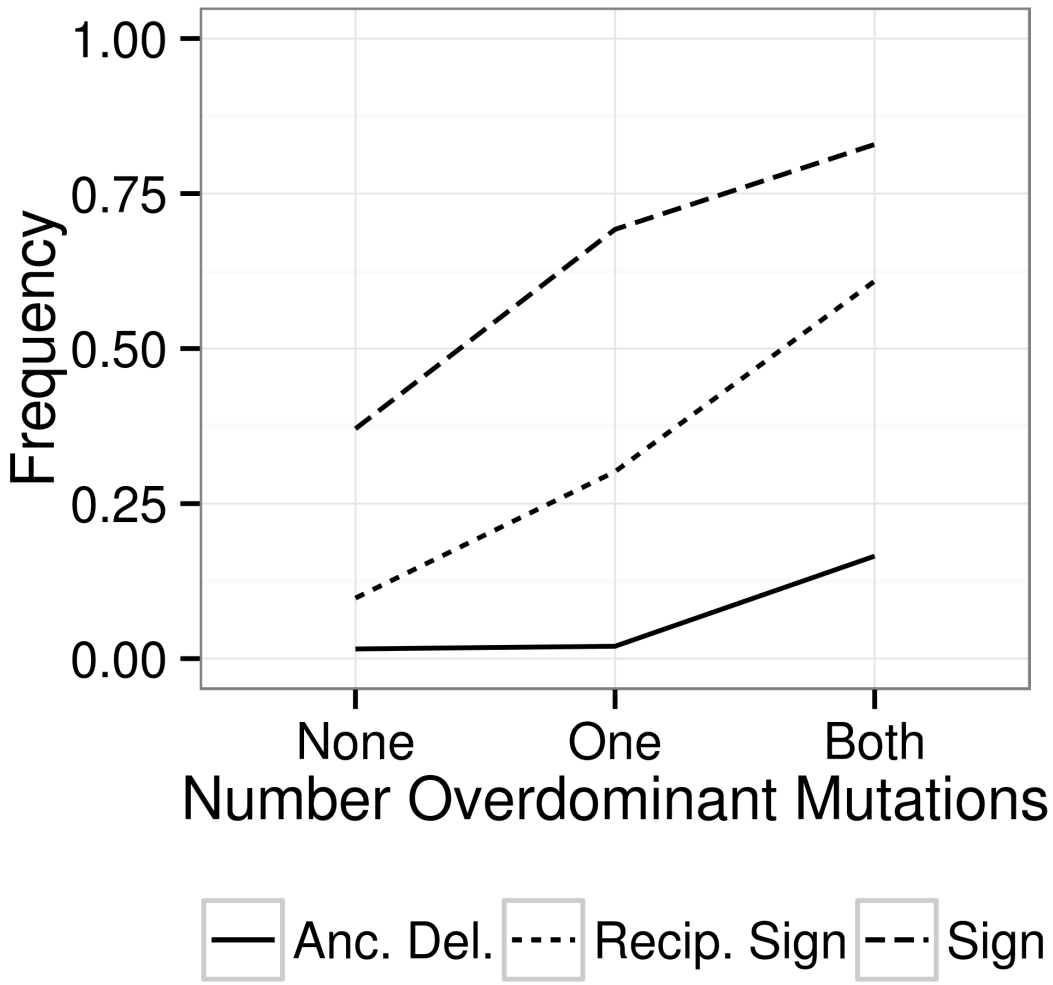

Figure 3. Epistasis is computed between pairs of initial mutations in independent adaptive walks. Frequency of sign epistasis, reciprocal sign epistasis and ancestral deleterious mutations in diploid and haploid adaptive walks as a function of the number of overdominant mutations present in each mutation pair.

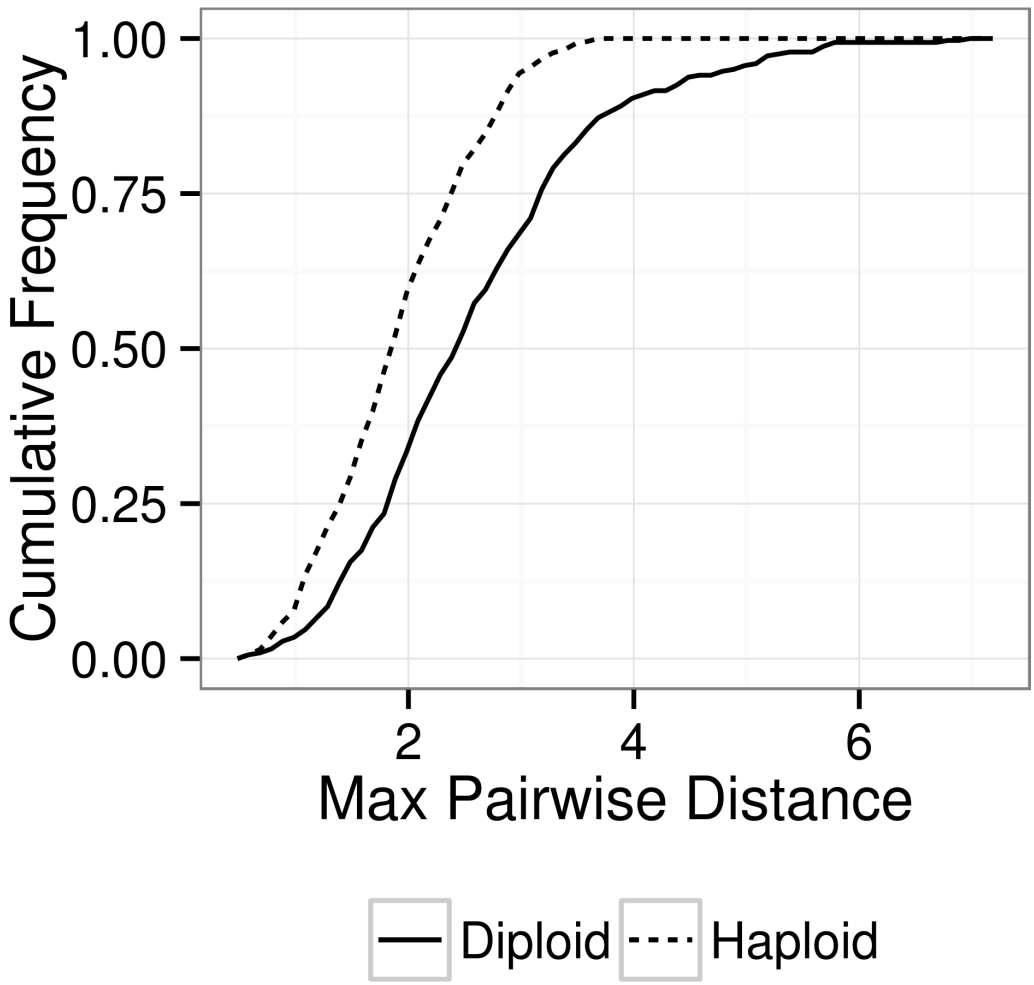

Figure 4. Cumulative distribution of the maximum phenotypic distance between independent adaptive walks, excluding the ancestral state. This is a measure of the phenotypic repeatability of independent walks on the same evolutionary landscape.The maximum phenotypic distance in haploids is significantly less than in diploids (Welch's t test $p<<10^{-10}$).

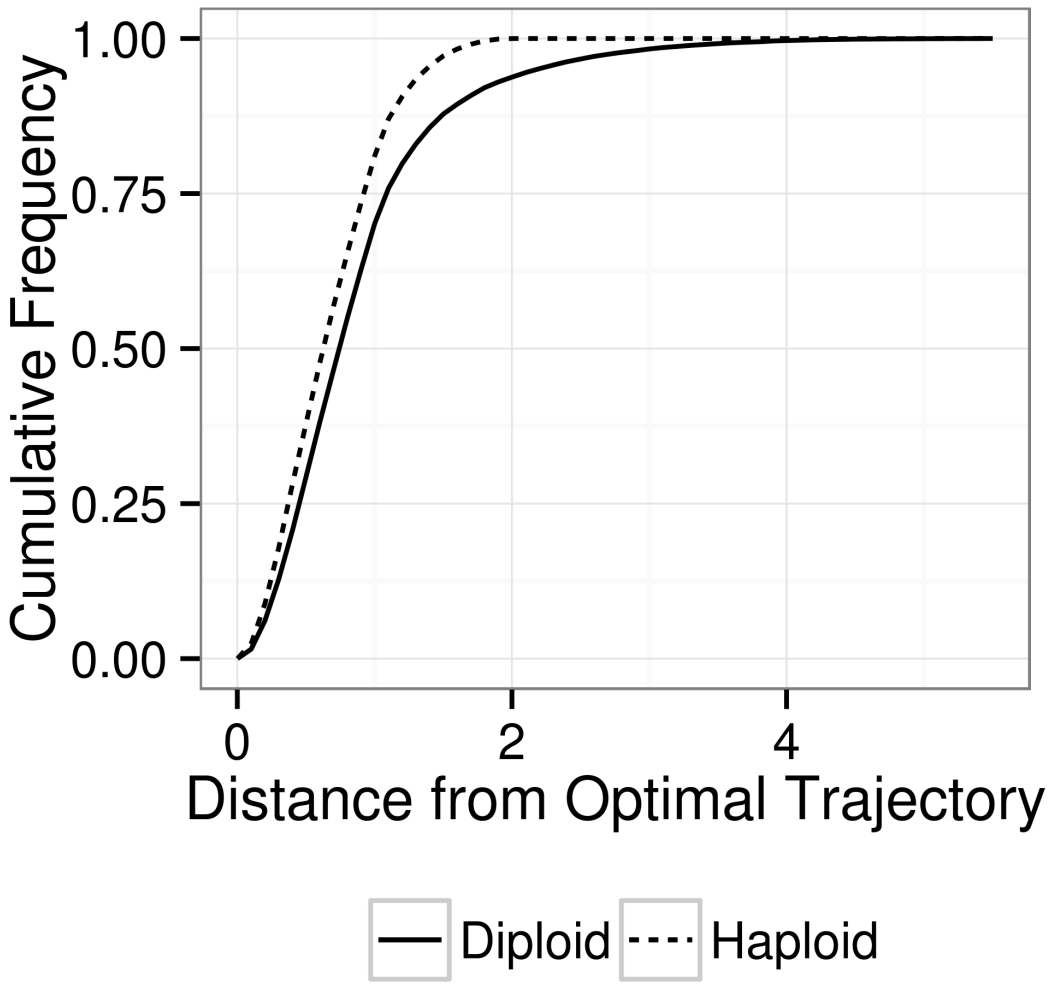

Figure 5. Cumulative distribution of the maximum distance from the optimal trajectory of adaptive walks regardless of number of mutations. This is a measure of the phenotypic optimality of walks. The maximum distance from the optimal trajectory in haploids is significantly less than in diploids (Welch's t test $p<<10^{-10}$).

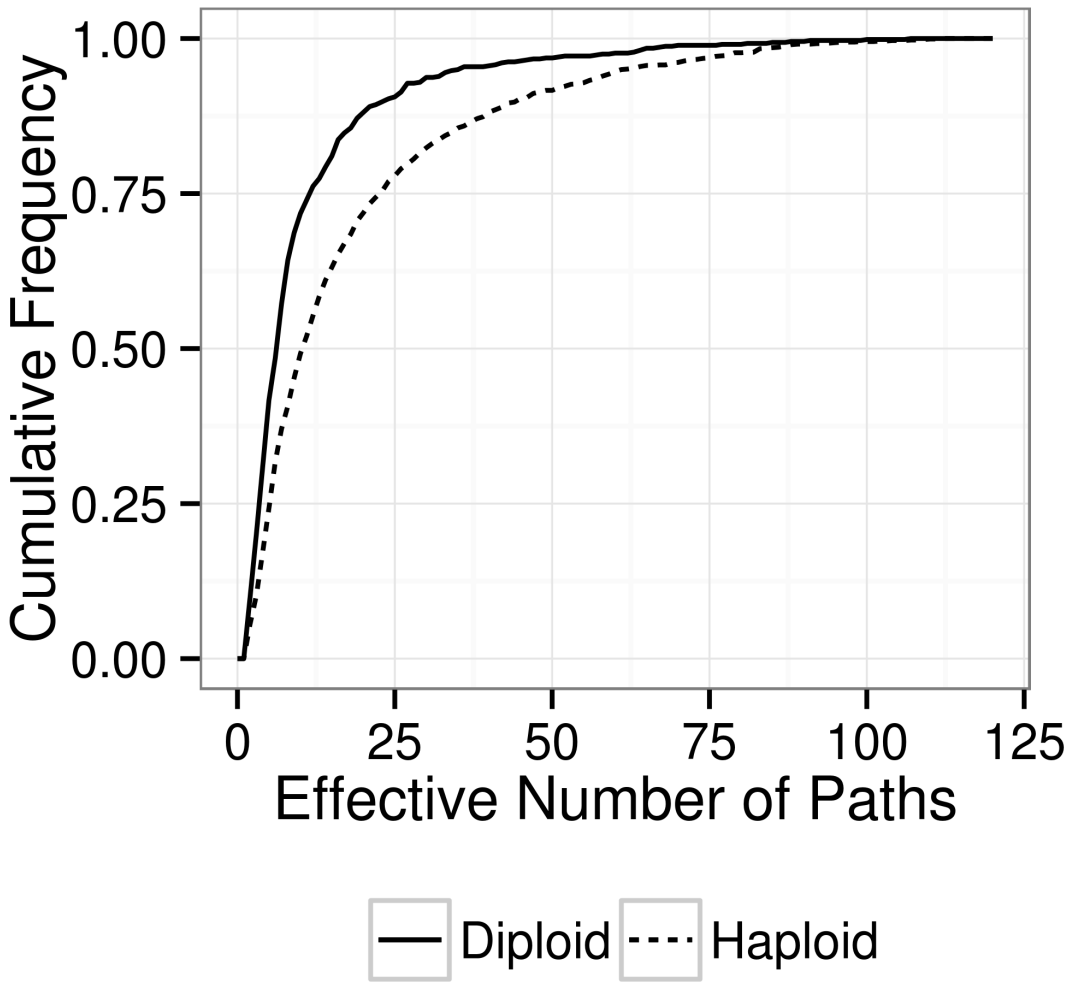

Figure 6. Cumulative distribution of the effective number of paths for adaptive walks with five mutations. This is a metric of mutational repeatability of evolution. Each mutation is introduced into the ancestral background in every possible order, and the number of viable mutational orders, weighted by their probabilities, determines the effective number of paths. The effective number of paths in haploids is significantly greater than in diploids (Welch's t test $p<<10^{-10}$).

Table 1. Pairwise Epistasis (n=5,000)

	Sign	Reciprocal Sign	Ancestral Deleterious
Haploids	52.3%	39.4%	25.1%
Diploids	53.0%	31.3%	5.7%

Table 2. Conditional Mutations

	All	Initial Mutation Not Overdominant	Initial Mutation Overdominant
Haploids (n=9478)	21.4%	NA	NA
Diploids (n=8855)	26.2%	18.2%	62.4%

Table 3. Summary of Backwards Predictability Analysis

	Diploid (n=645)	Haploid (n=1220)
Extra mutations	58.0%	7.5%
Observed Path Most Probable	16.7%	14.1%
Observed Path Top 50% Most Probable	80.8%	69.8%
Observed Path is Inviable	7.1%	0.0%
Final Homozygote Deleterious	0.13%	0.0%