SUPPLEMENTARY METHODS

Simulations: We model adaptive walks in diploid populations with Wright-Fisher simulations using Fisher's geometric model (FGM) as in Sellis et al. (2011). The simulations use code modified from Sellis et al. to allow for more than 2 dimensions. We perform 10,000 replicate simulations with population size $N = 5,000$. Simulations are conducted for 10,000 generations. Complete source code is available at `https://github.com/sunthedeep/Fisher-Geometric-Model`.

In FGM, alleles are represented as a vector in n-dimensional space (Figure 1a). We explore two models, one with two dimensions and one with 25 dimensions. We define the phenotype of a diploid individual as the midpoint of the two vectors of the constituent alleles (Sellis et al., 2011). This amounts to an assumption of phenotypic additivity of alleles, but not necessarily additivity of fitness. The population initially contains a single allele with a distance of 2 units from the optimum, and evolves on a symmetrical Gaussian fitness landscape with single phenotypic optimum at the origin. Fitness is computed using the function:

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

where x is the distance of the individuals phenotype to the optimum. The mutation rate is set to $\mu = 5 * 10^{-6}$ which results in one mutation every 20 generations on average in the population. The angle of the mutation vector is drawn from a spherically uniform distribution. The magnitude of the mutation vector is drawn from an exponential distribution. For the two dimensional regime, the mean of the mutational magnitude is 0.5, while for the 25 dimensional regime, the mean is set to 5. The mutational magnitudes were

chosen to generate sufficient numbers of adaptive walks both with and without overdominant mutations in simulations of both dimensionalities.

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. This is the set of mutations typically available for study in experimental systems. We limit our analysis to studying the first five mutations of each adaptive walk and ignore simulations with fewer than 5 mutations in order to control for the length of the adaptive walk when studying predictability. We partition our five-mutation adaptive walks into those that do and those that do not contain overdominant mutations to study the impact of balanced states on predictability.

Partitioning Walks: Throughout all of our analysis, we have separated walks with and without overdominant mutations. The methodology for this separation is as follows. For each FGM simulation, we have identified the most frequent allele at the end of the simulation, and isolated the first five mutations to occur on this allele. We first determine the time at which the fifth mutation exceeded 5% frequency in the population, which we use as a cutoff for eliminating alleles that have increased in frequency due to drift. All time-points in the simulation after this threshold time are no longer considered for analysis. Throughout the remainder of the simulation, we compute whether the alleles in the population at at least 1% frequency could be maintained as a balanced polymorphism using the method of Kimura (1956). If, at any time before the threshold time, the population contains a set of alleles that could be maintained in a stable polymorphic state, the walk is classified as containing at least one overdominant state.

Forward Predictability Analysis: We calculate the forward predictability of the adaptive trajectory using two metrics. In both of these metrics, we only consider

homozygous phenotypes. Our first metric, maximum pairwise distance, considers pairs of adaptive walks. We compute the maximum of the phenotypic distances between the observed single mutant phenotypes of the two adaptive walks, the double mutant phenotypes, the triple mutant phenotypes etc. Our second metric measures the maximal deviation from the optimal trajectory. For each adaptive walk, we compute the maximal phenotypic distance of any encountered (homozygous) phenotype from the line segment connecting the ancestral phenotype and the optimum.

Backward Predictability Analysis: 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 in the trajectory being generated on the appropriate background and successfully invading the population and reaching equilibrium. The probability of a mutation landing on the appropriate background is proportional to the frequency of the background. For example, if the ancestral allele is balanced with the 1-mutant allele, the probability of generation of a 2-mutant allele is proportional to the frequency of the 1-mutant allele.

The probability of a new allele (generated through mutation) invading and reaching a stable intermediate frequency or fixing from a single copy is calculated empirically through 10,000 Wright-Fisher (Fisher, 1930; Wright, 1931) simulations. These simulations are entirely separate from the FGM simulations used to generate the adaptive walks used throughout this work. In order to compute this probability, one first needs to define the expected equilibrium frequency of a new allele. This is complicated because new alleles can potentially balance with any of the existing alleles or fix in the population. We first ask whether the new allele can balance with any of preexisting alleles by determining whether the fitness of the heterozygous genotype is greater than the fitness of both homozygous

genotypes. If the allele can balance, we compute the equilibrium frequency of the new (derived) allele as

$$\frac{t}{s+t}$$

from Gillespie (2004) when $0 \leq s \leq 1$ and $0 \leq t \leq 1$, where an individual homozygous for the ancestral allele has a relative fitness of $1 - t$, the heterozygote has fitness 1 and the derived homozygote has fitness $1 - s$. If one of the two homozygous genotypes is the most fit state, we determine whether the new allele is either fixed or lost and set the expected equilibrium frequency of the new allele to 1 or 0, respectively. Using this equilibrium frequency, we can compute the mean fitness of the population at equilibrium for this pair of alleles. Through this process, we make a simplifying assumption that balanced states with more than two alleles are unlikely. This process is repeated to compute the mean fitness of every possible balanced state involving the new allele. We choose the state with the highest mean fitness as the equilibrium condition for this new allele. This can be either a balanced state or fixation or loss of this new allele. If the new allele is not present in the computed equilibrium condition with the highest mean fitness, we determine that the new allele cannot invade the population. Otherwise, if the new allele is able to invade the population, we compute the likelihood of reaching this equilibrium condition by asking how frequently the new allele (starting at a single copy) can get to 90% of its equilibrium frequency in our Wright-Fisher simulations.

We are forced to utilize empirical estimations through simulations and not the classical analytic solutions (Haldane, 1927; Kimura, 1962) 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. Our

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

We validate this method of computing the likelihood of a particular adaptive trajectory by testing whether the high probability trajectories are more likely to have been observed in our FGM simulations. We sort all trajectories by their computed probability of occurrence (excluding those with zero probability) and bin them into 40 equally-sized bins. We found that the median of the trajectory probabilities within a bin is significantly positively correlated with the number of trajectories in that bin that were observed in our original FGM simulations (Pearson $r^2 = 0.997$, $p << 10^{-10}$), suggesting that our method is truly capturing the likelihood of a trajectory taking place in the FGM simulations.

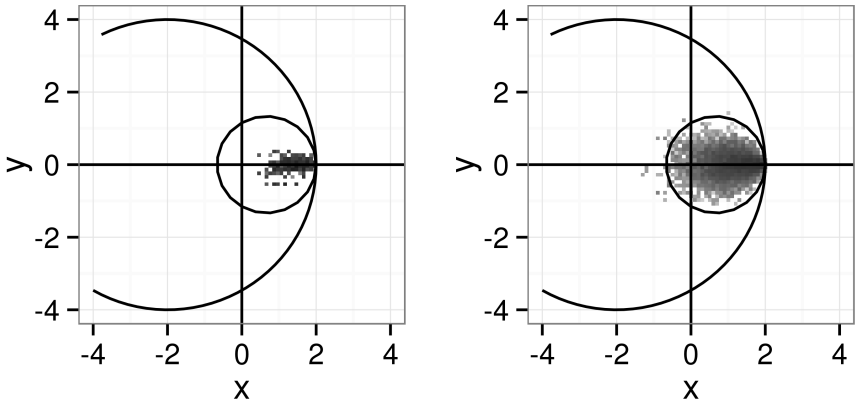
Note that in a traditional haploid model, where each successive mutation fixes in the population, there are $5! = 120$ possible orders of the five mutations observed in the simulated walk to generate the five-mutant allele observed at the end of the simulation. In our diploid model, however, each mutation can occur on any allele in the population where it has not already been introduced. Therefore, the same mutation can be introduced into the population multiple times, but it can occur only once on each allele. The probabilities of all viable mutational orders (where the 5-mutant allele is successfully reached) are then rescaled to add up to 1, to give the probability of a trajectory conditional on reaching the final 5-mutant allele. Mutation orders where the mutations were introduced on the final adapted allele in the same order, but different balanced intermediate alleles were encountered, are not distinguished from each other as these would be indistinguishable in natural systems.

We define the effective number of trajectories as

$$\frac{1}{\sum p^2}$$

where p is the rescaled probability for each viable trajectory in a given five-mutation system. The effective number of trajectories is defined to be 0 when there are no viable trajectories. This is similar to the effective number of alleles in a population (Kimura1964), the predictability metric of Roy (2009) and the entropy metric of Palmer et al. (2013). 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. This provides a single metric of the diversity of mutational orders that are possible and summarizes the backward predictability of the adaptive walk.

127 SUPPLEMENTARY FIGURES



128

129 Figure S1. Phenotypic distribution of homozygous alleles in the first two phenotypic
130 dimensions as in Figure 1b,c for FGM simulations conducted using 25 dimensions and a
131 mean mutational magnitude of 5.

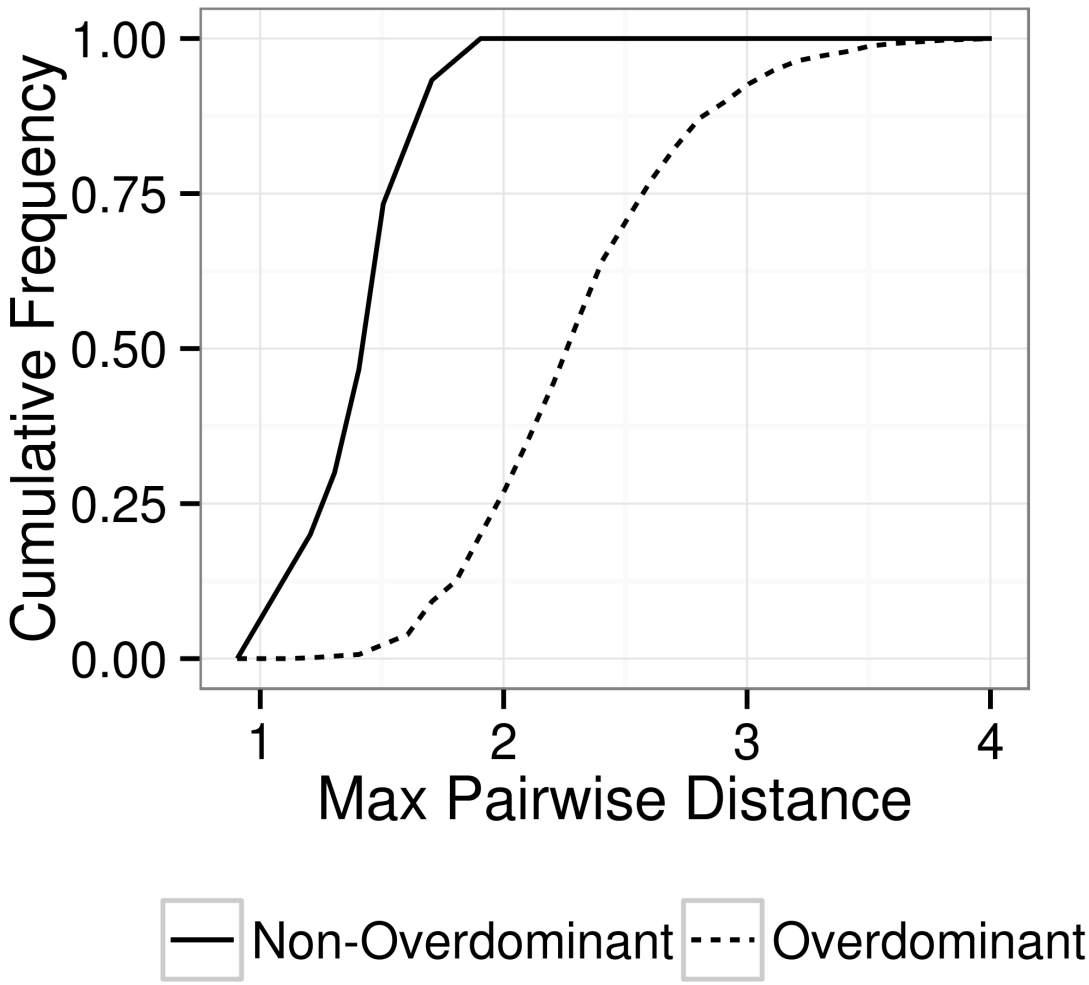


Figure S2. Maximum pairwise distance metric to study forward predictability as in Figure 2 for FGM simulations conducted using 25 dimensions and a mean mutational magnitude of 5. Kolmogorov-Smirnov test $p \ll 10^{-10}$

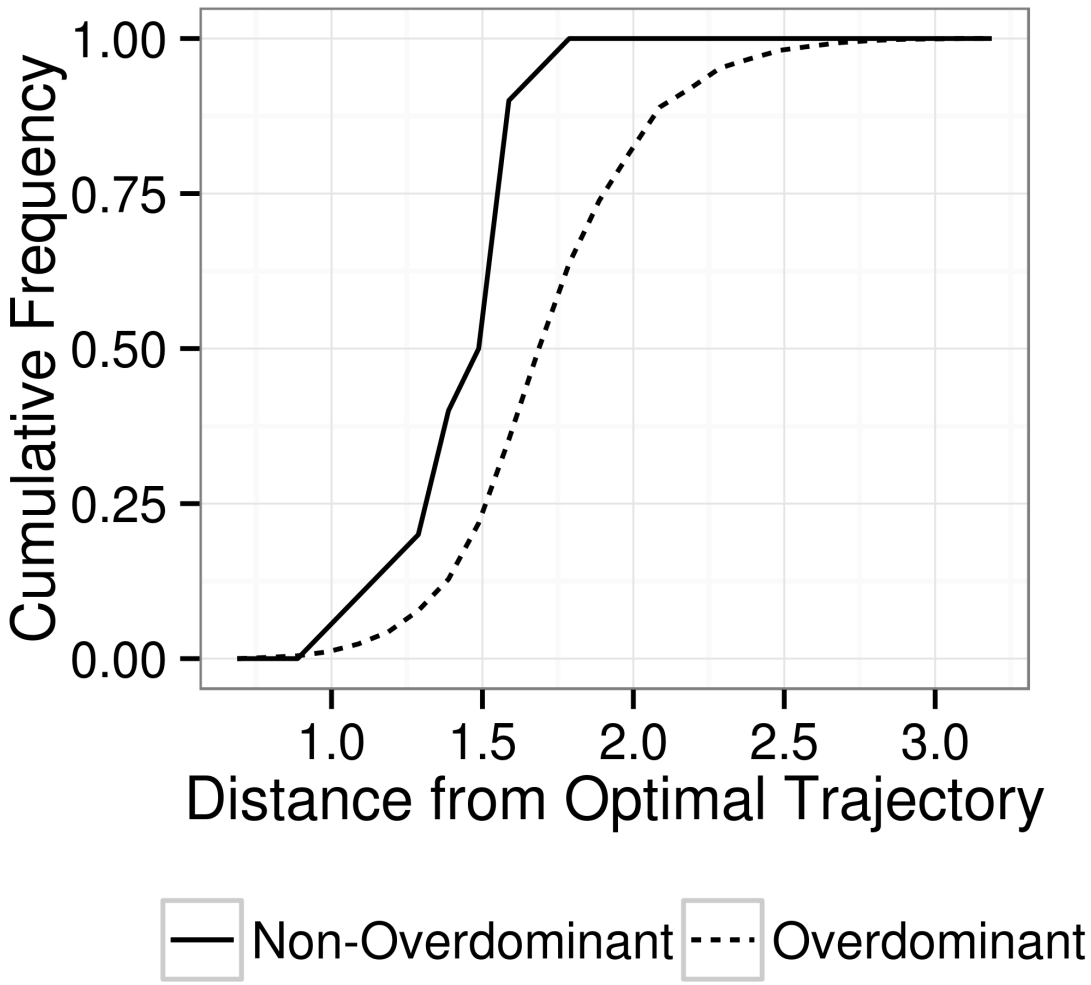


Figure S3. Maximum distance from the optimal trajectory metric to study forward predictability as in Figure 3 for FGM simulations conducted using 25 dimensions and a mean mutational magnitude of 5. Kolmogorov-Smirnov test $p = 10^{-6}$

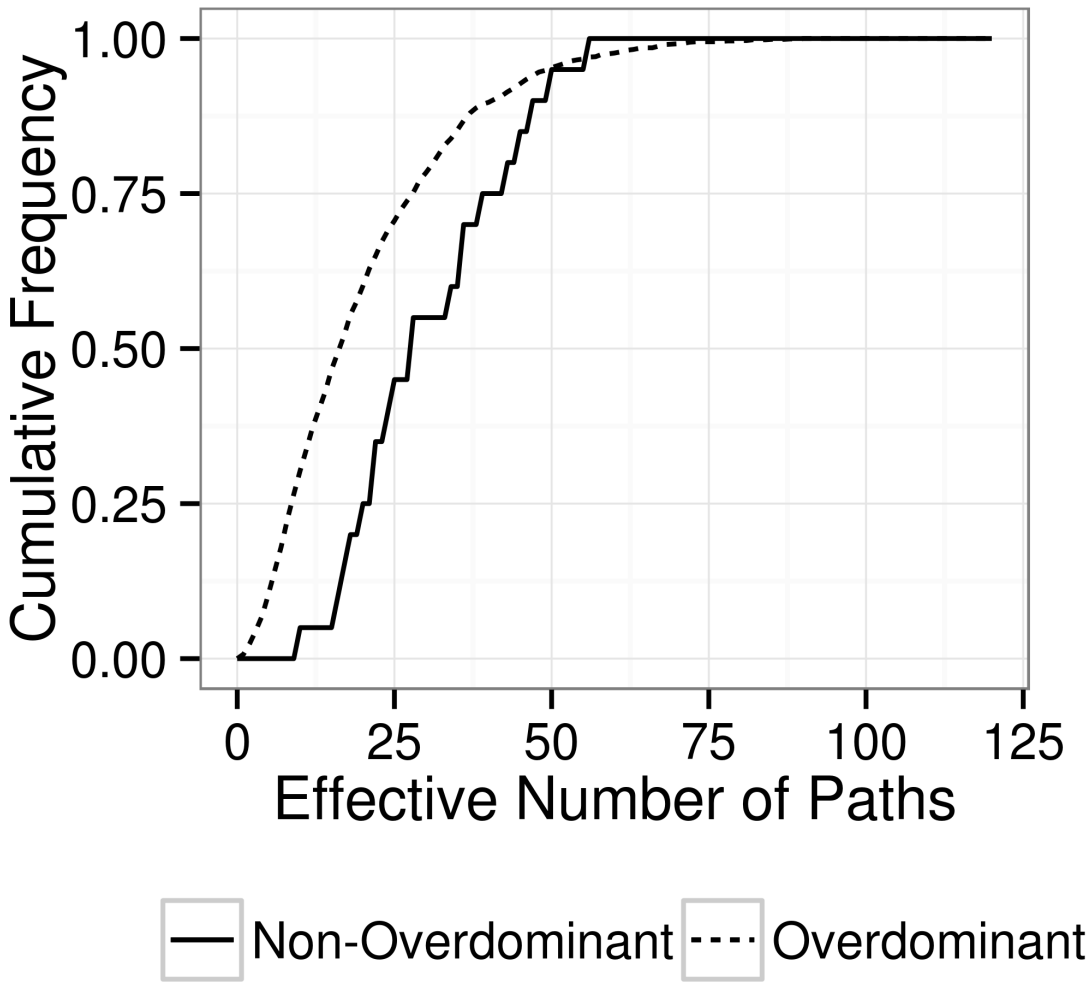


Figure S4. Effective number of paths statistic to study backward predictability as in Figure 4 for FGM simulations conducted using 25 dimensions and a mean mutational magnitude of 5. Kolmogorov-Smirnov test $p = 0.03$

Literature Cited

Fisher, R. (1930). *The genetical theory of natural selection* (1st ed.). Oxford: Oxford at the Clarendon Press.

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

Haldane, J. B. S. (1927, October). A Mathematical Theory of Natural and Artificial Selection, Part V: Selection and Mutation. *Mathematical Proceedings of the Cambridge Philosophical Society 23*(07), 838–844.

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

Kimura, M. (1962). On the probability of fixation of mutant genes in a population. *Genetics* (391).

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

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. *Proceedings of the National Academy of Sciences of the United States of America 2011*, 1–6.

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

Wright, S. (1931, March). Evolution in Mendelian Populations. *Genetics 16*(2), 97–159.