Predictability of evolution depends nonmonotonically on population size

Ivan G. Szendro, Jasper Franke, J. Arjan G. M. de Visser, and Joachim Krug

To gauge the relative importance of contingency and determinism in evolution is a fundamental problem that continues to motivate much theoretical and empirical research. In recent evolution experiments with microbes, this question has been explored by monitoring the repeatability of adaptive changes in replicate populations. Here, we present the results of an extensive computational study of evolutionary predictability based on an experimentally measured eight-locus fitness landscape for the filamentous fungus Aspergillus niger. To quantify predictability, we define entropy measures on observed mutational trajectories and endpoints. In contrast to the common expectation of increasingly deterministic evolution in large populations, we find that these entropies display an initial decrease and a subsequent increase with population size N, governed, respectively, by the scales Nμ and Nμ², corresponding to the supply rates of single and double mutations, where μ denotes the mutation rate. The amplitude of this pattern is determined by μ. We show that these observations are generic by comparing our findings for the experimental fitness landscape to simulations on simple model landscapes.

Evolutionary adaptations arise from an intricate interplay of deterministic selective forces and random reproductive or mutational events, and the relative roles of these two types of influences on the outcome of evolution has been subject to long-standing controversy with significant philosophical implications (1, 2). Although the vision of “replaying the tape of life” on Earth or on some extrasolar planet remains confined to the realm of imagination (3, 4), evolution experiments with microbial populations have begun to address predictability of adaptation on a microevolutionary scale (5–9). In particular, strong signatures of parallel evolution have been observed in the context of the evolution of antibiotic resistance in pathogens, a finding that is of direct relevance to strategies of drug design and deployment (10–14). As lack of knowledge of crucial parameters (e.g., the frequency of beneficial mutations) in such experiments prevents forward predictions, predictability is used in a weaker, a posteriori sense implying repeatability of evolutionary trajectories in replicate populations. For this reason, the two terms will often be used interchangeably in the following (15).

The repeatability of adaptive trajectories is expected to depend on the genetic constraints imposed by epistatic interactions as well as on parameters such as population size N, mutation rate μ, and the typical scale s of selection coefficients (16–18). To be specific, consider a population evolving in the regime of strong selection and weak mutation (SSWM), where mutations are so rare that normally not more than one mutant is present simultaneously and the population can be represented as a single entity that performs an adaptive walk in the space of genotypes (19–21). Such walks are constrained to move uphill in fitness (strong selection, Ns ≫ 1) in single mutational steps (weak mutation, Nμ ≪ 1). As a consequence, a mutualistic pathway connecting two genotypes is selectively accessible in the SSWM regime only if fitness increases in each step (22). A number of recent studies of empirical fitness landscapes have shown that, in most cases, only a small fraction of possible adaptive pathways are accessible in this sense, which implies a dramatic enhancement of evolutionary predictability (10, 11, 15, 23–27). Moreover, the statistical weights of different accessible trajectories often vary widely, further narrowing the range of possibilities to a small number of dominant evolutionary pathways (10, 11, 15). In the SSWM regime, the likelihood of a given trajectory can be quantified straightforwardly in terms of the product of the relative fixation probabilities for individual mutational steps (10).

With increasing N, the simultaneous presence of multiple mutant clones becomes likely and clonal interference sets in (28–31). Clonal interference introduces a bias favoring mutations of large effect (32, 33), thus bringing the dynamics closer to the “greedy” limit, in which the mutation of largest effect is fixed deterministically in each step (34, 35). Although this in itself tends to reduce the heterogeneity of evolutionary trajectories (12, 35–37) and thus increases predictability, it is countered by the increasing availability of genotypes carrying multiple mutations. For sufficiently large populations, the crossing of small fitness valleys (which is completely suppressed in the SSWM limit) becomes relatively facile (38–40), opening up a host of previously inaccessible pathways and leading to a greater degree of randomness in the dynamics. The resulting overall effect on evolutionary repeatability in large populations is hard to assess without detailed analysis, and is expected to depend significantly on the structure of the underlying fitness landscape.

The objective of this article is to explore how the predictability of evolutionary dynamics depends on population parameters, primarily population size and mutation rate, in the presence of realistic epistatic interactions. To this end, we performed extensive simulations of standard asexual population dynamics of Wright–Fisher type on an empirical eight-locus fitness landscape obtained experimentally for the assexual filamentous fungus Aspergillus niger (26). We provide two definitions of adaptive pathways, which can be applied across all evolutionary regimes of interest, and reduce to the familiar adaptive walks in the SSWM regime. Probabilities of pathways and endpoints are then accumulated in a large number of independent runs, and their repeatability is quantified through the entropies of these empirical probability distributions. As usual, high predictability is signaled by low values of the entropies.

Our central result is that the entropies of evolutionary trajectories and endpoints vary nonmonotonically with population size and mutation rate. The variation with population size is governed by the parameters Nμ and Nμ², which describe the supply of single and double mutants, respectively, and it becomes more pronounced with decreasing μ. Simulations on the empirical A. niger fitness landscapes have shown that, in

www.pnas.org/cgi/doi/10.1073/pnas.1213613110
landscape are complemented by a study using a class of model landscapes with tunable roughness (26, 41), which display the same type of behavior.

Results

Path Types and Arrow Plots. There are different ways of defining the path taken by an adapting population in a genotypic fitness landscape, which are generally not equivalent but may yield complementary information. Here, we focus primarily on lines of descent (LODs), which represent the lineages that arrive at the most populated genotype at the final time (see below for a precise definition). Similar definitions of paths have been introduced previously (see, e.g., refs. 42 and 43). In addition, we will make use of the information supplied by the paths defined as the time ordered sets of genotypes that at some time contain the largest subpopulation. We will call such a path the path of the maximum (POM). The POMs have been studied extensively in the context of deterministic mutation-selection models (44). Note that single steps in POMs, in contrast to those in LODs, can connect states that are separated by an arbitrary number of point mutations.

To gain a better understanding of the factors that determine the shape of the paths, we find it convenient to introduce arrow plots representing ensembles of paths realized up to time \( T \) (Fig. 1). Details of the construction are explained in the caption. Note that the choice of the final time \( T \) is, up to a certain point, arbitrary, as the population dynamics will never terminate completely. Here, we generally choose \( T \) such that the population has time to find at least some local fitness maximum. Only for the smallest population sizes, where the dynamics become very slow, do we observe trajectories that do not terminate at a local maximum.

The LODs are obtained as follows: After a fixed time \( T \), the most populated state is determined. The last step of the path is then defined as the connection between this state and the state from which it arose for the last time by mutation. By “arose,” we mean that the target state was unpopulated before the mutation occurred. A given genotype may undergo several episodes of “colonization” and extinction that are stored by the algorithm, and the last episode before the colonization of the final state is used to construct the step. Subsequent steps of the path are constructed analogously, starting from the latest ancestor state determined, i.e., we search for the state from which the latest ancestor arose for the last time before giving rise to the next genotype. The protocol is repeated until the starting point of the simulations is reached.

Note that the paths generated in this way do not include all paths explored by the population, nor all of the paths that contribute to the production of the final state. Rather, they represent the “stepwise fastest” paths through which the final state may have been accessed. The assumption made here is that this path should normally be the one responsible for creating the first mutants on the final state. Once these mutants have been created, selection will dominate the further evolution. Thus, the supply of additional mutants through other paths should play a minor role. If there are several paths with similar probabilities for being the fastest in this sense, they should all show up when the numerical experiment is repeated many times.

The POMs, however, are constructed by keeping track of the most populated genotype at every generation (Fig. S1). Note that maxima do not need to move between adjacent genotypes but can jump to states at Hamming distance larger than unity. Such events have sometimes been referred to as leapfrog events (28) (for experimentally observed examples, see, e.g., refs. 45 and 46). We depict them by wavy lines. By comparing LODs and POMs, we can thus obtain information about whether fitness valleys have been crossed by sequential fixation or by “stochastic tunneling” (38, 39). In the latter case, the deleterious mutation is not fixed, but the population on the deleterious state survives long enough for a secondary mutant of higher fitness to arise. As we will see later, tunneling is negligible as long as \( N \) is small compared with a threshold scaling as \( 1/\mu^2 \), but becomes dominant for larger \( N \).

Population Size Dependence of Typical Paths. Although the main focus of this article is on the statistical analysis of repeatability, it is instructive to first elucidate the effects of population size on the shape of evolutionary trajectories by means of a few typical examples. For this purpose, we refer to Fig. 1, where ensembles of LODs are shown that start from one of the four viable states at Hamming distance \( d = 7 \) from the global optimum (GO) in the \( A. niger \) landscape (see Materials and Methods for details on the landscape, Fig. S1 for the corresponding representation in terms of POMs, and Figs. S2–S4 for LODs starting from different

---

Fig. 1. Arrow plots of LODs obtained from an ensemble of 1,000 runs over \( T = 2^{15} = 32,768 \) generations, starting from genotype 250 of the empirical \( A. niger \) dataset. Population size \( N \) varies from \( 10^5 \) to \( 9 \times 10^5 \) in \( D \), and the mutation rate is \( \mu = 10^{-7} \). The possible states (genotypes) are ordered such that the ascissa provides the Hamming distance, \( d \), of a genotype to the global optimum (GO), which is located at the origin. The Hamming distance is the number of single mutational steps separating two genotypes. States with the same \( d \) are distributed equidistantly along the horizontal direction. The circles represent endpoints of paths; the darker their filling, the more often the corresponding state was the most populated one at final time \( T \). The thicknesses of the arrows are directly proportional to the fraction of times a certain step has been taken when the numerical experiment was repeated. The arrows always point from the ancestor to the descendant. The lines with arrows at both ends mean that the step was taken in both directions, although not necessarily in the same realization. The dashed arrows indicate that the descendant’s fitness is lower than that of the ancestor.
initial genotypes). The figure was generated with a mutation rate of $\mu = 10^{-5}$, and the data were accumulated over 1,000 realizations of the process. Note that, in the *A. niger* landscape, the wild type is the fittest state (the GO) and that, on average, the mutants are less fit than the mutations they have incorporated (26, 47). This is why most observed steps run in the direction of decreasing $d$.

For the population parameters used in Fig. 1A and Fig. S1A, SSWM behavior is expected. The population is mainly monomorphic, i.e., only single mutants appear on the background of the presently dominant genotype and fix with probabilities that are proportional to their selection coefficient but independent of $N$ (19–21). Hence all steps that lead to fitter states are realized with comparable probabilities, provided their fitness values are not too different. At the same time, the fixation probability for deleterious mutations is exponentially small in $N$ (48), making transitions to less fitt states very unlikely (Fig. S5B). This leads to a large number of realized paths and endpoints and highly unpredictable dynamics.

As the population size increases, several nearest neighbor mutants of the currently most populated genotype are present simultaneously, leading to competition between different mutants. Fitter mutations will be more commonly selected. Thus, in this regime, the dynamics becomes greedier and more deterministic (17, 35). The corresponding pronounced increase in predictability results in a dramatic thinning of the graph of adaptive pathways when going from Fig. 1A and Fig. S1A (with $N = 10^4$) to Fig. 1B and Fig. S1B (with $N = 1.5 \times 10^5$). Direct evidence for the increasing greedy nature of the dynamics is provided in Fig. S5A, which shows how the fraction of mutational steps that go to the fittest neighbor grows from $f_{g0} \approx 0.63$ for $N = 2 \times 10^3$ to $f_{g0} \approx 0.88$ for $N = 2 \times 10^5$.

As $N$ is increased further, the number of first step mutants, including deleterious ones, becomes larger and therefore second step mutants are created more frequently. If these second step mutants are sufficiently fit, they can eventually take over the population, effectively tunneling through (38) (or leaping over) the intermediate state. Such events yield a mechanism for crossing fitness valleys that becomes increasingly important for increasing $N$, as can be verified in Fig. 1 and Figs. S1–S4 and S5B. As long as only a few second step mutants are produced by this mechanism, the dynamics becomes again less deterministic, as it depends sensitively on which mutants are randomly created and the number of possible second step mutations is enormous.

Although some indication of this effect can be seen in the comparison between Fig. 1C and Fig. 1D, it is brought out more clearly by the quantitative analysis that we turn to next.

**Entropy Analysis.** When quantifying the degree of determinism of the evolutionary dynamics, it is important to distinguish between the repeatability of endpoints and of the paths taken, as well as between different types of paths. That this distinction matters is easily understood in the context of infinite population sizes. In that limit, the population always finds the GO and this optimal state always takes over the population. Thus, with respect to endpoints, the dynamics becomes totally deterministic. However, in the same limit, all possible paths (in the sense of LODs) to the GO will be taken, and the predictability of LODs should be low for very large $N$. In contrast, the most populated genotype follows a unique path (POM) in the infinite population limit (44).

To study the determinism of the dynamics on more quantitative grounds, it is convenient to define entropies with respect to the endpoints and the paths taken, respectively. The standard choice for the entropy function is $S = -\sum p_i \ln p_i$, where the sum runs over all endpoints (paths) and $p_i$ is the probability to observe a certain endpoint (path). The $p_i$ values are approximated by the fraction of times an endpoint (path) was observed among replicate simulation runs. The entropy is more appropriate to quantify determinism than just counting the number of endpoints or paths observed as it includes information about how often each outcome occurs. Note that the findings to be presented in the following do not depend strongly on the specific choice of the entropy function. Largely equivalent results are obtained for similar observables such as the repeatability measure $\sum p_i^2$ used in refs. 49 and 15.

It is important to notice that the observed ensemble of pathways generally depends strongly on the initial state, as is apparent when comparing the arrow plots in Fig. 1 to Figs. S2–S4 (the corresponding entropies are shown in Fig. S6 A and B). Although this effect is interesting in itself, here we focus on investigating how entropies behave on average when considering ensembles of equivalent starting points. To maximize the number of possible starting points on the *A. niger* landscape, we consider all paths that start at one of the 46 viable genotypes at Hamming distance $d = 4$ from the GO. To illustrate the role of the scales $N\mu$ and $N\mu^2$, we calculate the entropies for a broad range of mutation rates.

In Fig. 2, we plot the average endpoint entropy for starting points at Hamming distance $d = 4$, $(S_d)_{d=4}$, obtained in the following way: First, the entropy was determined separately for each starting point by carrying out 100 independent evolutionary runs up to time $T = 2^{10}$. Subsequently, the entropies were averaged over the different starting points, and the procedure was repeated for different values of $\mu$ and $N$. Apart from the case with the largest mutation rate ($\mu = 10^{-3}$), one observes an initial decrease of the entropy followed by a subsequent rise with increasing $N$ (Fig. 2A). This can be explained by means of the qualitative arguments given in the last section: The initial decrease of the entropy, i.e., increase of determinism of the dynamics, is due to the competition between single mutants causing the dynamics to become greedier, whereas the subsequent increase is a consequence of the increased appearance of double mutants. Fig. S5A shows that the fraction $f_{g0}$ of greedy steps goes through a maximum around the same value of $N$ at which the entropy is minimal.

The initial transition toward greedier dynamics depends on the production of nearest neighbor mutants, the supply rate of which is proportional to $N\mu$. In contrast, the subsequent increase of the entropy due to the appearance of double mutants is linked to their production rate $\sim N\mu^2$. The separation between these two
scales becomes more pronounced the smaller the mutation rate \( \mu \), and correspondingly the minimum value reached by the entropy decreases with decreasing \( \mu \), as is clearly seen in Fig. 2. To make apparent the importance of the scales \( N_f \) and \( N_f' \), in Fig. 2 B and C we plot the entropy as a function of \( N_f \) and \( N_f' \), respectively. The approximate collapse of the decreasing parts of the curves in Fig. 2B and of the increasing parts in Fig. 2C immediately affirms the roles played by the two scales. The lack of an increase of the entropy for the largest mutation rate considered here is most likely due to the lack of clear separation of the two scales.

We also calculated the averaged entropy with respect to the paths, \( \langle S_{\text{fi}} \rangle = \frac{1}{N_f} \sum_{\text{paths}} S_{\text{fi}} \), for the same ensemble of pathways (Fig. S7). Despite the expected distinct behaviors of the two quantities in the limit \( N \to \infty \), we find essentially the same \( N \) dependence as for \( \langle S \rangle = \frac{1}{N_f} \sum_{\text{paths}} S \). This reflects the fact that, even for the largest values of \( N \) and \( \mu \) used here, populations are likely to get trapped at local fitness maxima. New paths that open when increasing \( N \) often lead to formerly unexplored local maxima, which implies that an increase of the number of explored paths is strongly correlated with the number of endpoints.

**Finding the Fittest State.** In the SSWM regime, adaptation proceeds through single mutational steps moving uphill in fitness, and the fixation probability is independent of \( N \) or \( \mu \) (19–21). As a consequence, the statistical weights of different evolutionary trajectories are also independent of \( N \), and the SSWM regime should therefore appear as a plateau at small population sizes in the graphs showing pathway or endpoint entropies as a function of \( N \). The fact that no such plateau is observed in Fig. 2 and Fig. S7 shows that clonal interference already plays an appreciable role in the considered range of parameters, and that smaller populations or mutation rates would be needed to fully realize the SSWM regime.

Other quantities, however, seem to be more robust with respect to a certain level of clonal interference. As an example, we show in Fig. 3 the probability \( P_{\text{GO}} \) for the largest subpopulation to end up on the GO. The figure shows \( P_{\text{GO}} \) as a function of \( N \) averaged over starting points at a given Hamming distance \( d \) from the GO. The averaging is necessary, as the probability strongly depends on the specific starting point (Fig. S6C). For each starting point, 100–1,000 runs were carried out over \( T = 2^{17} \) generations.

The probability of finding the fittest state has a plateau for small \( N \) that coincides rather well with the SSWM value indicated by the horizontal dashed lines. The deviations that are particularly pronounced for \( d = 4 \) and 5 are most likely due to valley crossings that happen with a low probability for small populations but are prohibited within the SSWM approximation. When \( P_{\text{GO}} \) is small, these valley crossings, albeit very rare, may open up additional mutational pathways that are not accessible to SSWM dynamics because they contain at least one fitness decreasing step, thus increasing \( P_{\text{GO}} \) over the SSWM value. With increasing \( N \), clonal interference sets in, making the dynamics more greedy and thus more deterministic. As this implies a decrease in the number of different paths that are explored, the probability to find the GO decreases below the SSWM level. This effect is more pronounced the further away the starting point is from the GO, as the probability for the greedy dynamics to miss the GO by leading the population to a suboptimal local fitness maximum increases. Only when \( N \) is increased further to such large values that double mutants are regularly produced, the dynamics becomes again more stochastic, leading to a higher number of explored paths and thus to a higher \( P_{\text{GO}} \) exceeding the SSWM value. As for the entropy measures discussed previously, the variation of \( P_{\text{GO}} \) with population size is distinctly nonmonotonic.

Apart from the probability for finding the fittest state, it is also of interest to study through which mutational pathways this state is reached. Here, we are particularly interested in the role played by paths along which fitness increases monotonically [monotonically increasing pathways (MIPs)]. These are the only paths that are accessible to adaptation in the SSWM regime and have therefore been at the focus of much recent theoretical and empirical work on fitness landscapes (10, 11, 22–27). What we would like to clarify is whether (or when) such paths are actually the dominating ones when it comes to finding the fittest state, and to what extent they are realized by the dynamics.

To address these questions, we identified all MIPs that start at Hamming distance \( d = 4 \) from the GO, restricting ourselves to direct paths along which the distance to the GO decreases at every step. As a first measure, we computed the fraction of MIPs among all observed LODs that reach the GO, \( f_{\text{MIP/LOD}} \). This quantity was averaged over 100 realizations from each starting point with Hamming distance 4 from the GO from which at least one such path exists (Fig. 4A). One finds that at small \( N \), almost all successful paths are monotonically increasing in fitness. However, as \( N \) increases, \( f_{\text{MIP/LOD}} \) decreases rapidly, showing that the MIPs become increasingly less relevant for adaptation. The dashed line in Fig. 4A represents the ratio of the number of MIPs to the total number of direct paths, equal to \( d! \), averaged over all starting points at \( d = 4 \). Values of \( f_{\text{MIP/LOD}} \) below this line indicate that MIPs are selected even less frequently than would be expected if all direct paths were equally likely.

Furthermore, we have measured the fraction \( f_{\text{MIP}} \) of MIPs that are actually observed within the 100 simulational runs (Fig. 4B). This quantity displays a nonmonotonic dependence on \( N \) that can be explained in similar terms as for the entropies. It should, however, be noticed that, even at small \( N \), less than 70% of all existing MIPs are observed. Thus, the sheer existence of, in principle, easily accessible paths leading to high fitness genotypes does not guarantee that they are actually realized by the dynamics. This can also be concluded from the low values of \( P_{\text{GO}} \) observed in Fig. 3 for small \( N \).

**Comparison with Model Landscapes.** To demonstrate that the results described so far are not caused by the idiosyncrasies of the specific empirical landscape used in this work, we carried out simulations on a family of random model landscapes tuned to reproduce the overall features of the *A. niger* fitness data set. The model we consider is a slight variation of the rough Mount Fuji (RMF) model (26, 27) originally introduced in ref. 41. Within the

---

**Fig. 3.** Probability \( P_{\text{GO}} \) for the largest subpopulation to end up on the GO within \( T = 2^{17} \) generations vs. population size, for all Hamming distances \( d \) for which viable starting genotypes exist. The horizontal dashed lines show the SSWM predictions. The mutation rate was \( \mu = 10^{-7\frac{1}{16}} \).
according to the following:

\[ w_i = -c_1 d_i + c_2 + \xi_i, \]  

where \( d_i \) is the Hamming distance of \( i \) to a reference state whose fitness is set to 1 and will be the GO, \( c_1 \) and \( c_2 \) are constants, and \( \xi_i \) is a Gaussian random variable with mean zero and SD \( \sigma \). The constants \( c_1 \) and \( c_2 \) were obtained from the \( A. \) \( niger \) fitness data as follows: First, we averaged over all fitness values at a given Hamming distance \( d_i \geq 1 \) from the GO, including the states with zero fitness corresponding to nonviable genotypes (26). Then a straight line was fitted to the averaged values plotted against Hamming distance. The slope yields \( c_1 \approx 0.064 \), and for the intercept we obtain the estimate \( c_2 \approx 0.730 \). The variance of the fitness values yields the estimate \( \sigma^2 \approx 0.091 \) for the variance of the \( \xi_i \). Only values \( w_i < 1 \) are accepted, to ensure that the GO is located at the reference genotype. In cases when Eq. 1 yields a negative value, the corresponding fitness is set to zero. In principle the nonviable genotypes in the landscape could be modeled explicitly, e.g., along the lines of ref. 26. However, we prefer to keep the model simple and do not include a separate treatment of nonviable states.

The observed quantitative features have been reproduced by simulations of the RMF model over a broad range of parameters, but here we focus on the specific “\( A. \) \( niger \)” parameter set described above, which optimally matches the empirical landscape (see Fig. S8 for results covering a broader range of parameters). In Fig. 5 and Fig. S9, we plot the quantities obtained from simulations of adaptation on this model landscape. We considered \( \binom{8}{4} = 70 \) different starting points at Hamming distance \( d = 4 \) from the GO, and 100 independent runs from each of them were carried out. For each of the starting points, a new fitness landscape was created using Eq. 1.

Fig. 5 and Fig. S9A show that the entropies obtained for the model landscapes display a similar nonmonotonic dependence on \( N \) and \( \mu \) as the empirical landscape. Again, rescaling \( N \) by \( \mu \) and \( \mu^2 \), respectively, leads to an approximate data collapse of the decreasing and increasing parts of the entropy curves, respectively. Although comparison with Fig. 2 and Fig. S7 reveals that the values of the entropies and the positions of the respective minima are not quantitatively recovered by the model, the qualitative behavior is well reproduced.

Fig. S9B depicts a similar comparison for the probability to reach the state with the highest fitness. Because \( P_{GO} \ll 1 \) in most cases, this quantity is strongly affected by rare events and displays massive fluctuations between different realizations of the RMF model landscape. Averaging over realizations is therefore not appropriate for the comparison with the \( A. \) \( niger \) landscape. Instead, in Fig. S9B, we display data obtained for individual landscape realizations, which show that the overall shape of the variation of \( P_{GO} \) with \( N \) is reproduced by the model. Moreover, the supplementary results in Fig. S8 show that the nonmonotonic variation of the entropy with population size persists whenever the fitness landscape is sufficiently rugged, and disappears only when the limiting case of a smooth, additive landscape is approached.

**Discussion**

The repeatability of evolutionary trajectories in replicate populations is determined jointly by the distribution of the fitness effects of beneficial mutations, by their epistatic interactions, and by the rate at which they appear in the population. Whereas previous work has addressed primarily the first two determinants of evolutionary predictability (14, 18, 22, 26, 49, 50), here we focused on the effect of mutation supply mediated by the population size \( N \) and the mutation rate \( \mu \). By performing simulations on an experimentally measured fitness landscape, we ensured a realistic representation of the distribution of mutational effects and their epistatic interactions.

Our key observation is that, because of the distinct roles played by the supply rate of single (\( \sim N \)) and double (\( \sim N^2 \)) mutations, evolutionary predictability as quantified by the entropy measures \( S_e \) and \( S_p \) varies nonmonotonically with population size. Simulation results for the RMF model suggest that this behavior is generic whenever the underlying fitness landscape is rugged with many local optima, as is often the case for empirically determined fitness landscapes (27, 51). Similar to earlier observations of an evolutionary advantage of small populations in complex fitness landscapes (35–37), the phenomenon depends crucially on the clonal interference among beneficial mutations and cannot be captured within the commonly used SSWM approximation. This also implies that the restriction of evolutionary accessibility to pathways with monotonically increasing fitness (MIPs) assumed in a number of recent studies (10, 22, 24, 26) may be of limited relevance to adaptation.

Although the endpoint entropy \( S_e \) is easier to access experimentally than the path entropy \( S_p \), at least partial information...
about adaptive pathways can be inferred from microbial evolution experiments (12, 13, 45). Parallel evolution has been observed on several occasions, but very few studies explicitly addressed the effect of population size on repeatability. Among those, one found an increase of genotypic diversity with increasing population size (45). Other experiments have addressed the population size dependence of phenotypic diversity on the level of fitness trajectories. One study using Escherichia coli found a pronounced reduction of the variability of fitness trajectories with increasing population size (36), but another study using Aspergillus nidulans found no effect (52). More experimental work under precisely controlled conditions is clearly needed to test the predictions of the present article.

Materials and Methods

**Empirical Fitness Landscape.** The construction of the A. niger strains and the measurement of their fitness values has been explained in detail elsewhere (47). The fitness landscape consists of the wild-type strain and combinations of eight marker mutations: fnwA1 (fawn-colored conidiospores), argH12 (arginine deficiency), pyrA5 (pyrimidine deficiency), leuA1 (leucine deficiency), pheA1 (phenylalanine deficiency), ldsY25 (leucine deficiency), oliC2 (oligomycin resistance), and crnB12 (chlorate resistance). Of the $2^8 = 256$ possible combinations, a total of 16 were found to be viable and assigned nonzero Wrightian fitness (26). Among these, there are four mutants with seven mutations each, i.e., incorporating all but one mutation. In the order in which the genotypes were presented in Table S1 of ref. 26, these are genotypes 250 (all but pyrA5), 251 (all but leuA1), 252 (all but pheA1), and 253 (all but ldsY25).

**Evolutionary Dynamics.** The simulations presented here were performed using standard Wright-Fisher dynamics according to the following algorithm:

i) Draw the number $n_i$ of mutation events in a generation from an exponential distribution with mean $\lambda = N\mu$, where $N$ is the population size, $L$ is the number of loci, and $\mu$ is the mutation rate.

ii) The $n_i$ mutations are distributed among the present mutations with probabilities corresponding to their frequencies. The possibility of individuals accumulating several mutations in a single time step is neglected. Mutations at all loci are chosen with equal probability.

iii) Selection is carried out in two steps. First, frequencies are evolved analytically according to $w_i = \frac{N(w_i^0)}{w_i^0}$, where $w_i$ denotes the frequency of the $i$th state before selection, i.e., at generation $t$, the $w_i$ denote the respective fitnesses, and $\mathbf{w} = \sum w_i$ denotes the mean fitness of the population.

iv) Finally, the frequencies at time step $t + 1$ are obtained by drawing $N$ individuals from a multinomial distribution with probabilities $p_i = w_i / \mathbf{w}$.

**ACKNOWLEDGMENTS.** We thank J.-M. Park for useful discussions. This work was supported by Deutsche Forschungsgemeinschaft within Sonderforschungsbereich 680.
Fig. S1. Arrow plots of POMs obtained for the same ensemble as in Fig. 1. When the maximum of the genotype distribution moves from one state to another, an arrow is drawn accordingly. Again, the thickness of an arrow is chosen directly proportional to the number of times the specific step is taken among the independent runs of the process, and the dashed lines correspond to a step along which fitness decreases. The wavy lines correspond to transitions between states that are separated by more than a single mutation. Population size $N$ varies from $10^2$ for $A$ to $9 \cdot 10^6$ for $D$.

Fig. S2. Arrow plots of LODs obtained from an ensemble of 1,000 runs starting from genotype 251 with mutation rate $\mu = 10^{-5}$ and total evolution time $T = 2^{15}$. Population size $N$ varies from $10^2$ for $A$ to $9 \cdot 10^6$ for $D$. 

Szendro et al. 10.1073/pnas.1213613110
Fig. S3. Arrow plots of LODs obtained from an ensemble of 1,000 runs starting from genotype 252 with mutation rate $\mu = 10^{-5}$ and total evolution time $T = 2^{15}$. Population size $N$ varies from $10^2$ for $A$ to $9 \cdot 10^6$ for $D$.

Fig. S4. Arrow plots of LODs obtained from an ensemble of 1,000 runs starting from genotype 253 with mutation rate $\mu = 10^{-5}$ and total evolution time $T = 2^{15}$. Population size $N$ varies from $10^2$ for $A$ to $9 \cdot 10^6$ for $D$. 

Szendro et al. www.pnas.org/cgi/content/short/1213613110
Fig. S5. (A) The fraction of steps to the fittest neighbor on the paths, $f_{gs}$, and (B) the fraction of steps to less fit neighbors on the paths, $f_{ds}$, are plotted vs. $N$. The figures correspond to $T = 2^{15}$, $\mu = 10^{-5}/16$, and show averages over all starting points with $d = 4$ and 100 runs per starting point. The initial increase of $f_{gs}$ with $N$ can be explained by clonal interference getting more frequent. For larger $N$, $f_{gs}$ decreases, showing that fitter states that do not lie in the direct neighborhood become accessible. Note that the maximum of $f_{gs}$ coincides with the minimum of entropy in Fig. S7, $f_{gs}$ increases monotonically with $N$ indicating that paths crossing less fit states get progressively more accessible. Note that $f_{ds}$ is much smaller than $f_{gs}$ for all $N$ considered here, and, therefore, such tunneling events cannot be the only cause of the decrease of $f_{gs}$. However, one should keep in mind that tunneling through deleterious states is just the most extreme case of eluding the fittest neighbors to access fitter states at larger distances and that the frequency of leaping events over suboptimal, but beneficial, states will increase in a comparable manner.

Fig. S6. Entropies with respect to endpoints (A) and paths (B), as well as the probability for the largest subpopulation to end up on the GO (C), are plotted for the four starting points at $d = 7$ from the GO. Each data point corresponds to 1,000 runs with $\mu = 10^{-5}$ and $T = 2^{15}$. Note the strong dependence on the starting point. Missing data points in C mean that, in 1,000 runs, the largest subpopulation never reached the fittest state.
Fig. S7. Same as Fig. 2 but for entropies with respect to paths.

Fig. S8. This figure explores the properties of eight-locus RMF landscapes over a broad range of parameters. To ensure that the fitness of the least-fit genotype remains positive, the parameters in Eq. 1 are chosen as $c_1 = -s$, $c_2 = 8s + 1$, where $s$ is the mean slope of the fitness function. The Hamming distance $d_i$ in Eq. 1 is measured with respect to a reference genotype, which is likely to become the GO for large $s$ but is otherwise treated as any other state, i.e., its fitness is determined through Eq. 1. For $s = 0$, the landscape reduces to the house of cards (HoC) model where fitness values are independent, identically distributed random variables (1–4), whereas for $s = 0$ and $d > 0$ the landscape is purely additive. All simulations were run over $T = 2^{15}$ generations with mutation rate $\mu = 10^{-5}$ and averaging was carried out over all possible starting points at distance $d = 4$ from the reference genotype, as described in the main text. In A, we fixed $N = 1,024$ and $\sigma = 0.02$ and studied the influence of the slope $s$ on the entropies. For $s / \sigma < C_2$, the two entropies are roughly independent of $s$, indicating that the populations get stuck at some local maxima that they cannot escape within the time $T$. For $s / \sigma > C_2$, the populations may sometimes escape from these maxima and start to explore a larger part of the FL, resulting in an increase of the entropies. However, as the slope is increased even further, the population is able to access the global optimum, which causes the endpoint entropy $\langle S_e \rangle_{d=4}$ to decline. At the same time, there are few or no local maxima, such that all 4! direct paths to the global maximum become accessible and are comparably likely to be realized. Hence one expects the path entropy to saturate at $\langle S_p \rangle_{d=4} = \ln (4!)$ (dashed line), which compares nicely to the numerical data. The other two panels show (B) the endpoint entropy $\langle S_e \rangle_{d=4}$ and (C) the path entropy $\langle S_p \rangle_{d=4}$ as a function of $N$, for the HoC model ($s = 0$, $\sigma = 0.02$), an additive landscape ($s = 1$, $\sigma = 0$), and an intermediate case ($s = 1/32$, $\sigma = 0.02$). For the HoC model, both entropies vary nonmonotonically with $N$ in qualitative agreement with our observations for the experimental landscape. In the additive landscape, allowing for sufficient time, all populations reach the global optimum along any of the 4! possible pathways, which implies that $\langle S_e \rangle_{d=4} \to 0$ and $\langle S_p \rangle_{d=4} = \ln (4!)$ independent of $N$. In the intermediate case with $s = 1/32$ and $\sigma = 0.02$, the path entropy $\langle S_p \rangle_{d=4}$ displays a nonmonotonic behavior similar to but less pronounced than for the HoC model, whereas the endpoint entropy decreases nonmonotonically with increasing $N$. These landscapes are significantly smoother than the A. niger landscape used in main text, with $s / \sigma \sim 1.56$ compared with $s / \sigma \sim 0.21$ for the empirical landscape. Correspondingly, there are few local maxima at which the population can get trapped. Therefore, the onset of valley crossing events makes it more likely for the population to reach the GO, rather than opening up paths to other local maxima, and $\langle S_e \rangle_{d=4}$ continues to decrease.

Fig. S9. (A) Same as Fig. 5, but for entropies with respect to paths. (B) $P_{\text{GO}}$ is plotted vs. $N$ for the experimental and four realizations of the RMF model landscape with $A. niger$ parameters. All curves were obtained with $\mu = 10^{-5}/16$, averaging over all starting points at distance $d = 4$ from the GO and 100 runs per starting point. The model landscapes are constructed as described in the main text. The characteristic shape of $P_{\text{GO}}$ observed for the empirical landscape, with a minimum at intermediate $N$, is also a common feature for the model system. Note, however, that the depth of the minimum of $P_{\text{GO}}$ depends strongly on the specific realization of the landscape and that there are cases where no such minimum is observed.