Supporting Information

Szendro et al. 10.1073/pnas.1213613110

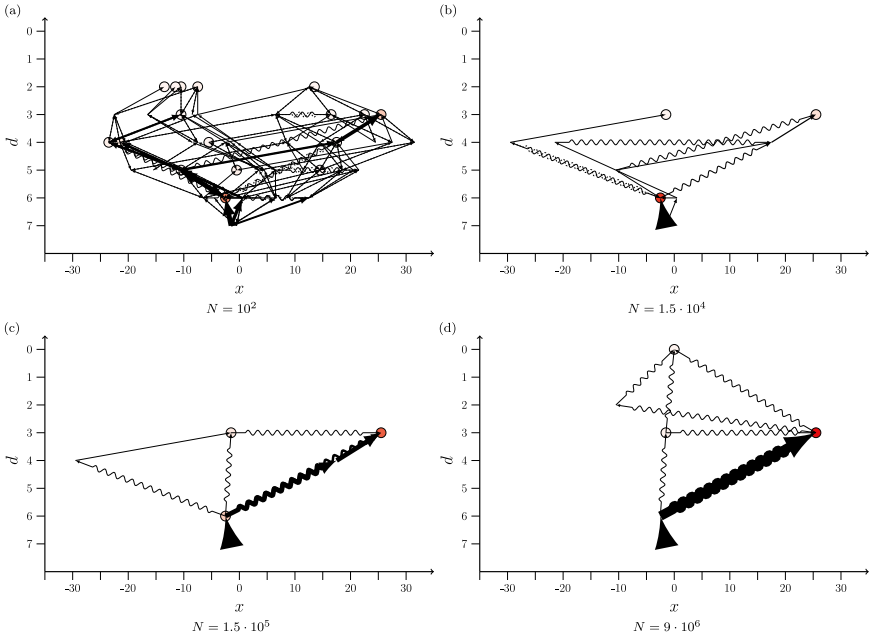

Fig. S1. Arrow plots of POMs obtained for the same ensemble as in Fig. 1. When the maximum of the genotype distribution moves from one state to another, an arrow is drawn accordingly. Again, the thickness of an arrow is chosen directly proportional to the number of times the specific step is taken among 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.

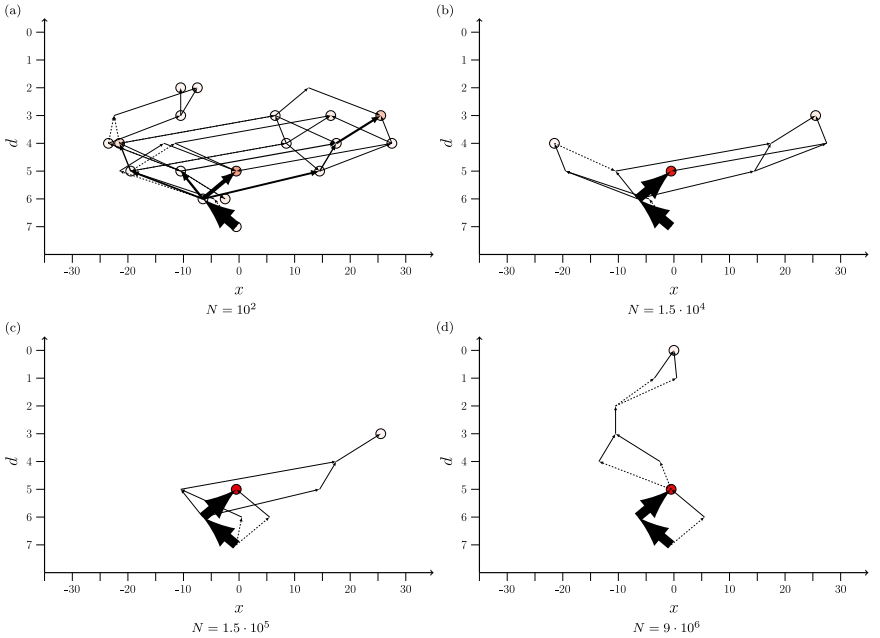

Fig. S2. Arrow plots of LODs obtained from an ensemble of 1,000 runs starting from genotype 251 with mutation rate $\mu = 10^{-5}$ and total evolution time $T = 2^{15}$. Population size N varies from 10^2 for A to $9{\cdot}10^6$ for D.

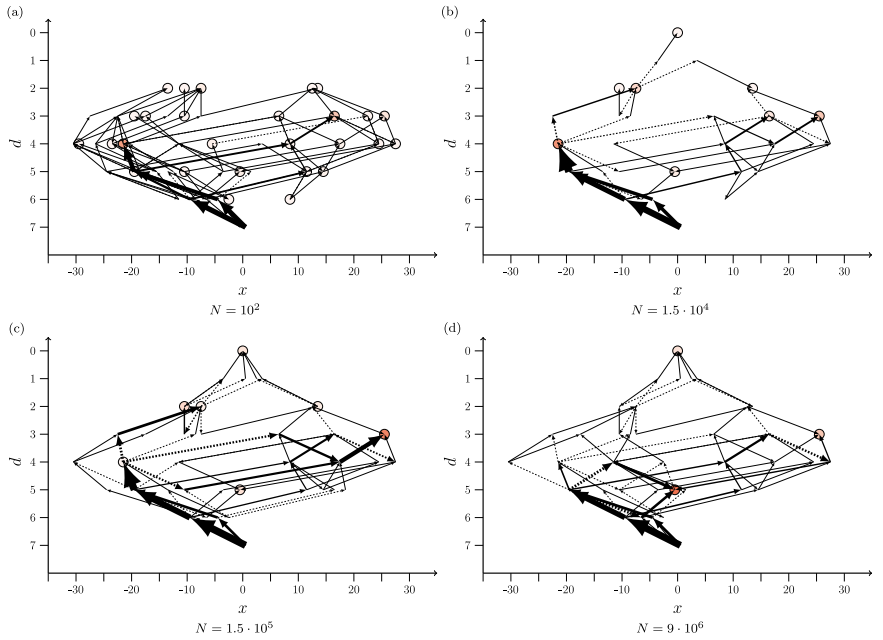

Fig. S3. Arrow plots of LODs obtained from an ensemble of 1,000 runs starting from genotype 252 with mutation rate $\mu = 10^{-5}$ and total evolution time $T = 2^{15}$. Population size N varies from 10^2 for A to $9{\cdot}10^6$ for D.

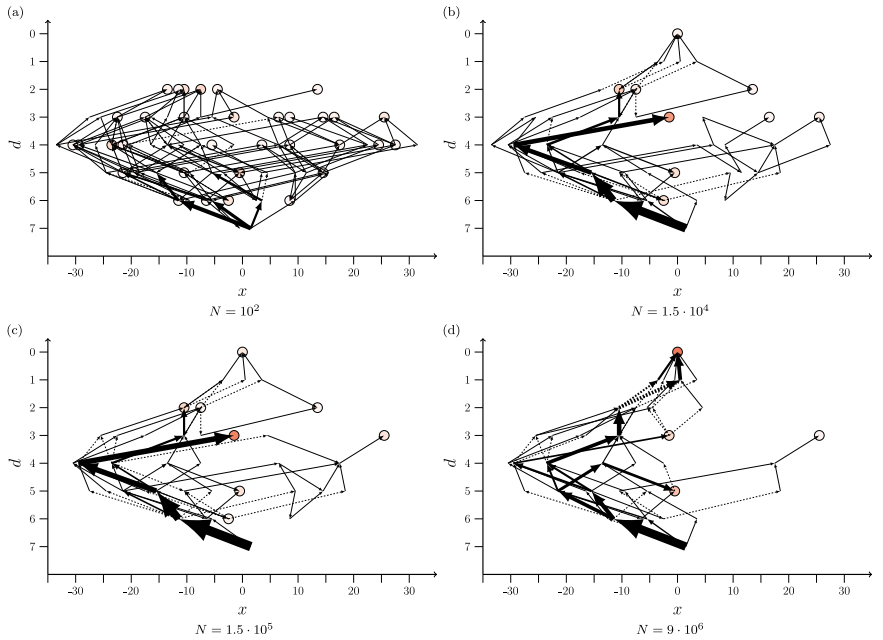

Fig. S4. Arrow plots of LODs obtained from an ensemble of 1,000 runs starting from genotype 253 with mutation rate $\mu = 10^{-5}$ and total evolution time $T = 2^{15}$. Population size N varies from 10^2 for A to $9{\cdot}10^6$ for D.

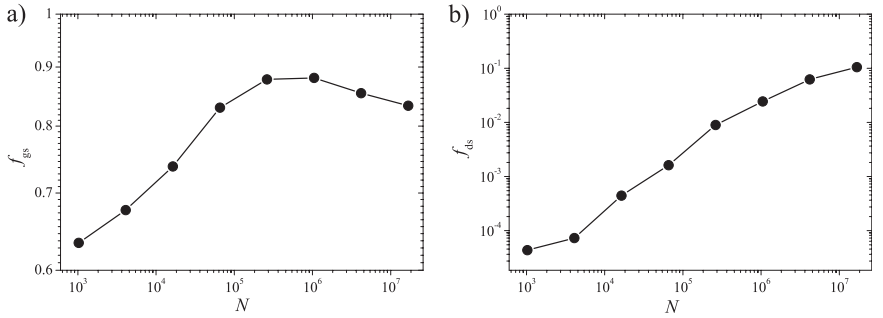

Fig. S5. (*A*) The fraction of steps to the fittest neighbor on the paths, f_{gs}, and (*B*) the fraction of steps to less fit neighbors on the paths, f_{ds}, are plotted vs. *N*. The figures correspond to $T = 2^{15}$, $\mu = 10^{-5}/16$, and show averages over all starting points with $d = 4$ and 100 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_{ds} 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.

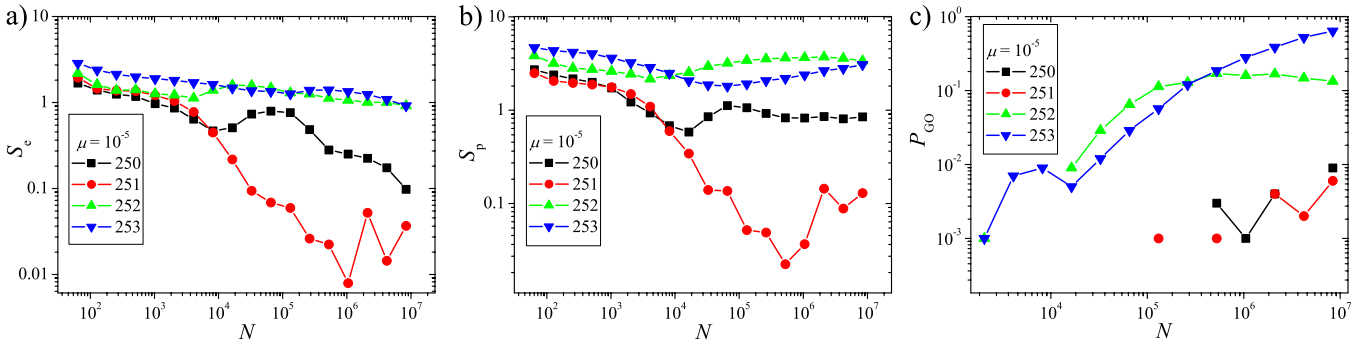

Fig. S6. Entropies with respect to endpoints (*A*) and paths (*B*), as well as the probability for the largest subpopulation to end up on the GO (*C*), are plotted for the four starting points at $d = 7$ from the GO. Each data point corresponds to 1,000 runs with $\mu = 10^{-5}$ and $T = 2^{15}$. Note 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.

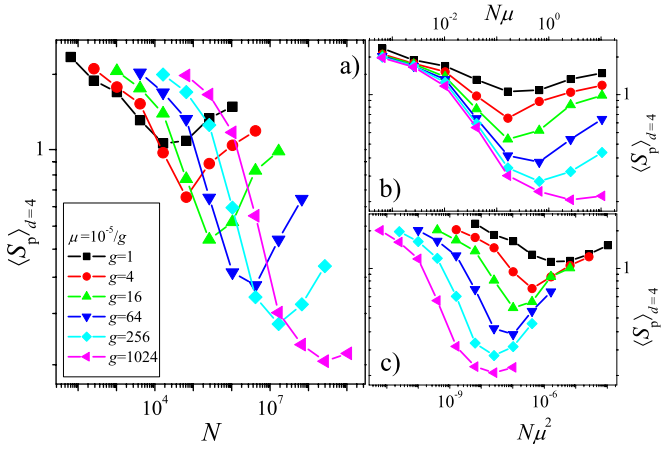

Fig. S7. Same as Fig. 2 but for entropies with respect to paths.

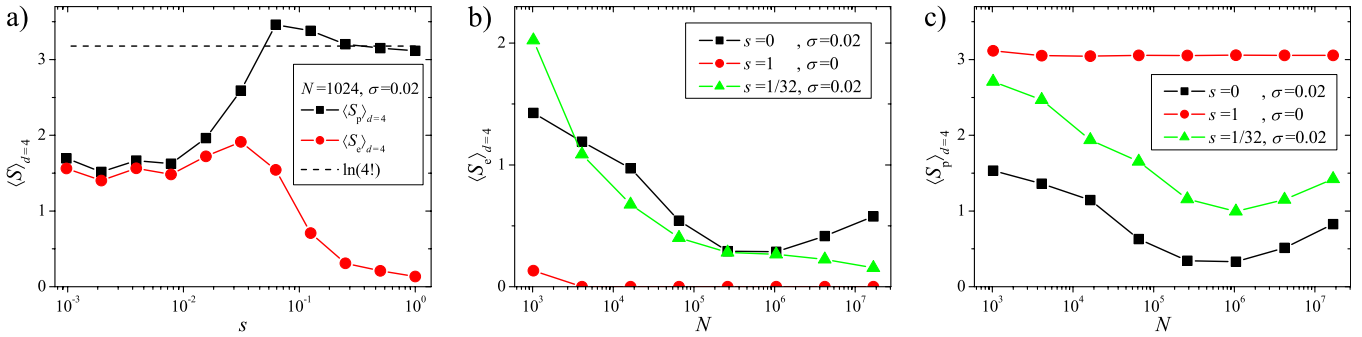

Fig. S8. This figure explores the properties of eight-locus RMF landscapes over a broad range of parameters. To ensure that the fitness of the least-fit genotype remains positive, the parameters in Eq. **1** are chosen as $c_1 = s, c_2 = 8s + \frac{2}{3}$, where s is the mean slope of the fitness function. 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**. Thus, for $s = 0$, the landscape reduces to the house of cards (HoC) model where fitness values are independent, identically distribution random variables (1–4), whereas for $\sigma = 0$ and $s > 0$ 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 $s \ll \sigma$, 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 $s \sim \sigma$, 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 $\langle S_p \rangle_{d=e} = \ln(4!)$ 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 monotonically 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.

1. Jain K, Krug J (2007) Deterministic and stochastic regimes of asexual evolution on rugged fitness landscapes. *Genetics* 175(3):1275–1288.
2. Carneiro M, Hartl DL (2010) Colloquium papers: Adaptive landscapes and protein evolution. *Proc Natl Acad Sci USA* 107(Suppl 1):1747–1751.
3. Franke J, Klözer A, de Visser JAGM, Krug J (2011) Evolutionary accessibility of mutational pathways. *PLoS Comput Biol* 7(8):e1002134.
4. Szendro IG, Schenk MF, Franke J, Krug J, de Visser JAGM (2012) Quantitative analyses of empirical fitness landscapes. arXiv:1202.4378.

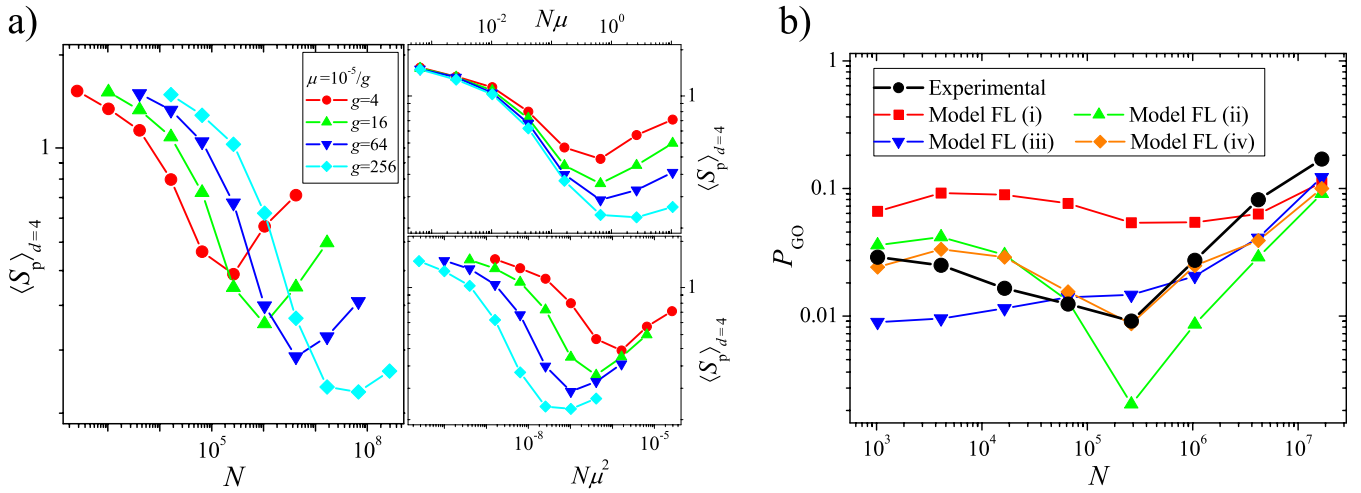

Fig. S9. (*A*) Same as Fig. 5, but for entropies with respect to paths. (*B*) P_{GO} is plotted vs. *N* for the experimental and four realizations of the RMF model landscape with *A. niger* parameters. All curves were obtained with $\mu = 10^{-5}/16$, averaging over all starting points at distance $d = 4$ 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_{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_{GO} depends strongly on the specific realization of the landscape and that there are cases where no such minimum is observed.