RESEARCH ARTICLE SUMMARY

EVOLUTIONARY BIOLOGY

A rugged yet easily navigable fitness landscape

Andrei Papkou, Lucia Garcia-Pastor, José Antonio Escudero, Andreas Wagner*

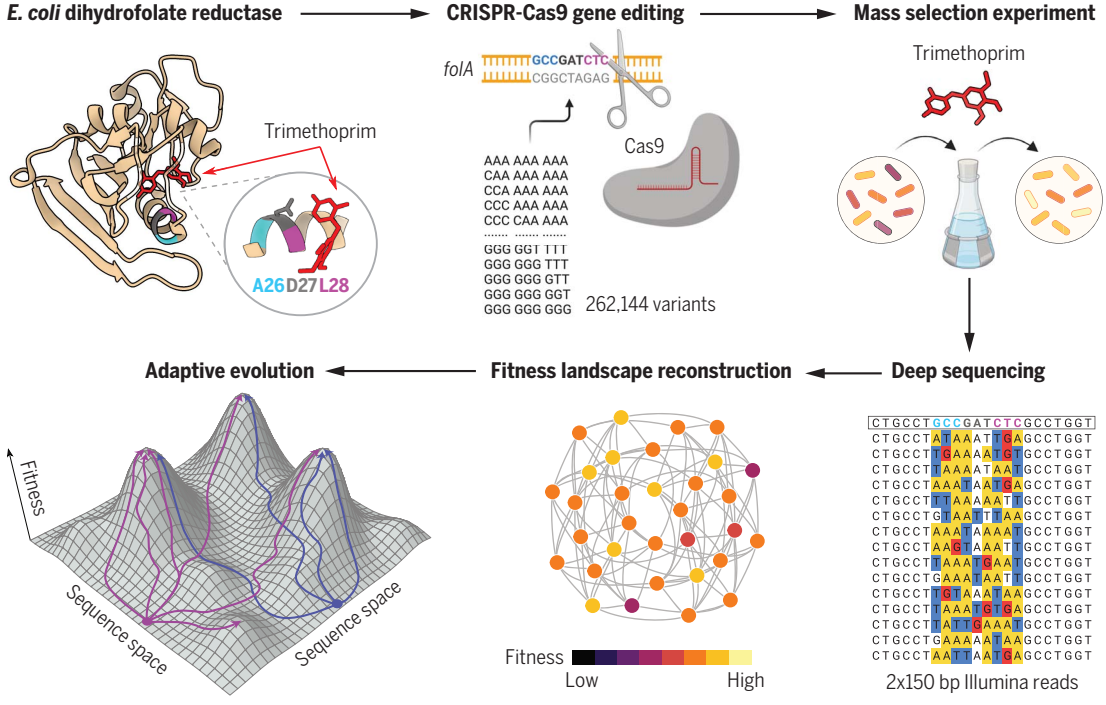
INTRODUCTION: The fitness landscape is a foundational concept in evolutionary biology that has also served to study complex optimization problems in multiple other disciplines. It is an analog to a physical landscape in which a location corresponds to a genotype, and the elevation at that location corresponds to the fitness of an organism with this genotype. Darwinian evolution can be viewed as an exploration of such a landscape by evolving organisms, in which the highest peaks correspond to the best-adapted organisms. When Sewall Wright coined the landscape concept in 1932, he was concerned that biological fitness landscapes may have an astronomical number of peaks, most of which may have low fitness. In such landscapes, evolving populations are likely to become trapped on low fitness peaks from which natural selection cannot help them escape. For almost 80 years after Wright's discovery, virtually all work on landscapes remained theoretical, and even though experimental landscape studies are becoming more frequent now, we still do not know whether rugged landscapes impair adaptive evolution.

RATIONALE: To tackle this fundamental question experimentally, we created a large biological fitness landscape (>260,000 mutants) by CRISPR-Cas9 gene editing of the key *Escherichia coli* metabolic gene *folA*, which encodes dihydrofolate reductase. We mapped the fitness landscape of this enzyme by exhaustively mutating nine nucleotides at three amino acid positions that can confer resistance to the clinical antibiotic trimethoprim. We passaged sixfold replicated mutant libraries of all *folA* variants in an antibiotic-containing environment and used deep sequencing to obtain fitness estimates for nearly 99.7% of all sequence variants. Our nearly combinatorially complete data allowed us to determine the ruggedness of this high-dimensional landscape. We identified its fitness peaks, their basins of attraction, and evolutionarily accessible paths to these peaks. To find out whether landscape ruggedness impairs adaptive evolution, we simulated the evolutionary dynamics on this landscape under various population genetics scenarios.

RESULTS: We found that the landscape is h… ly rugged. It has 514 fitness peaks, most of w… have low fitness. Nonetheless, the landscape has multiple properties of a smooth landscape. These include an abundance of monotonically fitness-increasing paths to high fitness peaks, large basins of attraction of these peaks, and easy reachability of these peaks by >75% of evolving populations. Furthermore, most evolving populations can access multiple high fitness peaks. All 74 high fitness peaks effectively share one enormous basin of attraction (104,496 variants). This leads to low predictability of evolution on the molecular level because each population can take multiple alternative paths that lead to different high fitness peaks. High fitness peaks remain accessible under various evolutionary dynamics on the landscape.

CONCLUSION: Our work shows that adaptive evolution on realistic high-dimensional and rugged fitness landscapes may be easier than commonly thought. Our finding calls for new and improved theory to understand the counterintuitive geometry of realistic high-dimensional fitness landscapes. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: andreas.wagner@ieu.uzh.ch
Cite this article as A. Papkou *et al.*, *Science* **382**, eadh3860 (2023). DOI: 10.1126/science.adh3860



Empirical fitness landscape of dihydrofolate reductase. We edited the *E. coli* genome to create a fitness landscape of all 64^3 codons encoding three consecutive amino acids (A, Ala; D, Asp; L, Leu) of the protein dihydrofolate reductase. We measured the fitness of each genotype in this landscape in the presence of the antibiotic trimethoprim using a mass selection experiment and deep sequencing. Even though the landscape is highly rugged, adaptive evolution can find the highest peaks from most starting locations via short and abundant fitness-increasing paths. [Created with BioRender.com]

RESEARCH ARTICLE

EVOLUTIONARY BIOLOGY

A rugged yet easily navigable fitness landscape

Andrei Papkou[1], Lucia Garcia-Pastor[2], José Antonio Escudero[2], Andreas Wagner[1,3,4]*

Fitness landscape theory predicts that rugged landscapes with multiple peaks impair Darwinian evolution, but experimental evidence is limited. In this study, we used genome editing to map the fitness of >260,000 genotypes of the key metabolic enzyme dihydrofolate reductase in the presence of the antibiotic trimethoprim, which targets this enzyme. The resulting landscape is highly rugged and harbors 514 fitness peaks. However, its highest peaks are accessible to evolving populations via abundant fitness-increasing paths. Different peaks share large basins of attraction that render the outcome of adaptive evolution highly contingent on chance events. Our work shows that ruggedness need not be an obstacle to Darwinian evolution but can reduce its predictability. If true in general, the complexity of optimization problems on realistic landscapes may require reappraisal.

The fitness landscape is a nearly century-old foundational concept in evolutionary biology (*1*). Its influence extends to multiple other disciplines, including ecology (*2*), synthetic biology (*3*), chemistry (*4*), computer science (*5*, *6*), the social sciences (*7*), and engineering (*8*). It is an analogy to a physical landscape, in which individual spatial locations correspond to genotypes, and the elevation at each location corresponds to the genotype's fitness. The best-adapted genotypes occupy the highest peak(s) of such a landscape. A population evolving by natural selection explores such a landscape, and natural selection drives the population uphill to the nearest peak (*1*, *9*, *10*).

A fitness landscape can be single-peaked or multipeaked ("rugged") (*11*). Ruggedness can pose a fundamental challenge to evolution's ability to find a landscape's highest peaks, because a population evolving under the influence of natural selection can only travel on accessible paths through the landscape, that is, paths in which each mutational step increases fitness (*9*, *12*). The reason is that natural selection favors high fitness genotypes and does not allow a population to traverse low fitness valleys between a local peak of intermediate fitness and nearby higher fitness peaks (*9*, *13*).

Theory predicts that in highly rugged landscapes, most evolving populations will become trapped at local peaks of low fitness (*14*). The relationship between landscape ruggedness and peak accessibility has been studied with various theoretical models developed in the 20th century (*15*), such as the *NK* (*16*), "house of cards" (*17*), and "rough Mount Fuji" models (*18*), which may not capture important features of empirical landscapes (*19*, *20*). Because most research on adaptive landscapes remained theoretical until recent decades (*9*, *11*), we still know little about the ruggedness, and even less about the accessibility, of high fitness peaks in empirical fitness landscapes.

Early experimental studies to map empirical landscapes measured the fitness of $<10^3$ genotypes (*21*–*31*). Later studies mapped up to 10^5 genotypes that were generated by random mutagenesis of a reference (wild-type) genotype. The resulting fitness data are dense around the wild type but sparse everywhere else because of missing information on double, triple, and higher-order mutants (*32*, *33*). In other words, such data do not fulfill the important requirement of combinatorial completeness, which means that the fitness of all allele combinations at the mutagenized loci must be known to permit an exhaustive search of evolutionary paths. Some studies achieved combinatorial completeness by quantifying biochemical properties that may be correlated with fitness, such as the binding of biological molecules in vitro (*34*–*38*). Only the most recent works have created large and combinatorially complete fitness landscapes in microorganisms such as *Saccharomyces cerevisiae* and *Escherichia coli*. However, these works focused on other aspects of landscapes, such as the prevalence of higher-order epistasis (*39*, *40*), the molecular principles underlying genotype-phenotype relationships (*41*), the allosteric effects of mutations (*42*), and the evolution of specificity in protein interactions (*43*, *44*).

Here, we address the fundamental question of the relationship between ruggedness and peak accessibility by mapping a large and combinatorially complete in vivo fitness landscape of the *E. coli folA* gene, which encodes the essential metabolic enzyme and antibiotic resistance protein dihydrofolate reductase (DHFR). We find that this landscape is rugged, but its ruggedness does not preclude evolving populations from accessing high fitness peaks.

[1]Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland. [2]Departamento de Sanidad Animal and VISAVET Health Surveillance Centre, Universidad Complutense de Madrid, Madrid, Spain. [3]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [4]The Santa Fe Institute, Santa Fe, NM, USA.
*Corresponding author. Email: andreas.wagner@ieu.uzh.ch

Experimental design and reproducible fitness measurements

We performed CRISPR-Cas9 (*45*) deep mutagenesis to edit the *folA* gene on the bacterial chromosome, randomizing nine nucleotide positions in a part of the gene that is both conserved and implicated in the evolution of antibiotic resistance (Fig. 1A). The result is a combinatorially complete library of almost 4^9 (262,144) DNA genotypes. They include all possible (64^3) combinations of codons that encode three successive amino acids of DHFR (wild-type sequence: 26A-27D-28L) (fig. S1). Missense mutations at these positions provide high resistance to the clinically important antibiotic trimethoprim, which inhibits DHFR. We selected these three positions because they frequently acquire resistance mutations in experimental evolution (*46*, *47*) and because their proximity to each other facilitates gene editing.

We exposed a population of *E. coli* cells expressing this library to a sublethal dose of trimethoprim and measured the fitness of library members through deep sequencing in a sixfold replicated mass-selection experiment. The resulting data comprise *folA* variant frequencies before and after selection for 99.7% (261,382/262,144) of all possible variants (*48*). Variant frequencies were highly consistent between replicates (Pearson's pairwise correlation coefficient r between replicates: $0.946 \leq r \leq 0.999$; fig. S2). We used these frequencies to calculate the fitness of all DHFR variants relative to the wild type. Population genetic theory shows (*49*) that it is best to represent all fitness values on a natural logarithmic scale (*48*). On this scale, the fitness of the wild type has a value of 0, and that of a variant with relative fitness 1 corresponds to an exponential growth rate that is $e^1 \approx 2.718$ times higher.

The fitness values we measured are highly consistent with values obtained in a smaller independent experiment ($N = 250$ variants, Pearson's $r = 0.972$, $P = 2 \times 10^{-158}$; fig. S3, A and B). Additional experiments confirmed that the genetic background of our *E. coli* strain did not substantially alter the relative fitness of our DHFR variants (fig. S3, C and D) (*48*). Finally, to validate our fitness estimates, we isolated 30 DHFR variants and measured their growth rate and resistance to trimethoprim in single cultures expressing individual DHFR variants. We found that relative fitness was highly correlated with the growth rate and resistance observed in single cultures (figs. S3 to S5; Pearson's $r = 0.993$, $P = 2.3 \times 10^{-27}$ for growth rate; Pearson's $r = 0.987$, $P = 1.1 \times 10^{-23}$ for resistance).

Functional DHFR variants vary widely in fitness and contain few amino acids at a key position

Most DHFR variants (93%, 243,303/261,332) have very low fitness (Fig. 1B). The distribution of their fitness values is consistent with that of variants with two stop codons (fig. S6).

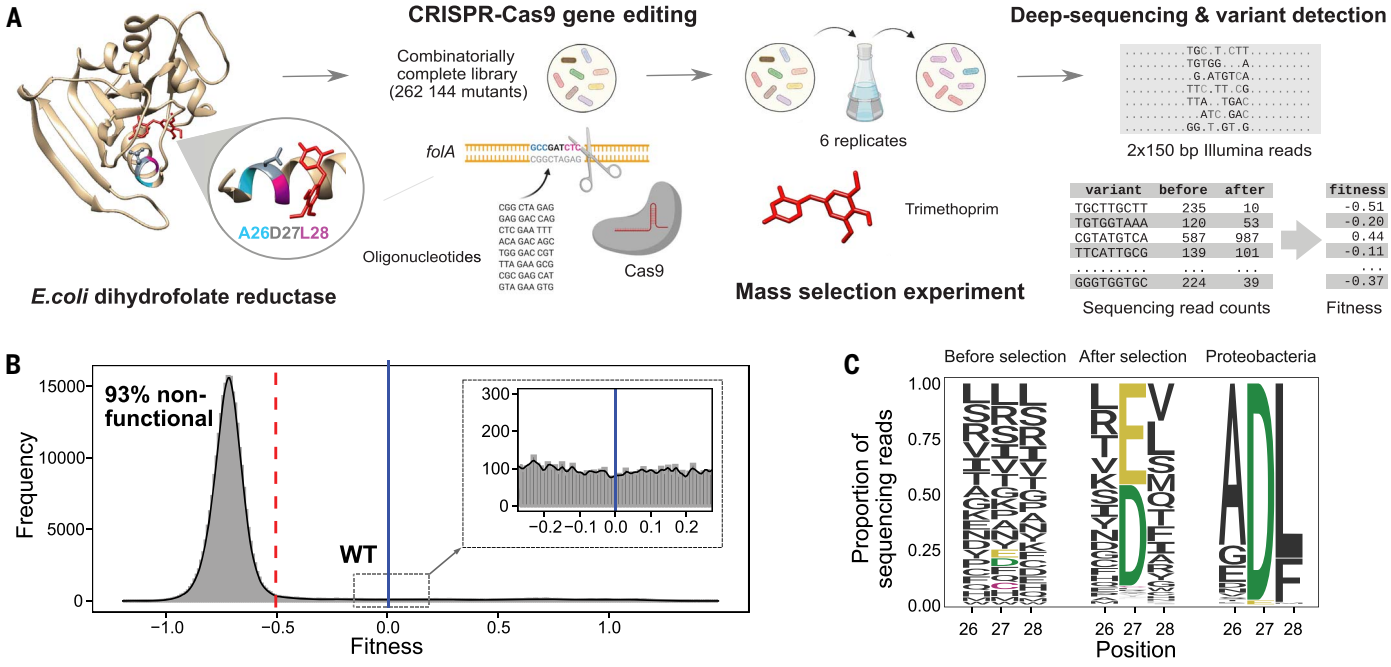


Fig. 1. Creation of combinatorially complete library and distribution of fitness effects. (**A**) Experimental design. We used CRISPR-Cas9 gene editing to create a library of DHFR variants. We targeted a 9-nt segment of the *folA* gene on the *E. coli* chromosome, which encodes three amino acids of DHFR (A26-D27-L28; A, Ala; D, Asp; L, Leu). In the folded protein, these three amino acids lie inside the substrate binding pocket (Protein Data Bank ID6XG5). We edited the segment with degenerate oligonucleotides to obtain a library comprising 99.7% of all theoretically possible DHFR genotypes for this segment. We used this library to perform a mass selection experiment by growing six parallel cultures at the half-inhibitory concentration of trimethoprim (0.4 μg/ml). We amplified the variable DHFR region and performed Illumina paired-end deep sequencing. We determined sequencing read counts for each variant before and after selection and used them to calculate relative fitness. [Created with BioRender.com] (**B**) The distribution of fitness effects in the library. The dashed red line indicates the cutoff value for nonfunctional variants (−0.508), the blue solid line marks the fitness of the wild-type variant, and the insets show the tail of the distribution near the wild type. N = 261,332 variants. The inset indicates that the distribution has a heavy tail near the wild type. (**C**) The frequency of amino acids at positions 26, 27, and 28. The panels show the frequency of amino acids in the library before selection (left panel) and after selection (middle panel). The right panel shows the corresponding amino acid frequencies in DHFRs from proteobacteria, based on an alignment of 5000 orthologous *folA* sequences.

Because DHFR catalytic activity bears a nearly linear relationship with *E. coli* fitness (*50*, *51*), we conclude that DHFR is inactive in these variants. We thus refer to these variants as nonfunctional, to distinguish them from the remaining 18,029 functional variants. The functional variants show highly variable fitness. Among them are high fitness variants identical to several previously characterized *folA* mutants (*52*) with clinically relevant trimethoprim resistance levels that exceed wild-type resistance by two orders of magnitude (fig. S7). Almost all functional variants have a negatively charged aspartic or glutamic acid at position 27 (Asp^{27} or Glu^{27}; fig. S8A). This amino acid position is highly conserved in DHFR (*53*) and highly prevalent in an alignment of 5000 orthologous DHFR proteins from proteobacteria (Fig. 1C). Consistent with a previous study that reported a functional mutant with a cysteine at position 27 (Cys^{27}) (*54*), we identified many functional Cys^{27} variants, even though the Cys^{27} allele is absent from the 5000 proteobacterial sequences we studied. Despite the importance of position 27, it does not solely determine the sequence-function relationship, which depends on interactions between different positions. For example, Asp^{27} and Glu^{27} alleles confer high fitness in combination with different sets of amino acids at positions 26 and 28 (fig. S9).

The fitness landscape is rugged

To study the DHFR fitness landscape, we represented our data as a network or graph, in which each node (vertex) represents a DHFR variant and is associated with the variant's fitness. Any two DHFR variants that differ at one nucleotide position are immediate (one-mutant) neighbors connected by an edge, which represents a single mutational step. An evolutionary path through this network consists of several consecutive mutational steps. We restricted such paths to functional DHFR variants by removing all nonfunctional variants whose neighbors comprise only other nonfunctional variants. In contrast, we retained nonfunctional variants that have at least one functional neighbor in the network, reasoning that rare mutations may create a functional variant from a nonfunctional neighbor.

Fifty-two percent of DHFR variants (135,178) in this network are contained in the largest connected subgraph [or "giant component" (*55*)], which constitutes the fitness landscape we analyze. Even though functional variants make up only 13% of genotypes (18,019/135,178) of this landscape, they form a densely connected part of the landscape (fig. S8, B and C), with as many functional-to-functional edges (50%, 161,015/324,044) as nonfunctional-to-functional edges. Almost 95% of functional variants have the maximal possible number of 27 (9 × 3) one-mutant neighbors, whereas nonfunctional variants have only one neighbor on average (fig. S8, B and C).

Next, we quantified our principal indicator of ruggedness, the number of peaks (*9*, *11*, *14*), that is, the number of DHFR variants that have higher fitness than all their one-mutant neighbors. We found that the landscape has 514 peaks and is thus rugged (Fig. 2A). To compare, a maximally rugged (uncorrelated) random *NK* landscape would contain a similar number of peaks (*14*) (see supplementary text section S1). Most of these peaks (408/514) have low fitness, meaning that they are less fit than the wild type (fitness < 0). Thirty-three peaks have intermediate fitness (from 0 to 1) and are enriched with

Cys^{27} variants. The remaining 73 peaks have high fitness (>1) and consist exclusively of Asp^{27} and Glu^{27} variants (Fig. 2, A and B). Only one Asp^{27} peak has a fitness of <1, which equals 0.87. For simplicity, we will refer to all Asp^{27} and Glu^{27} peaks (including this peak) as high fitness peaks.

Because the location of different peaks in a landscape may affect their evolutionary accessibility, we sought to determine whether peaks are close together in the landscape. We did so by computing the genetic (nucleotide) distance between all pairs of the 514 peaks, that is, the minimal number of single-nucleotide changes that are needed to convert one peak into the other. We compared the resulting distance distribution with that of all pairs of 514 variants chosen at random from the landscape. Their mean distances are very similar and lie within 1% percent of each other (d = 6.67 for peaks versus d = 6.61 for random variants, two-sided Kolmogorov-Smirnov test D = 0.03, $P < 10^{-308}$, N = 131,841; Fig. 2C). This pattern also persists if we consider amino acid distances instead of nucleotide distances (Fig. 2D). Most importantly, the pattern extends to the high fitness peaks (i.e., Glu^{27} and Asp^{27}; fig. S10). In sum, the DHFR landscape has many fitness peaks that are scattered across the landscape (fig. S11).

Fitness peaks are highly accessible

Considering the ruggedness of the landscape, one might expect that any one peak may only be accessible from a small fraction of variants. To find out, we first examined for each variant and peak whether accessible paths exist from the variant to the peak. Such paths consist only of beneficial (fitness-increasing) mutational steps, because fundamental population genetic principles dictate that weakly deleterious mutations are unlikely to go to fixation in an organism such as *E. coli*, with its large effective population size of 10^8 individuals (*56*).

More specifically, we determined the size of each peak's basin of attraction—the total number of variants from which the peak is accessible. The size of this basin varies considerably depending on peak fitness and which amino acid is present at the critical position 27 (Fig. 3A). In our dataset, low fitness peaks generally have small basins, with a median size of only 28 variants (0.02% of all variants). Intermediate fitness peaks (Cys^{27}) have larger basins, with a median size of 7667 variants (5.7%). Most notably, high fitness peaks (with Glu^{27} and Asp^{27}) have very large basins, whose median size comprised the majority (69% or 93,597) of variants. In general, peaks with higher fitness had significantly larger basins of attraction (Spearman's $\rho = 0.61$, $P = 6.8 \times 10^{-55}$, N = 514; fig. S12).

The large basin size of the Asp^{27} or Glu^{27} peaks indicates their accessibility. However, even though accessible paths to any one high fitness peak may exist, they may be few compared with the total number of paths. To exclude this possibility, we focused on all variants within a peak's basin of attraction and enumerated all shortest paths (accessible and inaccessible) from each variant to the peak. Not unexpectedly, the number of total paths increases exponentially with a variant's distance from a peak (Fig. 3B). Among these paths, the fraction of accessible paths is very high at modest distances to a peak but decreases with increasing path length (Fig. 3C). For example, at two, three, four, and five mutational steps away from a peak, 86, 62, 39, and 21% of all paths are accessible, respectively. At the distance of nine mutational steps, only 1% of paths remain accessible.

In a perfectly smooth landscape, the length of the shortest accessible path from any one variant to a high fitness peak equals the genetic distance between the variant and the peak. In contrast, in a rugged landscape, even the shortest accessible path may meander through the landscape and thus be much longer than this genetic distance (*52*). However, despite our landscape's ruggedness, this is not the case. Specifically, the mean length of the shortest accessible path between a variant and a high fitness peak (mean ± SD = 6.65 ± 1.9 mutations) is less than one mutation longer than in a smooth landscape (6.06 ± 1.22 mutations, two-sided Kolmogorov-Smirnov test D = 0.15, $P < 10^{-308}$, N = 6,748,190) (Fig. 3D). Thus, high fitness peaks have large basins of attraction containing many short and accessible paths.

Adaptive evolution can easily reach high fitness peaks via short paths

Even though high fitness peaks appear highly accessible, their proportion (14%) is small compared with the majority of low fitness peaks. Would selection drive most evolving populations to one of the many (408) low fitness peaks? To find out, we first simulated adaptive evolution in the strong-selection weak-mutation regime (*48*, *57*). This choice is motivated by the extremely low mutation rate for the 9-nt (nucleotide) mutational target in the *E. coli* genome [(9 positions) × (2.2×10^{-10} mutations per position per generation)] (*58*). In this regime, adaptive evolution effectively becomes an adaptive walk (*48*).

We simulated 10^6 such walks, each starting from a randomly chosen DHFR variant (functional or nonfunctional). For all immediate neighbors that are only one mutational step away from this variant, we then calculated the fixation probability of the corresponding mutation. To this end, we used a well-established expression derived by Kimura for the probability that a new mutation sweeps through a population and becomes fixed (*48*, *59*). This expression takes into account the fitness difference of the mutant to the starting variant, as well as the influence of genetic drift, whose strength falls with increasing effective population size. Because *E. coli* has very large populations

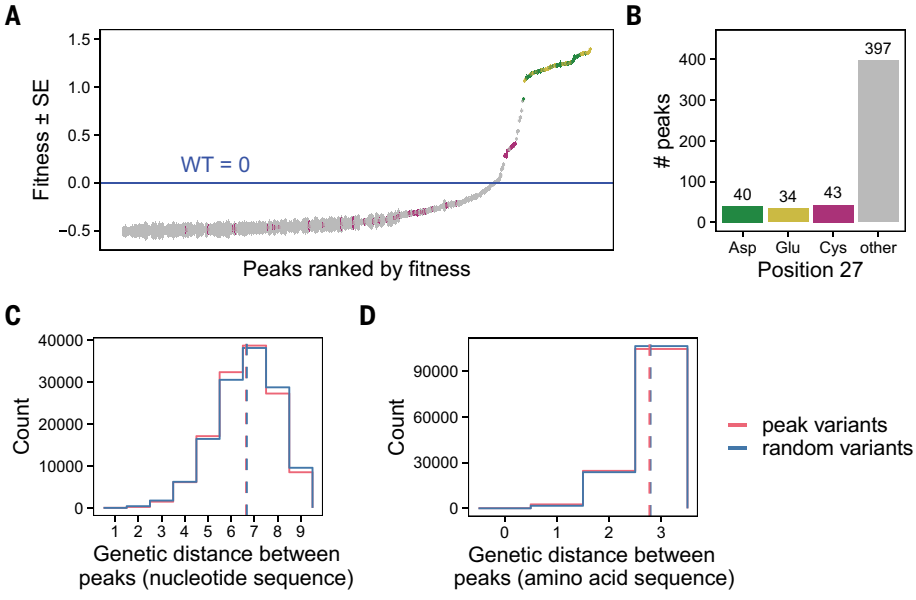


Fig. 2. Fitness peaks. (**A**) The distribution of fitness estimates and corresponding standard errors for 514 variants identified as peaks in the landscape. The solid blue line at y = 0 shows the fitness of the wild type. Colors indicate amino acid identity at position 27. (**B**) Frequencies of amino acids at position 27 for fitness peaks (N = 514). (**C**) Distribution of pairwise genetic distances between nucleotide sequences of 514 peaks (red) and between sequences of 514 randomly chosen variants (blue). The vertical lines show the means of the corresponding distributions. (**D**) As in panel (C) but based on amino acid sequences to calculate genetic distance.

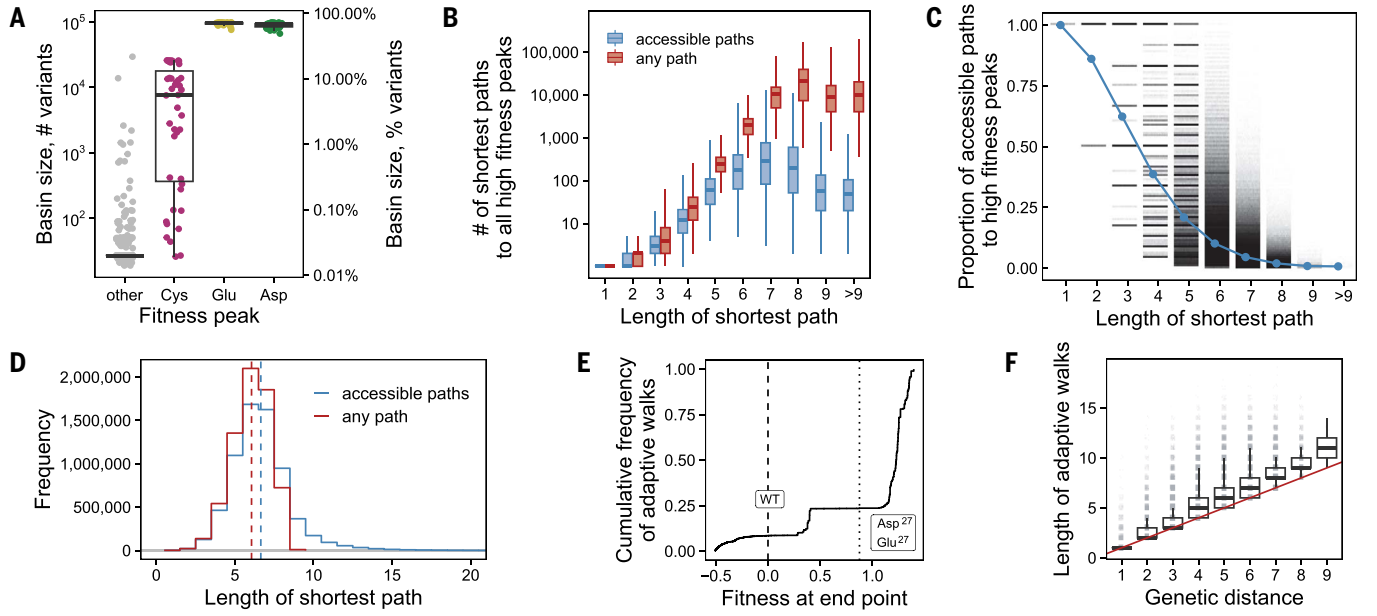


Fig. 3. High fitness peaks are easily accessible. (**A**) Basin size of fitness peaks depends on the amino acid at position 27 (horizontal axis). Basin size is shown both as the number of variants in the basin (left vertical axis) and as the percentage of the total number of the variants in the landscape (right vertical axis). (**B**) The landscape contains many paths to high fitness peaks. The vertical axis shows the total number of any shortest paths per variant to a high fitness peak. The horizontal axis shows the length of the shortest path. Red and blue boxplots summarize the number of paths for any shortest paths and accessible shortest paths. Each box spans the interquartile range (IQR), each horizontal line inside a box indicates the median value, and each whisker extends to the minimum or maximum value within a 1.5 IQR interval. The data values beyond the 1.5 IQR interval are not shown. N = 5,509,409,778 paths. (**C**) Proportion of accessible paths depends on path length. The vertical axis shows the proportion of accessible paths leading to high fitness peaks. Blue circles indicate the mean proportion of accessible paths at each length. N = 4,876,880 variant-peak pairs. (**D**) Length of shortest accessible paths. The blue line shows the length of shortest accessible paths leading from individual variants to each fitness peak. The red line shows the distribution for the length of any shortest path between corresponding variant-peak pairs (N = 6,748,190). (**E**) Cumulative distribution of fitness values reached by 10^6 adaptive walks starting from randomly selected variants. Dashed vertical line (x = 0): fitness of the wild type; dotted vertical line (x = 0.87): fitness of the lowest among all high fitness (Asp^{27}/Glu^{27}) peaks. (**F**) Length of adaptive walks. The vertical axis shows the length of adaptive walks that reached a high fitness peak. The horizontal axis shows the genetic distances between each starting variant and the attained peak. Red line ($y = x$): length of the shortest possible path to a high fitness peak, which is given by the genetic distance. Elements of boxplots are as in (B) (N = 765,181).

(>10^8 individuals), most fixation events are driven by selection. A population takes each mutational step with a probability that corresponds to this fixation probability. We repeated this procedure for every step of an adaptive walk until the walk reached a fitness peak.

Despite the predominance of low fitness peaks in our landscape, 76.5% of the adaptive walks reached a high fitness peak (Fig. 3E). These high fitness peaks exhibited trimethoprim resistance at least two orders of magnitude higher than the wild-type resistance (fig. S7). Analyzing the evolutionary trajectories leading to Asp and Glu peaks, we observed that 78% of walks began with acquiring a Glu^{27} or Asp^{27} allele, before converging to high fitness peaks (fig. S13, A and B). This suggests that the fitness advantage of Asp^{27} and Glu^{27} alleles, together with strong selection, drives populations to high fitness peaks (fig. S14). However, additional analysis revealed that nearly all accessible paths in our landscape lead to high fitness peaks, regardless of the selection gradients along these paths (see supplementary text section S2). On average, each variant has 1200.4 ± 1685.8 shortest accessible paths leading to high fitness peaks (mean ± SD, N = 134,662), compared with only 1.9 ± 5.7 paths to low fitness peaks (fig. S13D). Notably, adaptive walks tend to find peaks via relatively short paths, requiring only 5.6 ± 2.1 mutational steps (mean ± SD, N = 765,181) (Fig. 3F). Hence, most populations can easily reach high fitness peaks via short and abundant paths.

High fitness peaks are consistently accessible to Darwinian evolution across the different assumptions we tested in the adaptive walk simulations (fig. S15). However, these simulations overlook population dynamics and demographic stochasticity, which could influence peak accessibility by introducing clonal competition (*60*) or allowing populations to stochastically escape fitness peaks (*61*). To address this limitation, we performed individual-based simulations at realistic population sizes and mutation rates (*48*, *56*, *62*). When simulating evolution in 2600 populations for 100,000 discrete generations (fig. S16), we found that 73.1% of populations reach a high fitness peak (fig. S15), with the majority requiring fewer than 10,000 generations to do so (median ± interquartile range = 4870 ± 10,552 generations; fig. S17). These populations underwent 5.2 ± 1.8 selective sweeps (mean ± SD, N = 2034; fig. S15F), a number that agrees with the mean length of adaptive walks (5.6 ± 2.1 mutational steps). In sum, individual-based simulations confirm the high evolutionary accessibility of Asp and Glu peaks, despite the competition between clones (fig. S17, C to F) and rare cases of fitness peak escape (fig. S17, A and B).

Simultaneous accessibility of peaks leads to contingent evolution

Because high fitness peaks have enormous basins of attraction, we hypothesized that variants must be members of multiple basins. To quantify this overlap between different basins, we first determined the proportion of variants shared by any two basins. We found that Glu^{27} and Asp^{27} peaks shared 90.1 ± 6.4% (mean ± SD, N = 2701) of the variants in their basins (Fig. 4A). In contrast, the basins of the Cys^{27} peaks shared a much smaller proportion of their variants on average (mean ± SD = 22 ± 29%, N = 903). Furthermore, the basins of Asp and Glu peaks shared only 1.6 ± 1.6% (mean ± SD, N = 3182) variants with those of Cys^{27}

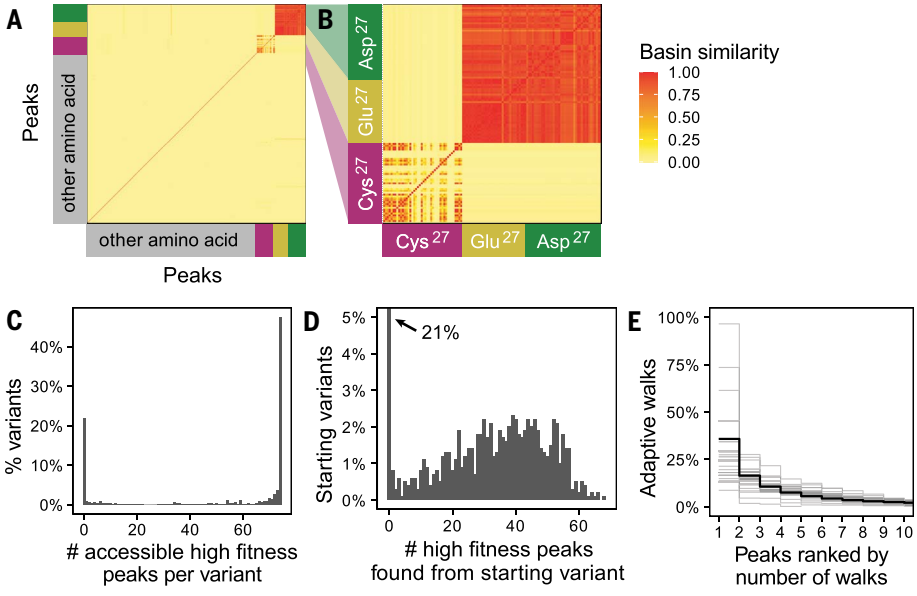


Fig. 4. Overlap of basins of attraction and evolutionary contingency. (**A**) The heatmap shows the fraction of variants shared by different basins of attraction. Each row and each column correspond to one peak, and peaks are classified according to the amino acid at position 27. A value of one (red) means that the basins of the corresponding peaks comprise identical sets of variants, and a value of zero (yellow) means that two basins do not share any variants. The results are presented as a symmetric matrix of pairwise fractional overlaps between pairs of all 514 peaks. (**B**) A magnified portion of the matrix including only peaks containing Asp^{27}, Glu^{27}, and Cys^{27}. The basins of high fitness peaks (Glu^{27} and Asp^{27}) share more than 90% of their variants. (**C**) Distribution of the number of high fitness peaks (Asp^{27}/Glu^{27}) that are accessible to each variant (N = 135,178). (**D**) The distribution shows the number of high fitness peaks discovered by a total of 1000 populations evolving through adaptive walks that start from the same variant. Data are based on 10^6 adaptive walks, such that 10^3 walks started from the same variant among 10^3 randomly chosen starting variants. Adaptive walks starting from 21% (209/1000) variants did not reach any high fitness peaks (vertical bar at x = 0). For the remaining starting variants, adaptive walks reached multiple high fitness peaks. (**E**) Adaptive walks preferentially attain some high fitness peaks. The vertical axis shows the percentage of adaptive walks (out of 1000) that reached the most preferred peaks (horizontal axis). Each gray line summarizes the data for 1000 adaptive walks started from the same variant. To reduce visual clutter, gray lines are shown for only 24 randomly selected starting variants (24 gray lines). The black bold line shows the mean percentage for all 1000 starting variants.

peaks (Fig. 4B). In other words, only a small proportion of variants had simultaneous access to both Asp^{27}/Glu^{27} and Cys^{27} peaks. Overall, 77% (104,496/135,176) of variants (functional and nonfunctional) had access to more than one high fitness peak. Notably, 47.5% of all variants in the landscape had simultaneous access to all 74 high fitness peaks (Fig. 4C).

Such simultaneous accessibility of multiple peaks can give rise to evolutionary contingency—the dependence of a historical process on chance events—because adaptive evolution starting from the same location can lead to different high fitness peaks. To test this hypothesis, we simulated 10^6 additional adaptive walks, such that 1000 walks started from each of 1000 randomly chosen starting variants. Adaptive walks originating from the same variant collectively reached 31 different high fitness peaks (mean ± SD = 29.5 ± 17.8 peaks, median 31; Fig. 4D), confirming that simultaneous peak accessibility renders the identity of an attained fitness peak contingent on chance events during adaptive evolution.

During adaptive evolution, not all high fitness peaks are equally likely to be found by a population starting from a particular variant. For instance, 36 ± 25% (mean ± SD) of adaptive walks starting from the same location in the landscape reach a single, most commonly attained high fitness peak (Fig. 4E). But even walks converging to the same peak are likely to use different paths (fig. S18), which is another manifestation of contingency in our landscape. In sum, the large size and highly overlapping basins of attraction of different peaks render evolution on our DHFR landscape highly contingent on stochastic events.

Discussion

The DHFR fitness landscape that we mapped through CRISPR-Cas9 gene editing is highly rugged, harboring 514 (mostly low) fitness peaks. At the same time, the landscape displays multiple properties expected from a smooth landscape, including an abundance of monotonically fitness-increasing paths to high fitness peaks, enormous basins of attraction of these peaks, and easy reachability of these peaks by most evolving populations. High peak accessibility in a rugged landscape contradicts the predictions of classical computational models of random fitness landscapes, such as the *NK* landscape (*16*). More biologically realistic models are needed to explain our findings. One such model requires a trade-off between fitness in the presence and absence of antibiotics, which is not consistent with our data (fig. S19) (*63*). Theoretical explanations of our observations thus remain an exciting task for future work.

Our metric of ruggedness—the number of peaks—correlates with other such metrics (*9*, *11*, *64*). Among the most widely used is the incidence of reciprocal sign epistasis (fig. S20A), which refers to a nonadditive interaction between mutations, in which a combination of two deleterious mutations produces a positive fitness effect. This type of epistasis causes nonmonotonic fitness changes along mutational paths and can create local fitness reduction that separates fitness peaks (*65*). However, reciprocal sign epistasis is necessary but not sufficient for the existence of multiple peaks (*20*, *65*, *66*). In our landscape, we found that 12.5% of mutant pairs show reciprocal sign epistasis (fig. S20B). This value falls within the range of other combinatorially complete landscapes (8 to 22%) (*40*, *64*, *67*), where peak accessibility has not been directly determined. In complex landscapes like these, the relationship between peak accessibility and reciprocal sign epistasis may also be complex, and simple proxies of ruggedness may be less useful indicators of landscape navigability than is commonly assumed.

Despite the presence of reciprocal sign epistasis, high fitness peaks remain accessible via short and abundant paths (Fig. 3). As a result, extradimensional bypasses (*38*, *68*, *69*) (i.e., indirect and longer paths that detour around a local fitness reduction) do not play a major role in rendering fitness peaks accessible (supplementary text section S3 and fig. S20, C and D). In our analyses, we initially focused on epistatic interactions between two nucleotide positions. When exploring epistasis at three or more positions (*70*), we identified the existence of such higher-order interactions (fig. S21, A to D). However, no more than three orders are needed to explain 93% of fitness variation in the landscape (fig. S21, E to K), and the strongest interactions frequently involve nucleotide positions within one codon (fig. S21, B

and C). Finally, we detected another type of epistatic interaction, known as diminishing returns epistasis. This epistasis manifests as a reduction in fitness gains from beneficial mutations as evolving populations approach the maximum fitness (fig. S22). Diminishing returns epistasis represents a form of global epistasis, as it universally affects most mutations at high fitness levels (*71*, *72*).

Because the effective population size of *E. coli* exceeds 10^8 individuals (*56*), selection is much stronger than genetic drift. For any organism with a large population size, even subtle differences in fitness can be selected upon. For example, synonymous mutations can have a measurable effect on fitness, because of codon-specific effects on mRNA stability, and on the rate and fidelity of translation (*73*–*76*). We thus considered all mutations (including synonymous mutations) in our landscape as non-neutral. However, we note that any existing technologies to measure fitness are limited by measurement error. Specifically, even in our sixfold replicated experiments with high sequencing coverage, we had low power to detect fitness differences of <5% (fig. S23). Such fitness differences exist between 11% of all mutational neighbors in the DHFR landscape. When we consider these differences to be effectively neutral, high fitness peaks remain accessible albeit through longer adaptive walks (fig. S24). Similarly, small population size does not substantially affect peak accessibility (fig. S24). Thus, our main findings are not affected by the neutrality assumption.

The simultaneous accessibility of multiple fitness peaks results in evolutionary contingency at the genotype level, because different populations arrive at different high fitness peaks. Previous studies highlighted that genetic drift, mutational stochasticity, and epistasis can create such contingency (*52*, *77*, *78*). Our observations show that such contingency can even arise when drift is negligible. Genotypic contingency does not preclude the predictability of phenotypic evolution (*46*, *52*, *77*, *79*, *80*), as most populations attained high fitness in a given environment.

However, different genotypes with similar fitness in one environment can have different fitness in another environment. This creates bifurcation points at which evolutionary trajectories can diverge in a new environment (*78*). Some of our high fitness genotypes illustrate this principle, as they differ in their fitness on a chemically modified version of trimethoprim (*81*). Whereas DHFR alleles Glu^{27} and Thr^{26} provide resistance against this modified antibiotic, allele Arg^{28} does not. All three alleles occur among the high fitness peaks of our landscape (Thr^{26} occurs in 15 peaks, Glu^{27} occurs in 34 peaks, and Arg^{28} occurs in 1 peak). Depending on the peak it started from, a population subject to this modified antibiotic would follow different evolutionary trajectories. Moreover, deformation of landscapes as a result of an environmental change may open up new evolutionary paths unavailable in a previous environment (*62*, *82*–*84*). Therefore, given the frequent interactions between genotype and environment (*85*), genotypic contingency is likely to cause phenotypic divergence (*78*).

Our landscape is one of the largest empirical landscapes currently available, but it covers only a small segment of a single gene, whose variation constitutes a tiny fraction of sequence space. This is an unavoidable limitation of any empirical landscape study. It also makes generalization difficult because the choice of mutational target can affect the properties of a reconstructed landscape (*86*). However, some features explaining the high navigability of our landscape may also apply to other landscapes. First, Glu^{27} and Asp^{27} peaks essentially share one enormous basin of attraction. The reason is that glutamate and aspartate are physiochemically similar amino acids that are encoded by similar codons in the standard genetic code. More generally, the genetic code has evolved to minimize the effect of mutations on the physicochemical properties of amino acids (*87*, *88*), suggesting that basins of attraction shared by adaptive peaks with functionally similar amino acids should be common in biological landscapes.

Second, we deliberately targeted a conserved gene that encodes a key metabolic enzyme in which a single mutation can lead to profound fitness and pleiotropic effects (*50*, *89*). Consequently, only 7% of variants resulted in a functional enzyme. Despite this constraint, functional variants formed a landscape with many fitness-increasing paths. If this phenomenon also exists in the landscapes of other conserved genes, it may help explain how pathogenic bacteria can evolve antibiotic resistance by altering essential protein targets of antibiotics (e.g., RNA polymerase, DNA gyrase, topoisomerase IV, and dihydropteroate synthase) (*90*). A landscape derived from a less-conserved gene would have harbored many more functional variants and thus potentially even more paths accessible to Darwinian evolution than we observed. In contrast, fitness landscapes in which amino acid positions have even stronger functional interactions (such as an enzyme's catalytic triad) might be evolutionarily more constrained. Future experiments using different mutational targets, organisms, and sampling design will show whether rugged yet highly navigable landscapes are typical or unusual.

When Sewall Wright coined the concept of an adaptive landscape nearly a century ago, he was concerned that multipeaked landscapes may prevent adaptive Darwinian evolution driven by natural selection (*1*). Simple theoretical models developed in the 20th century support this concern (*14*, *16*), but experimental evidence has been lacking. The landscape we studied shows that ruggedness need not impair Darwinian adaptation, even though it creates an enormous potential for contingent evolution. Our results suggest that we will need to refine our current theoretical understanding of the relationship between landscape ruggedness and navigability. Improved landscape theory will have to capture realistic properties of empirical landscapes, such as the sparsity of high-order epistasis (*19*), the adaptational tradeoffs between different fitness components (*63*), strong local correlation in fitness among genotypes (*41*), and dense connectivity of sequence space (*38*, *69*, *91*). Because applications of landscapes extend to different fields (*4*–*8*), including ecology (*2*), synthetic biology (*3*), and biomedicine (*92*), better data and theory on complex landscapes may also require other fields to reevaluate the challenges of optimization problems on their landscapes. Even though a small fraction of peaks may have high fitness, in landscapes like that of DHFR, these peaks can easily be discovered by blind Darwinian evolution.

Materials and methods summary

Full materials and methods are provided in the supplementary materials (*48*). In brief, we edited a 9-nt segment of the *E. coli* chromosomal gene *folA* using the no-SCAR protocol (*45*). To enable deep mutagenesis, we had to solve two problems. First, to allow cells with nonfunctional DHFR variants to grow, we integrated an inducible *dfrB9* gene, which encodes an alternative dihydrofolate reductase (unrelated to *E. coli* DHFR), into the chromosome (figs. S25 to S28). Second, to overcome a reduction in the efficiency of gene editing caused by mismatch DNA repair, we used a DNA repair–deficient strain of the *E. coli* K12 MG1655 (figs. S25 and S29). During gene editing, we transformed cells with a degenerate oligonucleotide encoding all possible codon combinations for positions 26, 27, and 28 of DHFR. We induced the expression of the guide RNA and of Cas9 and recovered cells by activating *dfrB9* expression. We stored the resulting DHFR mutant library at −70°C.

To measure fitness, we recovered the mutant library for 9 hours in M9-medium (0.4% glucose, 0.2% casamino acids) in the absence of *dfrB9* expression (fig. S30). Next, we inoculated 5×10^6 cells into fresh M9 medium (0.4% glucose, 0.2% casamino acids) containing 0.4 μg/ml of trimethoprim and incubated the resulting culture for 14 hours at 225 rpm and 30°C (fig. S31). We performed selection in six parallel replicate cultures. We isolated DNA from all replicates before and after trimethoprim selection and used it in a polymerase chain reaction to amplify a 214–base pair (bp) DNA fragment that spans the 9-nt mutated genome locus. We used a commercial sequencing service (150 bp paired-end Illumina NovaSeq

6000 at Eurofins Genomics, Germany) and implemented a custom analysis pipeline to count DHFR variants using a conservative detection threshold (=100 complete read pairs combined in all replicates before selection; fig. S32). We applied generalized linear regression to estimate the selection coefficient for each detected variant relative to the wild type (*48*).

We used fitness data and genetic distance information to reconstruct the DHFR fitness landscape as a graph (network), where nodes (vertices) represent variants and edges (links) represent single-nucleotide substitutions. In general, we only considered fitness-increasing edges accessible to Darwinian evolution. We determined fitness peaks (variants with only incoming edges), evolutionarily accessible paths (sequences of strictly fitness-increasing edges), basins of attraction (sets of all variants with evolutionarily accessible paths to a given peak), and pairwise overlaps among basins of attraction. For all the above analyses, we used the igraph v.1.3.4 library in R v.4.1.3 (*93*).

To study adaptive evolution in our landscape, we performed numerical simulations assuming strong selection and weak mutation, resulting in populations being genetically monomorphic most of the time (occupying a single genotype in the landscape), which allowed us to model evolution as an adaptive walk on the landscape (*57*). For each such walk, we chose as a starting genotype a random nonpeak variant in the graph. We then used Kimura's fixation probability (*48*, *59*, *94*) to stochastically draw a next variant among the genotype's one-mutation neighbors and repeated this process for, at most, 50 mutational steps (fixation events) or until the random walk had reached a fitness peak. In other random walk simulations, we either always selected the fittest neighbor (greedy walks), or we assumed a uniform fixation probability for all fitness-increasing mutations. We performed a total of 10^6 stochastic simulations, except for the deterministic greedy random walk, where we performed 134,664 such simulations using custom R code and the igraph library (*48*, *93*). Moreover, we performed 2600 individual-based simulations of haploid asexually reproducing populations for 100,000 generations using the simulation platform simuPop (*48*, *95*).

REFERENCES AND NOTES

1. S. Wright, "The roles of mutation, inbreeding, crossbreeding and selection in evolution," in vol. 1 of *Proceedings of the Sixth International Congress of Genetics*, D. F. Jones, Ed. (Genetics Society of America, 1932), pp. 356–366.
2. A. Sanchez-Gorostiaga, D. Bajić, M. L. Osborne, J. F. Poyatos, A. Sanchez, High-order interactions distort the functional landscape of microbial consortia. *PLOS Biol.* **17**, e3000550 (2019). doi: 10.1371/journal.pbio.3000550; pmid: 31830028
3. X. Yi, A. M. Dean, Adaptive landscapes in the age of synthetic biology. *Mol. Biol. Evol.* **36**, 890–907 (2019). doi: 10.1093/molbev/msz004; pmid: 30657938
4. B. M. R. Stadler, P. F. Stadler, Generalized topological spaces in evolutionary theory and combinatorial chemistry. *J. Chem. Inf. Comput. Sci.* **42**, 577–585 (2002). doi: 10.1021/ci0100898; pmid: 12086517
5. H. Mühlenbein, "Evolution in time and space - the parallel genetic algorithm" in *Foundations of Genetic Algorithms* (Morgan Kaufmann, 1991), pp. 316–337.
6. B. Manderick, M. K. de Weger, P. Spiessens, "The genetic algorithm and the structure of the fitness landscape" in *Proceedings of the 4th International Conference on Genetic Algorithms, San Diego, CA, USA, July 1991*, R. K. Belew, L. B. Booker, Eds. (Morgan Kaufmann, 1991), pp. 143–150.
7. D. A. Levinthal, Organizational adaptation and environmental selection-interrelated processes of change. *Organ. Sci.* **2**, 140–145 (1991). doi: 10.1287/orsc.2.1.140
8. V. K. Vassilev, J. F. Muller, T. C. Fogarty, "Digital circuit evolution and fitness landscapes" in vol. 2 of *Proceedings of the 1999 Congress on Evolutionary Computation-CEC99* (1999), pp. 1299–1306.
9. J. A. G. M. de Visser, J. Krug, Empirical fitness landscapes and the predictability of evolution. *Nat. Rev. Genet.* **15**, 480–490 (2014). doi: 10.1038/nrg3744; pmid: 24913663
10. I. Fragata, A. Blanckaert, M. A. Dias Louro, D. A. Liberles, C. Bank, Evolution in the light of fitness landscape theory. *Trends Ecol. Evol.* **34**, 69–82 (2019). doi: 10.1016/j.tree.2018.10.009; pmid: 30583805
11. I. G. Szendro, M. F. Schenk, J. Franke, J. Krug, J. A. G. M. de Visser, Quantitative analyses of empirical fitness landscapes. *J. Stat. Mech.* **2013**, P01005 (2013). doi: 10.1088/1742-5468/2013/01/P01005
12. D. M. Weinreich, R. A. Watson, L. Chao, Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* **59**, 1165–1174 (2005). doi: 10.1111/j.0014-3820.2005.tb01768.x; pmid: 16050094
13. C. Bank, Epistasis and adaptation on fitness landscapes. *Annu. Rev. Ecol. Evol. Syst.* **53**, 457–479 (2022). doi: 10.1146/annurev-ecolsys-102320-112153
14. S. Kauffman, S. Levin, Towards a general theory of adaptive walks on rugged landscapes. *J. Theor. Biol.* **128**, 11–45 (1987). doi: 10.1016/S0022-5193(87)80029-2; pmid: 3431131
15. P. Hegarty, A. Martinsson, On the existence of accessible paths in various models of fitness landscapes. *Ann. Appl. Probab.* **24**, 1375–1395 (2014). doi: 10.1214/13-AAP949
16. S. A. Kauffman, E. D. Weinberger, The *NK* model of rugged fitness landscapes and its application to maturation of the immune response. *J. Theor. Biol.* **141**, 211–245 (1989). doi: 10.1016/S0022-5193(89)80019-0; pmid: 2632988
17. J. F. C. Kingman, A simple model for the balance between selection and mutation. *J. Appl. Probab.* **15**, 1–12 (1978). doi: 10.2307/3213231
18. T. Aita, Y. Husimi, Fitness spectrum among random mutants on Mt. Fuji-type fitness landscape. *J. Theor. Biol.* **182**, 469–485 (1996). doi: 10.1006/jtbi.1996.0189; pmid: 8944894
19. F. J. Poelwijk, M. Socolich, R. Ranganathan, Learning the pattern of epistasis linking genotype and phenotype in a protein. *Nat. Commun.* **10**, 4213 (2019). doi: 10.1038/s41467-019-12130-8; pmid: 31527666
20. K. Crona, D. Greene, M. Barlow, The peaks and geometry of fitness landscapes. *J. Theor. Biol.* **317**, 1–10 (2013). doi: 10.1016/j.jtbi.2012.09.028; pmid: 23036916
21. E. R. Lozovsky *et al.*, Stepwise acquisition of pyrimethamine resistance in the malaria parasite. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 12025–12030 (2009). doi: 10.1073/pnas.0905922106; pmid: 19587242
22. A. I. Khan, D. M. Dinh, D. Schneider, R. E. Lenski, T. F. Cooper, Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* **332**, 1193–1196 (2011). doi: 10.1126/science.1203801; pmid: 21636772
23. P. E. O'Maille *et al.*, Quantitative exploration of the catalytic landscape separating divergent plant sesquiterpene synthases. *Nat. Chem. Biol.* **4**, 617–623 (2008). doi: 10.1038/nchembio.113; pmid: 18776889
24. D. W. Hall, M. Agan, S. C. Pope, Fitness epistasis among 6 biosynthetic loci in the budding yeast *Saccharomyces cerevisiae*. *J. Hered.* **101** (suppl. 1), S75–S84 (2010). doi: 10.1093/jhered/esq007; pmid: 20194517
25. J. da Silva, M. Coetzer, R. Nedellec, C. Pastore, D. E. Mosier, Fitness epistasis and constraints on adaptation in a human immunodeficiency virus type 1 protein region. *Genetics* **185**, 293–303 (2010). doi: 10.1534/genetics.109.112458; pmid: 20157005
26. L. Tan, S. Serene, H. X. Chao, J. Gore, Hidden randomness between fitness landscapes limits reverse evolution. *Phys. Rev. Lett.* **106**, 198102 (2011). doi: 10.1103/PhysRevLett.106.198102; pmid: 21668204
27. M. Lunzer, S. P. Miller, R. Felsheim, A. M. Dean, The biochemical architecture of an ancient adaptive landscape. *Science* **310**, 499–501 (2005). doi: 10.1126/science.1115649; pmid: 16239478
28. M. T. Reetz, J. Sanchis, Constructing and analyzing the fitness landscape of an experimental evolutionary process. *ChemBioChem* **9**, 2260–2267 (2008). doi: 10.1002/cbic.200800371; pmid: 18712749
29. K. M. Brown *et al.*, Compensatory mutations restore fitness during the evolution of dihydrofolate reductase. *Mol. Biol. Evol.* **27**, 2682–2690 (2010). doi: 10.1093/molbev/msq160; pmid: 20576759
30. H.-H. Chou, H.-C. Chiu, N. F. Delaney, D. Segrè, C. J. Marx, Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* **332**, 1190–1192 (2011). doi: 10.1126/science.1203799; pmid: 21636771
31. C. Bank, S. Matuszewski, R. T. Hietpas, J. D. Jensen, On the (un)predictability of a large intragenic fitness landscape. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 14085–14090 (2016). doi: 10.1073/pnas.1612676113; pmid: 27864516
32. C. A. Olson, N. C. Wu, R. Sun, A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain. *Curr. Biol.* **24**, 2643–2651 (2014). doi: 10.1016/j.cub.2014.09.072; pmid: 25455030
33. K. S. Sarkisyan *et al.*, Local fitness landscape of the green fluorescent protein. *Nature* **533**, 397–401 (2016). doi: 10.1038/nature17995; pmid: 27193686
34. W. Rowe *et al.*, Analysis of a complete DNA-protein affinity landscape. *J. R. Soc. Interface* **7**, 397–408 (2010). doi: 10.1098/rsif.2009.0193; pmid: 19625306
35. C. L. Warren *et al.*, Defining the sequence-recognition profile of DNA-binding molecules. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 867–872 (2006). doi: 10.1073/pnas.0509843102; pmid: 16418267
36. J. L. Payne, A. Wagner, The causes of evolvability and their evolution. *Nat. Rev. Genet.* **20**, 24–38 (2019). doi: 10.1038/s41576-018-0069-z; pmid: 30385867
37. J. I. Jiménez, R. Xulvi-Brunet, G. W. Campbell, R. Turk-MacLeod, I. A. Chen, Comprehensive experimental fitness landscape and evolutionary network for small RNA. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 14984–14989 (2013). doi: 10.1073/pnas.1307604110; pmid: 23980164
38. N. C. Wu, L. Dai, C. A. Olson, J. O. Lloyd-Smith, R. Sun, Adaptation in protein fitness landscapes is facilitated by indirect paths. *eLife* **5**, e16965 (2016). doi: 10.7554/eLife.16965; pmid: 27391790
39. V. O. Pokusaeva *et al.*, An experimental assay of the interactions of amino acids from orthologous sequences shaping a complex fitness landscape. *PLOS Genet.* **15**, e1008079 (2019). doi: 10.1371/journal.pgen.1008079; pmid: 30969963
40. J. Domingo, G. Diss, B. Lehner, Pairwise and higher-order genetic interactions during the evolution of a tRNA. *Nature* **558**, 117–121 (2018). doi: 10.1038/s41586-018-0170-7; pmid: 29849145
41. S.-T. Kuo *et al.*, Global fitness landscapes of the Shine-Dalgarno sequence. *Genome Res.* **30**, 711–723 (2020). doi: 10.1101/gr.260182.119; pmid: 32424071
42. A. J. Faure *et al.*, Mapping the energetic and allosteric landscapes of protein binding domains. *Nature* **604**, 175–183 (2022). doi: 10.1038/s41586-022-04586-4; pmid: 35388192
43. T. V. Lite *et al.*, Uncovering the basis of protein-protein interaction specificity with a combinatorially complete library. *eLife* **9**, e60924 (2020). doi: 10.7554/eLife.60924; pmid: 33107822
44. C. D. Aakre *et al.*, Evolving new protein-protein interaction specificity through promiscuous intermediates. *Cell* **163**, 594–606 (2015). doi: 10.1016/j.cell.2015.09.055; pmid: 26478181
45. C. R. Reisch, K. L. J. Prather, The no-SCAR (Scarless Cas9 Assisted Recombineering) system for genome editing in *Escherichia coli*. *Sci. Rep.* **5**, 15096 (2015). doi: 10.1038/srep15096; pmid: 26463009
46. E. Toprak *et al.*, Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nat. Genet.* **44**, 101–105 (2011). doi: 10.1038/ng.1034; pmid: 22179135
47. Y. T. Tamer *et al.*, High-order epistasis in catalytic power of dihydrofolate reductase gives rise to a rugged fitness landscape in the presence of trimethoprim selection. *Mol. Biol. Evol.* **36**, 1533–1550 (2019). doi: 10.1093/molbev/msz086; pmid: 30982891
48. Materials and methods are available as supplementary materials.
49. J. F. Crow, M. Kimura, *An Introduction to Population Genetics Theory* (Scientific Publishers, The Blackburn Press, 2010).

50. S. Thompson, Y. Zhang, C. Ingle, K. A. Reynolds, T. Kortemme, Altered expression of a quality control protease in *E. coli* reshapes the in vivo mutational landscape of a model enzyme. *eLife* **9**, e53476 (2020). doi: 10.7554/eLife.53476; pmid: 32701056
51. J. V. Rodrigues *et al*., Biophysical principles predict fitness landscapes of drug resistance. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E1470–E1478 (2016). doi: 10.1073/pnas.1601441113; pmid: 26929328
52. A. C. Palmer *et al*., Delayed commitment to evolutionary fate in antibiotic resistance fitness landscapes. *Nat. Commun.* **6**, 7385 (2015). doi: 10.1038/ncomms8385; pmid: 26060115
53. J. P. Volpato, J. N. Pelletier, Mutational 'hot-spots' in mammalian, bacterial and protozoal dihydrofolate reductases associated with antifolate resistance: Sequence and structural comparison. *Drug Resist. Updat.* **12**, 28–41 (2009). doi: 10.1016/j.drup.2009.02.001; pmid: 19272832
54. J. V. Rodrigues, E. I. Shakhnovich, Adaptation to mutational inactivation of an essential gene converges to an accessible suboptimal fitness peak. *eLife* **8**, e50509 (2019). doi: 10.7554/eLife.50509; pmid: 31573512
55. J. Clark, *A First Look at Graph Theory* (World Scientific, 2005).
56. M. Lynch *et al*., Genetic drift, selection and the evolution of the mutation rate. *Nat. Rev. Genet.* **17**, 704–714 (2016). doi: 10.1038/nrg.2016.104; pmid: 27739533
57. J. H. Gillespie, Some properties of finite populations experiencing strong selection and weak mutation. *Am. Nat.* **121**, 691–708 (1983). doi: 10.1086/284095
58. H. Lee, E. Popodi, H. Tang, P. L. Foster, Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc. Natl. Acad. Sci. U.S.A.* **109**, E2774–E2783 (2012). doi: 10.1073/pnas.1210309109; pmid: 22991466
59. M. Kimura, On the probability of fixation of mutant genes in a population. *Genetics* **47**, 713–719 (1962). doi: 10.1093/genetics/47.6.713; pmid: 14456043
60. C. B. Ogbunugafor, M. J. Eppstein, Competition along trajectories governs adaptation rates towards antimicrobial resistance. *Nat. Ecol. Evol.* **1**, 0007 (2016). doi: 10.1038/s41559-016-0007; pmid: 28812552
61. D. M. Weinreich, L. Chao, Rapid evolutionary escape by large populations from local fitness peaks is likely in nature. *Evolution* **59**, 1175–1182 (2005). pmid: 16050095
62. C. B. Ogbunugafor, C. S. Wylie, I. Diakite, D. M. Weinreich, D. L. Hartl, Adaptive landscape by environment interactions dictate evolutionary dynamics in models of drug resistance. *PLOS Comput. Biol.* **12**, e1004710 (2016). doi: 10.1371/journal.pcbi.1004710; pmid: 26808374
63. S. G. Das, S. O. Direito, B. Waclaw, R. J. Allen, J. Krug, Predictable properties of fitness landscapes induced by adaptational tradeoffs. *eLife* **9**, e55155 (2020). doi: 10.7554/eLife.55155; pmid: 32423531
64. S. Song, J. Zhang, Unbiased inference of the fitness landscape ruggedness from imprecise fitness estimates. *Evolution* **75**, 2658–2671 (2021). doi: 10.1111/evo.14363; pmid: 34554581
65. F. J. Poelwijk, S. Tănase-Nicola, D. J. Kiviet, S. J. Tans, Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. *J. Theor. Biol.* **272**, 141–144 (2011). doi: 10.1016/j.jtbi.2010.12.015; pmid: 21167837
66. M. Riehl, R. Phillips, L. Pudwell, N. Chenette, Occurrences of reciprocal sign epistasis in single- and multi-peaked theoretical fitness landscapes. *J. Phys. A Math. Theor.* **55**, 434002 (2022). doi: 10.1088/1751-8121/ac9938
67. C. Li, W. Qian, C. J. Maclean, J. Zhang, The fitness landscape of a tRNA gene. *Science* **352**, 837–840 (2016). doi: 10.1126/science.aae0568; pmid: 27080104
68. M. Conrad, The geometry of evolution. *Biosystems* **24**, 61–81 (1990). doi: 10.1016/0303-2647(90)90030-5; pmid: 2224072
69. S. F. Greenbury, A. A. Louis, S. E. Ahnert, The structure of genotype-phenotype maps makes fitness landscapes navigable. *Nat. Ecol. Evol.* **6**, 1742–1752 (2022). doi: 10.1038/s41559-022-01867-z; pmid: 36175543
70. D. M. Weinreich, Y. Lan, J. Jaffe, R. B. Heckendorn, The influence of higher-order epistasis on biological fitness landscape topography. *J. Stat. Phys.* **172**, 208–225 (2018). doi: 10.1007/s10955-018-1975-3; pmid: 29904213
71. S. Kryazhimskiy, D. P. Rice, E. R. Jerison, M. M. Desai, Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science* **344**, 1519–1522 (2014). doi: 10.1126/science.1250939; pmid: 24970088
72. J. Otwinowski, D. M. McCandlish, J. B. Plotkin, Inferring the shape of global epistasis. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E7550–E7558 (2018). doi: 10.1073/pnas.1804015115; pmid: 30037990
73. G. Boël *et al*., Codon influence on protein expression in *E. coli* correlates with mRNA levels. *Nature* **529**, 358–363 (2016). doi: 10.1038/nature16509; pmid: 26760206
74. J. B. Plotkin, G. Kudla, Synonymous but not the same: The causes and consequences of codon bias. *Nat. Rev. Genet.* **12**, 32–42 (2011). doi: 10.1038/nrg2899; pmid: 21102527
75. D. A. Drummond, C. O. Wilke, Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* **134**, 341–352 (2008). doi: 10.1016/j.cell.2008.05.042; pmid: 18662548
76. X. Shen, S. Song, C. Li, J. Zhang, Synonymous mutations in representative yeast genes are mostly strongly non-neutral. *Nature* **606**, 725–731 (2022). doi: 10.1038/s41586-022-04823-w; pmid: 35676473
77. Z. D. Blount, R. E. Lenski, J. B. Losos, Contingency and determinism in evolution: Replaying life's tape. *Science* **362**, eaam5979 (2018). doi: 10.1126/science.aam5979; pmid: 30409860
78. Z. D. Blount, C. Z. Borland, R. E. Lenski, Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 7899–7906 (2008). doi: 10.1073/pnas.0803151105; pmid: 18524956
79. J. R. Meyer *et al*., Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science* **335**, 428–432 (2012). doi: 10.1126/science.1214449; pmid: 22282803
80. H.-H. Chou, C. J. Marx, Optimization of gene expression through divergent mutational paths. *Cell Rep.* **1**, 133–140 (2012). doi: 10.1016/j.celrep.2011.12.003; pmid: 22832162
81. M. S. Manna *et al*., A trimethoprim derivative impedes antibiotic resistance evolution. *Nat. Commun.* **12**, 2949 (2021). doi: 10.1038/s41467-021-23191-z; pmid: 34011959
82. V. Mustonen, M. Lässig, From fitness landscapes to seascapes: Non-equilibrium dynamics of selection and adaptation. *Trends Genet.* **25**, 111–119 (2009). doi: 10.1016/j.tig.2009.01.002; pmid: 19232770
83. B. Steinberg, M. Ostermeier, Environmental changes bridge evolutionary valleys. *Sci. Adv.* **2**, e1500921 (2016). doi: 10.1126/sciadv.1500921; pmid: 26844293
84. E. S. King *et al*., Fitness seascapes are necessary for realistic modeling of the evolutionary response to drug therapy. bioRxiv 2022.06.10.495696 [Preprint] (2022); https://doi.org/10.1101/2022.06.10.495696.
85. C. Li, J. Zhang, Multi-environment fitness landscapes of a tRNA gene. *Nat. Ecol. Evol.* **2**, 1025–1032 (2018). doi: 10.1038/s41559-018-0549-8; pmid: 29686238
86. M. C. Whitlock, P. C. Phillips, F. B.-G. Moore, S. J. Tonsor, Multiple fitness peaks and epistasis. *Annu. Rev. Ecol. Syst.* **26**, 601–629 (1995). doi: 10.1146/annurev.es.26.110195.003125
87. S. J. Freeland, L. D. Hurst, The genetic code is one in a million. *J. Mol. Evol.* **47**, 238–248 (1998). doi: 10.1007/PL00006381; pmid: 9732450
88. H. Rozhoňová, C. Martí-Gómez, D. M. McCandlish, J. L. Payne, Protein evolvability under rewired genetic codes. bioRxiv 2023.06.20.545706 [Preprint] (2023); https://doi.org/10.1101/2023.06.20.545706.
89. S. Bershtein, J.-M. Choi, S. Bhattacharyya, B. Budnik, E. Shakhnovich, Systems-level response to point mutations in a core metabolic enzyme modulates genotype-phenotype relationship. *Cell Rep.* **11**, 645–656 (2015). doi: 10.1016/j.celrep.2015.03.051; pmid: 25892240
90. E. M. Darby *et al*., Molecular mechanisms of antibiotic resistance revisited. *Nat. Rev. Microbiol.* **21**, 280–295 (2023). doi: 10.1038/s41579-022-00820-y; pmid: 36411397
91. J. Franke, A. Klözer, J. A. G. M. de Visser, J. Krug, Evolutionary accessibility of mutational pathways. *PLOS Comput. Biol.* **7**, e1002134 (2011). doi: 10.1371/journal.pcbi.1002134; pmid: 21876664
92. D. Nichol *et al*., Steering evolution with sequential therapy to prevent the emergence of bacterial antibiotic resistance. *PLOS Comput. Biol.* **11**, e1004493 (2015). doi: 10.1371/journal.pcbi.1004493; pmid: 26360300
93. G. Csardi, T. Nepusz, The igraph software package for complex network research. *InterJournal* **Complex Systems**, 1695 (2006).
94. M. Kimura, *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, 1983).
95. B. Peng, M. Kimmel, simuPOP: A forward-time population genetics simulation environment. *Bioinformatics* **21**, 3686–3687 (2005). doi: 10.1093/bioinformatics/bti584; pmid: 16020469
96. A. Papkou, Archive of templatelib gitlab repository, version 1, Zenodo (2023); https://doi.org/10.5281/zenodo.8229020.
97. A. Papkou, Archive of gitlab repository to analyse amplicon sequence data from deep mutagenesis experiments, version 1, Zenodo (2023); https://doi.org/10.5281/zenodo.8229097.
98. A. Papkou, Code to reproduce data analyses and figures for manuscript "A rugged yet highly navigable landscape," version 1, Zenodo (2023); https://doi.org/10.5281/zenodo.8228920.

ACKNOWLEDGMENTS

We thank D. Pesce and C. Westmann for their help with establishing experimental protocols and J. Krug for helpful feedback. **Funding:** This work was supported by European Research Council grant agreement no. 739874 (A.W.), Swiss National Science Foundation grant 31003A_172887 (A.W.), the Zurich University Priority Research Program in Evolutionary Biology (A.W.), European Union's Horizon 2020 Program under grant agreement no. 101030711 (A.P.), European Research Council grant agreement no. 803375 (L.G.-P. and J.A.E.), and Programa de Atracción de Talento de la Comunidad de Madrid grants 2016-T1/BIO-1105 and 2020-5A/BIO-19726 (L.G.-P. and J.A.E.). **Author contributions:** A.P. and A.W. conceived the study and designed the experiments. A.P., L.G.-P., and J.A.E. carried out the experiments. A.P. and A.W. analyzed data. A.P. wrote computer code to carry out bioinformatic work, simulations, and analysis. A.W. and A.P. wrote the paper, which was edited by all authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Amplicon sequencing data are available from Sequence Read Archive under BioProject accession no. PRJNA998088. The computer code to design oligonucleotides for gene editing (https://doi.org/10.5281/zenodo.8229020) (*96*) and to process amplicon sequence data (https://doi.org/10.5281/zenodo.8229097) (*97*) is available from Zenodo. All data and code to reproduce results and figures are available from Zenodo (https://doi.org/10.5281/zenodo.8228920) (*98*). Bacterial strains and mutants will be made available under the terms of the Uniform Biological Material Transfer Agreement (UBMTA). This research was funded in whole or in part by the Swiss National Science Foundation (31003A_172887), a cOAlition S organization. The author will make the Author Accepted Manuscript (AAM) version available under a CC BY public copyright license.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adh3860
Materials and Methods
Supplementary Text
Figs. S1 to S32
Tables S1 to S4
References (*99*–*138*)
MDAR Reproducibility Checklist

Submitted 1 March 2023; accepted 29 September 2023
10.1126/science.adh3860

A rugged yet easily navigable fitness landscape

Andrei Papkou, Lucia Garcia-Pastor, José Antonio Escudero, and Andreas Wagner

Science **382** (6673), eadh3860. DOI: 10.1126/science.adh3860

Editor's summary

How many mutations does it take to move from one genetic fitness peak to another in a fitness landscape? Papkou *et al*. performed mutagenesis to survey the combinatorial genotypic space of nine nucleotides encoding three successive amino acids in a protein targeted by antibiotics in *Escherichia coli*. The authors found that most genotypes had low fitness, but that traveling between high fitness peaks required surprisingly few mutations. This work represents an exhaustive examination of more than 260,000 genotypes, surveying a nearly complete network of mutational paths to answer a long-standing question. —Corinne Simonti