Empirical fitness landscapes and the predictability of evolution

J. Arjan G.M. de Visser[1] and Joachim Krug[2]

Abstract | The genotype–fitness map (that is, the fitness landscape) is a key determinant of evolution, yet it has mostly been used as a superficial metaphor because we know little about its structure. This is now changing, as real fitness landscapes are being analysed by constructing genotypes with all possible combinations of small sets of mutations observed in phylogenies or in evolution experiments. In turn, these first glimpses of empirical fitness landscapes inspire theoretical analyses of the predictability of evolution. Here, we review these recent empirical and theoretical developments, identify methodological issues and organizing principles, and discuss possibilities to develop more realistic fitness landscape models.

Fitness
A measure of reproductive success of an organism that determines the change of the corresponding genotypic frequency in the population by natural selection.

Epistasis
Any kind of genetic interaction that leads to a dependence of mutational effects on the genetic background.

[1]*Laboratory of Genetics, Wageningen University, Droevendaalsesteeg 1, 6708PB Wageningen, The Netherlands.*
[2]*Institute for Theoretical Physics, University of Cologne, Zülpicher Str. 77, 50937 Köln, Germany.*
e-mails: arjan.devisser@wur.nl; krug@thp.uni-koeln.de
doi:10.1038/nrg3744
Published online 10 June 2014

A central topic in biology concerns how genotypes determine the phenotypes and functions of organisms that affect their evolutionary success (that is, fitness). Geneticists, developmental biologists and systems biologists study how organisms develop each generation from their genetic programmes, whereas evolutionary biologists seek to understand the evolutionary causes and consequences of the genotype–fitness map[1–4]. It is now accepted that the shape of the genotype–fitness map has fundamental effects on the course of evolution, and the map has a prominent role in many theories, including those concerning divergence and speciation, sex, genetic robustness and evolvability[3,5]. A particularly exciting consequence, which has motivated recent work and is the topic of this Review, is the effect of this map on the predictability of evolution[6,7].

To consider evolutionary consequences of the genotype–fitness map, Sewall Wright[8] introduced the concept of the fitness landscape (also known as the adaptive landscape), which is a visualization of this high-dimensional map, often in few dimensions. The iconic graphical rendering is a three-dimensional 'mountainous' landscape in which genotypes are organized in the x–y plane and fitness is plotted on the z axis (FIG. 1a). In such a landscape, evolution can be seen as 'walks' and adaptation as 'climbs' to higher positions on the fitness surface. Despite its intuitive appeal for the study of evolution, the impact of the fitness landscape concept on evolutionary biology has been limited owing to the lack of empirical information about the topography of real fitness landscapes. However, this situation is now changing. Experimentalists have begun to provide glimpses of real fitness landscapes by constructing and analysing mutants that represent all possible combinations of small sets of mutations that are involved in a single evolutionary lineage or in multiple lineages. In turn, the empirical work has triggered theoretical analyses of the predictability of evolution and its determinants.

We begin this Review with a brief historical account of the role of fitness landscapes in the study of evolution. We then discuss the methodological approaches and main findings of recent efforts to reveal information about the topography of real fitness landscapes. Only studies that present systematic analyses of mutational interactions are included. We do not address the large body of work on pairwise mutational interactions and on qualitative features of the fitness surface, which has been discussed elsewhere[4,9,10]. We then turn to theoretical analyses of pathway accessibility and predictability informed by this empirical work, discuss methods for studying larger-scale fitness landscapes and end with an outlook on the development of this emerging field based on current challenges.

Historical development of a concept

Wright's idea to explicitly consider the relationship between genotypic space and fitness came from his conviction that, different from Ronald Fisher's additive view of genetics, real fitness landscapes are likely to be complex owing to pervasive epistasis[8,11]. Grossly underestimating the number of genes and alleles in an average species, Wright realized the vastness of genotypic space:

Ruggedness
A measure of the complexity of fitness landscapes due to multidimensional epistasis. However, it is often used in a more restricted way to reflect the presence of multiple peaks.

Magnitude epistasis
Epistatic interactions that affect the magnitude but not the sign of mutational effects on fitness.

"with 10 allelomorphs in each of 1000 loci, the number of possible combinations is 10^{1000} which is a very large number" (REF. 8). Given these vast possibilities, he imagined that "it may be taken as certain that there will be an enormous number of widely separated harmonious combinations" (REF. 8). Consequently, in Wright's view, fitness landscapes have ruggedness owing to the presence of multiple fitness peaks that are separated by 'valleys' of low-fitness genotypes. He pictured this "field of gene combinations" in two dimensions (FIG. 1b), although he was aware that such low-dimensional pictures "are a very inadequate representation of such a field". In fact, in the same publication, Wright provided a more appropriate graphical rendering of the high-dimensional discrete genotypic space, which had been concurrently identified by J. B. S. Haldane[12] as a hypercube (BOX 1).

For many years, the fitness landscape concept underwent little further development. One reason was the general lack of understanding of the molecular basis of adaptation and hence of the relevant 'genetic units' of the genotypic space. This changed when John Maynard Smith introduced the notion of mutational pathways in protein space[13]. He used the analogy of a word game, in which a word must be converted into another word of the same length with the requirement that one letter is changed at a time and that all intermediates are meaningful words (for example, from 'word' via 'wore', 'gore' and 'gone' to 'gene'). Given the low per-base-pair mutation rate, he argued that proteins also adapt by a series of single amino acid changes, in which all intermediate states must be functional for a trajectory to be accessible by natural selection. Two decades later, on the basis of the notion of a discrete protein space, Stuart Kauffman and co-workers[14,15] developed their NK model for studying evolution in a fitness landscape with 'tunable' ruggedness (BOX 2).

By the start of the twenty-first century, still little was known about the topography of fitness landscapes in real organisms. Two initially independent developments changed the situation and paved the road towards empirical studies of fitness landscapes (FIG. 1c). First, growing genomic information allowed the prediction of ancestral genotypes, which stimulated studies that 'resurrected' and functionally analysed these ancestors to infer evolutionary explanations[16]. An early example was the analysis of lysozymes in game birds by synthesizing all combinations of ancestral and derived amino acids at three positions[17]. As all trajectories included at least one enzyme with thermodynamic stability outside the range of the extant proteins, it was concluded that the evolution of lysozymes must have been non-neutral. Recently, the construction and analysis of intermediates have become a regular procedure for testing scenarios in microbial evolution experiments[18].

Second, a popular model of the evolution of sex, which requires weak negative magnitude epistasis among deleterious mutations[19], motivated empirical work on epistasis among deleterious mutations. Although most studies analysed pairwise mutational interactions or the dependence of mean fitness on the number of mutations[20], a few studies systematically constructed all possible combinations of a handful of mutations[21,22]. Despite

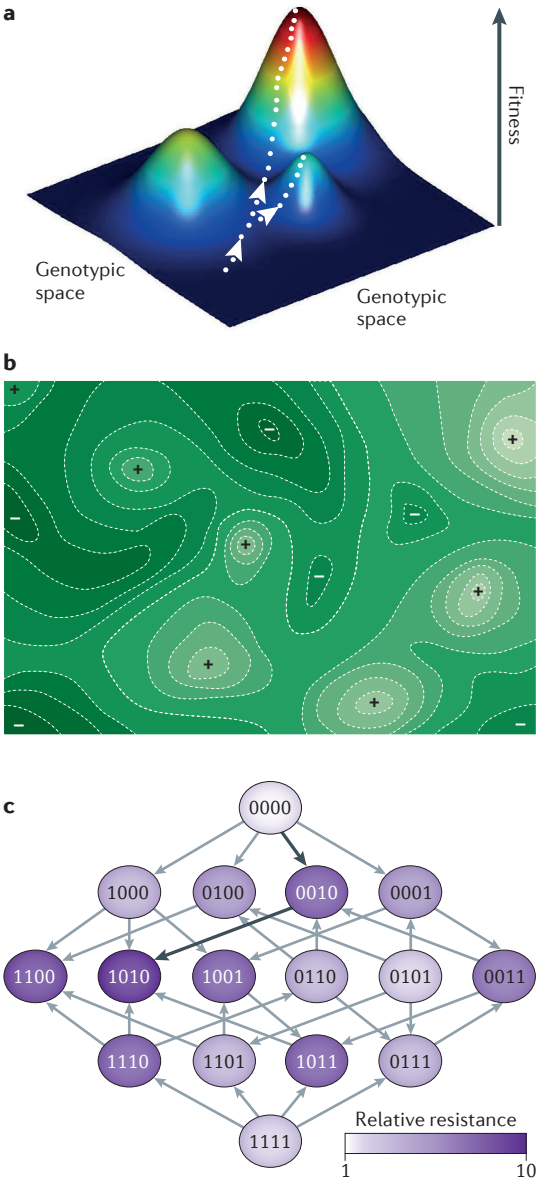

Figure 1 | **Development of the fitness landscape concept.** A fitness landscape can be visualized as a 'mountainous' landscape in three dimensions with genotypes arranged in the *x*–*y* plane and fitness on the *z* axis (part **a**). The landscape shown is rugged with three fitness peaks separated by fitness 'valleys', and two imaginary evolutionary trajectories are shown by white dots and arrows. Wright's two-dimensional "field of gene combinations" (REF. 8) is shown (part **b**). Fitness maximum and minimum are represented as "+" and "–", respectively; dotted lines are contours of equal fitness. A recent example of an empirical fitness landscape involves four mutations in the antibiotic resistance enzyme β-lactamase TEM1, which cause increased resistance to cefotaxime[44] (part **c**). Nodes represent the 2^4 (that is, 16) genotypes; 0 and 1 indicate wild-type and mutant amino acids, respectively. Arrows connect genotypes that differ by a single mutation and point towards genotypes with higher resistance. Bold black arrows indicate the 'greedy' walk (which substitutes the existing genotype with the largest-benefit mutation among the mutations available at each step) from wild-type (0000) to the global maximum (1010).

Box 1 | **Quantifying multidimensional epistasis**

The empirical data sets considered here comprise measurements of fitness or some related phenotype for combinations of L biallelic loci, which can be written as binary strings of variables (σ_i) taking two possible values: $\sigma = (\sigma_1, \sigma_2, \sigma_3, \ldots, \sigma_L)$. The resulting space of genotypes has the mathematical structure of an L-dimensional hypercube with 2^L corners. If some of the combinations are missing because the corresponding genotype is not viable or cannot be constructed, then the space is a subgraph of the hypercube. Any fitness function on the hypercube can be decomposed into products of the single-locus variables according to the following expansion[27,97]:

$$f(\sigma) = a^{(0)} + \sum a_i^{(1)}\sigma_i + \sum a_{ij}^{(2)}\sigma_i\sigma_j + \sum a_{ijk}^{(3)}\sigma_i\sigma_j\sigma_k + \ldots + a^{(L)}\sigma_1\sigma_2 \ldots \sigma_L \quad (1)$$

There are ${}^LC_k$ coefficients of type $a^{(k)}$ in this expansion, one for each subset of k of L loci. According to the binomial theorem, the total number of coefficients equals 2^L, which makes it evident that the mapping between fitness values and expansion coefficients is one-to-one. The first-order coefficient $a^{(1)}$ describes the linear, non-epistatic effects, the second-order coefficient $a^{(2)}$ denotes pairwise epistatic interactions and so on. In empirical data, one usually sets $\sigma_i = 0$ and $\sigma_i = 1$ to indicate the absence and presence of a mutation at locus i, respectively. However, for some purposes, the symmetrical choice $\sigma_i \in \{-1,1\}$ that treats all genotypes on an equal footing is preferable[98].

As the number of coefficients in expansion (1) is equal to the number of genotypes, the expansion provides a reduced description of the fitness landscape only if it happens to be sparse with many vanishing coefficients, for example, if interactions beyond a certain order are absent. A reduced description that applies more generally is obtained by summing the squares of coefficients of a given order, which yields the total weights of interactions of order n:

$$b^{(n)} = \sum \left| a_{i_1 \ldots i_n}^{(n)} \right|^2 \quad (2)$$

The amplitude spectrum $\{b^{(n)}\}$ is related by a linear transformation to the fitness correlation function $R(d) = \langle f(\sigma) f(\sigma') \rangle_{d_H(\sigma,\sigma')=d}$, which is obtained by averaging the product of fitness values over all pairs of genotypes at a constant Hamming distance d_H (REFS 99,100). To obtain a single number that characterizes the importance of epistatic interactions in the landscape, one sums the weights of orders $n > 1$ and normalizes by the sum over all weights[10]. A related overall epistasis measure is the roughness/slope ratio (r/s), which is defined as the standard deviation of the fitness values with respect to the best non-epistatic (that is, linear) fit divided by the average of the absolute values of the linear coefficients[30].

Importantly, epistasis measures that are derived from the interaction weights $\{b^{(n)}\}$ are not sensitive to the difference between magnitude and sign epistasis. On the basis of such measures alone, it is not possible to decide whether a given landscape is truly rugged or simply smoothly curved. They therefore need to be combined with quantities that are more directly related to the adaptive process, such as the number of fitness maxima (N_{max}) or the number of selectively accessible pathways (N_{paths}). Whereas these are global measures of sign epistasis that reflect the structure of the entire landscape, a local quantifier is obtained by considering all pairs of genotypes that differ at two loci and by determining the fraction of cases (f_{epi}) in which the two interact sign-epistatically[10,73].

the fact that these studies focused on the curvature of the fitness surface to test for negative unidimensional epistasis, they additionally generated complete information about the local topography on the basis of the multidimensional epistasis within these sets of mutations[4,23].

However, it was the realization by Daniel Weinreich and colleagues[24] of the evolutionary consequences of a particularly strong form of epistasis known as sign epistasis that motivated the recent surge of studies on empirical fitness landscapes. By reversing the fitness effects of mutations, sign epistasis introduces two sources of adaptive constraint: it causes some adaptive allele combinations to become 'isolated' by surrounding genotypes of low fitness[25] and reduces the number of accessible mutational trajectories under strong selection. This insight gave rise to the exciting prospect that evolution may be reproducible and perhaps even predictable (see below). The quantitative study of these implications requires complete information about the relevant part of the fitness landscape. Practically, this means that for a given pair of ancestral and derived genotypes, all 2^L possible combinations of the L mutations for which they differ must be constructed and their fitness or a proxy thereof measured (FIG. 1c). Weinreich and collaborators demonstrated the implications of sign epistasis by constructing and analysing a fitness landscape that involved five mutations in the β-lactamase TEM, which collectively gave rise to bacterial resistance to a novel antibiotic[26]. Only 18 of the 120 possible 5-step mutational trajectories from wild type to high-resistance enzyme were accessible under strong selection, and the single most likely trajectory would be used in almost half of the cases (discussed in detail below).

Empirical fitness landscapes

Type of available data. Empirical studies of fitness landscapes have used two different approaches. The first approach infers qualitative features of the topography, such as its ruggedness, either from patterns of parallel evolution in microbial evolution experiments or from the prevalence of sign epistasis among sets of constructed mutation pairs[4,9,10]. This approach probes a fairly large area of genotypic space, but the topographical information it reveals is incomplete and biased by the population dynamic regime used (see below). The second approach involves the systematic analysis of all possible combinations of a small, predefined set of mutations (FIG. 2). This approach explores a tiny part of genotypic space, but the information obtained is complete and allows the probability of mutational trajectories to be quantified and compared. Below, we focus on systematic studies that adopt the second approach and their use in analyses of evolutionary predictability.

Currently, there are <20 systematic studies of empirical fitness landscapes, but this number is rapidly growing[10,27]. These studies analyse interactions among three[17] to a maximum of nine mutations[28], which occur either in a single gene[17,26,28–38] or operon[39], or across genes in a bacterial[40,41], fungal[22,42] or fly genome[43]. The empirical landscapes can be classified on the basis of the source of the mutations involved (FIG. 2A). The largest class comprises studies of mutations that co-occur in extant genotypes which show a novel function of interest[28–31,34–37] (FIG. 2Aa) or in genotypes isolated from laboratory evolution experiments[26,33,40,41] (FIG. 2Ab). In a few studies, the collective effect of the mutations used was unknown (FIG. 2Ac), and mutations were selected on the basis of their individual negative[22,42,43] or positive effect on fitness[44]. In all but two studies[36,41], genotypes were analysed for a proxy of fitness, such as the maximum growth rate or level of antibiotic resistance, instead of fitness itself.

Emerging patterns and methodological issues. What do these studies tell us about the topography of real fitness landscapes? A recent meta-analysis of available data

Hamming distance
The distance between two genotypes measured by the number of mutations in which they differ.

Unidimensional epistasis
A description of epistasis based on the curvature of the relationship between average fitness and the number of mutations.

Multidimensional epistasis
Epistatic interaction that reflects the high-dimensional nature of genotypic space.

Sign epistasis
Epistatic interaction that affects the sign of mutational effects on fitness, such that a given mutation can be deleterious or beneficial depending on genetic background.

sets[10] compared the ruggedness on the basis of various statistical measures (BOX 1). Although these measures are sensitive to different types of epistasis (for example, magnitude or sign epistasis) at different genomic scales, they seem to correlate fairly well, which suggests that they reflect similar features of the topography[10] (FIG. 3). On the basis of these measures, two general observations can be made. First, the data sets show a variable, but on average substantial, level of ruggedness, even when no unidimensional epistasis is detected[4,10]. All fitness landscapes are more rugged than expected in the absence of epistasis but less so than expected if mutations had random effects across genetic backgrounds (FIG. 3). Second, several data sets[40,41,44] show on average diminishing returns epistasis among beneficial mutations, in which benefits are smaller in high fitness backgrounds than in low fitness backgrounds.

What can we say about the causes of the wide variation in ruggedness across empirical fitness landscapes? Clearly, too few data are available for drawing firm conclusions, particularly given the variation in methods, systems and types of mutations used. For example, no data exist for deleterious mutations that are known for their collective negative effect or occurrence in a single gene, and studies that involve beneficial mutations vary greatly in the scale of fitness effects. Nevertheless, some trends are worth mentioning, as they may help to interpret the growing data set.

A first trend is that mutations with known collective benefit show less ruggedness than mutations for which the combined effect is unknown (FIG. 3). This was shown in a recent comparative analysis of three sets of four mutations in the β-lactamase TEM1, in which mutations that increased resistance to cefotaxime by themselves showed stronger ruggedness than mutations that had a collective benefit[44]. This contrast is not surprising: a posteriori approaches (FIG. 2Aa,2Ab) are less likely to find much sign epistasis because the mutations have collectively been tested by selection, whereas mutations with only known individual benefits (FIG. 2Ac) do not suffer from this bias. Such a bias may be especially strong later during adaptation, when there is an increased likelihood that the paths that are not taken involve sign epistasis[45], and may also provide an explanation for the high ruggedness seen in landscapes of individually deleterious mutations[42]. The requirement for mutations to have individual benefits also introduces bias because it ignores the adaptive contribution of mutations that have deleterious or neutral effect in the ancestral background[44]. For an unbiased view, interactions among random mutations should be studied.

The same comparative analysis[44] also supports a second pattern: mutations of large effect show greater ruggedness than small-effect mutations. Among the two sets of four individually beneficial mutations in the β-lactamase TEM1, the large-effect mutations showed consistently higher ruggedness across different measures of epistasis than the small-effect mutations (FIG. 3). The authors explained this pattern with the fact that the nonlinear map that relates resistance to key enzyme properties has a greater impact on combinations of large-effect mutations than on combinations of small-effect mutations, which is consistent with predictions from various models[46–48].

A third trend is that intragenic landscapes have greater ruggedness than intergenic landscapes. This pattern is expected, given that epistasis among mutations in different genes is thought to result from functional (for example, metabolic) constraints, whereas intragenic epistasis may, in addition, result from structural constraints that affect enzymatic activity and folding stability[49,50]. The predicted trend is supported for landscapes that involve collectively beneficial mutations (FIG. 3), in which intragenic landscapes[26,35] show greater ruggedness than intergenic landscapes[40,41]. However, the available

Box 2 | **Models of fitness landscapes**

Epistatic interactions may arise at any stage of the mapping from genotype to phenotype to fitness, and theoretical models have correspondingly addressed the problem at different levels.

Random field models

In this approach, fitness landscapes are obtained as instances of a probabilistic algorithm that assigns fitness values directly to genotypes[99]. Such models allow one to systematically explore how different epistasis measures are interrelated and how landscape properties vary with dimensionality, and they can be used to parameterize and organize empirical data sets. For the comparison between models and empirical landscapes, it is often helpful to decompose the genotypic space into subgraphs that are spanned by subsets of loci and to treat the corresponding sublandscapes as instances of an ensemble with homogeneous statistical properties[10,42].

The simplest random field model assigns fitness values to genotypes independently from a fixed probability distribution[14]. Because of its similarity with a class of mutation–selection models introduced by Kingman[101], this is referred to as the House-of-Cards (HoC) model.

In the NK model introduced by Kauffman and Weinberger[15], each locus interacts with K other loci, where K has values between 0 and $L-1$, and L represents the total number of loci. Within each set of $K+1$ interacting loci, fitness values are assigned at random to the 2^{K+1} genotypes. For $K=0$, the landscape is additive, whereas $K=L-1$ corresponds to the maximally epistatic HoC model. Through the choice of interaction partners of a locus, different epistatic architectures can be implemented[65].

Rough Mount Fuji (RMF) models are obtained by combining a random HoC landscape with an additive landscape[30]. In the simplest version of the model, the additive selective advantage ($\bar{s}$) is the same for all loci. By varying $\bar{s}$ relative to the standard deviation of the HoC fitness values, the ruggedness of the landscape can be adjusted[42].

Sequence–structure maps

A paradigmatic example of an explicit genotype–phenotype map relates RNA sequences to their secondary structures predicted by complementary base pairing[88]. Explicit sequence–structure maps can also be constructed for proteins if the interactions between residues are suitably simplified, for example, by placing the protein onto a lattice[87]. Structure is then mapped to a fitness-like phenotype which, in the case of proteins, can be stability, abundance, folding robustness or affinity to a target ligand[90,102,103]. An important general feature of these genotype–phenotype maps is a high degree of redundancy (that is, many sequences map to the same structure), but there are also characteristic differences between RNA and protein landscapes[87,104].

Phenotype–fitness maps

A heuristic approach that is loosely based on Fisher's geometric model[61] explains epistasis by assuming that genotypes contribute additively to one or several phenotypes, which in turn map nonlinearly to fitness. In this way, sign epistasis can arise either from stabilizing selection towards a fitness optimum[105–107] or from two phenotypes that show a trade-off in their effect on fitness[44]. This approach is generally applicable whenever there is a reduction of dimensionality by passing from genotype to phenotype, but it is most powerful when it is based on explicit models of functional pathways[36,46,85,86,108].

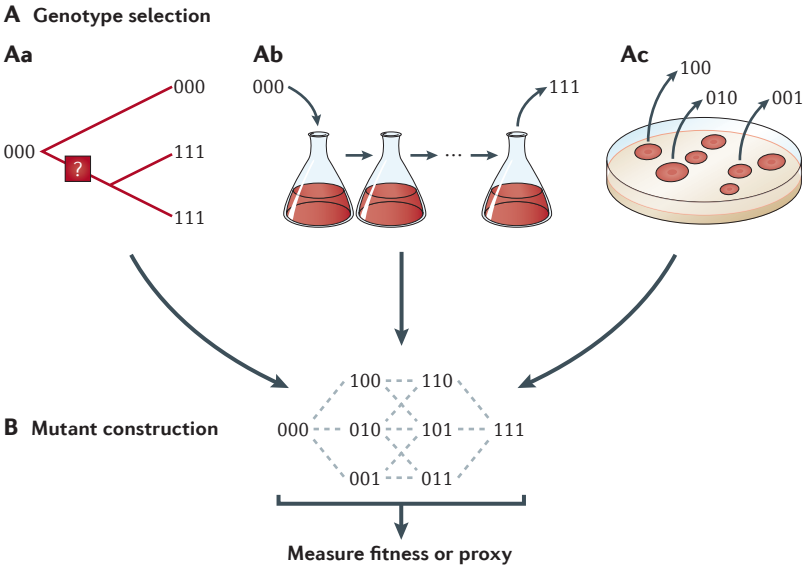

Figure 2 | **Approaches for the empirical study of fitness landscapes.** Experimental approaches for studying small-scale fitness landscapes share three essential components: a set of mutations of interest is identified (part **A**); mutants are constructed to carry all 2^L possible combinations of the L selected mutations (in this case, $L = 3$, and 0 and 1 indicate the absence and presence of the mutation, respectively) (part **B**); and the fitness or a fitness proxy (for example, antibiotic resistance) is measured for all genotypes. Mutations of interest can come from three different sources: from phylogenetic analyses that infer the ancestor of extant genotypes (part **Aa**); from microbial evolution experiments in which mutations co-occur in an evolving lineage (part **Ab**); or from sets of mutants that each carry a single mutation (part **Ac**), such as alternative mutations that cause antibiotic resistance. The a posteriori approaches (parts **Aa**, **Ab**) are less likely to find much sign epistasis because these mutations have collectively 'survived' the selective pressure, whereas the a priori approach (part **Ac**) does not suffer from this bias.

support is confounded by variation in both the effect sizes and the phenotypes affected by the mutations.

Evolutionary predictability

Whether evolution is inherently unpredictable (owing to chance events) or at least partly predictable is a fundamental question in biology and the topic of a longstanding debate[51,52]. Evolution experiments using microorganisms have begun to shed some light on this issue by probing the repeatability of adaptive trajectories in replicate populations[6,7]. Several studies have found extensive evidence for convergent or parallel evolution at the genetic level in which, repeatedly, the same or a similar outcome has evolved independently[53–55], whereas others have uncovered examples for the crucial dependence or historical contingency of evolutionary outcomes on random events, thus highlighting the intrinsic difficulty of evolutionary prediction[56,57]. A general finding of these experiments is that the amount of parallelism depends strongly on the organizational level; that is, there are higher frequencies of shared adaptive changes at the level of genes and metabolic pathways than at the level of nucleotide substitutions.

Repeatability in replicate experiments is a weak form of predictability, as the deterministic nature of the process can be ascertained only in retrospect. A high degree of repeatability is a necessary but insufficient condition for the more challenging task of forward prediction, which requires additional knowledge about the mechanistic basis of the traits under selection[58] and the adaptive process[59]. Below, we discuss the problem of evolutionary predictability from the perspective of the constraints imposed by the underlying fitness landscape and their interaction with the dynamics of the adapting population.

Pathway accessibility. The influence of the landscape structure is particularly clear-cut in the strong-selection–weak-mutation (SSWM) regime[60,61], in which the population can be treated as a monomorphic entity that is constrained to move uphill in fitness in single mutational steps. In this regime, the only mutational pathways available for adaptation are those along which fitness increases monotonically, and such paths have therefore been termed selectively accessible[24]. Focusing on pathways that terminate at the global fitness maximum, the authors established a one-to-one correspondence between evolutionary accessibility and sign epistasis: on a fitness landscape without sign epistasis, all direct paths from an arbitrary genotype to the global maximum are accessible and, conversely, any occurrence of sign epistasis implies that some paths are inaccessible. In this sense, pathway accessibility is a more stringent indicator for sign epistasis than the existence of multiple fitness peaks, which cannot be inferred from local epistatic interactions alone[25,62,63]. Unless stated otherwise, the following discussion is restricted to direct mutational pathways.

Many of the empirical examples reviewed above show substantially reduced accessibility, and the majority of pathways to the global maximum are inaccessible. This raises the questions of how accessible a 'typical' fitness landscape with sign epistasis should be, and to what extent pathway accessibility can be a quantitative metric of landscape ruggedness. For a baseline estimate, it is useful to consider the house-of-cards (HoC) model, in which fitness values of different genotypes are independent and identically distributed random variables (BOX 2). A simple combinatorial argument shows that the mean number of accessible paths from an arbitrary genotype to the global maximum is equal to one in the HoC model — a result that is independent of the distance to the peak[42]. However, at the same time, the proportion of landscapes without accessible pathways increases with increasing distance from the peak and approaches unity, which indicates that the mean number of accessible pathways is a poor estimate for the typical behaviour: in large HoC landscapes, the global optimum is typically not accessible via ascending paths. Interestingly, accessibility markedly increases if the initial genotype is assumed to have exceptionally low fitness[64].

Empirical fitness landscapes are usually smoother than those predicted by the HoC model (FIG. 3) and are better represented by models with tunable intermediate ruggedness, such as the NK model or the rough Mount Fuji (RMF) model (BOX 2). For the RMF model, it has been established that accessible paths to the global

Strong-selection–weak-mutation
(SSWM). A regime of population dynamics in which beneficial mutations are sufficiently rare to arise and fix independently, while selection is strong enough to prevent the fixation of deleterious mutations.

Direct paths
Shortest mutational pathways between genotypes, along which the distance to the target genotype decreases by one in each step. There are $d!$ direct paths between two genotypes at Hamming distance d.

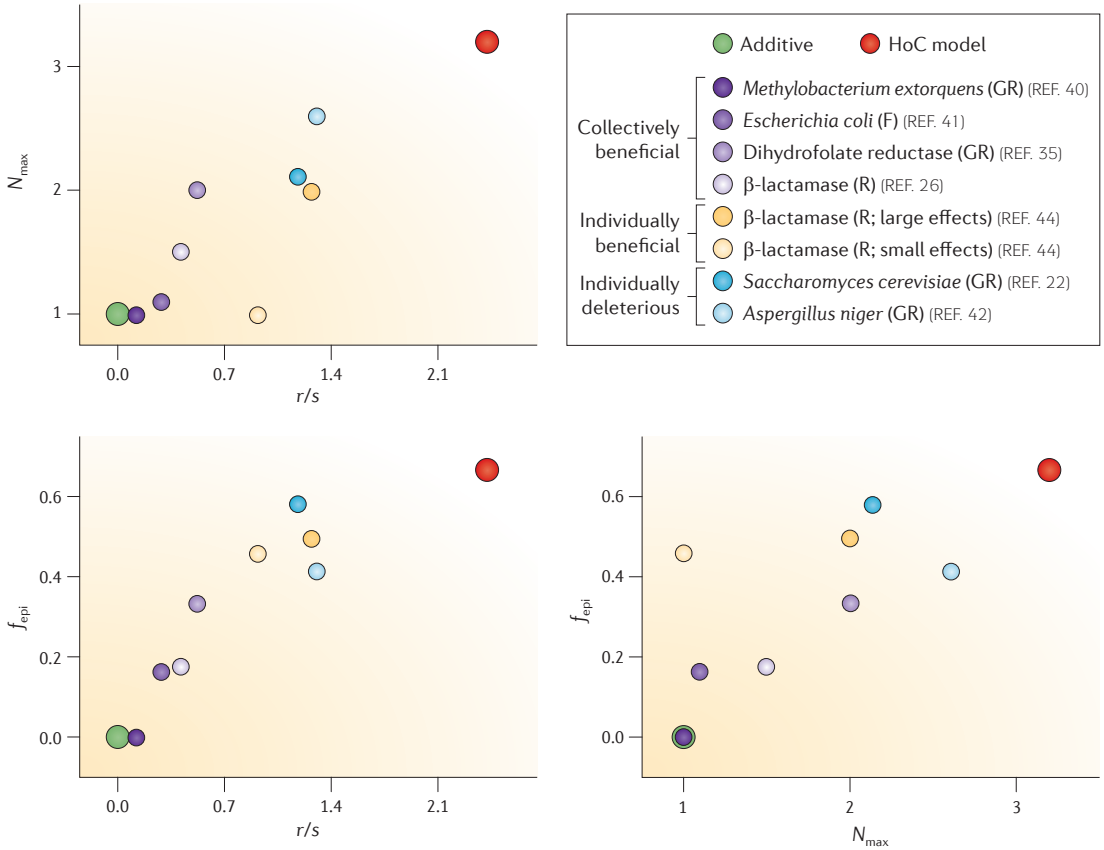

Figure 3 | **Trends in the ruggedness of empirical fitness landscapes.** Three measures that quantify the ruggedness are shown for a subset of eight available fitness landscapes: the number of fitness maxima (N_{max}), the fraction of mutation pairs with sign epistasis or reciprocal sign epistasis (f_{epi}) and the roughness/slope ratio (r/s) (BOX 1). For landscapes of different sizes to be comparable, all measures were calculated for subgraphs of size four[10]. Each plot compares two ruggedness measures for the eight landscapes, which belong to one of three classes with respect to the type of mutations involved: collectively beneficial mutations, individually beneficial mutations or individually deleterious mutations. Currently, no landscapes exist for collectively deleterious mutations. The 4–8 mutations of each landscape affect either a single gene or multiple genes in a range of microbial species, and fitness (F), growth rate (GR) or resistance (R) is measured. For comparison, expected values for a fully additive landscape (green) and a maximally rugged landscape (red; which is represented by the house-of-cards (HoC) model) are shown. Although the three measures capture different types of epistasis, they correlate reasonably well. The available empirical fitness landscapes show considerable ruggedness, especially if the combined fitness effect of mutations is unknown.

maximum exist with a probability of one for large genotypic dimensions[64]. Although this is consistent with the intuition that landscapes with less sign epistasis should be more accessible, available results for the NK model suggest that the architecture of epistatic interactions may be at least as important for evolutionary accessibility as the overall ruggedness[65]. However, at the level of resolution that is relevant for the comparison to empirical data, different models show similar behaviour. For example, the statistics of accessible pathways in the empirical *Aspergillus niger* landscape was found to be described equally well by the RMF and NK models[42]. Moreover, all models predict that the probability for a given pathway to be accessible decreases at least exponentially with its length. Thus, even if the existence of accessible pathways becomes increasingly likely when the genotypic dimensionality is large, they make up a vanishing proportion of all pathways, and it becomes more and more difficult for the adaptive process to actually 'find' them. This is consistent with the fact that adaptive walks under SSWM dynamics are typically very short[61].

Adaptive walks
Trajectories of monomorphic populations moving through genotypic space in single mutational steps, each of which increases fitness.

Pathway repeatability. Although adaptation in the SSWM regime is restricted to selectively accessible pathways, not all accessible paths that connect the same initial and final genotypes are equally likely to be realized by evolution. The probability of a single adaptive step is given by the fixation probability of the corresponding beneficial mutation, which is normalized by the sum of the fixation probabilities of all beneficial mutations that are available for the starting genotype[61]. As the steps are independent, the weight of the entire path is obtained as the product of the probabilities of the individual steps.

On the basis of this metric, two studies[26,35] concluded that most of the weight of the pathways to drug resistance considered was concentrated on a subset of the

accessible paths, thus further increasing the retrospective predictability in these systems. In the case of the 4 mutations that confer pyrimethamine resistance on the malaria parasite, the intermediate genotypes occurring along the 3 dominant pathways (of 24 direct paths) that make up close to 90% of the statistical weight were observed in natural populations[35]. Importantly, both studies used drug resistance as a proxy for fitness, which implies that a heuristic transformation from resistance to fitness was required to convert the experimental data into pathway weights. The choice of this transformation has a major influence on the inferred predictability. Using the pairwise repeatability measure (BOX 3), it was found that the effective number of pathways in the landscape of one of the studies[26] varied between 13 and 5 depending on the scheme used to convert antibiotic resistance to fitness[66], whereas the total number of accessible paths equals 18 in this case. If the restriction to direct pathways is relaxed, then the total number of accessible pathways increases to 27, whereas the effective number of pathways remains essentially unchanged[66]. This shows that, at least in this landscape, paths with back mutations carry little statistical weight.

Evolutionary predictability and population size. The weak mutation assumption of the SSWM regime requires the mutation supply rate to be sufficiently low for mutations to appear and fix in isolation. In large microbial populations, this assumption is typically violated, as multiple beneficial clones are present simultaneously and compete for fixation in a process known as clonal interference[67]. As a consequence, single adaptive steps become more repeatable because beneficial mutations of large effect are more likely to survive competition and thus have a greater chance of fixation than the SSWM expectation. A detailed analysis of this process for two competing mutations shows that the transition from SSWM dynamics to 'greedy' adaptation, in which the mutation of larger effect is fixed with certainty, precisely coincides with the onset of clonal interference[68]. A clear experimental signature of increased repeatability of fitness trajectories in large populations was reported in one study[69], in which replicate populations of *E. coli* were evolved in a complex nutrient medium at two different population sizes. Interestingly, these experiments showed that small populations occasionally reach higher fitness levels than large ones, and this was attributed to the greater adaptive heterogeneity of small populations[68].

However, in rugged fitness landscapes, the trend towards greater determinism due to clonal interference is counteracted by the fact that large population size also facilitates the crossing of fitness valleys[70]. The key element in the valley crossing mechanism, which is often referred to as stochastic tunnelling[71], is that a small portion of the population resides in the fitness valley long enough to give rise to the escape genotype. This process reflects that the two assumptions underlying the SSWM regime are interrelated in such a way that both of them break down in very large populations: with increasing mutation supply rate, the population is no longer monomorphic, and this enables it to access genotypes that are selected against[72].

The ability to cross fitness valleys releases the restriction of evolution to follow fitness-monotonic pathways and decreases adaptive repeatability in very large populations. Quantitative analysis shows that this effect becomes appreciable when population size (N) and mutation rate (μ) satisfy $N\mu^2 \sim 1$, whereas the onset of clonal interference that is associated with the transition to greedy adaptation occurs when $N\mu \sim 1$. Thus, when μ is small, there is a range of population sizes where $N\mu^2 << 1 << N\mu$, such that adaptation is restricted to fitness-monotonic pathways that are traversed in a deterministic and greedy manner, and that repeatability is maximal (FIG. 4). This regime was discovered in a numerical study of asexual adaptation on the empirical *A. niger* fitness landscape, in which repeatability of evolutionary pathways and endpoints was quantified using the Gibbs–Shannon entropy[73] (BOX 3), but experimental confirmation of the phenomenon is so far lacking. Further increasing the population size increases the standing genetic variation in the population, and one expects adaptation to become more deterministic, as the evolutionary trajectory is determined by the fitness landscape to a large extent.

The problem of scale

Although an important first step, the available data offer extremely coarse (that is, low-dimensional) and biased glimpses of the corresponding fitness landscapes. The information they contain rests on a handful of mutations that collectively or individually survived the 'sieve' of selection, or that were chosen for their observable phenotypic effect. However, mutations with unknown phenotypes and opposite fitness effects in the ancestral background also contribute during adaptation[6,10,24], and populations are subject to many more mutations than those that are eventually substituted[18,72]. To understand the consequences of these scale problems for evolutionary predictability, we need unbiased empirical fitness landscapes of realistic dimensionality. However, in this case, the limits of what can be done experimentally are rapidly met because the number of genotypes that must be analysed increases exponentially with the number of mutations included. How can we circumvent this practical barrier?

Moderate increases in scale may be expected from the application of new high-throughput technologies for the synthesis and functional analysis of large populations of small DNA or RNA molecules. For example, one study[74] synthesized all 4^{10} (that is, >1 million) 10-nucleotide DNA oligomers and measured their affinity for a fluorescently labelled allophycocyanin protein using highly parallel on-chip assays. The resulting affinity landscape was rugged with many local maxima. Another study[75] analysed the fitness of ~10^7 mutant RNA ligase ribozymes (that is, ~10^{-20} of all 4^{45} sequences) by using deep sequencing to estimate frequency changes during selection. A more recent study[76] assessed the binding affinity to GTP of >99.99% of all (~3×10^{14}) possible 24-nucleotide RNA molecules that survived selection and identified dozens of isolated fitness peaks with a fairly uniform distribution in sequence space. At these scales, a complete analysis of sequence space becomes

'Greedy' adaptation
An adaptive walk in which the available mutation of largest effect is fixed in each step.

Stochastic tunnelling
A mechanism for the crossing of fitness 'valleys', in which the escape genotype arises by mutation from a small valley population. This mechanism is different from that proposed by Wright for crossing valleys through the fixation of deleterious mutations, which happens only under weak selection.

clearly unfeasible, and information about the structure of the fitness landscape has to be extracted from sparse genotype samples that are possibly biased.

A systematic approach to this problem can be based on the decomposition of the fitness landscape into epistatic interactions of different order (BOX 1). For example, if epistatic effects are mainly due to pairwise interactions, then the number of coefficients in the expansion that need to be determined increases only quadratically rather than exponentially with the number of loci. This approach was used to analyse the *in vitro* fitness of 70,081 HIV sequences that covered 1,859 alleles at 404 amino acid positions[77]. As the total number of pairwise combinations of variants in this case much exceeds the number of samples, a regularization procedure has to be implemented to control overfitting. It was concluded that the incorporation of pairwise interactions improves the power to predict the fitness of individual genotypes, whereas the inclusion of higher-order interactions does not. Statistical exploration of the resulting landscape indicates that it is extremely rugged and has >25,000 local optima, but fitness correlations decay slowly on a scale that is comparable to the number of loci[78].

In a related study[79], the researchers analysed a previously obtained large-scale data set[80] for the transcriptional activity of 129,000 sequences of a 75-nucleotide region of the *E. coli lac* promoter. The landscape is essentially single peaked, and <3% of the pairwise interactions are non-zero. An important insight from this work is that fitness landscapes inferred in this way are generally subject to two counteracting biases: whereas regularization against overfitting reduces the apparent amount of epistasis, the extrapolation to genotypes that are far from the sampled region can introduce spurious ruggedness.

These exploratory studies tentatively suggest that fitness landscapes remain multipeaked at large scales, but the density of optima is much lower than in the small-scale examples discussed. Such a trend is, to some extent, intuitive because the inclusion of more loci requires more conditions to be satisfied for a genotype to remain a local fitness maximum, and hence optima become scarcer. This argument was given by Fisher in the 1930s in his correspondence with Wright[81] and is consistent with predictions from random field models (BOX 2). However, the consequences for the accessibility and predictability of mutational pathways on large-scale fitness landscapes are not obvious because pathway statistics do not always correlate simply with the density of fitness peaks[65].

Outlook

The recent empirical work on fitness landscapes is changing our understanding of the causes and constraints of evolution. At the small genomic scales considered so far, it is observed that sign epistasis is common, which reduces the number of accessible mutational pathways and leads to rugged landscapes with multiple fitness peaks. This has the following implications. First, population divergence and speciation may happen even in sympatry[5]. Second, new models may be needed to understand the evolutionary importance of sex and recombination in the presence of sign epistasis[50,82].

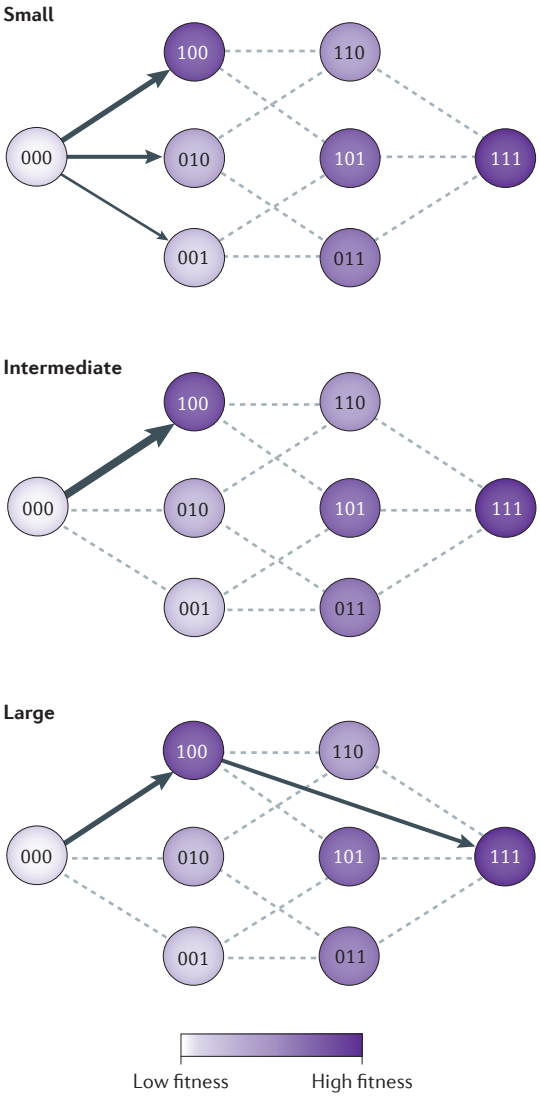

Figure 4 | **Evolutionary predictability is affected by population size.** Mutational trajectories are shown for populations of small, intermediate and large sizes on a rugged three-locus fitness landscape with global maximum 111 and local maximum 100. Nodes represent genotypes (in which 0 and 1 indicate the absence and presence of each of three mutations, respectively), and edges connect genotypes that differ in a single mutation. Arrows show mutational trajectories, which start from the wild type (000) and are realized in a limited time period that is sufficient for the fixation of a single mutation in populations of small or intermediate sizes. The width of the arrows indicates the repeatability of trajectories, which is a measure of evolutionary predictability. The small population is in the strong-selection–weak-mutation (SSWM) regime, in which different trajectories are realized owing to the chance occurrence and subsequent fixation of mutations, and predictability is therefore low. At intermediate population size, clonal interference causes the preferential fixation of the mutation with the largest benefit among the three available mutations, thereby maximizing predictability. In even larger populations, multiple mutations may sometimes fix simultaneously, which allows 'valley crossing' and decreases predictability.

Box 3 | **Measures for the predictability of mutational pathways**

Following earlier considerations of single mutational steps[109], one study[66] introduced a measure for pathway repeatability based on the strong-selection–weak-mutation (SSWM) path weights. Denoting the weight of the i^{th} path by W_i, the probability of observing this path twice in two replicate experiments is W_i^2 and, correspondingly, the sum $P_2 = \Sigma W_i^2$ is the probability of observing any pathway twice. This quantity is a simple metric for pathway repeatability that varies between $P_2 = 1$ for a single pathway and $P_2 = 1/n$ for n pathways that are equally likely. Thus, its inverse — $1/P_2$ — can be interpreted as the effective number of pathways that contribute to the process. A natural generalization of P_2 is the probability of observing the same pathway in $k > 2$ replicate experiments — $P_k = \Sigma W_i^k$, which similarly varies between $P_k = 1$ and $P_k = 1/(n^{k-1})$.

Another commonly used measure of pathway repeatability is the Gibbs–Shannon entropy of the distribution of path weights, which is defined as $S = -\Sigma W_i \ln(W_i)$ (REFS 73,103). The entropy takes on its maximal value $S = \ln(n)$ for n equally probable pathways and vanishes when all the weight is concentrated on a single path.

A recent study[102] introduced a refined measure of pathway predictability that takes into account the similarity between different accessible paths. The mean pathway divergence (D) is defined as $D = \Sigma W_i W_j d(i,j)$, where the distance $d(i,j)$ between two paths — i and j — is the average of the shortest Hamming distances from each point on path i to any point on path j.

The measures P_k and S can be used to quantify the repeatability of any replicate experiment in which the outcome belongs to a discrete set, for example, the genotypic endpoint of evolution, but they cannot be applied to continuous phenotypes such as fitness. To quantify the repeatability of fitness trajectories, one therefore has to resort to other measures of variation, such as the fitness variance or coefficient of variation, which are either evaluated at specific time points or averaged over the trajectory[69].

Third, sign epistasis alters the relative importance of genetic drift versus selection because drift may enhance long-term rates of adaptation on rugged fitness landscapes[8,69,83]. Finally and perhaps most excitingly, it implies that evolution has inherent predictability[6,7,26,51,73], although the implications for predictability have so far been explored exclusively in a retrospective way.

Understanding how evolutionary predictability is determined by the topography of fitness landscapes requires further empirical research. More data sets of the kinds discussed here are needed to confirm suggested causes of observed variation in topographies, such as the greater ruggedness of landscapes based on mutations of unknown collective effect, mutations of large effect and mutations that occur in the same gene. Empirical fitness landscapes are particularly desired for other model systems (for example, other proteins) and for classes of mutations for which we currently have no data, including collectively deleterious mutations, deleterious mutations in the same gene and individually beneficial mutations in different genes. By analysing mutations observed in different evolutionary lineages, the analysis of predictability can be extended from assessing the probability of mutational order in a single lineage towards assessing the likelihood of alternative trajectories. As indicated above, larger-scale empirical landscapes are especially in demand. It will be essential to develop systematic high-throughput methods to generate mutants[74,76,84] and to analyse their fitness[75,84] with sufficient replication to deal with the growing problem of measurement noise in large-scale analyses. Such data will allow analyses of the role of scale in determining pathway accessibility and predictability by considering these parameters in subgraphs of different sizes[42].

Theoretical work should complement the empirical developments at several fronts. Models that integrate specific topographical information from empirical fitness landscapes into simple statistical models (such as the NK and RMF models) are needed, and the statistical problems associated with inferring large-scale landscapes from sparse samples should be addressed[79]. Information from functional pathways[46,85,86] and the physicochemical properties of biomolecules[36,87–91] may be used to devise realistic genotype–fitness maps from first principles, for example, by inferring fitness from the stability, abundance and interactions between the molecules involved. Importantly, the apparent mismatch between intuitions gained from rather simple sequence–structure models with their extended neutral networks on the one hand[5,87,88] and the rugged landscapes obtained in empirical studies on the other hand needs to be better understood and ultimately resolved[83].

Finally, empirical fitness landscapes have the potential to tackle real-world problems. For example, common variants of avian A (H5N1) influenza virus require 3–5 amino acid substitutions to become transmissible via respiratory droplets between mammals. One study[92] modelled the potential of the virus to evolve into the transmissible type under different assumptions about the epistatic interactions among the substitutions. Although the topography of the actual fitness landscape is unknown in this case, the relevance of epistasis in influenza evolution has been established empirically[93]. Similarly, empirical fitness landscapes may help to find drug therapies that minimize the problem of antibiotic resistance evolution[34,94,95] and guide immunogen design for the treatment of HIV[96]. We expect that this approach may be helpful in other areas in which evolution of pathogens or cancer cells poses problems to therapies.

To summarize, more than 80 years after the inception of Wright's seminal concept, his intuition about the ruggedness of fitness landscapes has been forcefully vindicated. However, the task of exploring its consequences for the evolutionary processes that unfold across the scales of biological organization has only begun to be addressed.

1. Lehner, B. Genotype to phenotype: lessons from model organisms for human genetics. *Nature Rev. Genet.* **14**, 168–178 (2013).
2. Wagner, G. P. & Zhang, J. The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. *Nature Rev. Genet.* **12**, 204–213 (2011).
3. Phillips, P. C. Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Rev. Genet.* **9**, 855–867 (2008).
4. de Visser, J. A. G. M., Cooper, T. F. & Elena, S. F. The causes of epistasis. *Proc. R. Soc. B* **278**, 3617–3624 (2011).
5. Gavrilets, S. *Fitness Landscapes and the Origin of Species* (Princeton Univ. Press, 2004).
6. Achaz, G., Rodriguez-Verdugo, A., Gaut, B. S. & Tenaillon, O. The reproducibility of adaptation in the light of experimental evolution with whole genome sequencing. *Adv. Exp. Med. Biol.* **781**, 211–231 (2014).
7. Lobkovsky, A. E. & Koonin, E. V. Replaying the tape of life: quantification of the predictability of evolution. *Frontiers Genet.* **3**, 246 (2012).
8. Wright, S. The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proc. 6th Int. Congress Genet.* **1**, 356–366 (1932).
This paper introduces the concept of the fitness landscape as a key component of Wright's shifting balance theory.
9. Colegrave, N. & Buckling, A. Microbial experiments on adaptive landscapes. *BioEssays* **27**, 1167–1173 (2005).
10. Szendro, I. G., Schenk, M. F., Franke, J., Krug, J. & de Visser, J. A. G. M. Quantitative analyses of empirical fitness landscapes. *J. Stat. Mech.* P01005 (2013).
11. Wright, S. Evolution in Mendelian populations. *Genetics* **16**, 97–159 (1931).
12. Haldane, J. B. S. A mathematical theory of natural selection. Part VIII. Metastable populations. *Proc. Cambridge Philos. Soc.* **27**, 137–142 (1931).
13. Maynard Smith, J. Natural selection and the concept of a protein space. *Nature* **225**, 563–564 (1970).
This study presents the realization that genotypic space is discrete and that mutational pathways are only accessible when they pass through functional genotypes.
14. Kauffman, S. A. & Levin, S. Towards a general theory of adaptive walks on rugged landscapes. *J. Theor. Biol.* **128**, 11–45 (1987).
This is the first mathematical exploration of random fitness landscapes and their consequences for adaptation.
15. Kauffman, S. A. & Weinberger, E. D. The NK model of rugged fitness landscapes and its application to the maturation of the immune response. *J. Theor. Biol.* **141**, 211–245 (1989).
16. Harms, M. J. & Thornton, J. W. Evolutionary biochemistry: revealing the historical and physical causes of protein properties. *Nature Rev. Genet.* **14**, 559–571 (2013).
17. Malcolm, B. A., Wilson, K. P., Matthews, B. W., Kirsch, J. F. & Wilson, A. C. Ancestral lysozymes reconstructed, neutrality tested, and thermostability linked to hydrocarbon packing. *Nature* **345**, 86–89 (1990).
This is the first empirical analysis of a three-locus fitness landscape of lysozymes in game birds.
18. Barrick, J. E. & Lenski, R. E. Genome dynamics during experimental evolution. *Nature Rev. Genet.* **14**, 827–839 (2013).
19. Kondrashov, A. S. Deleterious mutations and the evolution of sexual reproduction. *Nature* **336**, 435–440 (1988).
20. Kouyos, R. D., Silander, O. K. & Bonhoeffer, S. Epistasis between deleterious mutations and the evolution of recombination. *Trends Ecol. Evol.* **22**, 308–315 (2007).
21. de Visser, J. A. G. M., Hoekstra, R. F. & van den Ende, H. Test of interaction between genetic markers that affect fitness in *Aspergillus niger*. *Evolution* **51**, 1499–1505 (1997).
22. Hall, D. W., Agan, M. & Pope, S. C. Fitness epistasis among 6 biosynthetic loci in the budding yeast *Saccharomyces cerevisiae*. *J. Hered.* **101**, S75–S84 (2010).
23. Kondrashov, F. A. & Kondrashov, A. S. Multidimensional epistasis and the disadvantage of sex. *Proc. Natl Acad. Sci. USA* **98**, 12089–12092 (2001).
24. Weinreich, D. M., Watson, R. A. & Chao, L. Perspective: sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* **59**, 1165–1174 (2005).
This paper formally introduces the concept of sign epistasis and proves its equivalence with limited pathway accessibility.
25. Poelwijk, F. J., Tanase-Nicola, S., Kiviet, D. J. & Tans, S. J. Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. *J. Theor. Biol.* **272**, 141–144 (2011).
26. Weinreich, D. M., Delaney, N. F., DePristo, M. A. & Hartl, D. L. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111–114 (2006).
This seminal study shows how sign epistasis limits the number of accessible trajectories on a five-locus fitness landscape of β-lactamase.
27. Weinreich, D. M., Lan, Y., Wylie, S. C. & Heckendorn, R. B. Should evolutionary geneticists worry about higher-order epistasis? *Curr. Opin. Genet. Dev.* **23**, 700–707 (2013).
28. O'Maille, P. E. *et al.* Quantitative exploration of the catalytic landscape separating divergent plant sesquiterpene synthases. *Nature Chem. Biol.* **4**, 617–623 (2008).
29. Lee, Y.-H., Dsouza, L. M. & Fox, G. E. Equally parsimonious pathways through an RNA sequence space are not equally likely. *J. Mol. Evol.* **45**, 278–284 (1997).
30. Aita, T., Iwakura, M. & Husimi, Y. A cross-section of the fitness landscape of dihydrofolate reductase. *Protein Engineer.* **14**, 633–638 (2001).
31. Bridgham, J. T., Carroll, S. M. & Thornton, J. W. Evolution of hormone-receptor complexity by molecular exploitation. *Science* **312**, 97–101 (2006).
32. Brown, K. M. *et al.* Compensatory mutations restore fitness during the evolution of dihydrofolate reductase. *Mol. Biol. Evol.* **27**, 2682–2690 (2010).
33. da Silva, J., Coetzer, M., Nedellec, R., Pastore, C. & Mosier, D. E. Fitness epistasis and constraints on adaptation in a human immunodeficiency virus type 1 protein region. *Genetics* **185**, 293–303 (2010).
34. Goulart, C. P. *et al.* Designing antibiotic cycling strategies by determining and understanding local adaptive landscapes. *PLoS ONE* **8**, e56040 (2013).
35. Lozovsky, E. R. *et al.* Stepwise acquisition of pyrimethamine resistance in the malaria parasite. *Proc. Natl Acad. Sci. USA* **106**, 12015–12030 (2009).
36. Lunzer, M., Miller, S. P., Felsheim, R. & Dean, A. M. The biochemical architecture of an ancient adaptive landscape. *Science* **310**, 499–501 (2005).
This study reconstructs a fitness landscape by analysing enzyme function as a phenotype that links genotype and fitness.
37. Novais, A. *et al.* Evolutionary trajectories of β-lactamase CTX-M-1 cluster enzymes: predicting antibiotic resistance. *PLoS Pathog.* **6**, e1000735 (2010).
38. Tan, L., Serene, S., Chao, H. X. & Gore, J. Hidden randomness between fitness landscapes limits reverse evolution. *Phys. Rev. Lett.* **106**, 198102 (2011).
39. de Vos, M. G. J., Poelwijk, F. J., Battich, N., Ndika, J. D. T. & Tans, S. J. Environmental dependence of genetic constraint. *PLoS Genet.* **9**, e1003580 (2013).
40. Chou, H.-H., Chiu, H.-C., Delaney, N. F., Segrè, D. & Marx, C. J. Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* **332**, 1190–1192 (2011).
41. Khan, A. I., Dinh, D. M., Schneider, D., Lenski, R. E. & Cooper, T. F. Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* **332**, 1193–1196 (2011).
42. Franke, J., Klözer, A., de Visser, J. A. G. M. & Krug, J. Evolutionary accessibility of mutational pathways. *PLoS Computat. Biol.* **7**, e1002134 (2011).
43. Whitlock, M. C. & Bourguet, D. Factors affecting the genetic load in *Drosophila*: synergistic epistasis and correlations among fitness components. *Evolution* **54**, 1654–1660 (2000).
44. Schenk, M. F., Szendro, I. G., Salverda, M. L. M., Krug, J. & de Visser, J. A. G. M. Patterns of epistasis between beneficial mutations in an antibiotic resistance gene. *Mol. Biol. Evol.* **30**, 1779–1787 (2013).
45. Draghi, J. A. & Plotkin, J. B. Selection biases the prevalence and type of epistasis along adaptive trajectories. *Evolution* **67**, 3120–3131 (2013).
46. Pumir, A. & Shraiman, B. Epistasis in a model of molecular signal transduction. *PLoS Comput. Biol.* **7**, e1001134 (2011).
47. Wilke, C. O. & Adami, C. Interaction between directional epistasis and average mutational effects. *Proc. R. Soc. B* **268**, 1469–1474 (2001).
48. You, L. & Yin, J. Dependence of epistasis on environment and mutation severity as revealed by *in silico* mutagenesis of phage T7. *Genetics* **160**, 1273–1281 (2002).
49. DePristo, M. A., Weinreich, D. M. & Hartl, D. L. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nature Rev. Genet.* **6**, 678–687 (2005).
50. Watson, R. A., Weinreich, D. M. & Wakeley, J. Genome structure and the benefits of sex. *Evolution* **65**, 523–536 (2010).
51. Conway Morris, S. *Life's Solution: Inevitable Humans in a Lonely Universe* (Cambridge Univ. Press, 2003).
52. Gould, S. J. *Wonderful Life: The Burgess Shale and the Nature of History* (W. W. Norton & Company, 1989).
53. Lang, G. I. *et al.* Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature* **500**, 571–574 (2013).
54. Tenaillon, O. *et al.* The molecular diversity of adaptive convergence. *Science* **335**, 457–461 (2012).
55. Woods, R., Schneider, D., Winkworth, C. L., Riley, M. A. & Lenski, R. E. Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **103**, 9107–9112 (2006).
56. Blount, Z. D., Barrick, J. E., Davidson, C. J. & Lenski, R. E. Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature* **489**, 513–518 (2012).
57. Salverda, M. L. M. *et al.* Initial mutations direct alternative pathways of protein evolution. *PLoS Genet.* **7**, e1001321 (2011).
58. Papp, B., Notebaart, R. A. & Pál, C. Systems-biology approaches for predicting genomic evolution. *Nature Rev. Genet.* **12**, 591–602 (2011).
59. Gerrish, P. J. & Sniegowski, P. D. Real time forecasting of near-future evolution. *J. R. Soc. Interface* **9**, 2268–2278 (2012).
60. Gillespie, J. H. Some properties of finite populations experiencing strong selection and weak mutation. *Am. Naturalist* **121**, 691–708 (1983).
61. Orr, H. A. The genetic theory of adaptation: a brief history. *Nature Rev. Genet.* **6**, 119–127 (2005).
62. Crona, K., Greene, D. & Barlow, M. The peaks and geometry of fitness landscapes. *J. Theor. Biol.* **317**, 1–10 (2013).
63. Whitlock, M. C., Phillips, P. C., Moore, F. B.-G. & Tonsor, S. J. Multiple fitness peaks and epistasis. *Annu. Rev. Ecol. Systemat.* **26**, 601–629 (1995).
64. Hegarty, P. & Martinsson, A. On the existence of accessible paths in various models of fitness landscapes. *Ann. Appl. Prob.* (in the press).
65. Schmiegelt, B. & Krug, J. Evolutionary accessibility of modular fitness landscapes. *J. Statist. Phys.* **154**, 334–355 (2014).
66. Roy, S. W. Probing evolutionary repeatability: neutral and double changes and the predictability of evolutionary adaptation. *PLoS ONE* **4**, e4500 (2009).
67. Gerrish, P. J. & Lenski, R. E. The fate of competing beneficial mutations in an asexual population. *Genetica* **102–103**, 127–144 (1998).
68. Jain, K., Krug, J. & Park, S.-C. Evolutionary advantage of small populations on complex fitness landscapes. *Evolution* **65**, 1945–1955 (2011).
69. Rozen, D. E., Habets, M. G. J. L., Handel, A. & de Visser, J. A. G. M. Heterogeneous adaptive trajectories of small populations on complex fitness landscapes. *PLoS ONE* **3**, e1715 (2008).
70. Weissman, D. B., Desai, M. M., Fisher, D. S. & Feldman, M. W. The rate at which asexual populations cross fitness valleys. *Theor. Popul. Biol.* **75**, 286–300 (2009).
71. Isawa, Y., Michor, F. & Nowak, M. A. Stochastic tunnels in evolutionary dynamics. *Genetics* **166**, 1571–1579 (2004).
72. Woods, R. J. *et al.* Second-order selection for evolvability in a large *Escherichia coli* population. *Science* **331**, 1433–1436 (2011).
This study experimentally shows the combined influence of epistasis and population dynamics on the outcome of evolution.
73. Szendro, I. G., Franke, J., de Visser, J. A. G. M. & Krug, J. Predictability of evolution depends non-monotonically on population size. *Proc. Natl Acad. Sci. USA* **110**, 571–576 (2013).

74. Rowe, W. *et al.* Analysis of a complete DNA–protein affinity landscape. *J. R. Soc. Interface* **7**, 397–408 (2010).
75. Pitt, J. N. & Ferré-D'Amaré, A. R. Rapid construction of empirical RNA fitness landscapes. *Science* **330**, 376–379 (2010).
76. Jiménez, J. I., Xulvi-Brunet, R., Campbell, G. W., Turk-MacLeod, R. & Chen, I. A. Comprehensive experimental fitness landscape and evolutionary network for small RNA. *Proc. Natl Acad. Sci.* **110**, 14984–14989 (2013).
This is an empirical analysis of the largest fitness landscape so far and involves > 10^{14} RNA molecules.
77. Hinkley, T. *et al.* A systems analysis of mutational effects in HIV-1 protease and reverse transcriptase. *Nature Genet.* **43**, 487–490 (2011).
This paper presents an early empirical fitness landscape of large dimensions for HIV-1 with fitness predictions for the many missing genotypes.
78. Kouyos, R. D. *et al.* Exploring the complexity of the HIV-1 fitness landscape. *PLoS Genet.* **8**, e1002551 (2012).
79. Otwinowski, J. & Nemenman, I. Genotype to phenotype mapping and the fitness landscape of the *E. coli lac* promoter. *PLoS ONE* **8**, e61570 (2013).
80. Kinney, J. B., Murugan, A., Callan, C. G. & Cox, E. C. Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. *Proc. Natl Acad. Sci.* **107**, 9158–9163 (2010).
81. Provine, W. B. *Sewall Wright and Evolutionary Biology* (Chicago Univ. Press, 1986).
82. de Visser, J. A. G. M., Park, S.-C. & Krug, J. Exploring the effect of sex on empirical fitness landscapes. *Am. Naturalist* **174**, S15–S30 (2009).
83. Wagner, A. Neutralism and selectionism: a network-based reconciliation. *Nature Rev. Genet.* **9**, 965–974 (2008).
84. Hietpas, R. T., Jensen, J. D. & Bolona, D. N. Experimental illumination of a fitness landscape. *Proc. Natl Acad. Sci. USA* **108**, 7896–7901 (2011).
85. Heckmann, D. *et al.* Predicting C_4 photosynthesis evolution: modular, individually adaptive steps on a Mount Fuji fitness landscape. *Cell* **153**, 1579–1588 (2013).
86. Perfeito, L., Ghozzi, S., Berg, J., Schnetz, K. & Lässig, M. Nonlinear fitness landscape of a molecular pathway. *PLoS Genet.* **7**, e1002160 (2011).
87. Chan, H. S. & Bornberg-Bauer, E. Perspectives on protein evolution from simple exact models. *Appl. Bioinformat.* **1**, 121–144 (2002).
88. Schuster, P. Prediction of RNA secondary structures: from theory to models and real molecules. *Rep. Progress Phys.* **69**, 1419–1477 (2006).
89. Mustonen, V., Kinney, J., Callan, C. G. & Lässig, M. Energy-dependent fitness: a quantitative model for the evolution of yeast transcription factor binding sites. *Proc. Natl Acad. Sci.* **105**, 12376–12381 (2008).
90. Heo, M., Kang, L. & Shakhnovich, E. I. Emergence of species in evolutionary "simulated annealing". *Proc. Natl Acad. Sci. USA* **106**, 1869–1874 (2009).
91. Wylie, S. C. & Shakhnovich, E. I. A biophysical protein folding model accounts for most mutational fitness effects in viruses. *Proc. Natl Acad. Sci. USA* **108**, 9916–9921 (2011).
92. Russell, C. A. *et al.* The potential for respiratory droplet-transmissible A/H5N1 influenza virus to evolve in a mammalian host. *Science* **336**, 1541–1547 (2012).
93. Gong, L. I., Suchard, M. A. & Bloom, J. D. Stability-mediated epistasis constrains the evolution of an influenza protein. *eLife* **2**, e00631 (2013).
94. Hall, B. G. Predicting evolution by *in vitro* evolution requires determining evolutionary pathways. *Antimicrob. Agents Chemother.* **46**, 3035–3038 (2002).
95. Palmer, A. C. & Kishony, R. Understanding, predicting and manipulating the genotypic evolution of antibiotic resistance. *Nature Rev. Genet.* **14**, 243–248 (2013).
96. Ferguson, Andrew, L. *et al.* Translating HIV sequences into quantitative fitness landscapes predicts viral vulnerabilities for rational immunogen design. *Immunity* **38**, 606–617 (2013).
97. Hansen, T. F. & Wagner, G. P. Modeling genetic architecture: a multilinear theory of gene interaction. *Theor. Popul. Biol.* **59**, 61–86 (2001).
98. Neher, R. A. & Shraiman, B. I. Statistical genetics and evolution of quantitative traits. *Rev. Modern Phys.* **83**, 1283–1300 (2011).
99. Stadler, P. F. & Happel, R. Random field models of fitness landscapes. *J. Math. Biol.* **38**, 435–478 (1999).
100. Neidhart, J., Szendro, I. G. & Krug, J. Exact results for amplitude spectra of fitness landscapes. *J. Theor. Biol.* **332**, 218–227 (2013).
101. Kingman, J. F. C. A simple model for the balance between selection and mutation. *J. Appl. Probabil.* **15**, 1–12 (1978).
102. Lobkovsky, A. E., Wolf, Y. I. & Koonin, E. V. Predictability of evolutionary trajectories in fitness landscapes. *PLoS Comput. Biol.* **7**, e1002302 (2011).
103. Palmer, M. E., Moudgil, A. & Feldman, M. W. Long-term evolution is surprisingly predictable in lattice proteins. *J. R. Soc. Interface* **10**, 20130026 (2013).
104. Ferrada, E. & Wagner, A. A comparison of genotype-phenotype maps for RNA and proteins. *Biophys. J.* **102**, 1916–1925 (2012).
105. Martin, G., Elena, S. F. & Lenormand, T. Distributions of epistasis in microbes fit predictions from a fitness landscape model. *Nature Genet.* **33**, 555–560 (2007).
106. Rokyta, D. R. *et al.* Epistasis between beneficial mutations and the phenotype-to-fitness map for a ssDNA virus. *PLoS Genet.* **7**, e1002075 (2011).
107. Pearson, V. M., Miller, C. R. & Rokyta, D. R. The consistency of beneficial fitness effects of mutations across diverse genetic backgrounds. *PLoS ONE* **7**, e43864 (2012).
108. Chou, H.-H., Delaney, N. F., Draghi, J. A. & Marx, C. J. Mapping the fitness landscape of gene expression uncovers the cause of antagonism and sign epistasis between adaptive mutations. *PLoS Genet.* **10**, e1004149 (2014).
109. Orr, H. A. The probability of parallel evolution. *Evolution* **59**, 216–220 (2005).

Acknowledgements
The authors thank I. Szendro for assistance with figure 3. They thank D. Weinreich and G. Achaz for sharing subsequently published manuscripts during the preparation phase of this Review, and four anonymous reviewers for constructive comments. This work was supported by Deutsche Forschungsgemeinschaft within SFB 680 "Molecular Basis of Evolutionary Innovation" and SPP 1590 "Probabilistic Structures in Evolution".

Competing interests statement
The authors declare no competing interests.

FURTHER INFORMATION
Magellan: Maps of genetical landscapes: http://wwwabi.snv.jussieu.fr/public/magellan/

ALL LINKS ARE ACTIVE IN THE ONLINE PDF