

ADAPTIVE LANDSCAPES ACROSS MULTIPLE SCALES



Malvika Srivastava
Doctoral Thesis Number: 30527

MALVIKA SRIVASTAVA

ADAPTIVE LANDSCAPES ACROSS MULTIPLE SCALES

Cover page artwork credits: Rinki Kumari. The cover page artwork depicts a Madhubani painting of the tree of life with DNA at its core. The art form originates from the Madhubani district in my home state of Bihar, India. It is traditionally painted using twigs, brushes and matchsticks along with natural pigments. It has been passed down through generations and is primarily practiced by rural women. Many thanks to my grandmother for getting this made on request.

DISS. ETH NO. 30527

ADAPTIVE LANDSCAPES ACROSS MULTIPLE SCALES

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES

(Dr. sc. ETH Zurich)

presented by

MALVIKA SRIVASTAVA

M.Sc., Universität zu Köln

born on 29th September 1995

accepted on the recommendation of

Dr. Joshua L. Payne, examiner

Prof. Dr. Ard A. Louis, co-examiner

Prof. Dr. Alex R. Hall, co-examiner

Prof. Dr. Claudia Bank, co-examiner

2024

To my parents and grandparents

SUMMARY

Visualizing evolution as a hill-climbing process on an *adaptive landscape* – which maps genotypes to their reproductive capability or *fitness* – has emerged as a valuable tool for studying evolution. The concept of an adaptive landscape itself has evolved from being a mere schematic diagram to facilitating directed evolution experiments and predicting the evolution of antibiotic resistance and the future evolution of SARS-CoV-2. This versatile tool allows for the study of evolution across various scales, by either mapping genotypes directly to fitness in the form of *fitness landscapes* or to intermediate phenotypes through *genotype-phenotype maps*. Recent technological advancements, such as high-throughput sequencing, have reignited interest in the field, by enabling the measurement of large empirical adaptive landscapes.

Despite considerable progress in our theoretical understanding of adaptive landscapes, many open questions remain. In this thesis, I address three such questions, at three levels of biological organisation – the level of DNA, the level of proteins and the level of development. In answering these questions, we not only gained theoretical insights about adaptive landscapes, but also biological insights about evolution at these different scales.

At the level of DNA, adaptive landscapes have been used to study the evolution of transcription factor binding sites (TFBS) of various eukaryotic transcription factors. Since it is difficult to quantify and measure the fitness of a genotype, a phenotype, such as binding affinity, is measured as a proxy for fitness. When there is selection for low or intermediate values of the phenotype, the genotype-phenotype map and the fitness landscape can look very different. In Chapter 2, we systematically quantified this *incongruence* between genotype-phenotype maps and fitness landscapes. On the theoretical front, we found a special genotype-phenotype map that would remain invariant under selection for low or intermediate phenotypic values and concluded that correlated genotype-phenotype maps would be more incongruent to their corresponding fitness landscapes, than uncorrelated genotype-phenotype maps. This knowledge provides intuition about when a genotype-phenotype map is a good approximation for the fitness landscape. On the biological front, we applied our findings to genotype-phenotype maps of TF-DNA interactions. Here, we found that selection for intermediate binding affinity in more than a 1000 genotype-phenotype maps aids adaptation, despite making the landscapes more rugged. In other words, we found an evolutionary benefit of low affinity binding sites, that could potentially explain their prevalence in the regulatory portfolios of a diversity of organisms such as bacteria, yeast and fly.

At the level of proteins, adaptive landscapes have been used to study the evolution of protein function. On a more fundamental level, it is not understood why there

are 20 standard amino acids. The number of allowed amino acids at each site in the protein sequence serves as a means of adding more dimensions to the fitness landscape. In Chapter 3, we were interested in understanding the evolutionary consequences of increasing the dimensionality of fitness landscapes, by changing the number of alleles at each site, or the *alphabet cardinality*. We concluded that the effect of increasing the cardinality on adaptive evolution depends on the correlations in the fitness landscape being considered – while it is detrimental to increase the cardinality in correlated fitness landscapes, it is beneficial to do so in uncorrelated fitness landscapes. This result also has implications for the general field of optimisation. We also applied our findings to three combinatorially complete empirical protein fitness landscapes with intermediate levels of correlations. For these landscapes, increasing the cardinality does not affect adaptive evolution significantly. We therefore concluded that topographical changes in the landscape cannot alone explain the expansion of the amino acid alphabet.

Finally, at the level of development, we were interested in studying the probability of emergence of complex phenotypes, such as the peacock's tail, from simple phenotypes through natural selection. To this end, in Chapter 4, we examined the genotype-phenotype map resulting from a toy model of development, namely Richard Dawkins' Biomorphs. We studied the *navigability* of Biomorphs fitness landscapes i.e. the prevalence of uphill paths in these landscapes. On the theoretical front, we established relationships between phenotypic evolvability and ruggedness, average peak height and navigability that hold for all genotype-phenotype maps. On the biological front, we found that emergence of complex phenotypes is rare, despite there being uphill paths between simple source phenotypes and complex target phenotypes. However, upon expanding the set of acceptable phenotypes, by allowing for some mismatches from the target phenotype, we reported an increased probability of evolving complex phenotypes.

ZUSAMMENFASSUNG

Die Vorstellung von Evolution als Bergsteigen auf einer adaptiven Landschaft, welche Genotypen auf ihre Reproduktionsfähigkeit oder Fitness abbildet, hat sich als wertvolles Werkzeug für die Erforschung von Evolution herauskristallisiert. Das Konzept von adaptiven Landschaften entwickelte sich von einem bloßen Schema zu einem etablierten Werkzeug – welches zum Design von Evolutionsexperimenten und zur Vorhersage der Evolution von Antibiotikaresistenz und SARS-CoV-2 verwendet wird. Die Vielfältigkeit von adaptiven Landschaften ermöglicht es Evolution auf verschiedenen Maßstäben zu untersuchen, indem Genotypen entweder direkt auf die Fitness, in Form von Fitness-Landschaften, oder auf intermediäre Phänotypen, durch Genotyp-Phänotyp-Karten, abgebildet werden. Neueste technologische Fortschritte, wie z. B. Highthroughput-Sequencing, haben das Interesse an diesem Gebiet neu entfacht, da sie die Messung großer empirischer adaptiver Landschaften ermöglichen.

Trotz erheblicher Fortschritte in unserem theoretischen Verständnis adaptiver Landschaften bleiben viele Fragen offen. Diese Arbeit befasst sich mit drei dieser Fragen auf drei verschiedenen Ebenen der biologischen Organisation – der Ebene der DNA, der Ebene der Proteine, und der Ebene der Entwicklung. Dabei haben wir nicht nur neue theoretische Erkenntnisse über adaptive Landschaften gewonnen, sondern auch biologische Erkenntnisse über Evolution auf diesen verschiedenen Ebenen. Weiterhin konnten wir auch die Grenzen von adaptiven Landschaften als Werkzeug zur Untersuchung der Evolution identifizieren.

Auf der Ebene der DNA wurden adaptive Landschaften genutzt, um die Evolution von Transkriptionsfaktor-Bindungsstellen (TFBS) verschiedener eukaryontischer Transkriptionsfaktoren (TF) zu untersuchen. Da es schwierig ist die Fitness eines Genotyps direkt zu messen wurden stattdessen stellvertretend die Bindungsaffinität bzgl. der Transkriptionsfaktoren gemessen. Wenn es eine Selektion für niedrige oder mittlere Werte des Phänotyps gibt, können die Genotyp-Phänotyp-Karte und die Fitness-Landschaft sehr unterschiedlich aussehen. In Kapitel 2 haben wir diese Diskrepanz zwischen Genotyp-Phänotyp-Karten und Fitnesslandschaften systematisch quantifiziert. Auf theoretischer Ebene haben wir eine spezielle Genotyp-Phänotyp-Karte gefunden, die bei der Selektion auf niedrige oder mittlere phänotypische Werte invariant bleibt, und schlussfolgern, dass korrelierte Genotyp-Phänotyp-Karten stärker mit den entsprechenden Fitnesslandschaften inkongruent sind als unkorrelierte Genotyp-Phänotyp-Karten. Dieses Wissen gibt Aufschluss darüber, wann eine Genotyp-Phänotyp-Karte eine gute Annäherung an die Fitness-Landschaft darstellt. Auf biologischer Ebene haben wir unsere Erkenntnisse auf Genotyp-Phänotyp-Karten von TF-DNA-Interaktionen angewandt. Dabei stellten wir fest,

dass Selektion für mittlere Bindungsaffinität in mehr als 1000 Genotyp-Phänotyp-Karten die Adaption fördert, obwohl die Landschaften dadurch zerklüfteter werden. Mit anderen Worten, wir haben einen evolutionären Vorteil von Bindungsstellen mit geringer Affinität gefunden, der möglicherweise ihre Verbreitung in den regulatorischen Portfolios einer Vielzahl von Organismen wie Bakterien, Hefe und Fliegen erklären könnte.

Auf der Ebene der Proteine haben wir adaptive Landschaften verwendet, um Evolution von Proteinfunktionen zu studieren. Es ist unklar warum es genau 20 Standardamino-säuren gibt. Die Anzahl der zulässigen Aminosäuren an jeder Stelle der Proteinsequenz dient als Mittel, um der Fitnesslandschaft weitere Dimensionen hinzuzufügen. In Kapitel 3 waren wir daran interessiert, die evolutionären Folgen einer Erhöhung der Dimensionalität von Fitnesslandschaften zu verstehen, indem wir die Anzahl der Allele an jeder Stelle oder die Kardinalität des Alphabets veränderten. Wir kamen zu dem Schluss, dass der Effekt einer erhöhten Kardinalität auf die adaptive Evolution von der Beschaffenheit der Fitnesslandschaft abhängt. In korrelierten Fitnesslandschaften ist eine erhöhte Kardinalität von Nachteil, während sie in unkorrelierten Fitnesslandschaften von Vorteil ist. Dieses Ergebnis hat auch Auswirkungen auf den allgemeinen Bereich der Optimierung. Wir haben unsere Erkenntnisse auch auf drei empirische, kombinatorisch vollständige Protein-Fitnesslandschaften mit mittlerem Korrelationsgrad angewandt. Für diese Landschaften hat erhöhte Kardinalität keinen signifikanten Einfluss auf adaptive Evolution. Wir kamen daher zu dem Schluss, dass topografische Veränderungen in der Landschaft allein die Erweiterung des Aminosäurealphabets nicht erklären können.

Auf der Ebene der Entwicklung haben wir abschließend analysiert wie durch natürliche Selektion aus einfachen Phänotypen komplexe Phänotypen entstehen können. Dafür untersuchten wir in Kapitel 4 die Genotyp-Phänotyp-Karten, welche aus einem Toy-Entwicklungs-Modell hervorgehen, nämlich Richard Dawkins' Biomorphs. Wir analysierten die Beschaffenheit der Biomorphs Fitnesslandschaften, d.h. wie viele Pfade bergauf (zu höherer Fitness) es in diesen Landschaften gibt. Wir haben eine Beziehung zwischen der phänotypischen Evolvierbarkeit und der Robustheit, der durchschnittlichen Gipfelhöhe und der Navigierbarkeit herstellen können, welche für alle Genotyp-Phänotyp-Karten gilt. Auf biologischer Ebene finden wir, dass komplexe Phänotypen selten entstehen, obwohl es zwischen einfachen Ausgangsphänotypen und komplexen Zielphänotypen viele bergauf gehende Pfade gibt. Wenn wir jedoch die Menge der akzeptablen Phänotypen erweiterten, indem wir einige Abweichungen vom Zielphänotyp zuließen, fanden wir eine höhere Wahrscheinlichkeit für die Entwicklung komplexer Phänotypen.

CONTENTS

1	Introduction	1
1.1	From Darwin to Wright	1
1.2	One map to explain it all?	1
1.3	From models to data	3
1.3.1	Theoretical landscapes	4
1.3.2	Empirical landscapes	6
1.4	Evolution across multiple scales	7
1.5	Evolution in action - case study of SARS-CoV-2	11
1.6	Thesis outline	12
	References	14
2	On the incongruence of genotype-phenotype and fitness landscapes	23
2.1	Introduction	25
2.2	Results	28
2.2.1	Landscape incongruence	29
2.2.2	Phenotype-to-fitness map	29
2.2.3	Two-locus biallelic genotype-phenotype landscapes	30
2.2.4	Multi-locus biallelic genotype-phenotype landscapes	34
2.2.5	Genotype-phenotype landscapes of transcription factor-DNA interactions	37
2.2.6	Evolutionary consequences	42
2.3	Discussion	46
2.4	Methods	49
	References	53
2.5	Supplementary information	62
2.5.1	Derivations	62
2.5.2	Proofs	64
2.5.3	Notes	66
2.5.4	Supplementary Figures	70
3	Alphabet cardinality and adaptive evolution	79
3.1	Introduction	80
3.2	Results	83
3.2.1	Theoretical fitness landscapes	83
3.2.2	Empirical fitness landscapes	88
3.3	Discussion	92
3.4	Methods	94
	References	97

3.5	Supplementary Information	103
4	Evolution of complex forms on Biomorphs genotype-phenotype-fitness maps	107
4.1	Introduction	108
4.2	Results	111
4.2.1	Defining genotype-phenotype-fitness maps	111
4.2.2	The link between local peaks and evolvability	112
4.2.3	The structure of the Biomorphs GP map does not always make the fitness landscapes navigable	117
4.2.4	Evolutionary navigability	122
4.3	Discussion	126
4.4	Methods	128
	References	133
4.5	Supplementary Information	136
5	Conclusions and future directions	149
5.1	Conclusions	149
5.2	Caveats	150
5.3	Unfinished puzzles and future directions	152
5.3.1	Trapping in changing environments	152
5.3.2	DNA methylation as an extra-dimensional bypass	153
5.3.3	Effect of sexual selection on the emergence of complexity	155
5.3.4	Effect of neutrality on adaptation	155
5.4	Outlook	156
	References	157

INTRODUCTION

O snail, climb Mt. Fuji, but slowly, slowly

— Kobayashi Issa

1.1 FROM DARWIN TO WRIGHT

In 1859, prompted by Alfred Russel Wallace's letter describing similar findings, Charles Darwin published an abstract of arguably the most influential work [1] on mankind's understanding of its origins, one that he had been ruminating on for years. The then revolutionary idea, that natural diversity has been shaped by the force of natural selection acting on randomly occurring heritable variation, took its time to be fully accepted. This was followed by Gregor Mendel's particulate theory of inheritance and advances in population genetics led by Ronald Fisher, J.B.S Haldane and Sewall Wright, which further solidified the theory and eventually culminated into the Modern Evolutionary Synthesis [2] that brought together naturalists, geneticists and paleontologists. Fascinatingly, all this progress happened before the discovery of the structure of DNA.

In 1932, Sewall Wright devised the schematic diagram of an adaptive landscape [3], to visualise evolution as a quest to climb adaptive peaks and elucidated conditions that allow a population to shift from one peak to another. While his goal was to explain the mathematical results of his paper [4] to a biological audience, his diagram ended up having a much larger impact than he would have anticipated. Ninety two years later, numerous theoretical results have been derived, various empirical adaptive landscapes have been measured and many a physicist has transitioned into an evolutionary biologist due to the intellectual appeal of the idea.

1.2 ONE MAP TO EXPLAIN IT ALL?

While Wright used the two-dimensional adaptive landscape solely as a visual metaphor to explain his shifting balance theory [5], the complete multidimensional fitness landscape holds the potential to explain [6], predict [7] and even direct [8] evolution. Formally, a fitness landscape F is defined as a map from the space of all genotypes, henceforth called the genotype space $G = \mathcal{A}^L$ to the space of real numbers,

$$F : G \rightarrow \mathbb{R}$$

where, \mathcal{A} is the alphabet of the genotype space and L is the length of the genotype. For instance, if the genotype space is composed of DNA sequences of length 8, $\mathcal{A} =$

$\{A, G, T, C\}$ and $L = 8$. Although this mathematical definition appears straightforward, capturing the intricacies of an organism, by assigning a single number to its genotype is unsurprisingly a challenging endeavour. However, for the purposes of this section, let's assume that it is possible to do.

Wright was well aware that non-additive interactions between mutations can lead to the existence of multiple adaptive peaks [3] or in other words, the local property of epistasis [9, 10] can influence the global property of the topography of the fitness landscape [7, 11]. This is due to a strong form of epistasis called sign epistasis [12] wherein the sign of the effect of a mutation depends upon the background on which it occurs. A fitness landscape can have multiple peaks only when sign epistasis occurs reciprocally between a pair of mutations, i.e. in the presence of reciprocal sign epistasis [13]. For instance, reciprocal sign epistasis can occur between costly resistance mutations in bacteria and compensate for the cost of resistance [14]. This can potentially make the bacterial fitness landscape rugged [15].

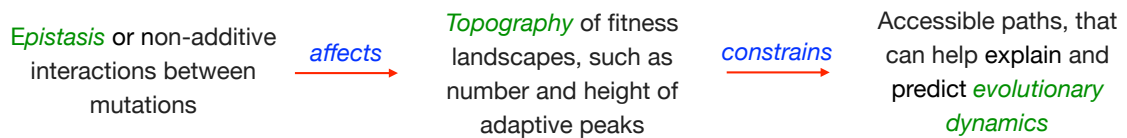


FIGURE 1.1: A flow chart linking epistasis to topography and evolutionary dynamics

A population evolving on a fitness landscape is more likely to traverse accessible paths comprising of a series of beneficial (or neutral) mutations, or in terms of Maynard Smith's word conversion metaphor [16], paths comprising of mutations that always lead to "meaningful words". The topography of the fitness landscape can thus constrain evolutionary trajectories, by reducing the number of accessible paths [17] and also determining the probability of realisation of these paths [18]. This enables us to understand and predict wide-ranging evolutionary phenomena, such as speciation [19], the evolution of sex [20] or the evolution of antibiotic resistance [17]. A flow chart summarising the connection between epistasis, topography of fitness landscapes and evolutionary dynamics is shown in Figure 1.1.

Since the definition of a fitness landscape doesn't rely on a particular genotype space or measure of fitness, one is free to choose the genotype space (e.g. DNA sequences, amino acid sequences or a sequence of a few genes) and the definition of fitness (e.g. growth rate or the binding affinity) depending upon the system of interest. Thus, the concept of a fitness landscape can be applied to multiple scales (see Figure 1.2). This can link molecular and systems biology to the study of evolution and ecology [6].

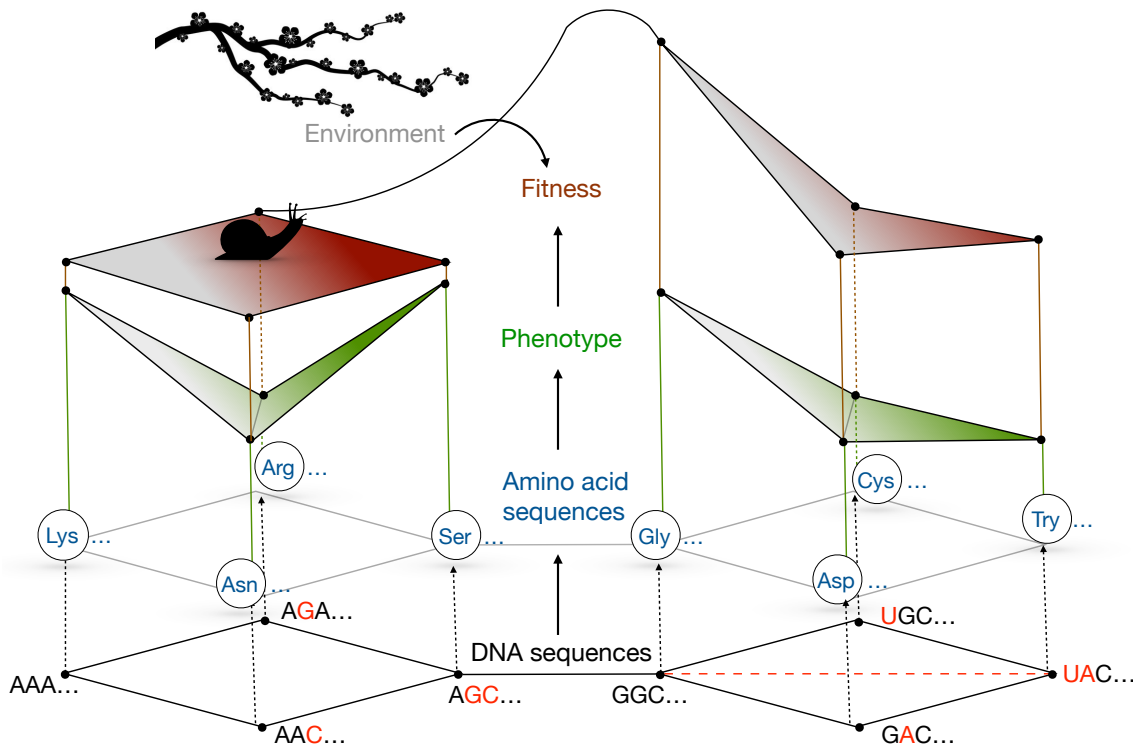


FIGURE 1.2: ANATOMY OF MT. FUJI. A schematic diagram showing a Mt. Fuji like adaptive landscape across multiple scales. The genotype space comprising of DNA sequences (shown in black), gets transcribed and translated into a genotype space comprising of amino acid sequences (shown in blue), this maps to some intermediate phenotype such as binding affinity or catalytic activity (shown in green) and then the phenotype maps to fitness (shown in maroon). The genotype-phenotype and genotype-fitness sub-landscapes shown on the left have different sign epistasis motifs and are thus *incongruent* [21], whereas those shown on the right have the same sign epistasis motifs and are thus *congruent*. Note how some amino acid sequences (such as the ones starting with Lysine and Serine) differing by a single amino acid are not mutational neighbours in the amino-acid genotype space due to the underlying genetic code [22]. The triangulation [23] imposed by the genotype-fitness sub-landscape on the right upon the DNA genotype space is shown with a red dashed line. The tree branch represents the external environment, which too can influence the landscape. The snail sits in the centre of the square as opposed to sitting on a vertex, indicating that the snail population is polymorphic and consists of a mixture of the genotypes flanking the square. The black line at the top connecting genotypes AGA... and GGC... is a schematic for the possible shape of the fitness landscape.

1.3 FROM MODELS TO DATA

Adaptive landscapes belong to two broad classes – *theoretical landscapes*, that rely on statistical or biophysical models to assign fitness values or phenotypes to genotypes

and *empirical landscapes*, that experimentally measure the fitness or some intermediate phenotype corresponding to each genotype. In this section, I discuss these two classes of adaptive landscapes.

1.3.1 *Theoretical landscapes*

The flexibility of the concept of an adaptive landscape led to theoretical models being developed at various levels of biological organisation. While some models directly map genotypes to fitness, others map genotypes to phenotypes or phenotypes to fitness. These different types of theoretical fitness landscapes are discussed below.

Genotype-fitness maps or fitness landscapes

The first models of fitness landscapes simply assigned random variables sampled from a fixed distribution to each genotype and came to be known as House-of-cards (HoC) fitness landscapes [24, 25]. These landscapes have completely uncorrelated fitness values and are characterised by high levels of epistasis and ruggedness. Then Stuart Kauffman introduced a slightly more sophisticated probabilistic model, in which the amount of epistasis could be tuned [26, 27]. This was done by controlling the number of interacting loci, $K+1$ in a genotype of length N and thus, this model is called the NK model. Another model with tune-able epistasis and therefore tune-able ruggedness is the Rough Mt. Fuji (RMF) model [28] that starts with a Mt. Fuji like additive landscape and adds a random component to the fitness value of each genotype. The relative importance of the random component can be adjusted by varying a roughness parameter. It is worth noting that while the HoC and NK model are isotropic landscapes, the RMF landscapes are anisotropic along the distance from the optimal genotype [29]. These statistical models can be used to establish general principles about landscape topography and accessible paths.

Another highly epistatic landscape called the Holey landscape was introduced by Sergey Gavrillets [19], in which each genotype is either neutral or lethal with some probability. This model helped in understanding the effect of higher dimensions on the prevalence of accessible paths, by demonstrating how higher dimensional ridges could enable bypasses around valleys. Apart from expanding our two-dimensional intuition about fitness landscapes, this model was also used to study speciation through genetic incompatibilities, much like it was done in the Bateson–Dobzhansky–Muller (BDM) model [30].

More recently, geometric properties of fitness landscapes, such as the triangulation induced by the fitness landscape on the convex hull of the genotype space, have been used to classify fitness landscapes, such that landscapes within the same class have similar topographical features and evolutionary dynamics [31–33] (see Figure 1.2). A

special class of landscapes that induce staircase triangulations [31] show the property of universal positive epistasis – i.e. upon representing bi-allelic genotypes as sets of loci that have been mutated, for any two genotypes σ and σ' , with $\sigma' \subset \sigma$ and a set τ of loci not contained in σ , universal positive epistasis guarantees that the fitness effect of σ mutating to $\sigma \cup \tau$ is always greater than the effect of σ' mutating to $\sigma' \cup \tau$ [23]. Furthermore, a similar concept of universal negative epistasis has also been developed, which satisfies the accessibility property of trade-off induced landscapes (wherein any peak genotype is accessible from all its subset and superset genotypes via all direct paths) [15] and explains the existence of rugged yet navigable fitness landscapes [29].

Lastly, there are also models of pervasive microscopic epistasis, that explain global evolutionary trends such as diminishing-returns and increasing-costs epistasis [34, 35].

Genotype-phenotype maps

Then there are biophysical models that describe the mapping between genotypes and specific phenotypes, for instance the mapping between RNA sequences and their secondary structure through free energy minimisation [36, 37], or between DNA sequences and their binding affinities to various transcription factors through the mismatch model [38] or more recently, landscapes in which genotypes are divided into constrained and unconstrained parts [39], motivated by the existence of exons and introns or genes and non-coding sequences. The last model gives rise to properties such as genotypic redundancy, phenotypic bias, positive correlation between phenotypic robustness and evolvability and shape-space covering [40] that result in genotype-phenotype maps with large connected neutral networks – a feature that is shared by many molecular genotype-phenotype maps [41]. Finally, there are also multi-level computational genotype-phenotype maps such as toyLIFE [42], that integrate various levels of phenotypes such as protein folding, gene regulation and metabolic networks and yet show similar genotype-phenotype map properties.

Phenotype-fitness maps

Finally, there are phenotype-fitness maps such as the Fisher's Geometric model [43] and the multi-linear model from quantitative genetics [44]. Fisher's Geometric model, in keeping with Fisher's views on evolution, assumes a single-peaked fitness function defined on a multi-dimensional trait space. Though the initial model was heuristic, it was later shown that such phenotype-fitness maps can be derived from first principles with minimal assumptions about the phenotypic space [45]. Moreover, it highlights that non-linearities at various levels of the fitness landscape can introduce epistasis in an otherwise additive genotype-phenotype map [21, 46, 47]. Further, it has also been used to match and explain empirical results, such as predicting the distribution of epistasis from the distribution of single mutation effects for *E. coli* and an RNA virus [48].

Thus, theoretical models of adaptive landscapes span the entire space of possibilities, from being completely uncorrelated to being fully correlated, from being purely statistical to being firmly grounded on biophysical principles, from merely describing the relationship between phenotype and fitness to describing the entire mapping from genotype to fitness and from explaining molecular evolution to explaining evolutionary ecology.

1.3.2 Empirical landscapes

Creating models of adaptive landscapes and analysing their evolutionary implications is certainly fascinating. But what do *real* adaptive landscapes look like? While small adaptive landscapes with less than ten loci have existed for more than two decades e.g., [12], these samples are inherently biased because they are constructed using mutations that are known to have either individual or combined beneficial effects. This can severely affect our conclusions about the extent of ruggedness and epistasis in these landscapes, primarily leading to an underestimation of sign epistasis [7]. Attempts to sample entire adaptive landscapes quickly become intractable due to the exponential growth of the number of possible genotypes, which becomes hyper-astronomical [49]. However, recent technological advancements such as high throughput sequencing technology have reinvigorated the field of adaptive landscapes, by allowing sampling of larger and complete empirical landscapes. These landscapes not only enable the testing of the validity of existing theoretical models, they can also inform the development of better models.

Some examples of large and nearly combinatorially complete landscapes include measurement of affinity of all 4^{10} 10-nucleotide DNA oligomers to allophycocyanin protein [50], measurement of binding affinity of all 20^4 mutants of GB1 protein to immunoglobulin [51] or the recently published fitness landscape of 4^9 nucleotide substitutions in *folA* gene in *E. coli* [52]. Missing fitness measurements can also be interpolated with an accuracy depending upon the order of epistasis that is relevant for determining the fitness [53].

Meta studies on empirical landscapes [7, 11] have revealed some of their common features, such as:

- Empirical landscapes have intermediate ruggedness, i.e. their ruggedness is somewhere between a Mt. Fuji and a HoC landscape. Further, ruggedness is greater for mutations whose combined effect is unknown and mutations that individually have a large effect.
- It is probable that mutations that occur in the same gene show higher epistasis as opposed to mutations occurring in different genes. However, this still remains to be tested systematically. Moreover, mutations occurring in genes located in the same pathway show more epistasis than mutations in random genes. Thus, adaptive

landscapes sampling a small contiguous portion of the genome are more likely to be epistatic and therefore rugged, than adaptive landscapes that combine different portions of the genome.

- Diminishing returns epistasis is a commonly observed phenomenon, where the beneficial effect of a mutation is smaller when it appears on a fitter genetic background. This can be attributed to idiosyncratic epistasis on the genotype–phenotype level.
- Empirical landscapes regularly show higher order epistasis, which can reduce predictability of the fitness landscape based on solely pairwise epistatic effects. However, the prevalence of higher order epistasis is still under debate.
- Epistasis observed in empirical landscapes is highly dependent upon the environment, perhaps due to changes in phenotype–fitness relationships that can lead to moving phenotypic optima – as has been proposed to occur in fitness seascapes [54].

1.4 EVOLUTION ACROSS MULTIPLE SCALES

Biology operates on multiple scales, starting at the level of DNA, where genetic information necessary for inheritance is stored. DNA undergoes transcription into mRNA, which in turn undergoes translation into proteins, which carry out essential cellular tasks. Proteins interact with one another and with DNA and RNA molecules, in order to regulate the expression of various genes and eventually orchestrate the development of an organism (see Figure 1.4 A). It is miraculous how well this process works, given the stochasticity due to the small number of molecules involved in each step of the process [55].

The characteristics of fitness landscapes across these different scales play an important role in shaping biological processes and can, therefore, help us understand why biology functions the way it does. For instance, a comparative analysis of TF-DNA genotype-phenotype maps and RNA and RNA binding protein genotype-phenotype maps showed that the TF-DNA genotype-phenotype map had higher evolvability and could thus bring forth phenotypic variation more quickly [56]. This can explain why transcriptional regulation, rather than post-transcriptional regulation mediated by RNA binding proteins, is more often utilised in adaptation and innovation [40]. Furthermore, the architecture of the intermediate genotype-phenotype maps can greatly influence evolutionary outcomes and studies have shown that phenotypic abundance i.e. the number of genotypes mapping to a phenotype, plays an equally important role as fitness in evolution [57, 58] and that explaining the direction of evolution requires both selective and generative arguments [59].

In the strong selection weak mutation regime, evolutionary dynamics gets reduced to an adaptive walk [27, 63]. In Figure 1.3, I show the outcome of 1000 random adaptive

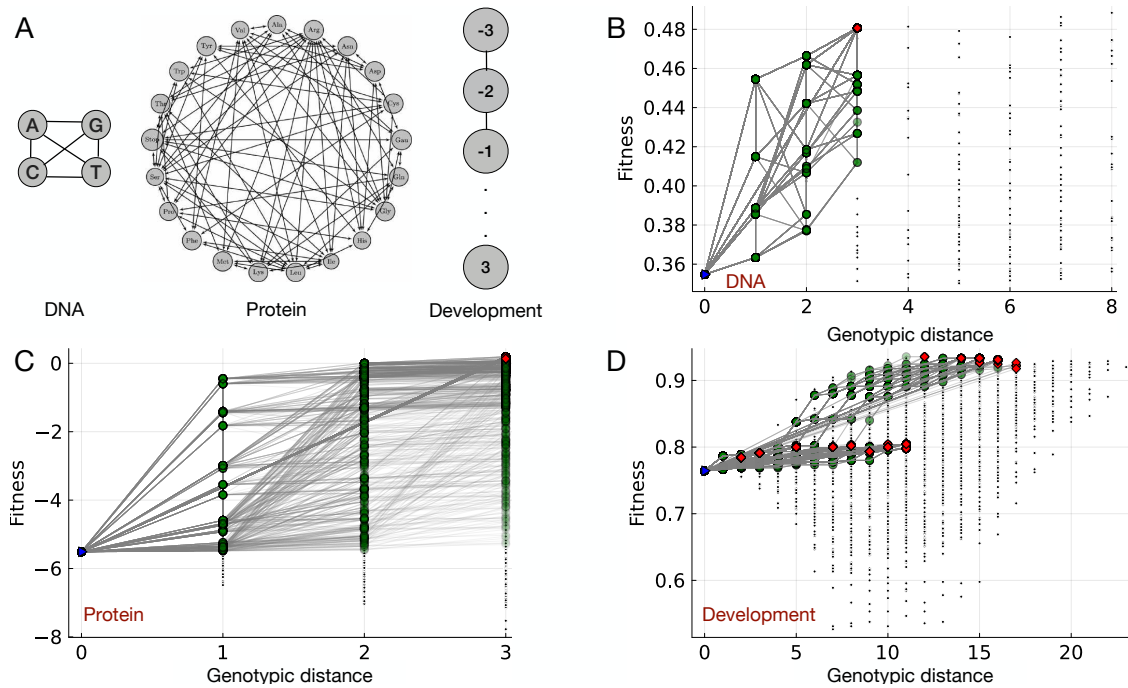


FIGURE 1.3: RANDOM ADAPTIVE WALKS ON ADAPTIVE LANDSCAPES ACROSS MULTIPLE SCALES: A. Allele graphs corresponding to the alphabets of DNA, protein and developmental genes. The DNA alphabet consists of the nucleotides {A,G,T,C}, the protein alphabet consists of the 20 standard amino acids and the alphabet of developmental genes consists of allowed values of genes at each site, as described in [60]. Results of 1000 random adaptive walks on fitness landscapes at the level of B. DNA sequences of length 8 [61], C. protein sequences of length 3 [62] and D. sequences of developmental genes of length 9 [60]. The black dots show the fitness of genotypes at a given distance from the reference genotypes. Genotypes highlighted in green were visited during the adaptive walks and the paths between these genotypes is shown with grey lines. Local peaks are highlighted in red and the initial genotype in blue.

walks on adaptive landscapes at three different biological scales – the level of TF-DNA interactions [61], the level of protein function [62] and the level of development [60]. The research presented in this thesis focuses on these three scales and therefore, this preliminary analysis will provide a glimpse of what is to come. The three landscapes have very different allele graphs (see Figure 1.3 A): while the DNA allele graph has four possible alleles and is fully connected, the amino acid allele graph has 20 possible alleles and is well connected with only a few dis-allowed amino acid substitutions due to the underlying genetic code [22], finally, the developmental allele graph is a sparsely connected path graph [64] with an intermediate number of alleles. Thus, in the examples considered here, the allele graph connectivity decreases as one goes up in biological scale. In terms of ruggedness, the DNA landscape has 2 peaks amongst $4^8 = 65,536$ sequences (Figure 1.3 B), the protein landscape has 5 peaks amongst $20^3 = 8,000$ sequences (Figure

1.3 C) and the developmental landscape has 19 peaks amongst $7^4 = 2,401$ sequences (Figure 1.3 D). This differing architecture results in very different evolutionary outcomes. At the level of DNA, while there is significant divergence in the evolutionary paths taken, all 1000 walks terminate on the same peak with a reasonably high fitness. Moreover, only a small fraction of the fitness landscape, at a small distance from the starting genotype is explored by the walks. At the level of protein, a high divergence in the walks is observed, however the walks terminate on only two of the highest peaks in the landscape. In contrast to the DNA landscape, here a significant fraction of the landscape is explored by the adaptive walks. This is due to the size and connectivity of the allele graph that ensures that even when the genotype space grows exponentially with the length of the sequence, the distances only grow linearly [40]. Finally, at the level of development, we see divergent trajectories but also divergent evolutionary end-points, although they primarily belong to two fitness classes. Further, as in the case of DNA sequences, a very small portion of the landscape is sampled during the walks.

Thus, it is evident that evolution can look very different across different levels, for instance in terms of its predictability or the fraction of genotype space it explores. In the following paragraphs, I discuss our current understanding of various aspects of adaptive landscapes, and consequently, evolution across different scales. The broad trends are summarised in Figure 1.4 B.

Knowledge across multiple scales

Since the number of interacting parts increases, as one goes up in scale from the molecular level to the organisimal level, it becomes increasingly difficult to study combinatorially complete adaptive landscapes at the corresponding levels. While we understand the mechanics of TF-DNA interactions and can measure the corresponding binding affinities [65] to a reasonably good degree, and we have computational packages to compute the lowest free energy secondary structure of RNA sequences [66]; the genotype-phenotype maps at higher levels such as that of gene regulatory networks [42] and body pattern development [60] are less well studied and also harder to measure accurately.

Neutrality across multiple scales

Motoo Kimura, the father of the neutral theory of evolution [67], suggested that genetic drift could be the primary driver of molecular evolution and that natural selection mostly acts on the phenotypic level. While most of the theoretical landscapes discussed above lack neutral mutations, empirical landscapes at almost all levels have some genotypic redundancy along with phenotypic correlations, that result in large connected neutral networks. For instance, the degeneracy in the genetic code leads to synonymous mutations

in protein landscapes and the robustness of gene regulatory networks leads to many gene combinations resulting in the same body plan. This causes the effect of mutations to be buffered across different levels of biological organisation. It can also strongly influence the topography of the resulting fitness landscape and confer phenotypic evolvability [68]. Prior work has shown that adding more layers to the mapping of genotypes to phenotypes enhances both robustness and evolvability [69]. Further, even non-degenerate fitness landscapes can be rendered effectively neutral in rapidly changing environments [54]. Thus, the neutrality at the level of the genotype-phenotype map cannot be neglected, especially at higher levels of biological organisation.

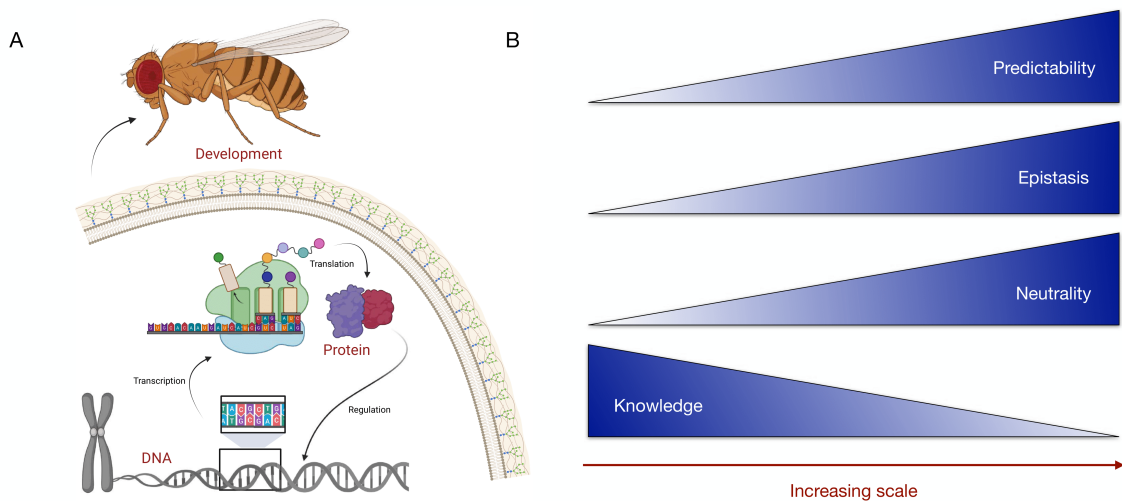


FIGURE 1.4: A. Biology at multiple scales: From DNA to Development. B. Broad trends in the levels of knowledge, neutrality, epistasis and predictability with increasing scale of the adaptive landscape.

Topography across multiple scales

One topographical feature that changes as we go up in biological scale is the connectivity of the allele graph of the genotype space. As mentioned above, the connectivity of the allele graph generally decreases as we go up in scale, which can increase the number of local peaks. Consequently, adaptation could be more constrained at higher levels of biological organisation. It is also known that the density of local peaks is lower for larger scale landscapes, and this correlates well with theoretical predictions [7]. As for propagation of epistasis across hierarchical levels, for metabolic networks it was observed that epistasis only gets stronger at higher levels and negative epistasis cannot be converted into positive epistasis. [70].

Predictability across multiple scales

The predictability of evolution depends upon the number of local peaks in the landscape. Many local peaks imply many possible destinations for populations starting at different initial conditions. Reasoning from our knowledge of the empirical datasets and the lower connectivity of allele graphs at higher levels, would lead us to conclude that landscapes at higher levels are more rugged and thus evolution is less predictable. However, the large neutral networks and bias in the arrival of variation at higher levels [59, 71] can render evolution more predictable. Moreover, some accessible paths having disproportionately higher probabilities of realisation can also enhance the predictability. Thus it is generally accepted that evolution is more predictable at the level of genes and metabolic pathways, than at the level of nucleotide substitutions [6, 7]. Mutational bias [72, 73] in nucleotide substitutions can however increase the predictability of these substitutions. Finally, predictability is highest at intermediate population sizes and lowest for landscapes of intermediate ruggedness [7, 11].

1.5 EVOLUTION IN ACTION - CASE STUDY OF SARS-COV-2

Few months before I started my Ph.D. in September 2020, the world became hostage to a nanometer sized virus and strongly validated Stephen Jay Gould's reservations about the iconography of the ladder [74], that places mankind at the pinnacle of evolutionary progress. The scale of the pandemic transferred textbook epidemiological and evolutionary theory to the newspapers, as we experienced the successive emergence of different viral strains over the next months.

Following a period of limited evolution over eight months, there was a swift emergence of divergent variants of concern — Alpha, Beta, and Gamma — each approximately 10-12 non-synonymous mutations away from their ancestral strains. Subsequently, the Omicron lineage surfaced, evolving gradually in comparison to the preceding variants of concern [75]. The evolutionary forces at play here included a mutation rate of roughly 10^{-6} mutations per base per replication cycle, host-mediated genome editing, recombination, genetic drift, and natural selection. This prompts the question of which viral phenotypes were being acted upon by natural selection. In the beginning of the pandemic, transmissibility of the virus was the primary driver of evolution. Thus, fitness could be increased by altering the spike protein to enhance binding to the ACE2 receptor, as was seen to be the case [75, 76]. After a certain fraction of the population acquired immunity, immune escape became an important objective of selection, since even a highly transmissible virus cannot infect an immune population. Omicron BA.1 variant was in fact found to have a higher re-infection rate due to mutations in the receptor binding domain (RBD). Thus, the phenotype under selection and therefore the fitness landscape, kept changing as the pandemic progressed.

Knowledge of the fitness landscape of the virus, with the correct phenotype as a proxy for fitness, could thus be used to understand the early evolution of the virus. One such study considered the bi-allelic landscape of all 15 point mutations between the spike protein of the ancestral strain and the Omicron BA.1 variant and chose the binding affinity to the ACE2 receptor as a proxy for fitness. This study found several epistatic interactions and identified compensatory mutations as primary drivers of SARS-CoV-2 evolution, while attributing the emergence of the variant to chronic infections [77]. Another recent study focused on immune escape and analysed the same genotype space, but used the binding affinities to 4 monoclonal antibodies as proxies for fitness [78]. They found that the evolution of escape to each antibody can largely occur independently, suggesting that there are many pathways to immune escape [78]. On the other hand, trade-offs between binding ACE2 and escaping antibodies were found to be much stronger. Lastly, higher order epistasis was found to be prevalent but only in one of the four antibody fitness landscapes.

Thus, so far, studying the local adaptive landscape of the virus with the relevant fitness determining traits, has helped in understanding the course of the pandemic. Further such investigations can help in examining the repeatability of the sequence of events described above and more importantly, the predictability of future events to ensure better preparedness. While deep-mutational scanning and machine learning approaches are being employed towards this end [79], it is by no measure an easy goal to achieve. Nevertheless, the COVID-19 pandemic, alongside other issues like the global prevalence of antibiotic resistance, has highlighted that understanding evolution is now an imperative for averting and controlling public health crises. I would like to conclude by quoting words attributed to Marie Curie, "*Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less*". Fortunately, we are better equipped than ever before to fulfil that objective.

1.6 THESIS OUTLINE

In this thesis, I address three knowledge gaps in our theoretical understanding of adaptive landscapes. While addressing each knowledge gap, I also arrive at biological insights at different levels of biological organisation – ranging from the level of DNA to the level of development.

The first knowledge gap pertains to the use of measurable phenotypes as proxies for fitness, as has most recently been done in studying the SARS-CoV-2 landscapes. In Chapter 2, we systematically study how using phenotypes as a proxy for fitness can lead to differences, or incongruence between the genotype-phenotype map and the fitness landscape, at both a local and a global level, especially under selection for low or intermediate values of the phenotype. Through this study, we establish some general principles of incongruence between genotype-phenotype maps and fitness landscapes. In

the second half of the chapter, we apply these general principles to genotype-phenotype maps at the level of DNA – in particular the system of TF-DNA interactions, that is essential for gene-regulation. We use a biophysically motivated landscape and more than a thousand empirical landscapes of TF-DNA interactions, to test the validity of the theoretical principles derived in the former part of the chapter. Finally, we use evolutionary simulations to provide a possible explanation for the prevalence of low-affinity binding sites in the regulatory portfolios of various organisms such as bacteria, yeast and fly.

The second knowledge gap concerns the impact of increasing the dimensionality of a fitness landscape, by increasing the number of alleles at each locus or the alphabet cardinality. While the extra-dimensional bypasses [80] arising due to the increased alphabet cardinality have been thought to aid adaptation – by opening up new accessible paths to the global optimum, the altered topography with increased ruggedness can potentially be detrimental to adaptation. In Chapter 3, we study the interaction of these two opposing factors, to derive conditions under which increasing the alphabet cardinality can be beneficial. The biological motivation behind this analysis was to understand the expansion of the amino acid alphabet and the existence of 20 standard amino acids, that seems to be conserved across all organisms. To this end, we analyse the effect of increasing the number of amino acids on adaptation, in three combinatorially complete, empirical protein fitness landscapes. Using this analysis, we show that the intuition gained from theoretical fitness landscapes and conventional measures such as landscape ruggedness, need not be indicative of evolutionary dynamics on empirical fitness landscapes.

The third knowledge gap was regarding the architecture of genotype-phenotype-fitness landscapes at higher levels biological organisation. In particular, we study Richard Dawkin’s Biomorphs, which is a toy model for body plan development, that was devised to explain the emergence of complex body plans. First, we establish general relationships between the topography of the landscape and the evolvability of phenotypes, which should also be applicable to a much wider class of genotype-phenotype maps with random fitness assignments. The biological motivation was to study the frequency of evolution of complex body plans from simple ones, in the Biomorphs fitness landscape. To do this, we examine the navigability of the landscape, i.e. the probability of occurrence of accessible paths between given source and target phenotypes, with a special focus on simple source phenotypes and complex target phenotypes. We also examine how frequently these accessible paths are utilised by evolving populations in various evolutionary regimes. We demonstrate that despite the existence of accessible paths, evolving populations seldom reach complex target phenotypes from simple source phenotypes.

Finally, in Chapter 5, I summarise the findings of this thesis and discuss future research directions that take into account additional features, such as changing environments, phenotypic plasticity and ecological interactions, with the goal to develop more comprehensive and widely applicable models of adaptive landscapes across multiple scales.

REFERENCES

1. Darwin, C. *On the Origin of Species by Means of Natural Selection* (London, 1859).
2. Huxley, J. *Evolution: The Modern Synthesis* (MIT Press, 2010).
3. Wright, S. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. *Proceedings of the 6th International Congress of Genetics*, 356 (1932).
4. Wright, S. Evolution in Mendelian Populations. *Genetics* **16**, 97 (1931).
5. Skipper, R. A. & Dietrich, M. R. in *The Adaptive Landscape in Evolutionary Biology* (Oxford University Press, 2013).
6. Fragata, I., Blanckaert, A., Louro, M. A. D., Liberles, D. A. & Bank, C. Evolution in the light of fitness landscape theory. *Trends in Ecology & Evolution* **34**, 69 (2019).
7. De Visser, J. A. G. M. & Krug, J. Empirical fitness landscapes and the predictability of evolution. *Nature Reviews Genetics* **15**, 480 (2014).
8. Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nature Reviews Molecular Cell Biology* **10** (2009).
9. Bateson, W. Heredity and variation in modern lights. *and modern science* (1909).
10. Phillips, P. C. Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews Genetics* **9**, 855 (2008).
11. Bank, C. Epistasis and Adaptation on Fitness Landscapes. *Annual Review of Ecology, Evolution, and Systematics* **53**, 457 (2022).
12. Weinreich, D., Watson, R. & Chao, L. Perspective: sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* **59**, 1165 (2005).
13. Poelwijk, F., Tănase-Nicola, S., Kiviet, D. & Tans, S. Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. *Journal of Theoretical Biology* **272**, 141 (2011).
14. Hall, A. R. & Maclean, R. C. Epistasis buffers the fitness effects of rifampicin-resistance mutations in *Pseudomonas aeruginosa*. *Evolution* **65**, 2370 (8 2011).
15. Das, S. G., Direito, S. O., Waclaw, B., J Allen, R. & Krug, J. Predictable properties of fitness landscapes induced by adaptational tradeoffs. *eLife*, e55155 (2020).
16. Maynard Smith, J. Natural selection and the concept of a protein space. *Nature* **225**, 563 (1970).
17. 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 (2006).
18. Manhart, M. & Morozov, A. Statistical Physics of Evolutionary Trajectories on Fitness Landscapes. *First-Passage Phenomena and Their Applications* (2013).

19. Gavrillets, S. *Fitness Landscapes and the Origin of Species*. Princeton Univ. Press (2004).
20. Misevic, D., Kouyos, R. D. & Bonhoeffer, S. Predicting the Evolution of Sex on Complex Fitness Landscapes. *PLoS Computational Biology* **5** (2009).
21. Srivastava, M. & Payne, J. L. On the incongruence of genotype-phenotype and fitness landscapes. *PLoS Computational Biology* **18(9)** (2022).
22. Rozhoňová, H., Martí-Gómez, C., McCandlish, D. M. & Payne, J. L. Robust genetic codes enhance protein evolvability. *PLOS Biology* **22**, 1 (2024).
23. Crona K. and Krug, J. & Srivastava, M. Geometry of fitness landscapes: peaks, shapes and universal positive epistasis. *Journal of Mathematical Biology* **86** (2023).
24. Kingman, J. On the properties of bilinear models for the balance between genetic mutation and selection. *Mathematical Proceedings of the Cambridge Philosophical Society* **81**, 443 (1977).
25. Kingman, J. A simple model for the balance between selection and mutation. *J. Appl. Probab.* **15**, 1 (1978).
26. Kauffman, S. A. & Weinberger, E. D. The NK model of rugged fitness landscapes and its application to the maturation of the immune response. *Journal of Theoretical Biology* **141**, 211 (1989).
27. Kauffman, S. & Levin, S. Towards a general theory of adaptive walks on rugged landscapes. *Journal of Theoretical Biology.* **128**, 11 (1987).
28. Aita, T., Uchiyama, H., Inaoka, T., Nakajima, M., T, K. & Husimi, Y. Analysis of a local fitness landscape with a model of the rough Mt. Fuji-type landscape: application to prolyl endopeptidase and thermolysin. *Biopolymers* **54**, 64 (2000).
29. Krug, J. & Oros, D. Evolutionary accessibility of random and structured fitness landscapes. *Journal of Statistical Mechanics: Theory and Experiment* **2024**, 034003 (2024).
30. Orr, H. A. The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics* **139**, 1805 (1995).
31. Beerenwinkel, N., Pachter, L. & Sturmfels, B. Epistasis and shapes of fitness landscapes. *Stat. Sin.* **17**, 1317 (2007).
32. Crona, K. in *Recent Advances in the Theory and Application of Fitness Landscapes* (eds Richter, H. & Engelbrecht, A.) 177 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2014).
33. Srivastava, M. *Epistasis, Shapes and Evolution* PhD thesis (Universität zu Köln, 2018).
34. Lyons, D. M., Zou, Z., Xu, H. & Zhang, J. Idiosyncratic epistasis creates universals in mutational effects and evolutionary trajectories. *Nature Ecology & Evolution* **4**, 1685 (2020).

35. Reddy, G. & Desai, M. M. Global epistasis emerges from a generic model of a complex trait. *eLife* **10** (eds Barkai, N., Petrov, D. A., Crona, K. & Krug, J.) e64740 (2021).
36. Fontana, W., Stadler, P., Bornberg-Bauer, E., Griesmacher, T., Hofacker, I. & et al. RNA folding and combinatorial landscapes. *Physical Review E* **47**, 2083 (1993).
37. Pitt, J. & Ferré-D'Amaré, A. Rapid Construction of Empirical RNA Fitness Landscapes. *Science* **330**, 376 (2010).
38. Haldane, A., Manhart, M. & Morozov, A. Biophysical fitness landscapes for transcription factor binding sites. *PLoS Computational Biology* **10** (2014).
39. Greenbury, S. F. & Ahnert, S. E. The organization of biological sequences into constrained and unconstrained parts determines fundamental properties of genotype–phenotype maps. *Journal of The Royal Society Interface* **255**, 279 (2015).
40. Manrubia, S., Cuesta, J. A., Aguirre, J., Ahnert, S. E., Altenberg, L., Cano, A. V., Catalán, P., Diaz-Uriarte, R., Elena, S. F., García-Martín, J. A., Hogeweg, P., Khatri, B. S., Krug, J., Louis, A. A., Martin, N. S., Payne, J. L., Tarnowski, M. J. & Weiß, M. From genotypes to organisms: State-of-the-art and perspectives of a cornerstone in evolutionary dynamics. *Physics of Life Reviews* **38**, 55 (2021).
41. Ahnert, S. E. Structural properties of genotype-phenotype maps. *Journal of The Royal Society Interface* **132** (2017).
42. Arias, C. F., Catalán, P., Manrubia, S. & Cuesta, J. A. toyLIFE: a computational framework to study the multi-level organisation of the genotype-phenotype map (2014).
43. Fisher, R. A. *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford. (1930).
44. Hansen, T. F. & Wagner, G. P. Modeling Genetic Architecture: A Multilinear Theory of Gene Interaction. *Theoretical Population Biology* **59**, 61 (2001).
45. Martin, G. Fisher's Geometrical Model Emerges as a Property of Complex Integrated Phenotypic Networks. *Genetics* **197**, 237 (2014).
46. Rokyta, D. R., Joyce, P., Caudle, S. B., Miller, C., Beisel, C. J. & Wichman, H. A. Epistasis between Beneficial Mutations and the Phenotype-to-Fitness Map for a ssDNA Virus. *PLoS Genetics* **7** (2011).
47. Hwang, S., Park, S.-C. & Krug, J. Genotypic Complexity of Fisher's Geometric Model. *Genetics* **206**, 1049 (2017).
48. Martin, G., Elena, S. & Lenormand, T. Distributions of epistasis in microbes fit predictions from a fitness landscape model. *Nature genetics* **39**, 555 (2007).

49. Louis, A. A. Contingency, convergence and hyper-astronomical numbers in biological evolution. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* **58**. Special Issue: Replaying the Tape of Life: Evolution and Historical Explanation, 107 (2016).
50. Rowe, W., Platt, M., Wedge, D., Day, P., Kell, D. & Knowles, J. Analysis of a complete DNA-protein affinity landscape. *Journal of The Royal Society Interface*, 397 (2010).
51. Wu, N., Dai, C. L., Olson, C. A., Lloyd-Smith, J. O. & Sun, R. Adaptation in protein fitness landscapes is facilitated by indirect paths. *eLife* **5**, e16965 (2016).
52. Papkou, A., Garcia-Pastor, L., Escudero, J. A. & Wagner, A. A rugged yet easily navigable fitness landscape. *Science* **382**, eadh3860 (2023).
53. Weinreich, D. M., Lan, Y., Wylie, C. S. & Heckendorn, R. B. Should evolutionary geneticists worry about higher-order epistasis? *Current Opinion in Genetics & Development* **23**. Genetics of system biology, 700 (2013).
54. Mustonen, V. & Lässig, M. From fitness landscapes to seascape: non-equilibrium dynamics of selection and adaptation. *Trends in Genetics* **25**, 111 (2009).
55. Suter, D. M. Transcription Factors and DNA Play Hide and Seek. *Trends in Cell Biology* **30**, 491 (2020).
56. Payne, J. L., Khalid, F. & Wagner, A. RNA-mediated gene regulation is less evolvable than transcriptional regulation. *Proceedings of the National Academy of Sciences* **115**, E3481 (2018).
57. Schaper, S. & Louis, A. A. The Arrival of the Frequent: How Bias in Genotype-Phenotype Maps Can Steer Populations to Local Optima. *PLoS One* **9**, e86635 (2014).
58. Martin, N. S., Schaper, S., Camargo, C. Q. & Louis, A. A. Non-Poissonian bursts in the arrival of phenotypic variation can strongly affect the dynamics of adaptation. *Molecular Biology and Evolution* (2024).
59. Salazar-Ciudad, I. & Cano-Fernández, H. Evo-devo beyond development: Generalizing evo-devo to all levels of the phenotypic evolution. *BioEssays* **45**, 2200205 (3 2023).
60. Martin, N. S., Camargo, C. Q. & Louis, A. A. Bias in the arrival of variation can dominate over natural selection in Richard Dawkins' biomorphs. *PLoS Computational Biology* (2024).
61. Aguilar-Rodríguez, J., Payne, J. & Wagner, A. A thousand empirical adaptive landscapes and their navigability. *Nature Ecology and Evolution* **1** (2017).
62. Lite, T.-L. V., Grant, R. A., Nocedal, I., Littlehale, M. L., Guo, M. S. & Laub, M. T. Uncovering the basis of protein-protein interaction specificity with a combinatorially complete library. *eLife* **9**, e60924 (2020).

63. McCandlish, D. M. & Stoltzfus, A. Modeling evolution using the probability of fixation: history and implications. *The Quarterly review of biology* **89**, 225 (2014).
64. Schmiegelt, B. & Krug, J. Accessibility Percolation on Cartesian Power Graphs. *Journal of Mathematical Biology* **86**, 46 (2023).
65. Berger, M. F. & Bulyk, M. L. Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nature Protocols* 2009 4:3 **4**, 393 (3 2009).
66. Lorenz, R., Bernhart, S. H., zu Siederdissen, C. H., Tafer, H., Flamm, C., Stadler, P. F. & Hofacker, I. L. ViennaRNA Package 2.0. *Algorithms for Molecular Biology* **6**, 1 (1 2011).
67. Kimura, M. *The neutral theory of molecular evolution* (Cambridge University Press, 1983).
68. Wagner, A. Robustness and evolvability: a paradox resolved. *Proceedings of the Royal Society B: Biological Sciences* **275**, 91 (2008).
69. Catalán, P., Wagner, A., Manrubia, S. & Cuesta, J. A. Adding levels of complexity enhances robustness and evolvability in a multilevel genotype–phenotype map. *Journal of The Royal Society Interface* **15**, 20170516 (2018).
70. Kryazhimskiy, S. Emergence and propagation of epistasis in metabolic networks. *Elife* **10**, e60200 (2021).
71. Dingle, K., Ghaddar, F., Šulc, P. & Louis, A. A. Phenotype Bias Determines How Natural RNA Structures Occupy the Morphospace of All Possible Shapes. *Molecular Biology and Evolution* **39**, msab280 (2021).
72. Cano, A. V. & Payne, J. L. Mutation bias interacts with composition bias to influence adaptive evolution. *PLoS Computational Biology* **16** (9 2020).
73. Cano, A. V., Gitschlag, B. L., Rozhoňová, H., Stoltzfus, A., McCandlish, D. M. & Payne, J. L. Mutation bias and the predictability of evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* **378**, 20220055 (2023).
74. Gould, S. *Wonderful Life: The Burgess Shale and the Nature of History* (W.W. Norton, 1989).
75. Markov, P. V., Ghafari, M., Beer, M., Lythgoe, K., Simmonds, P., Stilianakis, N. I. & Katzourakis, A. The evolution of SARS-CoV-2. *Nature Reviews Microbiology* | **21**, 361 (2023).
76. SARS-CoV-2 D614G spike mutation increases entry efficiency with enhanced ACE2-binding affinity.
77. Moulana, A., Dupic, T., Phillips, A. M., Chang, J., Nieves, S., Roffler, A. A., Greaney, A. J., Starr, T. N., Bloom, J. D. & Desai, M. M. Compensatory epistasis maintains ACE2 affinity in SARS-CoV-2 Omicron BA.1.

78. Moulana, A., Dupic, T., Phillips, A. M., Chang, J., Roffler, A. A., Greaney, A. J., Starr, T. N., Bloom, J. D. & Desai, M. M. The landscape of antibody binding affinity in SARS-CoV-2 Omicron BA.1 evolution. *eLife* **12** (ed van der Meer, J. W.) e83442 (2023).
79. Han, W., Chen, N., Xu, X., Sahil, A., Zhou, J., Li, Z., Zhong, H., Gao, E., Zhang, R., Wang, Y., Sun, S., Cheung, P. P.-H. & Gao, X. Predicting the antigenic evolution of SARS-COV-2 with deep learning.
80. Cariani, P. A. Extradimensional bypass. *Biosystems* **64**, 47 (2002).

GLOSSARY

accessible paths	Paths in which the fitness increases or remains constant in each mutational step.
additive landscape	A fitness landscape in which each locus has an independent additive contribution to the fitness.
allele graph	A graph showing all possible allelic states at each locus in a genotype with the edges connecting mutational neighbours.
alphabet cardinality	The number of possible allelic states at each locus in a genotype, e.g., it is 4 for DNA sequences and 20 for amino acid sequences.
epistasis	Non-additive interactions between effects of mutations.
evolvability	The number of novel phenotypes accessible by point mutations.
extra-dimensional bypass	A phenomenon in which a local maximum transforms into a saddle point upon increasing the dimensions, thereby allowing further adaptation.
fitness landscape	A map from the space of all possible genotypes of an organism to their corresponding reproductive capabilities. Often used interchangeably with the term "adaptive landscape".
genotype space	The space of all possible genotypes of an organism.
genotypic redundancy	Many genotypes mapping to the same phenotype.
higher order epistasis	When the effect of a mutation at locus is determined by interactions with two or more than two other loci.

incongruence	Differences in the local and global topographical features of two landscapes.
navigability	The average probability of existence of an accessible path between a given source and target phenotype.
negative epistasis	Combinations of mutations have fitness values lower than that expected from the sum of the fitness effects of single mutants.
phenotypic bias	Variation in the number of genotypes mapping to each phenotype.
phenotypic robustness	The average of the genotypic robustness (i.e. the number of neutral neighbours) over all genotypes mapping to a phenotype.
positive epistasis	Combinations of mutations have fitness values higher than that expected from the sum of the fitness effects of single mutants.
random adaptive walks	A walk on an adaptive landscape in which each step is taken by randomly choosing a fitter neighbour.
reciprocal sign epistasis	When two individually deleterious mutations jointly enhance fitness or vice-versa.
ruggedness	Measure of the number of local maxima or peaks in an adaptive landscape.
shape-space covering	A property of only requiring a small number of mutations to reach the most common phenotypes from a given phenotype.
triangulation	A triangulation of the L-cube is a subdivision of the cube into simplices (i.e. triangles if $L = 2$, tetrahedra if $L = 3$ etc).

ON THE INCONGRUENCE OF GENOTYPE-PHENOTYPE AND FITNESS LANDSCAPES

The real voyage of discovery consists not in seeking new landscapes, but in having new eyes.

— Marcel Proust

Published as: Srivastava M, Payne JL (2022) On the incongruence of genotype-phenotype and fitness landscapes. PLOS Computational Biology 18(9): e1010524.

<https://doi.org/10.1371/journal.pcbi.1010524>

Authors' contributions: M.S. : Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing; J.L.P. : Conceptualization, Funding acquisition, Investigation, Project administration, Supervision, Writing – review & editing

ABSTRACT

The mapping from genotype to phenotype to fitness typically involves multiple nonlinearities that can transform the effects of mutations. For example, mutations may contribute additively to a phenotype, but their effects on fitness may combine non-additively because selection favors a low or intermediate value of that phenotype. This can cause incongruence between the topographical properties of a fitness landscape and its underlying genotype-phenotype landscape. Yet, genotype-phenotype landscapes are often used as a proxy for fitness landscapes to study the dynamics and predictability of evolution. Here, we use theoretical models and empirical data on transcription factor-DNA interactions to systematically study the incongruence of genotype-phenotype and fitness landscapes when selection favors a low or intermediate phenotypic value. Using the theoretical models, we prove a number of fundamental results. For example, selection for low or intermediate phenotypic values does not change simple sign epistasis into reciprocal sign epistasis, implying that genotype-phenotype landscapes with only simple sign epistasis motifs will always give rise to single-peaked fitness landscapes under such selection. More broadly, we show that such selection tends to create fitness landscapes that are more rugged than the underlying genotype-phenotype landscape, but this increased ruggedness typically does not frustrate adaptive evolution because the local adaptive peaks in the fitness landscape tend to be nearly as tall as the global peak. Many of these

results carry forward to the empirical genotype-phenotype landscapes, which may help to explain why low- and intermediate-affinity transcription factor-DNA interactions are so prevalent in eukaryotic gene regulation.

AUTHOR SUMMARY

How do mutations change phenotypic traits and organismal fitness? This question is often addressed in the context of a classic metaphor of evolutionary theory — the fitness landscape. A fitness landscape is akin to a physical landscape, in which genotypes define spatial coordinates, and fitness defines the elevation of each coordinate. Evolution then acts like a hill-climbing process, in which populations ascend fitness peaks as a consequence of mutation and selection. It is becoming increasingly common to construct such landscapes using experimental data from high-throughput sequencing technologies and phenotypic assays, in systems such as macromolecules and gene regulatory circuits. Although these landscapes are typically defined by molecular phenotypes, and are therefore more appropriately referred to as genotype-phenotype landscapes, they are often used to study evolutionary dynamics. This requires the assumption that the molecular phenotype is a reasonable proxy for fitness, which need not be the case. For example, selection may favor a low or intermediate phenotypic value, causing incongruence between a fitness landscape and its underlying genotype-phenotype landscape. Here, we study such incongruence using a diversity of theoretical models and experimental data from gene regulatory systems. We regularly find incongruence, in that fitness landscapes tend to comprise more peaks than their underlying genotype-phenotype landscapes. However, using evolutionary simulations, we show that this increased ruggedness need not impede adaptation.

2.1 INTRODUCTION

Characterizing the relationship between genotype and phenotype is key to our understanding of evolution [1, 2]. For quantitative phenotypes, such as the expression level of a gene or the enzymatic activity of a protein, this relationship can be formalized as a genotype-phenotype landscape [3]. In such a landscape, genotypes represent coordinates in an abstract genotype space and their phenotype defines the elevation of each coordinate in this space [4]. The topographical properties of genotype-phenotype landscapes, such as their ruggedness, are influenced by epistasis [5] — non-additive interactions between mutations in their contribution to phenotype. These topographical properties have important evolutionary consequences, because they determine how mutation brings forth the phenotypic variation upon which selection acts [6, 7].

Technological advances are facilitating the construction and analysis of empirical genotype-phenotype landscapes at ever-increasing resolution, scale, and scope [8]. Example phenotypes include the enzymatic activity [9], binding affinity [10], allosteric profile [11], and fluorescence intensity of proteins [12], as well as exon inclusion levels [3], the expression levels of genes driven by regulatory elements [13], the expression patterns of gene regulatory circuits [14–16], and flux through metabolic pathways [17]. In analyzing the topographical properties of these landscapes and their evolutionary consequences, an assumption is often made that phenotype is a proxy for fitness [12, 18–24], thus rendering genotype-phenotype landscapes equivalent to fitness landscapes [25]. While this assumption may be justified under certain conditions, such as in directed protein evolution experiments [26], it is often the case that the relationship between phenotype and fitness is not so straightforward. For example, fitness may depend upon more phenotypes than those being assayed [27] or the relationship between phenotype and fitness may be inherently nonlinear, for example reflecting a tradeoff between the costs and benefits associated with a phenotype [28]. In the latter case, selection may favor a low or intermediate phenotypic value [29, 30]; e.g., an intermediate gene expression level, [31–33], enzyme efficiency [34] or protein production rate or activity [35]. Such non-linearities are a cause of epistasis [16, 36–38], and they can transform the effects of mutations as they map onto phenotype and fitness [37], thus rendering the topographical properties of a fitness landscape qualitatively different from those of its underlying genotype-phenotype landscape (Fig. 2.1A). While we do not doubt that workers in the field are well aware that the topographical properties of a fitness landscape can differ from those of its underlying genotype-phenotype landscape, a systematic study of these differences is lacking.

Selection for low or intermediate phenotypic values is especially relevant to transcription factor-DNA interactions [39]. Transcription factors are sequence-specific DNA binding proteins that help regulate gene expression. They do so by binding DNA sequences (transcription factor binding sites) in regulatory regions such as promoters and

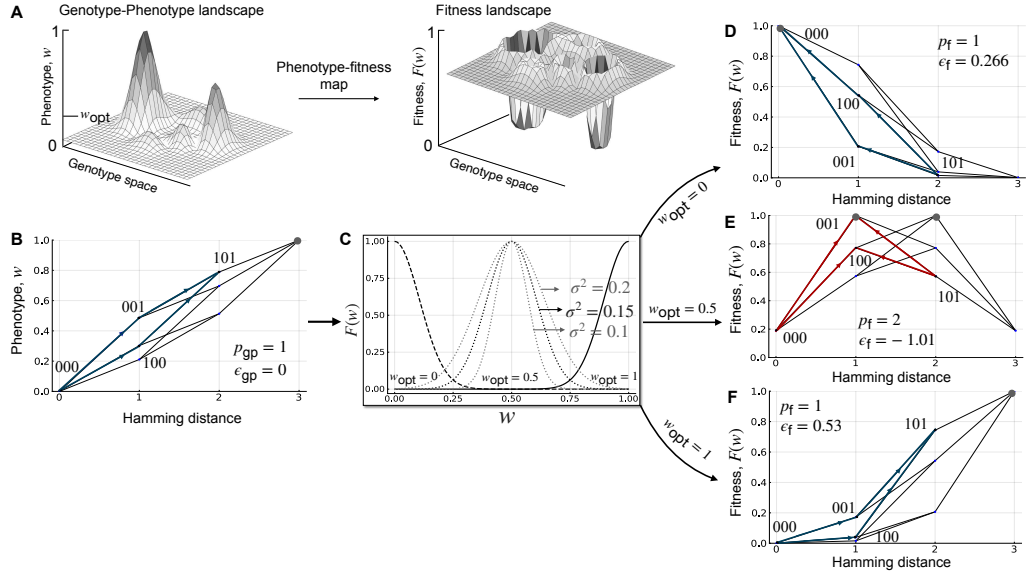


FIGURE 2.1: INCONGRUENCE. (A) Schematic illustration of how selection for an intermediate phenotypic value w_{opt} can make a genotype-phenotype landscape incongruent with the resulting fitness landscape. (B) An additive three-locus, biallelic genotype-phenotype landscape with a single peak (gray filled circle; $p_{\text{gp}} = 1$). One pairwise interaction is highlighted in blue. It exhibits no magnitude epistasis ($\epsilon_{\text{gp}} = 0$) or sign epistasis. (C) The Gaussian phenotype-fitness map (Eq. 2.1) is shown for three values of w_{opt} (dashed line, $w_{\text{opt}} = 0$; dotted line, $w_{\text{opt}} = 0.5$; solid line, $w_{\text{opt}} = 1$), with three values of σ shown for $w_{\text{opt}} = 0.5$. (D) Applying the Gaussian phenotype-fitness map with $w_{\text{opt}} = 0$ to the genotype-phenotype landscape results in a single-peaked fitness landscape (gray filled circle; $p_{\text{f}} = 1$). The same pairwise interaction from (B) is highlighted in blue. It exhibits positive epistasis ($\epsilon_{\text{gp}} = 0.266$), but no sign epistasis. (E) Applying the Gaussian phenotype-fitness map with $w_{\text{opt}} = 0.5$ to the genotype-phenotype landscape results in a multi-peaked fitness landscape (gray filled circles; $p_{\text{f}} = 2$). The same pairwise interaction from (B) is highlighted in red. It exhibits negative epistasis ($\epsilon_{\text{gp}} = -1.01$), as well as reciprocal sign epistasis. (F) Applying the Gaussian phenotype-fitness map with $w_{\text{opt}} = 1$ to the genotype-phenotype landscape results in a single-peaked fitness landscape (gray filled circle; $p_{\text{f}} = 1$). The same pairwise interaction from (A) is highlighted in blue. It exhibits positive epistasis ($\epsilon_{\text{gp}} = 0.53$), but no sign epistasis. In panels B,D-F, arrows point from genotypes with lower phenotypic or fitness values to genotypes with higher phenotypic or fitness values. The no sign epistasis motif is highlighted in blue and the reciprocal sign epistasis motif in red. Note the symmetry of landscapes in panels D and F.

enhancers to recruit or block the recruitment of RNA polymerase [40]. The regulatory effect of such a binding event depends in part on the affinity with which the DNA sequence is bound by the transcription factor [41, 42]. As such, binding affinity is an important molecular phenotype of transcription factor binding sites, upon which selection acts [43, 44]. While it is commonly assumed that selection increases binding affinity [21, 43–47], several lines of evidence suggest that selection for low or intermediate binding affinity also influences the evolution of transcription factor binding sites. For example, paralogous transcription factors often bind the same DNA sequences with high affinity, but different DNA sequences with low affinity [48, 49]. If an optimal gene expression pattern requires binding by just one of several transcription factor paralogs, such specificity can be achieved using low-affinity transcription factor binding sites, resulting in selection for low binding affinity [50]. Additional documented cases in which low-affinity binding sites play important regulatory roles include negative auto-regulation by high-copy number transcription factors in *Escherichia coli* [51], where high-affinity binding sites cause suboptimal noise suppression, and developmental patterning in *Ciona intestinalis* embryos [52, 53], where high-affinity binding sites cause deleterious ectopic gene expression patterns. Moreover, low-affinity binding sites are commonly observed in the regulatory portfolios of a diversity of organisms, including bacteria [51], yeast [54], fly [49, 55, 56], sea stars and sea urchins [57], as well as humans [58].

How incongruent are the topographies of genotype-phenotype and fitness landscapes when selection favors a low or intermediate phenotypic value? How does this depend on the ruggedness of the genotype-phenotype landscape? These are important questions, because the topography of a fitness landscape has implications for several evolutionary phenomena, including the evolution of genetic diversity [59], reproductive isolation [60], and sex [61], as well as the predictability of the evolutionary process itself [62]. How much we can learn about these phenomena from knowledge of a genotype-phenotype landscape depends on the genotype-phenotype landscape’s congruence with the fitness landscape. Despite decades of research on fitness landscapes [63, 64], these questions have not been addressed even in the context of classical theoretical models, such as Mt. Fuji [65], House-of-Cards [66], or NK landscapes [67]. They have also not been addressed in the context of biophysical models of genotype-phenotype landscapes or empirical genotype-phenotype landscapes. Here, we fill this knowledge gap by defining local and global measures of incongruence, which describe the topographical differences between a genotype-phenotype landscape and the corresponding fitness landscape when selection favors a low or intermediate phenotypic value. We use these measures to study incongruence in the context of the aforementioned theoretical models [65–67] and derive some fundamental results that are applicable to all empirical genotype-phenotype landscapes. We then consider the specific case of genotype-phenotype landscapes of transcription factor-DNA interactions, by first looking at an idealised biophysical model [68] and then taking a step further, by analysing 1,137 empirical genotype-phenotype landscapes, wherein

genotypes are transcription factor binding sites and the phenotype is a measure of relative binding affinity [21]. We study transcription factor-DNA interactions because there is strong biological motivation for studying selection for low or intermediate binding affinity, as discussed above, and because a large number of empirical genotype-phenotype landscapes of transcription factor-DNA interactions are publicly available [21], thus facilitating the statistical analysis of their incongruence.

2.2 RESULTS

We first present our measures of incongruence and the phenotype-to-fitness map used to incorporate the effect of selection for a low or intermediate phenotypic value. We use these measures and this map to study the incongruence of randomly generated genotype-phenotype landscapes, specifically two-locus biallelic landscapes and multi-locus biallelic landscapes. For the latter, we use NK landscapes [67], which include the corner cases of Mt. Fuji [65] ($K = 0$) and House-Of-Cards [66] ($K = N - 1$) landscapes (see Table 2.1 for a list of symbols). We then apply the principles learned from these model genotype-phenotype landscapes to genotype-phenotype landscapes of transcription factor-DNA interactions, first in the context of the mismatch model [68] (which enables us to study landscapes with more than two alleles per locus), and then in the context of empirical measurements of transcription factor-DNA interactions [21]. Finally, we study the evolutionary consequences of landscape incongruence.

w, w_{opt}	Phenotypic value, Optimal phenotypic value
σ	Strength of selection
$F(w)$	Fitness value
$\epsilon_{\text{gp}}, \epsilon_{\text{f}}$	Epistasis in the genotype-phenotype landscape (gp) and fitness landscape (f) respectively
$p_{\text{gp}}, p_{\text{f}}$	Number of peaks in the genotype-phenotype landscape (gp) and fitness landscape (f) respectively
L	Length of the sequence
K	Ruggedness parameter of the NK model
a	Number of alleles at each locus
m	Number of mismatches in the mismatch model
$\langle l \rangle$	Average length of adaptive walk
$\langle f \rangle$	Mean fitness at equilibrium

TABLE 2.1: LIST OF SYMBOLS

2.2.1 Landscape incongruence

Our goal is to quantify the topographical differences between a fitness landscape and its underlying genotype-phenotype landscape, when selection favors a low or intermediate phenotypic value. To do so, we define measures of landscape incongruence, at both a local and a global scale. At a local scale, we quantify differences in pairwise epistasis amongst loci in the genotype-phenotype landscape relative to the same loci in the fitness landscape (Methods). We do so by classifying the type of magnitude epistasis (ϵ) as additive (i.e., no epistasis; $\epsilon = 0$), positive ($\epsilon > 0$), or negative ($\epsilon < 0$). For a pair of loci, we then compare the type of epistasis in the genotype-phenotype landscape (ϵ_{gp}) to the type of epistasis in the fitness landscape (ϵ_f), and report whether this is the same in the two landscapes. We do the same for an additional classification of epistasis based on the absence or presence of sign epistasis. This results in three categories – no sign epistasis, simple sign epistasis, or reciprocal sign epistasis [69]. At a global scale, we quantify incongruence as the difference in the number of peaks in the fitness landscape (p_f) relative to the genotype-phenotype landscape (p_{gp}). Taken together, these local and global measures allow us to determine the extent to which selection for a low or intermediate phenotypic value increases or decreases the ruggedness of the fitness landscape, relative to the genotype-phenotype landscape.

2.2.2 Phenotype-to-fitness map

To study the effect of selection for low or intermediate phenotypic values, we use a Gaussian phenotype-to-fitness map $F(w)$ centred around an optimal phenotypic value w_{opt} ,

$$F(w) = \exp \left[- \left(\frac{w - w_{opt}}{\sigma} \right)^2 \right]. \quad (2.1)$$

The parameter σ determines the strength of selection by controlling how rapidly fitness decreases as the phenotype w deviates from the optimal phenotype w_{opt} . Increasing σ decreases the strength of selection, because it broadens the fitness map around w_{opt} and thus decreases the fitness differences between similar phenotypes. Similar maps are commonly used in evolutionary modeling frameworks, such as Fisher’s Geometric model [70, 71] and models of speciation [72], as well as in biophysical models of intermolecular interactions [29], including transcription factor-DNA interactions [46]. However, we emphasize that most of our measures of incongruence depend only on the rank ordering of phenotypic or fitness values, and are therefore independent of the exact shape of the phenotype-fitness map, so long as it is symmetric.

Fig. 2.1B-F illustrates the application of the phenotype-to-fitness map to a simple three-locus, biallelic genotype-phenotype landscape. This landscape is purely additive, and as a consequence, it exhibits no epistasis and has only one peak (Fig. 2.1B). Applying

the phenotype-to-fitness map (Fig. 2.1C) to this landscape can change the amount and type of epistasis, as well as the location and number of peaks in the resulting fitness landscape, relative to the genotype-phenotype landscape (Fig. 2.1D-F). It can therefore cause incongruence between genotype-phenotype and fitness landscapes. Whereas this schematic and our analyses below pertain to a single phenotype, extending our model to multiple phenotypes is straightforward, as we later discuss.

2.2.3 Two-locus biallelic genotype-phenotype landscapes

We first study incongruence using the simplest form of genotype-phenotype landscape that is capable of exhibiting epistasis: a two-locus biallelic landscape. We represent genotypes as binary strings of length $L = 2$ and randomly assign a phenotype w_i to each genotype i , which we draw from a uniform distribution between 0 and 1. We then apply the Gaussian phenotype-fitness map (Eq. 2.1) to generate the corresponding fitness landscape. We repeat this process 10,000 times for values of $w_{\text{opt}} \in [0, 1]$ (in increments of 0.01), and report the probability that the type of epistasis in the genotype-phenotype landscape is the same as in the fitness landscape. We first differentiate between no magnitude epistasis, positive epistasis, and negative epistasis, and then between no sign epistasis, simple sign epistasis, and reciprocal sign epistasis.

Incongruence in magnitude epistasis is highest when selection favors low phenotypic values

To determine whether the type of magnitude epistasis is the same in the genotype-phenotype landscape and fitness landscape, we calculate the product $\epsilon_{\text{gp}} \cdot \epsilon_{\text{f}}$, which will be positive when the two landscapes have the same type of epistasis. We use this product, rather than a discrete categorization, to ensure analytical tractability. We obtain an analytical expression for $\epsilon_{\text{gp}} \cdot \epsilon_{\text{f}}$ by assuming σ to be large (Supplementary Material, Derivation 1):

$$\epsilon_{\text{f}} \cdot \epsilon_{\text{gp}} = \frac{1}{\sigma^2} \left[2\epsilon_m \cdot \epsilon_{\text{gp}} + (2w_{\text{opt}} - \sum_i w_i) \cdot \epsilon_{\text{gp}}^2 \right], \quad (2.2)$$

where w_i is the phenotypic value of genotype i with $i \in \{0, 1\}^2$ and $\epsilon_m = w_{00}w_{11} - w_{01}w_{10}$, which is also known as multiplicative epistasis [73].

The probability that the type of epistasis is the same in the genotype-phenotype and fitness landscapes, $P(\epsilon_{\text{gp}} \cdot \epsilon_{\text{f}} > 0)$ can be computed using Monte-Carlo methods and yields results in good agreement with the randomly generated genotype-phenotype landscapes when σ is large. This is shown in Fig. 2.2, where three trends are immediately apparent. First, for large σ , the probability that the type of epistasis is the same in the genotype-phenotype landscape and the fitness landscape increases as the optimal phenotype w_{opt} increases, in agreement with the intuition that selection for large phenotypic values leaves the genotype-phenotype landscape mostly unchanged, except for the nonlinear rescaling introduced by the phenotype-fitness map. Second, even when

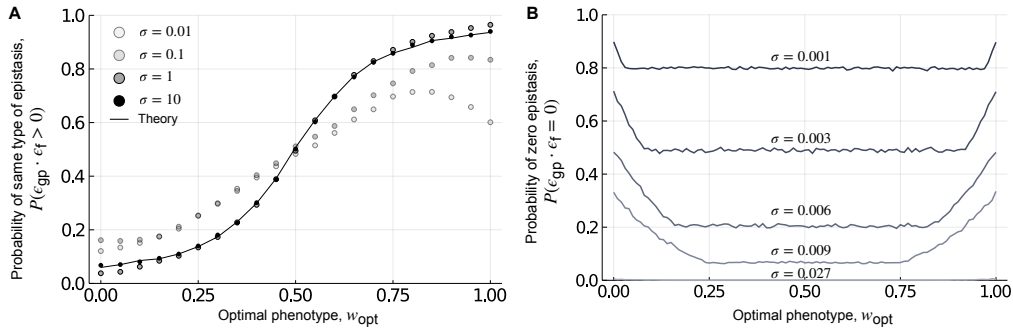


FIGURE 2.2: LOCAL INCONGRUENCE: MAGNITUDE EPISTASIS. (A) The probability of retaining the type of epistasis, shown in relation to the optimal phenotype w_{opt} . The black line shows the theoretical prediction and the dots show the results from randomly generated genotype-phenotype landscapes for different values of σ . The theoretical approximation agrees well with the results from randomly generated genotype-phenotype landscapes for large σ (i.e., $\sigma \geq 1$). (B) The probability of observing zero magnitude epistasis in the fitness landscape, shown in relation to w_{opt} , for different values of σ , which we selected to show the range of variation in the probability of observing zero epistasis.

$w_{opt} = 1$, the probability that the type of epistasis is the same in the genotype-phenotype landscape and the fitness landscape is less than one. The reason is the nonlinear rescaling introduced by the phenotype-to-fitness map does not guarantee conservation of the type of magnitude epistasis, even though it does preserve the rank ordering of fitness values. Results obtained with the randomly generated genotype-phenotype landscapes show this effect becomes even more pronounced as σ decreases. This means that as the strength of selection for w_{opt} increases, so does the likelihood of landscape incongruence. Finally, as σ decreases, the probability of retaining the type of epistasis from the genotype-phenotype landscape in the fitness landscape is not maximized at $w_{opt} = 1$, but rather at a smaller w_{opt} (e.g., at $w_{opt} \approx 0.8$ when $\sigma = 0.01$, Fig. 2.2A). The reason is as σ decreases, the fitness function becomes extremely narrow and more phenotypic values are mapped to zero fitness (within computer precision i.e. any fitness $< 5.0 \times 10^{-324} \approx 0$), resulting in cases where the genotype-phenotype landscape exhibits epistasis (i.e., $\epsilon_{gp} \neq 0$), but the fitness landscape does not (i.e., $\epsilon_f = 0$), because all fitness values are zero. Fig. 2.2B shows this is more likely to occur when σ is small and as w_{opt} approaches its extreme values of 0 or 1. In these randomly generated landscapes, the probability of obtaining negative epistasis is the same as the probability of obtaining positive epistasis, and the probability of conserving the type of epistasis is independent of the type of epistasis in the genotype-phenotype landscape. In sum, these results show that selection for low or intermediate phenotypic values can modify the genotype-phenotype landscape, such that the resulting fitness landscape exhibits a different type of magnitude epistasis, and

this effect is most pronounced when selection is strong and the optimal phenotypic value is low.

Incongruence in sign epistasis is highest when selection favors intermediate phenotypic values

Next, we categorized landscapes as exhibiting no sign epistasis, simple sign epistasis, or reciprocal sign epistasis. These three motifs are shown in the center panel of Fig. 2.3, where arrows point from genotypes with a lower phenotypic or fitness value to genotypes with a higher phenotypic or fitness value. Because the presence of sign epistasis only depends upon the partial ordering of the phenotypic values, we expect to retain the motif from the genotype-phenotype landscape in the fitness landscape as $w_{\text{opt}} \rightarrow 1$. We also expect to retain the motif as $w_{\text{opt}} \rightarrow 0$, because selecting for $w_{\text{opt}} = 0$ simply flips all the arrows from the genotype-phenotype landscape in the fitness landscape, which does not change the categorization of the motif. Thus, we expect the probability of retaining the motif from the genotype-phenotype landscape in the fitness landscape to be "U" shaped and symmetric about $w_{\text{opt}} = 0.5$. Fig. 2.3 shows the probability of retaining or changing the motif from the genotype-phenotype landscape in the fitness landscape for 10,000 randomly generated two-locus biallelic landscapes, grouped according to the motif in the genotype-phenotype landscape. These results confirm the expected "U" shape of the probability of retaining the type of epistasis from the genotype-phenotype landscape in the fitness landscape as w_{opt} is varied from 0 to 1.

Simple sign epistasis cannot be modified into reciprocal sign epistasis

When the genotype-phenotype landscape has the no sign epistasis motif, selection for an intermediate phenotypic value can transform the landscape into the simple sign epistasis motif or the reciprocal sign epistasis motif (Fig. 2.3A), in line with recent results on Fisher's Geometric model [74]. When the genotype-phenotype landscape has the simple sign epistasis motif, selection for an intermediate phenotypic value can transform the landscape into a no sign epistasis motif, but not into a reciprocal sign epistasis motif (Fig. 2.3B; Supplementary Material, Proof 1). Because reciprocal sign epistasis is a necessary condition for multiple peaks [5], this implies that genotype-phenotype landscapes with only simple sign epistasis motifs will always give rise to single peaked fitness landscapes, using the phenotype-fitness map considered here. Finally, when the genotype-phenotype landscape has the reciprocal sign epistasis motif, selection for an intermediate phenotypic value transforms the landscape into the no sign epistasis motif or the simple sign epistasis motif with equal probability (Fig. 2.3C; Supplementary Material, Proof 2). Moreover, the probability of retaining the motif from the genotype-phenotype landscape in the fitness landscape is always higher than the probability of changing it when the genotype-phenotype landscape has the simple sign epistasis motif (Fig. 2.3B) or the reciprocal sign epistasis motif (Fig. 2.3C). This is not always true when the genotype-phenotype

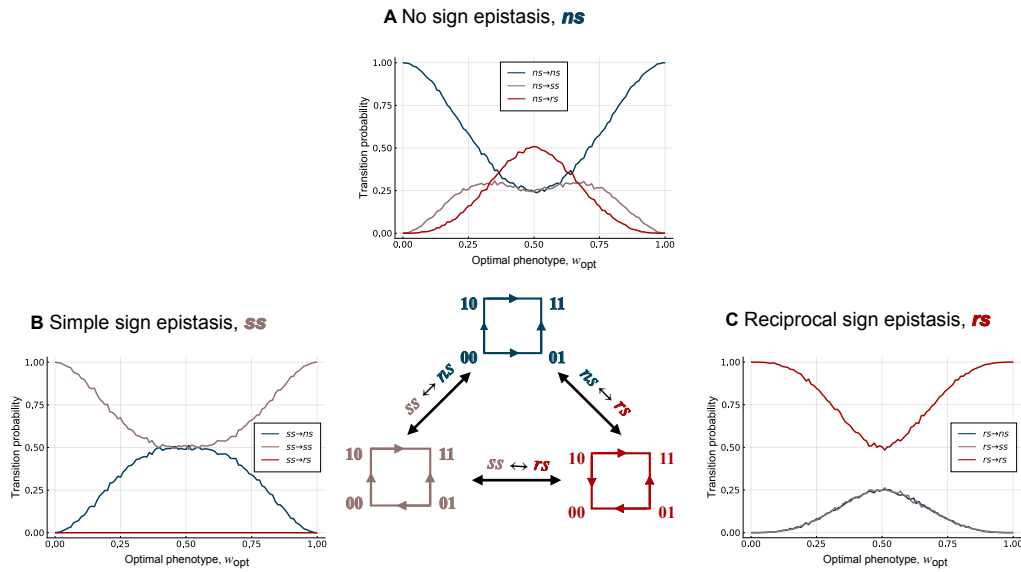


FIGURE 2.3: LOCAL INCONGRUENCE: SIGN EPISTASIS. The probability of retaining or changing the type of epistasis in the genotype-phenotype landscape, relative to the fitness landscape, shown in relation to the optimal phenotypic value w_{opt} . Data are grouped based on whether the genotype-phenotype landscapes exhibits (A) no sign epistasis (blue), (B) simple sign epistasis (brown), or (C) reciprocal sign epistasis (red). The colours of the lines represent the type of epistasis in the resulting fitness landscape. These results are independent of σ , because they only depend on the rank ordering of fitness values. Notice the “U” shape of the probability of retaining the type of epistasis in each panel.

landscape has the no sign epistasis motif (Fig. 2.3A), because at intermediate w_{opt} the landscape is most likely to transform into the reciprocal sign epistasis motif. Taken together, these results show that selection for intermediate phenotypic values can modify genotype-phenotype landscapes with no sign epistasis into fitness landscapes with sign epistasis and vice versa.

The inferences about pairwise interactions can be carried forward to multi-locus biallelic landscapes because their genotype spaces, which are L -dimensional hypercubes, are composed of two-dimensional squares. Due to the adjacency of squares, in the three-locus case, the motifs of four out of the six squares are sufficient to determine the motifs of the rest of the squares, and for any L , the motifs of only $2^{L-2} \cdot (L-1)$ squares are necessary to determine the motifs of all of the remaining squares. This is only a fraction $2/L$ of all the squares in the hypercube (because the total number of faces in an L -dimensional hypercube is $2^{L-2} \cdot \binom{L}{2}$), which is clearly minuscule for large L . However, pairwise interactions are not sufficient to predict peak patterns, which may result from higher-order interactions [75]. We study these in the next section.

2.2.4 Multi-locus biallelic genotype-phenotype landscapes

We use the NK model [67] to study multi-locus biallelic genotype-phenotype landscapes (Methods). In this model, each locus in a genotype of length L epistatically interacts with K other loci (whereas N is typically used to denote the number of loci in this model, we use L for consistency with the rest of our text). As corner cases, this model includes Mt. Fuji landscapes [65] when $K = 0$ and House-of-Cards landscapes [66] when $K = L - 1$. For each combination of L and K , we use this model to randomly generate a genotype-phenotype landscape. We then apply the Gaussian phenotype-fitness map (Eq. 2.1) to generate a fitness landscape. We repeat this process 10,000 times for $w_{\text{opt}} \in [0, 1]$ (in increments of 0.01), and report the average of the absolute change in the number of peaks, i.e., $\langle |p_f - p_{\text{gp}}| \rangle$, where p_f is the number of peaks in the fitness landscape and p_{gp} is the number of peaks in the genotype-phenotype landscape. This is our measure of global incongruence. We use the absolute value of the change in number of peaks so that we can average over many realisations of genotype-phenotype landscapes. However, since the sign of change is also important, we discuss that as well in the following sections.

Mt. Fuji landscapes

We begin with Mt. Fuji genotype-phenotype landscapes. Because these are single-peaked, selection for a low or intermediate phenotypic value can only maintain or increase the number of peaks from the genotype-phenotype landscape in the fitness landscape. Fig. 2.4A shows this change in the number of peaks, and Fig. 2.4B shows the probability that the number of peaks changes, in relation to w_{opt} for landscapes with $L = 2$ to $L = 8$ loci. These trends are symmetric about $w_{\text{opt}} = 0.5$, because Mt. Fuji landscapes are additive, so selecting for $w_{\text{opt}} = 0$ is equivalent to selecting for $w_{\text{opt}} = 1$ with regard to the change in the number of peaks. The reason is that selecting for $w_{\text{opt}} = 0$ flips all of the arrows in the fitness landscape, relative to the genotype-phenotype landscape, which changes the location of the peak, but does not change the number of peaks. This is illustrated in Fig. 2.1B,D. For a detailed explanation of the shape of the curves in Fig. 2.4A, see Supplementary Material, Note 1. More obvious is the increase in the number of peaks, and the probability that the number of peaks increases, as L increases, the latter converging to one for all values of w_{opt} except the extreme cases of $w_{\text{opt}} = 0$ and $w_{\text{opt}} = 1$. Note, however, that a high probability of increase in the number of peaks does not necessarily correspond to a high increase in the number of peaks, as can be seen from the different positions of the maxima in Fig. 2.4A,B (Supplementary Material, Note 1). For large L (> 10), the expected number of peaks in the fitness landscape increases exponentially with L [74]. In sum, these results show that selection for intermediate phenotypic values readily transforms Mt. Fuji genotype-phenotype landscapes, which

are smooth and single-peaked, into rugged fitness landscapes, and that this effect is most pronounced for large L and intermediate w_{opt} .

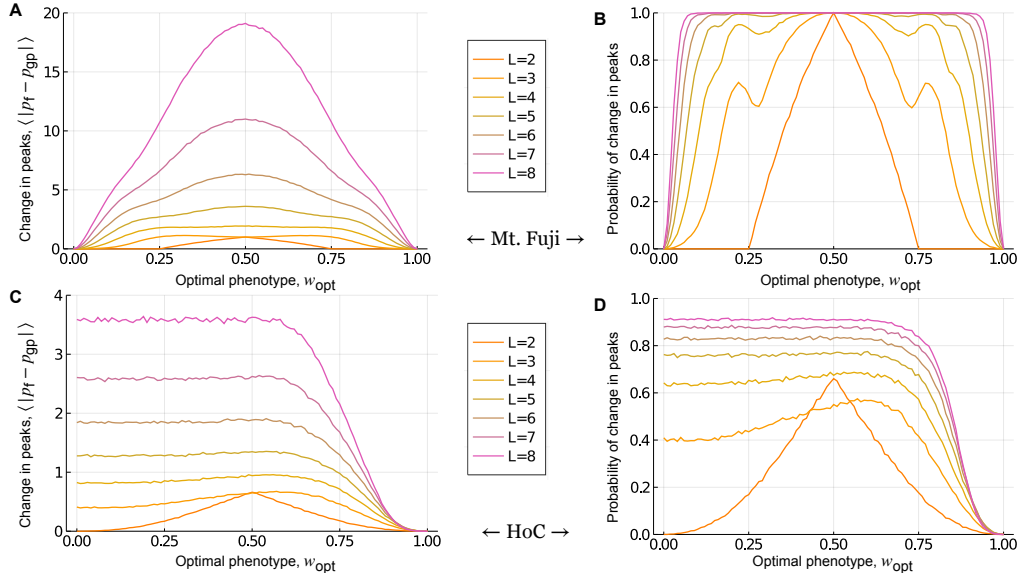


FIGURE 2.4: GLOBAL INCONGRUENCE : MT. FUJI AND HOUSE-OF-CARDS GENOTYPE-PHENOTYPE LANDSCAPES. The absolute change in the number of peaks and the probability that the number of peaks changes in the fitness landscape, relative to (A,B) Mt. Fuji and (C,D) House-of-Cards genotype-phenotype landscapes, shown in relation to w_{opt} for $L \in \{2, 3, 8\}$. These results are independent of σ , because they only depend on the rank ordering of fitness values.

House-of-Cards landscapes

We next study House-of-Cards genotype-phenotype landscapes. These landscapes are highly rugged, with an average of $\frac{2^L}{L+1}$ peaks [76], whereas the maximum possible number of peaks is 2^{L-1} . As such, selection for a low or intermediate phenotypic value can either increase or decrease the number of peaks in the fitness landscape, relative to the genotype-phenotype landscape. Fig. 2.4C shows this change in the number of peaks, and Fig. 2.4D shows the probability that the number of peaks changes, in relation to w_{opt} for landscapes with $L = 2$ to $L = 8$ loci. The change in the number of peaks is symmetric about $w_{\text{opt}} = 0.5$ for the two-locus case, where the number of peaks does not change as $w_{\text{opt}} \rightarrow 0$ or $w_{\text{opt}} \rightarrow 1$. However, this symmetry is lost for $L > 2$. The reason is that although the phenotype-fitness map flips all of the arrows in the fitness landscape when $w_{\text{opt}} = 0$, relative to the genotype-phenotype landscape, this does not guarantee conservation of the number of peaks. The number of peaks is jointly determined by the adjacent faces of the hypercube and thus, only very specific changes in the directions of arrows in the genotype-phenotype landscape guarantees conservation of the number of peaks in the fitness landscape (S. 2.4). However, in contrast to Mt.

Fuji genotype-phenotype landscapes, the magnitude of change in the number of peaks increases very little with L , despite an exponential increase in the maximum number of possible peaks. Moreover, the probability that the number of peaks changes is still less than one for large L . Finally, for large L , both the change in the number of peaks and the probability that the number of peaks changes are independent of w_{opt} , so long as w_{opt} is sufficiently less than one. This observation depends on the probability distribution used to generate these landscapes. Because the 2^L phenotypes in the NK model are drawn from a uniform distribution, nearly the same number of these phenotypes will be close to the optimal phenotype w_{opt} , so long as L is sufficiently large. Thus, averaging over all possible configurations of the genotype-phenotype landscape yields the same value of $\langle |p_f - p_{\text{gp}}| \rangle$ for every w_{opt} . For sufficiently large L , this value is given by (Supplementary Material, Derivation 2):

$$\langle |p_f - p_{\text{gp}}| \rangle \approx \sqrt{\frac{2^L \cdot (L - 1)}{2\pi(L + 1)^2}} \text{ for } L \gg 1 \quad (2.3)$$

So far we have focused on the absolute change $\langle |p_f - p_{\text{gp}}| \rangle$ in the number of peaks in the fitness landscape, relative to the genotype-phenotype landscape. For House-of-Cards genotype-phenotype landscapes, selection for a low or intermediate phenotypic value can either increase or decrease the number of peaks. We were therefore interested in finding out which outcome is more likely. While one might expect a decrease to be more likely, due to the extreme ruggedness of House-of-Cards genotype-phenotype landscapes, we find that the number of peaks is equally likely to increase or decrease (Supplementary Material, Proof 3). In sum, these results show that in House-of-Cards genotype-phenotype landscapes, selection for a low or intermediate phenotypic value increases or decreases the number of peaks in the fitness landscape with equal probability, and the severity as well as the probability of this change increases with L and is largely independent of w_{opt} .

Global incongruence decreases as the ruggedness of the genotype-phenotype landscape increases

Finally, we study NK genotype-phenotype landscapes, which bridge the gap between Mt. Fuji and House-of-Cards landscapes in terms of ruggedness, as K increases from 0 to $L - 1$. Fig. 2.5 shows the absolute change in the number of peaks in the fitness landscape relative to the genotype-phenotype landscape for genotypes of length $L = 5$ and $L = 8$, as K is increased from 0 to $L - 1$. Note the gradual transition from the trends observed for Mt. Fuji genotype-phenotype landscapes (Fig. 2.4A) to those observed for House-of-Cards genotype-phenotype landscapes (Fig. 2.4C) as K increases. From these trends, we conclude four principles of how the ruggedness of an NK genotype-phenotype landscape influences its incongruence with the fitness landscape. As an NK genotype-phenotype landscape becomes more rugged, incongruence (1) loses symmetry about $w_{\text{opt}} = 0.5$, (2) becomes less sensitive to w_{opt} , and (3) decreases in severity, at least in terms of the absolute change in the number of peaks. Finally, the probability

of increasing the number of peaks is always greater than or equal to the probability of decreasing the number of peaks, with the equality holding for House-of-Cards genotype-phenotype landscapes. This last principle is both intuitive and informative — it tells us that on average, selection for low or intermediate values is more likely to increase the ruggedness of a fitness landscape, relative to the genotype-phenotype landscape. Thus, selection for low or intermediate values is more likely to break than to create phenotypic correlations between mutationally similar genotypes as they map onto fitness, rendering fitness landscapes more rugged than their underlying genotype-phenotype landscapes. In the subsequent sections we address whether and how these results apply to genotype-phenotype landscapes with more than two alleles per locus, specifically in the context of a biophysical model and experimental measurements of transcription factor-DNA interactions.

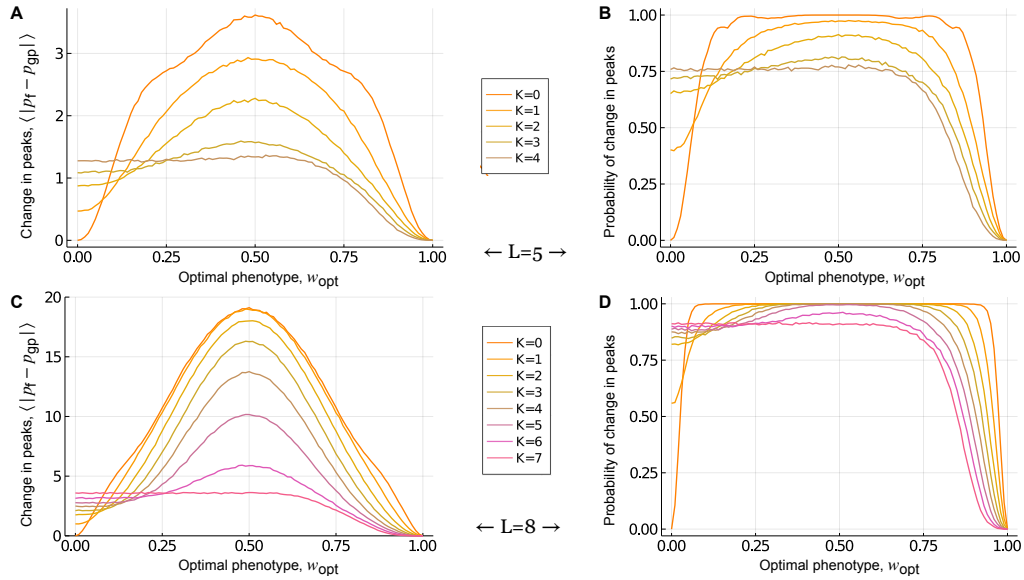


FIGURE 2.5: GLOBAL INCONGRUENCE : NK GENOTYPE-PHENOTYPE LANDSCAPES. The absolute change in the number of peaks in the fitness landscape, relative to the genotype-phenotype landscape, is shown in relation to w_{opt} for genotypes of length (A) $L = 5$ and (C) $L = 8$, as K increases from zero to $L - 1$. The corresponding probability of change in the number of peaks is shown in relation to w_{opt} for genotypes of length (B) $L = 5$ and (D) $L = 8$, as K increases from zero to $L - 1$. These results are independent of σ , because they only depend on the rank ordering of fitness values.

2.2.5 Genotype-phenotype landscapes of transcription factor-DNA interactions

Motivated by the common usage of low- and intermediate-affinity transcription factor binding sites in the regulatory portfolios of a diversity of organisms [49, 51, 54–57], we now study the incongruence of genotype-phenotype landscapes of transcription factor-

DNA interactions and the corresponding fitness landscapes generated after selection for low or intermediate phenotypic values. In these landscapes, genotypes represent DNA sequences — transcription factor binding sites — and the phenotype of a DNA sequence is the affinity with which it binds a transcription factor [21]. Because the regulatory effects of transcription factor-DNA interactions are partly determined by binding affinity [41, 42] and mutations to transcription factor binding sites can alter binding affinity [48, 77], the topographies of genotype-phenotype landscapes of transcription factor-DNA interactions have important implications for the evolution of gene regulation [21]. We study these landscapes using both a biophysical model and experimental measurements of transcription factor-DNA interactions. We focus on transcription factor binding sites of length $L = 8$, because this is the length of the binding sites assayed by protein binding microarrays [77, 78] — the data used to construct the empirical genotype-phenotype landscapes of transcription factor-DNA interactions [21].

The mismatch model

We first study genotype-phenotype landscapes of transcription factor-DNA interactions generated using the so-called mismatch model [46, 47, 68]. The key assumption of this model is that the binding energy of a DNA sequence is a linear function of the number of mismatches between the sequence (genotype) and a transcription factor’s consensus sequence — the sequence it binds with the highest affinity. Further, each mismatch is assumed to have the same energetic cost and these costs combine additively to determine binding energy. This model results in a Mt. Fuji-like, permutation-invariant genotype-phenotype landscape, wherein the phenotype only depends on the number of differences between the genotype and the consensus sequence, but not on which loci in the genotype differ from the consensus sequence. Although this is a simplified model, it provides an opportunity to study the effects of having more than two alleles per locus and serves as a bridge to our analyses of empirical transcription factor-DNA interactions.

To ensure that these results are comparable to the results on the empirical landscapes, we consider the negative of the binding energy as our phenotype, such that sequences that bind more strongly are assigned higher phenotypic values. We assume a phenotypic value of 1 for the genotype that is identical to the consensus sequence. For each mismatch between a genotype and the consensus sequence, we deduct a small positive value e , such that the phenotypic value of a genotype with m mismatches is $A_m = 1 - m \cdot e$. Due to the permutation invariance, the genotype-phenotype landscape is highly degenerate, such that the number of genotypes N_m in a mismatch class m is distributed according to the asymmetric binomial distribution:

$$N_m = (a - 1)^m \frac{L!}{(L - m)!m!}, \quad (2.4)$$

where a is the number of alleles per locus ($a = 4$ for transcription factor binding sites, because they are DNA sequences). Fig. 2.6A shows this distribution for transcription factor binding sites of length $L = 8$. Note that N_m is maximized at $m = 6$.

Fig. 2.6B shows the incongruence between genotype-phenotype landscapes and fitness landscapes of transcription factor-DNA interactions constructed using the mismatch

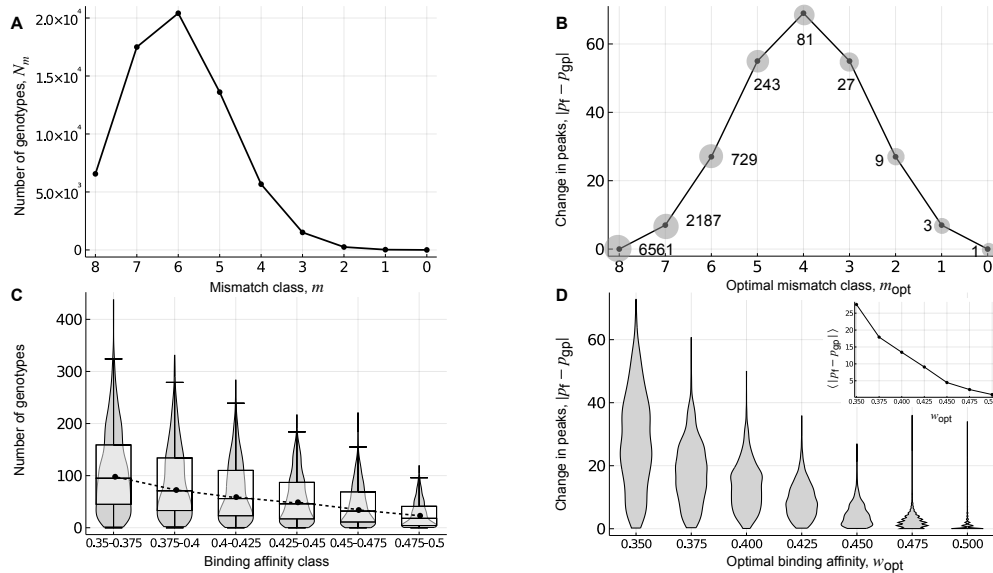


FIGURE 2.6: GLOBAL INCONGRUENCE : GENOTYPE-PHENOTYPE LANDSCAPES OF TRANSCRIPTION FACTOR-DNA INTERACTIONS. Landscapes constructed using (A,B) the mismatch model and (C,D) experimental measurements from protein binding microarrays for 1,137 eukaryotic transcription factors. (A) The number of genotypes, shown in relation to mismatch class. (B) The absolute change in the number of peaks in the fitness landscape, relative to the genotype-phenotype landscape, shown in relation to the optimal mismatch class m_{opt} . Labels indicate the number of genotypes per peak in the fitness landscape. Note the symmetry in the absolute change in the number of peaks around mismatch class $m_{opt} = 4$, as well as the tripling of the number of genotypes per peak for each increment in m_{opt} . The grey shaded circles are a schematic representation of the growing width of the peaks. (C) The number of genotypes per binding affinity class, where protein binding microarray E -scores are used as a proxy for relative binding affinity. Violin plots show the distribution, and box-and-whisker plots the 25-75% quartiles, across genotype-phenotype landscapes for the 1,137 transcription factors. Closed symbols and the dashed line denote the median of each distribution. (D) Violin plots of the distribution of the absolute change in the number of peaks in the fitness landscape, relative to the genotype-phenotype landscape, shown in relation to the optimal binding affinity w_{opt} for $\sigma = 0.15$. The inset shows the mean absolute change in the number of peaks, in relation to w_{opt} . The x-axes in (A,B) are arranged such that binding affinity increases when read from left to right, in qualitative agreement with the x-axes in (C,D). The results in panels A-C are independent of σ .

model, under selection for an optimal mismatch class m_{opt} , reported in terms of the absolute change in the number of peaks in the fitness landscape, relative to the genotype-phenotype landscape. Based on our analysis of Mt. Fuji genotype-phenotype landscapes, we anticipated asymmetric incongruence about the mismatch class $m_{\text{opt}} = 6$, because the distribution of the number of genotypes per mismatch class is asymmetric with a maximum at mismatch class $m = 6$ (Fig. 2.6A) and we expected all genotypes in the optimal mismatch class m_{opt} to be peaks. However, we observe symmetric incongruence about $m_{\text{opt}} = 4$ (Fig. 2.6B). This occurs because $a > 2$, which renders some genotypes in the same mismatch class mutational neighbours. Consequently, the peaks can be broad and include more than one genotype, thus resembling plateaus. Specifically, each genotype has $(a - 1)^m - 1$ mutational neighbours that are in the same mismatch class and are therefore part of the same peak. This leaves $\binom{L}{m}$ clusters of genotypes to be peaks in each mismatch class m , an expression that is maximized with $m = 4$ when $L = 8$, thus forcing symmetry in the absolute change in the number of peaks about $m_{\text{opt}} = 4$. However, the width of the peaks increases as $(a - 1)^m$, leading to a tripling of peak width for each increment in m (Fig. 2.6B). Thus, even in this idealised genotype-phenotype landscape, features of empirical landscapes of TF-DNA interactions begin to emerge, such as broad peaks.

Empirical landscapes

We now study genotype-phenotype landscapes of transcription factor-DNA interactions generated using experimental data from protein binding microarrays [78] (Methods). For all possible DNA sequences of length $L = 8$, these data include an enrichment score (E -score) ranging from -0.5 to 0.5 that serves as a proxy for relative binding affinity, such that higher E -scores correspond to higher binding affinities [77, 78]. We have previously used these data to construct genotype-phenotype landscapes for 1,137 eukaryotic transcription factors, in which the surface of each landscape was defined by the E -score [21]. Due to limitations in the reproducibility of E -scores across microarray designs for genotypes that are bound non-specifically or with very low affinity [77, 78], each genotype-phenotype landscape only includes DNA sequences with an E -score exceeding a threshold of 0.35, which corresponds to a false discovery rate of 0.001 [77]. As shown in our previous work [21], these landscapes tend to exhibit little, if any, reciprocal sign epistasis and therefore comprise few peaks. As such, they bear resemblance to the genotype-phenotype landscapes constructed using the mismatch model. An important difference, however, is that genotypes in the lower half of the E -score range are omitted from each landscape due to the reproducibility issues mentioned above. Fig. 2.6C shows the distributions of the number of genotypes across all 1,137 genotype-phenotype landscapes, grouped into six binding affinity classes. Whereas the lowest binding affinity class contains the most genotypes, we cannot determine if this is the true maximum, because we do not know what these distributions look like for lower binding affinity classes. However,

assuming the energetic contribution of each binding site to be additive [68], we expect lower binding affinity classes to have fewer genotypes and the maximum to occur at an intermediate binding affinity class, as can be seen in the mismatch model and other models in literature [43].

Fig. 2.6D shows the incongruence between these 1,137 empirical genotype-phenotype landscapes and their corresponding fitness landscapes, under selection for an optimal binding affinity w_{opt} , reported in terms of the absolute change in the number of peaks in the fitness landscape, relative to the genotype-phenotype landscape for $\sigma = 0.15$. Interestingly, the effect of changing σ is small (S2.5B Fig) because increasing σ not only decreases the range of variation of fitness values, but also decreases the uncertainty in these values (Methods), leaving the number of peaks largely unchanged. The mean and variance in the absolute change in the number of peaks decreases as w_{opt} increases to 0.5, in line with the intuition that selection for high w_{opt} generates fitness landscapes that are topographically similar to the underlying genotype-phenotype landscape. More precisely, the percentage of landscapes that show a change in the number of peaks decreases from around 99% to 40%, as we go from $w_{\text{opt}} = 0.35$ to $w_{\text{opt}} = 0.5$. As anticipated from our results with additive genotype-phenotype landscapes (Mt. Fuji and the mismatch model), selection for low or intermediate phenotypic values is more likely to increase than to decrease the number of peaks, although the relative fraction of increase depends upon w_{opt} – while around 99% of the landscapes show an increase in peaks for $w_{\text{opt}} = 0.35$, this value decreases to around 40% for $w_{\text{opt}} = 0.5$. Further, when we separately analyzed the single-peaked ($\approx 66.75\%$) and multi-peaked ($\approx 33.25\%$) genotype-phenotype landscapes, we found the single-peaked landscapes to be more incongruent (S2.5A Fig), in line with our results from the previous section. In sum, these results show that selection for low or intermediate phenotypic values tends to increase the ruggedness of a fitness landscape, relative to the underlying genotype-phenotype landscapes, rendering genotype-phenotype landscapes a poor proxy for fitness landscapes under such selection.

Mismatch model and empirical landscapes show different kinds of global incongruence

In contrast to genotype-phenotype landscapes constructed using the mismatch model, the incongruence of genotype-phenotype landscapes constructed using protein binding microarray data is highest for the lowest binding affinity class, rather than an intermediate class. There are three non-mutually exclusive explanations for this. First, these empirical landscapes are not purely additive [21], unlike the landscapes constructed with the mismatch model, so we do not expect perfect symmetry about an intermediate w_{opt} . Second, as previously mentioned, protein binding microarray data do not capture the full range of binding affinity, so the lowest binding affinity class in our data (E -score = 0.35), which contains the most genotypes (Fig. 2.6C), is unlikely to be the lowest binding affinity class, but rather an intermediate binding affinity class. Third, while the binding

affinity of a sequence is highly correlated with the binding affinities of its mutational neighbors [79], this correlation is not perfect, so neighboring genotypes that are in the same mismatch class may not have sufficiently similar binding affinities to be considered part of the same peak, unlike in the mismatch model.

There are two additional differences between the incongruence of the empirical genotype-phenotype landscapes and those constructed using the mismatch model that are worth highlighting. First, in the empirical landscapes, the average height of the peaks is maximised when selecting for intermediate binding affinities and is lowest when selecting for the highest affinity (S2.6A Fig), whereas peak height is independent of m_{opt} in the mismatch model. Second, in the empirical landscapes, peak width is maximized when selecting for the highest binding affinity ($w_{\text{opt}} = 0.5$) (S2.6B Fig), whereas it is maximized when selecting for the lowest binding affinity ($m_{\text{opt}} = 8$) in the mismatch model (Fig. 2.6B). These two differences in incongruence are important for understanding evolutionary simulations on these landscapes, which are the focus of the next section.

2.2.6 Evolutionary consequences

A key finding of our analyses so far is that selection for low or intermediate phenotypic values is more likely to increase than to decrease the number of peaks in the fitness landscape, relative to the genotype-phenotype landscape. Since the ruggedness of a fitness landscape has important implications for a diversity of evolutionary phenomena [59–62], we now study the evolutionary consequences of this finding. We do so using two metrics: (1) the length $\langle l \rangle$ of a greedy adaptive walk, averaged over all possible genotypes as starting points, and (2) the mean population fitness at equilibrium $\langle f \rangle$ under deterministic mutation-selection dynamics (i.e., assuming an infinite population size). These metrics provide complementary information to measures of landscape ruggedness, such as the number of peaks. To see why, consider our results with the mismatch model, in which the fitness landscape had one or more global peaks of equal height (Fig. 2.6A,B). We observed single-peaked fitness landscapes when the optimal mismatch class was $m_{\text{opt}} = 0$ or $m_{\text{opt}} = 8$, and the most rugged fitness landscapes when the optimal mismatch class was $m_{\text{opt}} = 4$. We also observed a tripling of the number of genotypes per peak as m_{opt} increased from 0 to 8, such that the landscape with $m_{\text{opt}} = 0$ comprised a single peak with one genotype, the landscape with $m_{\text{opt}} = 4$ comprised 70 peaks with 81 genotypes per peak, and the landscape with $m_{\text{opt}} = 8$ comprised 1 peak with 6,561 genotypes. Which landscape topography is most conducive to adaptive evolutionary change?

NK landscapes

On NK landscapes, the length of the greedy adaptive walk $\langle l \rangle$ is minimised when $w_{\text{opt}} = 0.5$ for $K < L - 1$, whereas for $K = L - 1$ (House-of-Cards landscapes), $\langle l \rangle$

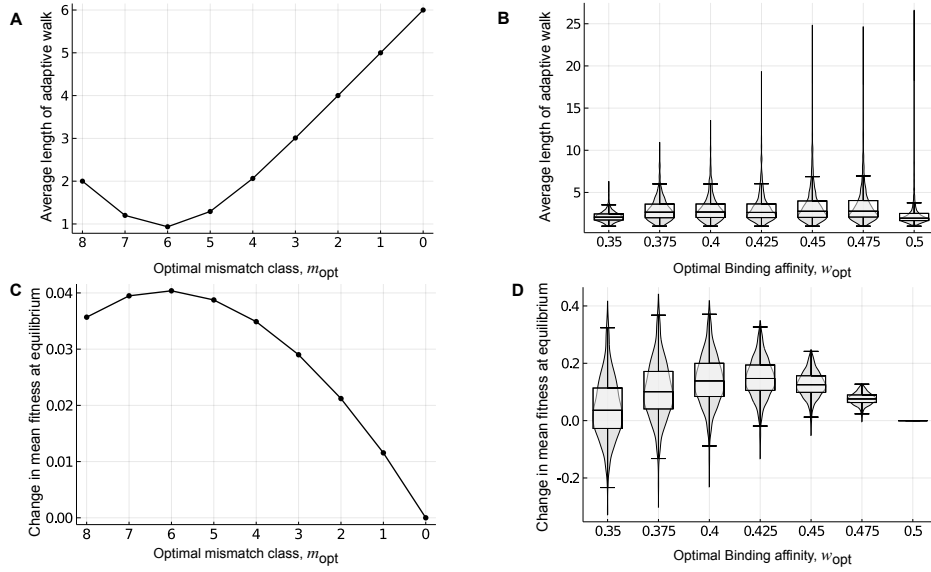


FIGURE 2.7: RUGGED FITNESS LANDSCAPES NEED NOT IMPEDE ADAPTATION. The average length of an adaptive walk $\langle l \rangle$ and the change in mean population fitness at equilibrium $\langle f \rangle$ is shown for landscapes of transcription factor-DNA interactions generated using (A,C) the mismatch model for $e = 0.05$ and (B,D) protein binding microarray data. In (B,D), violin plots show the distribution, and box-and-whisker plots the 25-75% quartiles, across the 1,137 empirical landscapes for each optimal binding affinity w_{opt} . The large variability of $\langle l \rangle$ at intermediate and high w_{opt} is a consequence of the random diffusion of the population on non-peak plateaus, which results in longer walks. In (C), $\mu = 0.1$, $\sigma = 1$ while in (D) $\mu = 0.1$, $\sigma = 0.15$.

is independent of w_{opt} . Moreover, the change in mean fitness at equilibrium is always positive, increases with L and K , and tends to be maximized at $w_{\text{opt}} = 0.5$ (Supplementary Material, Note 2). In sum, these results show that on biallelic landscapes, selection for an intermediate phenotypic value decreases the length of a greedy adaptive walk and increases mean fitness at equilibrium, despite increasing the overall ruggedness of the fitness landscape, relative to the genotype-phenotype landscape. Whereas previous work has shown that peak accessibility increases with alphabet cardinality (a) due to the existence of indirect paths [10, 80], we show below that despite this increased accessibility, results qualitatively similar to the biallelic case ($a = 2$) also hold for the multi-allelic case of the mismatch model and of the empirical landscapes of TF-DNA interactions ($a = 4$).

Mismatch model

In the fitness landscapes generated using the mismatch model, the length of the greedy adaptive walk averaged over all starting genotypes is given by

$$\langle l \rangle = \frac{\sum_{m=0}^L (a-1)^m \cdot \frac{L!}{(L-m)!m!} \cdot |m_{\text{opt}} - m|}{a^L}. \quad (2.5)$$

Fig. 2.7A shows this expression for $L = 8$ and $a = 4$. Due to the additive nature and degeneracy of this genotype-phenotype landscape, the fitness landscapes have monotonically decreasing fitness values as the mutational distance from any peak increases. Therefore, the length of any individual greedy adaptive walk is simply the absolute difference of the mismatch class of the starting genotype and that of the optimal mismatch class (i.e., $|m_{\text{opt}} - m|$). As such, the average length of the greedy adaptive walk is minimized when selecting for $m_{\text{opt}} = 6$ (Fig. 2.7A), because this maximizes the number of genotypes in adaptive peaks (Fig. 2.6B; 28 global peaks \times 729 genotypes per peak). If instead, we chose to start the greedy walk from only non-peak genotypes, the walk would still be minimized for $m_{\text{opt}} = 6$, because this class has the maximum number of genotypes that are Hamming distance one away from the peak genotypes. Recall that in each fitness landscape generated upon selection for m_{opt} , all the peaks are of the same height and therefore, the greedy walk always terminates on a global peak. This is in contrast with the NK landscapes, and as we will see below, the empirical landscapes.

For each m_{opt} , we next calculated the change in mean fitness at equilibrium under deterministic mutation-selection dynamics, relative to the fitness landscape generated for $m_{\text{opt}} = 0$. To do so, we exploit the permutation-invariance of these landscapes to group genotypes into a lower-dimensional state space defined by mismatch class (Methods). Specifically, we construct a transition matrix that defines the probability that a genotype from one mismatch class mutates into another, based on the frequency and fitness of genotypes in each mismatch class. We iterate this matrix until we reach steady state, which is guaranteed by the Frobenius-Perron theorem to be independent of initial conditions, because the matrix is irreducible [81]. Fig. 2.7C shows the change in mean fitness at equilibrium in relation to the optimal mismatch class $m_{\text{opt}} = 0$. Our first observation is that the change in mean fitness is always positive for $m_{\text{opt}} > 0$ when $\mu = 0.1$ and $\sigma = 1$, even though selection for such mismatch classes always causes an increase in the number of peaks in the fitness landscape, relative to the underlying genotype-phenotype landscape (Fig. 2.6B). Our second observation is that the change in mean fitness relative to m_{opt} is always unimodal. For example, selection for $m_{\text{opt}} = 6$ maximizes the change in mean fitness when $\mu = 0.1$ and $\sigma = 1$ (Fig. 2.7C). More generally, the m_{opt} that maximizes the change in mean fitness depends on the interplay between the mutation rate μ and the strength of selection σ . See S2.10 Fig for the phase diagram. When selection is strong or mutation is weak, mismatch class $m_{\text{opt}} = 8$ maximizes the change in mean fitness. As the strength of selection decreases or the mutation rate increases, the mismatch class that maximizes the change in mean fitness decreases from $m_{\text{opt}} = 8$ to $m_{\text{opt}} = 7$ and then to $m_{\text{opt}} = 6$, where it remains for weak selection or high mutation rates. This is because when σ is small (i.e., selection is strong), it is costly to step down from a peak and therefore, selecting for the class with the broadest peak ($m_{\text{opt}} = 8$) leads to the highest equilibrium mean population fitness. As σ increases and selection becomes weaker, it is no longer as costly to step down from a peak and therefore, selecting for

the class with the maximum number of genotypes in peaks ($m_{\text{opt}} = 6$) maximizes mean fitness at equilibrium. While this effect is reminiscent of the “survival of the flattest” [82] phenomenon, the crucial difference is that in the mismatch model, the peaks always have the same height and thus, there is no trade-off between peak height and width. Another way of altering the strength of selection is by changing e , the energetic cost of a mismatch. Larger e corresponds to stronger selection and the phase diagram changes accordingly (S2.11 Fig).

Empirical landscapes

The empirical genotype-phenotype landscapes are topologically more complex [79, 83] than the genotype-phenotype landscapes generated with the mismatch model, which are regular graphs (i.e., every genotype has $(a - 1) \cdot L$ mutational neighbors). We therefore used simulations to calculate the average length of a greedy adaptive walk $\langle l \rangle$ in these empirical landscapes [84], initiating the walks from all non-peak genotypes in the fitness landscape. Moreover, each binding affinity measurement (E -score) in the empirical landscapes is associated with a noise threshold that is used to determine whether two genotypes truly differ in phenotype (Methods). This noise threshold can cause landscapes to have large non-peak plateaus, in which many genotypes have indistinguishable fitness. We therefore modified the greedy adaptive walk such that when a non-peak plateau was encountered, we chose a random mutational neighbor of indistinguishable fitness for the next step in the walk, disallowing reversion mutations. We repeated this process until the plateau was traversed and a sequence with higher fitness was reached. Finally, we terminated the walk when a peak sequence was reached. Therefore, the walk was primarily a deterministic greedy walk, with some stochasticity due to the non-peak plateaus.

Fig. 2.7B shows $\langle l \rangle$, averaged over 100 simulations of the adaptive walk from each initial condition, in relation to w_{opt} . In contrast to the mismatch model, $\langle l \rangle$ is shortest for $w_{\text{opt}} = 0.35$ and $w_{\text{opt}} = 0.5$ and slightly higher for intermediate w_{opt} . This is because, as in the mismatch model, $\langle l \rangle$ is correlated with the total number of genotypes in peaks, which depends upon both the number of peaks and their widths. Whereas selecting for $w_{\text{opt}} = 0.35$ leads to the largest number of peaks (Fig. 2.6D), selecting for $w_{\text{opt}} = 0.5$ leads to the broadest peaks (S2.6B Fig), thus explaining the minimisation of $\langle l \rangle$ when selecting for these extreme phenotypes. We note that in the absence of plateaus (i.e., when the noise threshold is zero), $\langle l \rangle$ increases monotonically with w_{opt} , because in this case, $\langle l \rangle$ is inversely correlated with the number of peaks in the fitness landscape, which decreases monotonically with w_{opt} . Regardless of the noise threshold, when selecting for low or intermediate phenotypic values, most walks terminate at a local, rather than the global fitness peak. However, these local peaks tend to be nearly as high as the global peak, especially when selecting for intermediate phenotypic values (S2.9 Fig). These results hold for all noise thresholds, and are therefore not a consequence of the existence of

plateaus in the genotype-phenotype landscapes (see Supplementary Material, Note 2 for explanation).

Next, we simulated deterministic mutation selection dynamics (Methods). For each w_{opt} , Fig. 2.7D shows the change in mean fitness at equilibrium for $\mu = 0.1, \sigma = 0.15$, relative to the fitness landscape generated after selecting for $w_{\text{opt}} = 0.5$. As in the mismatch model, the change in mean fitness tends to be positive, despite the increase in the number of peaks caused by selection for low or intermediate phenotypic values (Fig. 2.6D). Moreover, the mean change in fitness at equilibrium is maximized when selection favors an intermediate phenotypic value ($w_{\text{opt}} = 0.425$ for $\mu = 0.1, \sigma = 0.15$). This can be explained by the average peak heights and widths that occur when selecting for intermediate phenotypic values (S2.6A and S2.6B Fig for $\sigma = 0.15$). Exactly which w_{opt} maximizes the change in mean fitness depends on the interplay between the mutation rate μ and the strength of selection σ (S2.12 Fig), converging on 0.425 for large μ and σ , similar to the convergence seen at $m_{\text{opt}} = 6$ for the mismatch model.

2.3 DISCUSSION

Non-linear relationships between phenotype and fitness have been observed in a diversity of biological systems [28, 31, 33, 35, 85, 86], often reflecting a trade-off between the costs and benefits of a particular phenotype, such as antibiotic resistance [87, 88]. It is well established that these non-linearities, either in the genotype-phenotype or phenotype-fitness map are a cause of epistasis [36–38, 74]. Moreover, when mutations have epistatic interactions in their contribution to phenotype, a non-linear phenotype-fitness map can change the form of these interactions from negative to positive, or vice versa [16], as well as introduce or remove sign epistasis [85]. Our work complements these empirical observations and expands upon previous theoretical work [74, 89], by systematically quantifying how and how often selection for a low or intermediate phenotypic value introduces or removes epistasis in the fitness landscape. Specifically, we show that the probability of changing the type of magnitude epistasis (e.g., positive to negative) is highest when selecting for low phenotypic values and the probability of introducing or removing sign epistasis is highest when selecting for intermediate phenotypic values. Further, we show that the simple sign epistasis motif cannot be converted into reciprocal sign epistasis, implying that genotype-phenotype landscapes with only simple sign epistasis motifs will remain single peaked and globally congruent to their corresponding fitness landscapes.

Another key finding of our analysis is that selection for low or intermediate phenotypic values is more likely to increase than to decrease the number of peaks, with the probability of the two types of change being equal only in House-of-Cards genotype-phenotype landscapes. This means that additive genotype-phenotype landscapes will tend to be incongruent with their fitness landscapes, whereas rugged genotype-phenotype land-

scapes will not. While increased landscape ruggedness is typically thought to frustrate the evolutionary process, because it limits the amount of adaptive phenotypic variation mutation can bring forth [4], our evolutionary simulations show this need not be the case. Specifically, we find that the rugged fitness landscapes caused by selection for low or intermediate phenotypic values comprise local adaptive peaks that are nearly as tall as the global adaptive peak. Moreover, these local peaks tend to be accessible from throughout the landscape via a small number of sequential mutations that monotonically increase fitness. As a result, the mean population fitness at equilibrium is almost always higher when selecting for low or intermediate phenotypic values than when selecting for a high phenotypic value.

Finally, while there have been several attempts at investigating genotype-phenotype-fitness landscapes in the past [74, 90, 91] – some models have been very specific to the system of interest and others are agnostic to any mechanistic details [92]. We tried to bridge this gap, by applying a Fisher’s Geometric model-like phenotype-fitness function to biophysically motivated and empirically determined genotype-phenotype landscapes. Further, our results on the mismatch model and the 1,137 landscapes of TF-DNA interactions may help to explain the prevalence of low- and intermediate-affinity binding sites in the control of gene expression. Prior work has suggested an entropic argument [93]: As with certain RNA secondary structures [94] or regulatory circuit motifs [95], low- and intermediate-affinity binding sites appear more frequently simply because they are more “findable”. That is, because a transcription factor binds more distinct DNA sequences with low or intermediate affinity than with high affinity, low- and intermediate-affinity binding sites are more likely to evolve to control gene expression. Our work complements this “arrival of the frequent” argument [96] by showing that low- and intermediate-affinity binding sites are not only more likely to arise *de novo* due to their increased frequency, selection for such sites also generates fitness landscapes that are more conducive to adaptation – in terms of increased mean fitness at equilibrium and decreased average length of adaptive walks, than fitness landscapes that were generated by selection for high affinity binding sites.

We made several modeling assumptions to simplify our analyses of transcription-factor DNA interactions, the relaxation of which may open new avenues for future research. First, we assumed a single nonlinearity in the relationship between genotype, phenotype, and fitness. As noted by Domingo et al. [37], “from transcription to RNA processing, translation, and protein folding and all the way up to protein activity and cellular fitness, there are many layers of biological organization where the effects of a mutation can be transformed.” To date, our knowledge of how such nonlinearities combine to modify genotype-phenotype landscapes is based on a small number of experimental studies (e.g., ref [16]). A systematic theoretical analysis could build off the work presented here, for example by incorporating the sigmoidal relationship between binding site occupancy and binding affinity in modeling transcription factor-DNA interactions [97], such that

fitness depends nonlinearly on occupancy, rather than affinity. Other nonlinearities, such as those caused by transcription factor cooperativity [97], could also be included.

Second, we assumed that selection acts directly on a single phenotype — binding affinity. While this assumption is common in models of the evolution of transcription factor binding sites [21, 43, 46] and is supported by empirical data [43, 44], the relationship between binding affinity and fitness is not so direct, because it is modulated by gene expression. Gene expression depends on a variety of factors, including the presence, arrangement, and affinities of binding sites for other competing or cooperating transcription factors in promoters and enhancers [98], as well as local sequence context [99], chromatin context [100], DNA methylation [101], and local transcription factor concentrations [102]. Existing modeling frameworks that relate the architecture of entire regulatory regions to gene expression patterns may provide a path forward [103], facilitating the study of landscape incongruence when fitness depends upon the multitude of molecular phenotypes characteristic of eukaryotic gene regulation. Alternatively, our modeling framework could be extended to include multiple phenotypes by defining fitness in terms of the differences between a vector of phenotypes and a vector of optimal phenotypes, rather than the scalars considered here. Incongruence could then be quantified between the fitness landscape generated by selecting for the highest phenotypic value of each phenotype and that generated by selecting for a combination of low and intermediate values of the phenotypes. Our results correspond to a special case of this extended model, wherein all phenotypes except one are exactly attuned to their optimal values.

Third, we assumed a Gaussian phenotype-fitness map, which is commonly employed in a diversity of modeling frameworks [29, 70, 72, 104], including those for transcription factor-DNA interactions [46]. Alternative symmetric phenotype-fitness maps (e.g., ref. [105]) will only affect our results quantitatively, because many of our findings, such as the changes in sign epistasis motifs and number of peaks, only depend on the rank ordering of fitness values. However, we expect asymmetric phenotype-fitness maps, such as those uncovered in experimental studies of biological systems such as viruses [30] and yeast [106], to affect our results qualitatively. For example, in our analyses of NK landscapes, we often observed symmetries in incongruence around an intermediate phenotypic value. These symmetries will almost certainly be lost. Understanding how asymmetric phenotype-fitness maps affect the incongruence of genotype-phenotype landscape is therefore an outstanding challenge.

Finally, our study may open new lines of research on dynamic genotype-phenotype and fitness landscapes [107–113]. For example, whereas we studied selection for a fixed phenotypic optimum, this optimum may in fact change in space or in time. Our results imply that even gradual changes in the phenotypic optimum may lead to abrupt changes in fitness landscape topography, which may have implications for an evolving population's ability to track this optimum and thus for population persistence and extinction. Moreover, because our measures of incongruence can be applied to any pair of land-

scapes so long as they are defined over the same set of genotypes, they are also applicable whenever a phenotype is mapped non-linearly to another phenotype. Ideally, we would be able to make inferences about phenotypic architecture based on the topographical properties of higher-level phenotypic or fitness landscapes – as was previously done for an antibiotic resistance phenotype [114]. However, because the phenotype-fitness map we study is not invertible, we can only make such inferences in limited cases. For instance, when the fitness landscape is single-peaked, we can make the probabilistic inference that the underlying genotype-phenotype landscape is also likely to be single-peaked, because the phenotype-fitness map is more likely to increase than to decrease the number of peaks. In contrast, when the fitness landscape has multiple peaks, we can only infer that the underlying genotype-phenotype landscape does not solely comprise simple sign epistasis motifs. Beyond that, we cannot infer the topographical properties of the underlying genotype-phenotype landscape. It could be smooth or rugged. Additionally, our measures may shed light on the kinds of topographical alterations induced by fluctuating environmental factors, such as DNA methylation [101] or the presence of protein partners [115] on transcription factor-DNA interactions. As our ability to experimentally interrogate such complexities in the relationship between genotype, phenotype, and fitness continues to improve, we anticipate a sharpened focus on landscape dynamics and their implications for the evolutionary process.

2.4 METHODS

Epistasis

In a genotype-phenotype landscape, we classify the type of magnitude epistasis between a pair of loci using the following linear combination of phenotypic values:

$$\epsilon_{gp} = w_{00} + w_{11} - w_{01} - w_{10}, \quad (2.6)$$

where w_i represents the phenotype of genotype $i \in \{0, 1\}^2$. When $\epsilon_{gp} = 0$, there is no magnitude epistasis, because the phenotypic effects of the two alleles combine additively; when $\epsilon_{gp} > 0$, there is positive epistasis, because the phenotypic effects of the two alleles are greater than expected based on their individual phenotypic effects; when $\epsilon_{gp} < 0$, there is negative epistasis, because the phenotypic effects of the two alleles are less than expected based on their individual phenotypic effects.

Analogously, in the fitness landscape, we classify the type of magnitude epistasis between a pair of loci using the following linear combination of fitness values:

$$\epsilon_f = F_{00} + F_{11} - F_{01} - F_{10}, \quad (2.7)$$

where F_i represents the fitness of the corresponding genotype i . As in the genotype-phenotype landscape, $\epsilon_f = 0$, $\epsilon_f > 0$, and $\epsilon_f < 0$ indicate additive, positive, and negative epistatic interactions among loci, respectively.

We use ϵ_{gp} and ϵ_f to calculate the fraction of pairs of loci that have the same type of epistasis in the genotype-phenotype landscape and the fitness landscape, which we determine as the product of ϵ_{gp} and ϵ_f . This is because the type of epistasis is the same in the two landscapes when $\epsilon_{gp} \cdot \epsilon_f > 0$. While it is theoretically possible for both ϵ_{gp} and ϵ_f to be 0, in which case the type of epistasis would be the same in the two landscapes yet the condition $\epsilon_{gp} \cdot \epsilon_f > 0$ would not be satisfied, this never happens in practice.

NK Landscapes

We constructed the NK landscapes using the adjacency neighbourhood scheme, wherein each locus i of a genotype of length L , interacts with K adjacent loci to the right of locus i and $0 \leq K \leq L - 1$. We used periodic boundary conditions, such that the L -th locus interacts with the first K loci, and so on.

The phenotype $w(\tau)$ of genotype τ is computed as the sum of the individual contributions of all loci, each of which depends on K other interacting loci:

$$w(\tau) = \sum_{i=1}^L f(\tau_i; \tau_i^1, \tau_i^2 \dots \tau_i^K), \quad (2.8)$$

where $f(\tau_i; \tau_i^1, \tau_i^2 \dots \tau_i^K)$ represents the contribution of the i -th locus, which depends on K other loci $\tau_i^1, \tau_i^2 \dots \tau_i^K$. We drew the contributions of each of the 2^{K+1} possible configurations from a uniform distribution between 0 and 1. Finally, we re-scaled the phenotypic values by subtracting the minimum and dividing by the maximum, such that they fell between 0 and 1.

Empirical genotype-phenotype landscapes of transcription factor-DNA interactions

We studied the incongruence of 1,137 genotype-phenotype landscapes of transcription factor-DNA interactions. The procedure for constructing these landscapes has been described elsewhere [21]. In brief, each landscape corresponds to a single transcription factor, the genotypes it contains represent DNA sequences of length $L = 8$ that specifically bind the transcription factor, and the phenotype of each sequence is a quantitative proxy for relative binding affinity, which defines the surface of the landscape. These phenotypes are reported as enrichment scores (E -scores) derived from protein binding microarrays. In each landscape, two genotypes are considered mutational neighbors if they differ by a single small mutation, specifically a point mutation or a small indel that shifts an entire contiguous binding site by a single base [79]. We performed all analyses on the dominant genotype network (i.e., the largest connected component of the network), which comprises the vast majority of genotypes in each landscape [21].

We used the Genonets Python package (version 0.31) to characterize the topographical properties of the empirical genotype-phenotype landscapes [116]. Specifically, we used this package to compute the number of peaks per landscape and the number of genotypes in the peaks of each landscape. These calculations rely on a noise threshold δ , which is used to determine whether two genotypes actually differ in phenotype. For each transcription factor, we used the value of δ reported in ref. [21], which was derived from a comparison of binding affinity measurements across two protein binding microarray designs.

While characterizing the topographies of the resulting fitness landscapes, we also had to transform δ following the rules of error propagation. Accordingly, the noise in fitness values, dF , depends on the noise in the phenotypic values $dw = \delta$ as follows:

$$dF = -2F \cdot \frac{w - w_{\text{opt}}}{\sigma^2} \cdot \delta. \quad (2.9)$$

To compute the number of peaks and their widths in the fitness landscapes, we adapted Genonets to specify different noise threshold values (dF) for each genotype.

Mutation-selection dynamics for the mismatch model

We grouped genotypes according to their mismatch class m and iterated a series of selection and mutation steps until the population reached an equilibrium distribution. In each selection step, the frequency of each mismatch class X_m was scaled by its fitness F_m and then normalized:

$$X_m^S(t+1) = \frac{F_m \cdot X_m(t)}{\sum_i F_i \cdot X_i}, \quad (2.10)$$

where $X_m^S(t+1)$ is the frequency of the mismatch class m after the selection step.

The selection step was followed by a mutation step, in which genotypes could either mutate within their mismatch class or mutate to an adjacent mismatch class. The total mutation probability is μ , so with probability $1 - \mu$, the genotype does not mutate. The frequency of each mismatch class was thus updated as

$$\begin{aligned} X_m^M(t+1) = & \left[\mu \cdot \frac{L - (m-1)}{L} \right] \cdot X_{m-1}^S(t+1) + \\ & \left[(1 - \mu) + \mu \cdot \frac{(a-2) \cdot m}{(a-1) \cdot L} \right] \cdot X_m^S(t+1) + \\ & \left[\mu \cdot \frac{m+1}{(a-1) \cdot L} \right] \cdot X_{m+1}^S(t+1), \end{aligned} \quad (2.11)$$

where $X_m^M(t+1)$ is the frequency of the mismatch class m after the mutation step and μ is the mutation probability. The mutation step implicitly accounts for the different number of genotypes per mismatch class. Finally, $X_m(t+1)$ was set to $X_m^M(t+1)$ and t was incremented. This process was repeated until equilibrium.

Mutation-selection dynamics for the empirical landscapes

The empirical landscapes comprise far fewer genotypes than the mismatch landscape. We therefore defined the state space of the empirical landscapes in terms of the individual genotypes in each landscape, merging each genotype with its reverse complement [79]. The mutational neighbors of each genotype were those that differed by a single small mutation, namely a point mutation or an indel that shifted an entire contiguous binding site by a single base [79]. In this analysis, we did not account for uncertainties in the fitness values, because these do not influence our results in this evolutionary regime (see below).

The recursion relation for mutation-selection dynamics in discrete time is

$$x_i(t+1) = \sum_j \mu^{d(\tau_i, \tau_j)} \cdot (1 - \mu)^{L-d(\tau_i, \tau_j)} \frac{f_j}{\bar{f}(\vec{x}, t)} x_j(t), \quad (2.12)$$

where \vec{x} is the vector of genotype frequencies, x_i is the frequency of the i th genotype, μ is the probability of a single point mutation or a small indel mutation, L is the length of the genotypes, $d(\tau_i, \tau_j)$ is the minimum of the mutational distance between genotypes τ_i and τ_j and between genotypes τ_i and the reverse complement of τ_j , f_i is the fitness of the i th genotype, and $\bar{f}(\vec{x}, t)$ is the mean population fitness at time t .

We linearized the dynamics by substituting $z_i(t) = \frac{x_i(t)}{\prod_{\tau=1}^{i-1} \bar{f}(\vec{x}, \tau)}$, yielding

$$z_i(t+1) = \sum_j \mu^{d(\tau_i, \tau_j)} (1 - \mu)^{L-d(\tau_i, \tau_j)} f_j z_j(t), \quad (2.13)$$

from which we retrieved the normalized genotype frequencies with $\vec{x} = \vec{z} / (\sum_i z_i)$. In matrix form, the dynamics are

$$\vec{z}(t+1) = M \cdot S \cdot \vec{z}(t), \quad (2.14)$$

where the mutation matrix M has elements $M_{ij} = \mu^{d(\tau_i, \tau_j)} \cdot (1 - \mu)^{L-d(\tau_i, \tau_j)}$ and the selection matrix S is a diagonal matrix with $S_{ii} = f_i$, where f_i is the fitness of sequence τ_i . We calculated the eigenvector corresponding to the largest eigenvalue of $M \cdot S$ to determine the equilibrium state of the dynamics, which is guaranteed by the Frobenius-Perron theorem to be unique and stable.

Because the fitness values have some uncertainty around them, we additionally performed a sensitivity analysis by adding Gaussian noise to the fitness values, with standard deviation equal to $1/3 \cdot dF$, such that 99.7% of the sampled fitness values fell within $F \pm \delta$. Our results are robust to these perturbations, with the exception of two parameter combinations – $\sigma = 0.1, \mu = 0.001$ and $\sigma = 0.1, \mu = 0.0025$. For small σ , dF is large and therefore such sensitivity is expected. However, the quantitative changes in our results were small, leading us to conclude that fitness uncertainties do not significantly influence our results in this evolutionary regime.

ACKNOWLEDGEMENTS

This work was supported by the Swiss National Science Foundation (Grant numbers PPOOP3_170604 and 310030_192541). We thank members of the Computational Biology Group at ETH, Michael Manhart, Joachim Krug, and David M. McCandlish for discussions.

DATA AVAILABILITY

All data used in this study are publicly available. See Aguilar-Rodriguez et al. (2017) "A thousand empirical adaptive landscapes and their navigability." *Nature Ecology & Evolution*, 1, 0045. Code used in this study is available [here](#).

REFERENCES

1. Alberch, P. From genes to phenotype: dynamical systems and evolvability. *Genetica* **84**, 5 (1991).
2. Lehner, B. Genotype to phenotype: lessons from model organisms for human genetics. *Nature Reviews Genetics* **14**, 168 (2013).
3. Julien, P., Miñana, B., Baeza-Centurion, P., Valcárcel, J. & Lehner, B. The complete local genotype–phenotype landscape for the alternative splicing of a human exon. *Nature Communications* **7** (2016).
4. Payne, J. L. & Wagner, A. The causes of evolvability and their evolution. *Nature Reviews Genetics* **20**, 24 (2019).
5. Poelwijk, F., Tănase-Nicola, S., Kiviet, D. & Tans, S. Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. *Journal of Theoretical Biology* **272**, 141 (2011).
6. Østman, B., Hintze, A. & Adami, C. Impact of epistasis and pleiotropy on evolutionary adaptation. *Proceedings of the Royal Society B: Biological Sciences* **279**, 247 (2012).
7. Weinreich, D., Watson, R. & Chao, L. Perspective: sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* **59**, 1165 (2005).
8. Kinney, J. B. & McCandlish, D. M. Massively Parallel Assays and Quantitative Sequence–Function Relationships. *Annual Review of Genomics and Human Genetics* **20**, 99 (2019).
9. Qiu, C., Erinne, O., Dave, J., Cui, P., Jin, H. & et al. High-Resolution Phenotypic Landscape of the RNA Polymerase II Trigger Loop. *PLoS Genetics* **12** (2016).

10. Wu, N., Dai, C. L., Olson, C. A., Lloyd-Smith, J. O. & Sun, R. Adaptation in protein fitness landscapes is facilitated by indirect paths. *eLife* **5**, e16965 (2016).
11. Tack, D., Tonner, P., Pressman, A., Olson, N., Levy, S., Romantseva, E., Alperovich, N., Vasilyeva, O. & Ross, D. The genotype-phenotype landscape of an allosteric protein. *Molecular Systems Biology* **17** (2021).
12. Sarkisyan, K. S., Bolotin, D. A., Meer, V., Usmanova, D. R., Mishin, S., Sharonov, G. V., Ivankov, D. N., Bozhanova, N. G., Baranov, M. S., Soylemez, O., Bogatyreva, N. S., Vlasov, P. K., Egorov, E. S., Logacheva, M. D., Kondrashov, S., Chudakov, D. M., Putintseva, E. V., Mamedov, I. Z., Tawfik, D. S., Lukyanov, K. A. & Kondrashov, F. A. Local fitness landscape of the green fluorescent protein. *Nature* **533**, 397 (2016).
13. De Boer, C. G., Vaishnav, E. D., Sadeh, R., Abeyta, E. L., Friedman, N. & Regev, A. Deciphering eukaryotic gene-regulatory logic with 100 million random promoters. *Nature Biotechnology* **38**, 56 (2019).
14. Lagator, M., Paixao, T., Barton, N. H., Bollback, J. P. & Guet, C. C. On the mechanistic nature of epistasis in a canonical cis-regulatory element. *eLife* **6**, e25192 (2017).
15. Schaerli, Y., Jimenez, A., Duarte, J., Mihajlovic, L., Renggli, J., Isalan, M., Sharpe, J. & Wagner, A. Synthetic circuits reveal how mechanisms of gene regulatory networks constrain evolution. *Molecular Systems Biology* **14** (2018).
16. Li, X., Lalic, J., Baeza-Centurion, P., Dhar, R. & Lehner, B. Changes in gene expression predictably shift and switch genetic interactions. *Nature Communications* **10** (2019).
17. Bassalo, M. C., Garst, A. D., Choudhury, A., Grau, W. C., Oh, E. J., Spindler, E., Lipscomb, T. & Gill, R. T. Deep scanning lysine metabolism in *Escherichia coli*. *Molecular Systems Biology* **14** (2018).
18. Schenk, M. F., Szendro, M. L., Salverda, I. G., Krug, J. & de Visser, J. A. G. Patterns of epistasis between beneficial mutations in an antibiotic resistance gene. *Molecular Biology and Evolution* **30**, 1779 (2013).
19. Gong, L. I., Suchard, M. A. & Bloom, J. D. Stability-mediated epistasis constrains the evolution of an influenza protein. *eLife* **2**, e00631 (2013).
20. Jiménez, J. I., Xulvi-Brunet, R., Campbell, G. W., Turk-MacLeod, R. & Chen, I. A. Comprehensive fitness landscape for RNA. *Proceedings of the National Academy of Sciences* **110**, 376 (2013).
21. Aguilar-Rodríguez, J., Payne, J. & Wagner, A. A thousand empirical adaptive landscapes and their navigability. *Nature Ecology and Evolution* **1** (2017).

22. Olson, C. A., Wu, N. C. & Sun, R. A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain. *Current Biology* **24**, 2643 (2014).
23. Hartman, E. C., Lobba, M. J., Favor, A. H., Robinson, S. A., Francis, M. B. & Tullman-Ercek, D. Experimental Evaluation of Coevolution in a Self-Assembling Particle. *Biochemistry* **58**, 1527 (2019).
24. Pitt, J. & Ferré-D'Amaré, A. Rapid Construction of Empirical RNA Fitness Landscapes. *Science* **330**, 376 (2010).
25. Wright, S. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. *Proceedings of the 6th International Congress of Genetics*, 356 (1932).
26. Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nature Reviews Molecular Cell Biology* **10** (2009).
27. Kinsler, G., Geiler-Samerotte, K. & Petrov, D. Fitness variation across subtle environmental perturbations reveals local modularity and global pleiotropy of adaptation. *eLife* **9**, e61271 (2020).
28. Dekel, E. & Alon, U. Optimality and evolutionary tuning of the expression level of a protein. *Nature* **436**, 588 (2005).
29. Lynch, M. & Hagner, K. Evolutionary meandering of intermolecular interactions along the drift barrier. *Proceedings of the National Academy of Sciences* **112** (2015).
30. Rokyta, D. R., Joyce, P., Caudle, S. B., Miller, C., Beisel, C. J. & Wichman, H. A. Epistasis between Beneficial Mutations and the Phenotype-to-Fitness Map for a ssDNA Virus. *PLoS Genetics* **7** (2011).
31. Duveau, F., Toubiana, W. & Wittkopp, P. J. Fitness Effects of Cis-Regulatory Variants in the *Saccharomyces cerevisiae* TDH3 Promoter. *Molecular Biology and Evolution* **34**, 2908 (2017).
32. Rest, J. S., Morales, C. M., Waldron, J. B., Opulente, D. A., Fisher, J., Moon, S., Bullaughey, K., Carey, L. B. & Dedousis, D. Nonlinear Fitness Consequences of Variation in Expression Level of a Eukaryotic Gene. *Molecular Biology and Evolution* **30**, 448 (2012).
33. Keren, L., Hausser, J., Lotan-Pompan, M., Slutskin, I., Alisar, H., Kaminski, S., Weinberger, A., Alon, U., Milo, R. & Segal, E. Massively Parallel Interrogation of the Effects of Gene Expression Levels on Fitness. *Cell* **166**, 1 (2016).
34. Labourel, F. & Rajon, E. Resource uptake and the evolution of moderately efficient enzymes. *Molecular Biology and Evolution* **38**, 3938 (2021).
35. Perfeito, L., Ghozzi, S., Berg, J., Schnetz, K. & Lässig, M. Nonlinear Fitness Landscape of a Molecular Pathway. *PLoS Genetics* **7** (2011).

36. Otwinowski, J., McCandlish, D. & Plotkin, J. Inferring the shape of global epistasis. *Proceedings of the National Academy of Sciences* **115**, E7550 (2018).
37. Domingo Espinós, J., Baeza-Centurion, P. & Lehner, B. The Causes and Consequences of Genetic Interactions (Epistasis). *Annual Review of Genomics and Human Genetics* **20**, 17.1 (2019).
38. Kryazhinskiy, S. Emergence and propagation of epistasis in metabolic networks. *eLife* **10**, e60200 (2021).
39. Crocker, J., Preger-Ben Noon, E. & Stern, D. The Soft Touch: Low-Affinity Transcription Factor Binding Sites in Development and Evolution. *Current Topics in Developmental Biology* **117**, 455 (2015).
40. Haberle, V. & Stark, A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nature Reviews Molecular Cell Biology* **19**, 621 (2018).
41. Shultzaberger, R., Malashock, D., Kirsch, J. & Eisen, M. The fitness landscapes of cis-acting binding sites in different promoter and environmental contexts. *PLoS Genetics* **6** (2010).
42. Sharon, E., Kalma, Y., Sharp, A., Raveh-Sadka, T., Levo, M., Zeevi, D., Keren, L., Yakhini, Z., Weinberger, A. & Segal, E. Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nature Biotechnology* **30**, 521 (2012).
43. Mustonen, V., Kinney J. and Callan, C. & Lässig, M. Energy-dependent fitness: a quantitative model for the evolution of yeast transcription factor binding sites. *Proceedings of the National Academy of the Sciences* **105**, 12376 (2008).
44. Haldane, A., Manhart, M. & Morozov, A. Biophysical fitness landscapes for transcription factor binding sites. *PLoS Computational Biology* **10** (2014).
45. Gerland, U. & Hwa, T. On the selection and evolution of regulatory DNA motifs. *Journal of Molecular Evolution* **55**, 386 (2002).
46. Berg, J., Willmann, S. & Lässig, M. Adaptive evolution of transcription factor binding sites. *BMC Evolutionary Biology* **4** (2004).
47. Tuğrul, M., Paixão, T., Barton, N. & Tkačik, G. Dynamics of Transcription Factor Binding Site Evolution. *PLoS Genetics* **11**, 273 (2015).
48. Weirauch, M., Yang, A., Albu, M., Cote, A., Montenegro-Montero, A. & Drewe, P. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431 (2014).
49. Crocker, J., Abe, N., Rinaldi, L., McGregor, A. P., Frankel, N., Wang, S., Alsawadi, A., Valenti, P., Plaza, S., Payre, F., Mann, R. S. & Stern, D. L. Low affinity binding site clusters confer HOX specificity and regulatory robustness. *Cell* **160**, 191 (2015).

50. Kribelbauer, J. F., Rastogi, C., Bussemaker, H. J. & Mann, R. S. Low-Affinity Binding Sites and the Transcription Factor Specificity Paradox in Eukaryotes. *Annual Review of Cell and Developmental Biology* **35**, 357 (2019).
51. Grönlund, A., Lötstedt, P. & Elf, J. Transcription factor binding kinetics constrain noise suppression via negative feedback. *Nature Communications* **4** (2013).
52. Farley, E., Olson, K., Zhang, W. & et al. Suboptimization of developmental enhancers. *Science* **350**, 325 (2015).
53. Farley, E., Olson, K., Zhang, W., Rokhsar, D. & Levine, M. Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. *Proceedings of the National Academy of Sciences* **113**, 6508 (2016).
54. Tanay, A. Extensive low-affinity transcriptional interactions in the yeast genome. *Genome Research*, 962 (2006).
55. Ramos, A. & Barolo, S. Low-affinity transcription factor binding sites shape morphogen responses and enhancer evolution. *Philosophical Transactions of the Royal Society B* **368** (2013).
56. Tsai, A., Muthusamy, A. K., Alves, M. R., Lavis, L. D., Singer, R. H., Stern, D. L. & Crocker, J. Nuclear microenvironments modulate transcription from low-affinity enhancers. *eLife* **6**, e28975 (2017).
57. Cary, G. A., Jarvela, A. M. C., Francolini, R. D. & Hinman, V. F. Genome-wide use of high- and low-affinity Tbrain transcription factor binding sites during echinoderm development. *Proceedings of the National Academy of Sciences* **114**, 5854 (2017).
58. Wang, J., Malecka, A., Trøen, G. & Delabie, J. Comprehensive genome-wide transcription factor analysis reveals that a combination of high affinity and low affinity DNA binding is needed for human gene regulation. *BMC genomics* **16** (2015).
59. Clarke, B. The Evolution of Genetic Diversity. *Proceedings of the Royal Society of London. Series B, Biological Sciences* **205**, 453 (1979).
60. Gavrillets, S. Fitness Landscapes and the Origin of Species. *Princeton Univ. Press* (2004).
61. Misevic, D., Kouyos, R. D. & Bonhoeffer, S. Predicting the Evolution of Sex on Complex Fitness Landscapes. *PLoS Computational Biology* **5** (2009).
62. De Visser, J. A. G. M. & Krug, J. Empirical fitness landscapes and the predictability of evolution. *Nature Reviews Genetics* **15**, 480 (2014).
63. Orr, H. A. A Minimum on the Mean Number of Steps Taken in Adaptive Walks. *Journal of Theoretical Biology* **220**, 241 (2003).
64. Nowak, S. & Krug, J. Analysis of adaptive walks on NK fitness landscapes with different interaction schemes. *Journal of Statistical Mechanics: Theory and Experiment* (2015).

65. Aita Takuyo and Husimib, Y. Fitness Spectrum Among Random Mutants on Mt. Fuji Type Fitness Landscape. *Journal of Theoretical Biology* **182**, 469 (1996).
66. Kingman, J. On the properties of bilinear models for the balance between genetic mutation and selection. *Mathematical Proceedings of the Cambridge Philosophical Society* **81**, 443 (1977).
67. Kauffman, S. A. & Weinberger, E. D. The NK model of rugged fitness landscapes and its application to the maturation of the immune response. *Journal of Theoretical Biology* **141**, 211 (1989).
68. Berg, O. G. & von Hippel, P. H. Selection of DNA Binding Sites by Regulatory Proteins. *Journal of Molecular Biology* **193**, 723 (1987).
69. Poelwijk, F., Kiviet, D., Weinreich, D. & Tans, S. Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* **445**, 383 (2007).
70. Fisher, R. A. *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford. (1930).
71. Tenaillon, O. The Utility of Fisher's Geometric Model in Evolutionary Genetics. *Annual Review of Ecology, Evolution, and Systematics* **45**, 179 (2014).
72. Dieckmann, U. & Doebeli, M. On the origin of species by sympatric speciation. *Nature* **400**, 354 (1999).
73. Gao, H., Granka, J. M. & Feldman, M. W. On the Classification of Epistatic Interactions. *Genetics* **184**, 827 (2010).
74. Hwang, S., Park, S.-C. & Krug, J. Genotypic Complexity of Fisher's Geometric Model. *Genetics* **206**, 1049 (2017).
75. Hordijk, W. & Stadler, P. Amplitude Spectra of Fitness Landscapes. *Advances in Complex Systems* **1**, 39 (1998).
76. Macken, C. A. & Perelson, A. S. Protein evolution on rugged landscapes. *Proceedings of the National Academy of the Sciences* **86**, 6191 (1989).
77. Badis-Breard, G., Berger, M., Philippakis, A., Talukder, S., Gehrke, A., Jaeger, S., Chan, E., Metzler, G., Vedenko, A., Chen, X., Kuznetsov, H., Wang, C.-F., Coburn, D., Newburger, D., Morris, Q., Hughes, T. & Bulyk, M. Diversity and Complexity in DNA Recognition by Transcription Factors. *Science* **324**, 1720 (2009).
78. Berger, M., Philippakis, A., Qureshi, A., He, F., Estep, P. & Bulyk, M. Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nature Biotechnology* **24**, 1429 (2006).
79. Payne, J. & Wagner, A. The robustness and evolvability of transcription factor binding sites. *Science* **343**, 875 (2014).

80. Zagorski, M., Burda, Z. & Waclaw, B. Beyond the Hypercube: Evolutionary Accessibility of Fitness Landscapes with Realistic Mutational Networks. *PLoS Computational Biology* **12** (2016).
81. Caswell, H. *Matrix population models* (Sinauer, Sunderland, Mass., 2001).
82. Wilke, C., Wang, J., Ofria, C., Lenski, R. & Adami, C. Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature* **412**, 331 (2001).
83. Aguilar-Rodriguez, J., Peel, L., Stella, M., Wagner, A. & Payne, J. The architecture of an empirical genotype-phenotype map. *Evolution* **72**, 1242 (2018).
84. Franke, J., Klozer, A., de Visser, J. & Krug, J. Evolutionary accessibility of mutational pathways. *PLoS Computational Biology* **7** (2011).
85. Chou, H.-H., Delaney, N., Draghi, J. & Marx, C. Mapping the Fitness Landscape of Gene Expression Uncovers the Cause of Antagonism and Sign Epistasis between Adaptive Mutations. *PLoS genetics* **10** (2014).
86. Kemble, H., Eisenhauer, C., Couce, A., Chapron, A., Magnan, M., Gautier, G., Nagard, H., Nghe, P. & Tenaillon, O. Flux, toxicity, and expression costs generate complex genetic interactions in a metabolic pathway. *Science Advances* **6** (2020).
87. Das, S. G., Direito, S. O., Waclaw, B., J Allen, R. & Krug, J. Predictable properties of fitness landscapes induced by adaptational tradeoffs. *eLife*, e55155 (2020).
88. Pinheiro, F., Warsi, O., Andersson, D. I. & Lässig, M. Metabolic fitness landscapes predict the evolution of antibiotic resistance. *Nature Ecology and Evolution* **5**, 677 (2021).
89. Blanquart, F., Achaz, G., Bataillon, T. & Tenaillon, O. Properties of selected mutations and genotypic landscapes under Fisher's geometric model. *Evolution* **68**, 3537 (2014).
90. Friedlander, T., Prizak, R., Barton, N. H. & Tkačik, G. Evolution of new regulatory functions on biophysically realistic fitness landscapes. *Nature Communications* **8** (2017).
91. Bershtein, S., Serohijos, A. W. & Shakhnovich, E. I. Bridging the physical scales in evolutionary biology: from protein sequence space to fitness of organisms and populations. *Current Opinion in Structural Biology* **42**, 31 (2017).
92. Fragata, I., Blanckaert, A., Louro, M. A. D., Liberles, D. A. & Bank, C. Evolution in the light of fitness landscape theory. *Trends in Ecology & Evolution* **34**, 69 (2019).
93. He, X., Duque, T. & Sinha, S. Evolutionary Origins of Transcription Factor Binding Site Clusters. *Molecular Biology and Evolution* **29**, 1059 (2011).
94. Dingle, K., Ghaddar, F., Šulc, P. & Louis, A. A. Phenotype Bias Determines How Natural RNA Structures Occupy the Morphospace of All Possible Shapes. *Molecular Biology and Evolution* **39**, msab280 (2021).

95. Xiong, K., Gerstein, M. & Masel, J. Differences in evolutionary accessibility determine which equally effective regulatory motif evolves to generate pulses. *Genetics* **219** (2021).
96. Schaper, S. & Louis, A. A. The Arrival of the Frequent: How Bias in Genotype-Phenotype Maps Can Steer Populations to Local Optima. *PLoS One* **9**, e86635 (2014).
97. Bintu, L., Buchler, N., Garcia, H., Gerland, U., Hwa, T., Kondev, J. & Phillips, R. Transcriptional regulation by the numbers: Models. *Current Opinion in Genetics and Development* **15**, 116 (2005).
98. Siggers, T. & Gordân, R. Protein–DNA binding: Complexities and multi-protein codes. *Nucleic Acids Research* **42**, 1 (2013).
99. Dror, I., Golan, T., Levy, C., Rohs, R. & Mandel-gutfreund, Y. A widespread role of the motif environment on transcription factor binding across diverse protein families. *Genome Research* **25**, 1268 (2015).
100. Xin, B. & Rohs, R. Relationship between histone modifications and transcription factor binding is protein family specific. *Genome Research* **28**, 321 (2018).
101. Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Syed, K. S., Das, P., Kivioja, T., Dave, K., Zhong, F., Nitta, K., Taipale, M., Popov, A., Ginno, P., Domcke, S., Yan, J., Schübeler, D., Vinson, C. & Taipale, J. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356** (2017).
102. Tsai, A., Galupa, R. & Crocker, J. Robust and efficient gene regulation through localized nuclear microenvironments. *Development* **147** (2020).
103. Duque, T., Samee, A. H., Kazemian, M., Pham, H., Brodsky, M. & Sinha, S. Simulations of Enhancer Evolution Provide Mechanistic Insights into Gene Regulation. *Molecular Biology and Evolution* **31**, 184 (2013).
104. Martin, G., Elena, S. & Lenormand, T. Distributions of epistasis in microbes fit predictions from a fitness landscape model. *Nature genetics* **39**, 555 (2007).
105. Roughgarden, J. Species packing and the competition function with illustrations from Coral Reef fish. *Theoretical population biology* **5**, 163 (1974).
106. Duveau, F., Hodgins-Davis, A., Metzger, B., Yang, B., Tryban, S., Walker, E., Lybrook, T. & Wittkopp, P. Fitness effects of altering gene expression noise in *Saccharomyces cerevisiae*. *eLife* **7**, e37272 (2018).
107. Vos, M. G. D., Dawid, A., Sunderlikova, V. & Tans, S. J. Breaking evolutionary constraint with a tradeoff ratchet. *Proceedings of the National Academy of Sciences* **112**, 14906 (2015).
108. Steinberg, B. & Ostermeier, M. Environmental changes bridge evolutionary valleys. *Science Advances* **2** (2016).

109. Hayden, E. J. & Wagner, A. Environmental change exposes beneficial epistatic interactions in a catalytic RNA. *Proceedings of the Royal Society B: Biological Sciences* **279**, 3418 (2012).
110. Li, C. & Zhang, J. Multi-environment fitness landscapes of a tRNA gene. *Nature Ecology and Evolution* **2**, 1025 (2018).
111. Gorter, F. A., Aarts, M. G., Zwaan, B. J. & Visser, J. A. G. D. Local fitness landscapes predict yeast evolutionary dynamics in directionally changing environments. *Genetics* **208**, 307 (2018).
112. Bank, C., Matuszewski, S., Hietpas, R. T. & Jensen, J. D. On the (un)predictability of a large intragenic fitness landscape. *Proceedings of the National Academy of Sciences* **113**, 14085 (2016).
113. Bajić, D., Vila, J., Blount, Z. & Sanchez, A. On the deformability of an empirical fitness landscape by microbial evolution. *Proceedings of the National Academy of Sciences* **115**, 201808485 (2018).
114. Zwart, M. P., Schenk, M. F., Hwang, S., Koopmanschap, B., Lange, N. d., van de Pol, L., Nga, T. T. T., Szendro, I. G., Krug, J. & de Visser, J. A. G. M. Unraveling the causes of adaptive benefits of synonymous mutations in TEM-1 β -lactamase. *Heredity* **121**, 406 (2018).
115. Barrera, L., Vedenko, A., Kurland, J., Rogers, J., Gisselbrecht, S., Rossin, E., Woodard, J., Mariani, L., Kock, K., Inukai, S., Siggers, T., Shokri, L., Gordân, R., Sahni, N., Cotsapas, C., Hao, T., Yi, S., Kellis, M., Daly, M. & Bulyk, M. Survey of variation in human transcription factors reveals prevalent DNA binding changes. *Science* **351**, 1450 (2016).
116. Khalid, F., Aguilar-Rodriguez, J., Wagner, A. & Payne, J. Genonets Server - A web server for the construction, analysis, and visualization of genotype networks. *Nucleic Acids Research* **44**, W70 (2016).
117. Baldi, P. & Rinott, Y. Asymptotic Normality of Some Graph-Related Statistics. *Journal of Applied Probability* **26**, 171 (1989).

2.5 SUPPLEMENTARY INFORMATION

2.5.1 Derivations

Derivation 1: Derivation of equation 2.2

The epistasis in the fitness landscape is given by equation 2.7. We assume σ to be large, such that $(w_i - w_{\text{opt}})^2/\sigma^2 \ll 1 \forall i \in \{0, 1\}^2$, an approximation that holds when selection is weak. Taylor expanding F_i and neglecting higher order terms, we find

$$F_i = \exp \left[- \left(\frac{w_i - w_{\text{opt}}}{\sigma} \right)^2 \right] \approx 1 - (w_i - w_{\text{opt}})^2/\sigma^2, \quad (2.15)$$

where $i \in \{0, 1\}^2$. Let $\Delta_i = w_i - w_{\text{opt}}$, then

$$\begin{aligned} \epsilon_f &\approx \frac{-1}{\sigma^2} (\Delta_{00}^2 + \Delta_{11}^2 - \Delta_{01}^2 - \Delta_{10}^2) \\ &= \frac{-1}{\sigma^2} [(\Delta_{00} + \Delta_{11})^2 - 2\Delta_{00}\Delta_{11} - (\Delta_{01} + \Delta_{10})^2 + 2\Delta_{10}\Delta_{01}] \\ &= \frac{-1}{\sigma^2} [(\Delta_{00} + \Delta_{11} + \Delta_{01} + \Delta_{10})(\Delta_{00} + \Delta_{11} - \Delta_{01} - \Delta_{10}) - 2(\Delta_{00}\Delta_{11} - \Delta_{10}\Delta_{01})] \\ &= \frac{-1}{\sigma^2} [(\sum_i \Delta_i) \epsilon_{\text{gp}} - 2\epsilon_{m,\Delta}] \end{aligned} \quad (2.16)$$

where $\epsilon_{m,\Delta} = \Delta_{00}\Delta_{11} - \Delta_{10}\Delta_{01} = \epsilon_{m,\Delta} = (w_{00} - w_{\text{opt}})(w_{11} - w_{\text{opt}}) - (w_{01} - w_{\text{opt}})(w_{10} - w_{\text{opt}}) \implies \epsilon_{m,\Delta} = \epsilon_m - w_{\text{opt}} \cdot \epsilon_{\text{gp}}$, and $\epsilon_m = w_{00}w_{11} - w_{01}w_{10}$ is the multiplicative epistasis. Substituting for $\epsilon_{m,\Delta}$, expanding the $\sum_i \Delta_i$ term and rearranging gives:

$$\begin{aligned} \epsilon_f &= \frac{1}{\sigma^2} [2\epsilon_m + (2w_{\text{opt}} - \sum_i w_i) \cdot \epsilon_{\text{gp}}] \\ \implies \epsilon_f \cdot \epsilon_{\text{gp}} &= \frac{1}{\sigma^2} [2\epsilon_m \cdot \epsilon_{\text{gp}} + (2w_{\text{opt}} - \sum_i w_i) \cdot \epsilon_{\text{gp}}^2]. \end{aligned} \quad (2.17)$$

Derivation 2: Derivation of equation 2.3

The absolute difference between the number of peaks in the genotype-phenotype landscape and the fitness landscape is

$$|p_f - p_{\text{gp}}| = \begin{cases} p_f - p_{\text{gp}}, & \text{if } p_f > p_{\text{gp}} \\ p_{\text{gp}} - p_f, & \text{if } p_f < p_{\text{gp}}. \end{cases} \quad (2.18)$$

We can rule out the case where $p_f = p_{\text{gp}}$, because the probability of change in the number of peaks tends towards 1 as L increases, so long as w_{opt} is sufficiently far from 1 (Fig 2.4D). This implies

$$\langle |p_f - p_{\text{gp}}| \rangle = \langle p_f - p_{\text{gp}} \rangle_{p_f > p_{\text{gp}}} P(p_f > p_{\text{gp}}) + \langle p_{\text{gp}} - p_f \rangle_{p_{\text{gp}} > p_f} P(p_{\text{gp}} > p_f), \quad (2.19)$$

where $\langle p_f - p_{\text{gp}} \rangle_{p_f > p_{\text{gp}}}$ is the conditional mean of $p_f - p_{\text{gp}}$, given that $p_f > p_{\text{gp}}$.

As shown below, the probability of increase in the number of peaks is equal to the probability of decrease in the number of peaks (Proof 3); i.e., $P(p_f > p_{gp}) = P(p_{gp} > p_f) = 0.5$. This implies

$$\langle |p_f - p_{gp}| \rangle = 0.5(\langle p_f - p_{gp} \rangle_{p_f > p_{gp}} + \langle p_{gp} - p_f \rangle_{p_{gp} > p_f}). \quad (2.20)$$

In genotype-phenotype landscapes generated by the House-of-Cards model, the phenotypic values of mutational neighbors are completely uncorrelated, so selection for a low or intermediate phenotypic value generates a fitness landscape in which the fitness values of mutational neighbors are also completely uncorrelated. Thus, the distribution of the number of peaks is the same for the genotype-phenotype landscape and the fitness landscape. Moreover, p_f is independent of p_{gp} because p_f depends upon the number of genotypes with phenotypes close to the optimal phenotypic value, and this number is independent of the number of genotypes with phenotypes close to 1. Therefore, p_{gp} and p_f are independent and identically distributed (i.i.d) random variables. Now as $L \rightarrow \infty$, the distribution of the number of peaks in the House-of-Cards model tends to a normal distribution [117] with mean $\frac{a^L}{(a-1) \cdot L + 1}$ and variance $\tau^2 = \frac{a^L \cdot ((a-1) \cdot L - (a-1))}{2((a-1) \cdot L + 1)^2}$ [76], where a is the number of alleles at each site and here $a = 2$. This implies the distribution of the random variable $p_f - p_{gp}$ is also normally distributed with mean zero and variance $2\tau^2$; i.e.,

$$p_f - p_{gp} := x \sim \mathcal{N}(0, 2\tau^2), \quad (2.21)$$

which is symmetric about zero, implying $\langle p_f - p_{gp} \rangle_{p_f > p_{gp}} = \langle p_{gp} - p_f \rangle_{p_{gp} > p_f}$. Thus,

$$\begin{aligned} \langle |p_f - p_{gp}| \rangle &= \langle p_f - p_{gp} \rangle_{p_f > p_{gp}} = \frac{1}{\sqrt{2\pi\tau^2}} \int_0^\infty x \cdot \exp\left[-\frac{x^2}{2\tau^2}\right] dx \\ &= \frac{\tau}{\sqrt{\pi}} = \frac{1}{\sqrt{\pi}} \sqrt{\frac{a^L \cdot ((a-1) \cdot L - (a-1))}{2((a-1) \cdot L + 1)^2}}, \end{aligned} \quad (2.22)$$

where $\tau'^2 = 2\tau^2$.

So, for $a = 2$, as in our biallelic landscapes,

$$\lim_{L \rightarrow \infty} \langle |p_f - p_{gp}| \rangle = \sqrt{\frac{2^L \cdot (L-1)}{2\pi(L+1)^2}}. \quad (2.23)$$

2.5.2 Proofs

Proof 1: Selection for w_{opt} cannot transform a simple sign epistasis motif into a reciprocal sign epistasis motif

In order to have a reciprocal sign epistasis motif:

- No two adjacent arrows can point in the same direction.
- No two parallel arrows can point in the same direction.

Because sign epistasis motifs cannot be cyclic, the two conditions above imply one another. The simple sign epistasis motif determines the complete rank ordering of the phenotypic values, unlike the other two motifs, which only give a partial ordering. Therefore, upon selecting for any w_{opt} , at least two arrows in the genotype-phenotype landscape remain the same as in the fitness landscape or reverse in direction together. This results in a pattern of arrows in the fitness landscape that always contradicts both the above mentioned conditions for reciprocal sign epistasis (see S2.1 Fig). ■

Proof 2: The phenotype-fitness map changes a reciprocal sign epistasis motif into a no sign epistasis motif or a simple sign epistasis motif with equal probability

The reciprocal sign epistasis motif does not determine the complete rank ordering of the four phenotypic values. In particular, it does not determine the ordering of the two peaks and the two valleys. Let the phenotypic value of the higher of the two peaks be w_{p_1} and that of the lower peak be w_{p_2} . Similarly, let the labels w_{v_1} and w_{v_2} refer to the higher and lower valleys, respectively. By design, these values lie on the real number line between 0 and 1. We define the "neighbourhood" of each of the phenotypic values as the region on the number line around the phenotypic value that is closer to it than to any other phenotypic value (S2.2 Fig). For transformation of the reciprocal sign epistasis motif into one of the other two motifs, w_{opt} must lie in the neighbourhood of either w_{p_2} or w_{v_1} and it must be second closest to the other (i.e., $F_{p_2} > F_{v_1} > \text{rest}$ or $F_{v_1} > F_{p_2} > \text{rest}$, where F_i represents the fitness corresponding to the phenotype w_i). If not, the reciprocal sign epistasis motif will be retained. So, for any w_{opt} , we only consider those configurations that satisfy this condition. Amongst these configurations, both phenotypic values are equally likely to be the closest to w_{opt} . Let's call the probability of either event happening $P_0 = P(F_{v_1} > F_{p_2} > \text{rest}) = P(F_{p_2} > F_{v_1} > \text{rest})$. We denote the transition probabilities amongst motifs as $P(x|rs)$, where x represents either the no sign epistasis motif, ns , or the simple sign epistasis motif, ss , and rs represents the reciprocal sign epistasis motif. Now, for the no sign epistasis motif to emerge, the fitness values of the peaks or valleys need to be "separated", while for the simple sign epistasis motif to emerge, they need to be "interspersed" (see S2.2), giving us the following:

$$P(ns|rs) = P(F_{v_1} > F_{p_2} > \text{rest}) \cdot P(F_{p_1} > F_{v_2}) + P(F_{p_2} > F_{v_1} > \text{rest}) \cdot P(F_{p_1} < F_{v_2}), \quad (2.24)$$

and

$$\begin{aligned} P(ss|rs) = & P(F_{v1} > F_{p2} > \text{rest}) \cdot P(F_{p1} < F_{v2}) + \\ & P(F_{p2} > F_{v1} > \text{rest}) \cdot P(F_{p1} > F_{v2}). \end{aligned} \quad (2.25)$$

Finally, using $P_0 = P(F_{v1} > F_{p2} > \text{rest}) = P(F_{p2} > F_{v1} > \text{rest})$ gives:

$$\begin{aligned} P(ns|rs) = & P_0 \cdot P(F_{p1} > F_{v2}) + P_0 \cdot P(F_{p1} < F_{v2}) \\ & \text{and} \\ P(ss|rs) = & P_0 \cdot P(F_{p1} < F_{v2}) + P_0 \cdot P(F_{p1} > F_{v2}) \\ \implies & P(ns|rs) = P(ss|rs). \blacksquare \end{aligned} \quad (2.26)$$

Proof 3: For House-of-Cards genotype-phenotype landscapes, selection for low or intermediate phenotypic values is equally likely to increase or decrease the number of peaks in the fitness landscape, relative to the genotype-phenotype landscape

For any number of loci L , the probability that selection for a low or intermediate phenotypic value causes an increase in the number of peaks in the fitness landscape, relative to the genotype-phenotype landscape, is

$$P_{\text{inc}} = \sum_{i>j} P(p_f = i | p_{\text{gp}} = j) \cdot P(p_{\text{gp}} = j), \quad (2.27)$$

whereas the probability of a decrease in the number of peaks is

$$P_{\text{dec}} = \sum_{i>j} P(p_f = j | p_{\text{gp}} = i) \cdot P(p_{\text{gp}} = i), \quad (2.28)$$

where $i, j \in \{1, 2, \dots, 2^{L-1}\}$, $P(p_f = j | p_{\text{gp}} = i)$ is the probability of having j peaks in the fitness landscape after selection for w_{opt} given the genotype-phenotype landscape has i peaks, and $P(p_{\text{gp}} = i)$ is the probability of i peaks in the genotype-phenotype landscape. The number of peaks in the fitness landscape, p_f , is independent of the number of peaks in the genotype-phenotype landscape, p_{gp} , because p_f depends upon the number of genotypes with phenotypes close to the optimal phenotypic value, and this number is independent of the number of genotypes with phenotypes close to 1. This implies:

$$P(p_f = i | p_{\text{gp}} = j) \cdot P(p_{\text{gp}} = j) = P(p_f = i) \cdot P(p_{\text{gp}} = j), \quad (2.29)$$

and

$$P(p_f = j | p_{\text{gp}} = i) \cdot P(p_{\text{gp}} = i) = P(p_f = j) \cdot P(p_{\text{gp}} = i). \quad (2.30)$$

Moreover, p_{gp} and p_f are identically distributed; i.e., $P(p_{\text{gp}} = i) = P(p_f = i)$

$$\implies P_{\text{inc}} = P_{\text{dec}} \blacksquare \quad (2.31)$$

2.5.3 Notes

Note 1: Number of maxima in the plot of $\langle |p_f - p_{gp}| \rangle$ versus w_{opt} for additive biallelic landscapes

Any biallelic, additive landscape can be relabelled such that the genotype with the lowest phenotypic value is represented by 000.....0 and the genotype with the highest phenotypic value by 111.....1. Moreover, the phenotypic values can be re-scaled such that genotype 000.....0 has phenotypic value zero and genotype 111.....1 has phenotypic value one. This yields a monotonically increasing genotype-phenotype landscape, wherein the phenotypic effects of every single mutant is larger than zero and they add up to 1; i.e.,

$$w_{100...0} + w_{010...0} + \dots + w_{000...1} = 1 \quad (2.32)$$

where $w_{100...0}$ is the phenotypic value of a single mutant with the mutation at the 1st site, and so on. Equation 2.32 implies that the vector of single mutant phenotypes belongs to a standard $L - 1$ -simplex³.

Two-locus case

A two-locus biallelic fitness landscape can have a maximum of two peaks. Because the additive genotype-phenotype landscape has only one peak, selection for a low or intermediate phenotypic value can increase the number of peaks in the fitness landscape, relative to the genotype-phenotype landscape, by at most one. Let the fitness of one of the single mutants be x , forcing the fitness of the other single-mutant to be $1 - x$. Due to the relabelling of the genotype-phenotype landscape, a two-peaked fitness landscape is only possible when the two single-mutants are peaks. This will occur when the two single-mutants have phenotypes that are closer to w_{opt} than the phenotypes of both genotypes 00 and 11. This leads to the following conditions on x :

$$1 - 2 \cdot w_{opt} < x < 2 \cdot w_{opt}, \forall w_{opt} < 0.5 \quad (2.33)$$

and

$$2 \cdot w_{opt} - 1 < x < 2 - 2 \cdot w_{opt}, \forall w_{opt} > 0.5. \quad (2.34)$$

From Eqs 2.33 and 2.34, we can infer the shape of the $L = 2$ curves in Fig 2.4A and 2.4B. Specifically, for $w_{opt} < 0.25$ and $w_{opt} > 0.75$, we get the absurd condition $0.5 < x < 0.5$, which implies there is no possible value of x that will lead to two peaks. Further, for $w_{opt} = 0.5$, $0 < x < 1$, which implies there will always be two peaks. For $0.25 \leq w_{opt} < 0.5$ or $0.5 < w_{opt} \leq 0.75$, the fitness landscape can have either one or two peaks.

Three-locus case

Here, the single-mutant phenotypic values belong to the 2-simplex formed by the points $(0,0,1)$, $(0,1,0)$ and $(1,0,0)$. Let (w_1, w_2, w_3) be an arbitrary point in the simplex,

³ An n dimensional simplex is the convex hull of $n + 1$ affinely independent points. E.g., a 0-simplex is a point, a 1-simplex is a line, a 2-simplex is a triangle, and a 3-simplex is a tetrahedron.

where w_i is the phenotypic value of the single mutant with a mutation at the i th site and $i \in \{1, 2, 3\}$. Although, a three-locus, biallelic landscape can have at most four peaks, the transformed additive genotype-phenotype landscapes can have at most three, due to the correlations between the phenotypic values. So for instance, if the three single mutants become peaks, then the triple mutant cannot be a peak because the double mutants will be assigned a higher fitness than the triple mutant. This is caused by correlations in the additive landscape, which makes the double-mutants closer in phenotypic value to the single-mutants than the triple mutant. Thus, it is easy to see that the only three peak configuration possible is one in which either all the single mutants become peaks or all the double mutants become peaks. For the former to be the case, each single mutant phenotypic value w_i must satisfy the following conditions:

$$(w_i - w_{\text{opt}})^2 < (0 - w_{\text{opt}})^2 \quad (2.35)$$

and

$$(w_i - w_{\text{opt}})^2 < (1 - w_j - w_{\text{opt}})^2 \quad \forall j \neq i. \quad (2.36)$$

While Eq. 2.35 necessitates the single mutant fitness values to be higher than the wild type fitness, Eq. 2.36 requires them to be higher than the neighbouring double mutant fitness values. On average, $w_i = 1/3 \forall i$, so on average, the simplex area enclosed by the above conditions will be maximised when $w_{\text{opt}} = 1/3$. A similar analysis can be done for the case when the three double mutants become peaks and in this case, the area is maximised when $w_{\text{opt}} = 2/3$. This explains the presence of the two maxima in Fig 2.4A at $w_{\text{opt}} = 1/3$ and $w_{\text{opt}} = 2/3$. Further, the dip at $w_{\text{opt}} = 0.5$ can be explained by the following argument: Assuming the single mutant phenotypic values are of the type $(x, y, 1 - x - y)$, the double mutant phenotypic value will be of the type $(x + y, 1 - x, 1 - y)$. At $w_{\text{opt}} = 0.5$, the wild type and triple mutant immediately become minima and leave an opportunity for the intermediate mutants to become peaks. However, if one of the single mutants (say with fitness x) is closest in fitness to 0.5, the other two cannot be peaks because the double mutant formed by combining these two mutations will have the same fitness i.e., $|x - 0.5| = |(1 - x) - 0.5|$. So at $w_{\text{opt}} = 0.5$, there are two peaks after selection for a low phenotypic value and therefore the net change is 1, which is smaller than what we see at $w_{\text{opt}} = 1/3, 2/3$.

Four and more locus cases

We could explicitly derive the shape of the curves for the $L = 2$ and $L = 3$ cases because it was easy to visualise the simplices. For $L \geq 4$, we can form an intuition about the expected result, by considering the average additive fitness landscape, which guides the behavior of the average change in the number of peaks. Because the only randomness in these landscapes arises from the effect of single mutations, which as described above, belong to the L -dimensional simplex, all we need to look at is the average L -dimensional vector of single mutation effects $\langle \vec{w}_1^L \rangle$ and the deviations around it. These random vectors belong to an L -dimensional Dirichlet distribution with all the concentration parameters

equal to one. It can be shown by symmetry that $\langle \bar{w}_1^L \rangle = (1/L, 1/L, \dots, 1/L)$. Therefore, on average, the fitness of a mutant with n mutations is n/L . As w_{opt} increases from zero, the number of peaks will gradually start to increase as w_{opt} gets closer the value of the fitness of single mutants. Then as w_{opt} approaches $1/L$, we encounter the first maxima because at this w_{opt} , all single mutants get to be peaks. Next, there is a slight decline due to competition between the single and double mutants as peaks (because these are mutational neighbours and therefore cannot both be peaks), followed by the emergence of the second maxima at $2/L$ when all the double mutants become peaks, and so on. Naturally, because the number of genotypes with n mutations increases as n reaches the value $L/2$, the height of the maxima increases accordingly. Thereafter, we observe a fall in $\langle |p_f - p_{\text{gp}}| \rangle$. Therefore, for a genotype of length L , there should be $L - 1$ maxima in the plot of $\langle |p_f - p_{\text{gp}}| \rangle$ vs. w_{opt} . However, the maxima are closely spaced and therefore merge into one another and cannot be distinguished from each other.

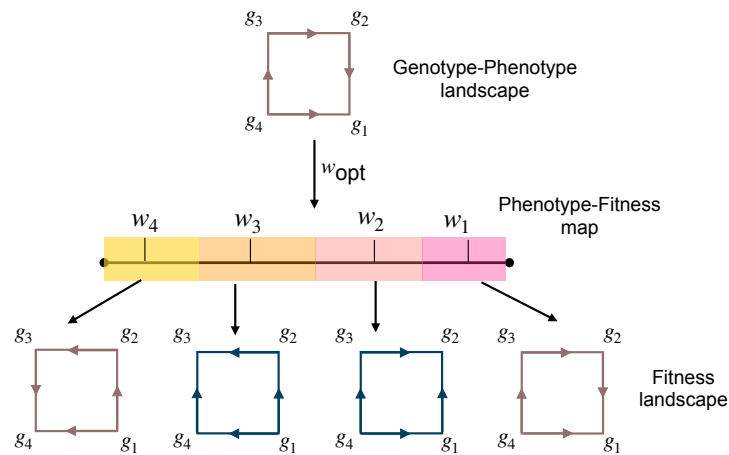
Note 2: Evolution on NK landscapes

S2.3A and S2.3C Fig shows the average length of a greedy adaptive walk $\langle l \rangle$ in NK landscapes with $L = 5$ and $L = 8$. For all $K < L - 1$, $\langle l \rangle$ is minimised when $w_{\text{opt}} = 0.5$, whereas for $K = L - 1$ (House-of-Cards landscapes), $\langle l \rangle$ is independent of w_{opt} . This can be understood by first considering additive genotype-phenotype landscapes ($K = 0$), wherein genotypes with a phenotype $w \approx 0.5$ are by definition, approximately located half-way between the wild-type and its antipodal sequence. Selecting for $w_{\text{opt}} = 0.5$ thus minimizes the average length of a greedy adaptive walk to these peak genotypes. As K increases, correlations among the phenotypes of mutationally-neighboring genotypes are gradually broken and the landscape becomes more rugged. Yet, genotypes with a phenotype $w \approx 0.5$ tend to remain near the center of the landscape. When K is increased to $L - 1$, the phenotypes of mutationally-neighboring genotypes become completely uncorrelated and monotonic increases in phenotype along mutational paths to peak genotypes become exceedingly rare, rendering $\langle l \rangle$ independent of w_{opt} . Thus the ‘‘U’’-shape of the curve of $\langle l \rangle$ versus w_{opt} flattens as K increases from 0 to $L - 1$. Interestingly, so long as $K > 0$, the average height of the peak reached by a greedy adaptive walk is also maximized at $w_{\text{opt}} = 0.5$, and these local peaks are $>97\%$ the height of the global peak for $\sigma = 1$ (S2.7 Fig), an observation we explain below.

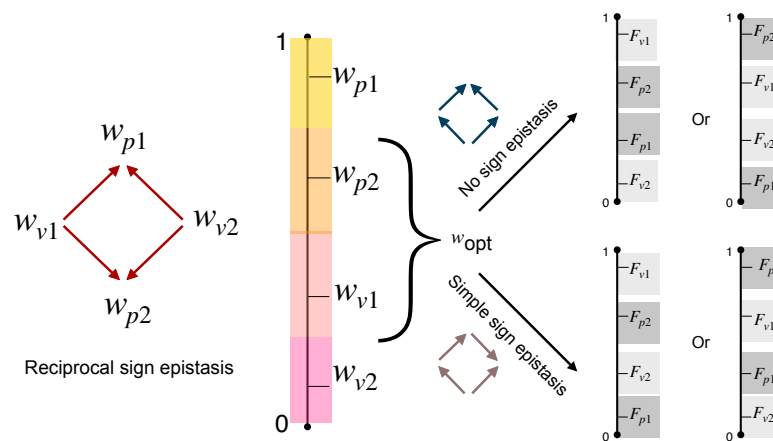
We next simulated deterministic mutation-selection dynamics. S2.3B and S2.3D Fig show the change in mean fitness at equilibrium when selecting for w_{opt} , relative to fitness at equilibrium when selecting for $w_{\text{opt}} = 1$, for $\sigma = 0.5$ and $\mu = 0.1$ (see S2.8 Fig for other values of σ and μ). This change is always positive, increases with L and K , and tends to be maximized at $w_{\text{opt}} = 0.5$ when $K > 0$. The latter occurs because the optimal phenotype $w_{\text{opt}} = 0.5$ is, on average, closer to all other phenotypic values in the landscape. For instance for $K = L - 1$, the sum $\sum_{i=1}^{2^L} (w_i - w_{\text{opt}})^2$ is minimized when $w_{\text{opt}} = \sum_{i=1}^{2^L} w_i / 2^L = \langle w \rangle = 0.5$. Thus, selecting for $w_{\text{opt}} = 0.5$ creates a fitness landscape

with higher average fitness than fitness landscapes under selection for any other w_{opt} . This also explains why the average height of local adaptive peaks reached is maximized at $w_{\text{opt}} = 0.5$ (S2.7 Fig). In sum, these results show that selection for an intermediate phenotypic value decreases the length of a greedy adaptive walk and increases mean fitness at equilibrium, despite increasing the overall ruggedness of the fitness landscape, relative to the genotype-phenotype landscape.

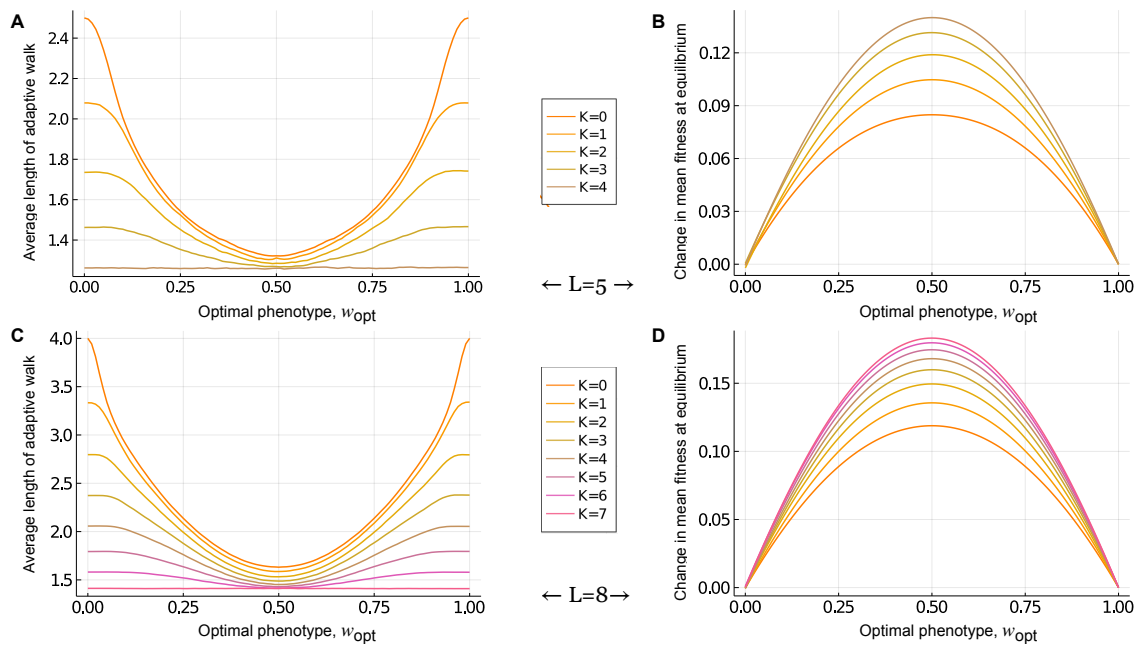
2.5.4 Supplementary Figures



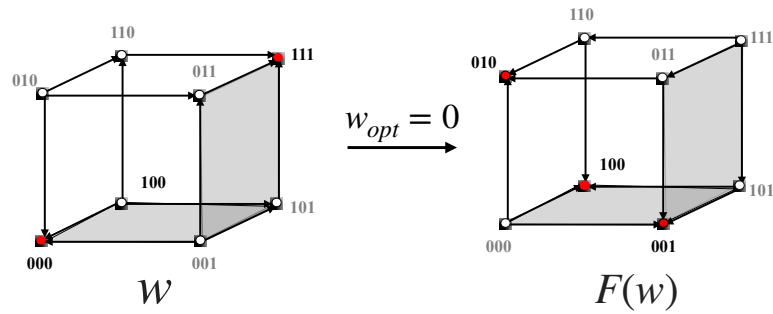
S 2.1: A simple sign epistasis motif (shown in brown) cannot be changed into a reciprocal sign epistasis motif, for any w_{opt} . The no sign epistasis motif is shown in blue. The remaining colours represent the neighbourhood of the phenotypic values (w_i) corresponding to each genotype (g_i), where $i \in \{1, 2, 3, 4\}$ and the genotypes are labelled in descending order, such that the genotype with the highest phenotypic value is labelled 1, and so on. The four bottom arrows point to the transformed motifs when w_{opt} belongs to the neighbourhood from which the arrow emanates.



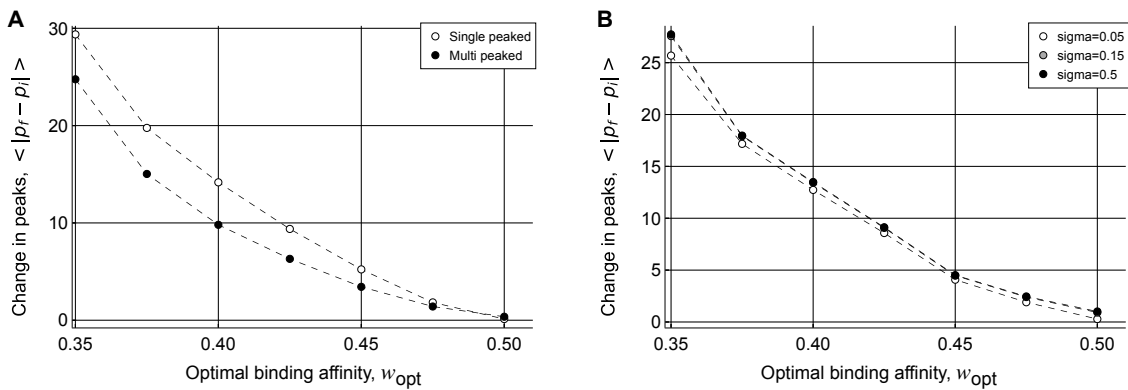
S 2.2: Selection for w_{opt} transforms the reciprocal sign epistasis motif (shown in red) into the no sign epistasis motif (shown in blue) and the simple sign epistasis motif (shown in brown) with equal probability. The neighbourhood of each phenotypic value is shown in a different colour. For the no sign epistasis motif to emerge (top), the fitness values need to be "separated", while for the simple sign epistasis motif to emerge (bottom), the fitness values need to be "interspersed".



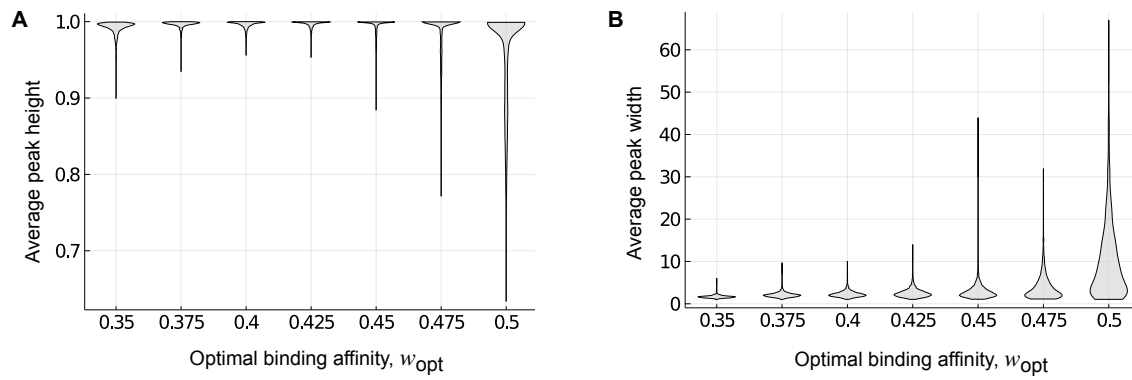
S 2.3: Evolutionary consequences of landscape ruggedness in NK landscapes. (A,C) The average length of a greedy adaptive walk and (B,D) the change in mean fitness at equilibrium, relative to fitness at equilibrium when selecting for $w_{opt} = 1$, are shown for (A,B) $L = 5$ and (C,D) $L = 8$. In (B,D), $\mu = 0.1$ and $\sigma = 0.5$.



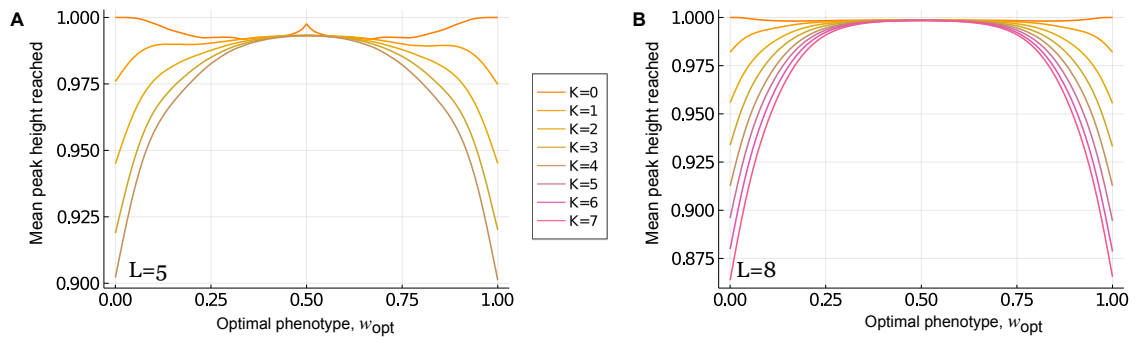
S 2.4: Selection for $w_{opt} = 0$ does not change the type of epistasis motif for any “square” in the fitness landscape, relative to the genotype-phenotype landscape, yet it can change the number and location of peaks. To understand how, consider that any two adjacent faces of the hypercube (e.g., gray faces above) are sufficient to determine whether the genotypes on their common edge are peaks or not. After selecting for $w_{opt} = 0$, the type of epistasis motif does not change in the adjacent faces, yet the number and location of the peaks on their common edge does change. Peak genotypes are shown in red. Arrows point from lower to higher phenotypic or fitness values.



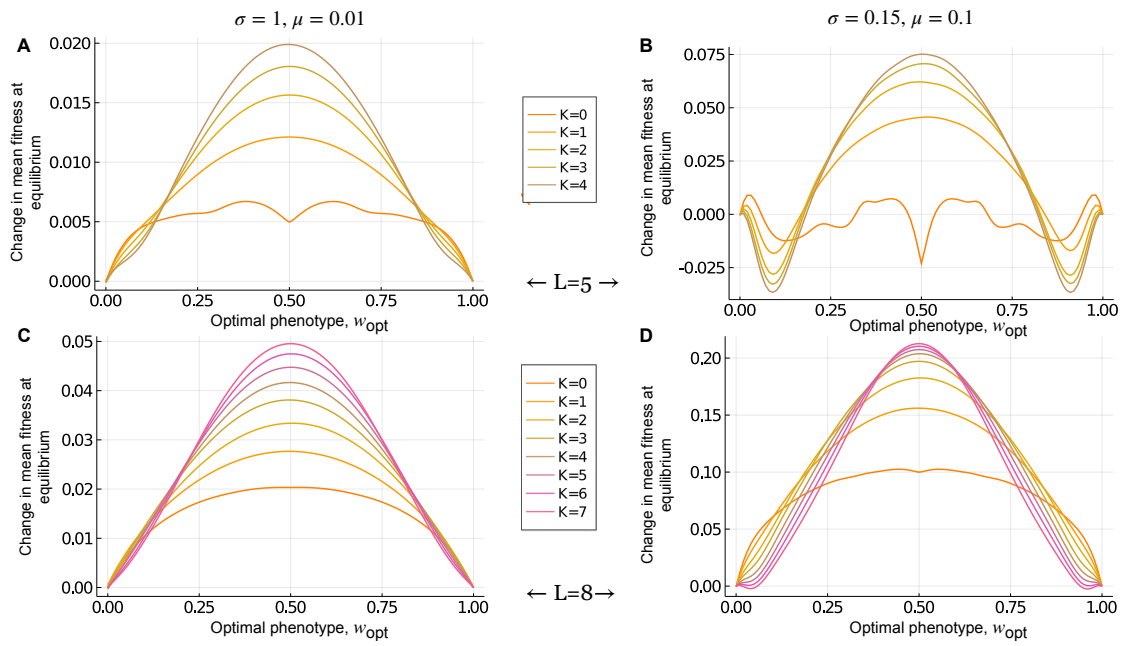
S 2.5: Mean absolute change in the number of peaks in the fitness landscape relative to the genotype-phenotype landscape, shown in relation to w_{opt} , for (A) single-peaked vs. multi-peaked genotype-phenotype landscapes and (B) three values of σ .



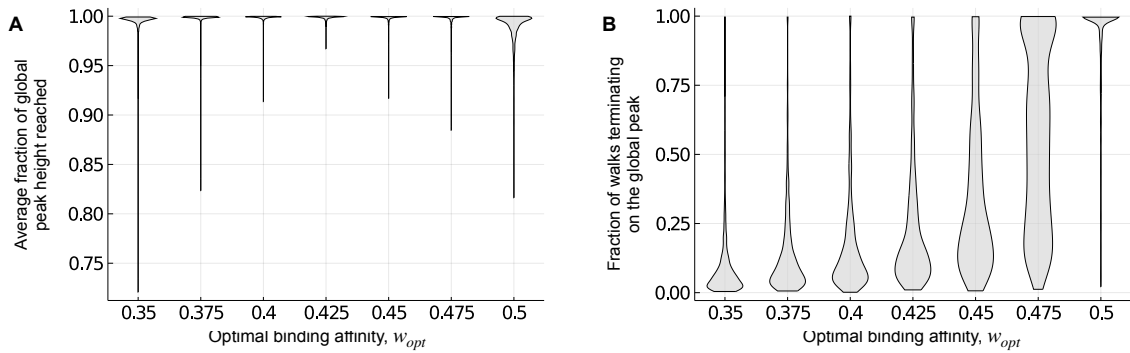
S 2.6: Average peak (A) height and (B) width for 1,137 empirical landscapes, shown in relation to the optimal binding affinity w_{opt} . Violin plots show the distribution across the landscapes for each w_{opt} . Data include both local and global peaks.



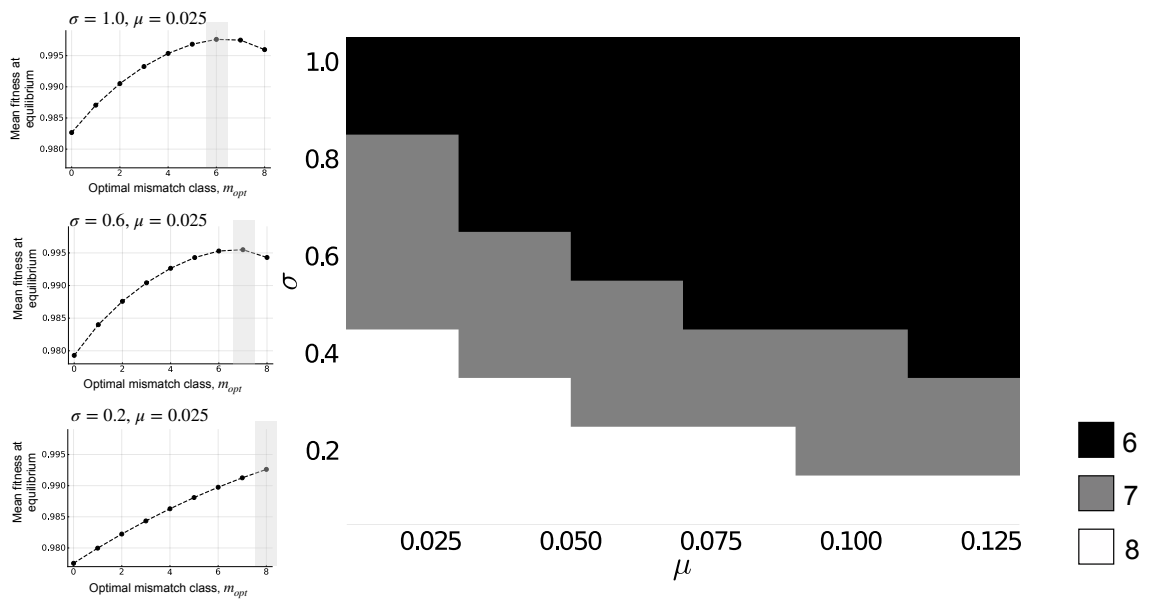
S 2.7: The average height of peaks reached by greedy adaptive walks in NK landscapes with $\sigma = 0.5$, shown in relation to w_{opt} , for (A) $L = 5$ and (B) $L = 8$, with $K = 0 \dots L - 1$.



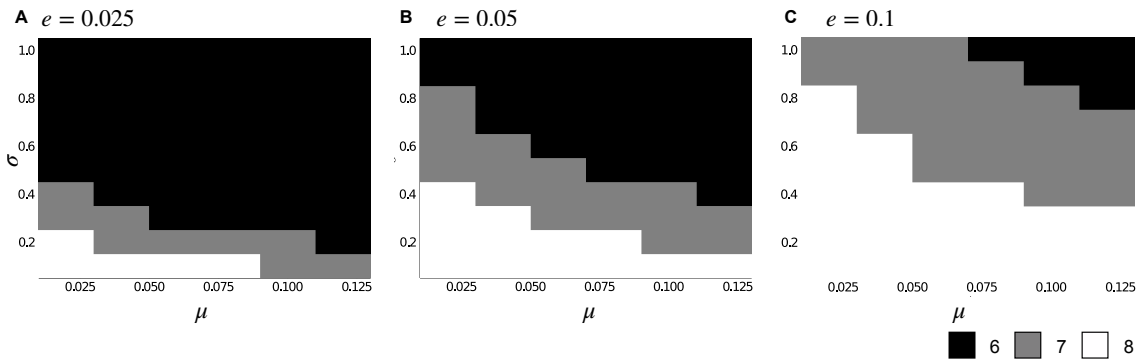
S 2.8: Change in mean fitness at equilibrium for NK landscapes with (A,B) $L=5$ and (C,D) $L=8$, shown in relation to w_{opt} , for different values of σ and μ .



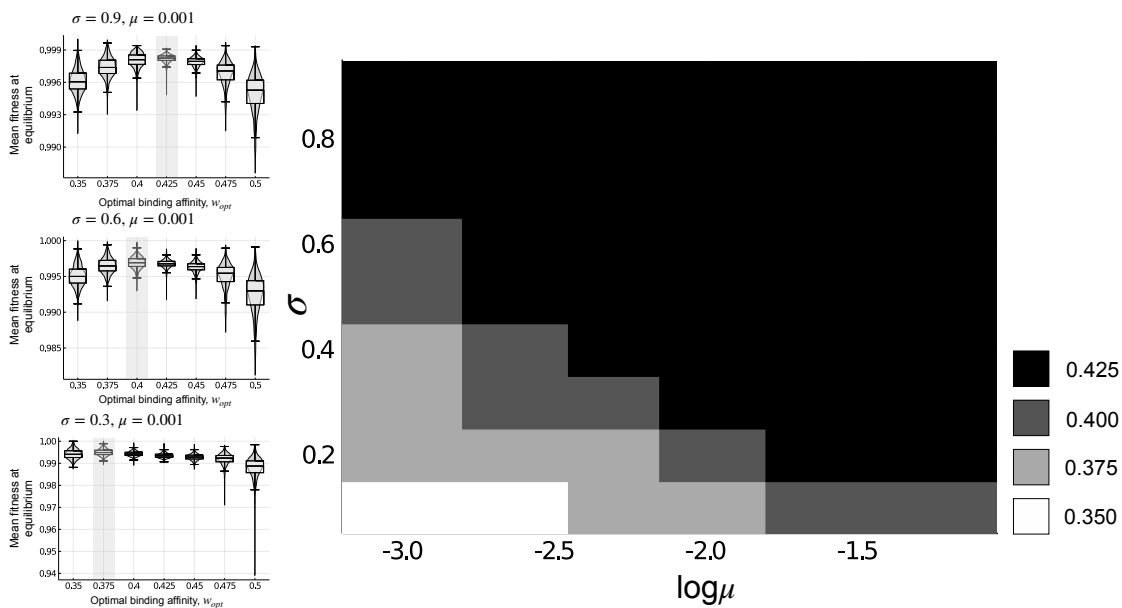
S 2.9: (A) Local peaks on which the adaptive walks terminate tend to be nearly as tall as the global peaks in the 1,137 empirical landscapes. Violin plots show the distribution of the fractional height of local peaks reached by greedy adaptive walks, relative to the height of the global peak, for each optimal binding affinity w_{opt} . (B) Violin plots show the distribution of the fraction of walks terminating on the global peak, for each optimal binding affinity w_{opt} .



S 2.10: The mismatch class m_{opt} that maximizes mean fitness at equilibrium is shown in relation to the strength of selection σ and the mutation rate μ . The three smaller panels show mean fitness at equilibrium in relation to m_{opt} for three combinations of σ and μ . The value of m_{opt} that maximizes mean fitness is indicated with a gray rectangle.



S 2.11: The mismatch class m_{opt} that maximizes mean fitness at equilibrium is shown in relation to the strength of selection σ and the mutation rate μ for three different values of the mismatch penalty e : (A) $e = 0.025$, (B) $e = 0.05$ and (C) $e = 0.1$.



S 2.12: The binding affinity w_{opt} that maximizes mean fitness at equilibrium is shown in relation to the strength of selection σ and the logarithm of the mutation rate ($\log \mu$) for the 1,137 empirical landscapes. The three smaller panels show the distributions of mean fitness at equilibrium as violin plots, in relation to w_{opt} , for three combinations of σ and μ . Box-and-whisker plots show the 25 – 75% quartiles. The value of w_{opt} that maximizes mean fitness is indicated. These results are robust to perturbations of the fitness values (Methods), with the exception of two parameter combinations – $\sigma = 0.1, \mu = 0.001$ and $\sigma = 0.1, \mu = 0.0025$. For these parameter combinations, the binding affinity w_{opt} that maximizes mean fitness at equilibrium changes from 0.35 to 0.385 (on average) and 0.375 respectively.

*If you do not change direction, you may end up
where you are heading.*

— Lao Tzu

Published as: Malvika Srivastava *et al* 2023 *J. Phys. A: Math. Theor.* **56** 455601

<https://iopscience.iop.org/article/10.1088/1751-8121/ad0200>

Authors' contributions: M.S., H.R. and J.L.P. designed research; M.S. and H.R. performed research; M.S., H.R. and J.L.P. analyzed data; and M.S., H.R. and J.L.P. wrote the paper.

ABSTRACT

One of the most fundamental characteristics of a fitness landscape is its dimensionality, which is defined by genotype length and alphabet cardinality - the number of alleles per locus. Prior work has shown that increasing landscape dimensionality can promote adaptation by forming new “uphill” mutational paths to the global fitness peak, but can also frustrate adaptation by increasing landscape ruggedness. How these two topographical changes interact to influence adaptation is an open question. Here, we address this question in the context of alphabet cardinality, using theoretical fitness landscapes with tuneable fitness correlations, as well as three empirical fitness landscapes for proteins. We find that the primary effect of increasing alphabet cardinality is the introduction of a new global fitness peak. Controlling for this effect, we find that increasing alphabet cardinality promotes adaptation on uncorrelated fitness landscapes, but frustrates adaptation on correlated fitness landscapes. The primary explanation is that the increased ruggedness that accompanies alphabet expansion is characterized by an increase in mean peak height on uncorrelated fitness landscapes, but a decrease in mean peak height in correlated fitness landscapes. Moreover, in two of the empirical fitness landscapes we observe no effect of increasing alphabet cardinality on adaptation, despite an increase in the number of peaks and a decrease in mean peak height, calling into question the utility of these common measures of landscape ruggedness as indicators of evolutionary outcomes.

3.1 INTRODUCTION

The metaphor of the fitness landscape has framed evolutionary thought for nearly a century [1]. In it, genotypes represent coordinates in a multi-dimensional genotype space, and fitness represents the “elevation” of each coordinate in this space. This defines the surface upon which populations evolve, with natural selection favoring “uphill” mutational steps, driving populations toward peaks of high fitness. This metaphor has inspired a diversity of theoretical models (e.g., the House-of-Cards, NK and Rough Mt. Fuji landscapes) [2–4], has been used to predict and direct adaptive evolution [5–8], and more recently has become the focus of many experimental studies, in systems ranging from macromolecules to metabolisms [9–17].

How an evolving population navigates a fitness landscape depends in part on population genetic conditions [18–20]. In the strong selection, weak mutation regime, any beneficial mutation that appears and escapes stochastic loss will go to fixation prior to the arrival of a subsequent mutation, such that the population remains essentially monomorphic at all times [21]. Adaptive evolution can then be thought of as a random adaptive walk, where populations traverse mutational paths in which fitness increases at each mutational step, ending on a local or global adaptive peak. Such paths are called accessible paths [14].

Many theoretical studies have focused on adaptive walks in fitness landscapes, revealing the important influence of topographical and topological properties, such as landscape ruggedness and dimensionality, on the dynamics and outcomes of evolutionary processes [9, 22–27]. For example, as landscape ruggedness increases, the formation of maladaptive valleys renders many mutational trajectories to high fitness peaks inaccessible. In contrast, as landscape dimensionality increases, so-called “extradimensional bypasses” emerge, which circumvent maladaptive valleys by forming indirect, accessible paths to high fitness peaks [28]. How landscape ruggedness and dimensionality interact to open or close mutational trajectories to adaptation remains poorly understood. One of the reasons is that past research has tended to focus on bi-allelic landscapes (e.g., the presence or absence of adaptive mutations at each locus), controlling landscape dimensionality by changing genotype length L [23, 24, 28, 29]. However, empirical adaptive landscapes are not bi-allelic, because their genotypes are DNA or amino acid sequences. The number of alleles per locus is therefore an additional, and important, facet of landscape dimensionality [9, 26].

The set of all possible alleles at any of a genotype’s loci makes up an alphabet \mathcal{A} , with cardinality $|\mathcal{A}|$ denoting the number of its constituent elements (Figure 3.1). For instance, for DNA sequences, $\mathcal{A} = \{A, C, G, T\}$ and $|\mathcal{A}| = 4$; for amino acid sequences, \mathcal{A} is the set of 20 standard amino acids and $|\mathcal{A}| = 20$. Allowable mutational transitions between alleles are defined by an “allele graph” [27], such that for DNA sequences the allele graph is complete, because any nucleotide can mutate into any other, and for amino acid

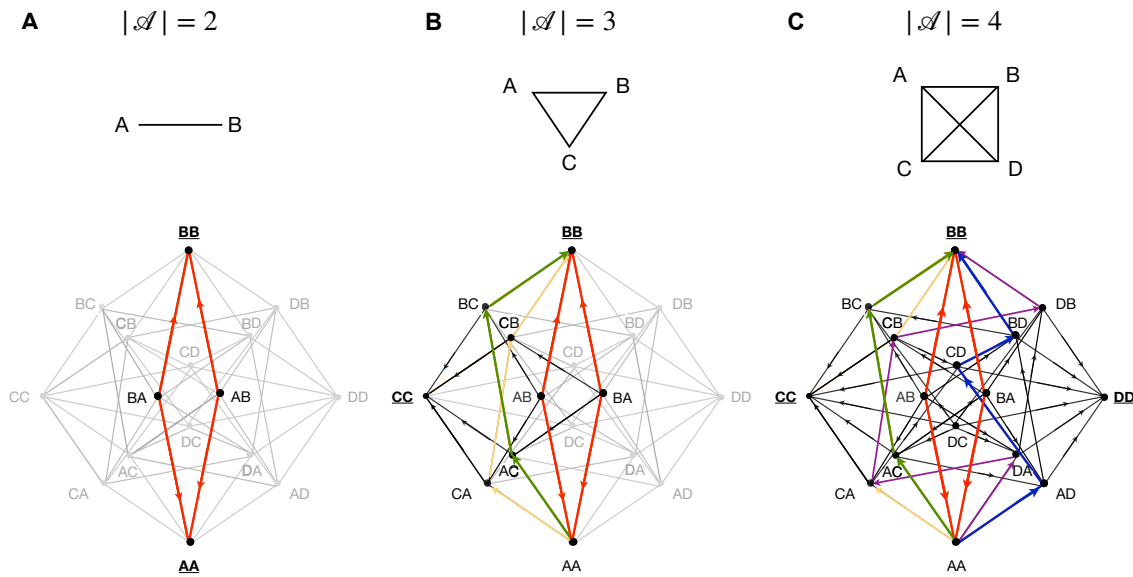


FIGURE 3.1: ALPHABET CARDINALITY, EXTRA-DIMENSIONAL BYPASSES, AND LANDSCAPE TOPOGRAPHY. A two-locus genotype space with (A) $\mathcal{A} = \{A, B\}$, $|\mathcal{A}| = 2$; (B) $\mathcal{A} = \{A, B, C\}$, $|\mathcal{A}| = 3$; and (C) $\mathcal{A} = \{A, B, C, D\}$, $|\mathcal{A}| = 4$. In the middle row, the motifs show the underlying complete allele graphs for different \mathcal{A} . The bottom row shows a fitness landscape for each $|\mathcal{A}|$, in which nodes represent genotypes and edges connect genotypes that differ at a single locus, with arrows pointing from low- to high-fitness genotypes. Direct paths from the source genotype AA to the target global peak genotype BB are shown in red and are both inaccessible. Accessible indirect paths are shown in the remaining colours. As alphabet cardinality increases, the number of accessible paths to the global peak increases, as extra-dimensional bypasses open up, and the number of peaks (shown with underlined genotypes) increases from 2 to 3. Note that the location and the height of peaks can change as well, because the new peak genotypes may have different fitness values.

sequences the allele graph is defined by the standard genetic code, under which missense mutations can change an amino acid into six alternative amino acids, on average [30].

Prior work on alphabet cardinality has revealed the formation of extra-dimensional bypasses that circumvent maladaptive fitness valleys [9], thus increasing the number of accessible paths to local and global fitness peaks [9, 26], as well as the sizes of their basins of attraction [9], and the probability that at least one accessible path exists from an initial genotype to the global peak [26, 27]. However, because these studies did not systematically fine-tune alphabet cardinality on landscapes with different levels of fitness correlations, several open questions remain, specifically pertaining to the interplay between cardinality, ruggedness, and the outcomes of adaptive walks. Whereas increasing cardinality can increase the prevalence of extra-dimensional bypasses to the global peak, it can also change other aspects of landscape topography, such as the height,

location, and number of local peaks [31, 32]. For instance, for the block model of NK landscapes, where each genotype is divided into B non-interacting blocks of size $K + 1$, the expected number of peaks is $\frac{|\mathcal{A}|^L}{[(|\mathcal{A}|-1) \cdot \frac{L}{B} + 1]^B}$, which grows approximately polynomially with increasing alphabet cardinality (for $B \ll L$) and approximately exponentially with decreasing fitness correlations (i.e., B) [32]. Such an increase in landscape ruggedness may offset any evolutionary benefit of extra-dimensional bypasses to the global peak if the landscape’s local peaks are of low or intermediate fitness. Conversely, if these local peaks are of high fitness, their emergence may prove no obstacle to adaptation, even if they do titrate evolving populations from the global peak. Relatedly, increasing cardinality not only increases the number of extra-dimensional bypasses to the global peak, but also to local peaks [9, 33], so whether evolving populations converge on the global or local peaks will depend on how these bypasses influence the relative sizes of the basins of attraction of these peaks. Finally, because extra-dimensional bypasses are by definition longer than direct paths, they are less likely to be utilized by evolving populations [34], which may instead follow direct accessible paths to emerging local peaks. Thus, calling a fitness landscape accessible based on the existence of at least one accessible path to the global peak [24, 26, 27] may belie a landscape’s true accessibility. For these reasons, it remains unclear how and to what extent increasing alphabet cardinality facilitates adaptive evolution.

Here, we study the effect of increasing alphabet cardinality on fitness landscapes of different topographies. First, we consider a spectrum of theoretical landscapes with tuneable levels of fitness correlations, ranging from completely correlated Mt. Fuji landscapes to completely uncorrelated House-of-Cards landscapes. Systematically varying alphabet cardinality and controlling for the emergence of the global adaptive peak, we study the impact of alphabet cardinality on landscape topography, specifically the number and heights of fitness peaks, as well as the number of extra-dimensional bypasses to these fitness peaks. Further, we measure the effect of increasing $|\mathcal{A}|$ on the mean fitness reached by random adaptive walks – a quantity we call the *final mean fitness*. We show that landscapes with different levels of correlation show characteristically different behavior upon increasing cardinality – whereas correlated fitness landscapes show a declining final mean fitness with increasing cardinality, uncorrelated fitness landscapes show an increasing final mean fitness with increasing cardinality.

Motivated by this observation, we then analyze the effect of increasing cardinality on three combinatorially complete empirical fitness landscapes with intermediate levels of fitness correlation. Using evolutionary simulations and controlling for the emergence of the global adaptive peak, we find that increasing cardinality does not lead to higher final mean fitness, even though it increases the number of accessible paths to the global peak. Whereas the results for the empirical fitness landscapes are qualitatively in agreement with the results for theoretical fitness landscapes with intermediate levels of fitness correlation, the theoretical fitness landscapes do not fully capture the behavior of two out

of the three empirical landscapes, thereby highlighting the limits of theoretical studies and the need for more experimental studies on fitness landscapes.

3.2 RESULTS

3.2.1 *Theoretical fitness landscapes*

To understand the effect of alphabet cardinality on landscapes of different topographies, we first analyzed Rough Mt. Fuji (RMF) landscapes [35, 36]. We generated the RMF landscapes by adding a normally distributed random variable with mean zero and variance r^2 to an additive Mt. Fuji landscape (see Methods), where the roughness parameter r controls the relative contribution of the random component and serves as a measure of the level of fitness correlation in the landscape. Thus, a landscape with $r = 0$ corresponds to a completely correlated additive landscape, whereas a landscape with $r = 1$ corresponds to an uncorrelated House-of-Cards-like landscape.

For proteins, the most obvious benefit of increasing cardinality is the expansion of the chemical toolkit of the amino acid alphabet, which may endow proteins with enhanced or entirely new functions [37, 38]. In the context of an adaptive landscape, this is equivalent to introducing a new fitness peak, which may be the global peak. Indeed, in RMF landscapes, increasing alphabet cardinality causes a monotonic increase in the global peak height (Supplementary Figure 3.1, top panel).

To study the influence of alphabet cardinality on landscape topography beyond the effect of introducing a new global peak, we hold the global peak constant throughout every alphabet expansion. Specifically, we confine our analyses to a subset of the 20! possible expansions of the alphabet in which the global peak genotype is always present. For example, consider a fitness landscape for a protein with $L = 4$ variable sites, in which the amino acid sequence WWLA is the fitness peak. In this case, we truncate the full landscape (i.e., $|\mathcal{A}| = 20$) to a landscape with $\mathcal{A} = \{W, L, A\}$ and analyze its topography. We then incrementally expand \mathcal{A} by randomly adding an allele from the remaining set and analyze the resulting landscape, continuing this process until all 20 alleles are included in \mathcal{A} . To understand the general consequences of increasing $|\mathcal{A}|$ in RMF landscapes, we studied an ensemble of 100 randomly generated RMF landscapes and simulated 200 expansions of \mathcal{A} for each of those landscapes. To enable comparison with the empirical protein landscapes in the following section, we present results for RMF landscapes with $L = 4$ in the main text and $L = 3$ in the supplementary text (Supplementary Figure 3.2).

Topographical analysis

We first studied the topographical properties of each class of landscapes, namely the number of peaks, their mean height, and the number of accessible paths that terminate at the global peak, out of all accessible paths (see Methods). Except for the purely additive case with $r = 0$, the number of local peaks increases monotonically with increasing $|\mathcal{A}|$ (Figure 3.2, first column). Note that the number of peaks also depends upon the size of the landscape and therefore, it is important to ascertain whether the increase in the number of peaks is due to the landscapes truly becoming more rugged or due to their increasing size alone. To do so, we also computed a landscape-size agnostic measure of ruggedness, the roughness-to-slope ratio, which measures the deviations of a fitness landscape from a linear model [39] (Supplementary Figure 3.3). The roughness-to-slope ratio also increases with increasing cardinality, confirming that the increase in the number of peaks is not solely due to the increasing landscape size. Further, regardless of the cause of the increase in the number of peaks, more peaks imply more sub-optimal endpoints for the adaptive walks, which frustrates adaptation. Thus, the number of peaks is indicative of the outcome of random adaptive walks, so we focus on this measure of ruggedness throughout the rest of the manuscript.

While the number of peaks increases monotonically regardless of the roughness parameter r , we observe that the trend in the mean fitness of these peaks depends on the level of correlation in these landscapes. For small values of r , i.e., strongly correlated landscapes, we observe monotonically declining mean peak height (Figure 3.2A, second column). In contrast, for large values of r , i.e., strongly uncorrelated landscapes, the mean peak height tends to increase with $|\mathcal{A}|$ (Figure 3.2D, second column). In line with prior literature [9, 26], the number of accessible paths to the global peak increases with increasing alphabet cardinality (Figure 3.2A-D, third column). However, the fraction of accessible paths that terminate on the global peak, out of all accessible paths, decreases with $|\mathcal{A}|$, regardless of the value of the roughness parameter r (except for the corner case of $r = 0$, in which case the fraction is independent of cardinality) (Supplementary Figure 3.4). The reason is that more local peaks emerge as alphabet cardinality increases, such that accessible paths are more likely to terminate on local peaks than on the global peak. In other words, although the absolute number of accessible paths to the global peak increases, the probability that a randomly chosen accessible path reaches the global peak decreases as more local peaks emerge in the landscape.

Why do the trends in mean peak height depend on the level of fitness correlation? For completely correlated landscapes (i.e. $r = 0$), the mutational effects of alleles at the various loci are completely independent of each other, so increasing $|\mathcal{A}|$ has no effect on the mean peak height because there is only one peak in the landscape (i.e., the global peak). By slightly decreasing the level of fitness correlation (e.g., when $r = 0.05$ or $r = 0.125$), we begin to introduce interactions between loci, yet the truncated landscape with only the

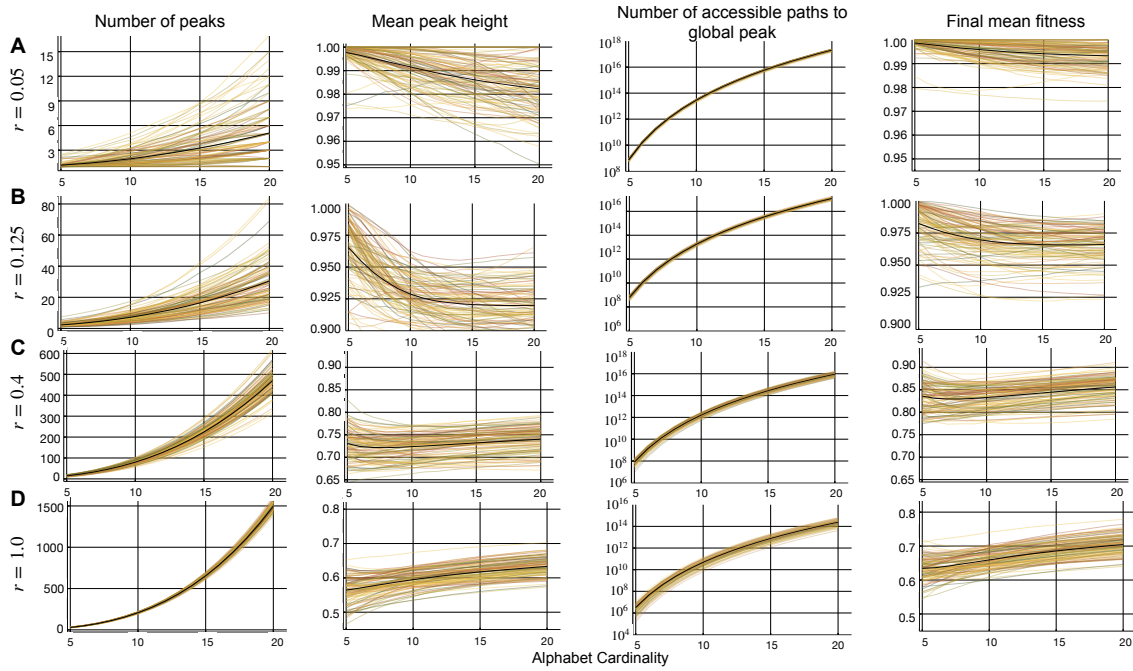


FIGURE 3.2: Topographical properties and evolutionary outcomes in relation to alphabet cardinality, for landscapes with different levels of fitness correlations. Each row corresponds to a different value of the roughness parameter r and each column to a topographical property or evolutionary outcome, which are indicated by column headers (these also define the corresponding y-axes). The number of accessible paths to the global peak only include paths starting from genotypes antipodal to the global peak. The coloured lines show results for 100 RMF landscapes and the black lines show the average over these landscapes. Data pertain to $|\mathcal{A}| \geq 5$.

global peak alleles remains very smooth, comprising only one or two peaks. Expanding the alphabet increases the number of sub-optimal peaks (Figures 3.2A and 3.2B, first column) and thus, the mean peak height decreases monotonically (Figures 3.2A and 3.2B, second column). In contrast, for landscapes with larger r , the truncated landscapes are already quite rugged and thus already have a low mean peak height. However, for intermediate levels of fitness correlation (e.g., when $r = 0.4$), we see a nearly constant mean peak height (Figure 3.2C, second column), while for uncorrelated landscapes (e.g., when $r = 1.0$) we find that it increases with $|\mathcal{A}|$ (Figure 3.2D, second column). This is because in uncorrelated landscapes, all the new mutational neighbours introduced due to the increased cardinality also compete to be peaks. Therefore, to be a peak, a genotype must be larger than all its $(|\mathcal{A}| - 1) \cdot L$ neighbors. It is well known that the expected value of the maximum of $(|\mathcal{A}| - 1) \cdot L + 1$ normally distributed random variables increases with $|\mathcal{A}|$, thus leading to higher mean peak height. The same constraint does not hold for landscapes with intermediate levels of fitness correlation, because a potential peak genotype does not have to compete with all its mutational neighbours; the existing correlations already ensure that some of them have lower fitness than the potential peak

genotype. In other words, for uncorrelated landscapes, the average threshold fitness value to be a peak increases with increasing $|\mathcal{A}|$, while it does not for landscapes with intermediate levels of fitness correlation.

Random adaptive walks

In the previous section, we observed that the number of peaks and the number of accessible paths terminating at the global peak both increase with $|\mathcal{A}|$, for all r (Figure 3.2, first and third column), but that the trends in mean peak height depend on r . Whereas an increase in landscape ruggedness may frustrate adaptation by trapping evolving populations on local peaks, especially if these peaks are of low fitness, increasing the accessibility of the global peak may promote adaptation by opening new mutational paths to high fitness. Which of these two topographical changes matters more for adaptation? To find out, we simulated random adaptive walks in the RMF landscapes and recorded the final mean fitness for each value of $|\mathcal{A}|$. We initialized the walks on all genotypes in the landscape, i.e., we simulated $20^4 = 160,000$ walks. In each step of a walk, the current genotype was replaced by a randomly chosen neighbor of higher fitness. This process was repeated until the walk reached a peak (Methods). We then averaged over the fitness values reached by each of the walks to calculate the final mean fitness. We chose to simulate random adaptive walks, because they are more likely to employ indirect paths than greedy adaptive walks [9].

Figure 3.2 (fourth column) shows that the effect of $|\mathcal{A}|$ on the final mean fitness depends on the roughness parameter r . For the purely additive landscapes ($r = 0$), the final mean fitness is independent of $|\mathcal{A}|$. For landscapes with small r , the final mean fitness decreases as a function of $|\mathcal{A}|$; however, as r increases, the trend flips and increasing $|\mathcal{A}|$ leads to increasing final mean fitness (Figure 3.2, fourth column). Interestingly, for each r , the final mean fitness curve resembles the mean peak height curve, except that it is shifted upwards (Figure 3.2, second and fourth column). This suggests that the emergence of local peaks may have a stronger influence on adaptation than the formation of extra-dimensional bypasses to the global peak. But why are the curves for final mean fitness higher than those for mean peak height?

Basins of attraction

We reasoned that the curves for final mean fitness may be higher than those for mean peak height because the random walks preferentially converge on higher peaks. We therefore computed the fraction of random walks terminating on each peak in the landscape and used it as a proxy for the basin of attraction of that peak.

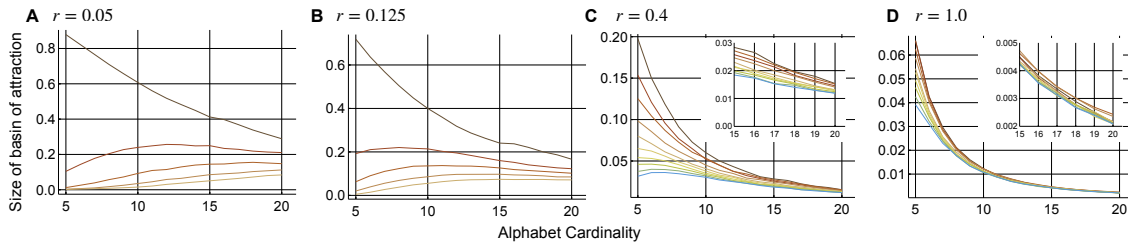


FIGURE 3.3: Basins of attraction of the (A,B) top five and (C,D) top ten peaks in RMF landscapes with varying values of the roughness parameter A) $r = 0.05$, B) $r = 0.125$, C) $r = 0.4$, and D) $r = 1.0$. Lines are rank ordered from top to bottom based on peak height. The results are averaged over 30 RMF landscapes. The insets in (C,D) pertain to $|\mathcal{A}| \geq 15$, showing how even at high cardinalities, the basins of attraction segregate based on fitness rank. Note the change in the y-axis limits across panels.

These data are shown in Fig. 3.3, which reveals four key trends. First, higher peaks tend to have larger basins of attraction, for all r and $|\mathcal{A}|$. Indeed, the fitness rank of a peak and the rank of the size of its basin of attraction are always strongly correlated (even at $|\mathcal{A}| = 20$, the Pearson correlation coefficient ranged between 0.7 and 1 for all r , with lower values for landscapes with larger r and perfect correlation for landscapes with $r = 0.05$ and $r = 0.125$). This is in line with prior work showing that genotypes are more likely to have an accessible path to a higher peak than to a lower peak [27]. Second, for any $|\mathcal{A}|$, the difference in the basin of attraction is larger for landscapes with small r than for landscapes with large r . Third, this difference decreases with $|\mathcal{A}|$ for all r . Fourth, for landscapes with small r , the basins of attraction of emerging local peaks grows, rather than shrink, as $|\mathcal{A}|$ increases. Together, these four trends explain why the curves for final mean fitness are higher than those for mean peak height. Had the basins of attraction been of equal size, these curves would be nearly identical. Indeed, the difference in the basins of attraction is smallest for landscapes with large r , where the final mean fitness curve most closely resembles the mean peak height curve, although it is still shifted slightly upwards due to the segregation of the basins of attraction based on peak height. Moreover, while for all r , the basin of attraction of the global peak decreases with $|\mathcal{A}|$ [31], the basins of attraction of the local peaks increase for landscapes with small r (Figure 3.3A and B), but decrease for landscapes with large r (Figure 3.3C and D). This implies that as $|\mathcal{A}|$ increases, local peaks become more important in landscapes with small r by attracting a higher proportion of the random walks, thus causing the final mean fitness to decline.

To further illustrate this point, we computed the entropy of the basins of attraction of all peaks in each landscape. This measure takes on its minimum value of zero when the landscape is single-peaked, and its maximum value of 1 when the landscape is multi-peaked and all peaks have basins of attraction of equal size. Supplementary Figure 3.5 shows that entropy increases with $|\mathcal{A}|$ for all r . However, the rate of increase is highest

for landscapes with intermediate r (e.g., $r = 0.125$), meaning that local peaks rapidly become significant attractors. Because these peaks tend to be of lower fitness than the global peak in landscapes with intermediate r , their emergence brings down the final mean fitness. In contrast, for landscapes with larger r , entropy is high and only increases marginally with $|\mathcal{A}|$. Thus the basins of attractions are more uniform in size. However, entropy is always less than one, due to the persistent correlation of a peak's height and its basin of attraction. Taken together, these trends explain why the curves for final mean fitness track those for mean peak height, yet remain shifted upwards.

3.2.2 Empirical fitness landscapes

In the previous section, we saw that the effect of increasing cardinality on adaptation depends upon the level of fitness correlation in the landscape, which is controlled by the roughness parameter r . Empirical landscapes usually exhibit intermediate levels of fitness correlation [9, 14, 40–42], and we focus on three such landscapes in the following section.

Data sets

We used three empirical fitness landscapes built from combinatorially-complete data sets, i.e., data that characterize protein phenotype (e.g., binding affinity) for all possible 20^L combinations of amino acids at a small number L of protein sites [9, 40]. Following the protein evolution literature [9, 43], we assume a direct mapping from phenotype to fitness, although this assumption is often violated [44].

The first data set pertains to streptococcal protein G domain B1. Wu et al. [9] created a library of all possible combinations of amino acids at 4 sites in GB1 ($20^4 = 160,000$ variants), specifically V39, D40, G41, and V54, which interact epistatically and influence the protein's binding affinity to immunoglobulin [15]. They assayed protein phenotype — binding affinity — for all sequence variants by measuring the relative frequency of each variant before and after selection for binding immunoglobulin, defining protein phenotype as a log-enrichment ratio relative to the wild type (Methods). In the following, we refer to this data set as GB1.

The second and third data sets pertain to the bacterial ParD-ParE toxin-antitoxin system. Lite et al. [40] created a library of all possible combinations of amino acids at 3 sites in the antitoxin ParD ($20^3 = 8,000$ variants), specifically D61, K64, and E80, which influence the protein's specificity for ParE toxins. They assayed protein phenotype — the ability to antagonize toxin — for two toxins, ParE2 and ParE3, by measuring the relative frequency of each variant before and after selection for antagonizing toxin, defining protein phenotype as a log-enrichment ratio relative to the wild type (see Methods). In the following, we refer to these data sets as ParE2 and ParE3, respectively.

For each of the three data sets we construct an adaptive landscape as follows: Each protein sequence of a given length ($L = 4$ for the GB1 data set, and $L = 3$ for the ParE2 and ParE3 data sets) is represented as a vertex in a mutational network. Two vertices are connected by an edge if their corresponding sequences differ in exactly one position, i.e., we assume that each amino acid can mutate into any other [9, 45]. Each vertex is labeled with the phenotype (binding affinity or growth rate) of its sequence, defining its “elevation” in the landscape. In order to reduce experimental noise, as well as impute missing values (6.6% of the GB1 variants failed to assay), we smoothed the data using empirical variance component regression [46] (Methods). The resulting landscapes comprised 17, 5, and 5 peaks for the GB1, ParE2, and ParE3 data sets, respectively. In the GB1 data set, the protein phenotype ranged from -8.28 (for genotype WPGI) to 2.52 (for genotype WWLA), while in the ParD datasets, the protein phenotype ranged from -7.78 (for genotype PIP) to 0.19 (for genotype ELK) in the presence of ParE2, and from -7.74 (for genotype APP) to 0.13 (for genotype DWE) in the presence of ParE3.

Topographical analysis and adaptive walks on empirical fitness landscapes

The results for the topographical analysis of the three empirical landscapes are shown in Figure 3.4 (columns 1-3). As expected based on the analysis of the RMF landscapes, the average number of peaks increases monotonically with increasing $|\mathcal{A}|$ for all three landscapes. Whereas the truncated GB1 landscape has approximately 2.5 peaks on average, the truncated ParE2 and ParE3 landscapes were almost always single-peaked. This reflects in the mean peak height, which shows little variation for the GB1 landscape, whereas it decreases for the ParE2 and ParE3 landscapes due to the emergence of local peaks. Finally, in accordance with the analyses on RMF landscapes, whereas the number of accessible paths to the global peak increases with $|\mathcal{A}|$ (Figure 3.4, third column), the fraction of accessible paths that terminate at the global peak, out of all accessible paths, decreases (see Supplementary Figure 3.6).

Also in line with our analysis of RMF landscapes with intermediate r , we observed that alphabet cardinality has little impact on the final mean fitness for all three empirical fitness landscapes. This occurs despite a decrease in mean peak height in the ParE2 and ParE3 fitness landscapes (Figure 3.4B-C, second column), a trend we did not observe in the RMF landscapes. To explain this discrepancy, we next look at the basins of attraction of each of the peaks.

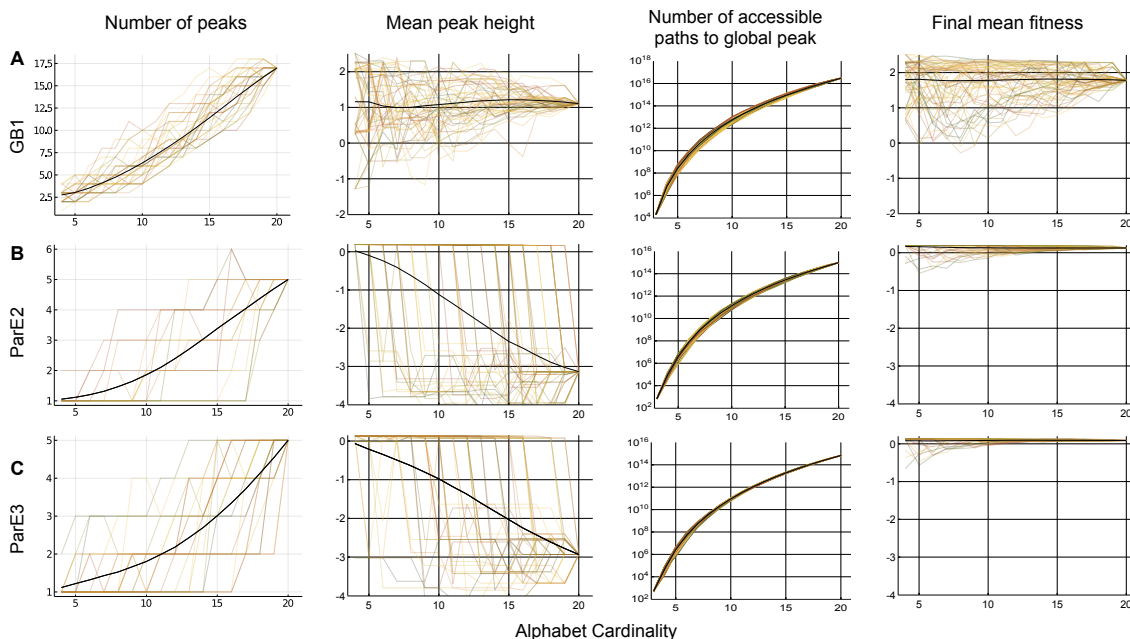


FIGURE 3.4: Topographical properties and evolutionary outcomes for three empirical fitness landscapes. Each row corresponds to a different landscape and each column to a topographical property or evolutionary outcome, which are indicated by column headers (these also define the corresponding y-axes). The coloured lines show results for approximately 10% of the alphabet expansions, which were randomly chosen, and the black lines show the results averaged over all 5,000 expansions.

Basins of attraction

In Figure 3.5, we show the basins of attraction of the top five peaks as a function of the alphabet cardinality, averaged over 3000 alphabet expansions. The results for the GB1 landscape (Fig. 3.5A) are similar to those obtained for RMF landscapes with intermediate level of correlation (Figure 3.3B) – fitter peaks have larger basins of attraction and the difference in the sizes decreases with increasing cardinality. The results for the ParE2 and ParE3 landscapes also resemble the $L = 3$ RMF landscape results (Supplementary Figure 3.7A), but with one crucial difference: The basins of attraction of the lowest three peaks in the ParE2 and ParE3 landscapes do not increase with increasing cardinality (Fig. 3.5B and C), a trend that we do not see in the RMF landscapes. This explains why despite showing very different trends in the mean peak height (Fig. 3.4, second column), all three empirical landscapes show similar trends in the variation of the final mean fitness (Fig. 3.4, fourth column). While in the GB1 landscape, this is due to the final mean fitness tracking the mean peak height, in the ParE2 and ParE3 landscapes, it is due to the sub-optimal peaks with very low fitness not affecting the final mean fitness, even though

they significantly decrease the mean peak height, and the second highest peak having a reasonably high fitness.

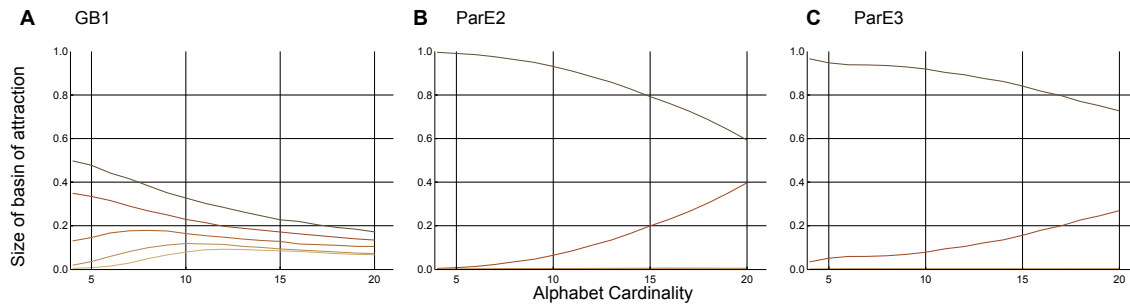


FIGURE 3.5: The basins of attraction of the top five peaks as a function of alphabet cardinality for empirical fitness landscapes of protein A) GB₁, B) ParE₂, and C) ParE₃. Lines are rank ordered from top to bottom based on peak height. For each alphabet cardinality, we generated 3000 random alphabet expansions.

Indirect paths promote adaptation when the number and location of peaks is preserved

Our analyses of the RMF and empirical landscapes have revealed that increasing cardinality increases both the number of local peaks and the number of accessible paths to the global peak. However, our evolutionary simulations were more strongly influenced by the emergence of local peaks, such that final mean fitness tracked mean peak height. This led us to wonder if there exists a scenario in which the emergence of indirect paths to the global peak clearly promotes adaptation.

To find out, we considered an artificial scenario in which we controlled not only for the emergence of the global peak, but also for the emergence of local peaks. To do so, we began our alphabet expansions from a truncated alphabet that contained all of the amino acids in all of the peaks in the full landscape (i.e., with $|\mathcal{A}| = 20$). This was not possible for GB₁, because the peaks in the full landscape comprise all 20 amino acids. However, it was possible for both the ParE₂ and ParE₃ landscapes, because their peaks comprise only nine amino acids.

Fig. 3.6 shows the outcomes of our evolutionary simulations on these landscapes. We observe a monotonic and non-saturating increase in final mean fitness with increasing alphabet cardinality. Thus, there exists a scenario in which the emergence of indirect paths to the global peak promotes adaptation, but this scenario is highly artificial. Under more realistic conditions, the emergence of local peaks that accompanies alphabet expansion has a stronger influence on adaptation than does the emergence of indirect paths to the global peak.

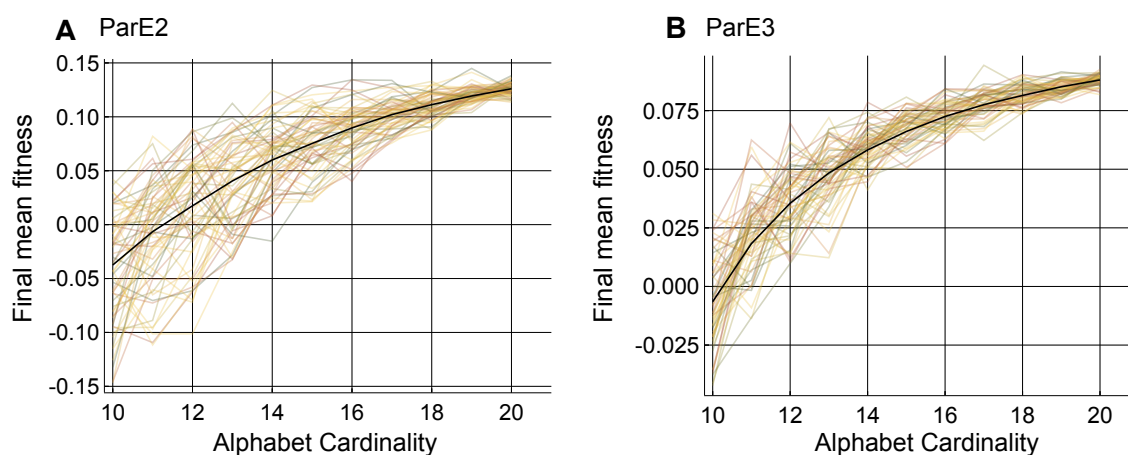


FIGURE 3.6: The final mean fitness as a function of the alphabet cardinality for empirical fitness landscapes of protein A) ParE2 and B) ParE3 when the number and location of peaks were preserved. The black lines show the variation in the mean fitness averaged over 1,000 possible expansions of the alphabet and the coloured lines show the variation for 50 randomly chosen expansions. We confined our analysis to only those alphabet expansions where the number of peaks was preserved. This was true for all ParE3 alphabet expansions and for approximately 68% of the ParE2 alphabet expansions.

3.3 DISCUSSION

Is dimensionality a blessing or a curse? For fitness landscapes, the answer may be a bit of both. On the one hand, increasing the length of a sequence or the cardinality of an alphabet can lead to the formation of new, high-fitness peaks or to the emergence of new mutational paths to high-fitness peaks. On the other hand, such increases in dimensionality can do the opposite; they can lead to the formation of new low-fitness peaks or to the emergence of new mutational paths to low-fitness peaks. Moreover, any increase in dimensionality necessarily increases the size of sequence space, which may slow adaptation if mutational paths to high-fitness peaks become very long.

For proteins, the most obvious advantage of increasing the size of the amino acid alphabet is expanding the chemical lexicon of the proteome. Different amino acids have different chemistries and adding to this chemical toolkit expands the space of functions available to proteins. Indeed, our analyses revealed that increasing cardinality caused a monotonic increase in the average height of the global fitness peak, revealing that the primary influence of alphabet cardinality on protein function is the introduction of protein variants with novel chemistries and improved functionalities. To probe the effects of alphabet cardinality on landscape topography beyond this primary effect, we studied an artificial scenario in which the global peak did not change during alphabet expansion. This has the advantage that it allows us to directly study secondary effects of alphabet

expansion on landscape topography, but it has the disadvantage that it puts distance between our analyses and biological reality. An additional limitation of our study is that we did not take into consideration the genetic code, which influences both landscape topography [47] and the likelihood of various alphabet expansions [48–51]. Moreover, we restricted our analyses to static fitness landscapes, although real fitness landscapes are often dynamic, for example due to fluctuating environmental conditions [44, 52] or due to frequency-dependent selection [53]. Finally, we confined our evolutionary simulations to the strong selection weak mutation (SSWM) regime and didn't allow for downhill steps and crossing of fitness valleys. How relaxing this constraint affects our results is a direction for future work. With these caveats in mind, our results have four key implications for fitness landscape research.

The first implication concerns landscape accessibility, which is usually defined based on the existence of a single accessible path between a source and a target genotype [26, 27, 29]. Accessibility, defined as such, has been shown to increase with increasing alphabet cardinality in both theoretical and empirical fitness landscapes [9, 26, 27], due to the emergence of so-called extra-dimensional bypasses [28]. However, because these bypasses are necessarily indirect and therefore require more mutations than direct paths, their probability of realisation can be very low. Moreover, there are typically many accessible paths emanating from any source genotype, and these often lead to non-target genotypes, such as those atop local fitness peaks. These observations indicate that landscape accessibility is more nuanced than existing metrics let on, and call for more holistic metrics that take into account a greater diversity of source-target pairs and the realisation probabilities of individual accessible paths.

The second implication concerns the interaction between epistasis and alphabet cardinality, and its influence on adaptation. Past work has suggested that increasing alphabet cardinality may promote adaptation by increasing the number of accessible paths to the global adaptive peak [9, 26], but may frustrate adaptation by increasing the number of local peaks in the landscape [54]. How these two topographical changes come together to influence adaptation remained unclear. Our analyses of the RMF landscapes suggest that the latter topographical change has a stronger influence on adaptation, because the outcomes of our evolutionary simulations more closely tracked the average height of fitness peaks than they did the number of accessible paths to the global fitness peak. Importantly, how the average height of fitness peaks changed with alphabet cardinality depended on the level of epistasis amongst loci. When epistasis was low, the average height of fitness peaks decreased upon alphabet expansion; when epistasis was high, the average height of fitness peaks increased upon alphabet expansion. These results suggest that increased alphabet cardinality can either promote or frustrate adaptation, depending primarily upon its influence on the average heights of fitness peaks, a topographical property that is modulated by the extent to which loci epistatically interact.

Relatedly, the number and heights of fitness peaks are often used to characterize the ruggedness of a fitness landscape, and to make predictions about how adaptation may play out on a particular landscape [6, 24, 55, 56]. Indeed, in the RMF landscapes we found that the average heights of fitness peaks was a reasonable predictor of the outcomes of our evolutionary simulations. In contrast, in the ParD landscapes, we found that the average heights of fitness peaks decreased substantially upon alphabet expansion, but that this had little impact on the outcomes of our evolutionary simulations. The reason is that the three lowest local peaks that emerged during alphabet expansion had negligible basins of attraction, such that our evolutionary simulations preferentially converged on the global peak and the second highest peak. Therefore, the third implication of our study is that the number and heights of fitness peaks can be an insufficient proxy for landscape ruggedness, particularly in the context of predicting the outcomes of evolutionary processes. Alternative proxies that take into account the basins of attraction of fitness peaks and their entropy are likely to be more useful.

Finally, our results may have practical implications. Recent advances in biotechnology have enabled the design and construction of living organisms with expanded genetic codes, i.e., genetic codes that include a 21st, non-standard amino acid [38, 57–59]. These expanded codes are commonly utilized in biophysical studies to elucidate the structure, function, and cellular localization of proteins (reviewed in [57]), and they are a promising tool in evolving and designing proteins with novel properties [59–62], including those with therapeutic applications [63–65]. To date, nearly 250 non-standard amino acids have been explored for these purposes [66]. By increasing the chemical diversity of the amino acid alphabet, these non-standard amino acids can introduce new high-fitness peaks in a protein’s fitness landscape [67–71], such as those representing dramatic increases in thermostability [72, 73]. However, an important lesson from our study is that alphabet expansion can induce multiple topographical changes to a protein’s fitness landscape, and some of these make finding the highest-fitness peaks more difficult.

3.4 METHODS

RMF landscapes

To generate the RMF landscapes, we first generated an additive (Mt. Fuji) landscape of cardinality 20 by sampling the fitness effect of each allele at each locus from a uniform distribution between 0 and 1. The additive fitness value of any genotype σ was then just the sum of the contributions of the alleles at each locus i.e., $f_{\sigma}^0 = \sum_{i=1}^L f_{i\sigma_i}$, where $f_{i\sigma_i}$ is the contribution of the allele σ_i at locus i . Then to each genotype σ ’s additive fitness f_{σ}^0 , we added a random component x that was scaled by the roughness parameter r , causing its total fitness to become

$$f_{\sigma} = f_{\sigma}^0 + r \cdot x,$$

with $x \sim \mathcal{N}(0, 1)$. By tuning the parameter r , we could generate the entire spectrum of fitness landscapes with varying levels of correlation between fitness values and varying levels of epistatic interactions, with $r = 0$ leading to a purely additive landscape (with no epistatic interactions) and $r \gg 1$ leading to a House-of-Cards landscape (with many epistatically interacting loci). All RMF landscapes were normalised to ensure that the fittest genotype had unit fitness. We used the same Mt. Fuji landscape (f_v^0) and added the random component x to it, for different values of r , in order to generate landscapes with varying levels of correlation that could still be compared with each other.

Data processing

For GB1, we downloaded the raw data from the Supplementary Tables in ref. [9]; for ParE2 and ParE3, we obtained the data directly from the study authors.

We computed the fitness of each variant following Rubin et al. [74]. In particular, for GB1, the fitness of variant v is equal to

$$f_v = \log \left(\frac{c_{v,\text{sel}} + \frac{1}{2}}{c_{\text{wt},\text{sel}} + \frac{1}{2}} \right) - \log \left(\frac{c_{v,\text{inp}} + \frac{1}{2}}{c_{\text{wt},\text{inp}} + \frac{1}{2}} \right),$$

where $c_{v,\text{sel}}$ is the count of variant v in the sample after selection for binding immunoglobulin, $c_{\text{wt},\text{sel}}$ the count of the wild type (VDGV) in the sample after selection for binding immunoglobulin, $c_{v,\text{inp}}$ the count of variant v in the input sample, and $c_{\text{wt},\text{inp}}$ the count of the wild type in the input sample. The variance of the estimate is given by

$$\sigma_v^2 = \frac{1}{c_{v,\text{inp}} + \frac{1}{2}} + \frac{1}{c_{v,\text{sel}} + \frac{1}{2}} + \frac{1}{c_{\text{wt},\text{inp}} + \frac{1}{2}} + \frac{1}{c_{\text{wt},\text{sel}} + \frac{1}{2}}.$$

For ParE2 and ParE3, two replicate measurements were reported for each variant. For each variant and each replicate, the fitness and variance were computed as described above for GB1. The fitness values of the two replicates were then combined to compute the fitness of variant v as a weighted average of the two replicates, with weights given by the inverse of the corresponding variance:

$$f_v = \frac{1}{\sigma_{v,1}^2} f_{v,1} + \frac{1}{\sigma_{v,2}^2} f_{v,2},$$

and the variance of variant v was computed as

$$\sigma_v^2 = \frac{1}{\frac{1}{\sigma_{v,1}^2} + \frac{1}{\sigma_{v,2}^2}},$$

where $f_{v,i}$ and $\sigma_{v,i}^2$ denote the fitness and variance of the i -th replicate, respectively.

These values of fitness and variance were used as an input for empirical variance component regression [46], a method to reduce experimental noise and impute missing variants. The smoothed landscapes were obtained as maximum a posteriori estimates.

Enumerating all accessible paths

To enumerate all accessible paths in the landscapes, we employed dynamic programming. In particular, the number of accessible paths starting in an antipodal sequence (relative to the global peak) and ending at sequence v was represented as vector $P(v)$. The i -th element of $P(v)$, $P(v)_i$, stores the number of accessible paths from any antipodal sequence to variant v that are of length i , $i = 0, \dots, M$, where M is the maximum possible length of the walks. The vectors were initialized to $P(v) = (1, 0, \dots, 0)$ for the antipodal sequences, and to $P(v) = (0, \dots, 0)$ for all other sequences. Iterating through the sequences in ascending order of their fitness, we updated $P(v)$ using the following rule:

$$P(v)_i = \sum_{n \in \mathcal{N}(v)} P(n)_{i-1} \mathbb{I}[f(n) < f(v)],$$

where by $\mathcal{N}(v)$ we denote the set of all neighbors of sequence v , i.e. the set of sequences that differ from v in exactly one position, $f(v)$ is the fitness of variant v , and $\mathbb{I}[f(n) < f(v)]$ is an indicator function which is 1 if $f(n) < f(v)$ and 0 otherwise. The number of accessible paths of length i ending at the global peak GP is then given by $P(\text{GP})_i$, the number of accessible paths of length i ending at a local peak is given by $\sum_{p \in \text{local peaks}} P(p)_i$. For computational reasons, we computed the number of accessible paths only up to length $M = 10$.

Adaptive walk simulations

For the adaptive walk simulations, adaptation under strong selection and weak mutation can be treated as a Markov chain [18]. We initialized the walks on each genotype in the landscape and used the equal fixation model [9, 75] wherein, the next step of the walk was to a randomly chosen fitter neighbour of the initial genotype. This same process was iterated until no fitter neighbours could be found and a local or global peak had been reached.

Entropy of the basins of attraction

The entropy of the basins of attraction of all peaks in each landscape was computed by normalising the Shannon entropy by the maximum possible Shannon entropy for a given number of peaks (n_p) i.e., $\log_2(n_p)$. For landscapes that only had one peak, we omitted the normalisation and let the entropy be zero.

ACKNOWLEDGEMENTS

This work was supported by the Swiss National Science Foundation (Grant numbers PPOOP3_170604 and 310030_192541). We thank Thuy-Lan Lite for kindly providing the counts data for the ParE2 and ParE3 data sets. We thank members of the Computational Biology Group at ETH for discussions and the anonymous reviewers for their useful inputs.

DATA AVAILABILITY

The data that support the findings of this study are openly available at the following URL: https://github.com/MSri95/alphabet_cardinality.

REFERENCES

1. Wright, S. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. *Proceedings of the 6th International Congress of Genetics*, 356 (1932).
2. Kingman, J. On the properties of bilinear models for the balance between genetic mutation and selection. *Mathematical Proceedings of the Cambridge Philosophical Society* **81**, 443 (1977).
3. Kauffman, S. A. & Weinberger, E. D. The NK model of rugged fitness landscapes and its application to the maturation of the immune response. *Journal of Theoretical Biology* **141**, 211 (1989).
4. Aita Takuyo and Husimib, Y. Fitness Spectrum Among Random Mutants on Mt. Fuji Type Fitness Landscape. *Journal of Theoretical Biology* **182**, 469 (1996).
5. Bloom, J. & Arnold, F. In the light of directed evolution: pathways of adaptive protein evolution. *Proceedings of the National Academy of Sciences* (2009).
6. De Visser, J. A. G. M. & Krug, J. Empirical fitness landscapes and the predictability of evolution. *Nature Reviews Genetics* **15**, 480 (2014).
7. Lässig, M., Mustonen, V. & Walczak, A. Predicting evolution. *Nature Ecology and Evolution* **1** (2017).
8. Wu, Z., Kan, S. B. J., Lewis, R. D., Wittmann, B. J. & Arnold, F. H. Machine learning-assisted directed protein evolution with combinatorial libraries. *Proceedings of the National Academy of Sciences* **116**, 8852 (2019).
9. Wu, N., Dai, C. L., Olson, C. A., Lloyd-Smith, J. O. & Sun, R. Adaptation in protein fitness landscapes is facilitated by indirect paths. *eLife* **5**, e16965 (2016).

10. Schenk, M. F., Szendro, M. L., Salverda, I. G., Krug, J. & de Visser, J. A. G. Patterns of epistasis between beneficial mutations in an antibiotic resistance gene. *Molecular Biology and Evolution* **30**, 1779 (2013).
11. Gong, L. I., Suchard, M. A. & Bloom, J. D. Stability-mediated epistasis constrains the evolution of an influenza protein. *eLife* **2**, e00631 (2013).
12. Jiménez, J. I., Xulvi-Brunet, R., Campbell, G. W., Turk-MacLeod, R. & Chen, I. A. Comprehensive fitness landscape for RNA. *Proceedings of the National Academy of Sciences* **110**, 376 (2013).
13. Sarkisyan, K. S., Bolotin, D. A., Meer, V., Usmanova, D. R., Mishin, S., Sharonov, G. V., Ivankov, D. N., Bozhanova, N. G., Baranov, M. S., Soylemez, O., Bogatyreva, N. S., Vlasov, P. K., Egorov, E. S., Logacheva, M. D., Kondrashov, S., Chudakov, D. M., Putintseva, E. V., Mamedov, I. Z., Tawfik, D. S., Lukyanov, K. A. & Kondrashov, F. A. Local fitness landscape of the green fluorescent protein. *Nature* **533**, 397 (2016).
14. Aguilar-Rodríguez, J., Payne, J. & Wagner, A. A thousand empirical adaptive landscapes and their navigability. *Nature Ecology and Evolution* **1** (2017).
15. Olson, C. A., Wu, N. C. & Sun, R. A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain. *Current Biology* **24**, 2643 (2014).
16. Hartman, E. C., Lobba, M. J., Favor, A. H., Robinson, S. A., Francis, M. B. & Tullman-Ercek, D. Experimental Evaluation of Coevolution in a Self-Assembling Particle. *Biochemistry* **58**, 1527 (2019).
17. Pitt, J. & Ferré-D'Amaré, A. Rapid Construction of Empirical RNA Fitness Landscapes. *Science* **330**, 376 (2010).
18. Orr, H. A. The Population Genetics of Adaptation: The Adaptation of DNA Sequences. *Evolution* **56**, 1317 (2002).
19. Orr, H. A. A Minimum on the Mean Number of Steps Taken in Adaptive Walks. *Journal of Theoretical Biology* **220**, 241 (2003).
20. Szendro, I. G., Franke, J., de Visser, J. A. G. M. & Krug, J. Predictability of evolution depends nonmonotonically on population size. *Proceedings of the National Academy of Sciences* **110**, 571 (2013).
21. Gillespie, J. H. A simple stochastic gene substitution model. *Theoretical Population Biology* **23**, 202 (1983).
22. Poelwijk, F., Kiviet, D., Weinreich, D. & Tans, S. Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* **445**, 383 (2007).
23. Franke, J., Klozer, A., de Visser, J. & Krug, J. Evolutionary accessibility of mutational pathways. *PLoS Computational Biology* **7** (2011).

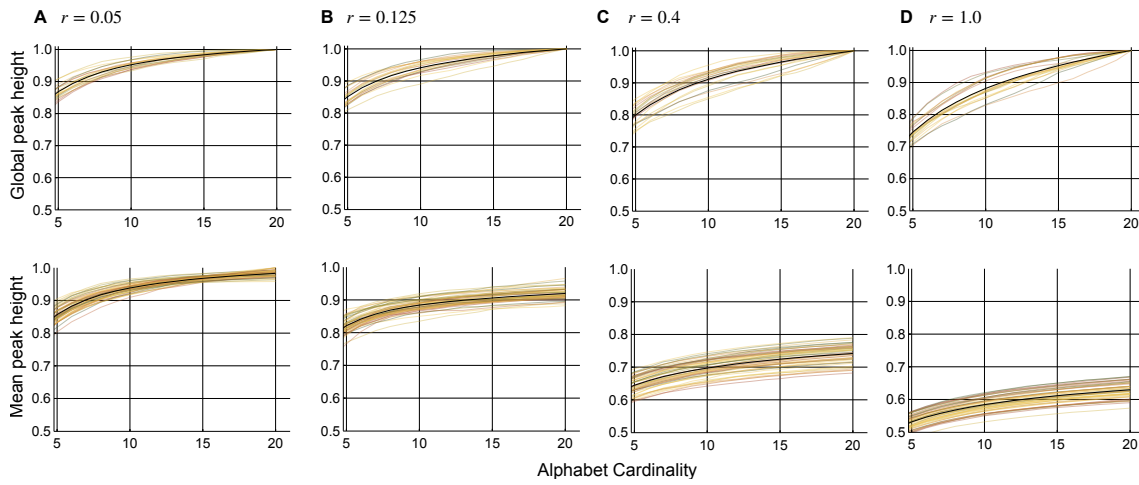
24. Franke, J. & Krug, J. Evolutionary Accessibility in Tunably Rugged Fitness Landscapes. *Journal of Statistical Physics* **148** (2012).
25. Hwang, S., Schmiegelt, B., Ferretti, L. & Krug, J. Universality Classes of Interaction Structures for NK Fitness Landscapes. *Journal of Statistical Physics* **172**, 226 (2018).
26. Zagorski, M., Burda, Z. & Waclaw, B. Beyond the Hypercube: Evolutionary Accessibility of Fitness Landscapes with Realistic Mutational Networks. *PLoS Computational Biology* **12** (2016).
27. Schmiegelt, B. & Krug, J. Accessibility Percolation on Cartesian Power Graphs. *Journal of Mathematical Biology* **86**, 46 (2023).
28. Gavrillets, S. Fitness Landscapes and the Origin of Species. *Princeton Univ. Press* (2004).
29. Greenbury, S., Louis, A. & Ahnert, S. The structure of genotype-phenotype maps makes fitness landscapes navigable. *Nature Ecology & Evolution* **6**, 1742 (2022).
30. Pines, G., Winkler, J. D., Pines, A. & Gill, R. T. Refactoring the Genetic Code for Increased Evolvability. *mBio* **8** (2017).
31. Kauffman, S. A. The Origins of Order: Self Organization and Selection in Evolution. *Oxford University Press* (1993).
32. Perelson, A. S. & Macken, C. A. Protein Evolution on Partially Correlated Landscapes. *Proceedings of the National Academy of Sciences*, 9657 (1995).
33. Palmer, A. C., Toprak, E., Baym, M., Kim, S., Veres, A., Bershtein, S. & Kishony, R. Delayed commitment to evolutionary fate in antibiotic resistance fitness landscapes. *Nature Communications* **6** (2015).
34. Manhart, M. & Morozov, A. V. Scaling properties of evolutionary paths in a biophysical model of protein adaptation. *Physical Biology* **12** (2015).
35. Aita, T., Uchiyama, H., Inaoka, T., Nakajima, M., T, K. & Husimi, Y. Analysis of a local fitness landscape with a model of the rough Mt. Fuji-type landscape: application to prolyl endopeptidase and thermolysin. *Biopolymers* **54**, 64 (2000).
36. Aita, T. & Husimi, Y. Adaptive walks by the fittest among finite random mutants on a Mt. Fuji-type fitness landscape II. Effect of small non-additivity. *Journal of Mathematical Biology* **41**, 207 (2000).
37. Ambrogelly, A., Palioura, S. & Söll, D. Natural expansion of the genetic code. *Nature Chemical Biology* **3**, 29 (2007).
38. De la Torre, D. & Chin, J. W. Reprogramming the genetic code. *Nature Reviews Genetics* **22**, 169 (2021).
39. Aita, T., Iwakura, M. & Husimi, Y. A cross-section of the fitness landscape of dihydrofolate reductase. *Protein Engineering, Design and Selection* **14**, 633 (2001).

40. Lite, T.-L. V., Grant, R. A., Nocedal, I., Littlehale, M. L., Guo, M. S. & Laub, M. T. Uncovering the basis of protein-protein interaction specificity with a combinatorially complete library. *eLife* **9**, e60924 (2020).
41. Greenbury, S. F., Schaper, S., Ahnert, S. E. & Louis, A. A. Genetic Correlations Greatly Increase Mutational Robustness and Can Both Reduce and Enhance Evolvability. *PLoS Computational Biology* **12**(3) (2016).
42. Szendro, I. G., Schenk, M. F., Franke, J., Krug, J. & de Visser, J. A. G. M. Quantitative analyses of empirical fitness landscapes. *Journal of Statistical Mechanics: Theory and Experiment* (2013).
43. Tokuriki, N. & Tawfik, D. S. Protein Dynamism and Evolvability. *Science* **324**, 203 (2009).
44. Srivastava, M. & Payne, J. L. On the incongruence of genotype-phenotype and fitness landscapes. *PLoS Computational Biology* **18**(9) (2022).
45. Zhou, J. & McCandlish, D. M. Minimum epistasis interpolation for sequence-function relationships. *Nature Communications* **11**, 1782 (2020).
46. Zhou, J., Wong, M. S., Chen, W.-C., Krainer, A. R., Kinney, J. B. & McCandlish, D. M. Higher-order epistasis and phenotypic prediction. *Proceedings of the National Academy of Sciences* **119**, e2204233119 (2022).
47. Firnberg, E. & Ostermeier, M. The genetic code constrains yet facilitates Darwinian evolution. *Nucleic Acids Res.* **41**, 7420 (2013).
48. Massey, S. E. A Neutral Origin for Error Minimization in the Genetic Code. *J Mol Evol* **67**, 510 (2008).
49. Di Giulio, M. An extension of the coevolution theory of the origin of the genetic code. *Biology Direct* **3**, 37 (2008).
50. Novozhilov, A. S. & Koonin, E. V. Exceptional error minimization in putative primordial genetic codes. *Biology Direct* **4**, 44 (2009).
51. Higgs, P. G. A four-column theory for the origin of the genetic code: tracing the evolutionary pathways that gave rise to an optimized code. *Biol Direct* **4**, 16 (1 2009).
52. Bajić, D., Vila, J., Blount, Z. & Sanchez, A. On the deformability of an empirical fitness landscape by microbial evolution. *Proceedings of the National Academy of Sciences* **115**, 201808485 (2018).
53. Dieckmann, U. & Doebeli, M. On the origin of species by sympatric speciation. *Nature* **400**, 354 (1999).
54. Macken, C. A. & Perelson, A. S. Protein evolution on rugged landscapes. *Proceedings of the National Academy of the Sciences* **86**, 6191 (1989).

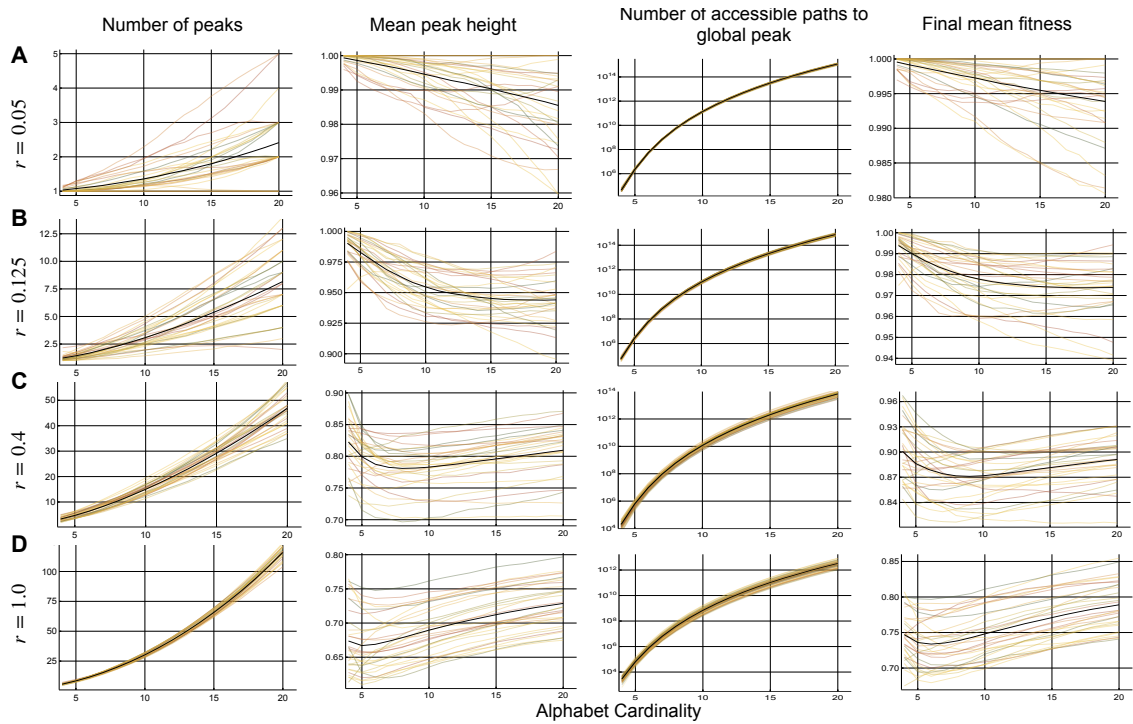
55. Poelwijk, F., Tănase-Nicola, S., Kiviet, D. & Tans, S. Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. *Journal of Theoretical Biology* **272**, 141 (2011).
56. Park, S.-C., Hwang, S. & Krug, J. Distribution of the number of fitness maxima in Fisher's geometric model. *Journal of Physics A: Mathematical and Theoretical* **53** (2020).
57. Liu, C. C. & Schultz, P. G. Adding New Chemistries to the Genetic Code. *Annual Review of Biochemistry* **79**. PMID: 20307192, 413 (2010).
58. Dumas, A., Lercher, L., Spicer, C. D. & Davis, B. G. Designing logical codon reassignment – Expanding the chemistry in biology. *Chemical Science* **6**, 50 (1 2015).
59. Nödling, A. R., Spear, L. A., Williams, T. L., Luk, L. Y. & Tsai, Y.-H. Using genetically incorporated unnatural amino acids to control protein functions in mammalian cells. *Essays in Biochemistry* **63**, 237 (2019).
60. Liu, C. C., Mack, A. V., Brustad, E. M., Mills, J. H., Groff, D., Smider, V. V. & Schultz, P. G. Evolution of Proteins with Genetically Encoded “Chemical Warheads”. *Journal of the American Chemical Society* **131**, 9616 (2009).
61. Liu, C. C., Choe, H., Farzan, M., Smider, V. V. & Schultz, P. G. Mutagenesis and Evolution of Sulfated Antibodies Using an Expanded Genetic Code. *Biochemistry* **48**, 8891 (2009).
62. Brustad, E. M. & Arnold, F. H. Optimizing non-natural protein function with directed evolution. *Current Opinion in Chemical Biology* **15**. Biocatalysis and Bio-transformation/Bioinorganic Chemistry, 201 (2011).
63. Romesberg, F. E. Discovery, implications and initial use of semi-synthetic organisms with an expanded genetic alphabet/code. *Philosophical Transactions of the Royal Society B: Biological Sciences* **378**, 20220030 (2023).
64. Liu, C. C. & Schultz, P. G. Recombinant expression of selectively sulfated proteins in *Escherichia coli*. *Nature Biotechnology* **24**, 1436 (2006).
65. Grünewald, J., Tsao, M.-L., Perera, R., Dong, L., Niessen, F., Wen, B. G., Kubitz, D. M., Smider, V. V., Ruf, W., Nasoff, M., Lerner, R. A. & Schultz, P. G. Immuno-chemical termination of self-tolerance. *Proceedings of the National Academy of Sciences* **105**, 11276 (2008).
66. Mayer-Bacon, C., Agboha, N., Muscalli, M. & Freeland, S. Evolution as a Guide to Designing xeno Amino Acid Alphabets. *International Journal of Molecular Sciences* **22** (2021).
67. Burke, A. J., Lovelock, S. L., Frese, A., Crawshaw, R., Ortmyer, M., Dunstan, M., Levy, C. & Green, A. P. Design and evolution of an enzyme with a non-canonical organocatalytic mechanism. *Nature* **570**, 219 (2019).

68. Windle, C. L., Simmons, K. J., Ault, J. R., Trinh, C. H., Nelson, A., Pearson, A. R. & Berry, A. Extending enzyme molecular recognition with an expanded amino acid alphabet. *Proceedings of the National Academy of Sciences* **114**, 2610 (2017).
69. Xiao, H., Nasertorabi, F., Choi, S.-h., Han, G. W., Reed, S. A., Stevens, R. C. & Schultz, P. G. Exploring the potential impact of an expanded genetic code on protein function. *Proceedings of the National Academy of Sciences* **112**, 6961 (2015).
70. Rogers, J. M., Passioura, T. & Suga, H. Nonproteinogenic deep mutational scanning of linear and cyclic peptides. *Proceedings of the National Academy of Sciences* **115**, 10959 (2018).
71. Drienovska, I. & Roelfes, G. Expanding the enzyme universe with genetically encoded unnatural amino acids. *Nature Catalysis* **3**, 1 (2020).
72. Li, J. C., Liu, T., Wang, Y., Mehta, A. P. & Schultz, P. G. Enhancing Protein Stability with Genetically Encoded Noncanonical Amino Acids. *Journal of the American Chemical Society* **140**, 15997 (2018).
73. Li, J. C., Nasertorabi, F., Xuan, W., Han, G. W., Stevens, R. C. & Schultz, P. G. A Single Reactive Noncanonical Amino Acid Is Able to Dramatically Stabilize Protein Structure. *ACS Chemical Biology* **14**, 1150 (2019).
74. Rubin, A. F., Gelman, H., Lucas, N., Bajjalieh, S. M., Papenfuss, A. T., Speed, T. P. & Fowler, D. M. A statistical framework for analyzing deep mutational scanning data. *Genome Biology* **18**, 150 (2017).
75. 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 (2006).

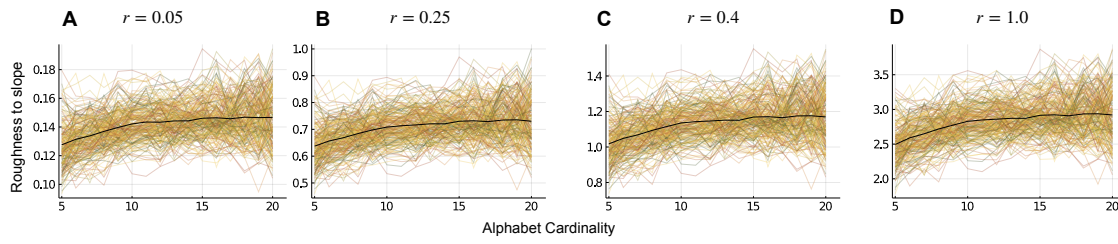
3.5 SUPPLEMENTARY INFORMATION



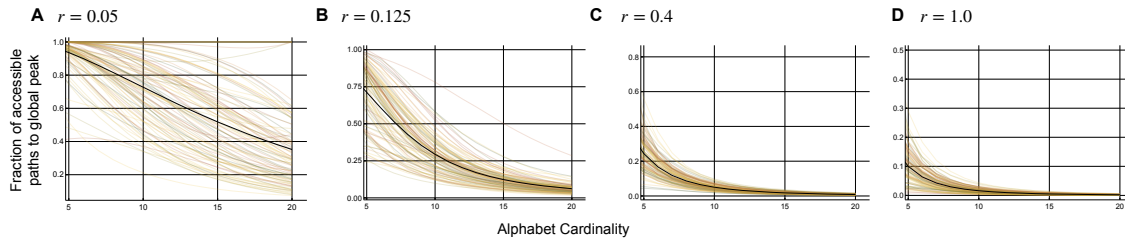
S3.1: The global peak height (top row) and the mean peak height (bottom row) as a function of the alphabet cardinality with A) $r = 0.05$, B) $r = 0.0125$, C) $r = 0.4$ and D) $r = 1.0$. In these landscapes, we did not control for the emergence of the global peak. The coloured lines show results for the individual landscapes which are averaged over 200 alphabet expansions. The black lines show the results averaged over 20 RMF landscapes (top panel) and 50 RMF landscapes (bottom panel).



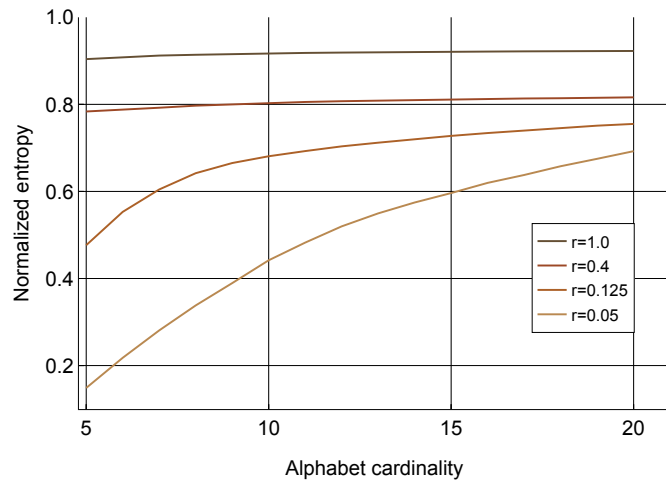
S3.2: The characteristic curves for a set of 50 $L = 3$ RMF landscapes with varying levels of correlation, with the magnitude of correlation determined by the roughness parameter r . The first column depicts the number of peaks in the landscapes, the second column depicts the mean peak height, the third column depicts the number of accessible paths to the global peak and the fourth column depicts the final mean fitness. The coloured lines show results for 30 RMF landscapes. Even for those landscapes that contained only 2 or 3 amino acids in their global peak, the expansion is shown from cardinality 5 onward such that the x axis remains the same for all landscapes. The black lines show the average over those landscapes.



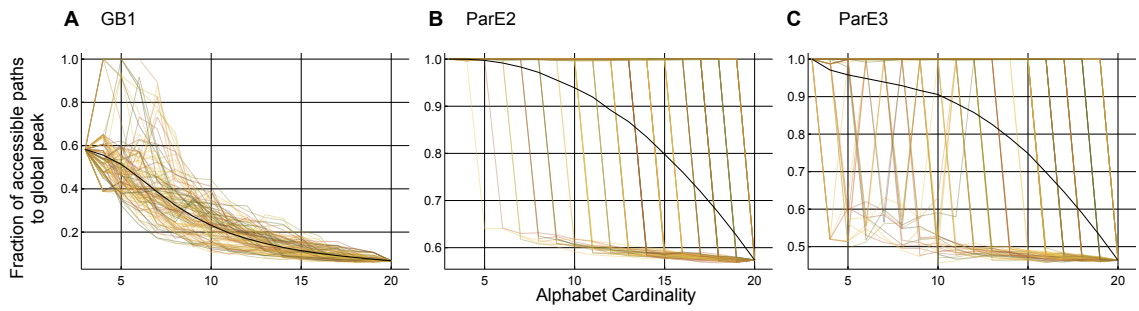
S3.3: The roughness-to-slope ratio for $L = 4$ RMF landscapes with varying levels of correlation, with the magnitude of correlation determined by the roughness parameter r . The coloured lines show results for 200 RMF landscapes averaged over 1000 alphabet expansions. The black line shows the average over the 200 RMF landscapes.



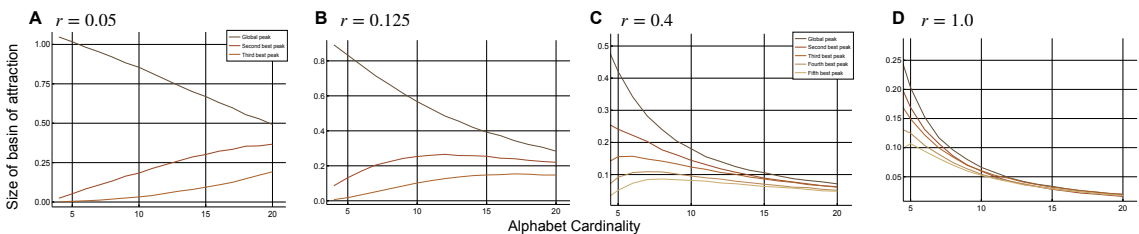
S3.4: Fraction of accessible paths to the global peak, out of all accessible paths starting in the antipodal genotypes of the global peak, for RMF landscapes with A) $r = 0.05$, B) $r = 0.125$, C) $r = 0.4$ and D) $r = 1.0$.



S3.5: Normalised Shannon entropy of the basins of attraction of all the peaks in RMF landscapes with different values of the roughness parameter r .



S3.6: Fraction of accessible paths to the global peak, out of all accessible paths starting in the antipodal genotypes of the global peak, for empirical fitness landscapes A) GB1, B) ParE2, and C) ParE3.



S3.7: Basins of attraction of the highest three (A, B) or five (C, D) peaks averaged over 10 RMF landscapes with $L = 3$ and A) $r = 0.05$, B) $r = 0.125$, C) $r = 0.4$, and D) $r = 1.0$. Lines are rank ordered from top to bottom based on peak height.

EVOLUTION OF COMPLEX FORMS ON BIOMORPHS GENOTYPE-PHENOTYPE-FITNESS MAPS

*I took the one less traveled by, And that has
made all the difference.*

— Robert Frost

In preparation as: Srivastava M, Louis AA and Martin N. (2024) Increased evolvability can decrease ruggedness in genotype-phenotype-fitness maps.

Authors' contributions: M.S., N.M. and A.A.L designed research; M.S. and N.M. performed research; M.S., N.M. and A.A.L analyzed data; and M.S., N.M. and A.A.L wrote the paper.

ABSTRACT

Understanding the dynamics of evolution requires knowledge of the genotype-phenotype-fitness (GPF) map. Here we study how the architecture of the underlying genotype-to-phenotype (GP) map interacts with the phenotype-to-fitness map to influence the topographical properties of and evolutionary dynamics on a GPF map based on Richard Dawkins' Biomorphs, a toy model of morphological evolution. We treat two distinct phenotype to fitness assignments: In the uncorrelated map each phenotype is assigned a random fitness, which leads to maximal ruggedness. In the correlated map, the fitness increases linearly with phenotypic proximity to a target phenotype. For the uncorrelated map, we derive analytic relationships between the number and heights of fitness peaks and the number of novel phenotypes accessible in a single mutation from the peak phenotype, i.e. its evolvability. GP maps with many low-evolvability phenotypes have a higher number of peaks and peaks with larger phenotype evolvability are typically higher in fitness. These relationships should hold for a much wider class of GPF maps with random fitness assignments. Interestingly, while the Biomorphs GP map shares many global structural properties with molecular GP maps, it differs in how it shapes the topography of fitness landscapes. For correlated landscapes it is more likely to generate a multi-peaked structure, and it does not promote navigability in uncorrelated fitness maps. Although using the correlated fitness mapping significantly enhances navigability, the navigability enhancing accessible paths are seldom realised in evolutionary simulations.

Finally, we show that this difficulty of navigating the fitness landscapes is significantly reduced if we expand the set of target phenotypes to ‘similar enough’ phenotypes.

4.1 INTRODUCTION

Evolution is a two-step process. First, mutations in genotypes generate novel phenotypic variation. Second, natural selection acts on this variation to enhance the frequency of fitter phenotypes in a population. To understand evolution we must work out how mutations in genotypes are mapped to changes in fitness. This mapping can be represented by a structure called the fitness landscape [1].

It is well established that the ruggedness of fitness landscapes, measured by the number and height of local peaks, influences the dynamics of adaptation, by determining whether uphill paths to the global optimum exist and whether evolving populations will get stuck on sub-optimal local peaks [2–5]. These uphill paths to a given target are called *accessible paths* [6].

The presence of accessible paths in statistical models of fitness landscapes, such as the House-of-cards, Mt. Fuji or the NK model [7–10] have been extensively studied [11–13]. While the structure of these genotype-fitness (GF) maps are meant to emulate the full genotype-phenotype-fitness (GPF) map, they only treat the intermediate genotype-phenotype (GP) map implicitly. In recent years, there has been a lot of progress in studying GP maps in isolation [14–16], showing that many biological systems exhibit similar structure in how they map genotypes to phenotypes. For example, GP maps have neutral correlations [17], meaning that genotypes mapping to the same phenotype are more likely to be mutationally connected, leading to large *neutral components*, which are sets of connected genotypes that map to the same phenotype [18]. A recent study of several genotype-phenotype-fitness (GPF) maps [19] found that ubiquitous connected neutral components greatly increased the *navigability*, defined as the average probability that an accessible path exists between randomly chosen source and target genotypes on a fitness landscape. Furthermore, they found that the prevalence of local peaks was highly dependent on the GP map model. In addition, the accessible paths were likely to be utilised during evolutionary dynamics.

While molecular GPF maps were shown to have high navigability, this raises the question of whether other GPF maps – such as those at higher levels of biological organization – exhibit the same levels of navigability. We investigate this question on Richard Dawkins’ Biomorphs GP map, which is a toy model of body plan development [21, 22] that was recently shown to share many structural features with molecular GP maps [20]. As shown in Figure 4.1, in this model, numeric genotypes define constituent vectors, which are then combined in a recursive developmental process to generate the phenotype. This simple process is capable of producing a rich array of forms, such as those resembling insects and plants.

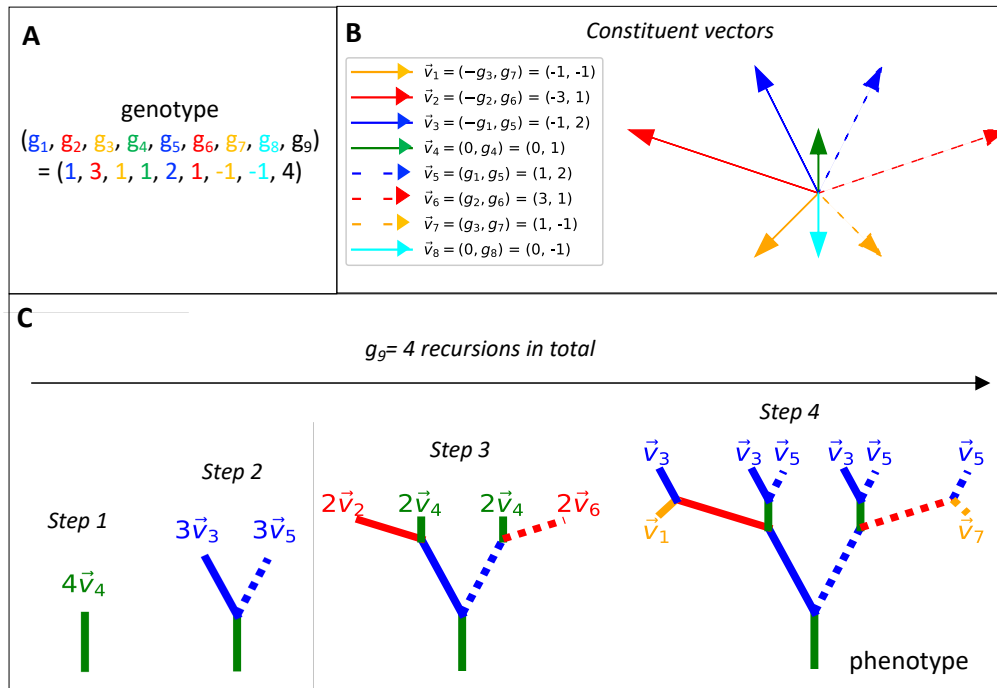


FIGURE 4.1: THE BIOMORPHS GENOTYPE-PHENOTYPE MAP A. The genotype is a set of nine integers and is used to define the constituent vectors that ultimately produce the Biomorphs phenotype. We use a fixed range of values for each locus ($-3 \leq g_n \leq 3$, for $1 \leq n \leq 8$, and $1 \leq g_9 \leq 8$). B. The integers at the first eight loci (labeled g_1 to g_8) are used to define the eight two-dimensional constituent vectors. C. The 2D Biomorphs phenotype is constructed recursively from these vectors, with the number of recursions set by the integer at the ninth locus, g_9 . For example in the figure above, we have $g_9 = 4$ and thus there are four recursions: In Step 1, the recursion index i is set to 4 and we start constructing the Biomorph with the vector $i \cdot \vec{v}_{g_9}$. Then in the next step, we set the recursion index i to $i - 1$ and add vectors, $i \cdot \vec{v}_{g_9-1}$ and $i \cdot \vec{v}_{g_9+1}$ to the end point of the structure from Step 1. This process is repeated, by reducing i in each step and adding vectors $i \cdot \vec{v}_{j\pm 1}$ to the end point of each vector (\vec{v}_j) added in the previous step, until $i = 1$. Since vector \vec{v}_8 would be first used in the fifth recursion, it is not utilised in this Biomorph. We ensure that the vector index $1 \leq j \leq 8$ by using the convention $i \rightarrow 8$, when $i = 0$ and $i \rightarrow 1$ when $i = 9$. Figure adapted from [20].

Dawkins used this model to illustrate how to evolve a desired shape by many successive small fitness increasing steps. In particular, he demonstrated how to gradually steer the evolution of Biomorphs towards particular desired shapes by manually picking out mutants at each generation that most closely resembled his desired target shape. This echoes Darwin's explanation for the emergence of complex forms such as the human eye: "...if numerous gradations from a simple and imperfect eye to one complex and perfect can be shown to exist, each grade being useful to its possessor...then the difficulty of believing that a perfect and complex eye could be formed by natural selection, though insuperable

by our imagination, should not be considered as subversive of the theory" [23]. Inspired by Dawkins' explorations, we will study in detail the frequency and properties of such series of beneficial mutations that can transform a simple Biomorphs phenotype to a complex one. In other words, we are interested in quantifying the existence of accessible paths between a given source and a target phenotype in the Biomorphs GPF map.

Analysis of the GPF map can build on detailed exhaustive analyses of the first part, the GP map, including the Biomorphs GP map [16, 20]. Many GP map concepts such as many genotypes mapping to the same phenotype (redundancy), a highly non-uniform distribution of the number of genotypes mapping to a phenotype and high mutational robustness of phenotypes are applicable to the Biomorphs GP map [14, 20]. These properties lead to the formation of large neutral components in the GP map [18]. Further, there is a positive correlation between the robustness and *evolvability* of phenotypes – i.e. the number of novel phenotypes accessible in a single mutation from the neutral set of a given phenotype [24]. Moreover, it was recently established that the frequency-complexity relationship of Biomorphs phenotypes, like many other input-output maps, displays simplicity bias, i.e. phenotypes with large neutral sets have low complexity and outputs with high complexity have small neutral sets [20, 25, 26]. However, the Biomorphs GP map also has some peculiarities, such as a sparsely connected *allele graph*, which encodes the connectivity between all possible alleles at a single locus [12]. In the Biomorphs GP map, each locus can only mutate to its nearest neighbour in the allele graph. This raises the question of how these different properties of the Biomorphs GP map cumulatively influence the topography and navigability of the resulting fitness landscapes.

In this paper, we apply the genotype-phenotype-fitness framework to the Biomorphs GP map, by building on two extreme phenotype-fitness maps from ref [19] – a completely uncorrelated map and a linearly correlated map (see Figure 4.2 and Methods, both adapted). We then examine the topographical features of the resulting fitness landscapes, particularly the number and height of the peaks and relate these to the properties of the genotype-phenotype map. In particular we find new analytic links to the evolvability and establish two general principles that can be applied to all genotype-phenotype-fitness maps with uncorrelated phenotype-fitness functions: (1) fitness landscapes of GP maps with low evolvability phenotypes are likely to have higher ruggedness and (2) local peaks with higher evolvabilities should be fitter on average.

Next, we focus on the *landscape navigability* of the two classes of fitness landscapes, by checking for the existence of accessible paths between source-target pairs belonging to different complexity classes. We also relate this quantity to the evolvability of the phenotypes in the GP map.

Finally, we study the *evolutionary navigability* of the fitness landscapes, by quantifying how often existing accessible paths are realised in evolutionary processes. We find that although the correlated fitness landscape has a high landscape navigability, its evolutionary navigability is very low, indicating that even if navigable paths exist, they

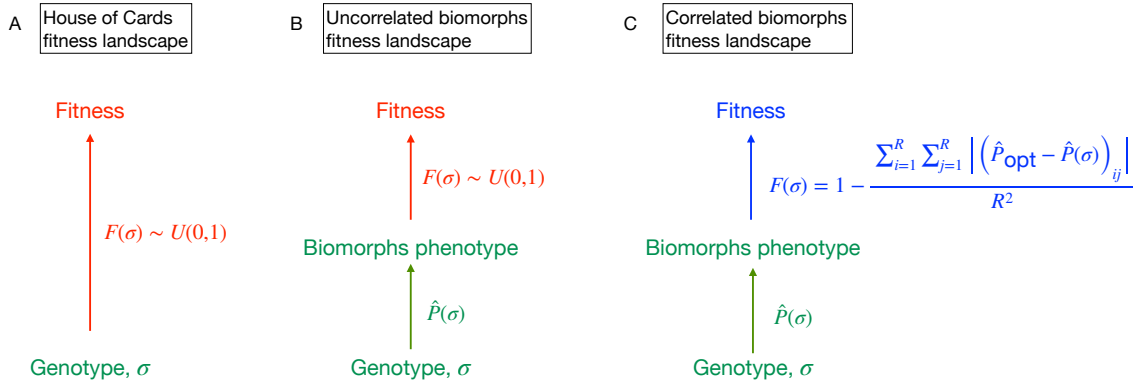


FIGURE 4.2: THREE MODELS OF FITNESS LANDSCAPES. A. The House-of-cards fitness landscape ignores the intermediate phenotypic layer and assigns fitness values directly to genotypes by sampling from a uniform distribution $U(0,1)$ with fitness between 0 and 1. The uncorrelated (B.) and correlated (C.) fitness landscapes are based on the Biomorphs GP map, by mapping each genotype first to its phenotype and then using a phenotype-fitness map. This fitness is assigned randomly from a uniform distribution in the uncorrelated landscape and depends linearly on the distance to the target phenotype in the correlated landscape. $\hat{P}(\sigma)$ is the coarse-grained array representing the phenotype and $F(\sigma)$ is the fitness value assigned to genotype σ respectively. \hat{P}_{opt} is the coarse-grained array representing the optimal phenotype and R is the dimension of the array. In this paper, $R = 30$.

may not be used. Finally, we study how broadening our definition of a ‘successful’ target phenotype can enhance the evolutionary navigability.

4.2 RESULTS

4.2.1 Defining genotype-phenotype-fitness maps

The choice of the phenotype-fitness map plays a crucial role in determining the properties of the fitness landscape [27]. In this paper, we employ three models, as illustrated in Figure 4.2:

- The **House-of-Card fitness landscape**, where each genotype is simply assigned a random fitness value. This well-studied and analytically tractable [5, 7, 12] system generates a null model for a fitness landscape defined without a GP map.
- The **uncorrelated Biomorphs fitness landscape** [19], where each genotype is first mapped to its Biomorphs phenotype, and then each phenotype assigned a random fitness. This ‘worst-case’ phenotype-fitness relationship [19], is useful due to its analytic tractability.

- The **correlated Biomorphs fitness landscape** [19], where each genotype is first mapped to its Biomorphs phenotype, and then each phenotype assigned a fitness depending on its phenotypic distance from a fixed target phenotype. This is closer to what Dawkins had in mind, when he evolved Biomorphs through artificial selection.

Our computational treatment of Biomorphs used coarse-graining of the phenotypes, to ensure that only visually distinct Biomorphs are classed as different phenotypes. This led to each 2D Biomorph being represented by a 30×30 pixel array (see Methods).

4.2.2 *The link between local peaks and evolvability*

We start by analysing an important determinant of evolutionary dynamics: the existence of local fitness peaks [5, 28], which can trap evolving populations and prevent them from reaching the global fitness optimum. We first study the simpler, uncorrelated fitness landscapes to develop analytic predictions, and then turn to the more complex correlated fitness landscapes.

The simplest approach is to directly map genotypes to random fitness values as in the House-of-Cards fitness landscape in Figure 4.2A. For a standard Hamming graph, this has the advantage that one can directly calculate the expected number of fitness peaks, which is known to be $\frac{a^L}{(a-1) \cdot L + 1}$, where a is the alphabet cardinality and L is the sequence length [29]. The Biomorphs model has $a \approx 7$ and sequence length $L = 9$. If we assume that its topology is a standard Hamming graph, this would lead to $\approx 7 \times 10^5$ local peaks.

The next step up in complexity is the ‘uncorrelated’ landscape in Figure 4.2B, where each genotype is first mapped to a phenotype according to the Biomorphs GP map. The predicted number of local peaks will change compared to the House-of-Cards prediction for two contrasting reasons. First, the redundancy and robustness in the mapping from genotype to phenotype, will introduce fitness plateaus: if two neighbouring genotypes map to the same phenotype, they will also have the same fitness. As noted by Greenbury et al. [19], such neutral plateaus can enhance the navigability of a landscape, and reduce the number of fitness peaks. Second, the allele graph of the Biomorphs GP map (see Figure 4.7 A) differs from the standard Hamming graph, and this increases the number of peaks. This increase can be understood by realising that the allele at each locus can only change by ± 1 , implying that the maximum number of mutational neighbours of any genotype is $2 * L$, as opposed to $(a - 1) * L$ in a Hamming graph. Thus, the expected number of peaks in a House-of-Cards landscape with this special allele graph is

$$\langle \text{peaks} \rangle \geq \frac{a^L}{2 \cdot L + 1} \gg \frac{a^L}{(a - 1) \cdot L + 1}.$$

Therefore, while the existence of fitness plateaus in the Biomorphs fitness landscape decreases the number of peaks, the sparse connectivity of the allele graph increases the number of peaks.

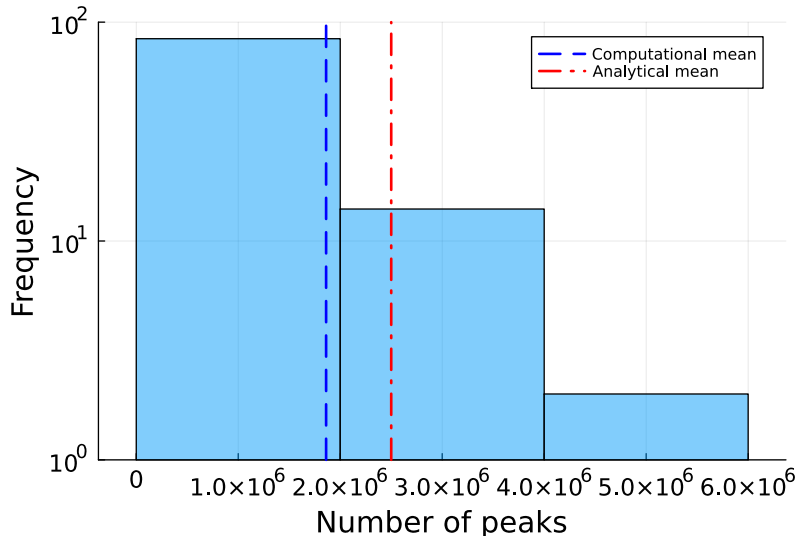


FIGURE 4.3: NUMBER OF LOCAL PEAKS IN THE UNCORRELATED FITNESS LANDSCAPE FROM FIGURE 4.2B: Distribution of the number of peaks for 100 realisations of uncorrelated fitness landscapes on the computational Biomorphs GP map. The mean number of peaks is shown with the blue dashed line, while the analytical estimate of the mean is shown with a red dot-dashed line. The analytical estimate is of the correct order of magnitude, but overestimates the true value because the analytic model underestimates the evolvability values in the GP map [20].

Despite these complications, the number of peaks in the uncorrelated Biomorphs landscape can be estimated based on one property of the genotype-phenotype map, the *neutral component (NC) evolvability* ϵ_p , defined as the number of novel phenotypes that are accessible by point mutations from a single NC [18]. A NC is a local peak if and only if it is fitter than its ϵ_p phenotypic neighbours. The probability of this constellation, if the fitness values of all phenotypes are drawn from the same distribution (regardless of the distribution) is

$$\frac{1}{\epsilon_p + 1}.$$

This expression thus quantifies the probability that a NC is a local peak, simply in terms of its evolvability. The NC evolvability takes into account the combined effects of the allele graph and the effect of large neutral plateaus, which tend to have higher evolvabilities [24]. The larger the value of ϵ_p , the less likely it is for the particular NC to form a fitness peak.

To estimate the mean number of local peaks on the entire fitness landscape we simply sum over the individual probabilities of peaks to obtain:

$$\langle \text{peaks} \rangle \approx \sum_{NCs} \frac{1}{\epsilon_p + 1} \quad (4.1)$$

In a general GP map, evolvability values in the sum would have to be found computationally. In the Biomorphs model, however, we can use an analytic approximation to the GP map from ref [20]. This analytic approximation is possible because the ninth locus in the Biomorphs genotype determines which of the first eight loci contribute to the construction of the phenotype. The contributing loci are constrained, since altering them alters the phenotype. Those loci that do not contribute are unconstrained and can be mutated without altering the phenotype. The analytically tractable model does not coarse-grain over phenotypes as the computational model does, and so it has a larger number of distinct phenotypes than the computational model. The analytical model gives a parametric expression for ϵ_p [20] (see Methods). Then the expected number of peaks can be estimated as:

$$\langle \text{peaks} \rangle \approx \sum_{g_9} N(g_9) \cdot \frac{1}{\epsilon_p(g_9) + 1} \quad (4.2)$$

where $N(g_9)$ is the number of NCs with a fixed g_9 . Equation 4.2 also indicates that NCs with low $\epsilon_p(g_9)$ will have a higher mean number of peaks.

From equation 4.2, we analytically calculate that there will be 2.5×10^6 peaks on average in the Biomorphs landscape (see Supplementary Figure 4.1 for the variation in the number of peaks with the maximum allowed value of g_9). Our analytic estimates are consistent with computational analyses of the Biomorphs GP map with an uncorrelated phenotype-fitness function: based on 100 uncorrelated fitness landscapes, we find 1.86×10^6 peaks on average (see Figure 4.3 for the full distribution). We expect our analytic estimate to be larger than what we measure computationally because the analytic model has a higher number of phenotypes and lower evolvabilities, as compared to the computational treatment [20]. However, the order of magnitude is correct, and this analytic result helps us qualitatively understand the tight link between NC evolvability and the number of peaks.

Having analysed the number of local peaks, we next turn to the heights of local peaks, which indicate how fit a population would be if it was trapped on these peaks. As expected, the average peak height is much higher than the average fitness of a phenotype (which is 0.5) and the distribution of peak heights is highly skewed towards high fitness values (Figure 4.4 A). Further, we expect a positive correlation between peak height and the evolvability of the NC that forms the peak (henceforth called the *peak evolvability*) because NCs with high evolvability need to out-compete more neighbours to be a peak. As explained above, for a NC with phenotype p and evolvability ϵ_p to be a peak, it must have a fitness higher than its ϵ_p neighbours. Since the fitness value of each NC is

sampled from a uniform distribution $U(0,1)$, the distribution of the height of a peak with evolvability ϵ_p (see Supplementary material for proof) is given by:

$$p_{f_{max}}(x) = (\epsilon_p + 1) \cdot x^{\epsilon_p} \implies \langle f_{max} \rangle = \frac{\epsilon_p + 1}{\epsilon_p + 2} \quad (4.3)$$

Thus, peaks with high evolvability also tend to have higher fitness. In Figure 4.4 B, we show the relationship between the peak height and peak evolvability for one uncorrelated fitness landscape and juxtapose it with the analytic expression in equation 4.3 from above (see Supplementary Figure 4.2 for the distribution of peak evolvability). We see an excellent agreement between the theoretical prediction and the computational data. While this expression is specific to the fitness distribution used, it can also be derived for other distributions.

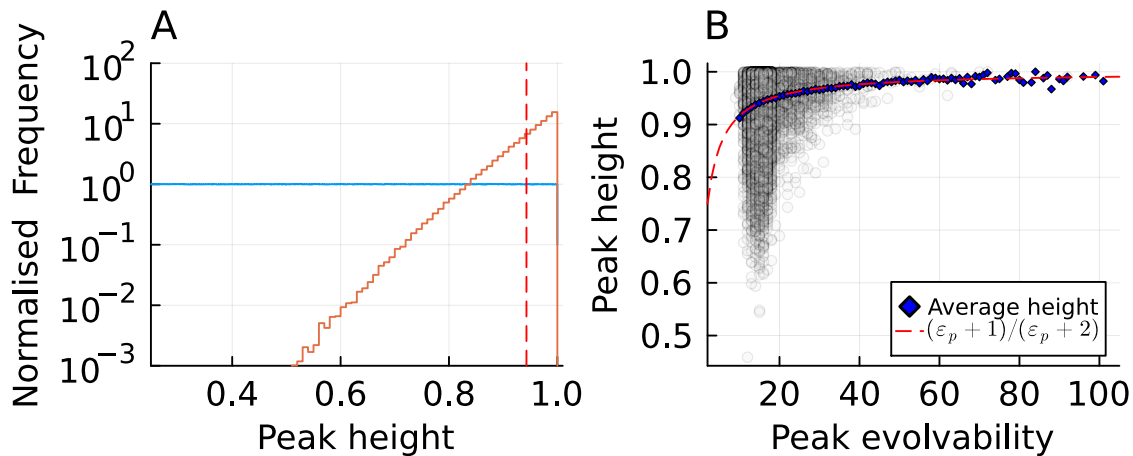


FIGURE 4.4: LOCAL PEAKS IN THE UNCORRELATED LANDSCAPE ARE HIGH AND EVOLVABILITY-DEPENDENT: A. Distribution of peak height (orange) and the distribution of fitness over all NCs in the landscape (blue), which is a uniform distribution for a single realisation of the uncorrelated fitness landscape. The dashed red line depicts the mean peak height. B. Peak height vs peak evolvability: computational data for a single realisation of the uncorrelated landscape is shown in grey, a moving average over the data for a given evolvability is shown with blue dots and the theoretical expectation is shown with a red dashed line.

Having analysed the topography of the analytically tractable uncorrelated fitness landscapes, we next turn to the linearly correlated one in Figure 4.2C, which we can only treat computationally. In the correlated landscape, similar phenotypes have similar fitness (see Methods), and so we expect a smoother landscape [19] with fewer peaks. Indeed, we find $(3.0 \pm 1.4) \times 10^5$ peaks on average when a simple phenotype is chosen as the global peak and $(3.4 \pm 1.2) \times 10^5$ peaks on average when a complex phenotype is chosen as the global peak (see Figure 4.5 A). These numbers are lower than for the uncorrelated landscapes. Further, we expect the landscapes with a complex phenotype as global peak to have higher ruggedness. This is because firstly, complex phenotypes

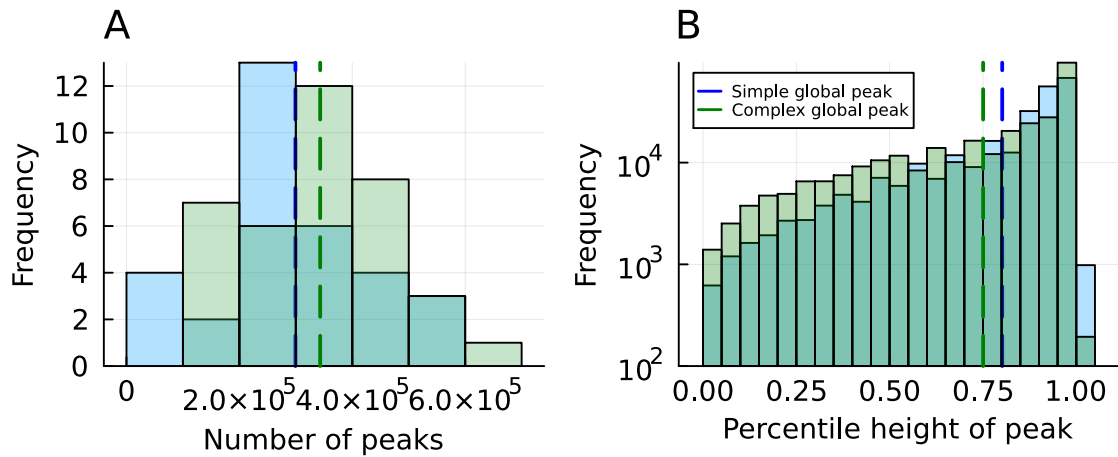


FIGURE 4.5: CORRELATED LANDSCAPES WITH COMPLEX GLOBAL PEAKS HAVE HIGHER RUGGEDNESS AND LOWER MEAN PEAK FITNESS: A. Distribution of number of peaks in 32 correlated landscapes with simple global peaks (blue) and 37 correlated landscapes with complex global peaks (green). B. Distribution of percentile peak height relative to the fitness of all phenotypes in the landscape, for a single correlated landscape with a simple global peak (blue) and a single correlated landscape with a complex global peak (green). Dotted lines show the distribution means.

have lower evolvabilities and are thus more likely to be local peaks and secondly, due to their phenotypic proximity to the global peak, they are also assigned higher fitness, which further increases the likelihood of them being peaks. This result for linearly correlated landscapes clearly demonstrates the importance of the GP map in influencing the topography of the fitness landscape. While one often imagines single peaked landscapes when thinking of linearly correlated fitness functions, the resulting landscape can in fact have hundreds of thousands of peaks, if it is the phenotype, not the genotype that linearly influences fitness. This principle is illustrated in Figure 4.6.

While we cannot make analytic predictions as we do in equation 4.1 for the correlated landscapes, we can computationally estimate the mean peak evolvability for correlated landscapes with simple global peaks, finding it to be roughly 17.4, which is slightly higher than the average evolvability of the genotype-phenotype map (16.5). On the other hand, the mean peak evolvability for correlated landscapes with complex global peaks is roughly 16.6 (see Supplementary Figure 4.3). We expect this trend because peaks comprising of simple phenotypes would have larger NCs and thus higher evolvabilities.

In Figure 4.5 B, we show the distributions of percentile heights of local peaks for a single correlated landscape with a simple global peak and a single correlated landscape with a complex global peak. The average height of local peaks is higher for the former (0.8) compared to the latter (0.75). This is also evident from the tails of the two distributions in Figure 4.5 B. We expect this trend to hold generally because as mentioned before, NCs with simple phenotypes have higher evolvabilities, which implies NCs must have

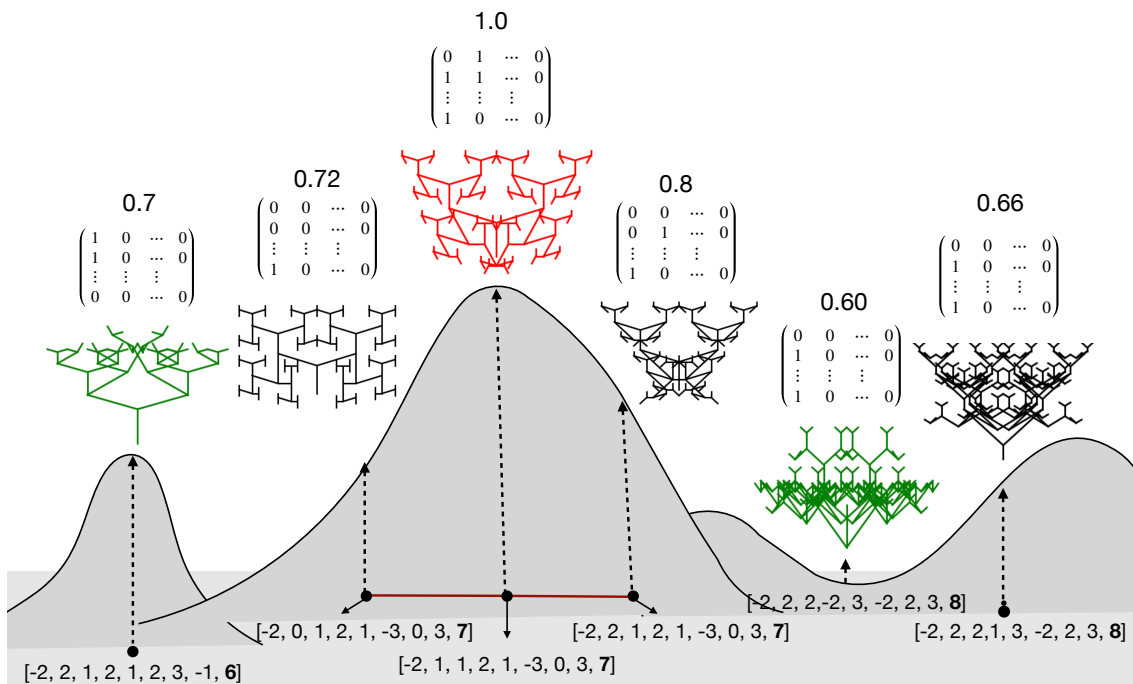


FIGURE 4.6: PEAKS CAN EXIST EVEN IN THE CORRELATED LANDSCAPE: A schematic diagram of a slice of the correlated Biomorphs fitness landscape from Figure 4.2C. The schematic shows that the structure of the GP map can make the landscape rugged, despite having a linear phenotype-fitness function. The genotypes corresponding to the phenotypes are shown below the phenotypes and neighbouring genotypes (those at a genotypic distance of 1) are connected with red lines. Neighbouring genotypes have similar looking phenotypes. The difference between the 30×30 coarse-grained arrays corresponding to the phenotype of interest and the optimal phenotype (shown in red) is used to calculate the fitness value of the phenotypes. Local peaks and valleys are shown in green.

higher fitness to be a peak. Supplementary Figure 4.4 shows that much like uncorrelated landscapes, this is indeed the case.

4.2.3 *The structure of the Biomorphs GP map does not always make the fitness landscapes navigable*

Our knowledge of local peaks helps us understand where populations might become trapped, but it gives no information on whether a global fitness optimum can still be reached without crossing a fitness valley. This aspect is quantified by the landscape navigability, $\langle \psi \rangle$ which is defined as the average probability of existence of an accessible

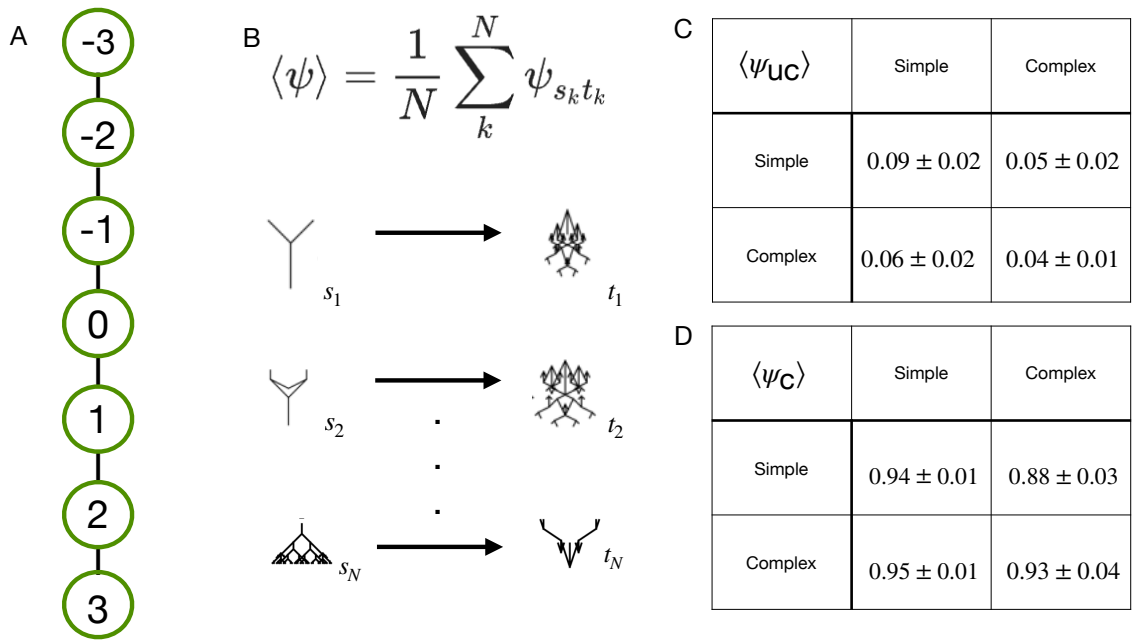


FIGURE 4.7: LANDSCAPE NAVIGABILITY: A. The allele graph for the first eight loci of the Biomorphs GP map, showing the connectivity between all possible alleles at a locus. Following [20], we restricted the alleles at each locus within a fixed range of values ($-3 \leq g_i \leq 3$ for $1 \leq i \leq 8$ and $1 \leq g_9 \leq 8$). B. Definition of landscape navigability as the average probability of having an accessible path between N source (s) and target (t) genotypes [19]. C. The navigability for uncorrelated landscapes $\langle \psi_{uc} \rangle$. D. The navigability for linearly correlated landscapes $\langle \psi_c \rangle$. The error values are the Bernoulli standard error following [19]. These numbers are based on 1000 source-target pairs per category.

path, between a given source, s_k and target, t_k genotype in the landscape ($\psi_{s_k t_k} = 1$, if an accessible path exists and $\psi_{s_k t_k} = 0$, if it does not) [19], i.e.,

$$\langle \psi \rangle = \frac{1}{N} \sum_{k=1}^N \psi_{s_k t_k} \quad (4.4)$$

Greenbury et al. [19] showed that properties of molecular GP maps such as large neutral sets lead to enhanced navigability over a simple genotype-fitness map. They found high navigability even in the worst-case scenario of completely uncorrelated phenotype-fitness maps such as that shown in Figure 4.2B. Will a similar enhanced navigability emerge in the Biomorphs GP map?

Unlike the case of the molecular GP maps, we find that the Biomorphs GP map with an uncorrelated fitness landscape is not very navigable. By calculating the navigability of 4000 source-target pairs (details in the Methods) we observe an average navigability of 0.06 ± 0.007 , which is much smaller than what we would expect purely based on the GP map's relatively high phenotypic robustness [19]. Since source and target complexity correlate with the sizes of their neutral sets and could be a confounding factor, we also

computed the navigability conditioned upon the complexity of the source and target phenotypes. The results are shown in Figure 4.7 C. We observed that navigability is low for all combinations of source-target complexity.

One reason for the uncorrelated Biomorphs fitness landscape's exceptionally low navigability is its unique allele graph, which is similar to the path graph discussed in [12, 13]. As shown in Figure 4.7 A, although there are seven alleles at each of the first eight loci and eight alleles at the ninth locus (not shown), the alleles can only mutate to neighbouring alleles. This significantly reduces the connectivity of the allele graph. The sparse connectivity of the allele graph makes biological sense, since each locus encodes a gene and it is only realistic to assume that single mutations make small changes to the gene function. The effect of the allele graph is to reduce the average number of accessible paths, by limiting the total number of paths between a given source and target at a distance l . The total number of possible paths is reduced by a factor of $\frac{1}{\prod_{i=1}^l d_i!}$, where d_i is the number of mutations that need to occur at locus i . Due to the structure of the allele graph, these d_i mutations must occur in a sequential order and cannot be permuted. Since the fitness values are randomly assigned in the uncorrelated landscape, the probability of the path being accessible is $\frac{1}{l!}$ [6] and thus, the mean number of accessible paths becomes:

$$\langle n_A \rangle = \frac{1}{l!} \frac{l!}{\prod_{i=1}^l d_i!} = \frac{1}{\prod_{i=1}^l d_i!} \ll 1 \quad (4.5)$$

This is much smaller than the mean number of accessible paths in House-of-cards landscapes with fully connected allele graphs, which is equal to one [6].

As predicted from other GP maps [19], upon introducing correlations in the phenotype-fitness function, we see a drastic increase in the navigability amongst all source target pairs belonging to various complexity classes. The average landscape navigability with the correlated fitness function approximately turns out to be 0.94. These results are presented in Figure 4.7 D.

In the next section, we study the properties of the navigable pairs that were found in both the landscapes.

Navigable source-target pairs

In order to understand if the source-target pairs, for which a navigable path exists are somewhat special, we investigate this subset in more detail.

To compute the landscape navigability of the given source target pairs, we made a conceptual abstraction that allowed us to find accessible paths and to also make analytic arguments. This involved examining the fitness landscapes at the level of a network, in which NCs are nodes and the mutational connections between them are edges. The network of NCs contains complete information about the existence of accessible paths, since any genotype in given a NC is accessible from any other genotype in that NC via neutral steps.

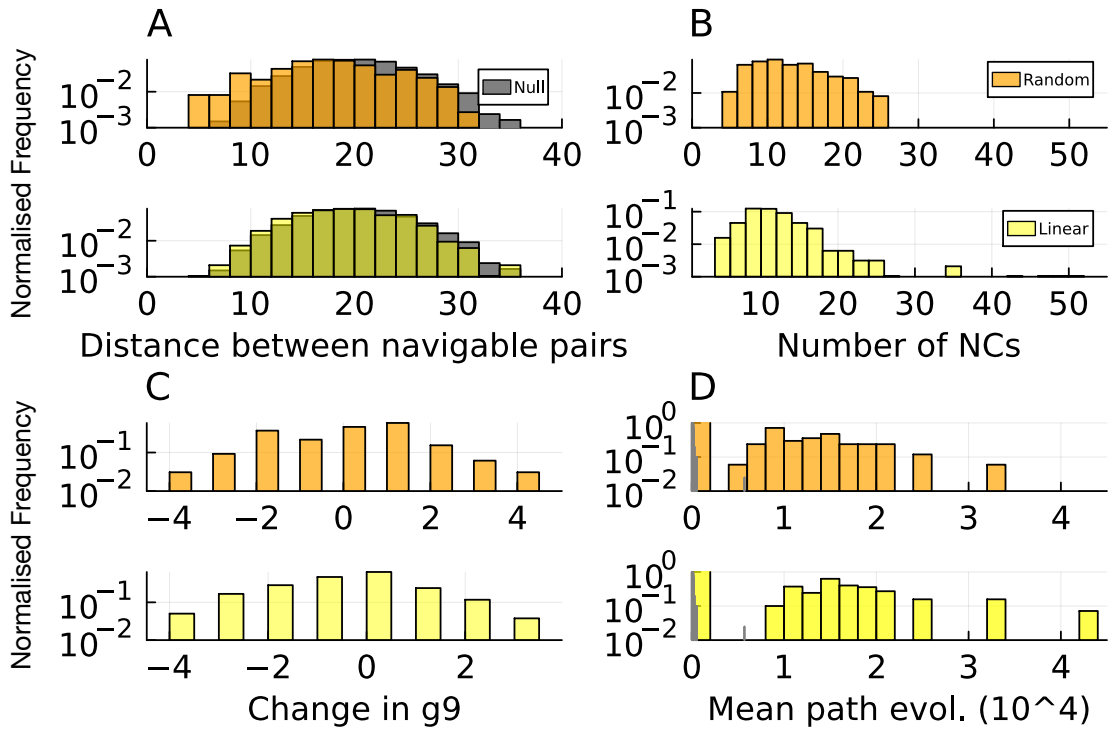


FIGURE 4.8: STATISTICS ON NAVIGABLE PATHS: A. Normalised distributions of distance between navigable pairs in the uncorrelated (orange) and linearly correlated (yellow) landscapes. The distribution of genotypic distance between random source-target pairs generated by shuffling the navigable source and target pairs is shown in grey. B. Normalised distributions of the number of neutral components (NCs) encountered along navigable paths in the uncorrelated (orange) and linearly correlated (yellow) landscapes. C. Normalised distributions of difference in value at ninth locus between navigable pairs in the uncorrelated (orange) and linearly correlated (yellow) landscapes. D. Normalised distributions of the mean path evolvability of navigable paths in the uncorrelated (orange) and linearly correlated (yellow) landscapes. The grey distribution indicates the normalised distribution of evolvability of the entire GP map. The normalisation is performed such that the area under the distribution is one. The data is based on the navigable pairs found out of the pairs used in Figure. 4.7.

In Figure 4.8, we show some statistics about the navigable pairs. First, we start by considering the genotypic distance between the two phenotypes. Since we perform our navigability test on the level of NCs instead of genotypes (see Methods), the shortest navigable path is not known and we use a proxy. We compute the genotypic distance between navigable source-target pairs by randomly sampling a genotype belonging to the each phenotype, since we start the search from a random genotype belonging to the source phenotype and end the search as soon as we encounter a genotype with the target phenotype. The average genotypic distance between navigable pairs is 17.4 in the uncorrelated landscapes and 18.9 in the linearly correlated landscapes and the

distributions look very similar to each other (see Figure 4.8 A). For reference, we also show the distribution of genotypic distance between random source-target pairs generated by shuffling the navigable source and target pairs (in grey). The average distance between a randomly chosen source-target pair is 19.7, which shows that the navigable pairs are closer than any two randomly chosen source-target pairs. The maximum genotypic distance in the Biomorphs landscape is 56.

Moreover, we could compute the average number of NCs encountered along a navigable path, which serves as an additional proxy for the length of the accessible path, and can be obtained directly from our network. The average number of NCs encountered is 12.35 in the uncorrelated landscapes and 11.35 in the linearly correlated landscapes. Thus, accessible paths traverse similar number of NCs in either landscape. However, the linearly correlated landscapes also have a much longer tailed distribution (see Figure 4.8 B).

In Figure 4.8 C we show that the distribution of the change in the value at the ninth locus (g_9), is roughly symmetric for uncorrelated landscapes, with the average change being -0.08. For the linearly correlated landscapes, the average change is -0.58. Since g_9 is a proxy for the complexity of the phenotype [20], we see that navigable paths are more likely to decrease complexity than increase it in correlated landscapes. However, this statement cannot be generalised to other GP maps since it could be an artefact of the Biomorphs GP map, wherein changing g_9 significantly changes the complexity. This could also explain why the distance between navigable pairs is shorter than average, since changing g_9 is a quick way to significantly alter the phenotype.

Finally, we also looked at the mean evolvability of the NCs encountered along a navigable path. The results are shown in Figure 4.8 D. While the average mean path evolvability for linearly correlated landscapes (9813.2) was higher than uncorrelated landscapes (8327.9), the interesting finding is that the mean path evolvability for either of the landscapes is significantly higher than the mean evolvability of the landscape (which is 16.5). This means that navigability is facilitated via high evolvability NCs. This result can be generalised to any landscape, since high-evolvability NCs have many non-neutral neighbours by definition, which means that they have a larger number of potentially uphill directions that can ultimately constitute a navigable path to the target phenotype. Therefore, landscapes with high-evolvability NCs will have high navigability. We can formalise this assertion by calculating the probability that none of the paths between a chosen source (s) and target (t) is navigable in a uncorrelated landscape i.e.

$$P(\psi_{s,t} = 0) \geq \prod_{\text{path} \in \text{paths}} P(\text{encountering a local peak on path})$$

$$\approx \prod_{\text{path} \in \text{paths}} \left[\frac{1}{\epsilon_{p1} + 1} + \frac{\epsilon_{p1}}{(\epsilon_{p1} + 1)(\epsilon_{p2} + 1)} + \dots \right]$$

where, $\text{path} = \{p_i\}$, p_i is the i th NC along the path, ϵ_{p_i} is the evolvability of that NC and the i th term in the square brackets represents the probability of encountering

a peak at the i th step along the path. If $\epsilon_{p_i} \gg 1 \forall p_i \in \text{path}$, it is easy to see that $P(\psi_{s,t} = 0) \rightarrow 0 \implies P(\psi_{s,t} = 1) \approx 1$. Moreover, the total number of paths i.e. $|\text{paths}|$, also increases multiplicatively when we encounter high evolvability NCs and this further reduces $P(\psi_{s,t} = 0)$ by increasing the number of terms in the product. Thus, if paths with high-evolvability NCs exist between the source and target, the source-target pair is very likely to be navigable.

4.2.4 Evolutionary navigability

As has been previously observed [19, 30], the mere existence of an accessible path from a source to a target does not guarantee that the path will be utilised during an evolutionary process – the path may be too long or just one of many uphill paths emanating from the source genotype. Therefore, it is important to additionally compute the evolutionary navigability – i.e. the frequency of reaching the target in an evolutionary process, given that there exists an accessible path. We will start with the strong-selection-weak-mutation regime, where the population typically occupies a single genotype at a time [31]. This regime is not only simple to analyse, it is also a regime, where fitness valleys take a long time to be crossed [32], which makes the navigability of a landscape particularly important. Rather than using a Wright-Fisher model, we represent the population by a single random walker whose steps are weighted by a fitness-dependent fixation probability [33] and we restrict the walker to neutral or fitness-increasing steps to focus on uphill paths only (see Methods).

To assess the evolutionary navigability of the Biomorphs fitness landscapes, we first simulated random adaptive walks between navigable source target pairs. For the uncorrelated landscapes, we observed zero evolutionary navigability upon simulating 100 evolutionary walks between navigable source and target pairs obtained from the analysis shown in Figure 4.7. For the correlated landscapes, we observed an average evolutionary navigability of 0.005, implying that evolutionary simulations between navigable source-target pairs arrive at the target in only 0.5% of the cases. This is in stark contrast to the landscape navigability for correlated landscapes which was around 94%. Figure 4.9 A shows these results in more detail: The evolutionary navigability is very variable between different source-target pairs: Most pairs have zero evolutionary navigability, which means that despite the existence of a fitness-increasing path, the target was not reached in any of the 100 attempts. However, for some source-target pairs, the evolutionary navigability was found to be as high as 0.5, indicating that 50 out of 100 populations reached the target. This trend is robust when we segregated the source-target pairs by their complexity, showing there is little dependence of evolutionary navigability on the complexity of the pairs (see Supplementary Figures 4.6, 4.7, 4.8 and 4.9). We also looked at the evolutionary navigability as a function of distance between the navigable source-target pair (see Figure 4.10) and saw that the evolutionary navigability decreases with increasing distance

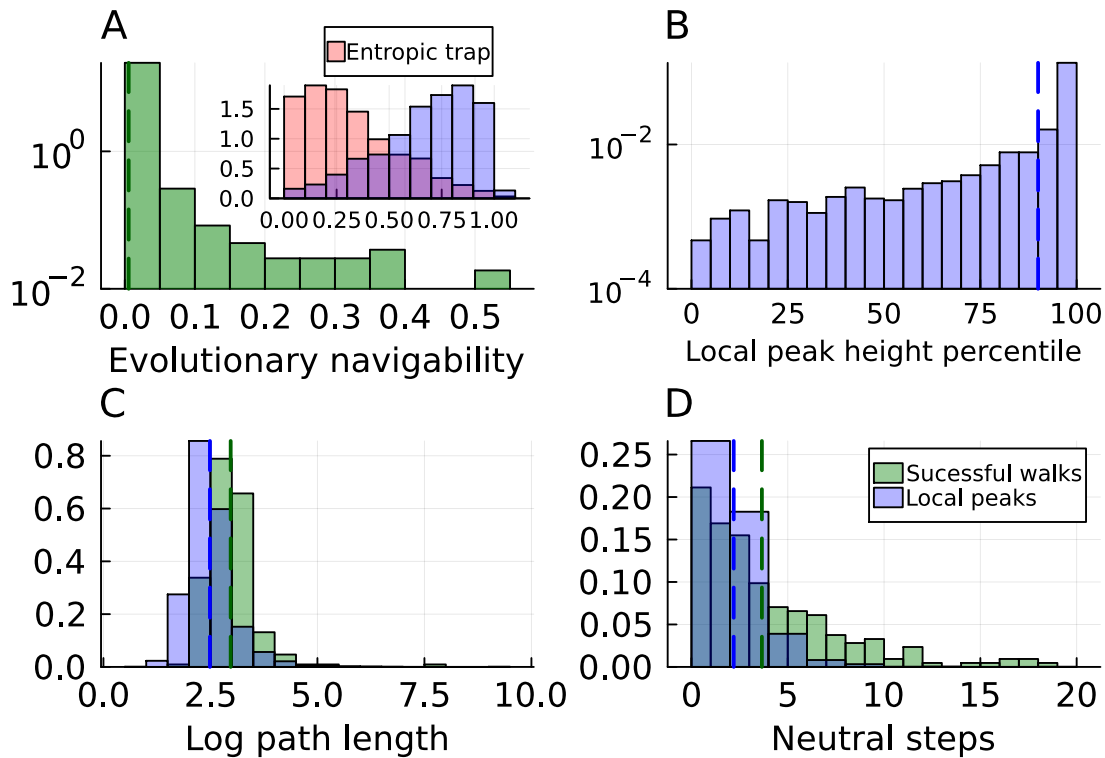


FIGURE 4.9: ARE ACCESSIBLE TARGETS IN THE CORRELATED LANDSCAPE REACHED IN THE SSWM REGIME? Histograms showing normalised distributions of A. evolutionary navigability and the causes of non-navigability (inset), B. local peak height percentile reached by unsuccessful walks that did not end up in entropic traps, C. natural logarithm of the path length of successful walks and unsuccessful walks that ended up on local peaks and D. number of neutral steps utilised in successful walks and unsuccessful walks that ended up on local peaks for linearly correlated landscapes. Distribution means are indicated by dashed vertical lines. The data is based on the navigable pairs on the correlated landscape out of the pairs used in Figure. 4.7.

between the navigable source-target pairs, highlighting the fact that long accessible paths need not be evolutionarily relevant due to their small probability of realisation.

We investigate the reasons behind the low evolutionary navigability in Figure 4.9A (inset): while on average, 32% of the walks terminated in an *entropic trap* (meaning the walk continued beyond the threshold of 10^6 steps), 68% of the walks terminated on a local peak. In Figure 4.9 B, we show the distribution of height of the local peaks reached by the unsuccessful walks in the latter case. To put the peak heights in context, we report them as a percentile of the phenotype-fitness distribution of the landscape (which depends on the target phenotype, see Supplementary Figure 4.11). The average local peak height percentile is roughly 90%, which shows that the majority of the local peaks are of high fitness.

Path statistics

The average path length of the random adaptive walks was roughly 31 steps for the successful walks and 21 steps for the unsuccessful walks terminating on local peaks, but the distributions are skewed with long tails (note that in Figure 4.9 C we show the log path length). Thus on average, the successful walks are longer than unsuccessful walks terminating on local peaks, indicating that longer paths could help in reaching the target phenotype.

Finally, in Figure 4.9 D, we show the average number of neutral steps employed. The successful walks use 3.6 neutral steps on average whereas walks terminating on local peaks use 2.2 neutral steps. However, the distributions again have long tails and some successful walks use up to 20 neutral steps. Compared to the total path length, the random adaptive walks do not utilise many neutral steps, but there is a positive correlation between the number of neutral steps used and the log path length (see Supplementary Figure 4.12).

Expanding the set of target phenotypes

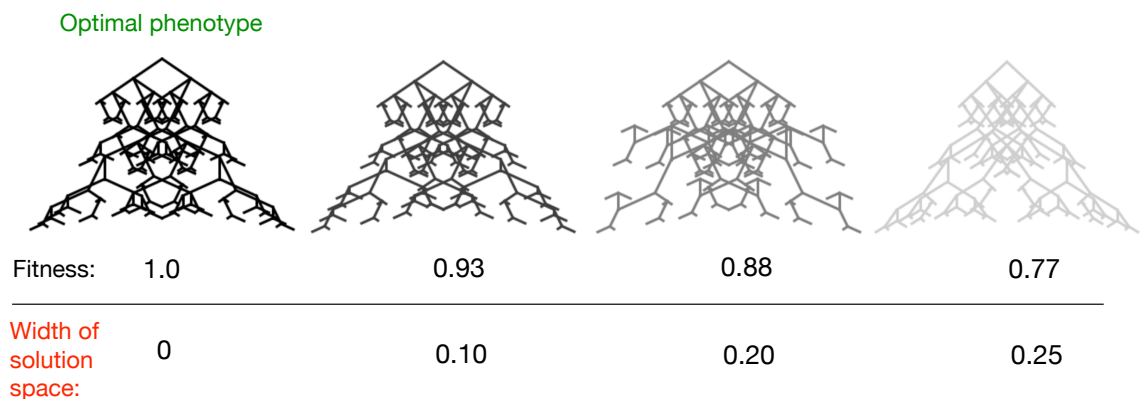


FIGURE 4.10: Optimal phenotype and example phenotypes within a broader ‘solution space’ as an example. Decreasing opacity of the phenotype shows declining fitness and increasing deviation from the optimal phenotype.

Since the goal of evolution is not to generate optimal phenotypes, but to just produce phenotypes that are good enough to ensure survival, we also considered broader ‘solution spaces’, where every walk that reaches a certain fitness threshold counts as a success, even if it never reaches the global optimum (see Figure 4.10). We investigated how this affects the evolutionary navigability and the path statistics. The results are summarised in Figure 4.11.

In Figure 4.11 A, we see that the average evolutionary navigability increases significantly with the width of the ‘solution space’. The fraction of walks terminating on local peaks and entropic traps monotonically decreases with the width of the ‘solution space’,

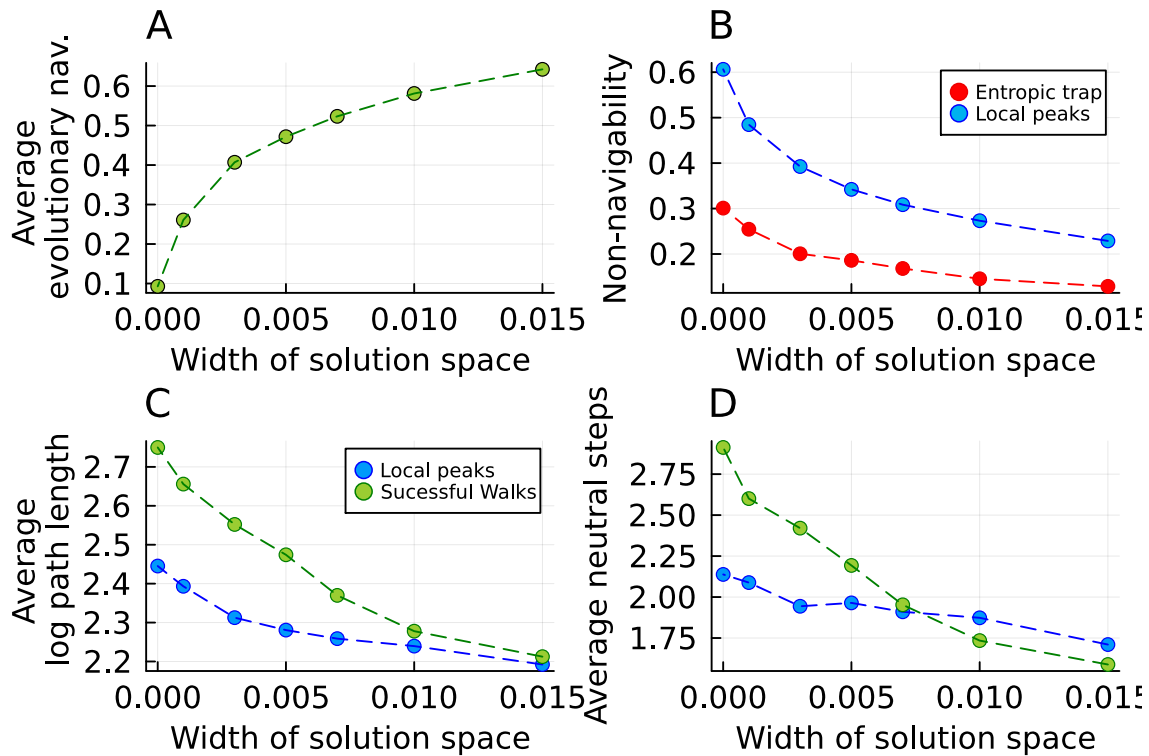


FIGURE 4.11: HAVING A WIDER ‘SOLUTION SPACE’ GIVES SHORTER AND MORE SUCCESSFUL WALKS: Plots showing the variation of A. averaged evolutionary navigability, B. the averaged causes of non-navigability, C. the averaged logarithm of the path length for successful and failed walks terminating on local peaks and D. the averaged number of neutral steps for successful and unsuccessful walks terminating on local peaks, as a function of the width of the ‘solution space’ measured as a fitness percentile for linearly correlated landscapes. Due to computational constraints, a fraction (0.15) of the navigable pairs considered in Figure 4.9 were used to generate this figure.

with the fraction at local peaks always remaining slightly higher than the fraction at entropic traps (Figure 4.11 B). The average path length of both successful walks and walks terminating on local peaks decreases with the width of the ‘solution space’ (Figure 4.11 C). Finally, the average number of neutral steps employed also decreases with the width of the ‘solution space’ (Figure 4.11 D). This result is intuitive since the addition of phenotypes to the solution space will create additional solutions closer to any initial genotype and can be reached in fewer steps.

Simulations with polymorphic populations

Finally, we looked at the evolutionary navigability with polymorphic populations. As expected, we see a much increased evolutionary navigability with an increased mutation rate, which allows the population to spread in the genotypic space. For $N = 1000$, $\mu =$

0.01 and $T = 5000$ generations, we observe an evolutionary navigability of approximately 55% for a subset of the navigable pairs considered in Figure 4.9. Upon increasing the number of generations to $T = 10000$, the evolutionary navigability rises to 61%. This increased navigability could originate for multiple reasons, including a higher potential of such populations to leave local peaks by crossing fitness valleys [32] and their faster exploration of neutral sets [34]. This however also comes at the cost of an increased mutational load due to population not being confined to the global peak.

4.3 DISCUSSION

We explored the architecture and navigability of the genotype-phenotype-fitness landscapes of the Biomorphs GP map under uncorrelated and correlated phenotype-fitness assignments. We found that both the uncorrelated and the correlated Biomorphs fitness landscapes have many local peaks, with the uncorrelated fitness landscapes having $\approx 10\times$ higher number of local peaks than the correlated landscapes. The number of local peaks in a landscape depends on the features of the GP map, such as neutrality [19] and the allele graph [12]. These effects can be encapsulated by a simple relationship, wherein the probability that a NC is a peak in the uncorrelated landscape only depends on its evolvability ϵ_p as $1/(\epsilon_p + 1)$. This implies that GP maps with many low-evolvability components are likely to have many peaks. The Biomorphs GP map, where most phenotypes have small neutral sets and low evolvability [20], falls into this category. This dependence on one particular property of a GP map may explain why the prevalence of peaks was found to be highly dependent on the GP map in a previous work [19]. Note that it is not the phenotype evolvability that matters, but that of a mutationally connected part of a phenotype's neutral set, which can be much smaller [18]. Testing these links between evolvability and the prevalence of fitness peaks for other GP maps is a topic for future work.

The importance of evolvability for landscape ruggedness prompts the question of what determines evolvability in GP maps. In this regard, it has been shown that high-evolvability components, where the entire component has higher evolvability than individual genotypes, requires a specific form of epistasis [35]. Intuitively, the argument is as follows: if the joint evolvability of two genotypes (A and B) is higher than their individual evolvabilities, then they must have distinct phenotypes in their respective mutational neighbourhoods. This is only possible if at least one mutation in genotype A generates a phenotype that cannot be reached at all from genotype B , which means that the effect of a mutation must depend on the genotype to which it is applied, indicating the presence of epistasis. Contrary to the intuition that epistasis always increases ruggedness [28, 36, 37], this type of epistasis can make genotype-phenotype-fitness landscapes less rugged, by increasing the evolvability.

Moreover, we also establish an analytical relationship between the height and the evolvability of peaks in the uncorrelated landscape and showed that peaks with higher evolvability must be fitter on average. In the Biomorphs GP map, the effect size is small since peak fitness scales as $(\epsilon_p + 1)/(\epsilon_p + 2)$ when fitness values are sampled the uniform distribution, and typical evolvabilities are $\gtrsim 10$. However, the range of evolvabilities may be broader in other maps and this would have evolutionary implications: since low-evolvability NCs are typically small [18] and thus less likely to be reached in evolutionary processes [34], this would imply that higher peaks are easier to reach. Such a positive relationship between peak height and navigability would be consistent with other examples of fitness landscapes [38], even though the details and underlying principles may be different.

The analytic results so far have been for the uncorrelated fitness landscape, but even the correlated landscapes have hundreds of thousands of local peaks. The ruggedness in the correlated fitness landscapes is due to the structure of the underlying GP map and this is perhaps the first example of a GP map that exhibits rugged multi-peaked structure despite the presence of a phenotype-fitness map that is linear in the distance to the target phenotype.

Next, we turned to the navigability of the Biomorphs landscape. We found low navigability in the uncorrelated case. This is consistent with recent work showing that the path graph allele graph reduces the probability of having accessible paths in uncorrelated fitness landscapes [12]. We find that the average genotypic distance between navigable pairs is shorter than the average distance between two randomly chosen genotypes, indicating that navigable pairs tend to be closer in genotypic space. We also found an analytic expression for the navigability of a source-target pair based on a simpler abstraction of the GP map: a graph that represents each NC as a node and mutational connections as edges. High-evolvability components play the role of high-degree nodes in this network, and thus promote navigability. In agreement with this theoretical argument, we found that navigable paths pass through phenotypes with high evolvability and moreover that GP maps with high evolvability phenotypes will be highly navigable.

To study whether an existing navigable path to a given target is actually adopted during evolutionary simulations, we computed the evolutionary navigability of the correlated landscape in the strong selection weak mutation regime. Surprisingly, the evolutionary navigability turned out to be lower than 1%, meaning that simulations did not reach the target despite the existence of an accessible path. This is in part due to the evolutionary walks terminating at local peaks and partly due to the lack of convergence after 10^6 steps. The successful walks were longer and utilised more neutral steps than the walks terminating at local peaks. Further, the local peaks themselves had a fitness percentile of approximately 90%. This indicates that the correlated landscapes belong to the class of rugged yet navigable landscapes [38, 39] that are characterised by multiple local peaks of decently high fitness. Thus, we revisited the setup, where only a single

global optimum was counted as a ‘successful’ outcome, and instead allowed for some variability in the target fitness by tolerating a few mismatches to the target phenotype. As expected, doing so greatly increased the evolutionary navigability, while decreasing the average path length and the number of neutral steps utilised. It is important to note that downhill steps were forbidden in these simulations, and it remains to be seen if some peaks can be escaped if this restriction is relaxed.

Finally, we also computed the evolutionary navigability under polymorphic population dynamics and we saw that having $NL\mu \gg 1$ significantly increases the evolutionary navigability between navigable pairs and this also increases with increasing the number of generations we simulate. This may be due to the higher number of genotypes explored, or the higher likelihood of crossing fitness valleys, which was not allowed in our random walk simulations.

Overall, we have shown, how a direct link can be drawn from GP map properties, like evolvabilities, to the navigability of a landscape, and then to the dynamics of complex multi-step adaptive processes. Future work, should apply these calculations to other GP maps, to find out whether the high number of local peaks is a feature only of the Biomorphs model or applies more broadly.

The analysis in this paper can be extended in various directions. It remains to be seen how the landscape and evolutionary navigability results change when the correlated phenotype-fitness map is modified for instance, to a Gaussian phenotype-fitness map, or to a phenotype-fitness map in which certain features of the phenotype, such as the length of the branches are more important than others and are given more weight in the fitness function. Moreover, the relationships between ruggedness, peak height, navigability and evolvability need to be made more concrete for the correlated Biomorphs fitness landscapes. Further, the recent work on the likelihood of different phenotypic transitions, depending on their biased appearance [34] and non-Poissonian bursts [40], could be used to study the adaptive trajectories. Finally, the role of neutral steps in enhancing the navigability is still not fully understood. Whereas they have been invoked to explain enhanced evolvability, here we find scant utilisation in the successful random adaptive walks.

4.4 METHODS

Generating the GP map

Following [20], we included all genotypes within a fixed range of values for each locus ($-3 \leq g_i \leq 3$ for $1 \leq i \leq 8$ and $1 \leq g_9 \leq 8$) and generated their computationally predicted phenotypes. This computational treatment uses some coarse-graining, to ensure that only visually clearly distinct Biomorphs are classed as different phenotypes. This is achieved by projecting each 2D image onto a 30×30 pixel grid. Then coinciding line

segments were merged, the phenotype was placed on the grid and scaled and centered. Next the lines contained within each pixel were recorded and then coarse-grained based on a threshold. The qualitative characteristics of the GP map were found to be robust to changes in the dimensions of the pixel grid (i.e., 30) and the coarse-graining threshold (i.e., 20%) [20].

Generating the fitness landscapes

- *Uncorrelated fitness landscapes*

The uncorrelated GPF landscapes, which were first used for the RNA GP map [19], were generated by simply assigning a random number sampled from a uniform distribution between 0 and 1 to each phenotype. The target phenotype however was always assigned a fitness value of one, whereas the source phenotype fitness was variable and randomly sampled like the other phenotypes.

- *Correlated fitness landscapes*

Following work on RNA [19], we also created a fitness landscape based on phenotypic distances to a target. For the Biomorphs, we defined the *phenotypic similarity* between any two phenotypes as the normalised Hamming distance between the coarse grained matrices encoding the phenotypes. The linearly correlated fitness landscapes were then defined, such that fitness values $F(p)$ depended linearly on the phenotypic similarity between the phenotype in question (p) and the optimal phenotype (p^{opt}), i.e.,

$$F(p) = 1 - \frac{1}{N^2} \sum_{ij} (1 - \delta_{P_{ij}, P_{ij}^{\text{opt}}})$$

where, the phenotypes p and p_{opt} are encoded as $N \times N$ coarse-grained matrices P and P_{opt} (here $N = 30$) and $\delta_{P_{ij}, P_{ij}^{\text{opt}}}$ is the Kronecker delta function, which equals 1 when $P_{ij} = P_{ij}^{\text{opt}}$ and 0 when $P_{ij} \neq P_{ij}^{\text{opt}}$. Thus, our fitness function penalises deviations from the optimal phenotype in its coarse-grained representation, and is bounded between $0 \leq F(p) \leq 1$.

Analytic estimate of number of peaks

The analytical model of the Biomorphs GP map gives a parametric expression for $\epsilon_p = 2 \cdot (9 - n_u(g_9)) - 1 + 7^{n_u(g_9) - n_u(g_9+1)}$, where $n_u(g_9)$ is the number of unconstrained positions in the genotype, i.e. the positions that do not alter the phenotype upon being mutated and it is a function of g_9 , the value at the ninth locus [20]. Then the expected number of peaks can be estimated as:

$$\langle \text{peaks} \rangle \approx \sum_{g_9} N(g_9) \cdot \frac{1}{\epsilon_p(g_9) + 1}$$

where $N(g_9)$ is the number of NCs with a fixed g_9 , which is twice the number of phenotypes with a fixed g_9 due to axial symmetry, i.e. $N(g_9) = 2 \cdot 7^8 / n_p(g_9)$, with $n_p(g_9) = 2 \cdot 7^{n_u(g_9)}$ being the size of a neutral set for a given g_9 .

Computational enumeration of the peaks

For an estimate of the number of peaks in the various fitness landscapes, we first computed the neutral components (NCs) in the GP map – i.e. connected networks of genotypes that map to the same phenotype. We then used this information to look at the neighbourhood of each NC and checked if the given NC had a higher fitness than all its neighbouring NCs. If true, we added the given NC to the list of peaks.

Analytic derivation: distribution of the height of peaks with evolvability ϵ_p

For a NC with evolvability ϵ_p to be a peak in an uncorrelated fitness landscape, it must have the highest fitness amongst $\epsilon_p + 1$ fitness values. Since the fitness values are drawn from a uniform distribution $U(0, 1)$, the cumulative distribution function of the largest of $\epsilon_p + 1$ i.i.d uniform random variables $f_1, f_2 \dots f_{max}$ is given by:

$$P_{f_{max}}(x) = P(f_{max} \leq x)$$

Since $f_1 < f_2 < \dots < f_{max}$,

$$= P(f_1 \leq x) \cdot P(f_2 \leq x) \dots P(f_{max} \leq x) = P_f(x)^{\epsilon_p + 1}$$

because f_i s are i.i.d. random variables.

$$\implies p_{f_{max}}(x) = (\epsilon_p + 1) \cdot P_f(x)^{\epsilon_p} \cdot p_f(x)$$

where $p_{f_{max}}(x)$ is the probability density function. Since f_i s are uniform random variables,

$$p_{f_{max}}(x) = (\epsilon_p + 1) \cdot x^{\epsilon_p} \implies \langle f_{max} \rangle = \frac{\epsilon_p + 1}{\epsilon_p + 2}$$

Complexity of phenotypes

To estimate the descriptive complexity of phenotypes in this coarse-grained representation, we employed the block decomposition method as described in [41], which is tailored for 2D binary arrays. Due to the axial symmetry of all Biomorphs, we analyzed only one

half of each phenotype. The block decomposition method was applied with its default parameters, except for the boundary conditions. Instead of the default setting, which disregards boundary pixels in complexity calculations, we opted for the sliding window approach. See [20] for details.

Landscape Navigability

Following Greenbury et al. [19], we define landscape navigability as the average probability that a randomly chosen phenotype pair have at least one accessible path (i.e. an uphill path) between them.

- *Method 1*

We use a breadth-first search algorithm in genotype space from Greenbury et al. [19]. This algorithm has three possible outcomes: If a navigable path is identified, or all branches of the search process terminate without hitting the target, we know if a navigable path exists or not. However, when a computational threshold T is reached, the process is terminated without a certain answer. Thus, we do not use this method in the main text, but we do use it in the SI (see Supplementary Figure 4.5) since it can return the genotypes on the navigable path if the target is reached.

- *Method 2*

A large fraction of our navigability searches got terminated because they reached the termination threshold T in Method 1. Since we did not know whether the terminated searches would have resulted in navigability, our landscape navigability results could potentially be biased. To deal with this issue, we decided to zoom out and simplify the computations by looking at NCs (defined above) and not at individual genotypes. By definition, it is possible to neutrally mutate between any two genotypes belonging to a NC, so if one genotype on the NC is accessible from a source, all genotypes on that NC are accessible. Thus, we treated each NC as a single entity. We defined a graph, where every node is a NC and every edge connects two NCs that have a mutational connection. Then, we used a breadth-first-search algorithm on this graph to test for the existence of accessible paths. As an initial condition, we randomly sampled a genotype belonging to our source phenotype and then identified which NC it belonged to. This computational simplification helped us get rid of the termination threshold (T).

We used method 2 to compute all the results in the main text and method 1 to compute the results in the supplementary. Both the methods produced identical results in examples where method 1 gave a clear outcome (navigable or not, rather than terminated).

Source-target pairs

For the complexity specific landscape navigability computations, we randomly sampled the top 0.1% and bottom 0.1% of phenotypes ordered according to their complexity. For each source-target class, we sampled 1000 source-target pairs. These were used to generate the data in Figure 4.7. Out of these pairs, we selected those pairs that had a navigable path between them. These navigable pairs were then used to produce Figures 4.8, 4.9 and 4.10.

Monomorphic simulations of evolving populations

To simulate monomorphic evolutionary dynamics in the strong-selection weak mutation regime (SSWM) [33], we simulated a random walker in the sequential fixation model [19, 31]. At each step, we picked a mutational neighbour with higher fitness with a probability equal to its haploid Kimura fixation probability [42]. This genotype became the next monomorphic genotype of the population. We continued this process until a peak was reached or the termination threshold was reached (here, $T = 10^6$). For the results in the text, we used a population size of 100, but the results hold qualitatively even with other population sizes such as $N = 10$ (see Supplementary Figure 4.13).

Polymorphic simulations of evolving populations

We used the Wright-Fisher dynamics to simulate the evolution of populations in the polymorphic regime. The population sizes and mutation rates used are specified in the text.

ACKNOWLEDGEMENTS

MS was supported by the Swiss National Science Foundation (Grant number PP00P3_170604 and the PhD mobility grant). NM acknowledges support of Issachar Fund, the Spanish Ministry of Science and Innovation through the Centro de Excelencia Severo Ochoa (CEX2020-001049-S, MCIN/AEI /10.13039/501100011033) and to the EMBL partnership, of the Generalitat de Catalunya through the CERCA programme, and Grant JDC2022-049526-I funded by MCIN/AEI/10.13039/501100011033 and by European Union NextGenerationEU/PRTR. We thank members of the Louis research group at Oxford for discussions.

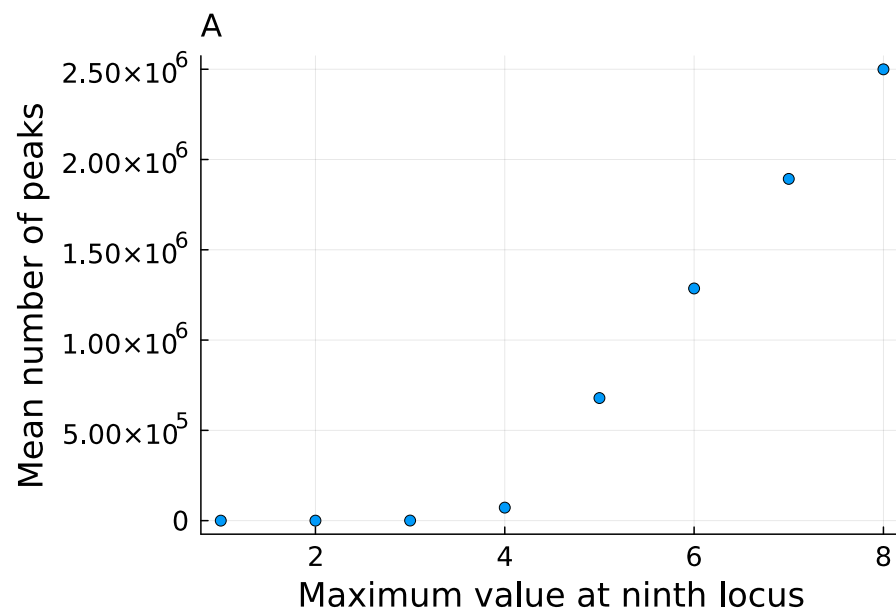
REFERENCES

1. Wright, S. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. *Proceedings of the 6th International Congress of Genetics*, 356 (1932).
2. Kauffman, S. & Levin, S. Towards a general theory of adaptive walks on rugged landscapes. *Journal of Theoretical Biology*. **128**, 11 (1987).
3. Weinreich, D., Watson, R. & Chao, L. Perspective: sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* **59**, 1165 (2005).
4. 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 (2006).
5. De Visser, J. A. G. M. & Krug, J. Empirical fitness landscapes and the predictability of evolution. *Nature Reviews Genetics* **15**, 480 (2014).
6. Franke, J., Klozer, A., de Visser, J. & Krug, J. Evolutionary accessibility of mutational pathways. *PLoS Computational Biology* **7** (2011).
7. Kingman, J. On the properties of bilinear models for the balance between genetic mutation and selection. *Mathematical Proceedings of the Cambridge Philosophical Society* **81**, 443 (1977).
8. Kingman, J. A simple model for the balance between selection and mutation. *J. Appl. Probab.* **15**, 1 (1978).
9. Aita Takuyo and Husimib, Y. Fitness Spectrum Among Random Mutants on Mt. Fuji Type Fitness Landscape. *Journal of Theoretical Biology* **182**, 469 (1996).
10. Kauffman, S. A. & Weinberger, E. D. The NK model of rugged fitness landscapes and its application to the maturation of the immune response. *Journal of Theoretical Biology* **141**, 211 (1989).
11. Hwang, S., Schmiegelt, B., Ferretti, L. & Krug, J. Universality Classes of Interaction Structures for NK Fitness Landscapes. *Journal of Statistical Physics* **172**, 226 (2018).
12. Schmiegelt, B. & Krug, J. Accessibility Percolation on Cartesian Power Graphs. *Journal of Mathematical Biology* **86**, 46 (2023).
13. Krug, J. & Oros, D. Evolutionary accessibility of random and structured fitness landscapes. *Journal of Statistical Mechanics: Theory and Experiment* **2024**, 034003 (2024).
14. Ahnert, S. E. Structural properties of genotype-phenotype maps. *Journal of The Royal Society Interface* **132** (2017).
15. Greenbury, S. F. & Ahnert, S. E. The organization of biological sequences into constrained and unconstrained parts determines fundamental properties of genotype-phenotype maps. *Journal of The Royal Society Interface* **12**, 279 (2015).

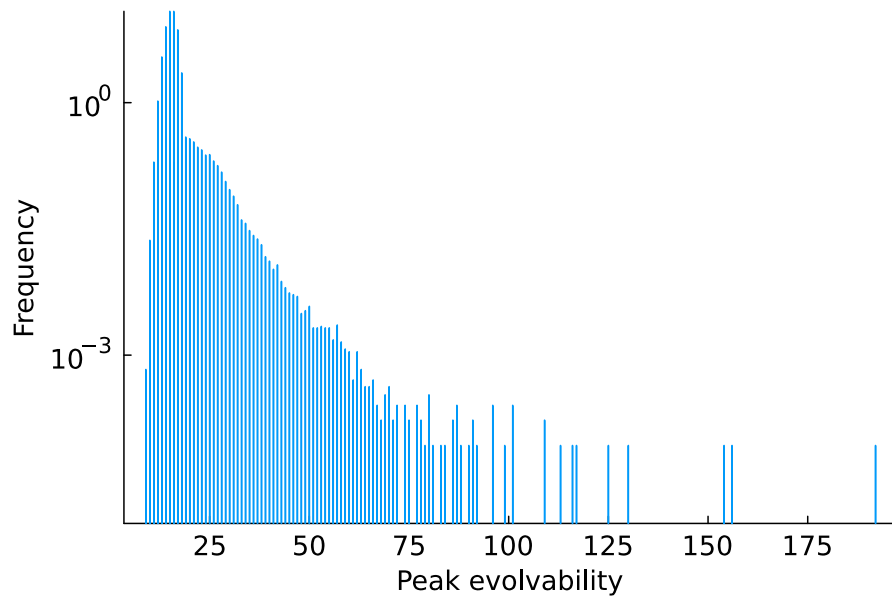
16. Manrubia, S., Cuesta, J. A., Aguirre, J., Ahnert, S. E., Altenberg, L., Cano, A. V., Catalán, P., Diaz-Uriarte, R., Elena, S. F., García-Martín, J. A., Hogeweg, P., Khatri, B. S., Krug, J., Louis, A. A., Martin, N. S., Payne, J. L., Tarnowski, M. J. & Weiß, M. From genotypes to organisms: State-of-the-art and perspectives of a cornerstone in evolutionary dynamics. *Physics of Life Reviews* **38**, 55 (2021).
17. Greenbury, S. F., Schaper, S., Ahnert, S. E. & Louis, A. A. Genetic Correlations Greatly Increase Mutational Robustness and Can Both Reduce and Enhance Evolvability. *PLoS Computational Biology* **12**(3) (2016).
18. Schaper, S., Johnston, I. G. & Louis, A. A. Epistasis can lead to fragmented neutral spaces and contingency in evolution. *Proceedings of the Royal Society B: Biological Sciences* **279**, 1777 (2012).
19. Greenbury, S., Louis, A. & Ahnert, S. The structure of genotype-phenotype maps makes fitness landscapes navigable. *Nature Ecology & Evolution* **6**, 1742 (2022).
20. Martin, N. S., Camargo, C. Q. & Louis, A. A. Bias in the arrival of variation can dominate over natural selection in Richard Dawkins' biomorphs. *PLoS Computational Biology* (2024).
21. Dawkins, R. *The Blind Watchmaker* (1986).
22. Dawkins, R. The Evolution of Evolvability. *Artificial Life. The Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems* (2003).
23. Darwin, C. *On the Origin of Species by Means of Natural Selection* (London, 1859).
24. Wagner, A. Robustness and evolvability: A paradox resolved. *Proceedings of the Royal Society B: Biological Sciences* **275**, 91 (1630 2008).
25. Dingle, K., Camargo, C. & Louis, A. Nature Communications. *Molecular Biology and Evolution* **9** (2018).
26. IG, J., K, D., SF Greenbury CQ, C., JPK, D., SE, A. & AA, L. Symmetry and simplicity spontaneously emerge from the algorithmic nature of evolution. *Proceedings of the National Academy of Sciences* **119** (2022).
27. Srivastava, M. & Payne, J. L. On the incongruence of genotype-phenotype and fitness landscapes. *PLoS Computational Biology* **18**(9) (2022).
28. Aguilar-Rodríguez, J., Payne, J. & Wagner, A. A thousand empirical adaptive landscapes and their navigability. *Nature Ecology and Evolution* **1** (2017).
29. Macken, C. A. & Perelson, A. S. Protein evolution on rugged landscapes. *Proceedings of the National Academy of the Sciences* **86**, 6191 (1989).
30. Srivastava, M., Rozhoňová, H. & Payne, J. L. Alphabet cardinality and adaptive evolution. *Journal of Physics A: Mathematical and Theoretical* **56**, 455601 (45 2023).
31. McCandlish, D. M. & Stoltzfus, A. Modeling evolution using the probability of fixation: history and implications. *The Quarterly review of biology* **89**, 225 (2014).

32. Gokhale, C. S., Iwasa, Y., Nowak, M. A. & Traulsen, A. The pace of evolution across fitness valleys. *Journal of Theoretical Biology* **259**, 613 (2009).
33. Gillespie, J. H. A simple stochastic gene substitution model. *Theoretical Population Biology* **23**, 202 (1983).
34. Schaper, S. & Louis, A. A. The Arrival of the Frequent: How Bias in Genotype-Phenotype Maps Can Steer Populations to Local Optima. *PLoS One* **9**, e86635 (2014).
35. Weiß, M. & Ahnert, S. E. Phenotypes can be robust and evolvable if mutations have non-local effects on sequence constraints. *Journal of The Royal Society Interface* **15**, 20170618 (2018).
36. Poelwijk, F., Tănase-Nicola, S., Kiviet, D. & Tans, S. Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. *Journal of Theoretical Biology* **272**, 141 (2011).
37. Bank, C. Epistasis and Adaptation on Fitness Landscapes. *Annual Review of Ecology, Evolution, and Systematics* **53**, 457 (2022).
38. Papkou, A., Garcia-Pastor, L., Escudero, J. A. & Wagner, A. A rugged yet easily navigable fitness landscape. *Science* **382**, eadh3860 (2023).
39. Das, S. G., Direito, S. O., Waclaw, B., J Allen, R. & Krug, J. Predictable properties of fitness landscapes induced by adaptational tradeoffs. *eLife*, e55155 (2020).
40. Martin, N. S., Schaper, S., Camargo, C. Q. & Louis, A. A. Non-Poissonian bursts in the arrival of phenotypic variation can strongly affect the dynamics of adaptation. *Molecular Biology and Evolution* (2024).
41. Zenil, H., Hernández-Orozco, S., Kiani, N., Soler-Toscano, F., Rueda-Toicen, A. & Tegnér, J. A. Decomposition Method for Global Evaluation of Shannon Entropy and Local Estimations of Algorithmic Complexity. *Entropy* **20** (2018).
42. Kimura, M. On the probability of fixation of mutant genes in a population. *Genetics* **47**, 713 (1962).

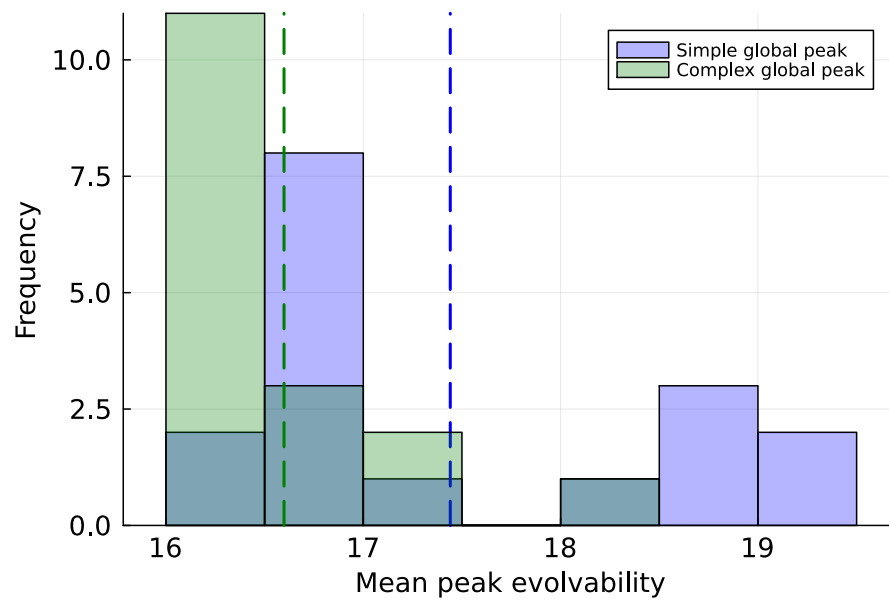
4.5 SUPPLEMENTARY INFORMATION



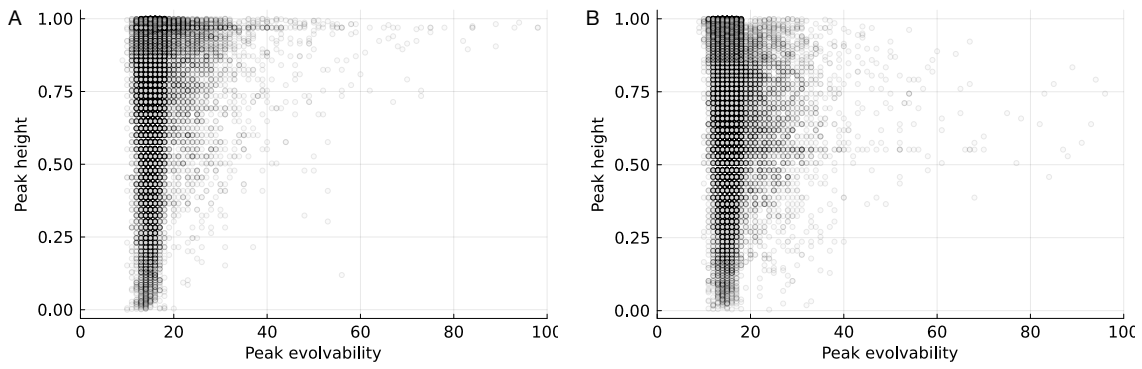
SUPPLEMENTARY FIGURE 4.1: Mean number of peaks in the uncorrelated fitness landscape of the analytically tractable Biomorphs GP map as a function of maximum g_9 , which is a proxy for complexity of the phenotype.



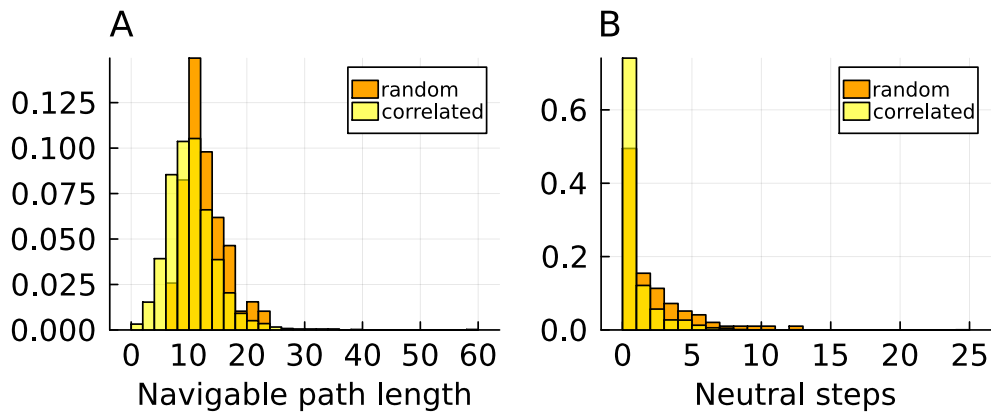
SUPPLEMENTARY FIGURE 4.2: The distribution of evolvability of the peaks in one example of an uncorrelated fitness landscape.



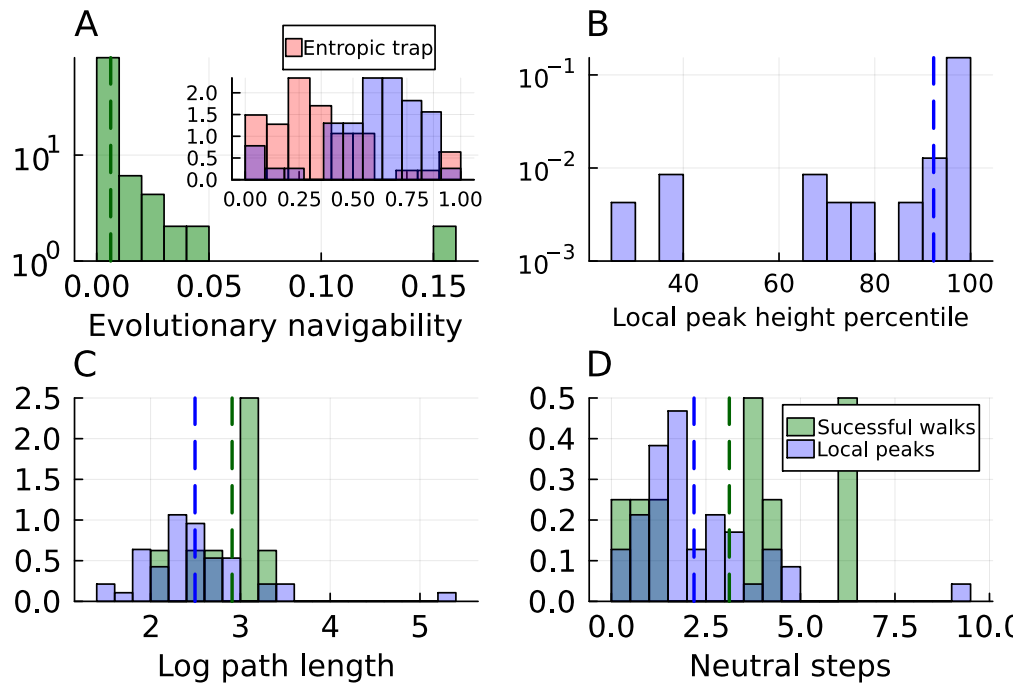
SUPPLEMENTARY FIGURE 4.3: The distribution of mean peak evolvability of the peaks in correlated fitness landscapes with simple (blue) and complex (green) global peaks.



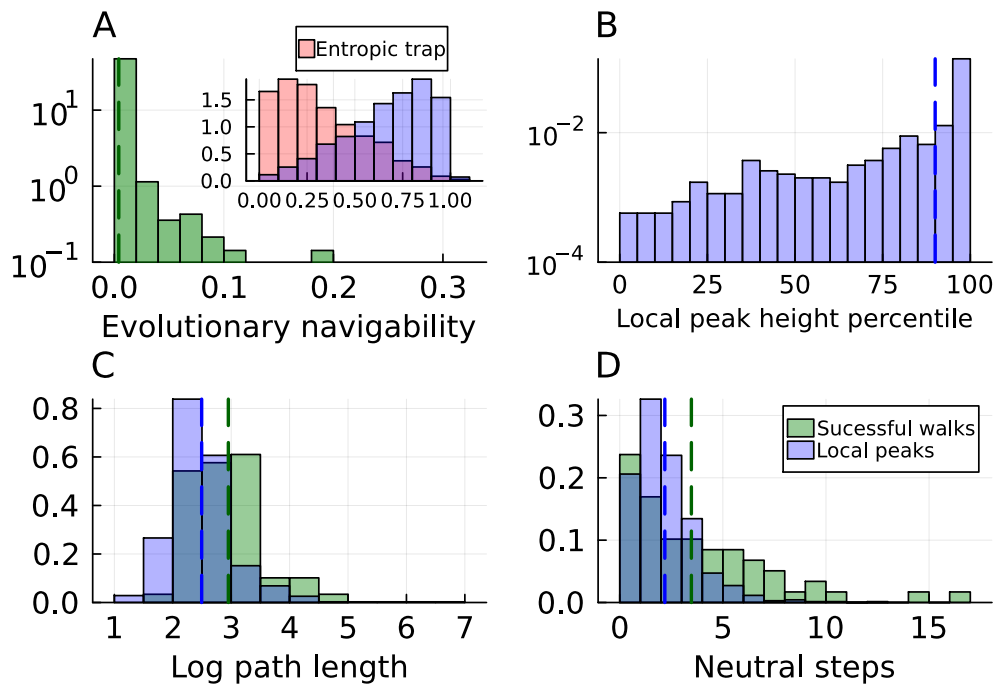
SUPPLEMENTARY FIGURE 4.4: The variation in peak height as a function of peak evolvability for a correlated landscape with A. a simple global peak and B. a complex global peak.



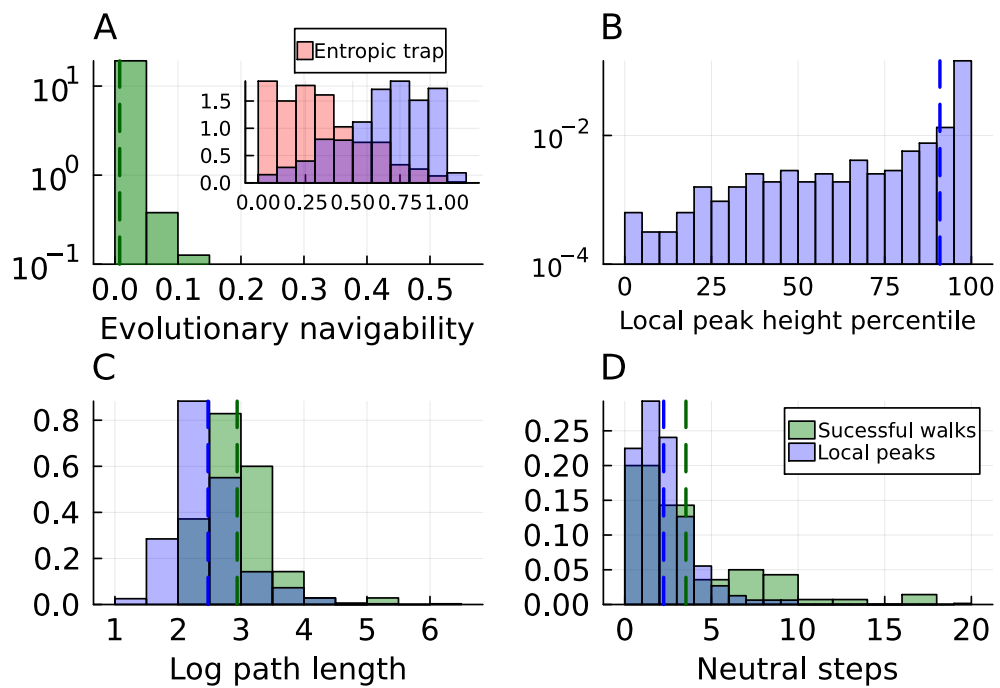
SUPPLEMENTARY FIGURE 4.5: Distributions of (A) path length and (B) number of neutral steps between navigable pairs found on random and correlated fitness landscapes.



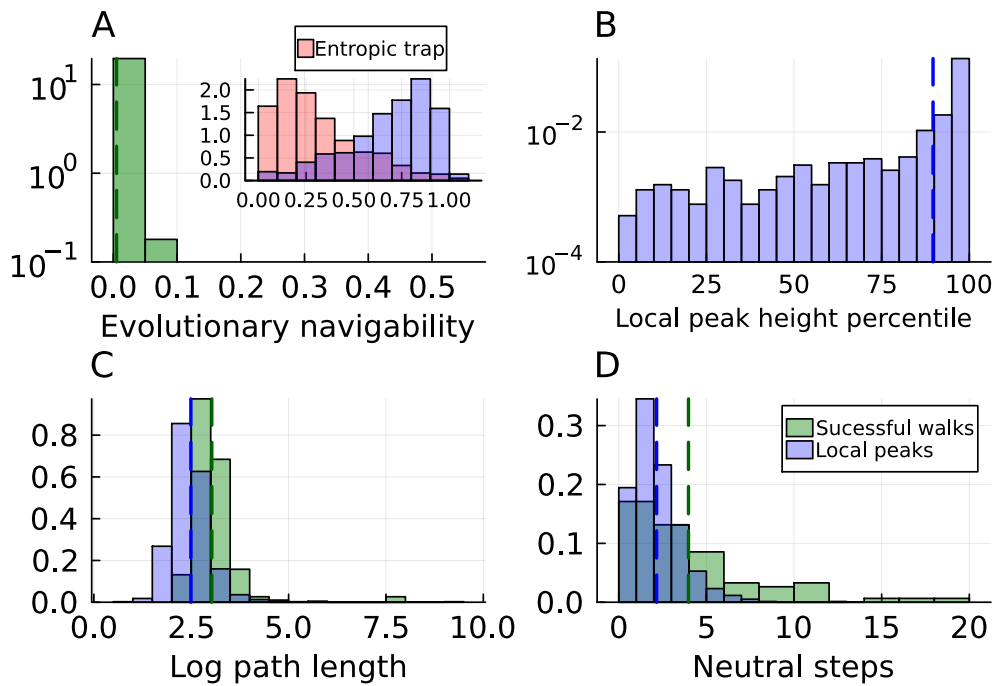
SUPPLEMENTARY FIGURE 4.6: Complex source, complex target: Histograms showing distributions of A. Evolutionary navigability and the causes of non-navigability (inset), B. Local peak height reached by failed walks, C. Logarithm of the path length for successful and failed walks that ended up on local peaks and D. Neutral steps for successful and failed walks that ended up on local peaks for linearly correlated landscapes. Distribution means are indicated by dashed vertical lines.



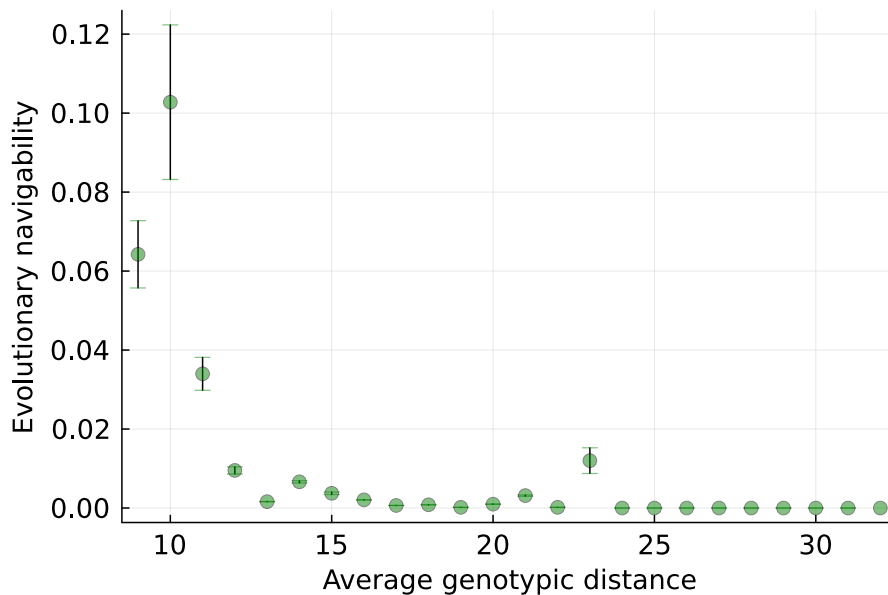
SUPPLEMENTARY FIGURE 4.7: Complex source, simple target: showing distributions of A. evolutionary navigability and the causes of non-navigability (inset), B. local peak height reached by failed walks, C. logarithm of the path length for successful and failed walks that ended up on local peaks and D. neutral steps for successful and failed walks that ended up on local peaks for linearly correlated landscapes. Distribution means are indicated by dashed vertical lines.



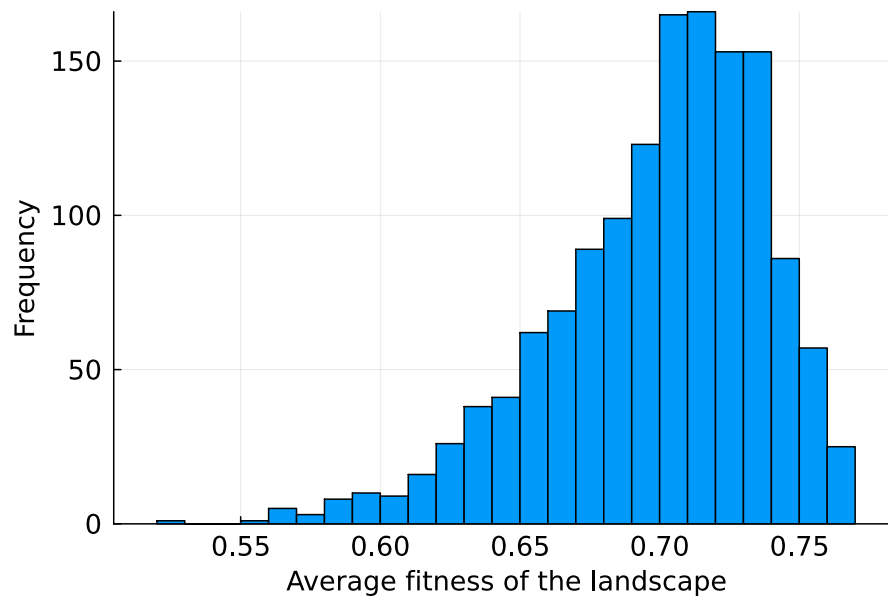
SUPPLEMENTARY FIGURE 4.8: Simple source, simple target: Histograms showing distributions of A. evolutionary navigability and the causes of non-navigability (inset), B. local peak height reached by failed walks, C. logarithm of the path length for successful and failed walks that ended up on local peaks and D. neutral steps for successful and failed walks that ended up on local peaks for linearly correlated landscapes. Distribution means are indicated by dashed vertical lines.



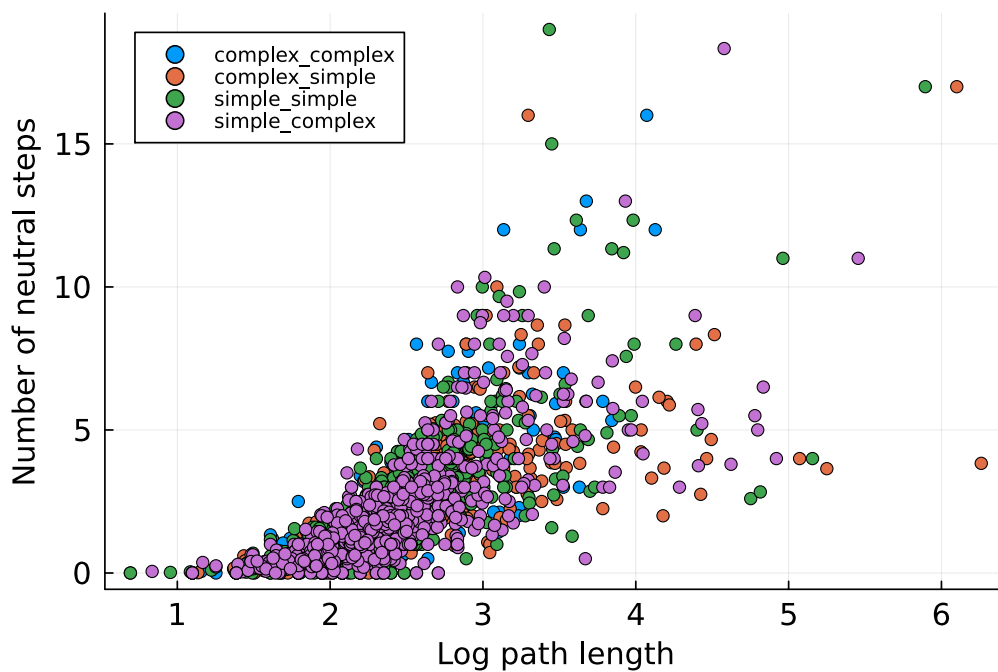
SUPPLEMENTARY FIGURE 4.9: Simple source, complex target: Histograms showing distributions of A. evolutionary navigability and the causes of non-navigability (inset), B. local peak height reached by failed walks, C. logarithm of the path length for successful and failed walks that ended up on local peaks and D. neutral steps for successful and failed walks that ended up on local peaks for linearly correlated landscapes. Distribution means are indicated by dashed vertical lines.



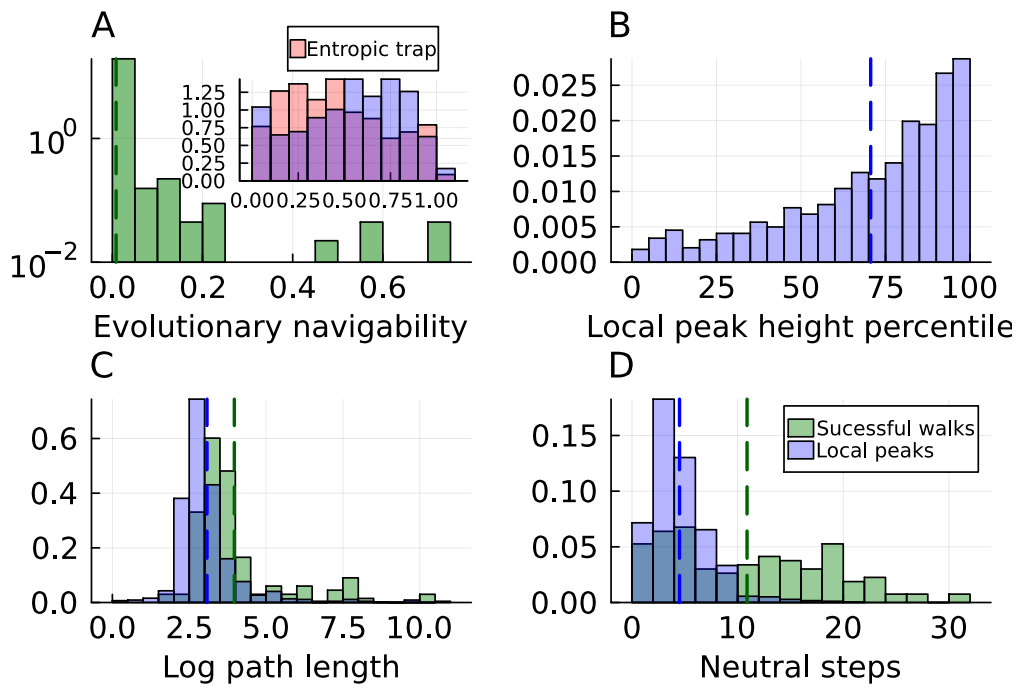
SUPPLEMENTARY FIGURE 4.10: Dependence of evolutionary navigability on the distance between navigable source-target pairs for correlated fitness landscapes.



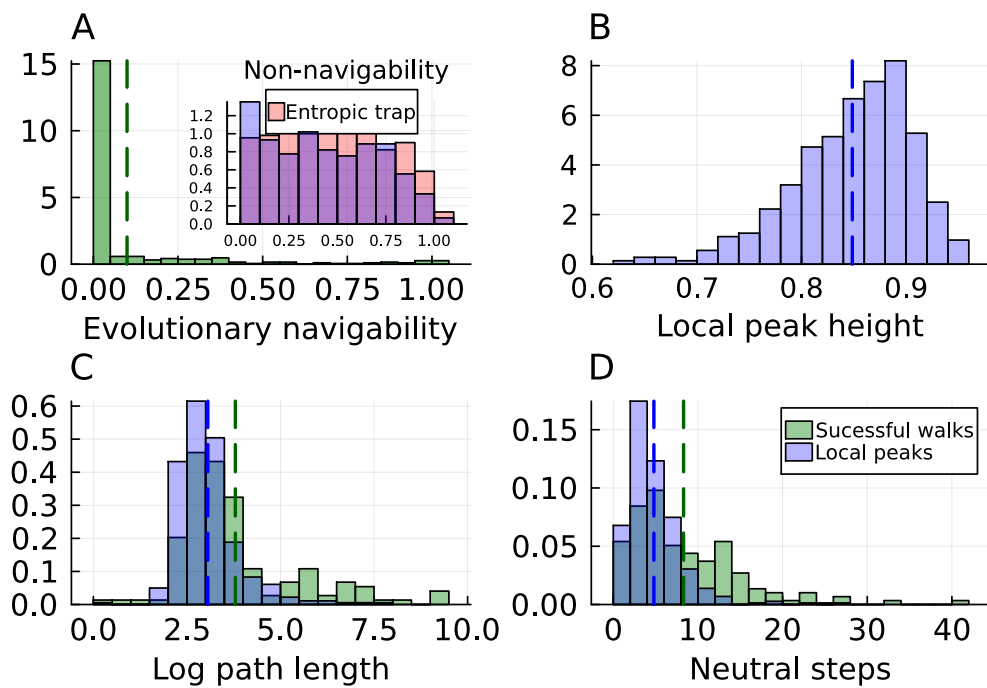
SUPPLEMENTARY FIGURE 4.11: Histogram showing the distribution of the average fitness of 2400 correlated fitness landscapes.



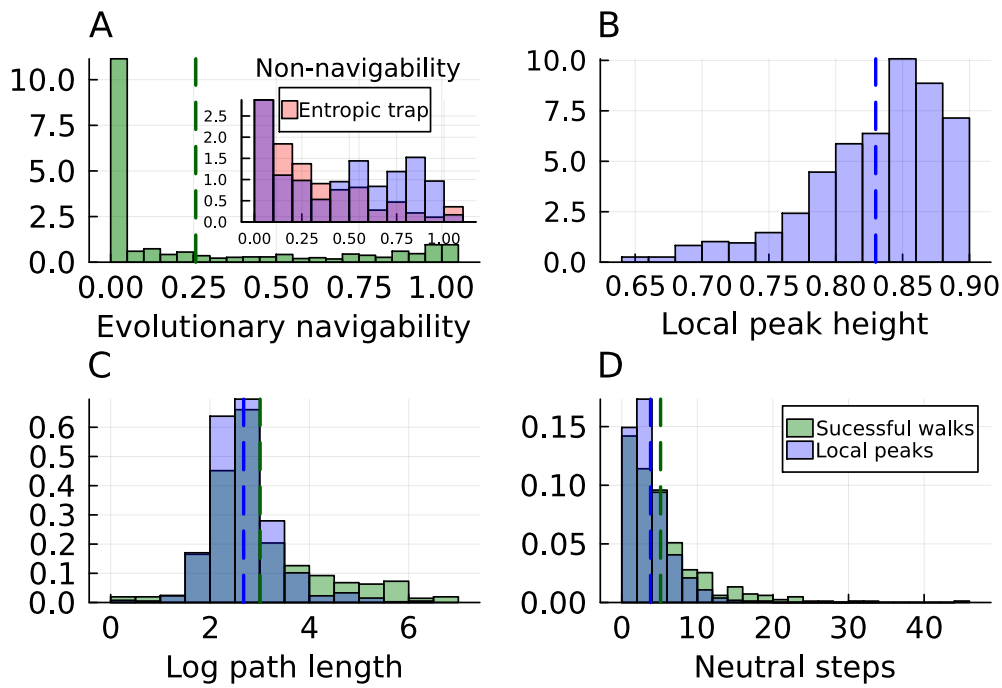
SUPPLEMENTARY FIGURE 4.12: Number of neutral steps as a function of the path length segregated by complexity of source and target.



SUPPLEMENTARY FIGURE 4.13: Histograms showing distributions of A. evolutionary navigability and the causes of non-navigability (inset), B. local peak height reached by failed walks, C. natural logarithm of the path length for successful and failed walks that ended up on local peaks and D. neutral steps for successful and failed walks that ended up on local peaks for linearly correlated landscapes. Distribution means are indicated by dashed vertical lines. Population size $N = 10$.



SUPPLEMENTARY FIGURE 4.14: Width of solution space = 0.95: Histograms showing distributions of A. evolutionary navigability and the causes of non-navigability (inset), B. local peak height reached by failed walks, C. logarithm of the path length for successful and failed walks that ended up on local peaks and D. neutral steps for successful and failed walks that ended up on local peaks for linearly correlated landscapes. Distribution means are indicated by dashed vertical lines.



SUPPLEMENTARY FIGURE 4.15: Width of solution space = 0.9: showing distributions of A. evolutionary navigability and the causes of non-navigability (inset), B. local peak height reached by failed walks, C. logarithm of the path length for successful and failed walks that ended up on local peaks and D. neutral steps for successful and failed walks that ended up on local peaks for linearly correlated landscapes. Distribution means are indicated by dashed vertical lines.

CONCLUSIONS AND FUTURE DIRECTIONS

*And when you have reached the mountain top,
then you shall begin to climb.*

— Khalil Gibran

5.1 CONCLUSIONS

In this thesis, I addressed three open questions motivated by adaptive landscapes at three levels of biological organization, ranging from the level of DNA to the level development. These landscapes had very different architectures and comprised of genotypes of different lengths, alphabet cardinalities and allele graph connectivities. Moreover, they measured very different phenotypes. The range of questions that could be addressed in this thesis, highlights the versatility of the concept of an adaptive landscape.

A common thread running through the chapters is that the topography of an adaptive landscape, measured by the level of correlations in the landscape, had a significant influence on the answers to our questions. In Chapter 2, we found that correlated genotype-phenotype landscapes have higher incongruence than uncorrelated landscapes. In Chapter 3, we saw that increasing the alphabet cardinality increases both, the landscape ruggedness and the number of accessible paths to the global peak in all landscapes, however, it only aids adaptation in uncorrelated fitness landscapes. Finally, in Chapter 4, we observed that the Biomorphs landscapes with correlated phenotype-fitness maps are much more navigable than the Biomorphs landscapes with uncorrelated maps.

The biological insights gained from this thesis also tie in well together. In Chapter 2, we learnt about the evolution of transcription factor (TF) binding sites, particularly, why we should not always expect to find the consensus binding sequence of the TF, at the DNA binding sites. The binding of the TF to its corresponding binding site, regulates the expression of genes, which code for different proteins. These proteins perform various essential functions in the cell. In Chapter 3, we examined how protein function can be optimised by including more amino acids in the protein alphabet. Proteins and DNA interact in gene regulatory networks to orchestrate the development of an organism. Changes in these interactions can alter the body plan of an organism. This can give rise to novel features, that can help the organism survive in diverse and complex environments. In Chapter 4, we saw that the emergence of complex body plans in the Biomorphs fitness landscapes – through this multi-level mapping, that connects changes in DNA to changes in body plans – is rare.

While the quotes that lead each chapter of this thesis may or may not be autobiographical, they certainly tell the story of the findings in this thesis. I started out by describing evolution as a quest to climb a mountain and elucidated how adopting such a view can be useful in understanding biological phenomena at various scales. In Chapter 2, I showed how we do not always need to explore new landscapes, to find answers to our questions, sometimes, applying a fresh perspective to the existing landscape is enough to unveil the answers. Realising that there could be selection for low or intermediate binding affinities, is akin to such a shift in perspective, which led us to a possible explanation for the prevalence of low affinity binding sites in the regulatory portfolios of several organisms. In Chapter 3, I focused on how having more directions to move in, can open up new paths, both towards and away from the goal. Yet, when the path is fraught with obstacles, as is the case in uncorrelated fitness landscapes, having more directions to move in and changing direction often, is imperative to reaching the goal of optimised protein function. In Chapter 4, I showed how the classic adage of taking the road not taken, by traversing paths that have a low probability of realisation, can help in reaching the target of increased phenotypic complexity. The final quote that leads this chapter is indicative of what I intend to do next: discuss the shortcomings of the adaptive landscape approach and then suggest directions in which future research can move.

5.2 CAVEATS

While the adaptive landscape is such a promising concept in principle, in practice, it is virtually impossible to measure, especially at higher levels of biological organisation. We can only ever sample a small fraction of the space and depending upon our choice of genotypes, we can end up with a very biased representation of the actual landscape [1]. The lack of consensus on whether traits are modular or omnigenic [2] further complicates the issue, since fitness can depend on multiple traits which can have trade-offs between them. Thus, an accurately measured, combinatorially complete adaptive landscape is a lofty, distant dream. Thus, conclusions drawn based on in-vitro samples of adaptive landscapes ought to be interpreted with caution. Even if we manage to obtain an unbiased sample of an adaptive landscape, environmental fluctuations would constantly alter the adaptive landscape and therefore, conclusions about evolutionary dynamics can only be made for short time periods during which the environment remains constant.

Another shortcoming of the adaptive landscape approach is that rugged fitness landscapes have been considered synonymous with hindered adaption [3]. While this is a reasonable association, ruggedness need not always be detrimental to adaptation. I noticed the limitations of this generalisation throughout the work presented in this thesis. In Chapter 2, we predicted and exhibited that selection for low or intermediate binding affinities increases the ruggedness of both theoretical fitness landscapes and empirical fitness landscapes of TF-DNA interactions. Based on this observation, we expected

populations to be poorly adapted under selection for low or intermediate phenotypic values. On the contrary, we saw that populations evolving on these landscapes had higher mean population fitness than populations evolving on landscapes with selection for other binding affinities. This was because the local peaks in the fitness landscape had a reasonably high fitness. Thus, selection for intermediate binding affinities resulted in rugged yet navigable fitness landscapes. Recently, other such landscapes have been reported in literature [4, 5], which further strengthens the lack of a direct correlation between ruggedness and hindered adaptation. Furthermore, in Chapter 3, we saw that increasing the alphabet cardinality increases the ruggedness in the landscape. The trends observed for the ParE2 and ParE3 landscapes confirmed this result, however the increased ruggedness, which was also accompanied by a steep decline in average peak fitness, did not impede adaptation. This was because the newly introduced peaks had very small *basins of attraction*, i.e. the fraction of random adaptive walks terminating on these peaks. Thus to summarise, ruggedness due to local peaks of reasonably high fitness does not impede adaptation, since all the alternative destinations are almost equally good. Interestingly, ruggedness due to local peaks of very low fitness also does not impede adaptation, since these local peaks are accessible from a very small fraction of the genotype space. Thus, ruggedness need not be as big an impediment as we are used to believing, and the existence of rugged yet navigable landscapes should not be a surprise, but rather an expectation.

Another common limitation that I encountered, was in the focus on the existence of accessible paths in fitness landscapes [6]. For instance, increasing the alphabet cardinality [7, 8] is considered a boon for adaptation, due to the emergence of new accessible paths to the global peak. Moreover, the navigability of fitness landscapes is quantified based on the existence of accessible paths between a given source and target [9]. While it is certainly illuminating to quantify the number of accessible paths to the target of choice, it should not be used as a metric to quantify the probability of reaching a given target. In both Chapters 3 and 4, we showed how solely using this metric can be misleading. Other metrics, such as the fraction of accessible paths from a given source to the target, sizes of the basins of attraction or the evolutionary navigability are more informative in this regard.

Finally, the majority of work on adaptive landscapes has been on the genotype to fitness map, which overlooks the intermediate genotype to phenotype map. Prior work has shown that this intermediate level can be very consequential in determining various properties of the adaptive landscape [9]. I demonstrate this in Chapter 4, by showing that the topography and the navigability of adaptive landscapes are reasonably influenced by the underlying genotype-phenotype map.

5.3 UNFINISHED PUZZLES AND FUTURE DIRECTIONS

In this section, I discuss some possible directions for future research, along with some preliminary results that I obtained in those directions.

5.3.1 *Trapping in changing environments*

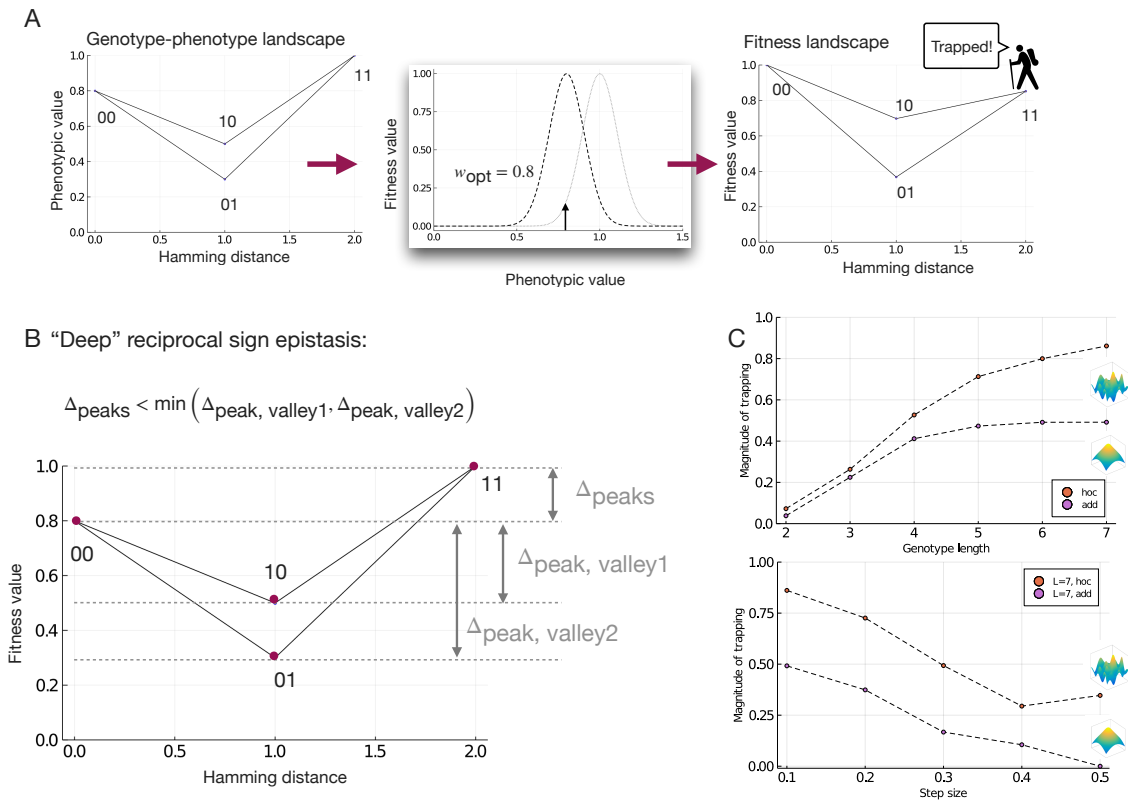


FIGURE 5.1: A. An example of a population getting trapped upon environmental change that alters the optimal phenotypic value to 0.8 from 1. This results in the global peak sequence "11" becoming a local peak, resulting in a population with sub-optimal fitness. B. Condition for deep reciprocal sign epistasis, expressed in terms of the fitness difference between the two peaks (Δ_{peaks}) and the fitness differences between the lower peak and each of the valleys ($\Delta_{peak, valley1}$ and $\Delta_{peak, valley2}$). C. The magnitude of trapping increases with the length of the genotype and decreases with the step size of change in the optimal phenotypic value, which is a proxy for the sensitivity of the adaptive landscape to the environment.

Changing environments can influence adaptation by decreasing the fraction of beneficial mutations [10] and preventing populations from reaching the fitness peak [11]. A gradually changing environment can be modelled using a phenotype–fitness map that was discussed in Chapter 2 with a slowly moving optima, such that the population has

enough time to adapt to the change in the environment. The advantage of this approach is that from Chapter 2, we already know what fitness landscapes look like under selection for a range of phenotypic values in a given genotype-phenotype map. Under such a framework, a population adapted to a given environment can get trapped at a local peak and end up with sub-optimal fitness, as the environment changes the optimal phenotypic value and thus the shape of the fitness landscape. During my Ph.D., I derived some preliminary results that suggest a necessary condition for trapping under gradually changing environments.

Specifically, my preliminary results show that "deep" reciprocal sign epistasis – which is an extreme form of reciprocal sign epistasis (see Figure 5.1 B), is a necessary condition for population trapping. This kind of reciprocal sign epistasis can frequently arise upon selection for intermediate phenotypic values in additive landscapes. Further, my results also show that the magnitude of trapping (i.e. the number of times the population gets trapped as the optimal phenotypic value sweeps the entire range of possible values), increases with increasing genotype length and decreases with increasing sensitivity to the environment (as measured by the step size of change in the optimal phenotypic value, see Figure 5.1 C).

To the best of my knowledge, such concrete topographical conditions for population trapping in changing environments have not been quantified before. Moreover, many questions remain to be examined, such as the effect of increasing the alphabet cardinality on trapping, formulating measures of trapping that take into account the difference in fitness values between global peak and the local peak on which the population is trapped, allowing the population size to vary to study extinction dynamics etc.

5.3.2 *DNA methylation as an extra-dimensional bypass*

Epigenetic modifications such as DNA methylation have long been identified as a means to acquire phenotypic plasticity to cope with variable environments [12]. Such modifications have also been observed to be heritable over at least a few generations [13]. Thus, the effect of changing environments on adaptive landscapes cannot fully be understood without taking into account DNA methylation. Recently, the effect of DNA methylation on the binding affinities of hundreds of TFs has been measured [14]. This data showed that DNA methylation can both increase and decrease the binding affinity and thus, TFs can be classified into "methyl-plus" and "methyl-minus" classes based on whether their binding affinity is increased or decreased upon methylation of the binding sequence. Expanding the TF-DNA genotype-phenotype maps to include methylated DNA sequences and their respective binding affinities to various TFs can be done by considering methylated cytosine as a new element in the DNA alphabet (see Figure 5.2 for possible allele graph topologies). This will make it possible to construct more comprehensive genotype-phenotype maps. More importantly, this can help us identify

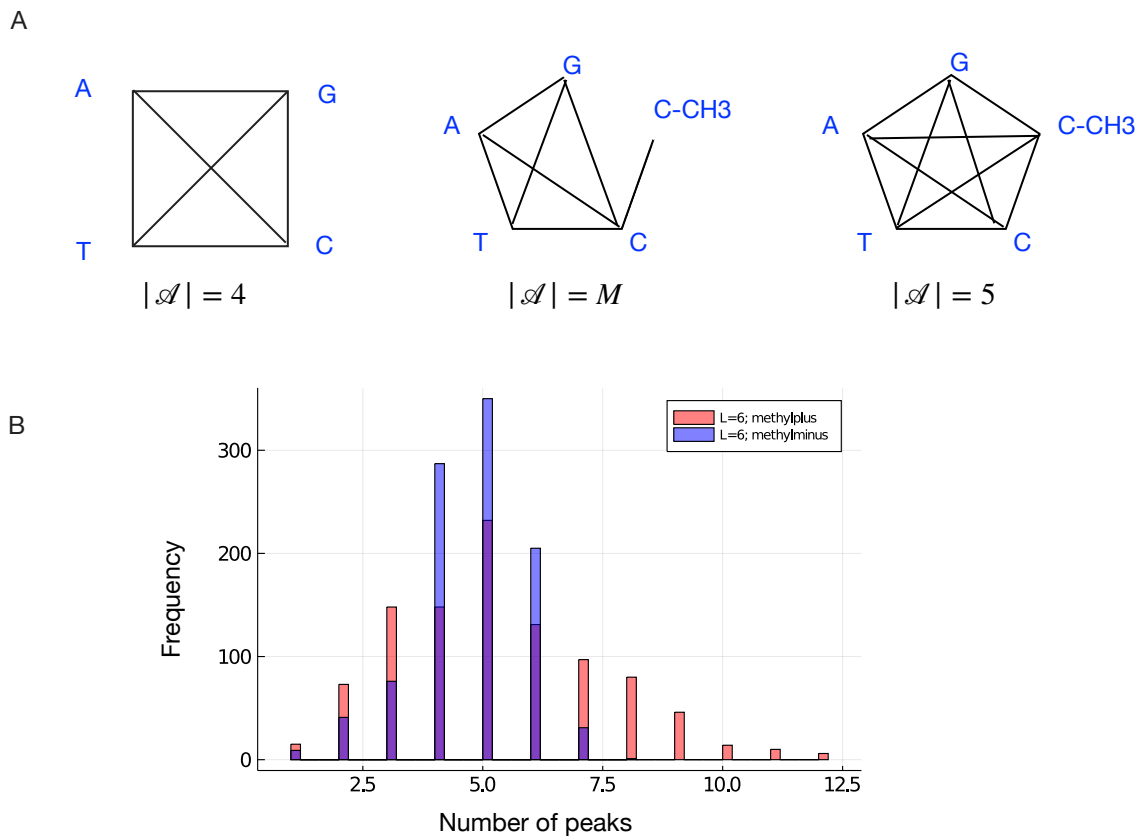


FIGURE 5.2: A. The DNA allele graph after including methylated cytosine as a new element in the DNA alphabet. The allele graph is an intermediate between the fully connected allele graphs of cardinality 4 and 5. B. Distribution of number of peaks post methylation for binding affinity model landscapes of methyl-plus and methyl-minus TFs.

conditions under which DNA methylation helps populations in dealing with fluctuating environments and perhaps also explain the patterns of DNA methylation observed in different cells of various organisms.

My preliminary results show that models of genotype-phenotype maps of both methyl-plus and methyl-minus TFs show increased ruggedness upon the addition of methylated sequences in the genotype space, however, landscapes of TFs of the former class show greater ruggedness than those belonging to the latter class (see Figure 5.2 B). In this case, I also saw that the increased ruggedness is accompanied by hindered adaptation and populations are more likely to reach sub-optimal targets in the expanded landscapes. Whether these landscapes aid adaptation in fluctuating environments still remains to be evaluated.

5.3.3 *Effect of sexual selection on the emergence of complexity*

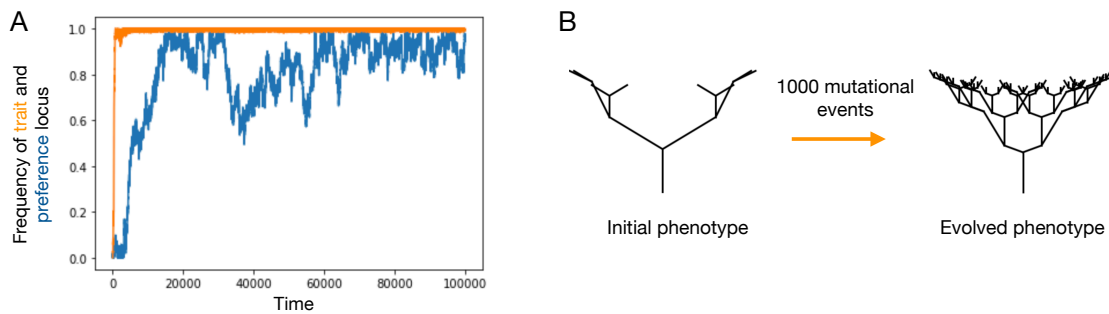


FIGURE 5.3: A. Frequencies of the male trait (in orange) and female preference (in blue) allele of a complex trait as a function of time. B. The starting phenotype and the resulting phenotype after 1000 mutational events.

Explaining the emergence of complex body forms such as the peacock's tail on the Biomorphs fitness landscape, would be impossible without including ecological effects such as sexual selection. Upon incorporating a model of sexual selection [15], by adding an additional locus to the Biomorphs genotype encoding for female preference, my preliminary results showed that even traits with low fitness can fix in the population when preferred by females due to Fisherian runaway selection [16] (see Figure 5.3). Female choice can also increase the frequency of emergence of complex traits, but it remains to be seen how often this happens and under what conditions.

5.3.4 *Effect of neutrality on adaptation*

While it is known that the existence of neutral correlations in a landscape promotes landscape navigability [9], its impact on evolutionary navigability is not clear. To examine the effect of neutral correlations on the accessibility of a target genotype, I analysed House-of-cards (HoC) fitness landscapes [17] of various dimensions and randomly introduced fitness correlations of varying degrees. I did this by forcing a fraction of the neighbours of a chosen genotype to have the same fitness as the genotype. I could vary the degree of "neutrality" in the landscape by gradually increasing the number of neutral components that were introduced.

I measured the landscape navigability between a poorly fit source ($f_s = 0.1$) and the global peak target with unit fitness ($f_t = 1$) and averaged over 1000 distinct ways of "neutralising" the landscape at each level of neutrality. I observed an interesting trend in Figure 5.4 A – the HoC landscapes seem to belong to two different classes: those that show increased landscape navigability upon increasing neutrality and those that show the reverse trend. However, on average, there is an increase in navigability with increasing neutrality as predicted by prior work [9]. On the other hand, in Figure 5.4

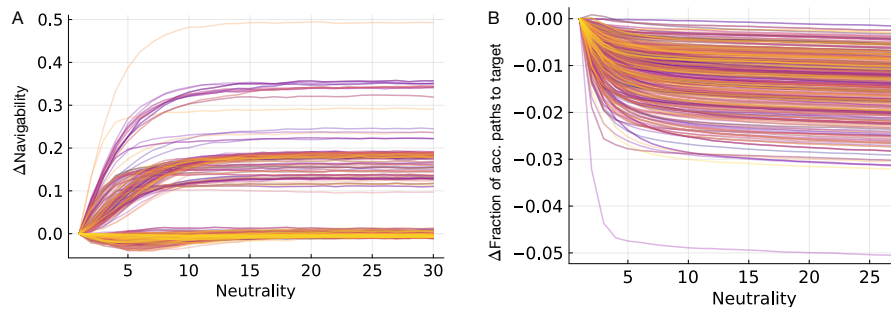


FIGURE 5.4: A. Change in navigability, B. Change in fraction of accessible paths to the target genotype as a function of neutrality for HoC landscapes with $L = 4, a = 4$. Each line shows the result for a different HoC landscape averaged over 1000 ways of "neutralising" the landscape.

B, I observed that the fraction of accessible paths to a target genotype decreases with increasing neutrality, which signifies that the probability of reaching the target decreases with increasing neutrality. These results were robust to changes in the dimensions of the fitness landscapes.

Which features of fitness landscapes lead to the clustering seen in 5.4 A and what do the results in 5.4 B imply about the utilisation of neutral steps by evolving populations? These are still open questions that require further investigations to be answered.

5.4 OUTLOOK

Despite the many limitations and criticisms of the concept of an adaptive landscape [18], its continued use to study evolution, is proof that this concept is here to stay. It is an excellent tool to develop intuition and test theories about the evolution of biological processes across multiple scales. There are still several unexplored and interesting extensions of adaptive landscapes, such as those that include epigenetics and ecological factors, to generate more realistic models that can help us solve the many unanswered questions in biology.

REFERENCES

1. Bank, C. Epistasis and Adaptation on Fitness Landscapes. *Annual Review of Ecology, Evolution, and Systematics* **53**, 457 (2022).
2. Manrubia, S., Cuesta, J. A., Aguirre, J., Ahnert, S. E., Altenberg, L., Cano, A. V., Catalán, P., Diaz-Uriarte, R., Elena, S. F., García-Martín, J. A., Hogeweg, P., Khatri, B. S., Krug, J., Louis, A. A., Martin, N. S., Payne, J. L., Tarnowski, M. J. & Weiß, M. From genotypes to organisms: State-of-the-art and perspectives of a cornerstone in evolutionary dynamics. *Physics of Life Reviews* **38**, 55 (2021).
3. De Visser, J. A. G. M. & Krug, J. Empirical fitness landscapes and the predictability of evolution. *Nature Reviews Genetics* **15**, 480 (2014).
4. Das, S. G., Direito, S. O., Waclaw, B., J Allen, R. & Krug, J. Predictable properties of fitness landscapes induced by adaptational tradeoffs. *eLife*, e55155 (2020).
5. Papkou, A., Garcia-Pastor, L., Escudero, J. A. & Wagner, A. A rugged yet easily navigable fitness landscape. *Science* **382**, eadh3860 (2023).
6. Schmiegelt, B. & Krug, J. Accessibility Percolation on Cartesian Power Graphs. *Journal of Mathematical Biology* **86**, 46 (2023).
7. Wu, N., Dai, C. L., Olson, C. A., Lloyd-Smith, J. O. & Sun, R. Adaptation in protein fitness landscapes is facilitated by indirect paths. *eLife* **5**, e16965 (2016).
8. Zagorski, M., Burda, Z. & Waclaw, B. Beyond the Hypercube: Evolutionary Accessibility of Fitness Landscapes with Realistic Mutational Networks. *PLoS Computational Biology* **12** (2016).
9. Greenbury, S., Louis, A. & Ahnert, S. The structure of genotype-phenotype maps makes fitness landscapes navigable. *Nature Ecology & Evolution* **6**, 1742 (2022).
10. Connallon, T. & Clark, A. G. The distribution of fitness effects in an uncertain world. *Evolution* **69**, 1610 (2015).
11. Trubenová, B., Krejca, M. S., Lehre, P. K. & Kötzting, T. Surfing on the seascape: Adaptation in a changing environment. *Evolution* **73**, 1356 (2019).
12. Stajic, D. & Bank, C. in *Phenotypic Switching* (eds Levine, H., Jolly, M. K., Kulkarni, P. & Nanjundiah, V.) 281 (Academic Press, 2020).
13. Stajic, D. & Jansen, L. E. T. Empirical evidence for epigenetic inheritance driving evolutionary adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences* **376**, 20200121 (2021).

14. Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P. K., Kivioja, T., Dave, K., Zhong, F., Nitta, K. R., Taipale, M., Popov, A., Ginno, P. A., Domcke, S., Yan, J., Schübeler, D., Vinson, C. & Taipale, J. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**, eaaj2239 (2017).
15. Kirkpatrick, M. Sexual Selection and the Evolution of Female Choice. *Evolution* **36**, 1 (1982).
16. Barnett, V. in *Encyclopedia of Evolutionary Psychological Science* (eds Shackelford, T. K. & Weekes-Shackelford, V. A.) 3149 (Springer International Publishing, Cham, 2021).
17. Kingman, J. On the properties of bilinear models for the balance between genetic mutation and selection. *Mathematical Proceedings of the Cambridge Philosophical Society* **81**, 443 (1977).
18. Frank, S. A. in *The Adaptive Landscape in Evolutionary Biology* (Oxford University Press, 2013).

ACKNOWLEDGEMENTS

I would be far from where I am today, were it not for the kindness, generosity, encouragement, support, love and blessings of so many people I've been incredibly fortunate to encounter in my life.

I would like to start by expressing my deepest gratitude to my supervisor, Joshua Payne, for believing in me and giving me the opportunity to pursue my doctoral studies in his group. Josh, your words of encouragement went a long way and greatly motivated me, whenever I needed it the most. I thoroughly enjoyed our scientific discussions and have greatly benefited from your feedback. I cannot thank you enough for being so approachable, understanding and empathetic throughout the course of these four years. You enthusiastically supported everything I wanted to do – from attending the complex systems summer school in Santa Fe to pursuing a part of my Ph.D. in Oxford. You built a group of great researchers and really took very good care of us!

Next, I would like to thank the community I got to be a part of. I really cherished my interactions with the amazing members of the CB group – Alejandro, Alex, Hana, Magda and Paco, our in-office discussions (and digressions) never ended and I only feel joy while looking back at our time together. I learnt a lot from each one of you. I am also very grateful for my interactions with all the kind and intelligent members of the TB group, who were always there to chat and help. I would especially like to mention Berit and H el ene, for being such good companions during the many lockdowns and even afterwards. I would also like to thank Sebastian, Rita, Roli and Dani for helping out with basically anything and making my tenure at ETH a lot more comfortable. I owe a lot of gratitude to Ard, for giving me the opportunity to conduct the final part of my Ph.D. at Oxford and his talented group members, especially Nora, who was an absolute delight to work with. I would also like to extend my gratitude to my master's supervisor, Prof. Dr. Joachim Krug, for his continued feedback and support throughout my Ph.D. journey. I am also deeply grateful to my esteemed examination committee members, especially Prof. Dr. Alex Hall and Prof. Dr. Claudia Bank, for kindly taking out the time to assess my work. This section would not be complete without mentioning my other friends in Zurich, Jiduan – my sweet, inspiring flatmate cum little sister, Ankit Bhaiyya – who really helped me settle down here, Amruta – with whom I savoured many a cup of cappuccino and Divvya – whose cozy apartment always felt like a second home.

Last, but most certainly not the least, I would like to thank my family and friends. My incredible parents, for giving me everything, and my sister, for being the best (and most annoying) companion I could have asked for. Your constant faith in me, helped me believe I can do anything and what greater gift can anyone receive than this? My

amazing grandparents – each extremely evolved and inspiring in their own way, I look up to you, and your love and blessings propel me forward everyday. My partner brought into my life a new family, full of wonderful people and I am truly and deeply grateful to them for supporting all my dreams. Finally, where do I find words to thank my partner, who moved thousands of kilometers to support the completion of my Ph.D. and stood by me through every phase of this journey? Jai, you make me a much better person and might I add: *thou art more lovely and more temperate than a summer's day!*

CURRICULUM VITAE

PERSONAL DATA

Name Malvika Srivastava
Date of Birth September 29, 1995
Place of Birth Patna, India
Citizen of India

EDUCATION

Doctor of Philosophy, ETH Zürich, Zürich, 2020-present
Ph.D. mobility program at the *University of Oxford, 2023-2024*
Evolutionary Biology

Master of Science in Physics, University of Cologne, Cologne, 2019
Overall Grade: 1.0 ¹
Specialisation: Statistical and Biological Physics

Bachelor of Science in Physics, St. Stephen's College, Delhi, 2016
Overall Marks: 90.7%
First Division

All India Senior School Certificate Examination, Notre Dame Academy, Patna, 2013
Overall Marks: 96.6%
Amongst top 1% in the country

WORK EXPERIENCE

Peer review for interdisciplinary journals

- Bulletin of Mathematical Biology (BMAB) and European Physical Letters (EPL)

Co-organiser of the Zurich Interaction Seminar (March 2021- March 2022)

- Current research in *ecology and evolutionary biology*

¹ Grading scale: 1.0 - 1.5 very good, 1.6 - 2.5 good, 2.6 - 3.5 satisfactory, 3.6 - 4.0 sufficient, 4.1 - 5.0 fail

- Jointly hosted by the University of Zürich and ETH Zürich

Supervision of masters students at ETH Zürich

- Term paper on the origins of life (Autumn 2022)
- Evolutionary consequences of DNA methylation (Spring 2022)
- Evolution of pioneer transcription factors (Spring 2021)

Teaching Fellow at Ashoka University, Sonapat (August 2019 - December 2019)

- Statistical Physics under *Prof. Dr. Gautam Menon*
- Quantum mechanics lab course under *Prof. Dr. Pramoda Kumar*

Tutor at University of Cologne, Cologne (October 2017 - October 2018)

- Experimentalphysik I

ACADEMIC AWARDS AND GRANTS

- Received the *Alma Dal Co Foundation grant* for ECRs to visit Harvard University and attend the first Alma Dal Co Foundation lecture in March, 2024.
- Received the *Swiss National Foundation (SNF) Ph.D. mobility grant* to pursue a part of my doctoral studies at the University of Oxford from May, 2023-May, 2024.
- Got selected for a *NSF Research Coordination Network grant funded scholarship* to contribute to the origins of life (OoL) group within the Complex Systems Summer School (CSSS) at Santa Fe from June-July, 2022.
- Qualified the *Physics National Eligibility Test (UGC-NET)* in 2019, with all India rank 15 and became eligible for both Junior Research Fellowship (JRF) and lectureship.
- *Deutscher Akademischer Austauschdienst (DAAD)* scholarship for study completion in the year 2018.
- *Bonn-Cologne graduate school of physics and astronomy (BCGS) scholarship* for the years 2016-18.
- *Usha Ltd. Prize for the best student in Physics* at the St. Stephen's College convocation ceremony 2016.
- *Innovation in Science Pursuit for Higher Research (INSPIRE) Scholarship* for the years 2013-16 by the Department of Science and Technology, Government of India.

SUMMER SCHOOLS & CONFERENCES

- *6th Bangalore School on Population Genetics and Evolution* at the International Centre for Theoretical Sciences (ICTS), Bengaluru in February 2024.
- *Predicting Evolution conference* at European Molecular Biology Laboratory (EMBL) at Heidelberg in July 2023.
- *Congress of the European Society for Evolutionary Biology (ESEB)* at Prague in August 2022.
- *Complex Systems Summer School* at Santa Fe during June-July 2022.
- *Integrated Science Lab (IceLab) camp* at Granö Beckasin in September 2021.
- *EMBO virtual Workshop: Predicting Evolution* in June 2021.
- *Probabilistic Structures in Evolution* summer school at the Reisenburg Castle, Günzburg in June 2018.
- *3rd Bangalore School on Population Genetics and Evolution* at the International Centre for Theoretical Sciences (ICTS), Bengaluru in March 2018.
- Intensive week on *Experiments and simulations on biological systems* in March 2017 at the University of Cologne.
- *Physics of Life monsoon school* in June 2015 at the National Centre for Biological Sciences, Bengaluru.

PROGRAMMING & COMPUTER SKILLS

Julia, Python, C++, Mathematica, Java, R, AnyLogic, Tableau, MySQL, HTML & MS Office.

LANGUAGES

English (IELTS Score 8.0), Hindi (Mother Tongue), Deutsch (level A2).

OTHER INTERESTS

Sports, yoga, painting and dancing.

PUBLICATIONS

- Srivastava, M., Rozhoňová, H. & Payne, J. (2023). **Alphabet cardinality and adaptive evolution** . Journal of Physics A: Mathematical and Theoretical Volume 56, Number 45 <https://iopscience.iop.org/article/10.1088/1751-8121/ad0200/meta>
- Crona, K., Krug, J. & Srivastava, M. (2023). **Geometry of fitness landscapes: Peaks, shapes and universal positive epistasis**. Journal of Mathematical Biology 86 <https://link.springer.com/article/10.1007/s00285-023-01889-6>
- Srivastava, M. & Payne, J. (2022). **On the incongruence of genotype-phenotype and fitness landscapes** PLoS Computational Biology 18(9):e1010524 <https://doi.org/10.1371/journal.pcbi.1010524>
- Srivastava, M (2018). **Epistasis, Shapes and Evolution**. Masters thesis, Universität zu Köln
<https://kups.ub.uni-koeln.de/9494/1/Main.pdf>

