The structure of genotype-phenotype maps makes fitness landscapes navigable

Sam F. Greenbury,[1,2,3] Ard A. Louis,[4] and Sebastian E. Ahnert[5,6]

[1]*Theory of Condensed Matter Group, Cavendish Laboratory, University of Cambridge, UK*
[2]*NIHR Imperial Biomedical Research Centre, ITMAT Data Science Group, Imperial College London, UK*
[3]*Department of Metabolism, Digestion and Reproduction, Imperial College London, UK*
[4]*Rudolf Peierls Centre for Theoretical Physics, University of Oxford, UK*
[5]*Department of Chemical Engineering and Biotechnology, University of Cambridge, Philippa Fawcett Drive, Cambridge CB3 0AS, United Kingdom*
[6]*The Alan Turing Institute, British Library, 96 Euston Road, London NW1 2DB, United Kingdom*

Fitness landscapes are often described in terms of 'peaks' and 'valleys', implying an intuitive low-dimensional landscape of the kind encountered in everyday experience. The space of genotypes, however, is extremely high-dimensional, which results in counter-intuitive structural properties of genotype-phenotype maps. Here we show that these properties, such as the presence of pervasive neutral networks, make fitness landscapes navigable. For three biologically realistic genotype-phenotype map models—RNA secondary structure, protein tertiary structure and protein complexes—we find that, even under random fitness assignment, fitness maxima can be reached from almost any other phenotype without passing through fitness valleys. This in turn implies that true fitness valleys are very rare. By considering evolutionary simulations between pairs of real examples of functional RNA sequences, we show that accessible paths are also likely to be utilised under evolutionary dynamics. Our findings have broad implications for the prediction of natural evolutionary outcomes and for directed evolution.

I. INTRODUCTION

Ever since they were first introduced in Sewall Wright's foundational paper [1], fitness landscapes have become an enduring and central concept in evolutionary biology [2–6]. In particular, a low-dimensional picture of fitness 'peaks' and fitness 'valleys' has played an important role in shaping intuition around evolutionary dynamics. A key prediction is that a population must typically traverse an unfavourable valley of lower fitness to move from one fitness peak to another. But, as already pointed out by many since [4, 7–11], the space of genotypes is typically extremely high dimensional. As illustrated in Fig. 1, what appears to be a fitness valley in a lower-dimensional landscape could be easily bypassed when dimensions are added [9–11].

Three key open questions are: 1) Does the low-dimensional picture of fitness valleys hold for realistic high-dimensional genotype spaces? And if we define *accessible paths* of point mutations between a low fitness phenotype and a high fitness phenotype as those with monotonically increasing fitness, 2) what properties of biological systems facilitate their presence and 3) are such paths sufficiently common that they can easily be found by an evolving population?

One way forward is to consider empirical fitness landscapes, where much recent progress has been made [5, 12], particularly for molecular phenotypes [5, 13–21]. This body of work has yielded important insights, such as the role of local epistatic interactions in sculpting evolutionary paths [22–24]. Nevertheless, ruling out high-dimensional bypasses is difficult in empirical studies because genotype spaces, which grow exponentially as K^L for alphabet size K and genotype length L, are almost always unimaginably vast [25]. They are also highly connected since distances are linear; two genotypes are at most L point mutations away, but are connected by up to $L!$ shortest possible paths given the L mutations may occur in any order. For example, even for a very short $L = 20$ strand of RNA, there are up to $20! \approx 2 \times 10^{18}$ paths between any two genotypes. Empirical landscapes can typically only ever sample a small fraction of the full genotype space, so what appears to be an isolated fitness peak, may in fact be accessible via pathways not included in the experiment.

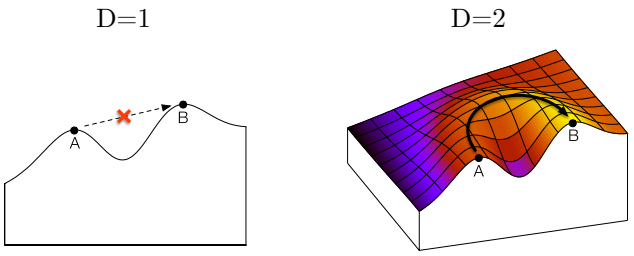


FIG. 1. Illustration of how increasing dimensionality D of the genotype space can affect the navigability and presence of valleys in a fitness landscape.

A different strand of work, which can in principle address questions of global accessibility, has focused on model genotype-to-fitness landscapes [3, 6, 10, 11, 26, 27]. If fitness is assigned randomly to genotypes, as in Kingman's 'house of cards' model [28], then the probability of finding accessible paths is small. If instead there are correlations between fitness and the genotypes, then, depending on details of the model, accessible paths can indeed be common [11, 29]. These correlations are often expressed in terms of ruggedness - a more rugged model has fewer correlations between genotypes and fitness, and so is less navigable. While much progress has been made in this literature, it is not clear how well these models capture the true correlations of biological fitness landscapes.

Here we take a different approach, and build upon recent advances showing that many realistic genotype-phenotype (GP) maps share generic structural features that can en-

hance navigability [30–32]. In contrast to the genotype-to-fitness models studied by others (see above), we consider the genotype-to-phenotype-to-fitness map by inserting the GP map as an additional intermediate step that provides the non-random organisation of the mapping from genotypes to fitness. This means correlations in fitness are naturally incorporated as a consequence of the GP map, rather than through an assumption explicitly parameterised.

One commonality, with important implications for evolutionary dynamics, is the existence of large neutral networks of genotypes that map to the same phenotype [21, 33]. Another is that the mutational robustness ρ_p of a phenotype p (defined as the mean probability that a point mutation leaves the phenotype unchanged) is much larger than what one would expect from a naïve uncorrelated model, where $\rho_p \approx f_p$, the fraction of genotypes that map to phenotype p. Instead, as genotypes from the same neutral network are highly *correlated* [30], $\rho_p \propto -\log f_p$, a generic feature [31, 32, 34–36], that produces orders of magnitude larger robustness than the naïve expectation. Such large robustness means that neutral networks are easily navigable, providing access to a large amount of potential variation [31, 32, 37, 38].

We first explore features of several specific GP maps that affect the navigability and the ruggedness of the landscape: redundancy (large neutral sets), frequency of the unfolded or trivial phenotype, neutral correlations (robustness) and high-dimensionality. We next investigate accessible paths for functional RNA (fRNA) phenotypes identified *in vivo* from the fRNA database [39]. Finally, we explore whether accessible paths are utilised in evolutionary dynamics under a wide range of dynamical regimes. We consider both random and non-random fitness assignments exploring the additional role of non-neutral phenotypic correlations. Our findings demonstrate that generic structural properties shared across many maps from genotype to phenotype dramatically enhance the navigability of fitness landscapes with important implications for evolutionary dynamics.

II. RESULTS

A. Several well-studied genotype-phenotype maps induce navigable fitness landscapes

To concretely measure the effects of different properties of GP maps on the navigability of fitness landscapes, we consider several well-known systems in detail, including the RNA secondary structure GP map for lengths $L = 12$ and $L = 15$ (RNA12, RNA15) [40–46] representing the RNA sequence's minimum free energy folded secondary structure, the Polyomino lattice self-assembly maps ($S_{2,8}$, $S_{3,8}$) [30, 47, 48] modelling the topology of protein quaternary structure assembled from interacting constituent tiles, and several hydrophobic-polar (HP) lattice protein models for folding of a sequence into a tertiary structure (two compact models, HP5x5 and HP3x3x3, and two non-compact ones, HP20 and HP25) [49–51]. See Section IV A and Fig. 6 for further descriptions of these maps.

TABLE I. All GP maps studied with their properties (base K, sequence length L, number of phenotypes N_P, proportion of genotypes with the deleterious phenotype f_{del}, average redundancy $\log_{10} R$, mean genotypic robustness $\langle \rho_g \rangle$) and estimate with standard error of navigability $\langle \psi \rangle \pm SE(\langle \psi \rangle)$. RNA, Polyomino and compact HP GP maps all have navigable fitness landscapes ($\langle \psi \rangle > 0.6$) under random fitness assignment. By contrast, non-compact HP models have very low navigability ($\langle \psi \rangle \leq 0.013$).

GP map	K	L	N_P	f_{del}	$\log_{10} R$	$\langle \rho_g \rangle$	$\langle \psi \rangle \pm SE(\psi)$
RNA12	4	12	58	0.854	4.6	0.465	0.966 ± 0.002
RNA15	4	15	431	0.650	5.9	0.482	0.978 ± 0.002
$S_{2,8}$	8	8	14	0.537	5.8	0.487	0.913 ± 0.008
$S_{3,8}$	8	12	147	0.800	8.0	0.489	0.919 ± 0.003
HP5x5	2	25	549	0.816	4.1	0.285	0.995 ± 0.001
HP3x3x3	2	27	49,807	0.939	2.2	0.115	0.669 ± 0.005
HP20	2	20	5,311	0.976	0.7	0.102	0.004 ± 0.001
HP25	2	25	107,337	0.977	0.9	0.099	0.013 ± 0.001

We performed computational experiments in which fitness is assigned to phenotypes randomly, and two phenotypes are chosen randomly from the set of all phenotypes as the 'source' and 'target'. This is a worst-case scenario that highlights the effect of the correlations between genotypes and phenotypes on fitness. The key property we study is the navigability $\langle \psi \rangle$, defined as:

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

over a set of N source-target pairs (s_k, t_k), where ψ_{ij} is the probability that single-point mutation steps with monotonically increasing fitness (an accessible path) exist from a genotype of phenotype i to a genotype of phenotype j. In other words, the navigability $\langle \psi \rangle$ is the average probability of an accessible path between phenotype pairs in the fitness landscape see IV B 4).

In Table I, we report navigability for each GP map. The value of $\langle \psi \rangle$ is greater than 0.6 for all the GP maps we consider, apart from the non-compact HP models HP20 and HP25. The non-compact HP models have a navigability $\langle \psi \rangle \leq 0.013$. In light of these differences we next investigate what generic structural properties of GP maps promote navigability.

B. Common properties of GP maps are associated with navigability

1. *GP maps with fewer phenotypes, fewer deleterious genotypes and greater genotypic robustness are more navigable*

The first property we consider is the redundancy R of a GP map, defined as the average number of genotypes per non-deleterious phenotype (see Eq. (1)), which is closely related to the average size of the neutral networks. Next we consider the deleterious frequency f_{del}, defined as the fraction of genotype space that does not map to a well-defined phenotype. In

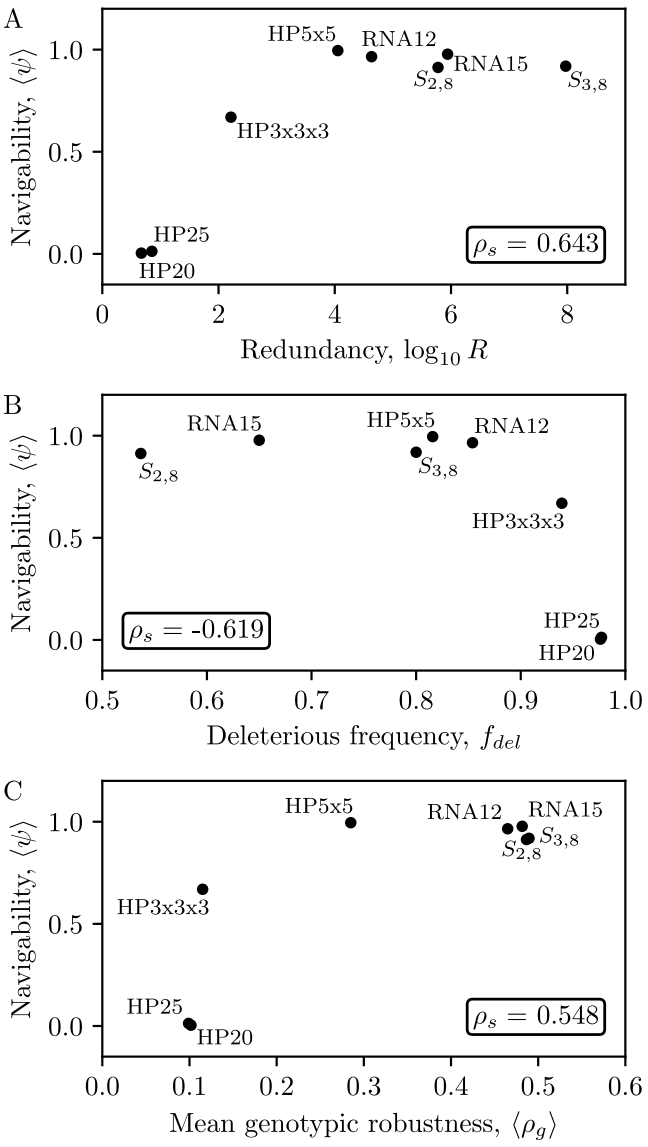


FIG. 2. Navigability of each GP map is plotted in relation to (A) redundancy $\log_{10} R$, (B) deleterious frequency f_{del} and (C) mean genotypic robustness $\langle\rho_g\rangle$. Spearman's rank correlation coefficient ρ_s is shown in each. We find that there is a positive association between navigability and redundancy, a negative association with respect to deleterious frequency for large f_{del} and a positive association with genotypic robustness $\langle\rho_g\rangle$.

the case of RNA secondary structure, the deleterious phenotype would correspond to the unfolded RNA strand (i.e. the absence of any secondary structure). In the HP model it corresponds to the absence of a unique folded ground state. In the Polyomino model it corresponds to unbounded or non-deterministic assembly. Finally, we measure the the mean genotypic robustness $\langle\rho_g\rangle$, defined as the mean proportion of genotypic neighbours that have the same phenotype averaged over the non-deleterious genotypes. This provides a measure of local neutral connectivity.

In Fig. 2 we plot navigability against redundancy, deleterious frequency and mean genotypic robustness with the numerical values provided in Table I and association measured with Spearman's rank correlation coefficient ρ_s. Taking the GP maps together without system-specific considerations, we observe a general increase in navigability for greater redundancy ($\rho_s = 0.643$), smaller f_{del} ($\rho_s = -0.619$) and greater genotypic robustness ($\rho_s = 0.762$).

The results across different GP maps provide some intuition for factors that determine navigability. With decreasing redundancy, it becomes more difficult to access all phenotypes as they begin to occupy smaller fractions of the overall space. As f_{del} increases, more neighbours of a given genotype will have a fitness of 0, therefore localising phenotypes to smaller components in the GP map, increasing the likelihood of each genotype having no neighbouring genotypes with greater fitness. Mean genotypic robustness provides an overall aggregate measure of the connectivity of the neutral networks. HP3x3x3 presents an example of particular interest by maintaining navigability ($\langle\psi\rangle = 0.669$) with less redundancy ($\log_{10} R = 2.2$), large deleterious frequency ($f_{del} = 0.939$) and low genotypic robustness ($\langle\rho_g\rangle = 0.115$). The two non-compact HP models appear to be just below the thresholds that allow for navigability.

2. *Positive neutral correlations increase navigability*

As seen above, robustness plays a key role in enhancing navigability. For a null model, where genotypes map randomly to phenotypes, $\rho_p \sim f_p$ and average robustness is typically extremely low. High robustness therefore corresponds to strong neutral correlations: if a genotype maps to a specific phenotype, the probability that genotypes one mutation away also map to the same phenotype is highly enhanced [30]. As mentioned before, it is widely observed that $\rho_p \sim -\log f_p$, a scaling first pointed out for the RNA map [52], but expected to be universal [44], because it naturally arises from a picture of constrained and unconstrained portions of genotype sequences [34–36]. Interestingly, we can break naturally occurring correlations by taking two genotypes g_1 and g_2 at random and assigning the phenotype of g_1 to g_2 and vice versa. Such random swaps remove the intrinsic local correlations that are intrinsic. Increasing the total number of swaps s reduces the correlations. We define in Eq. (9) a natural measure $c(s)$ of the amount of decorrelation caused by the swaps in terms of the frequency f_p averaged across the phenotypes of the GP map for a given number of swaps s. When $c(s) = 1$, the correlations are equal to the original GP map, and when $c(s) = 0$, the correlations are that of the randomised null model.

In Fig. 3A, we plot how navigability varies with $c(s)$ in $S_{2,8}$, RNA12, HP5x5 and HP3x3x3 GP maps, a subset of the GP maps from the previous section that are both small enough to be tractable here, and have sufficiently large navigability such that the effect of reducing correlations and dimensionality may be sizeable. All four GP maps, on average, show greater navigability for greater $c(s)$ with an approximately linear decay in navigability with decreasing $c(s)$, saturating at a lower value specific to each GP map: 0.378 ± 0.005 for

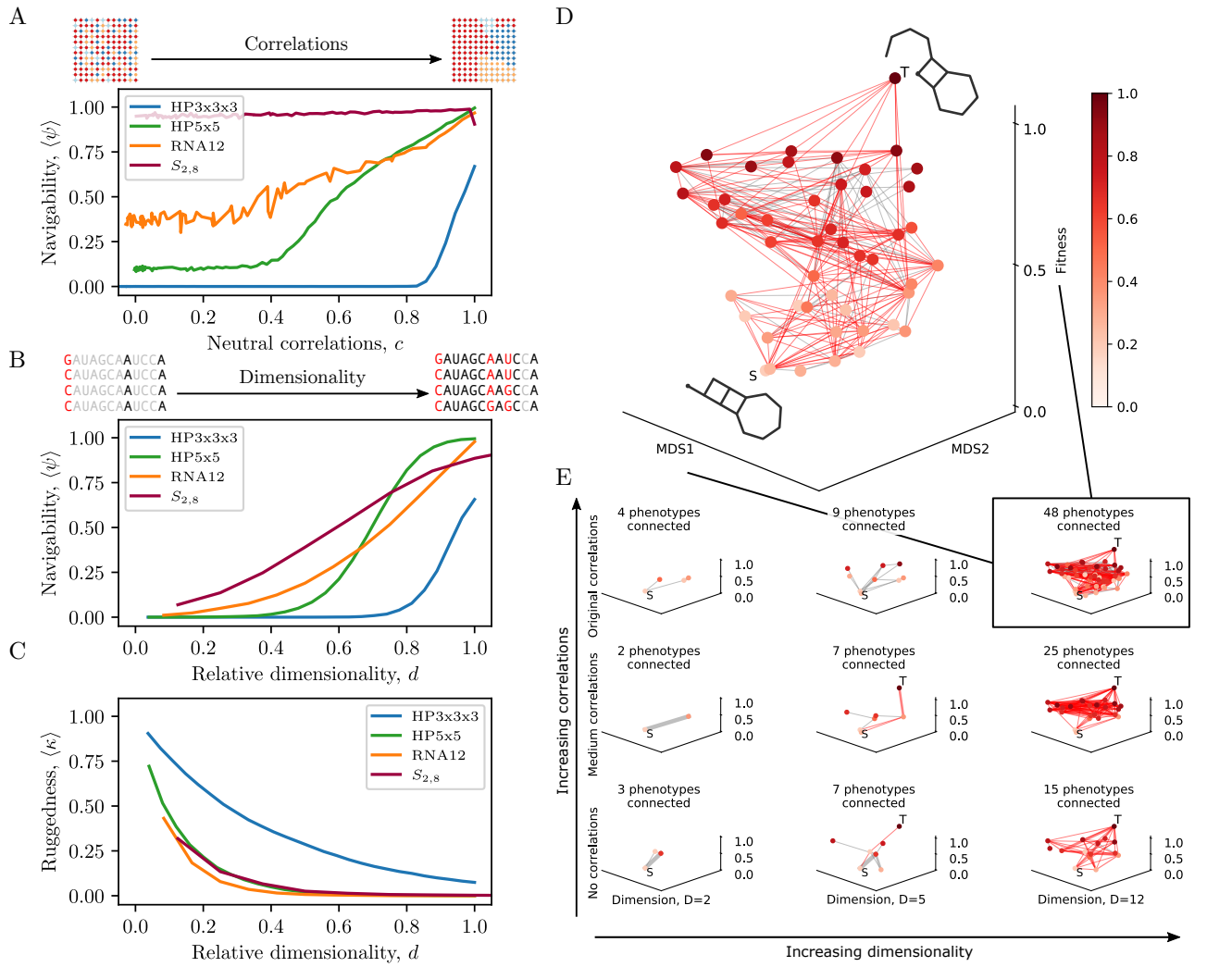


FIG. 3. **(A) Navigability emerges as positive neutral correlations are added to HP3x3x3, HP5x5, RNA12 and $S_{2,8}$ GP maps.** The level of neutral correlations is adjusted through genotype swaps, and the extent of positive neutral correlations after s swaps is measured on a scale c between the original GP map ($c = 1$) and the random null model's correlations ($c = 0$). A caricature of the genotype space, coloured according to phenotypes, is shown for low neutral correlations (top left) and high neutral correlations (top right). **(B) Greater dimensionality of the GP map increases navigability for $S_{2,8}$, HP3x3x3, HP5x5 and RNA12 GP maps.** During the search from a randomly chosen source phenotype to a target phenotype, we only allow D ($d = D/L$) of the total L bases to be mutated to explore genotype space. A caricature of a sequence with grey bases ($L - D$) not mutable, black bases mutable (D) and red bases varying across sequences, is depicted for low dimensionality (top left, $d = 3/12$) and high dimensionality (top right, $d = 11/12$). The GP maps show differing tolerance with respect to navigability under a change in dimensionality, $S_{2,8}$ permitting navigability for low dimensionality significantly more than HP3x3x3, for example. **(C) With increasing dimensionality, landscape ruggedness decreases.** We measure landscape ruggedness $\langle\kappa\rangle$ as the average proportion of all genotypes encountered that are local fitness maxima (no neutral neighbours or neighbours with increased fitness). Ruggedness decreases in all GP maps as dimensionality increases, but the level of ruggedness is GP map dependent. **(D) Accessible and inaccessible paths between two RNA12 phenotypes.** An example is illustrated of the search for an accessible path in a specific random instance of a fitness landscape for all $N_P = 58$ RNA12 phenotypes. Phenotypes are nodes and the edges are possible transitions between genotypes of those phenotypes given the random fitness assignments. Edges that are red are transitions that may lead to the target phenotype from the source phenotype. Inaccessible transitions are shown in grey. The vertical axis is fitness. The horizontal plane is a two-dimensional embedding of the phenotype space of RNA12 derived through a multidimensional scaling (MDS) that uses the pairwise Hamming distances between the dot-bracket representations of the phenotypes. Proximity in the horizontal plane corresponds to similar dot-bracket phenotypes. **(E) The navigability of the phenotype network for three levels of correlations (original, medium, and no correlations) and three levels of dimensionality ($D = 2, 6, 12$).** Navigability and connectivity in the phenotypic network visibly increases with both increasing correlations and dimensionality.

RNA12, 0.100 ± 0.003 for HP5x5, 0.000 ± 0.000 for HP3x3x3, and 0.949 ± 0.002 for $S_{2,8}$, substantial reductions apart from for $S_{2,8}$. In $S_{2,8}$, the navigability $\langle\psi\rangle$ takes a greater value for the decorrelated GP map ($c < 1$) than for the original one ($c = 1$). This is because not all phenotypes are directly accessible from each other in the original GP map. However,

a slight randomisation increases phenotype inter-connectivity due to the fact that the number of phenotypes for $S_{2,8}$ is smaller than the number of local mutations ($N_P < (K-1)L$). Importantly, we expect that in GP maps of longer sequence length L, the role of positive neutral correlations will become even more pronounced. We explore this in Section II C with respect to fRNA phenotypes.

3. *Large dimensionality increases navigability and decreases ruggedness*

We now examine the effect of *dimensionality* of the GP map. The dimensionality of the entire GP map is defined as L, the length of the sequence. During the search for an accessible path from the source to target phenotype, all bases can be mutated, making use of the full dimensionality of the GP map. We can, however, reduce the dimensionality of the search by allowing only a random set of D sites (where $D < L$) to be mutated during a given search for an accessible path from source to target. We then consider $\langle\psi\rangle$ as a function of the *relative dimensionality* $d = D/L \ \forall D \in \{1, ..., L\}$.

In Fig. 3B, we plot navigability $\langle\psi\rangle$ as a function of d. Decreasing dimensionality severely reduces the navigability of fitness landscapes, with a sigmoidal relationship between $\langle\psi\rangle$ and d. All the curves show an increase from low navigability to high navigability as $d \to 1$ of the full GP map. The critical value of d, and general scale and shape, is different across the four GP maps indicating a complex dependence on other GP map properties.

In addition to identifying an accessible path during the search from source to target, we also count the number of genotypes that do not have a neutral neighbour or neighbour with greater fitness. In other words, the proportion of genotypes that are local fitness peaks, therefore providing a measure of landscape *ruggedness*. The average proportion of genotypes that are local fitness peaks across source-target phenotype pairs and fitness assignments in a given GP map, is represented as $\langle\kappa\rangle$. In Fig. 3C, the ruggedness for each relative dimensionality $d = D/L$ is plotted in the same four GP maps. We observe increasing dimensionality reduces ruggedness and, as relative dimensionality drops below a certain level, ruggedness sharply increases. Of note is HP3x3x3, where ruggedness is greater at a given relative dimensionality than for the other GP maps. Where all bases may mutate at $d = 1$, around 7 in 100 genotypes are local peaks ($\langle\kappa\rangle = 0.07$) but navigability remains high ($\langle\psi\rangle = 0.66$), demonstrating that partially rugged landscapes can still be navigable.

We illustrate an example of a source-target search in a schematic of the RNA12 GP map in Fig. 3D. We choose a random source and target pair and, during the search for an accessible path, keep track of all phenotypes encountered, their fitness and any transition between phenotypes. Each phenotype is represented as a node, edges as transitions between phenotypes, and the value on the vertical axis as the fitness. The $N_P = 58$ phenotypes of this GP map are assigned coordinates in the horizontal plane using multidimensional scaling (MDS) based on the pairwise Hamming distance between phenotypes [53]. This allows phenotypes that are similar to each other to be located in similar parts of the MDS1-MDS2 plane. The source and target phenotypes are labelled 'S' and 'T' respectively, edges that may form accessible paths are coloured red, and the remaining edges grey. This depiction of the fitness landscape immediately shows that it is highly connected with many accessible paths.

In Fig. 3E, with the same schematic source-target pair and fitness assignments as Fig. 3D, we illustrate the joint effect of neutral correlations and dimensionality on connectivity and navigability of the phenotype network for three different degrees of correlation (no correlations, some correlations, original correlations) and three different dimensionalities ($D = 2, 6, 12$). The top right of the 9 plots is the original GP map that is also shown enlarged in Fig. 3D. In the case of $D = 2$, the dimensionality in which fitness valleys are often visualised in the literature, phenotypic connectivity is sparse, making the landscape unnavigable. The increase in navigability with increases in both dimensionality and correlations highlights that both the correlation structure of the underlying GP map, and the high-dimensional nature of the evolutionary search, are essential for navigability.

C. Navigability of functional RNA fitness landscapes

Next we focus on the RNA secondary structure GP map by specifically choosing source and target phenotypes that have been observed in nature. This is important as only a small subset of all possible phenotypes are typically seen in real biological systems [48, 54] and it is navigability among this subset that has most relevance for evolutionary processes.

1. *Fitness valleys are not observed between short fRNAs*

We sample RNA secondary structures from the functional RNA database (fRNAdb) [39]. We consider pairs of fRNA phenotypes from the database with sequence length L, assigning a random fitness $0 \leq F_{\text{source}} < 1$ and $F_{\text{target}} = 1$, with random uniform assignment of fitness for all non-trivial phenotypes found during the search process. We consider the range $L \in [20, 40]$, which is larger than the model GP maps we studied more exhaustively. We perform two distinct types of search by either permitting or preventing neutral mutations in exploring a given genotype's mutational neighbourhood. This provides a means to directly measure the role of neutral correlations in facilitating navigability for larger L. Additionally we test two different fitness assignment schemes: (a) random as previously and (b) using a given phenotype's dot-bracket Hamming distance to the target phenotype (see Section IV B 2 for further details). The former ignores non-neutral phenotypic correlations in the GP map, while the latter introduces local phenotypic fitness correlations under the assumption that similar dot-bracket phenotypes have more similar genotypes. As the sequence length increases the number of phenotypes grows as $N_P \approx 1.76^L$ [55] producing a large computational overhead to track all phenotypes and genotypes

TABLE II. The navigability $\langle\psi\rangle$ for length $L = 20 - 40$ fRNAs, the number of phenotypes in the fRNA database, the proportion of runs that are aborted α and the estimated navigability $\langle\psi\rangle$. Results for simulations with and without neutral mutations are shown in the left-hand and right-hand sets of columns respectively, with random and Hamming fitness assignment in the top and bottom sets of rows respectively. For non-aborted runs with neutral mutations permitted, random observed fRNA landscapes are almost completely navigable. When neutral mutations are prohibited, navigability is severely reduced, but still substantial. Hamming fitness assignment has $\langle\psi\rangle \approx 1$ both with and without neutral mutations. Random assignment with neutral mutations had $T = 2 \times 10^6$, while runs with no neutral mutations and for Hamming fitness assignment (denoted with *) have computational threshold $T = 2 \times 10^4$ due to computational limitations.

			Neutral mutations		No neutral mutations*	
Fitness	L	N_P^{fRNAdb}	$\langle\psi\rangle \pm SE(\psi)$	α	$\langle\psi\rangle \pm SE(\psi)$	α
Random	20	14350	0.987 ± 0.004	0.002	0.583 ± 0.016	0.063
	25	12958	0.999 ± 0.001	0.015	0.628 ± 0.017	0.233
	30	42195	1.000 ± 0.000	0.065	0.520 ± 0.022	0.481
	35	752	1.000 ± 0.000	0.190	0.539 ± 0.025	0.603
	40	662	1.000 ± 0.000	0.517	0.273 ± 0.027	0.736
Hamming*	20	14350	1.000 ± 0.000	0.001	0.992 ± 0.003	0.007
	25	12958	1.000 ± 0.000	0.009	0.998 ± 0.001	0.021
	30	42195	1.000 ± 0.000	0.031	0.999 ± 0.001	0.053
	35	752	1.000 ± 0.000	0.050	1.000 ± 0.000	0.049
	40	662	1.000 ± 0.000	0.099	0.998 ± 0.002	0.078

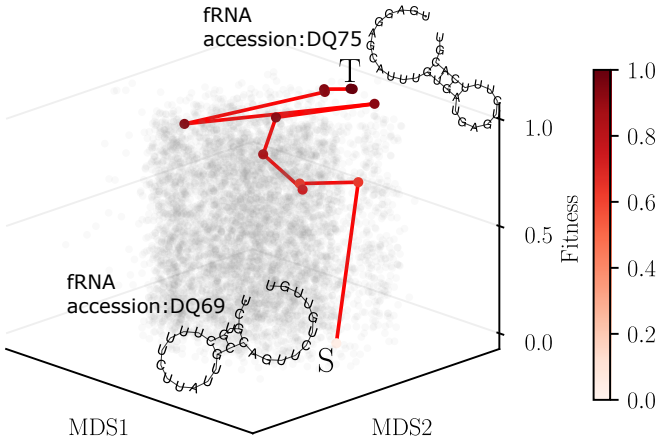


FIG. 4. **Example of an accessible path for a specific $L = 30$ fRNA source-target pair.** As introduced in Fig. 3D, phenotypes are nodes whose coordinates are derived from a multidimensional scaling (MDS) embedding of the phenotype similarities based on Hamming distance, while the vertical axis is fitness. We show the vast extent of phenotypes discovered during the search as 'grey' nodes, a shortest accessible path connecting the source and target phenotypes with red edges, and the phenotypes along this path shaded in proportion to fitness. The example illustrates the interconnected nature of the fitness landscapes for a concrete fRNA example, where the properties of the GP map are key in facilitating navigability.

encountered during a search. The computational threshold T is the maximum number of genotypes whose neighbourhoods are searched before the search is *aborted*, the proportion of which is defined as α. In Section IV F, we describe other computational details necessary to measure navigability for larger L due to computational limitations.

In Table II, the navigability $\langle\psi\rangle$ for fitness landscapes with fRNA of sequence length $L = 20 - 40$ is reported along with the proportion of searches that were aborted and whether or not neutral mutations were permitted. With neutral mutations allowed, $\langle\psi\rangle \approx 1$, suggesting that fitness landscapes with fRNAdb source and targets are highly navigable. For $L > 30$ the proportion of aborted searches increases, leading to the greater potential for this estimate to be biased. However, there is a strong indication that with a greater computational threshold, similarly large navigability would be achieved at even larger L fRNA landscapes due to the observed scaling of $\langle\psi\rangle$ with the computational threshold (see Section A). When Hamming fitness assignment is used, $\langle\psi\rangle = 1$ and aborted runs are rare demonstrating that phenotypic correlations (like genotypic correlations) enhance navigability.

Where neutral mutations are disallowed, we find that navigability is markedly reduced below unity, although still substantially greater than zero ($\langle\psi\rangle \in [0.273, 0.628]$). This finding is intriguing as it highlights that positive neutral correlations are important, but not essential, for the existence of accessible paths in this system. A possible explanation lies in the vast number of phenotypes $N_P \approx 1.76^L$ available in the GP map, coupled with its high dimensionality. As fitness is randomly assigned and novel variation is only a few mutations away, there is a pool of non-neutral phenotypes with possibly larger fitness, potentially within a small mutational radius. Additionally, given the fRNAdb is occupied with highly frequent phenotypes [55], the source and target themselves will have greater robustness and therefore larger neutral spaces that may be found. With Hamming fitness assignment, the reduction in navigability is only marginal suggesting that phenotypic correlations can overcome dramatically diminished neutral correlations.

In Fig. 4, we use the representation introduced in Fig. 3D to illustrate an accessible path in fRNA. For the successful traversal between a specific source and target fRNA, we see a vast array of background, 'greyed out' phenotypes discovered during the search for an accessible path, as well as a shortest accessible path connecting 10 different phenotypes with the node colour and their vertical axis coordinate showing their fitness. This illustration further highlights the hyper-connectedness and high-dimensional bypasses present in fRNA GP maps that are afforded through exponentially increasing redundancy, positive neutral correlations, and high dimensionality. The phenotype network also serves again as an alternative depiction of the fitness landscape in which the effect of GP map structure on the course of potential evolutionary explorations may be grasped more intuitively.

Summarising our results, we have demonstrated that fRNA GP maps have navigable fitness landscapes up to $L = 30$ fRNA , and likely up to $L = 40$ given observed scaling with increased computational time.. They are highly likely to be navigable for even larger *in vivo* fRNAs due to the observed scaling of both the GP map properties and navigability with

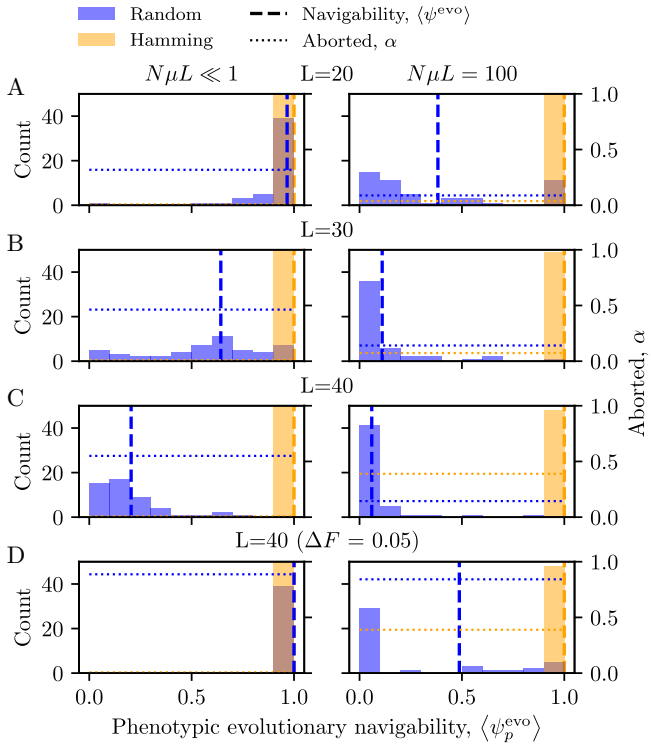


FIG. 5. **Navigability under evolutionary dynamics with source and target phenotypes sampled from the fRNA database with random and Hamming fitness assignments under monomorphic (left) and polymorphic (right) evolutionary dynamics.** (A-C) The evolutionary navigability for 50 different target fRNA phenotypes p are illustrated using histograms of $\langle\psi_p^{\text{evo}}\rangle$ for each random target fRNA phenotype at lengths $L = 20, 30, 40$. The blue shaded histogram show the proportion of successful searches for random fitness assignment, and the orange histogram for Hamming fitness assignment. Evolutionary navigability is shown as vertical dashed lines and the proportion of aborted searches as horizontal dotted lines. Hamming fitness assignment has high navigability across all lengths in both regimes. Random fitness assignment has higher navigability at $L = 20$, decreasing with increasing L. (D) Identical to the (C) but now allowing small decreases in fitness of $\Delta F = 0.05$ along the adaptive path, which increases navigability in random fitness landscapes. Histogram bar heights may not sum to $N_t = 50$ due to some phenotypes p having all searches aborted yielding no estimate for $\langle\psi_p^{\text{evo}}\rangle$.

respect to the computational threshold. Neutral mutations drastically increase and non-neutral phenotypic correlations enhance navigability, but neither solely determine the presence of accessible paths.

2. *Evolutionary dynamics make use of accessible paths between fRNAs*

Having considered whether accessible paths *exist* in a variety of GP maps, we next consider whether these accessible paths are *found* under evolutionary dynamics. It is conceivable that, while accessible paths to the true fitness maximum exist in a fitness landscape, there are so many alternative paths leading to local fitness maxima, that a population will become trapped, necessitating passage across a fitness valley to reach the fittest phenotype.

Under evolutionary dynamics the adaptive path taken may be dependent on population mutation rate ($N\mu L$, with N population size, μ point mutation rate and L sequence length). We therefore explored both monomorphic ($N\mu L \ll 1$) and polymorphic $N\mu L \gg 1$ regimes in the main text, with Section B and Section C further investigating the role of population size and mutation rates.

For monomorphic evolutionary dynamics, we simulated evolution with a sequential fixation model [56] combined with Kimura's fixation probability for a haploid population [57]. For polymorphic evolutionary dynamics, we simulated a Wright-Fisher model, implemented via a genetic algorithm. Further details are provided in Section II C 2. As in the previous Section II C, we again consider both random and Hamming fitness assignments.

We chose $N_s = 20$ source phenotypes for each of $N_t = 50$ target phenotypes, with the population initialised to a clonal population of genotypes that map to the source phenotype. The fitness of the target was set to 1. The *adaptive path* was measured during evolutionary search. In the monomorphic case, this was the sequence of genotypes (and their phenotypes) that fixed, while for polymorphic dynamics, the change in fitness of the population's majority phenotype (more than 50% of genotypes) was measured. Analogously to landscape navigability, we define *evolutionary navigability* $\langle\psi^{\text{evo}}\rangle$ as the average probability that the adaptive path reaches a target phenotype from a source phenotype via an accessible path, with the *phenotypic evolutionary navigability* $\langle\psi_p^{\text{evo}}\rangle$ as the probability of that an adaptive path to a specific target phenotype p is an accessible one. We have previously required that an accessible path did not have any decrease in fitness expressed with a *tolerance* to a maximum decrease of $\Delta F = 0$ between phenotypes along the path. Here, we also measure the effect of relaxing this constraint.

In Fig. 5A-C(left) we plot histograms of evolutionary navigability in monomorphic ($N\mu L \ll 1$ at $N = 100,000$) and polymorphic ($N\mu L = 100$ at $N = 100$) dynamical regimes for fRNA source and target phenotypes at sequence lengths $L = 20, 30, 40$. For monomorphic dynamics, we find that the Hamming fitness assignment has $\langle\psi_p^{\text{evo}}\rangle = 1$ across all phenotypes with negligible aborted runs. Random fitness assignment has high navigability at $L = 20$ but decreasing with increasing L. Across all L the aborted fraction is sizeable $\alpha \in [0.318, 0.550]$. Fig. 5D(left) relaxes the requirement for an accessible path to have a maximum fitness decrease of $\Delta F = 0.05$. We find random fitness assignment can be navigable for some phenotypes ($\langle\psi_p^{\text{evo}}\rangle = 1$) but with an increase in aborted runs ($\alpha = 0.887$) where it remains uncertain. In Fig. 5A-C(right), a similar pattern is observed for polymorphic evolutionary dynamics, with Hamming fitness assignment having $\langle\psi_p^{\text{evo}}\rangle = 1$ but with an increasing proportion of aborted runs for increasing L. Random assignment has reasonable $\langle\psi_p^{\text{evo}}\rangle = 0.302$ at $L = 20$ but becomes negligible at $L = 40$. In Fig. 5D(right), again, allowing a tolerance

of $\Delta F = 0.05$ regains navigability for some phenotypes with random fitness assignment (overall $\langle \psi^{\text{evo}} \rangle = 0.487$) but with a large proportion aborted ($\alpha = 0.842$). We explore tolerance further in Section E by identifying a ΔF sufficient to generate navigability.

Navigability under monomorphic dynamics are sensitive to population size N with typically lower navigability in smaller populations (see Section B). Navigability under polymorphic evolutionary dynamics are sensitive to population mutation rate $N\mu L$, with lower evolutionary navigability observed for smaller population mutation rates in the Wright-Fisher model (Section C). Echoing our investigation into the role of neutral mutations in Section II C, we also considered monomorphic evolutionary dynamics with and without neutral mutations where we also found that neutral mutations enhance evolutionary navigability (Section D). To gain insight beyond the computational limits incurred with increasing L, in Section F, we explore navigability of coarse-grained fRNA 'shape' phenotypes [58]. These have recently been shown to possess similar GP map properties of redundancy, bias [59] and neutral correlations [60] that we have shown earlier to be associated with and facilitate navigability. With this model we find evolutionary navigability can be attained in the monomorphic setting at lengths $L = 60, 100, 140$, and also enhanced by the Levenshtein fitness assignment, an equivalent of the Hamming fitness assignment related to phenotypic correlations.

Our evolutionary simulations have importantly shown that neutral genotypic correlations and phenotypic correlations are sufficient to allow evolution to find accessible paths in the fitness landscape for certain conditions in both monomorphic and polymorphic regimes. Additionally, when these properties are jointly available (i.e. Hamming fitness assignment with neutral mutations), they facilitate navigability under a very broad range of dynamical regimes.

III. DISCUSSION

Our main contribution is to explicitly include the phenotype as an intermediate step between genotype and fitness, and therefore implicitly include generic properties such as redundancy and correlations which dramatically increase the navigability of fitness landscapes. We demonstrated for a wide range of evolutionary dynamical regimes, that biological systems can be navigable, even when fitness is assumed to be distributed randomly. When fitness correlations based on phenotypic similarity are incorporated, navigability is enhanced even further Our conclusions, that true fitness valleys are probably rare, should be relevant for a broad scope of issues in biological evolution.

Open questions remain: Firstly, our computational explorations only allow for relatively small systems to be studied. However, there is evidence in our findings to suggest that navigability will hold at larger L too: 1) We found navigability to be monotonic for increasing L in the RNA we studied; 2) The deleterious fraction decreases monotonically with L for RNA; 3) While the number of sequences grows exponentially ($N_P \sim 1.76^L$), as the number of sequences grows exponentially as 4^L, the average redundancy R will grow exponentially too at $R \sim 2.27^L$. Given robustness scales with frequency, the average genotype's robustness will also grow meaning that genotypes encountered along paths will have more neutral dimensions available; 4) Importantly, phenotypes found *in vivo* are taken from a tiny fraction of phenotypes with the largest neutral sets and largest robustness [45, 55, 59], a phenomenon which may hold much more widely [48] and which should greatly enhance navigability. 5) We have mainly studied a worst-case scenario with random assignment of fitness to phenotypes. For the fRNA strands, we also studied a fitness landscape based on Hamming distance between structures, showing that improved navigability drastically. While much less is known about such *phenotypic fitness correlations*, they are likely to exist more generally, and so enhance navigability. Taken together these arguments suggest that landscapes at larger L should have accessible paths and be navigable.

Another issue to consider is that the model systems we study all describe some form of self-assembly, where we assign fitness to the physical structure alone. This will not always hold for all biological systems. For example, where a specific sequence is necessary to facilitate binding of a protein, an additional sequence constraint is imposed on top of that required to specify the structure. This additional specificity potentially reduces both the redundancy of the phenotype and the dimensionality available for accessing alternate genotypes.

Our findings support work on the role of high-dimensionality in promoting accessibility [4, 7–11], as well as attempts to create an up-to-date metaphor for evolutionary adaptation [61], but moves well beyond the current literature by demonstrating both the generality across multiple systems and the presence of navigability with either random fitness assignments to phenotypes, or ones grounded in phenotypic similarity. A fuller understanding of the role of the GP map in structuring the high dimensional fitness landscape could provide vital insights into areas such as the arrival of drug resistance [62, 63], the mutational progressions of cancer [64]. In particular, understanding the fitness landscapes in cancer is notoriously challenging due to the difficulty of inferring the fitness of mutants [65]. Introducing the notion of a mapping from genotypes to phenotypes and studying generic properties such as genetic correlations and redundancy may provide new insights into cancer evolution. Another example of particular current interest is found in viruses such as influenza or SARS-CoV-2 where mutations across a multitude of sites (high dimensionality) leads to variants (phenotypes) that evade host immune responses. Understanding whether accessible paths are afforded to such pathogenic viruses could provide important insights into their progression and population dynamics.

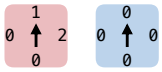
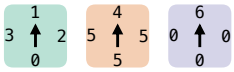
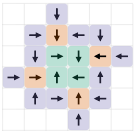
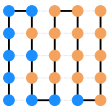
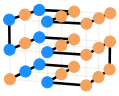
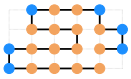
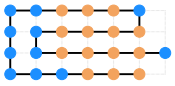
Category	GP map	Genotype	Phenotype	Phenotype encoding
RNA RNA sequence is folded to minimum free energy secondary stucture represented with dot-bracket or shape representation	**RNA12**	GGGGGAAAACCC		(((.....)))
	RNA15	GGGACCAAAGGUCCC		(((((((...)))))))
	fRNA30	UGGAACGAGUUCUUGGAAUGAAUCCCAUGC		.(((....((((...))))...)))..... Shape level 1: _[_[]_]_ Shape level 3: [[]]
Polyomino Tile assembly kit with interacting pairs (1–2, 3–4, 5–6, with 0 and 7 neutral) self-assembles to polyomino with grid coordinates	$S_{2,8}$	**1200 0000** 		2D tile coordinates: **0,1 1,1** **0,0 0,1**
	$S_{3,8}$	**1203 4555 6000** 		2D tile coordinates: **3,0 4,1 3,2 3,3 2,4** **1,1 0,2 4,2 4,3 3,4** **2,1 1,2 1,3 5,3 4,4** **3,1 2,2 2,3 1,4 2,5**
HP (compact) Hydrophobic (H) – Polar (P) protein sequences fold with H–H neighbour interactions contributing to the minimum energy fold on a compact grid	**HP5x5**	PPPPPPHPHPPHHHHHHHHPHHHHH		2D Lattice directions (U=up, D=Down, L=Left, R=Right): **UUUURDDDDRUUUURDDDDRUUUU**
	HP3x3x3	HPPHHPHHHHHHPHHPHPPHPHHPHHH		3D Lattice directions (U=up, D=Down, L=Left, R=Right, F=Forward, B=Back): **FFRBBRFFUBBLFFLBBUFFRBBRFF**
HP (non-compact) Hydrophobic (H) – Polar (P) protein sequences fold on grid without compact restriction	**HP20**	HHHHPPHHHHHHPHHPHPPH		**LLLLURRRULLURRRDRDL**
	HP25	PHHHHPPHHHHPHHHPPPPPPPHHH		**LLLLLURRRRULLLLLDDDRRRRR**

FIG. 6. **Depiction of the different biological systems, specific GP maps considered and example genotype, phenotype and encoding of phenotype.**

IV. METHODS

A. Self-assembly GP maps

We consider three GP maps for different systems of biological self-assembly: the RNA secondary structure GP map [41] for secondary structure of RNA sequences, the HP lattice model for protein tertiary structure [49, 66] and the Polyomino model for protein quaternary structure [47]. The phenotype in each is solely related to the assembled structure. The GP maps have been extensively studied and compared in ref. [30] and are shown in Fig. 6. We summarise their details:

- *RNA secondary structure*: Genotypes are sequences where each position is one of the four RNA nucleotide bases (an alphabet $\mathcal{A} = \{A, C, G, U\}$). Phenotypes are the secondary structure bonding pattern of the minimum free energy fold of the genotype, represented with the dot-bracket notation [41], apart from in Section F where 'RNA shapes' are used instead [58]. We use the Vienna package [41] (version 1.8.5) with default parameters to convert RNA sequences to dot-bracket secondary structures. GP maps are represented as RNAL with sequences of length L. Fig. 6 illustrates three example GP maps at $L = 12, 15, 30$.

- *HP lattice model*: Genotypes are sequences where each position is an amino acid base classified as either hydrophobic or polar (an alphabet $\mathcal{A} = \{H, P\}$) [49, 66]. Phenotypes are the minimum energy fold of the genotype, restricting the fold to occur on either a square or cubic lattice, with the energetics determined by interactions between neighbours on the lattice that are non-adjacent in the sequence. We represent folds with a string describing the moves that are required to construct fold on the lattice with the basis: 'Up', 'Down', 'Right', 'Left' for 2D lattices, and additionally 'Forward' and 'Back' for 3D lattices. We follow refs. [50, 51] and consider energetic interactions between non-adjacent pairs to have values $E_{\mathrm{HH}} = -1$, with $E_{\mathrm{HP}} = E_{\mathrm{PP}} = 0$, where H are hydrophobic and P are polar amino acids. If a sequence has a unique minimum energy structure, its phenotype is that structure, otherwise it is considered degenerate and not defined. We consider both the non-compact GP map and

Term and description	Representation
Alphabet	$\mathcal{A}$
Genotype space	$\mathcal{G}$
Phenotype space	$\mathcal{P}$
Alphabet size	$\lvert\mathcal{A}\rvert = K$
Genotype length	L
Dimensionality: the number of bases that may be mutated in the genotype	D
Relative dimensionality	$d = D/L$
1-mutant genotypes of genotype g	$\sigma(g)$
Number of 1-mutation neighbours of genotype g	$\lvert\sigma(g)\rvert = (K-1)L$
Number of genotypes	$N_G = K^L$
Number of phenotypes	N_P
GP map	$M : \mathcal{G} \to \mathcal{P}$
Phenotype redundancy: the number of genotypes that map to phenotype p	n_p
Phenotype frequency: the fraction of genotypes that map to phenotype p	f_p
Genotypic robustness: the average proportion of mutational neighbours that are of the same phenotype for a random non-deleterious genotype	ρ_g
Phenotypic robustness: the average proportion mutational neighbours that are of the same phenotype for a random genotype of phenotype p	ρ_p
GP map redundancy: the average number of genotypes per non-deleterious phenotype	$R = K^L(1 - f_{del})/(N_P - 1)$
Fitness of phenotype p	F_p
Presence of at least one accessible path (monotonic in fitness) with a binary event indicator	ψ
Probability of an accessible path existing between phenotype p and q	ψ_{pq}
Navigability: average probability of an accessible path for phenotype pairs of a GP map	$\langle\psi\rangle$
Evolutionary navigability: average probability that an adaptive path is an accessible one under evolutionary dynamics on a GP map	$\langle\psi^{\text{evo}}\rangle$
Phenotypic evolutionary navigability: average probability that an adaptive path is an accessible one under evolutionary dynamics on a GP map towards specific target phenotype p	$\langle\psi_p^{\text{evo}}\rangle$
Ruggedness	$\langle\kappa\rangle$
Neutral correlations	c
RNA GP map for genotype sequence length L	RNAL
Polyomino GP map with N_t kit tiles and N_c interface types	S_{N_t,N_c}
HP lattice GP map for genotype sequence length L	HPL
HP lattice GP map for compact phenotypes in the grid box $l \times w \times h$	HPlxwxh

TABLE III. **Terminology.** A summary of terms and their representations used in the paper. The first column (left) provides the term used and its description, while the second column (right) has the corresponding mathematical symbol and equation where relevant.

compact GP maps. The former identifies the minimum energy fold among all folds of a given length and is referred to as HPL. The latter only considers the set of compact structures as possible folds and is referred to as HPlxw for 2D lattices (e.g. HP5x5) and HPlxwxh for 3D lattices (e.g. HP3x3x3). The compact HP model only allows folds that fit within the prescribed grid (e.g. either 5x5 or 3x3x3 here). These maximally compact subsets aim to capture the globular nature of *in vivo* proteins [67], vastly reducing the number of folds at a given length while being more faithful to observed protein structure topology. Fig. 6 depicts examples from the two compact (HP3x3x3 and HP5x5) and two non-compact (HP20 and HP25) GP maps studied here.

- *Polyomino model*: The Polyomino GP map represents protein quaternary structure on a 2D square lattice, with constituent tiles from assembly kit placed where interactions are present. Genotypes represent an assembly kit of N_t tiles, where each edge of the tile may have one of N_c interface types denoted by integers. Here we follow refs. [30, 47] and consider the GP maps S_{N_t,N_c}, specifically $S_{2,8}$ and $S_{3,8}$. We use $N_c = 8$ with bases from an alphabet $\mathcal{A} = \{0, 1, 2, 3, 4, 6, 7\}$ for each tile edge. Interactions are only allowed between $1 \leftrightarrow 2$, $3 \leftrightarrow 4$, $5 \leftrightarrow 6$, with 0 and 7 being neutral. The genotype sequence is transformed from a sequence of bases and encoded in blocks of four clockwise around each assembly kit tile. To construct the phenotype from the assembly kit, the first encoded tile is used to 'seed' the assembly, with subsequent tile places made at randomly available points of interaction with assembly kit tiles that may be placed on the lattice. The assembly process terminates upon no available placements remaining or if the structure becomes unbounded. The assembly process is repeated $k = 200$ times with the final polyomino compared across the ensemble of assemblies. The phenotype is the unique bounded shape across the ensemble of assemblies, allowing for rotations, with a classification of unbounded or non-deterministic (*UND*) otherwise.

The GP maps may be further characterised by their genotype sequence length L, base K, number of genotypes $N_G =$

K^L and number of phenotypes N_P. The redundancy n_p of a given phenotype p is the number of genotypes that map to p and this is normalised by the size of the genotype space to give the frequency $f_p = n_p/K^L$. The overall redundancy R of a GP map is defined as the average number of genotypes per non-deleterious phenotype:

$$R = K^L(1 - f_{del})/(N_P - 1) \tag{1}$$

We provide Table III to summarise the characteristic properties used to differentiate the GP maps.

A particular feature of all three GP maps is a single phenotype that is of a different nature to the others: for RNA secondary structure this is the unfolded 'trivial' structure, the HP lattice model it is sequences that have a degenerate minimum energy state and for the Polyomino model it is when there is either unbounded or non-deterministic growth (UND). We refer to this phenotype here as the *deleterious* or *del* phenotype as, in each GP map, we consider it low fitness due to the non-specificity of the structural phenotype. We assign a fitness of zero for del throughout this work. While this is a strong assumption, given the large-scale dominance of the del phenotype in Polyomino and HP GP maps, we expect this assumption to exacerbate the presence of valleys rather than introducing a bias towards navigability.

B. Measuring landscape navigability

1. Definitions and formulation

In order to establish the presence of fitness valleys in a fitness landscape, we consider whether it is possible to reach the fittest phenotype from any given point in the genotype space via a path where the fitness increases monotonically defined as an *accessible path* [11, 68]. *Landscape navigability* has previously been defined as the proportion of accessible paths to a given genotype from all other genotypes [17]. To briefly summarise, here we specifically define the navigability as the average probability that a randomly chosen phenotype pair have at least one accessible path between them, given a fitness assignment process to phenotypes. We denote accessibility with ψ, where $\psi = 1$ indicates the presence of at least one accessible path between two phenotypes for a specific set of fitness assignments, and $\psi = 0$ indicating no accessible paths. When $\psi = 0$, a fitness valley must be traversed between the phenotypes. With this notation, we use $\langle\psi\rangle$ to represent navigability of fitness landscapes for a given GP map.

2. Fitness landscapes

In conjunction with the GP map M, a fitness landscape instance is defined by the set of phenotype fitnesses $\mathcal{F} := \{F_{p_i}\}_{i=1}^{N_P}$, with i denoting the i^{th} indexed phenotype p_i. We refer to the *source* phenotype p and *target* phenotype q in the search for an accessible path from $p \to q$. We consider two fitness assignments in this paper:

- *Random fitness*: random samples $F_{p_i} \sim \text{Uniform}(0, 1)$ with target phenotype q having $F_q = 1$
- *Hamming distance*: where the similarity of phenotype p compared to a phenotype q is measured by the number of matching positions in the aligned phenotype string representation given by $F(p, q) = 1 - \sum_j^L \delta(p^{(j)}, q^{(j)})/L$, where $p^{(j)}$ is the string character representing phenotype p at the j^{th} base position and $F(p, q)$ is the fitness of phenotype p compared to a target phenotype q

$F_{del} = 0$ for all fitness assignments.

3. Navigability estimation

The probability of an accessible path ($\psi = 1$) between a source phenotype p and target phenotype q, given a random fitness landscape instance $\mathcal{F}$, is deterministic with a binary outcome. We can define the probability of ψ more explicitly as a function of p, q and $\mathcal{F}$ as follows:

$$\psi(p, q, \mathcal{F}) := P(\psi = 1 | p \to q, \mathcal{F}) \tag{2}$$

where

$$\psi(p, q, \mathcal{F}) = \begin{cases} 1 & \text{if at least one accessible path exists} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

We can take the expectation over $\mathcal{F}$ yielding the mean probability of an accessible path from p to q as:

$$\psi_{pq} = E_{\mathcal{F}}[\psi(p, q, \mathcal{F})] \tag{4}$$

With this notation, we can define the *navigability* for the GP map as the expectation over Eq. (4) for phenotypes p and q sampled uniformly at random:

$$\langle\psi\rangle = E_{p,q}[\psi_{pq}] \tag{5}$$

We can estimate this probability of reaching a given target phenotype q from a uniform randomly chosen source phenotype p by computationally measuring $\psi(p, q, \mathcal{F})$ for N_s randomly chosen sources for each of N_t randomly chosen targets, with a new random fitness landscape instance $\mathcal{F}$ for each pair. During the practical estimation, it is convenient to understand the outcome of the search as:

$$\psi(p_{st}, q_t, \mathcal{F}_{st}) = \begin{cases} 1 & \text{at least one accessible path} \\ 0 & \text{no accessible path, not aborted} \\ \text{NA} & \text{no accessible path, aborted} \end{cases}$$

where searches are aborted if they extend beyond a computational threshold of genotypes encountered T. An estimate of the navigability $\langle\psi\rangle$ can be written as:

$$\langle\psi\rangle = \frac{1}{N_c} \sum_{t=1}^{N_t} \sum_{s=1}^{N_s} I_T(s, t)\psi(p_{st}, q_t, \mathcal{F}_{st}) \tag{6}$$

where p_{st} and q_t are the source and target phenotypes of s^{th} source for the t^{th} target, with $I_T(s,t) := I\left(\psi(p_{st}, q_t, \mathcal{F}_{st}) \neq \text{NA}\right)$ an indicator for whether the run was not aborted, and therefore the number of completed runs is $N_c = \sum_{t,s} I_T\left(s,t\right)$ with the aborted proportion α:

$$\alpha = 1 - \frac{N_c}{N_t N_s} \tag{7}$$

The estimate of the navigability of a fitness landscape with GP map has an associated Bernoulli standard error (derived from an estimate of the corrected sample standard deviation):

$$SE(\langle\psi\rangle) = \sqrt{\frac{\langle\psi\rangle\left(1-\langle\psi\rangle\right)}{N_c - 1}} \tag{8}$$

We next describe in more detail the computational algorithm for estimating $\langle\psi\rangle$.

4. Navigability estimation algorithm

For a given source and target phenotype, in each random landscape instance, we perform the following computational algorithm to measure ψ. We first provide some definitions:

- GP map M: is a function $M : \mathcal{G} \to \mathcal{P}$ where $\mathcal{G}$ is the space of genotypes and $\mathcal{P}$ is the space of phenotypes, such that we can write the phenotype p of genotype g as $p = M(g)$
- Dimensionality: We define the set of sequence positions that may be mutated as $\mathcal{D}$, with the size of $|\mathcal{D}|$ being the dimensionality D. When $|\mathcal{D}| = L$ all base positions are mutable. Relative dimensionality is defined as the dimensionality relative to sequence length $d = D/L$
- Alphabet: sequences have a set of $\mathcal{A}$ possible letters at a given site. The size of $|\mathcal{A}| = K$ is the base.
- u_0 contains genotypes whose 1-mutant neighbours are yet to be considered in a given search for an accessible path
- u_1 contains genotypes that have already had their 1-mutant neighbours considered in a given search for an accessible path

The algorithm proceeds with a Breadth First Search (BFS):

1. A random genotype g that maps to the source phenotype is chosen and added to u_0
2. Set the first element of u_0 as g
3. For base $a \in \mathcal{A}$ at position j and for each position $j \in \mathcal{D}$, measure genotype neighbour g' and phenotype $p' = M(g')$
4. If $F_{p'} \geq F_p$ and $g' \notin u_1$, add g' to u_0
5. Move g from u_0 to u_1
6. If $|u_0| = 0$ or $|u_0|+|u_1| > T$ (computational threshold) or the target phenotype is found, return 'aborted' or ψ respectively. Otherwise return to step 2

The algorithm finishes with either u becoming empty, or the combined size of u_0 and u_1 becoming larger than a predefined threshold T (introduced in Section IV B 1), beyond which computational progress may become unfeasible. We discard these aborted runs from the measurement of navigability $\langle\psi\rangle$ using the indicator function I_T of the previous section (Section IV B 3).

As described in Eq. (6) we pick N_s source phenotypes uniformly at random for each of the N_t target phenotypes also chosen at random. We set $N_t = 20$ and $N_s = 50$. The uncertainty in the estimate of the navigability $\langle\psi\rangle$ is reported as the standard error $SE(\langle\psi\rangle)$ across the ensemble of measurements.

C. Removing correlations

In order to measure the effect of positive neutral correlations [30], we perform genotype swaps and then repeat the measurement of $\langle\psi\rangle$. This process involves constructing a new GP map M_s from the original GP map $M_{s=0} := M$ where s is the number of pairs of genotypes whose phenotype's have been swapped. More precisely, a swap involves selecting two genotypes g_1 and g_2 with uniform random probability and setting $M_s(g_1) = M_{s-1}(g_2)$ and $M_s(g_2) = M_{s-1}(g_1)$. It follows that $M_{s\to\infty}$ is the uncorrelated random null model GP map with no positive neutral correlations as used in ref. [30]. As shown in ref. [30], the random null model has $\rho_p \approx f_p$ when there are no positive neutral correlations. Therefore, we additionally define the correlations c present in a given GP map M_s by comparing the logarithm of the average robustness-to-frequency ratio in a given GP map against the original GP map, generating a scale for measuring correlations in M_s:

$$c(s) = \frac{\log_{10}\left\langle \frac{\rho_p(s)}{f_p(s)} \right\rangle_p}{\log_{10}\left\langle \frac{\rho_p(0)}{f_p(0)} \right\rangle_p} \tag{9}$$

where for $s = 0$ we have $c(0) = 1$, and for $\lim_{s\to\infty} c(s) \approx 0$ the expectation for the random model. Therefore, the scale yields positive values for c where there is, on average, greater robustness than frequency. The process of removing correlations gradually from the original GP map ($s = 0$) to the random null model ($s \to \infty$) provides a range over which the relationship between positive neutral correlations and navigability may be considered in GP maps. We measure the navigability of $S_{2,8}$, RNA12, HP3x3x3 and HP5x5 by taking 100 evenly spaced values for s on the range $s = [0, K^L]$ and measuring $\langle\psi\rangle$ and $c(s)$ for each.

D. Restricting dimensionality

To measure the role of dimensionality we restrict the dimensionality of a search for an accessible path from source to target by only allowing a set of $\mathcal{D}$ randomly chosen positions along the sequence to be mutated in the 1-mutant neighbour measurement in Step 3 of the navigability algorithm above (Section IV B 4). The dimensionality D is the number of positions that may be mutated $|\mathcal{D}|$, and the relative dimensionality $d := D/L$. When $D = L$ we have the original dimensionality, while for $D = 1$ only a single sequence position may be mutated. The GP map M itself is not changed under this dimensional restriction but rather the connectivity of genotypes and therefore the connectivity of the fitness landscape.

We measure the navigability of $S_{2,8}$, RNA12, HP3x3x3 and HP5x5 by taking evenly spaced values for D on the range $D \in [1, L]$.

E. Measuring ruggedness

For fitness landscapes, related to navigability is the concept of landscape *ruggedness*. We measure $\kappa(g)$, whether a genotype is a local fitness maximum, during the search from source to target. The average proportion of genotypes that are local fitness maxima provides a measure of ruggedness [26]. Whether a genotype g is a local fitness peak is determined by the fitness of all accessible 1-mutant neighbours g', such that:

$$\kappa(g) = \begin{cases} 1 & \text{if } F_{M(g')} < F_{M(g)} \forall g' \in \sigma(g) \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

where we have the function $\sigma(g)$ which returns the set of 1-mutants of genotype g. We calculate the ruggedness for a landscape by taking the average of $\kappa(g)$ over all genotypes and all source-target pairs once the search has completed. We denote the ruggedness as $\langle\kappa\rangle$.

F. Navigability in the functional RNA database

In Section II C, we examine navigability in a specific subset of RNA phenotypes, namely those that are found in the functional RNA database (fRNAdb) [39]. For a given length we use all phenotypes in proportion to their occurrence in the fRNAdb apart from the trial structure which we exclude as it is assigned zero fitness here. We randomly choose $N_t = 50$ targets with $N_s = 20$ randomly chosen sources from this set.

In order to examine navigability between functional RNAs, we must consider sequences longer than $L = 15$. In doing so, we introduce additional computational overhead given the increasing neutral set size resulting in the condition $|u_0|+|u_1| > T$ being more likely to be met. Therefore to maximise the number of non-aborted runs, we perform a modified Depth-First Search (DFS) where we attempt to greedily follow paths of increasing gradient until we reach the max fit phenotype. If the path fails, instead of moving back one step as in a standard DFS, we go all the way back to the start of the walk and pick an unexplored neighbour with the lowest fitness to begin a new uphill walk. In this way, we maximise the exploration of new phenotypes by always starting our deep walks from the lowest point while still maintaining the ability to perform long walks during the search.

We write the modified DFS algorithm explicitly as:

1. A random genotype g that maps to the source phenotype is chosen and added to u_0.
2. Set the first element of u_0 as g, and $p = M(g)$
3. For each alternative base $a \in \mathcal{A}$ at position j and for each position j in $\mathcal{D}$, measure genotype neighbour g' and phenotype $p' = M(g')$
4. If any g' has $F_{p'} > F_p$ and $g' \notin u_1$ and $g' \notin u_0$, add g' to front of u_0 and return to step 2
5. If any g' have $p = p'$ and $|u_0| = 1$, add one such neutral case to the back of u_0 if $g' \notin u_0$ and $g' \notin u_1$
6. Move g from u_0 to u_1
7. If $|u_0| = 0$ or $|u_0|+|u_1| > T$ (computational threshold) or the target phenotype is found, return 'aborted' or ψ respectively. Otherwise return to step 2.

We note that for searches where neutral mutations are not permitted as part of the search, step 5 of the above is omitted.

In terms of computational time, on a single Intel Xeon core at 2.8GHz a single search for a target with $T = 2 \times 10^6$ took on average 0.9 minutes for $L = 20$, 1.3 hours for $L = 30$ and 19.1 hours for $L = 40$. With $T = 2 \times 10^4$, the times were on average 0.1 minutes for $L = 20$, 3.0 minutes for $L = 30$ and 19.5 minutes for $L = 40$.

G. Navigability estimation under evolutionary dynamics

We measured fitness landscape navigability as the average probability that a given source-target pair could be connected by way of an accessible path. We extend this definition to the more strict requirement of *evolutionary navigability* where the evolutionary dynamics of a population is considered instead of just the existence of an accessible path in crossing the fitness landscape.

1. Monomorphic evolutionary dynamics

We model monomorphic evolutionary dynamics with a sequential fixation model [56], assuming that the rate of mutation is much less than the time it takes for mutants to reach fixation once they have arisen. Under this model, the sequence of fixation can be treated as a Markov chain, with the adaptive path of the population essentially following a biased random walk.

Following the formalism of ref. [56], and assuming that the neighbouring genotypes $\sigma(g)$ of genotype g will be produced

at equal rates, the probability that mutant genotype h will be the next to fix after genotype g, is given by:

$$P(g,h) = \frac{P_{\text{fix}}(s(h,g),N)}{\sum_{g' \in \sigma(g)} P_{\text{fix}}(s(g',g),N)} \tag{11}$$

where the probability P_{fix} that a given mutant arising in a haploid population of size N is given by Kimura's equation [57]:

$$P_{\text{fix}}(s,N) = \frac{1-\exp(-2s)}{1-\exp(-2Ns)} \tag{12}$$

with $s(g',g) = F_{g'}/F_g - 1$ as the relative fitness of genotype g' to genotype g.

To computationally implement these dynamics for a given source-target pair of phenotypes p and q respectively, with fitness assignment function F (either random or Hamming, Section IV B 2), we perform the following algorithm up to a limit of T iterations:

1. Set genotype g as the source genotype and its phenotype p corresponding to randomly chosen entry from fRNA database, calculating its fitness F_p using the fitness assignment function
2. For each neighbouring genotype g' in the set $\sigma(g)$ of neighbours of g, calculate their phenotype $p' = M(g')$
3. Calculate the fitness of each neighbour $F_{p'}$ and the $P_{\text{fix}}(s(g',g),N)$
4. Randomly choose a neighbour genotype g' in proportion to $P_{\text{fix}}(s(g',g),N)$
5. Set $g \leftarrow g', t \leftarrow t+1$
6. Return to step 2 if $M(g') \neq q$ and $t < T$. Otherwise terminate.

We performed the evolutionary search for $N_s = 20$ sources for each of $N_t = 50$ targets randomly chosen from the fRNA database at lengths $L = 20, 30, 40$, with both random and Hamming distance fitness assignment (Section IV B 2. A computational limit of $T = 50,000$ sequential fixations was used. On a single Intel Xeon core at 2.8GHz a single search from source to target took on average 0.4 minutes for $L = 20$, 5.1 minutes for $L = 30$ and 30.7 minutes for $L = 40$.

2. *Non-monomorphic evolutionary dynamics*

For non-monomorphic evolutionary dynamics, we modelled the evolutionary process using Wright-Fisher dynamics [69, 70]. This directly involved simulating a population of N genotypes and updating this population every generation with genotypes chosen for reproduction in proportion to their fitness, with point mutations applied.

For a given source-target pair of phenotypes p and q respectively, we use the following algorithm:

1. Set genotype g as the source genotype and its phenotype p corresponding to randomly chosen entry from the fRNA database
2. Make N copies of g constructing the population $\Gamma_{t=0}$ at time $t = 0$
3. For subsequent times $0 < t \leq T$ (with T as the computational limit/maximum number of generations), we repeat the following:
   (a) For i^{th} genotype g_i of the population Γ_t, calculate the phenotype p_i and its fitness F_{p_i}
   (b) Sample N genotypes at random with probability $F_{p_i}/\sum_k F_{p_k}$ with replacement from Γ_t, constructing a temporary population of genotypes Γ'_t
   (c) For each base position j for each genotype i of Γ'_t, apply a random mutation with Bernoulli probability μ (point mutation rate). Where a mutation is applied to g_{ij}, a random alternative base to the current is chosen from $\{A,C,G,U\}\backslash g_{ij}$ with uniform probability.
   (d) Set the population at time $t+1$ from the mutated temporary population: $\Gamma_{t+1} \leftarrow \Gamma'_t$

We performed the evolutionary search for $N_s = 20$ sources for each of $N_t = 50$ targets randomly chosen from the fRNA database at lengths $L = 20, 30, 40$, with both random and Hamming distance fitness assignment (Section IV B 2. A computational limit of $T = 20,000$ generations and a population size of $N = 100$ was used. Population mutation rates of $N\mu L = 1$ (intermediate), $N\mu L = 10$ and $N\mu L = 100$ (polymorphic as $N\mu L \gg 1$) were investigated.

On a single Intel Xeon core at 2.8GHz, a single simulation of a population of size $N = 100$ for $T = 10,000$ generations took on average 1.3 minutes for $L = 20$, 5.6 minutes for $L = 30$ and 16.7 minutes for $L = 40$.

3. *Estimating evolutionary navigability*

To quantify navigability under evolutionary dynamics we need to define the *adaptive path* from source to target. For monomorphic evolutionary dynamics, this is the genotypes (and their corresponding phenotype's and fitness) along the Markov chain of sequential fixations. For non-monomorphic evolutionary dynamics, we measure whether the population Γ_t has a majority phenotype with proportion greater than 50%, otherwise recording a null value, leading to a sequence of majority phenotypes and their corresponding fitnesses during the search. An accessible path is an adaptive path that reaches the target with monotonic fitness changes along the adaptive path. We defined *evolutionary navigability* for a given GP map as the average probability that an adaptive path was an accessible path given the evolutionary dynamics, GP map and fitness assignment.

To estimate this computationally, we record two binary properties of the search: 1) whether the target was discovered

('Successful') 2) whether the adaptive path only increased in fitness ('Monotonic'). We record whether the population took an accessible path by enumerating the cases:

$$\psi^{\text{evo}} = \begin{cases} 1 & \text{Successful AND Monotonic} \\ 0 & \text{Successful AND not Monotonic} \\ 0 & \text{not Successful AND not Monotonic} \\ \text{NA} & \text{not Successful AND Monotonic} \end{cases}$$

with the evolutionary navigability then estimated over a k-indexed ensemble of searches as:

$$\langle \psi^{\text{evo}} \rangle = \frac{1}{N_c} \sum_k \psi^{\text{evo}}_{s_k,t_k} I\left(\psi^{\text{evo}}_{s_k,t_k} \neq \text{NA}\right) \tag{13}$$

with $N_c = \sum_k I\left(\psi^{\text{evo}}_{s_k,t_k} \neq \text{NA}\right)$ counting the searches where it is certain that the search will be via an accessible path or not. As in Eq. (7), the proportion of searches aborted is given by $\alpha = 1 - \frac{N_c}{N_t N_s}$. We additionally define the *phenotypic evolutionary navigability* $\left\langle \psi^{\text{evo}}_p \right\rangle$ for an individual phenotype p as an ensemble where $t_k = p \, \forall k$, such that:

$$\left\langle \psi^{\text{evo}}_p \right\rangle = \frac{1}{N_c} \sum_k \psi^{\text{evo}}_{s_k,p} I\left(\psi^{\text{evo}}_{s_k,p} \neq \text{NA}\right) \tag{14}$$

providing a means to investigate the distribution conditional on the target phenotype p as well as overall navigability.

V. DATA AVAILABILITY

The dataset containing functional RNA (fRNAdb) used in this paper is available at: `https://doi.org/10.18908/lsdba.nbdc00452-001`. The GP maps analysed are available under Section VI.

VI. CODE AVAILABILITY

The ViennaRNA package (1.8.5), RNAshape package `https://anaconda.org/bioconda/rnashapes` and custom C++ and Python source code was used to construct GP maps and perform computational simulations. The source code is available at: `https://github.com/sgreenbury/gp-maps-nav`.

VII. ACKNOWLEDGEMENTS

The authors would like to thank Marcel Weiß for helpful discussions and insights.

VIII. AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: SFG, AAL, SEA. Performed the experiments: SFG. Analysed the data: SFG, AAL, SEA. Supervised the work: SEA. Wrote the paper: SFG, AAL, SEA.

[1] Sewall Wright. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. In *Proceedings 6th International Congress on Genetics*, volume 1, pages 356–366, 1932.

[2] Stuart A. Kauffman. *The origins of order: Self-organization and selection in evolution*. Oxford university press, 1993.

[3] Erik Svensson and Ryan Calsbeek. *The adaptive landscape in evolutionary biology*. Oxford University Press, 2012.

[4] Massimo Pigliucci. Landscapes, surfaces, and morphospaces: what are they good for. *The adaptive landscape in evolutionary biology*, pages 26–38, 2012.

[5] J Arjan GM de Visser and Joachim Krug. Empirical fitness landscapes and the predictability of evolution. *Nature Reviews Genetics*, 15(7):480–490, 2014.

[6] Inês Fragata, Alexandre Blanckaert, Marco António Dias Louro, David A Liberles, and Claudia Bank. Evolution in the light of fitness landscape theory. *Trends in ecology & evolution*, 34(1):69–82, 2019.

[7] Ronald Aylmer Fisher. *The genetical theory of natural selection*. Clarendon Press, 1958.

[8] Robert M May. *Stability and complexity in model ecosystems*. Princeton university press, 1973.

[9] Michael Conrad and Werner Ebeling. M.V. Volkenstein, evolutionary thinking and the structure of fitness landscapes. *BioSystems*, 27(3):125–128, 1992.

[10] Sergey Gavrilets. *Fitness landscapes and the origin of species (MPB-41)*. Princeton University Press, 2004.

[11] Jasper Franke, Alexander Klözer, J. Arjan G. M. de Visser, and Joachim Krug. Evolutionary Accessibility of Mutational Pathways. *PLoS Computational Biology*, 7(8):e1002134, 2011.

[12] Suman G. Das, Susana O.L. Direito, Bartlomiej Waclaw, Rosalind J. Allen, and Joachim Krug. Predictable properties of fitness landscapes induced by adaptational tradeoffs. *eLife*, 9:1–24, 2020.

[13] Daniel M. Weinreich, Richard A. Watson, and Lin Chao. Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution*, 59(6):1165–1174, 2005.

[14] Maurício Carneiro and Daniel L. Hartl. Adaptive landscapes and protein evolution. *Proceedings of the National Academy of Sciences*, 107(suppl. 1):1747–1751, 2010.

[15] Nicholas C. Wu, Lei Dai, C. Anders Olson, James O. Lloyd-Smith, and Ren Sun. Adaptation in protein fitness landscapes is facilitated by indirect paths. *eLife*, 5:1–21, 2016.

[16] Claudia Bank, Sebastian Matuszewski, Ryan T. Hietpas, and Jeffrey D. Jensen. On the (un)predictability of a large intragenic fitness landscape. *Proceedings of the National Academy of Sciences of the United States of America*, 113(49):14085–14090,

2016.

[17] José Aguilar-Rodríguez, Joshua L. Payne, and Andreas Wagner. A thousand empirical adaptive landscapes and their navigability. *Nature Ecology & Evolution*, 1(2):0045, 2017.

[18] Júlia Domingo, Guillaume Diss, and Ben Lehner. Pairwise and higher-order genetic interactions during the evolution of a tRNA. *Nature*, 558(7708):117–121, 2018.

[19] Jia Zheng, Joshua L Payne, and Andreas Wagner. Cryptic genetic variation accelerates evolution by opening access to diverse adaptive peaks. *Science*, 365(6451):347–353, 2019.

[20] Victoria O. Pokusaeva, Dinara R. Usmanova, Ekaterina V. Putintseva, Lorena Espinar, Karen S. Sarkisyan, Alexander S. Mishin, Natalya S. Bogatyreva, Dmitry N. Ivankov, Arseniy V. Akopyan, Sergey Ya Avvakumov, Inna S. Povolotskaya, Guillaume J. Filion, Lucas B. Carey, and Fyodor A. Kondrashov. An experimental assay of the interactions of amino acids from orthologous sequences shaping a complex fitness landscape. *PLOS Genetics*, 15(4):e1008079, 2019.

[21] Andreas Wagner. *Life Finds a Way: What Evolution Teaches Us about Creativity*. Oneworld Publications, 2019.

[22] Frank J Poelwijk, Daniel J Kiviet, Daniel M Weinreich, and Sander J Tans. Empirical fitness landscapes reveal accessible evolutionary paths. *Nature*, 445(7126):383–386, 2007.

[23] Alexander E Lobkovsky and Eugene V Koonin. Replaying the tape of life: quantification of the predictability of evolution. *Frontiers in genetics*, 3:246, 2012.

[24] Daniel L. Hartl. What can we learn from fitness landscapes? *Current Opinion in Microbiology*, 21:51–57, 2014.

[25] Ard A Louis. 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*, 58:107–116, 2016.

[26] Stuart Kauffman and Simon Levin. Towards a general theory of adaptive walks on rugged landscapes. *Journal of Theoretical Biology*, 128(1):11 – 45, 1987.

[27] Marcin Zagorski, Zdzislaw Burda, and Bartlomiej Waclaw. Beyond the hypercube: evolutionary accessibility of fitness landscapes with realistic mutational networks. *PLoS computational biology*, 12(12):e1005218, 2016.

[28] John FC Kingman. A simple model for the balance between selection and mutation. *Journal of Applied Probability*, 15(1):1–12, 1978.

[29] Bjørn Østman and Christoph Adami. Predicting Evolution and Visualizing High-Dimensional Fitness Landscapes. In *Recent Advances in the Theory and Application of Fitness Landscapes*, pages 509–526. Springer, 2014.

[30] Sam F. Greenbury, Steffen Schaper, Sebastian E. Ahnert, and Ard A. Louis. Genetic Correlations Greatly Increase Mutational Robustness and Can Both Reduce and Enhance Evolvability. *PLoS Computational Biology*, 12(3):1–27, 2016.

[31] S. E. Ahnert. Structural properties of genotype–phenotype maps. *Journal of The Royal Society Interface*, 14(132):20170275, 2017.

[32] Susanna Manrubia, José A. Cuesta, Jacobo Aguirre, Sebastian E. Ahnert, Lee Altenberg, Alejandro V. Cano, Pablo Catalán, Ramon Diaz-Uriarte, Santiago F. Elena, Juan Antonio García-Martín, Paulien Hogeweg, Bhavin S. Khatri, Joachim Krug, Ard A. Louis, Nora S. Martin, Joshua L. Payne, Matthew J. Tarnowski, and Marcel Weiß. From genotypes to organisms: State-of-the-art and perspectives of a cornerstone in evolutionary dynamics. *Physics of Life Reviews*, 38:55–106, 2021.

[33] Erik van Nimwegen, James P. Crutchfield, and Martijn Huynen. Neutral evolution of mutational robustness. *Proceedings of the National Academy of Sciences*, 96(17):9716–9720, 1999.

[34] S F Greenbury and S E Ahnert. The organization of biological sequences into constrained and unconstrained parts determines fundamental properties of genotype–phenotype maps. *Journal of The Royal Society Interface*, 12(113):20150724, 2015.

[35] Susanna Manrubia and José A. Cuesta. Distribution of genotype network sizes in sequence-to-structure genotype-phenotype maps. *Journal of the Royal Society Interface*, 14(129), 2017.

[36] Marcel Weiß and Sebastian E Ahnert. Phenotypes can be robust and evolvable if mutations have non-local effects on sequence constraints. *Journal of The Royal Society Interface*, 15(138):20170618, 2018.

[37] Andreas Wagner. Robustness and evolvability: a paradox resolved. *Proceedings of the Royal Society B: Biological Sciences*, 275(1630):91–100, 2008.

[38] Steffen Schaper, Iain G. Johnston, and Ard A. Louis. Epistasis can lead to fragmented neutral spaces and contingency in evolution. *Proceedings of the Royal Society B: Biological Sciences*, 279(1734):1777–1783, 2012.

[39] Taishin Kin, Kouichirou Yamada, Goro Terai, Hiroaki Okida, Yasuhiko Yoshinari, Yukiteru Ono, Aya Kojima, Yuki Kimura, Takashi Komori, and Kiyoshi Asai. fRNAdb: A platform for mining/annotating functional RNA candidates from non-coding RNA sequences. *Nucleic Acids Research*, 35(SUPPL. 1):145–148, 2007.

[40] Peter Schuster, Walter Fontana, Peter F. Stadler, and Ivo L. Hofacker. From sequences to shapes and back: A case study in RNA secondary structures. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 255(1344):279–284, 1994.

[41] I.L. Hofacker, W. Fontana, P.F. Stadler, L.S. Bonhoeffer, M. Tacker, and P. Schuster. Fast folding and comparison of rna secondary structures. *Monatshefte für Chemie/Chemical Monthly*, 125(2):167–188, 1994.

[42] Walter Fontana. Modelling 'evo-devo' with RNA. *BioEssays*, 24(12):1164–1177, 2002.

[43] Matthew C. Cowperthwaite, Evan P. Economo, William R. Harcombe, Eric L. Miller, and Lauren Ancel Meyers. The ascent of the abundant: How mutational networks constrain evolution. *PLoS Comput Biol*, 4(7):e1000110, 07 2008.

[44] Jacobo Aguirre, Javier M. Buldú, Michael Stich, and Susanna C. Manrubia. Topological structure of the space of phenotypes: The case of RNA neutral networks. *PLoS ONE*, 6(10):e26324, 10 2011.

[45] Steffen Schaper and Ard A. Louis. The arrival of the frequent: How bias in genotype-phenotype maps can steer populations to local optima. *PLoS ONE*, 9(2):e86635, 02 2014.

[46] Andreas Wagner. *The origins of evolutionary innovations: a theory of transformative change in living systems*. Oxford University Press, 2011.

[47] Sam F. Greenbury, Iain G. Johnston, Ard A. Louis, and Sebastian E. Ahnert. A tractable genotype–phenotype map modelling the self-assembly of protein quaternary structure. *Journal of The Royal Society Interface*, 11(95), 2014.

[48] Iain G Johnston, Kamaludin Dingle, Sam F Greenbury, Chico Q Camargo, Jonathan PK Doye, Sebastian E Ahnert, and Ard A Louis. Symmetry and simplicity spontaneously emerge from the algorithmic nature of evolution. *Proceedings of the National Academy of Sciences*, 119(11):e2113883119, 2022.

[49] Ken A. Dill. Theory for the folding and stability of globular proteins. *Biochemistry*, 24(6):1501–1509, 1985.

[50] Anders Irbäck and Carl Troein. Enumerating designing sequences in the HP model. *Journal of Biological Physics*, 28(1):1–15, 2002.

[51] Evandro Ferrada and Andreas Wagner. A comparison of genotype-phenotype maps for RNA and proteins. *Biophysical Journal*, 102(8):1916 – 1925, 2012.

[52] Thomas Jörg, Olivier Martin, and Andreas Wagner. Neutral network sizes of biological RNA molecules can be computed and are not atypically small. *BMC Bioinformatics*, 9(1):464, 2008.

[53] Ingwer Borg and Patrick JF Groenen. *Modern multidimensional scaling: Theory and applications*. Springer Science & Business Media, 2005.

[54] Kamaludin Dingle, Fatme Ghaddar, Petr Šulc, and Ard A Louis. Phenotype bias determines how natural RNA structures occupy the morphospace of all possible shapes. *Molecular Biology and Evolution*, 2021. msab280.

[55] Kamaludin Dingle, Steffen Schaper, and Ard A. Louis. The structure of the genotype–phenotype map strongly constrains the evolution of non-coding RNA. *Interface Focus*, 5(6):20150053, 2015.

[56] David M. McCandlish and Arlin Stoltzfus. Modeling Evolution Using the Probability of Fixation: History and Implications. *The Quarterly Review of Biology*, 89(3):225–252, 2014.

[57] Motoo Kimura. ON THE PROBABILITY OF FIXATION OF MUTANT GENES IN A POPULATION. *Genetics*, 47(6):713–719, 1962.

[58] Robert Giegerich, Björn Voß, and Marc Rehmsmeier. Abstract shapes of RNA. *Nucleic Acids Research*, 32(16):4843–4851, 2004.

[59] Kamaludin Dingle, Fatme Ghaddar, Petr Šulc, and Ard A Louis. Phenotype Bias Determines How Natural RNA Structures Occupy the Morphospace of All Possible Shapes. *Molecular Biology and Evolution*, 39(1):1–11, jan 2022.

[60] Nora S Martin and Sebastian E Ahnert. Insertions and deletions in the RNA sequence–structure map. *Journal of The Royal Society Interface*, 18(183), 2021.

[61] Pablo Catalán, Clemente F. Arias, Jose A. Cuesta, and Susanna Manrubia. Adaptive multiscapes: An up-to-date metaphor to visualize molecular adaptation. *Biology Direct*, 12(1):1–15, 2017.

[62] C. Brandon Ogbunugafor, C. Scott Wylie, Ibrahim Diakite, Daniel M. Weinreich, and Daniel L. Hartl. Adaptive Landscape by Environment Interactions Dictate Evolutionary Dynamics in Models of Drug Resistance. *PLoS Computational Biology*, 12(1):1–20, 2016.

[63] Daniel Nichol, Mark Robertson-Tessi, Alexander R.A. Anderson, and Peter Jeavons. Model genotype–phenotype mappings and the algorithmic structure of evolution. *Journal of the Royal Society Interface*, 16(160), 2019.

[64] Ramon Diaz-Uriarte. Cancer progression models and fitness landscapes: A many-to-many relationship. *Bioinformatics*, 34(5):836–844, 2018.

[65] Calum Gabbutt and Trevor A. Graham. Evolution's cartographer: Mapping the fitness landscape in cancer. *Cancer Cell*, 39(10):1311–1313, 2021.

[66] Kit Fun Lau and Ken A. Dill. A lattice statistical mechanics model of the conformational and sequence spaces of proteins. *Macromolecules*, 22(10):3986–3997, 1989.

[67] Hao Li, Robert Helling, Chao Tang, and Ned Wingreen. Emergence of preferred structures in a simple model of protein folding. *Science*, 273(5275):666–669, 1996.

[68] Daniel M Weinreich. Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins. *Science*, 312(5770):111–114, 2006.

[69] Warren John Ewens. *Mathematical Population Genetics: Theoretical Introduction*, volume 1. Springer, New York, NY, 2004.

[70] Lorens A. Imhof and Martin A. Nowak. Evolutionary game dynamics in a wright-fisher process. *Journal of Mathematical Biology*, 52(5):667–681, 2006.

[71] Vladimir I Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet physics doklady*, 10(8):707–710, 1966.

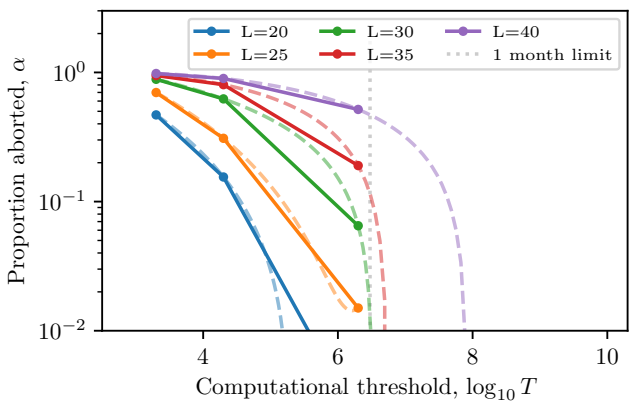


FIG. 7. **Proportion of estimations aborted for four different thresholds for different fRNA length L.** Dashed lines provide quadratic interpolations to illustrate potential computational thresholds for which a given abortion threshold may be reached if the fit holds for extrapolation. As a guide, we highlight the computational limit corresponding to one month of chronological time given available computational resources.

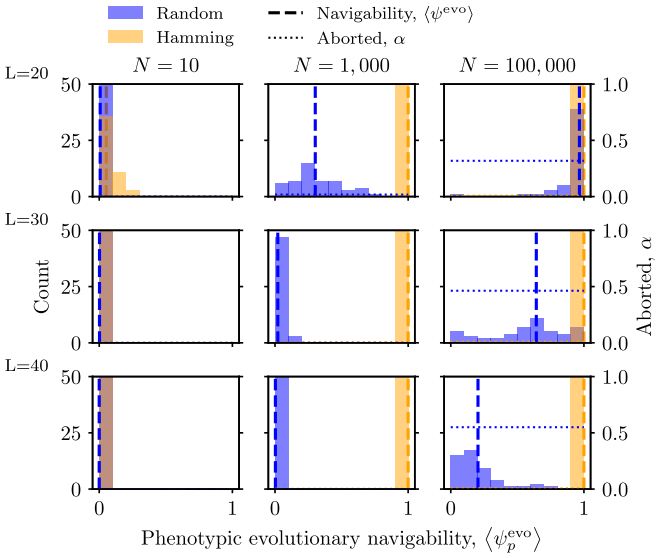


FIG. 8. **Navigability under monomorphic evolutionary dynamics for fRNA $L = 20, 30, 40$ (rows) for random and Hamming fitness assignments at different population sizes $N = 10, N = 1,000, N = 100,000$ (columns).** Blue shaded histogram show the proportion of successful searches for random fitness assignment, and the orange histogram for Hamming fitness assignment. Mean navigability is shown as vertical dashed lines and the proportion of aborted searches as horizontal dotted lines. Evolutionary navigability increases with increasing population size, is found under both random and Hamming fitness assignment, but with Hamming fitness assignment providing greater navigability at $N = 1,000$ and for $N = 100,000$ with fewer aborted searches.

Appendix A: Impact of computational thresholds on discovery of estimation of navigability

To allow us to consider the plausibility of navigable landscapes for longer fRNA ($L > 20$), we explore the effect of changing the computational threshold T (Section IV B 1) at which the search for an accessible path is aborted. We test four orders of magnitude for the threshold $|u_0| + |u_1| < T$ condition: $N_{\rm thresh} = \{2 \times 10^3, 2 \times 10^4, 2 \times 10^5$ and $2 \times 10^6\}$. In each case we attempt $N_t = 50$ target phenotypes and for each target $N_s = 20$ source phenotypes and attempt to identify an accessible path, where we record whether a search was successful, unsuccessful or aborted.

In Fig. 7 the proportion of runs that are aborted respectively for the different thresholds against the length of the fRNA sequences. The change in the proportion of aborted runs is pertinent for understanding both how navigability changes when increasing the threshold and also what level of T is required to be able to reasonably estimate navigability for a given length L. For all lengths L and thresholds T, navigability $\langle \psi \rangle \approx 1$, showing that almost all non-aborted runs have accessible paths. Extrapolating this observation we should expect high navigability for longer length $L > 30$ if greater computation resource were available. With respect to the required computational thresholds for a given length L, we observe, very roughly, that around 50% aborted proportion is reached for $L = 20$ at $T = 2 \times 10^3$, for $L = 35$ at $T = 2 \times 10^5$ and $L = 40$ at $T = 2 \times 10^6$. Extrapolating with quadratic fits we could hypothesise that the aborted threshold could be reduced to 10% for $L = 40$ at between $[2 \times 10^7, 2 \times 10^8]$.

Appendix B: The effect of population size on navigability in monomorphic evolutionary dynamics

In Section II C 2, we considered monomorphic evolutionary dynamics with $N = 100,000$. Here, we examine the dependence of evolutionary navigability on population size. Fig. 8 shows results for random and Hamming fitness assignment at $N = 10$; $1,000$; $100,000$ demonstrating that population size critically determines navigability. Small population sizes ($N = 10$) show very low navigability with the emergence of higher navigability at $N = 1,000$ and $100,000$. Remarkably, for Hamming fitness, $\langle \psi^{\rm evo} \rangle = 1$ for $N = 1,000$ and $N = 100,000$ for $L = 20, 30, 40$ and negligible aborted searches with $\alpha < 0.010$ across all cases. For random fitness assignment, the increase in navigability with increasing population size is less stark: at $N = 1,000$ the navigability for $L = 20$ is non-negligible at $\langle \psi^{\rm evo} \rangle = 0.302$, with a gradual reduction observed ($\langle \psi^{\rm evo} \rangle = 0.021$ for $L = 30$ and $\langle \psi^{\rm evo} \rangle = 0.003$ for $L = 40$). However, for $N = 100,000$ at $L = 20, 30, 40$, the random fitness assignment also has sizeable navigability with $\langle \psi^{\rm evo} \rangle = 0.966, 0.642, 0.204$ respectively. However, it is notable that an increasing proportion are aborted ($\alpha = 0.318, 0.463, 0.550$ at $L = 20, 30, 40$).

The increase in navigability with increasing population size is intuitive: Eq.(12) has a dependence on N such that decreases become exponentially less likely for larger N. For

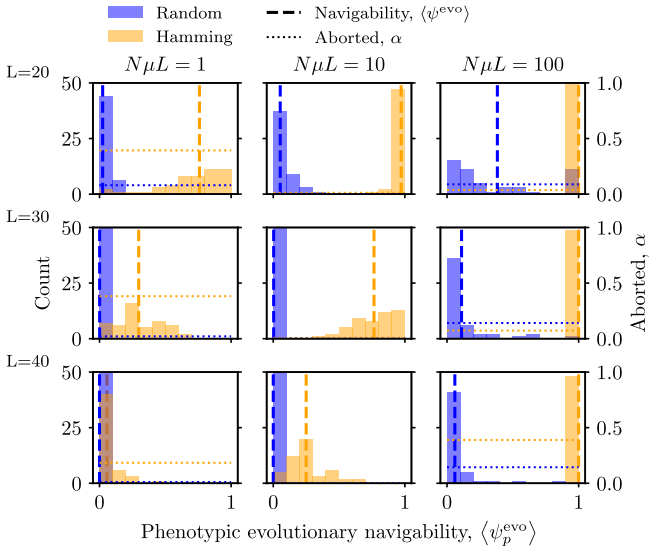


FIG. 9. **Navigability under non-monomorphic evolutionary dynamics for fRNA $L = 20,\ 30,\ 40$ (rows) for random and Hamming fitness assignments at different population mutation rates $N\mu L = 1,\ 10,\ 100$ (columns) a $N = 100$.** Blue shaded histogram show the proportion of successful searches for random fitness assignment, and the orange histogram for Hamming fitness assignment. Mean navigability is shown as vertical dashed lines and the proportion of aborted searches as horizontal dotted lines. Evolutionary navigability increases with increasing population mutation rate for both random and Hamming fitness assignment, but with Hamming fitness assignment providing greater navigability and with fewer aborted searches.

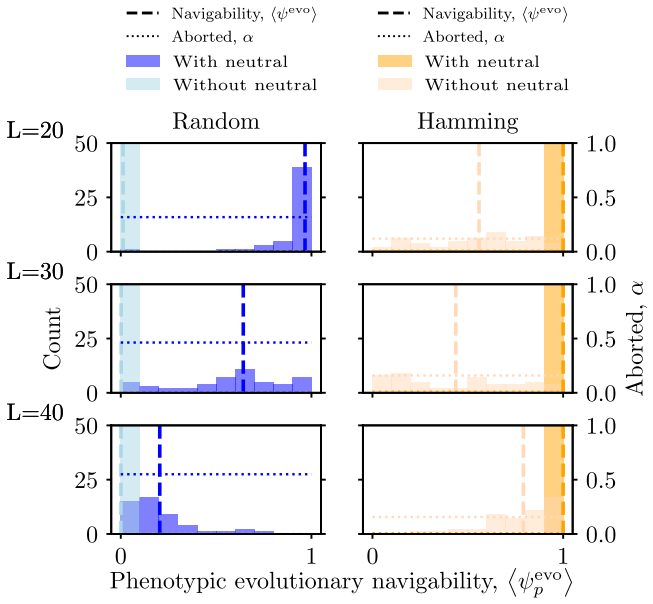


FIG. 10. **Evolutionary navigability $\langle\psi^{\text{evo}}\rangle$ for fRNA under monomorphic evolutionary dynamics with and without neutral mutations permitted.** Histograms of evolutionary navigability by target is plotted for $N_t = 50$ targets estimated from $N_s = 20$ sources for RNA sequences of length $L = 20$ (top), $L = 30$ (middle) and $L = 40$ (bottom), and random (left) and Hamming (right) fitness assignments. $N = 100,000$ in all dynamics. Random fitness assignment are in blue, Hamming fitness assignment in orange with the versions that did not allow neutral mutations shown in lighter shades. Vertical dashed lines indicate the overall evolutionary navigability $\langle\psi^{\text{evo}}\rangle$ and the horizontal dotted lines indicate the proportion of runs α that were aborted. Random fitness assignment decays to low navigability for all lengths, while Hamming fitness assignment is depleted but remains non-negligible across all lengths.

example, for novel phenotype q with F_q = 0.999 compared to $F_p = 1.0$, $s = -0.001$. The resulting probability of fixation for $N = 10$; 1000; 100000 are $P_{\text{fix}} = 0.10, 3 \times 10^{-4}, 3 \times 10^{-90}$ demonstrating a large reduction in the probability of fixation for mutants that have even just a small $s = 0.1\%$ decrease in fitness.

Appendix C: The effect of population mutation rate on navigability in non-monomorphic evolutionary dynamics

In Section II C 2, we considered navigability for polymorphic evolutionary dynamics at $N\mu L = 100$ for $N = 100$, using a Wright-Fisher model (see Section IV G) with populations evolving over $T = 20,000$ generations. Here, we extend the analysis to lower population mutation rates $N\mu L = 10$ and $N\mu L = 1$. These represent somewhat 'in-between' monomorphic ($N\mu L \ll 1$) and polymorphic ($N\mu L \gg 1$) dynamical regimes.

In Fig. 9, we plot evolutionary navigability for $N_s = 20$ fRNA sources and each of $N_t = 50$ targets at $L = 20,\ 30,\ 40$ (rows) for $N\mu L = 1,\ 10,\ 100$ (columns). At $N\mu L = 100$, Hamming fitness assignment has $\langle\psi^{\text{evo}}\rangle = 1$ but with an increasing fraction aborted with increasing L, reaching $\alpha = 0.389$ at $L = 40$. For random fitness assignment at $N\mu L = 100$, for $L = 20,\ 30,\ 40$, evolutionary navigability decreases with $\langle\psi^{\text{evo}}\rangle = 0.383,\ 0.111,\ 0.060$ respectively. At $N\mu L = 10$, evolutionary navigability with Hamming fitness assignment begins to decrease as L increases ($\langle\psi^{\text{evo}}\rangle = 0.972, 0.765, 0.250$ for $L = 20, 30, 40$), and for $N\mu L = 1$ the drop is even more pronounced ($\langle\psi^{\text{evo}}\rangle = 0.759, 0.298, 0.056$ for $L = 20, 30, 40$), with increasing fractions aborted. For random fitness assignment at $N\mu L = 10$ and $N\mu L = 1$, evolutionary navigability is negligible with $\langle\psi^{\text{evo}}\rangle < 0.024$ and a relatively low aborted proportion α.

Our results here indicate that, with the Wright-Fisher model and definition of majority phenotype for defining the phenotypes along the adaptive path, decreasing population mutation rate leads to less navigability, with the effect more pronounced at larger L and for random fitness assignment. Further research would be beneficial to understand whether this approach is effectively capturing the adaptive path at intermediate population mutation rates.

Appendix D: The role of neutral mutations for navigability in monomorphic evolutionary dynamics

We explore the role of neutral mutations in facilitating evolutionary navigability for fRNA at varying sequence length ($L = 20, 30, 40$) under different fitness functions (random and Hamming fitness assignment).

As demonstrated in Section II C and Table II, positive neutral correlations manifest through mutational robustness with neutral mutations playing a crucial role in facilitating the presence of accessible paths. A natural extension is to ask what role these play under the evolutionary dynamics? Monomorphic evolutionary dynamics provide a simple experimental setting in which we explore this. To do so, we modify the probability of transitioning to neighbouring mutant genotypes h from the current fixed genotype g, to have zero probability if they are neutral (i.e. $M(g) = M(h)$). We can rewrite this by modifying Eq. (12):

$$P_{\text{fix}}\left(s(g', g), N\right) = \begin{cases} \frac{1-\exp(-2s)}{1-\exp(-2Ns)} & M(g') \neq M(g) \\ 0 & M(g') = M(g) \end{cases} \quad \text{(D1)}$$

such that neutral cases where $M(g') = M(g)$ are assigned $P_{\text{fix}} = 0$.

In Fig. 10, we show the evolutionary navigability under these dynamics for three sequence lengths $L = 20$ (top) $L = 30$ (middle), $L = 40$ (bottom) measured across $N_t = 50$ targets and $N_s = 20$ sources per target for both random (left) and Hamming fitness assignment to phenotypes at a population size $N = 100,000$ (as in Fig. 5(left)). We focus on this case to explore the effect of removing neutral mutations because navigability is present for both fitness assignments at this population size.

For random fitness assignments, we observe that neutral mutations are essential for finding accessible paths with $\langle \psi^{\text{evo}} \rangle = 0$ for all lengths examined. The number of aborted runs also decreases to zero as steps of decreasing fitness are more likely to have been taken even when the target is not found. This is in stark contrast the the actual availability of accessible paths when neutral mutations are prohibited (Table II) where $\langle \psi \rangle > 0.273$ for $L = 20 - 40$, demonstrating their criticality under random fitness.

For Hamming fitness assignment the effect is less profound with $\langle \psi^{\text{evo}} \rangle = 0.559$ for $L = 20$, $\langle \psi^{\text{evo}} \rangle = 0.437$ for $L = 30$ and $\langle \psi^{\text{evo}} \rangle = 0.791$ for $L = 40$. However, the fact that there is a decrease from $\langle \psi^{\text{evo}} \rangle$ is of note as it identifies that phenotypic correlations through the Hamming distance are not sufficient to guarantee evolutionary navigability under monomorphic evolutionary dynamics, with the joint role of positive neutral correlations and phenotypic correlations driving complete evolutionary navigability.

Appendix E: Navigability with tolerance to fitness decreases

In Section II C 2, we introduced the possibility of allowing paths to possess tolerance to small decreases in fitness ΔF. Here, we extend this further by asking the question: what is the smallest allowable decrease in absolute fitness for an evolutionary path to be more likely to be accessible than not? That is, for $\langle \psi^{\text{evo}} \rangle > 0.5$, what ΔF must we have on average? As before, the allowable decrease in fitness along a path for is ΔF, with the requirement for strictly monotonic paths being where $\Delta F = 0$. We therefore recompute $\langle \psi^{\text{evo}} \rangle (\Delta F)$ as a function of ΔF in intervals of 0.025 on the range $[0, 1]$, and record $\Delta F_{min} = \text{argmin}_{\Delta F} \langle \psi^{\text{evo}} \rangle (\Delta F) > 0.5$.

TABLE IV. Upper bound for fitness tolerance ΔF_{min} at which $\langle \psi^{\text{evo}} \rangle > 0.5$ with aborted proportion α in brackets.

Length	Fitness	$N\mu L \ll 1$	$N\mu L = 100$
20	Random	0.050 (0.124)	0.025 (0.281)
20	Hamming	0.000 (0.001)	0.000 (0.037)
30	Random	0.050 (0.537)	0.050 (0.806)
30	Hamming	0.000 (0.001)	0.000 (0.073)
40	Random	0.075 (0.883)	0.075 (0.909)
40	Hamming	0.000 (0.000)	0.000 (0.389)

In Table E, we show the values for ΔF_{min} for monomorphic $N\mu L \ll 1$ and polymorphic ($N\mu L \gg 1$) evolutionary dynamics, along with the remaining aborted fraction at this threshold at a population size of $N = 100$ (comparable with main results Fig. 5).

The results show that for the monomorphic case with random fitness assignment, at $L = 20$ and $L = 30$, if the fitness is allowed to decrease by an absolute value of 0.050 along the path $\langle \psi^{\text{evo}} \rangle > 0.5$, while for $L = 40$ this critical value is greater at 0.075. However, the fraction aborted is of note with $\alpha = 0.124$, $\alpha = 0.537$ and $\alpha = 0.883$ for $L = 20, 30, 40$ respectively. For the monomorphic case, the Hamming fitness function has $\Delta F_{min} = 0$ across each length.

For the polymorphic case with random fitness assignment, similar thresholds less than 0.1 are observed with $\Delta F_{min} = 0.025, 0.050, 0.075$ for $L = 20, 30, 40$ respectively, but again with substantial aborted values of $\alpha = 0.281, 0.806, 0.909$ respectively.

Our results here further support indications in Fig. 5D that only small relaxation in the allowable fitness decrease along the adaptive path can lead to a large increase in navigability.

Appendix F: Evolutionary navigability with RNA shapes for longer fRNA sequences

We have been limited to studying sequences of limited length (up to $L = 40$ for fRNA) due to the vast and exponentially growing number of phenotypes ($N_P \approx 1.76^L$) at a given length L. However, an alternative approach for studying larger L is to instead use coarse-grained representations of the 'dot-bracket' phenotype, considering the 'shape' rather than exact description of the secondary structure. This 'RNAshape' approach was introduced by Giegerich *et al.* [58] and has been successfully demonstrated to possess similar GP map properties such as bias [59] and robustness relationships [60]. Here we make use of the 'RNAshapes' description to allow much longer sequences to be considered with respect to evolution-

TABLE V. Example genotype, dot-bracket and RNA shapes for a random fRNA with $L = 60$.

	Sequence/structure/shape
Genotype	UUGAACCGUAUGCAUCAAAAAGUGCAUGUACGGUUCCUAAGGGAUAAAAAUUUACCCUAC
Phenotype (dot-bracket)	..(((((((((((((((.....)))))..))))))))).((((.(((....)))))))..
Phenotype (shape level 1)	_[[]_]_[_[]]_
Phenotype (shape level 2)	[[]_][_[]]
Phenotype (shape level 3)	[[]][[]]
Phenotype (shape level 4)	[][]
Phenotype (shape level 5)	[][]

TABLE VI. Number of sequences and phenotypes at each coarse-grained level in the fRNA database.

L	Genotypes	dot-bracket	Shape level 1	Shape level 3
60	354	344	209	38
80	566	532	428	100
100	939	886	736	247
120	3006	2945	2654	912
140	291	284	283	253
160	1617	1590	1574	1351

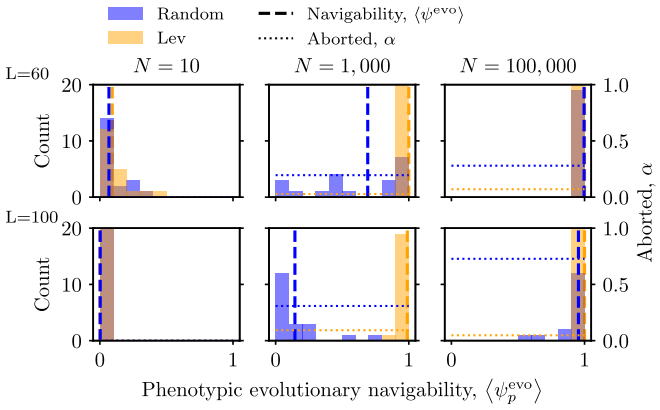


FIG. 11. **Evolutionary navigability $\langle\psi^{\text{evo}}\rangle$ with RNA shape level 1 phenotypes with fRNA for varying population size and sequence length.** Histograms of evolutionary navigability by target is plotted for $N_t = 50$ targets estimated from $N_s = 20$ sources for RNA sequences of length $L = 20$ (top), $L = 30$ (middle) and $L = 40$ (bottom), and for population sizes $N = 10$ (left), $N = 1,000$ (middle) and $N = 100,000$ (right). Results with random fitness assignment are in blue with Hamming fitness assignment in orange. Vertical dashed lines indicate the overall evolutionary navigability $\langle\psi^{\text{evo}}\rangle$ and the horizontal dotted lines indicate the proportion of runs α that were aborted. Landscapes are typically navigable even with random fitness for $N = 1,000$ and $N = 100,000$. Larger compared to smaller population sizes and Hamming compared to random fitness assignment lead to greater navigability.

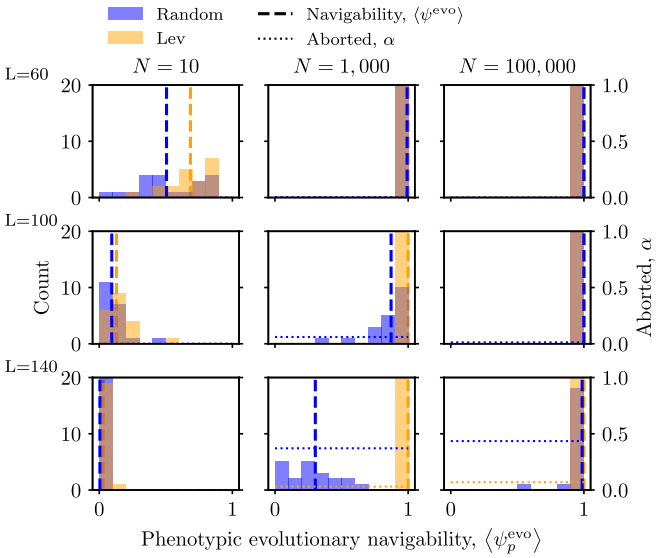


FIG. 12. **Evolutionary navigability $\langle\psi^{\text{evo}}\rangle$ with RNA shape level 3 phenotypes with fRNA for varying population size and sequence length.** Histograms of evolutionary navigability by target is plotted for $N_t = 50$ targets estimated from $N_s = 20$ sources for RNA sequences of length $L = 60$ (top), $L = 100$ (middle) and $L = 140$ (bottom), and for population sizes $N = 10$ (left), $N = 1,000$ (middle) and $N = 100,000$ (right). Results with random fitness assignment are in blue with Levenshtein fitness assignment in orange. Vertical dashed lines indicate the overall evolutionary navigability $\langle\psi^{\text{evo}}\rangle$ and the horizontal dotted lines indicate the proportion of runs α that were aborted.

ary navigability to provide evidence as to what we might expect for navigability in larger fRNA. We extend our analysis to consider RNA phenotypes sampled from the fRNAdb at $L = 60$, 100, 140 (where $N_P \approx 5 \times 10^{14}, 3 \times 10^{24}, 2 \times 10^{34}$ respectively, demonstrating the intractability), using the RNA 'shape' as the phenotype instead of the 'dot-bracket' structure. Table F illustrates an example $L = 60$ fRNA structure for different levels of coarse-grained shape. Table F provides counts of the number of unique sequence/structures/shapes at different lengths, with longer lengths intuitively getting closer to the number of unique sequences.

To find the shape for a given genotype, we first map each sequence to the dot-bracket phenotype before a further step to the RNA shape (level=1 and level=3 as used in [59]). We then assign fitness either randomly as we do with the dot-bracket, or based on a distance measure from the target to incorporate phenotypic correlations. For a given length RNA sequence, as the shape structure is no longer of fixed length or with specific alignment, the Hamming distance is not an appropriate

measure for measuring the distance from a target. Instead, we choose the Levenshtein distance between two strings which can accommodate variable length and non-specific alignment by measuring distance in terms of the minimum number of insertions, deletions or substitutions to map one string to the other [71]. We define the Levenshtein distance between shape p and q, $lev(p, q)$ and to use the distance measure as a fitness measure, we calculate the fitness:

$$F_{lev}(p, q) = 1 - \frac{lev(p, q)}{\max(L(p), L(q))} \tag{F1}$$

where $L(p)$ is the length of shape p and by definition the maximum value for $\max lev(p, q) := max(L(p), L(q))$.

Due to the additional computational expense of folding larger genotypes, we restrict the monomorphic evolutionary dynamics to a maximum number of generations $T = 1,000$ but other parameters and methods remain identical to Section IV G 1.

In Fig. F and Fig. 12, we depict the measured evolutionary navigability for $N_s = 20$ sources towards $N_t = 20$ targets under monomorphic evolutionary dynamics for a range of N for shape level 1 and shape level 3 phenotypes respectively, corresponding to increasing levels of coarse-grained representation. As for shorter fRNA, the same pattern is observed. Shape based fitness function (Levenshtein here rather than Hamming) lead to greater evolutionary navigability than random fitness assignment. This highlights the role of the phenotypic correlations. Additionally, larger population sizes lead to greater navigability due to the decreased probability of lower fitness mutants fixing under the Kimura fixation probability (Eq. (12)). Lastly, as sequence length increases, while navigability remains almost certain for $N = 100,000$ under either fitness function, for $N = 1,000$ there is a decrease to low navigability for random fitness. For $N = 10$, both fitness functions see a reduction to low navigability as sequence length increases. Of note among these results is that $\langle \psi^{\text{evo}} \rangle \approx 1$ for random fitness assignment at very large $L = 140$ for large population size $N = 100,000$.

By using a further coarse-grained representation of fRNA phenotypes, we find that longer fRNA sequences can also have evolutionary navigability without fitness valleys under evolutionary dynamics, following similar relationships with population size N, sequence length L and type of fitness assignment. This adds further support to our argument of the generality of the critical roles of redundancy in the GP map leading to large-scale neutral correlations, the fitness function incorporating phenotypic correlations and larger population sizes making the fixation of mutants with reduced fitness less likely to occur along adaptive paths, all facilitating more navigable fitness landscapes under evolutionary dynamics.