


Phenotype Bias Determines How Natural RNA Structures Occupy the Morphospace of All Possible Shapes

Kamaludin Dingle,¹ Fatme Ghaddar,¹ Petr Šulc ² and Ard A. Louis^{*3}

¹Centre for Applied Mathematics and Bioinformatics, Department of Mathematics and Natural Sciences, Gulf University for Science and Technology, Hawally, Kuwait

²School of Molecular Sciences and Center for Molecular Design and Biomimetics at the Biodesign Institute, Arizona State University, Tempe, AZ, USA

³Rudolf Peierls Centre for Theoretical Physics, University of Oxford, Oxford, United Kingdom

***Corresponding author:** E-mail: ard.louis@physics.ox.ac.uk.

Associate editor: Julian Echave

Abstract

Morphospaces—representations of phenotypic characteristics—are often populated unevenly, leaving large parts unoccupied. Such patterns are typically ascribed to contingency, or else to natural selection disfavoring certain parts of the morphospace. The extent to which developmental bias, the tendency of certain phenotypes to preferentially appear as potential variation, also explains these patterns is hotly debated. Here we demonstrate quantitatively that developmental bias is the primary explanation for the occupation of the morphospace of RNA secondary structure (SS) shapes. Upon random mutations, some RNA SS shapes (the frequent ones) are much more likely to appear than others. By using the RNASHapes method to define coarse-grained SS classes, we can directly compare the frequencies that noncoding RNA SS shapes appear in the RNACentral database to frequencies obtained upon a random sampling of sequences. We show that: 1) only the most frequent structures appear in nature; the vast majority of possible structures in the morphospace have not yet been explored; 2) remarkably small numbers of random sequences are needed to produce all the RNA SS shapes found in nature so far; and 3) perhaps most surprisingly, the natural frequencies are accurately predicted, over several orders of magnitude in variation, by the likelihood that structures appear upon a uniform random sampling of sequences. The ultimate cause of these patterns is not natural selection, but rather a strong phenotype bias in the RNA genotype–phenotype map, a type of developmental bias or “findability constraint,” which limits evolutionary dynamics to a hugely reduced subset of structures that are easy to “find.”

Key words: evolution, phenotype bias, morphospace, RNA structure.

Introduction

Darwinian evolution proceeds in two steps. First, random changes to genotypes can lead to new heritable phenotypic variation in a population. Next, natural selection ensures that variation with higher fitness is more likely to dominate a population over time. Much of evolutionary theory has focused on this second step. By contrast, the study of variation has been relatively underdeveloped (Smith et al. 1985; Wagner and Altenberg 1996; Gould 2002; Laland et al. 2014; McCandlish and Stoltzfus 2014; Wagner 2014; Love 2015; Charlesworth et al. 2017; Stoltzfus 2019; Uller et al. 2018; Svensson and Berger 2019; Uller and Laland 2019; Jablonski 2020). If the variation is unstructured, or *isotropic*, then this lacuna would be unproblematic. As Stephen J. Gould wrote in a critique of those who make this implicit assumption (Gould 2002):

Under these provisos, variation becomes raw material only – an isotropic sphere of potential about the modal form of a species . . . [only] natural selection . . . can manufacture substantial, directional change.

In other words, if a variation is isotropic, then evolutionary trends should primarily be rationalized in terms of natural selection. On the other hand, if there are strong anisotropic developmental biases, then structure in the arrival of variation may well play an important explanatory role in the biological phenomena we observe today. Although the discussion of how anisotropic variation affects adaptive evolutionary outcomes has moved on significantly from the days of Gould's critique, primarily due to the growth of the field of evo-devo (Love 2015), it remains a source of significant contention (Laland et al. 2014; McCandlish and Stoltzfus 2014; Love 2015; Charlesworth et al. 2017; Stoltzfus 2019; Uller et al. 2018; Svensson and Berger 2019; Uller and Laland 2019; Jablonski 2020).

Unraveling whether a long-term evolutionary trend in the past was primarily caused by biased variation is not straightforward. It often means answering counterfactual questions (Louis 2016) such as: What kind of variation could have occurred but did not due to bias? An important analysis tool for such questions was pioneered by Raup (1966), who plotted

three key characteristics of coiled snail shell shapes in a diagram called a *morphospace*, finding that only a relatively small fraction of all possible shapes were realized in nature. This concept can be generalized to almost any combination of phenotypic characters (McGhee 2007). The fundamental reason for the anisotropic occupation of a morphospace could simply be some form of contingency, where the evolution started at one point in the morphospace and did not have enough time to fully explore the space. Or it could be some more predictable cause, such as natural selection disfavoring certain characteristics, or else developmental bias favoring certain types of variation (Uller et al. 2018; Jablonski 2020).

One way to make progress on these big questions in evolutionary theory is to study genotype–phenotype (GP) maps that are tractable enough to provide access to the full spectrum of possible variation (Ahnert 2017; Manrubia et al. 2021) so that counterfactuals (Louis 2016) can be explored. In this article, we follow this strategy, employing the well-known GP mapping from RNA sequences to secondary structures (SS), to explain in detail how noncoding RNA (ncRNA) found in nature populates the morphospace of all possible RNA SS shapes.

RNA is a versatile molecule. Made of a sequence of four different nucleotides (AUCG) it can both encode information as messenger RNA (mRNA), or play myriad functional roles as ncRNA (Mattick and Makunin 2006). This ability to take a dual role, both informational and functional, has made it a leading candidate for the origin of life (Gilbert 1986). The number of identified functional ncRNA types has grown rapidly over the last few decades, driven in part by projects such as ENCODE Project Consortium (2012) and Palazzo and Lee (2015). Well-known examples include transfer RNA (tRNA), catalysts (ribozymes), structural RNA (most famously rRNA in the ribosome), and RNAs that mediate gene regulation such as micro RNAs (miRNA) and riboswitches.

The function of ncRNA is intimately linked to the three-dimensional (3D) structure that linear RNA strands fold into. Although much effort has gone into the sequence to 3D structure problem for RNA, it has proven to be stubbornly recalcitrant to an efficient solution (Thiel et al. 2017; Miao et al. 2020). By contrast, a simpler challenge, predicting the RNA SS which describes the bonding pattern of a folded RNA, and which is, therefore, a major determinant of tertiary structure, is much easier to solve (Lorenz et al. 2011; Schuster et al. 1994). A combination of computational efficiency and accuracy has made RNA SS a popular model for studying basic principles of evolution (Hofacker et al. 1994; Schuster et al. 1994; Fontana 2002; Knight et al. 2005; Wagner 2005, 2011; Smit et al. 2006; Cowperthwaite et al. 2008; Jorg et al. 2008; Stich et al. 2008; Aguirre et al. 2011; Schaper and Louis 2014; Dingle et al. 2015; Greenbury et al. 2016; García-Martín et al. 2018; Weiß and Ahnert 2018; Oliver et al. 2019).

An important driver of the growing interest in GP maps is that they allow us to open up the black box of variation—to explain, via a stripped-down version of the process of development, how changes in genotypes are translated into changes in phenotypes. Unfortunately, it remains much harder to establish how the patterns typically observed in

studies of GP maps (Ahnert 2017; Manrubia et al. 2021) translate into evolutionary outcomes, because natural selection must then also be taken into account. For GP maps, this means attaching fitness values to phenotypes which is difficult because fitness is hard to measure and is of course dependent on the environment, and so fluctuates.

Progress can still be made by simply ignoring fitness differences, and comparing patterns in nature directly to patterns in the arrival of phenotypic variation generated by uniform random sampling of genotypes, which is also known as “genotype sampling” or *G-sampling* (Dingle et al. 2015). For example, Smit et al. (2006) followed this strategy and found that G-sampling leads to almost identical nucleotide composition distributions for SS motifs such as stems, loops, and bulges as found for naturally occurring structural rRNA. In a similar vein, Jorg et al. (2008) calculated the neutral set size (NSS), defined as the number of sequences that fold to a particular structure, using a Monte-Carlo based sampling technique. For the length-range, they could study ($L = 30$ to $L = 50$), they found that natural ncRNA from the fRNAdb database (Mituyama et al. 2009) had much larger than average NSS. More recently, Dingle et al. (2015) developed a method that makes it possible to calculate the NSS, as well as the distributions of a number of other structural properties, for a much wider range of lengths. They found, for lengths ranging from $L = 20$ up to $L = 126$, that the distribution of NSS sizes of natural ncRNA—calculated by taking the sequences found in the fRNAdb, folding them to find their respective SS, and then working out its NSS using the estimator from Jorg et al. (2008)—was remarkably similar to the distribution found upon G-sampling. A similar close agreement upon G-sampling was found for several structural elements, such as the distribution of the number of helices, and also for the distribution of the mutational robustness, confirming earlier work on much smaller samples (Fontana et al. 1993).

An alternative to G-sampling is to use uniform random sampling of phenotypes, so-called *P-sampling*. If all phenotypes are equally likely to occur under G-sampling, then its outcomes will be similar to P-sampling. If, however, there is a bias toward certain phenotypes under G-sampling, an effect we will call *phenotype bias*, then the two sampling methods will lead to different results. When the authors of Dingle et al. (2015) calculated the distributions of structural properties such as the number of stems or the mutational robustness under P-sampling, they found large differences compared with natural RNA in the fRNAdb. The fact that G-sampling yields distributions close to those found for natural ncRNA, whereas the counterfactual under P-sampling does not, suggests that bias in the arrival of variation is strongly affecting evolutionary outcomes in nature. As illustrated schematically in figure 1a, such a bias toward shapes that appear frequently as potential variation can lead to natural RNA SS taking up only a small fraction of the total morphospace of possible RNA shapes. Here we treat the morphospace more abstractly, but this pattern would carry through with more traditional morphospaces (Raup 1966) that utilize specific axes to describe phenotypic characteristics or RNA.

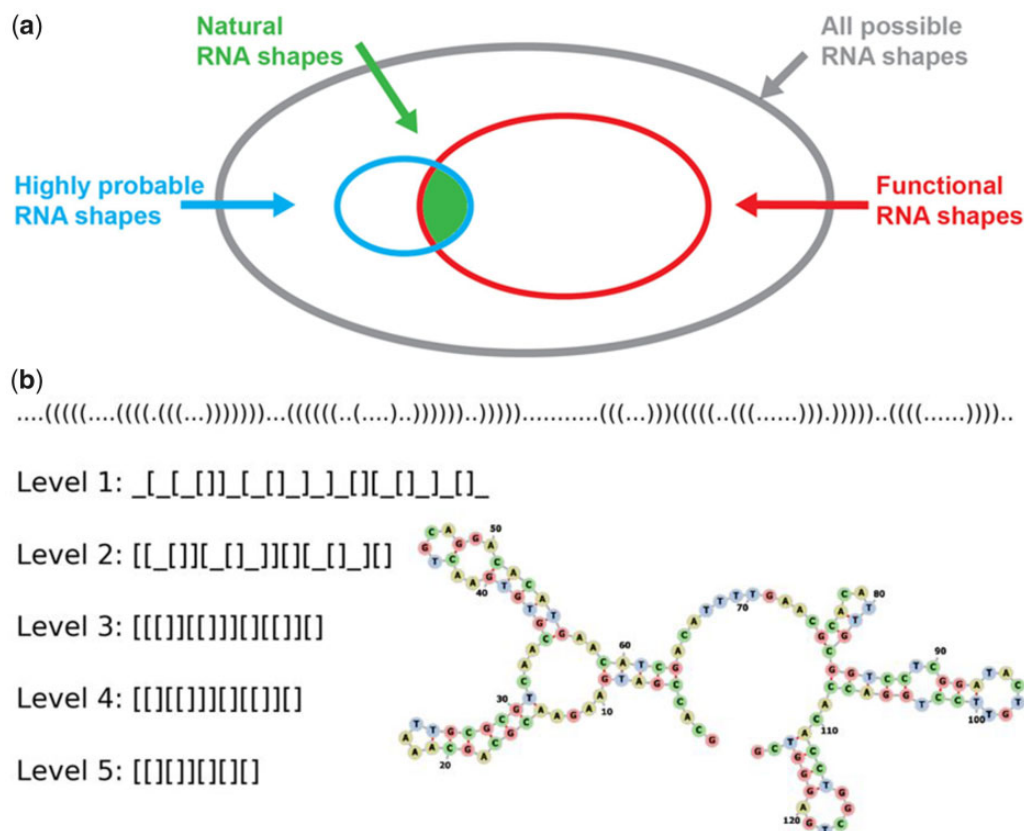


Fig. 1. (a) Conceptual diagram of the RNA SS shape morphospace: The set of all potentially functional RNA is a subset of all possible shapes. In this article we show that natural RNA SS shapes only occupy a minuscule fraction of the morphospace of all possible functional RNA SS shapes because of a strong phenotype bias which means that only highly probable (high-frequency) shapes are likely to appear as potential variation. We quantitatively predict the identity and frequencies of the natural RNA shapes by randomly sampling sequences for the RNA SS GP map. (b) RNA coarse-grained shapes: An illustration of the dot-bracket representation and five levels of more coarse-grained abstracted shapes for the 5.8S rRNA (length $L = 126$), a ncRNA. Level 1 abstraction describes the nesting pattern for all loop types and all unpaired regions; level 2 corresponds to the nesting pattern for all loop types and unpaired regions in external loop and multiloop; level 3 is the nesting pattern for all loop types, but no unpaired regions; level 4 is the helix nesting pattern and unpaired regions in external loop and multiloop; and level 5 is the helix nesting pattern and no unpaired regions.

Nevertheless, the evidence presented so far for this picture of a strong bias in the arrival of variation has been indirect, and only for distributions over SS structures because individual SS rarely appear more than once in the fRNAdb. Moreover, the measurements have often needed theoretical input, in that they used theoretical estimates for the NSS of individual sequences in the ncRNA databases. To conclusively address big questions related to the role of bias in evolutionary outcomes, a more direct measure is needed.

To achieve this goal of directly measuring frequencies, we first note that any tiny change to the bonding pattern of a full SS, illustrated by the dot-bracket notation in figure 1b, means a new SS. In practice, however, many small differences are often found in homologs, suggesting that these differences are not critical to function. To capture this intuition that larger-scale “shape” is more important than some of the finer features captured by the full dot-bracket notation, Giegerich et al. (2004) defined a five-level hierarchical abstract representation of SS. At each nested level of description, the SS shape is more coarse-grained, as illustrated in figure 1b. By grouping together shapes with similar features, frequencies of ncRNA

shapes can be directly measured from a given database. Here, we mainly use the large, popular, and up-to-date RNAcentral (RNAcentral Consortium 2021) database.

In this article, we show that the frequency f_p with which abstract shapes are found in the RNAcentral database is accurately predicted by frequencies f_p^G that they are found for G-sampling. We then discuss what these results mean in light of the longstanding controversies about developmental bias.

Results

Nature Only Uses High Frequency Shapes, Which Are Easily Found

We computationally generated random RNA sequences for lengths $L = 40, 55, 70, 85, 100, 126$, and then folded them to their SS using the Vienna package (Lorenz et al. 2011), which, as for other closely similar packages based on thermodynamics (Mathews et al. 2004), such as RNAstructure (Reuter and Mathews 2010) or Unafold (Markham and Zuker 2008), is thought to be accurate for the relatively short RNAs we study here (Methods). Next, we used the RNA abstract shapes

method (Giegerich et al. 2004; Janssen and Giegerich 2015) (see fig. 1b), to classify the folded SS into separate abstract structures. Similarly, we also took natural ncRNA sequences from the popular RNAcentral database (RNAcentral Consortium 2021), folded these, and used the RNA abstract shape method to assign structures to them (see Methods).

To compare the G-sampled RNA structures to the natural structures, a balance must be struck between being detailed enough to capture important structural aspects, but not so detailed that for a given data set very few repeated shapes are found, making it impossible to obtain reliable frequencies. Considering our data sets, we use level 3 for all RNA of length $L = 40$ and $L = 55$ and level 5 for $L \geq 70$. In supplementary figures S1 and S2, Supplementary Material online, we include all five other levels for $L = 55$, finding essentially the same results.

Figure 2 shows the shape frequencies f_p^G found by G-sampling, ranked from most frequent to least frequent (blue dots). The frequencies, or equivalently the NSS of these structures, vary by many orders of magnitude. The shapes which also appear in the RNAcentral database have been highlighted (yellow circles). Natural ncRNA shapes employ a tiny subset of the most frequent structures. Interestingly, a remarkably small number of random sequences, on the order of 10^3 – 10^6 independent random samples, is enough to find essentially all shapes at these levels of abstraction found in the RNAcentral database for the lengths studied here. For a sense of scale, there are $4^{126} \approx 7 \times 10^{75}$ sequences of length 126, so that we are sampling on the order of 1 in 10^{70} th of the total space and still finding all the structures at

the coarse-graining levels chosen. Note that this fraction decreases as the coarse-graining level increases. For example, in supplementary figure S1, Supplementary Material online, where we show that for $L = 55$ strands, for which there are $4^{55} \approx 10^{33}$ total possible sequences, we need on the order of 10^7 sequences for level 1, up to 10^4 sequences for level 5. These numbers of samples all remain remarkably tiny fractions of the total.

To further quantify just how small a subset of the total morphospace has been explored by nature, we use asymptotic analytic estimates of the total set of possible structures from table 1 of Nebel and Scheid (2009) (but see also earlier results in Lorenz et al. [2008]). These predict $s_{3_L} \approx 1.85 \times 1.46^L \times L^{-\frac{3}{2}}$ for level 3 and $s_{5_L} \approx 2.44 \times 1.32^L \times L^{-\frac{5}{2}}$ for level 5, where we have taken results pertaining to minimum hairpin length of 3, and minimum ladder length of 1 (which is consistent with the options we used in the Vienna folding package). From these equations, we estimate $s_{3,40} \approx 10^4$, $s_{3,55} \approx 10^7$, $s_{5,70} \approx 10^6$, $s_{5,85} \approx 10^8$, $s_{5,100} \approx 10^9$, and $s_{5,126} \approx 10^{12}$. By contrast, in the RNAcentral database we find, at level 3, 18 structures for $L = 40$ and 63 for $L = 55$. At level 5 we find 16, 25, 35, and 68 independent structures for $L = 70, 85, 100$ and 126 respectively. We provide a direct illustration in figure 3 where the top 183 level 3 structures found for $L = 55$ (after 5×10^6 samples) are shown together with the 63 found in nature. The structures employed by natural ncRNA take up an incredibly small fraction of the whole morphospace of possible structures. Moreover, the relative fraction explored decreases rapidly with increasing length.

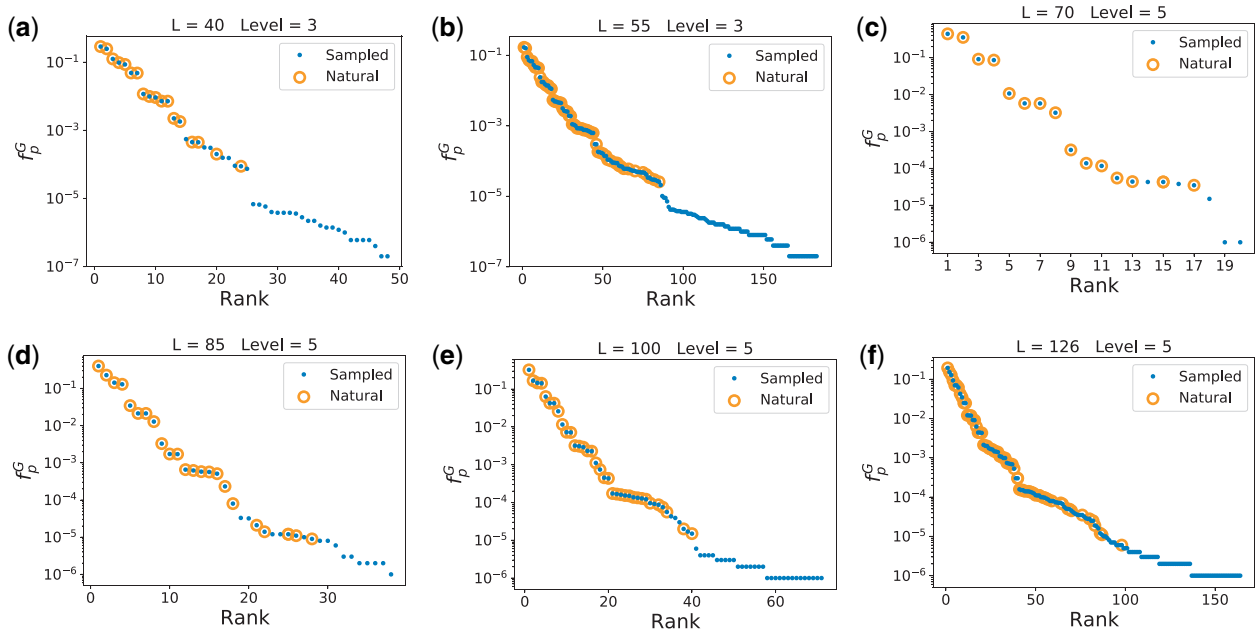


FIG. 2. Nature selects highly frequent structures. The frequency f_p^G (blue dots) of each abstract shape, calculated by random sampling of sequences (G-sampling), is plotted versus the rank. Yellow circles highlight which of the randomly generated shapes were also found in the RNAcentral database. Panels (a–f) are for $L = 40, 55, 70, 85, 100, 126$, respectively. The number of natural shapes are 18, 63, 16, 25, 35, and 68 in order of ascending length, whereas the numbers of possible shapes in the full morphospace are many orders of magnitude larger, ranging from $\approx 10^4$ possible level 3 shapes for $L = 40$ to $\approx 10^{12}$ level 5 shapes for $L = 126$. The shapes in nature are all from remarkably small fraction of possible structures that have the highest f_p^G or equivalently the highest NSS. The natural shapes found in the database appear upon relatively modest amounts of random sampling of sequences.

1	((()))	21	(())	41	(())	61	(())	81	(())	101	((()))	121	(())	141	((()))	161	((()))	181	(())
2	((()))	22	(())	42	(())	62	(())	82	(())	102	((()))	122	(())	142	((()))	162	(())	182	(())
3	((()))	23	(())	43	_	63	(())	83	(())	103	((()))	123	(())	143	((()))	163	(())	183	((()))
4	(())	24	((()))	44	(())	64	((()))	84	(())	104	((()))	124	(())	144	(())	164	((()))		
5	(())	25	((()))	45	(())	65	(())	85	(())	105	((()))	125	(())	145	(())	165	(())		
6	(())	26	((()))	46	(())	66	((()))	86	(())	106	((()))	126	(())	146	(())	166	((()))		
7	(())	27	((()))	47	((()))	67	(())	87	(())	107	((()))	127	(())	147	((()))	167	((()))		
8	((()))	28	((()))	48	((()))	68	((()))	88	(())	108	((()))	128	(())	148	((()))	168	(())		
9	(())	29	((()))	49	((()))	69	(())	89	(())	109	((()))	129	(())	149	((()))	169	((()))		
10	(())	30	((()))	50	((()))	70	((()))	90	(())	110	((()))	130	(())	150	((()))	170	((()))		
11	((()))	31	((()))	51	(())	71	(())	91	((()))	111	((()))	131	(())	151	(())	171	((()))		
12	((()))	32	(())	52	((()))	72	(())	92	((()))	112	(())	132	((()))	152	((()))	172	((()))		
13	(())	33	((()))	53	((()))	73	((()))	93	((()))	113	(())	133	(())	153	((()))	173	((()))		
14	(())	34	((()))	54	((()))	74	(())	94	((()))	114	((()))	134	((()))	154	((()))	174	((()))		
15	((()))	35	(())	55	((()))	75	((()))	95	((()))	115	((()))	135	((()))	155	((()))	175	((()))		
16	(())	36	((()))	56	((()))	76	((()))	96	(())	116	((()))	136	(())	156	((()))	176	(())		
17	(())	37	((()))	57	((()))	77	((()))	97	((()))	117	((()))	137	((()))	157	((()))	177	((()))		
18	(())	38	((()))	58	((()))	78	(())	98	((()))	118	((()))	138	((()))	158	((()))	178	((()))		
19	((()))	39	((()))	59	((()))	79	((()))	99	((()))	119	((()))	139	(())	159	((()))	179	((()))		
20	(())	40	((()))	60	((()))	80	((()))	100	((()))	120	((()))	140	((()))	160	((()))	180	((()))		

Fig. 3. Shape array for $L = 55$ RNA at level 3, showing the 183 shapes found by sampling 5×10^6 random sequences, in order of their rank by frequency f_p^G . The 63 naturally occurring level 3 shapes from the RNAcentral database are highlighted in yellow, demonstrating that only a small fraction of the total morphospace of shapes is occupied by RNAs found in nature, and that these are all highly frequent structures. We estimate that there are on the order of 10^7 possible level 3 structures for $L = 55$ RNA, so that this array only shows a tiny fraction of the total morphospace of shapes.

Frequencies of Shapes in Nature Can Be Predicted from Random Sampling

Figure 4 demonstrates that the G-sampled frequency of shapes correlates closely with the natural frequency of shapes, for a range of lengths. In [supplementary figure S2, Supplementary Material](#) online, we show for $L = 55$ that similar results are found for different levels of shape abstraction so that this result is not dependent on the level of coarse-graining.

We note that there is an important assumption in our interpretation, which is that the frequency with which structures are found in the RNAcentral database is similar to the frequency with which they are found in nature. To first order, it is reasonable to assume that this is true, as the databases are typically populated by finding sequences that are conserved in genomes, a process that should not be too highly biased. Moreover, the good correlation between the f_p^G and f_p found here provides additional a posteriori evidence for this assumption as it would be hard to imagine how such close agreement could obtain if there were strong man-made biases in the database. Nevertheless, there are structures that have been the subject of greater researcher interest, and one may expect them to be deposited in the database with higher frequency. We give one example in [figure 4c](#) of an outlier that is overrepresented (with high confidence) compared with our prediction, namely the shape $[\] [\] [\]$, which includes the classic clover-leaf shape of transfer RNA.

Further, we show in [supplementary section C, Supplementary Material](#) online, that qualitatively similar rank and correlation plots ([supplementary fig. S3, Supplementary Material](#) online) appear using the popular Rfam database (Kalvari, Argasinska, et al. 2018; Kalvari, Nawrocki, et al. 2018), where structures are determined not by folding, but by a consensus alignment procedure. We also show in [supplementary materials S5 and S6, Supplementary Material](#) online, that our results are robust to changes in CG bias, or when we include other suboptimal structures that are close enough in energy to be accessed by thermal fluctuations. The similar behavior we find across structure prediction methods, strand lengths, and databases would be extremely odd if artificial biases were strong on average in the natural databases. Hence, we believe that our main findings are unlikely to be due to database biases, although at a finer scale there may very well be biases, such as the one we present for tRNA, that are observable, and possibly an interesting source of new insight. Finally, here we have used relatively short RNA for the purposes of computationally tractability and accuracy, in forthcoming work we study much longer RNA, as well as utilizing different methods.

Discussion

We first recapitulate our main results below under three headings and discuss their implications for evolutionary theory.

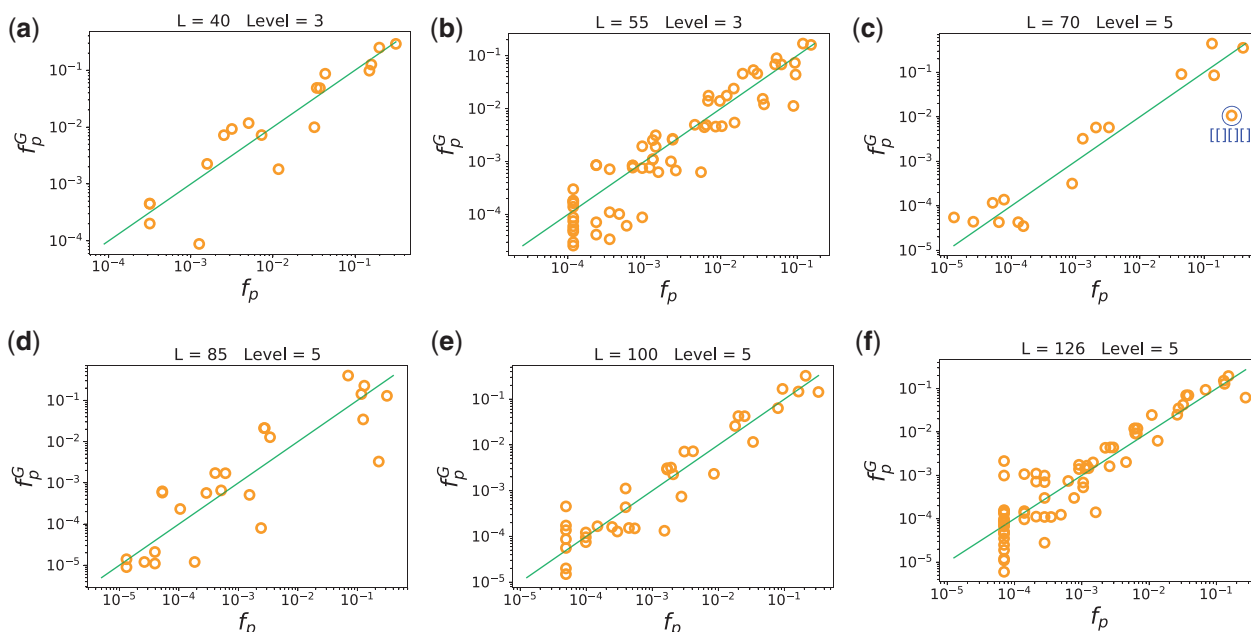


FIG. 4. The frequency of shapes in nature correlates with the frequency of shapes from random sampling. Yellow circles denote the frequencies f_p of natural RNA from RNAcentral. The green line denotes $x = y$, that is natural and sampled frequencies coincide. The log frequency upon G-sampling f_p^G correlates well with f_p ; (a) $L = 40$ Pearson $r = 0.92$; (b) $L = 55$ $r = 0.93$; (c) $L = 70$ $r = 0.94$; (d) $L = 85$ $r = 0.86$; (e) $L = 100$ $r = 0.95$; (f) $L = 126$ $r = 0.92$; and all correlations have P -value $< 10^{-6}$. We also highlight a blue structure, namely t-RNA for $L = 70$ which has been the subject of extra scientific interest, and is hence overrepresented in the database.

Nature Only Utilizes a Tiny Fraction of the RNA SS Phenotypic Variation That Is Potentially Available

Besides being an interesting fact about the natural world, this result has implication for synthetic biology as well. There is a vast morphospace (McGhee 2007) of structures that nature has not yet sampled. If these could be artificially created, then they could be mined for new and potentially intriguing functions.

Remarkably Small Numbers of Sequences Are Needed to Recover the Full Set of Abstract Shapes in the RNAcentral Database

This effect is enhanced by the fact that we have coarse-grained the SS to allow for direct comparisons. As shown in the supplementary section A, Supplementary Material online, for finer descriptions of the SS, more sequences are needed to obtain all natural structures, but the numbers remain remarkably small.

For a sense of the scale of the tiny numbers of sequences needed to produce the full spectrum of structures found in nature, consider that the total number of sequences N_G grows exponentially with length as $N_G = 4^L$. This scaling implies unimaginably vast numbers of possible sequences, even for modest RNA lengths. For example, all individual sequences of length $L = 77$ together would weigh more than the earth, whereas the mass of all combinations of length $L = 126$ would exceed that of the observable universe (Louis 2016). Such hyper-astronomically large numbers have been used to argue against the possibility of evolution producing viable phenotypes, based on the claim that the space is too vast to search through. See the Salisbury-Maynard Smith controversy

(Salisbury 1969; Smith 1970) for an iconic example of this trope. And it is not just evolutionary skeptics who have made such claims. In an influential essay, Francois Jacob wrote (Jacob 1977):

The probability that a functional protein would appear de novo by random association of amino acids is practically zero.

A similar argument could be made for RNA. Our results suggest instead that a surprisingly small number of random sequences is enough to generate all the basic RNA structures needed for life in all its diversity. This finding is relevant for the RNA world hypothesis, since it suggests that relatively small numbers of sequences are needed to facilitate primitive life. In the same vein, it helps explain why random RNAs can already exhibit a remarkable amount of function (Neme et al. 2017), similarly to what is suggested for proteins in the rapidly developing field of de novo gene birth (Begun et al. 2007; de la Peña and García-Robles 2010; Tautz and Domazet-Lošo 2011; Wilson et al. 2017).

The Frequency with Which Structures Are Found in Nature Is Remarkably Well Predicted by Simple G-sampling

This result is perhaps the most surprising of the three because these G-sampling ignores natural selection. It is widely thought that structure plays an important part in biological function, and so should be under selection.

The key to understanding results (A)–(C) above can be found in one of the most striking properties of the RNA SS GP map, namely strong phenotype bias which manifests in the

enormous differences in the G-sampled frequencies (or equivalently the NSS) of the SS (Schuster et al. 1994). For example, for $L = 20$ RNA, the largest system for which exhaustive enumeration was performed (Schaper and Louis 2014), the difference in the f_p^G between the most frequent SS phenotype and the least frequent SS phenotype was found to be 10 orders of magnitude. For $L = 100$ this difference was estimated to be over 50 orders of magnitude (Dingle et al. 2015; García-Martín et al. 2018). Such phenotype bias also explains why G-sampling and P-sampling are so different (Dingle et al. 2015): a small fraction of high-frequency phenotypes take up the majority of the genotypes, and thus dominate under G-sampling.

Evolutionary modeling that takes strong bias in the arrival of variation into account is rare. Population-genetic models that do include new mutations typically consider a genotype-to-fitness map, which often includes an implicit assumption that all phenotypes are equally likely to appear as potential variation, something akin to P-sampling. A notable exception is a work by Yampolsky and Stoltzfus (2001) which has been applied, for example, to the effect of mutational biases (Stoltzfus and McCandlish 2017; Cano and Payne 2020).

For the specific case of RNA, however, the effect of strong phenotype bias was treated explicitly in Schaper and Louis (2014), where it was shown for evolutionary dynamics simulations ranging from the low mutation monomorphic to the high mutation polymorphic regimes that the mean rate ϕ_{pq} at which new variation p appears in a population made up of phenotype q can be quite accurately approximated as $\phi_{pq} \approx (1 - \rho_q)f_p^G$, where ρ_q is the mean mutational robustness of genotypes mapping to q . In other words, the average local rate at which variation p appears in an evolving population closely tracks the global frequency f_p^G , which is exactly what G-sampling measures.

Although it is not so controversial that biases could affect outcomes under neutral mutation (Lynch 2007), the strongest disagreements in the field center around the effect of bias in adaptive mutations (Laland et al. 2014; McCandlish and Stoltzfus 2014; Love 2015; Charlesworth et al. 2017; Stoltzfus and McCandlish 2017; Stoltzfus 2019; Uller et al. 2018; Svensson and Berger 2019; Uller and Laland 2019; Cano and Payne 2020; Jablonski 2020). Since RNA structure is thought to be adaptive, the main question to answer is how phenotype bias affects RNA evolution when natural selection is also at work. In Schaper and Louis (2014), the authors explicitly treat cases where phenotype bias and fitness effects interact. They provide calculations of an effect called the *arrival of the frequent*, where the enormous differences in the rate at which variation arrives implies that frequent phenotypes are likely to fix, even if other higher fitness, but much lower frequency phenotypes are possible in principle. This same effect has also been observed in the evolutionary modeling of gene regulatory networks (Catalán et al. 2020). To avoid confusion, we note that the arrival of the frequent is fundamentally different from the survival of the flattest (Wilke et al. 2001), which is a steady-state effect. There, two phenotypes compete, and at high mutation rates, the one with the largest neutral set size can dominate in a population, even if its fitness is smaller. By

contrast, the arrival of the frequent is a nonergodic effect in the sense that it is not about a steady-state with competing phenotypes in a population. Instead, it is about large differences in the rate at which variation appears. Indeed, it can be shown (Schaper and Louis 2014) for the RNA GP map, that to first order, the number of generations T_p at which variation on average first appears in a population scales as $T_p \propto 1/f_p^G$ in both the high and the low mutation regimes. Since f_p^G varies over many orders of magnitude, on a typical evolutionary time-scale T , only a limited amount of variation (typically that with $T_p \lesssim T$) can appear. Variation can only fix if it appears in a population. Therefore natural selection acts on SS variation that has been heavily presculpted by the GP map (Dingle et al. 2015).

The close agreement between G-sampling frequencies and measured frequencies of natural ncRNA suggests that once an SS is found that is good enough, natural selection mainly works by further refining parts of the sequence for function, rather than significantly altering the SS. Taken together, these arguments suggest that the arrival of the frequent picture, which is fundamentally about strongly anisotropic variation, provides a mechanism that rationalizes all three main classes of observations above.

Nevertheless, given the wide diversity of possible fitness functions that will have played a role in the emergence of the different RNA structures found in the RNAcentral database, the argument above that strong phenotype bias determines the outcomes of evolutionary dynamics in such a predictable way may still seem quite surprising. There is, however, a fascinating connection between the arrival of the frequent effect in evolution, and a related behavior in the dynamics of optimization in deep neural networks (DNNs). One way of thinking about DNNs is as mappings from the (adjustable) parameters of the DNN (the genotypes), to functions (the phenotypes), which describe how input data maps to a DNN output. The volume of parameters mapping to a particular function is directly proportional to the probability that this function obtains upon random sampling of parameters, and is analogous to the NSS in GP maps. As was found for the RNA GP maps, the mapping from parameters to DNN functions can be hugely biased (Valle-Pérez et al. 2018; Mingard et al. 2019).

Of course, DNNs are not trained by randomly sampling parameters, just as an evolutionary process does not use G-sampling either. Instead, the most popular way to optimize DNNs is by using stochastic gradient descent (SGD) (Bottou et al. 2018) which follows the contours of a complex loss-landscape, much as evolution follows a fitness-landscape over time. For such highly biased systems, the arrival of the frequent phenomenology predicts that functions with a large volume of parameters mapping to them are much more likely to be found by an optimiser than functions with a smaller volume of parameters are. Interestingly, it was recently shown for several DNNs and data sets (Mingard et al. 2021) that the probability that the SGD optimizer converges on a particular function is well approximated by the (Bayesian) probability that this function obtains upon random sampling of DNN parameters, which is directly analogous to G-sampling. Since

this phenomenology was observed for multiple systems and loss functions, it suggests that a mechanism much like the arrival of the frequent works robustly for these highly biased systems also. As for the RNA system, where G-sampling provides a good first-order prediction of the frequencies of RNA structures found in nature, so for DNNs, random sampling of parameters provides a good first-order prediction of the frequencies that a DNN converges to a particular function when it is optimized by SGD. This analogy between optimization on a loss landscape for DNNs and evolutionary dynamics on fitness landscapes strengthens our hypothesis that the arrival of the frequency mechanism can explain why, for highly biased GP maps, multiple evolutionary scenarios produce outputs with probabilities given by G-sampling.

It is interesting to consider whether our arguments that strong phenotype bias affects adaptive evolution can shed light on a related controversy around mutational biases. For example, [Stoltzfus and McCandlish \(2017\)](#) argued that transition–transversion mutation bias in the arrival of mutations can affect the frequency of adaptive amino acid substitutions. This conclusion was criticized by [Svensson and Berger \(2019\)](#), who argued that the bias may not be large enough to overcome fitness differences, and that there may be alternative adaptive arguments for the codon substitution patterns observed in [Stoltzfus and McCandlish \(2017\)](#). The basic arguments behind mutational biases having an effect in adaptive evolution are similar in spirit to our arguments for phenotype bias, but there are also differences. Phenotype bias is about the rate at which phenotypes arise, and here we treat all mutations as being equally likely, whereas mutational bias captures inhomogeneities in the rate at which mutations arise along a genome.

Mutation bias is also typically much smaller than phenotype bias ([Stoltzfus and McCandlish 2017](#); [Cano and Payne 2020](#); [Gomez et al. 2020](#)). The global differences in the f_p^G are enormous. Even in the presence of adaptive forces, this “findability constraint” limits the evolutionary process to a tiny subset of high-frequency phenotypes in the morphospace. Within the subset of phenotypes that are found, however, the relative differences in frequencies are relatively small (on the order of 4 to 5 orders of magnitude in range). As long as phenotypes are findable, one might think that adaptive forces could overwhelm the developmental bias, which manifests in differential rates in the arrival of variation ([Laland et al. 2014](#); [Charlesworth et al. 2017](#); [Svensson and Berger 2019](#)). Interestingly, we nevertheless observe a fairly close correlation between f_p and f_p^G which suggests that averaged across many evolutionary scenarios, relatively small differences in the arrival of variation, such as those expected under mutational bias ([Stoltzfus and McCandlish 2017](#); [Cano and Payne 2020](#); [Gomez et al. 2020](#)), may indeed affect adaptive evolution.

Strong phenotype bias is also consistent with SELEX experiments ([Ellington and Szostak 1990](#); [Tuerk and Gold 1990](#)), where artificial selection for RNA function can, with a relatively small amount of material, lead to the repeated convergent evolution of the same structures. Famous examples include RNA aptamers ([Vu et al. 2012](#)) and the hammerhead ribozyme ([Salehi-Ashtiani and Szostak 2001](#)), which also

shows convergence in nature ([de la Peña and García-Robles 2010](#)). In light of the unimaginably small portion of the hyper-astronomically large sequence spaces these experiments explore, this convergent evolution seems highly surprising. But when we consider the strong phenotype bias, then a possible explanation emerges. SELEX experiments rely on artificial selection to refine sequences and hone in on a particular function. Although natural selection is the ultimate reason why a particular *function* emerges (such as self-cleaving catalytic activity for the hammerhead ribozyme), we hypothesize that the same *structures* emerge because of phenotype bias. After all, multiple structures could, in principle, produce the same function. In other words, to use Mayr’s famous ultimate-proximate distinction ([Mayr 1961](#); [Laland et al. 2011](#); [Scholl and Pigliucci 2015](#)) for RNA SS, phenotype bias is the ultimate, and not merely the proximate cause of the evolutionary convergence of the structures found in SELEX experiments and in nature. The idea that developmental biases could help explain convergence is not new, but we believe that the type of phenotype bias we are proposing here has not yet been seriously considered as a cause of convergence.

How is phenotype bias related to the broader literature on the developmental bias? To first order, phenotype bias is just another way of expressing developmental bias: certain phenotypes are more likely than others to appear upon mutations. Early work mainly considered developmental bias as a constraint ([Smith et al. 1985](#)), in that it limits what kind of variation natural selection can work on. The phenotype bias we observe can be viewed that way. Whether it can also act as a developmental drive ([Arthur 2001](#)) that facilitates adaptive evolution would hinge on there being advantages to the kinds of structures that it favors. Indeed, G-sampled RNA structures are on average different from P-sampled structures. For example, they have higher mutational robustness, and fewer stems ([Dingle et al. 2015](#)). So, there is a bias toward these characteristics, which may be adaptive.

Where RNA phenotype bias differs the most from classic examples of developmental bias such as the universal pentadactyl nature of tetrapod limbs, is that the latter are thought to occur because evolution took a particular turn in the past that locked in a developmental pathway, most likely through shared ancestral regulatory processes ([Jablonski 2020](#)). If one were to rerun the tape of life again, then it is conceivable that a different number of digits would be the norm. By contrast, phenotype bias predicts that the same spectrum of RNA shapes would appear, populating the morphospace in the same way. It is true that given enough time, a larger set of RNA shapes could appear, but the exponential nature of phenotype bias implies that orders of magnitude more time are needed to see linear increases in the number of potential shapes, so that we can be pretty confident that a broadly similar spectrum of shapes would appear again.

It is also interesting to compare phenotype bias to adaptive constraints. For example, there are many scaling laws observed in nature. One of the most dramatic is Kleiber’s law which states that the metabolic rate of organisms scales as their mass to the 3/4 power, and which has been shown to hold over a remarkable 27 orders of magnitude ([West and](#)

Brown 2005)! The morphospace of metabolic rates and masses is therefore highly constrained. Such scaling laws can be understood in an adaptive framework from the interaction between various basic physical constraints (West and Brown 2005), rather than from biases in the arrival of variation. Phenotype bias also arises from a fundamental physical process (Dingle et al. 2018) and limits the occupation of the RNA morphospace. But it provides, by contrast, a nonadaptive explanation for the constraint. At this level, it may be closest in spirit to some constraints that are postulated in biological or process structuralism (Thompson 1942) since the phenotype bias “findability constraint” arises from the GP map itself.

In the literature, nomenclature around developmental biases and evolutionary constraints is not completely settled, and that ambiguity affects our discussions above. Phenotype bias is interesting in this regard because on a larger scale it is perhaps most naturally described as a findability constraint, whereas on the smaller scale of those phenotypes that are found, the close agreement between f_p and f_p^G is perhaps most naturally described as a developmental bias or a developmental drive.

Finally, the fact that G-sampling does such a good job at predicting the likelihood that SS structures are found in nature also has implications for the study of selective processes in RNA structure (Rivas et al. 2017; Schlick and Pyle 2017). We propose here that signatures of natural selection should be measured by considering deviations from the null-model provided by G-sampling. Our current work has been on relatively short sequences, where simple SS folding algorithms based on thermodynamics are thought to work reasonably well. For longer sequences, other more sophisticated methods that include, for example, information from evolutionary covariance (Rivas et al. 2020), may be needed.

In conclusion, although the RNA sequence to SS map describes a pared down case of development, this simplicity is also a strength. Just as in the fields of chemistry and physics, where the hydrogen atom provides an important model system because it is so easily solvable, so the RNA SS GP map could be viewed as the “hydrogen atom of developmental biology.” The fact that it is so tractable allows us to explore counterfactual questions (Louis 2016) such as: what kind of phenotypic variation is possible in principle, but did not appear due to phenotypic bias. This system thus provides, to our knowledge, the cleanest evidence yet for developmental bias strongly affecting evolutionary outcomes.

Many other GP maps also show strong phenotype bias (Dingle et al. 2018; Manrubia et al. 2021). An important question for future work will be whether there is a universal structure to this phenotype bias that holds more widely and whether it also has such a clear effect on evolutionary outcomes in other biological systems. In this context, we note a recent proposal (Johnston et al. 2021) that applies a result related to the coding theorem of algorithmic information theory (AIT) (Dingle et al. 2018) to predict that GP maps should be generically biased toward phenotypes with low descriptive (Kolmogorov) complexity. In close analogy to what we found for natural RNA here, such simplicity bias was

demonstrated directly for protein quaternary structures in nature as well as for a related polyomino model for protein quaternary structures (Ahnert et al. 2010; Johnston et al. 2011; Greenbury et al. 2014), and also in a gene regulatory network. If it is indeed the case that strong simplicity bias is common in nature, and if, as also suggested in (Johnston et al. 2021), the arrival of the frequent mechanism is important for the evolutionary dynamics of this much wider set of systems, then the conclusions for evolutionary causation driven by strong phenotype bias we draw here for RNA should hold much more widely in nature.

Materials and Methods

Folding RNA

We use the popular Vienna package (Lorenz et al. 2011), based on the Turner model thermodynamics (Mathews et al. 2004), to fold sequences to structures, with all parameters set to their default values (e.g., the temperature $T = 37^\circ\text{C}$). This method, much like others in its class, is thought to be especially accurate for shorter RNA. The numbers of random samples were 5×10^6 for $L = 40$ and $L = 55$, and 10^6 for $L = 70, 85, 100, 126$. For G-sampling, we choose random sequences, and fold each one. Sequences from the RNAcentral database were folded using the Vienna package with the same parameters as above, after removing any duplicate sequences.

Abstract Shapes

RNA SS can be abstracted in standard dot-bracket notation, where brackets denote bonds, and dots denote unbonded pairs. To obtain coarse-grained abstract shapes (Janssen and Giegerich 2015) of differing levels we used the RNashapes tool available at <https://bibiserv.cebitec.uni-bielefeld.de/rnashapes> (last accessed August 2021) and the Bioconda rnashapes package available at <https://anaconda.org/bioconda/rnashapes> (last accessed August 2021). The option to allow single-bonded pairs was selected, to accommodate the Vienna folded structures which can contain these.

Natural Sequences

For each length, we took all available natural noncoding RNA sequences from the RNAcentral database (RNAcentral Consortium 2021). Any repeated sequences were discarded. The sequence secondary structure predictions were made by the Vienna package, and then abstract RNA shapes for each structure were obtained.

The numbers of natural sequences, numbers of shapes, and fractions of natural shapes found by random sampling, were:

- $L = 40$: 3,160 sequences, yielding 18 unique shapes at level 3 (18/18 shapes were found by random sampling);
- $L = 55$: 8,619 sequences, yielding 63 unique shapes at level 3 (63/63 found);
- $L = 70$: 78,075 sequences, yielding 16 unique shapes at level 5 (16/16 found);

$L = 85$: 76,143 sequences, yielding 25 unique shapes at level 5 (24/25 found);
 $L = 100$: 20,314 sequences, yielding 35 unique shapes at level 5 (35/35 found);
 $L = 126$: 14,318 sequences, yielding 68 unique shapes at level 5 (68/68 found).

In total 224/225 (i.e., >99%) of the shapes in the database were found by relatively modest sampling.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We thank David McCandlish for discussions and for suggesting the H-atom analogy for RNA GP maps as models for development, and Michael Matthies for help with figure 1a.

Data Availability

The data underlying this article are available in RNACentral database at <https://rnacentral.org>. The data underlying this article will be shared on reasonable request to the corresponding author.

References

- Aguirre J, Buldú JM, Stich M, Manrubia SC. 2011. Topological structure of the space of phenotypes: the case of RNA neutral networks. *PLoS One* 6(10):e26324.
- Ahnert S, Johnston I, Fink T, Doye J, Louis A. 2010. Self-assembly, modularity, and physical complexity. *Phys Rev E Stat Nonlin Soft Matter Phys.* 82(2 Pt 2):026117.
- Ahnert SE. 2017. Structural properties of genotype–phenotype maps. *J R Soc Interface.* 14(132):20170275.
- Arthur W. 2001. Developmental drive: an important determinant of the direction of phenotypic evolution. *Evol Dev.* 3(4):271–278.
- Begun DJ, Lindfors HA, Kern AD, Jones CD. 2007. Evidence for de novo evolution of testis-expressed genes in the *Drosophila yakuba/Drosophila erecta* clade. *Genetics* 176(2):1131–1137.
- Bottou L, Curtis FE, Nocedal J. 2018. Optimization methods for large-scale machine learning. *Siam Rev.* 60(2):223–311.
- Cano AV, Payne JL. 2020. Mutation bias interacts with composition bias to influence adaptive evolution. *PLoS Comput Biol.* 16(9):e1008296.
- Catalán P, Manrubia S, Cuesta JA. 2020. Populations of genetic circuits are unable to find the fittest solution in a multilevel genotype–phenotype map. *J R Soc Interface.* 17(167):20190843.
- Charlesworth D, Barton NH, Charlesworth B. 2017. The sources of adaptive variation. *Proc R Soc B.* 284(1855):20162864.
- Cowperthwaite M, Economo E, Harcombe W, Miller E, Meyers L. 2008. The ascent of the abundant: how mutational networks constrain evolution. *PLoS Comput Biol.* 4(7):e1000110.
- de la Peña M, García-Robles I. 2010. Ubiquitous presence of the hammerhead ribozyme motif along the tree of life. *RNA.* 16(10):1943–1950.
- Dingle K, Camargo CQ, Louis AA. 2018. Input–output maps are strongly biased towards simple outputs. *Nat Commun.* 9(1):761.
- Dingle K, Schaper S, Louis AA. 2015. The structure of the genotype–phenotype map strongly constrains the evolution of non-coding RNA. *Interface Focus* 5(6):20150053.
- Ellington AD, Szostak JW. 1990. In vitro selection of RNA molecules that bind specific ligands. *Nature* 346(6287):818–822.
- ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414):57–74.
- Fontana W. 2002. Modelling ‘evo-devo’ with RNA. *Bioessays* 24(12):1164–1177.
- Fontana W, Konings DA, Stadler PF, Schuster P. 1993. Statistics of RNA secondary structures. *Biopolymers* 33(9):1389–1404.
- García-Martín JA, Catalán P, Manrubia S, Cuesta JA. 2018. Statistical theory of phenotype abundance distributions: a test through exact enumeration of genotype spaces. *Europhys Lett.* 123(2):28001.
- Giegerich R, Voss B, Rehmsmeier M. 2004. Abstract shapes of RNA. *Nucleic Acids Res.* 32(16):4843–4851.
- Gilbert W. 1986. Origin of life: the RNA world. *Nature* 319(6055):618.
- Gomez K, Bertram J, Masel J. 2020. Mutation bias can shape adaptation in large asexual populations experiencing clonal interference. *Proc Biol Sci.* 287(1937):20201503.
- Gould SJ. 2002. The structure of evolutionary theory. Cambridge (MA): Harvard University Press.
- Greenbury SF, Johnston IG, Louis AA, Ahnert SE. 2014. A tractable genotype–phenotype map modelling the self-assembly of protein quaternary structure. *J R Soc Interface.* 11(95):20140249.
- Greenbury SF, Schaper S, Ahnert SE, Louis AA. 2016. Genetic correlations greatly increase mutational robustness and can both reduce and enhance evolvability. *PLoS Comput Biol.* 12(3):e1004773.
- Hofacker I, Fontana W, Stadler P, Bonhoeffer L, Tacker M, Schuster P. 1994. Fast folding and comparison of RNA secondary structures. *Monatsh Chem.* 125(2):167–188.
- Jablonski D. 2020. Developmental bias, macroevolution, and the fossil record. *Evol Dev.* 22(1–2):103–125.
- Jacob F. 1977. Evolution and tinkering. *Science* 196(4295):1161–1166.
- Janssen S, Giegerich R. 2015. The RNA shapes studio. *Bioinformatics* 31(3):423–425.
- Johnston IG, Ahnert SE, Doye JP, Louis AA. 2011. Evolutionary dynamics in a simple model of self-assembly. *Phys Rev E Stat Nonlin Soft Matter Phys.* 83(6 Pt 2):066105.
- Johnston IG, Dingle K, Greenbury SF, Camargo CQ, Doye JPK, Ahnert SE, Louis AA. 2021. Symmetry and simplicity spontaneously emerge from the algorithmic nature of evolution. *bioRxiv.* <https://doi.org/10.1101/2021.07.28.454038>.
- Jorg T, Martin O, Wagner A. 2008. Neutral network sizes of biological RNA molecules can be computed and are not atypically small. *BMC Bioinformatics.* 9(1):464.
- Kalvari I, Argasinska J, Quinones-Olvera N, Nawrocki EP, Rivas E, Eddy SR, Bateman A, Finn RD, Petrov AI. 2018. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.* 46(D1):D335–D342.
- Kalvari I, Nawrocki EP, Argasinska J, Quinones-Olvera N, Finn RD, Bateman A, Petrov AI. 2018. Non-coding RNA analysis using the Rfam database. *Curr Protoc Bioinformatics.* 62(1):e51.
- Knight R, De Sterck H, Markel R, Smit S, Oshmyansky A, Yarus M. 2005. Abundance of correctly folded RNA motifs in sequence space, calculated on computational grids. *Nucleic Acids Res.* 33(18):5924–5935.
- Laland K, Uller T, Feldman M, Sterelny K, Müller GB, Moczek A, Jablonka E, Odling-Smee J, Wray GA, Hoekstra HE, et al. 2014. Does evolutionary theory need a rethink? *Nature* 514(7521):161–164.
- Laland KN, Sterelny K, Odling-Smee J, Hoppitt W, Uller T. 2011. Cause and effect in biology revisited: is Mayr’s proximate–ultimate dichotomy still useful? *Science* 334(6062):1512–1516.
- Lorenz R, Bernhart SH, Zu Siederdisen CH, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. Vienna package 2.0. *Algorithms Mol Biol.* 6(1):26.
- Lorenz WA, Ponty Y, Clote P. 2008. Asymptotics of RNA shapes. *J Comput Biol.* 15(1):31–63.
- Louis AA. 2016. Contingency, convergence and hyper-astronomical numbers in biological evolution. *Stud Hist Philos Biol Biomed Sci.* 58:107–116.
- Love AC. 2015. Conceptual change in biology. Vol. 307. Boston: Springer.
- Lynch M. 2007. The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc Natl Acad Sci U S A.* 104(Suppl 1): 8597–8604.

- Manrubia S, Cuesta JA, Aguirre J, Ahnert SE, Altenberg L, Cano AV, Catalán P, Diaz-Uriarte R, Elena SF, García-Martín JA, et al. 2021. From genotypes to organisms: state-of-the-art and perspectives of a cornerstone in evolutionary dynamics. *Phys Life Rev.* 38(55):55–106.
- Markham NR, Zuker M. 2008. UNAFold. In: Keith JM, editor. *Bioinformatics. Methods in molecular biology*TM. Vol. 453. Clifton (NJ): Springer. p. 3–31.
- Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH. 2004. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci U S A.* 101(19):7287–7292.
- Mattick JS, Makunin IV. 2006. Non-coding RNA. *Hum Mol Genet.* 15(Suppl 1):R17–R29.
- Mayr E. 1961. Cause and effect in biology. *Science* 134(3489):1501–1506.
- McCandlish DM, Stoltzfus A. 2014. Modeling evolution using the probability of fixation: history and implications. *Q Rev Biol.* 89(3):225–252.
- McGhee G. 2007. *The geometry of evolution: adaptive landscapes and theoretical morphospaces*. Cambridge, United Kingdom: Cambridge University Press.
- Miao Z, Adamiak RW, Antczak M, Boniecki MJ, Bujnicki JM, Chen S-J, Cheng CY, Cheng Y, Chou F-C, Das R, et al. 2020. RNA-puzzles round iv: 3D structure predictions of four ribozymes and two aptamers. *RNA* 26(8):982–995.
- Mingard C, Skalse J, Valle-Pérez G, Martínez-Rubio D, Mikulik V, Louis AA. 2019. Neural networks are a priori biased towards boolean functions with low entropy. *arXiv preprint arXiv:1909.11522*.
- Mingard C, Valle-Pérez G, Skalse J, Louis AA. 2021. Is SGD a Bayesian sampler? Well, almost. *J Mach Learn Res.* 22(79):1–64.
- Mituyama T, Yamada K, Hattori E, Okida H, Ono Y, Terai G, Yoshizawa A, Komori T, Asai K. 2009. The functional RNA database 3.0: databases to support mining and annotation of functional RNAs. *Nucleic Acids Res.* 37(Database):D89–D92.
- Nebel ME, Scheid A. 2009. On quantitative effects of RNA shape abstraction. *Theory Biosci.* 128(4):211–225.
- Neme R, Amador C, Yildirim B, McConnell E, Tautz D. 2017. Random sequences are an abundant source of bioactive RNAs or peptides. *Nat Ecol Evol.* 1(6):1–7.
- Oliver CG, Reinharz V, Waldispühl J. 2019. On the emergence of structural complexity in RNA replicators. *RNA* 25(12):1579–1591.
- Palazzo AF, Lee ES. 2015. Non-coding RNA: what is functional and what is junk? *Front Genet.* 6:2.
- Raup DM. 1966. Geometric analysis of shell coiling: general problems. *J Paleontol.* 40:1178–1190.
- Reuter JS, Mathews DH. 2010. Rnastructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics.* 11(1):129–129.
- Rivas E, Clements J, Eddy SR. 2017. A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat Methods.* 14(1):45–48.
- Rivas E, Clements J, Eddy SR. 2020. Estimating the power of sequence covariation for detecting conserved RNA structure. *Bioinformatics* 36(10):3072–3076.
- RNAcentral Consortium. 2021. RNAcentral 2021: secondary structure integration, improved sequence search and new member databases. *Nucleic Acids Res.* 49(D1):D212–D220.
- Salehi-Ashtiani K, Szostak J. 2001. In vitro evolution suggests multiple origins for the hammerhead ribozyme. *Nature* 414(6859):82–83.
- Salisbury FB. 1969. Natural selection and the complexity of the gene. *Nature* 224(5217):342–343.
- Schaper S, Louis AA. 2014. The arrival of the frequent: how bias in genotype-phenotype maps can steer populations to local optima. *PLoS One* 9(2):e86635.
- Schlick T, Pyle AM. 2017. Opportunities and challenges in RNA structural modeling and design. *Biophys J.* 113(2):225–234.
- Scholl R, Pigliucci M. 2015. The proximate–ultimate distinction and evolutionary developmental biology: causal irrelevance versus explanatory abstraction. *Biol Philos.* 30(5):653–670.
- Schuster P, Fontana W, Stadler P, Hofacker I. 1994. From sequences to shapes and back: a case study in RNA secondary structures. *Proc Biol Sci.* 255(1344):279–284.
- Smit S, Yarus M, Knight R. 2006. Natural selection is not required to explain universal compositional patterns in rRNA secondary structure categories. *RNA* 12(1):1–14.
- Smith JM. 1970. Natural selection and the concept of a protein space. *Nature* 225:563–564.
- Smith JM, Burian R, Kauffman S, Alberch P, Campbell J, Goodwin B, Lande R, Raup D, Wolpert L. 1985. Developmental constraints and evolution: a perspective from the mountain lake conference on development and evolution. *Q Rev Biol.* 60(3):265–287.
- Stich M, Briones C, Manrubia SC. 2008. On the structural repertoire of pools of short, random RNA sequences. *J Theor Biol.* 252(4):750–763.
- Stoltzfus A. 2019. Understanding bias in the introduction of variation as an evolutionary cause. In: Uller T, Laland KN, editors. *Evolutionary Causation: Biological and Philosophical Reflections*. Boston: MIT Press.
- Stoltzfus A, McCandlish DM. 2017. Mutational biases influence parallel adaptation. *Mol Biol Evol.* 34(9):2163–2172.
- Svensson EI, Berger D. 2019. The role of mutation bias in adaptive evolution. *Trends Ecol Evol.* 34(5):422–434.
- Tautz D, Domazet-Lošo T. 2011. The evolutionary origin of orphan genes. *Nat Rev Genet.* 12(10):692–702.
- Thiel BC, Flamm C, Hofacker IL. 2017. RNA structure prediction: from 2D to 3D. *Emerg Top Life Sci.* 1(3):275–285.
- Thompson D. 1942. *On growth and form*. Cambridge, United Kingdom: Cambridge University Press.
- Tuerk C, Gold L. 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249(4968):505–510.
- Uller T, Laland K. 2019. *Evolutionary causation: biological and philosophical reflections*. Vol. 23. Cambridge (MA): The MIT Press.
- Uller T, Moczek AP, Watson RA, Brakefield PM, Laland KN. 2018. Developmental bias and evolution: a regulatory network perspective. *Genetics* 209(4):949–966.
- Valle-Pérez G, Camargo CQ, Louis AA. 2018. Deep learning generalizes because the parameter-function map is biased towards simple functions. In: Conference paper at ICLR 2019. Available from: <https://openreview.net/forum?id=rye4g3AqFm>
- Vu MM, Jameson NE, Masuda SJ, Lin D, Larralde-Ridaura R, Lupták A. 2012. Convergent evolution of adenosine aptamers spanning bacterial, human, and random sequences revealed by structure-based bioinformatics and genomic selex. *Chem Biol.* 19(10):1247–1254.
- Wagner A. 2005. *Robustness and evolvability in living systems*. Princeton (NJ): Princeton University Press.
- Wagner A. 2011. *The origins of evolutionary innovations: a theory of transformative change in living systems*. Oxford, United Kingdom: Oxford University Press.
- Wagner A. 2014. *Arrival of the fittest: solving evolution's greatest puzzle*. New York: Penguin.
- Wagner GP, Altenberg L. 1996. Perspective: complex adaptations and the evolution of evolvability. *Evolution* 50(3):967–976.
- Weiße M, Ahnert SE. 2018. Phenotypes can be robust and evolvable if mutations have non-local effects on sequence constraints. *J R Soc Interface.* 15(138):20170618.
- West GB, Brown JH. 2005. The origin of allometric scaling laws in biology from genomes to ecosystems: towards a quantitative unifying theory of biological structure and organization. *J Exp Biol.* 208(Pt 9):1575–1592.
- Wilke C, Wang J, Ofria C, Lenski R, Adami C. 2001. Digital organisms: survival of the flattest. *Nature* 412(6844):331–333.
- Wilson BA, Foy SG, Neme R, Masel J. 2017. Young genes are highly disordered as predicted by the preadaptation hypothesis of de novo gene birth. *Nat Ecol Evol.* 1(6):1–6.
- Yampolsky L, Stoltzfus A. 2001. Bias in the introduction of variation as an orienting factor in evolution. *Evol Dev.* 3(2):73–83.