



Research

Cite this article: 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: 20150053.
<http://dx.doi.org/10.1098/rsfs.2015.0053>

One contribution of 12 to a theme issue
'Are there limits to evolution?'

Subject Areas:

biophysics, computational biology,
systems biology

Keywords:

genotype–phenotype map, arrival of
variation, robustness, bias in development

Author for correspondence:

Ard A. Louis
e-mail: ard.louis@physics.ox.ac.uk

[†]Present address: Bayer Technology Services GmbH, BTS-TD-ET-CSM, Building B106, 51368 Leverkusen, Germany.

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rsfs.2015.0053> or via <http://rsfs.royalsocietypublishing.org>.

The structure of the genotype– phenotype map strongly constrains the evolution of non-coding RNA

Kamaludin Dingle^{1,2,3}, Steffen Schaper^{1,†} and Ard A. Louis¹

¹Rudolf Peierls Centre for Theoretical Physics, University of Oxford, Oxford OX1 3NP, UK

²Systems Biology DTC, University of Oxford, Oxford, UK

³Department of Mathematics and Natural Sciences, Gulf University for Science and Technology, Block 5, West Mishref, Kuwait

The prevalence of neutral mutations implies that biological systems typically have many more genotypes than phenotypes. But, can the way that genotypes are distributed over phenotypes determine evolutionary outcomes? Answering such questions is difficult, in part because the number of genotypes can be hyper-astronomically large. By solving the genotype–phenotype (GP) map for RNA secondary structure (SS) for systems up to length $L = 126$ nucleotides (where the set of all possible RNA strands would weigh more than the mass of the visible universe), we show that the GP map strongly constrains the evolution of non-coding RNA (ncRNA). Simple random sampling over genotypes predicts the distribution of properties such as the mutational robustness or the number of stems per SS found in naturally occurring ncRNA with surprising accuracy. Because we ignore natural selection, this strikingly close correspondence with the mapping suggests that structures allowing for functionality are easily discovered, despite the enormous size of the genetic spaces. The mapping is extremely biased: the majority of genotypes map to an exponentially small portion of the morphospace of all biophysically possible structures. Such strong constraints provide a non-adaptive explanation for the convergent evolution of structures such as the hammerhead ribozyme. These results present a particularly clear example of bias in the arrival of variation strongly shaping evolutionary outcomes and may be relevant to Mayr's distinction between proximate and ultimate causes in evolutionary biology.

1. Introduction

Many questions about the limits of evolution hinge not only on what has happened in natural history, but also on *counterfactuals*: what did not happen, but perhaps could have. Re-run the tape of life [1,2] and what parts of phenotype space—the set of all possible phenotypes [3]—would be occupied? Typically, only a miniscule fraction of the phenotype space has been explored throughout natural history. The reasons given for this phenomenon usually combine adaptive arguments: some parts of phenotype space yield higher fitness than others, with arguments based on contingency: nature has not had time to explore all of phenotype space. However, an evolutionary search does not occur by uniform sampling over phenotypes, but rather by random mutations in the space of genotypes. Does this basic fact alter the way that phenotype space is explored and occupied? To answer such whole genotype–phenotype (GP) map questions is difficult, in part because the number of possible genotypes typically grows exponentially with length, and so rapidly becomes unimaginably vast [4–6], and in part because biological systems with sufficiently tractable GP maps are rare.

One system where progress can be made is the mapping from sequences to the structure of RNA. Although simpler than many biological phenotypes, RNA is interesting and important, because it can fulfil multiple roles both as an information carrier (e.g. messenger RNA and viral RNA) and as functional non-coding RNA

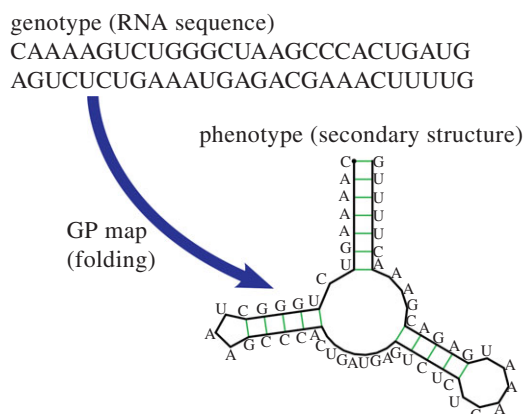


Figure 1. Schematic of the mapping from an RNA sequence to an SS. Here, for an $L = 55$ type III hammerhead ribozyme with three stems. Any sequence that folds to the same SS topology is considered to map to the same SS phenotype. (Online version in colour.)

(ncRNA) [7], including chemically active catalysts (e.g. ribozymes) and key structural components of larger self-assembled structures (e.g. ribosomal RNA). One reason RNA is so versatile is that it can fold into complex three-dimensional structures. The bonding pattern of these structures is called the secondary structure (SS), which is an important determinant of the three-dimensional structure and biological activity of ncRNA molecules, and so can be treated as a (simplified) phenotype in its own right [8,9] (figure 1).

Fast algorithms exist to predict the free energy minimum SS for a given sequence, and these are thought to be fairly accurate, especially for shorter strands [10,11]. Moreover, extensive databases exist for functional ncRNA [7]. For these reasons, this computationally tractable yet biologically relevant sequence to structure mapping is uniquely suited for investigating ‘whole genotype–phenotype map’ properties that may point towards general principles relevant for a wider class of systems, and has been extensively studied over the past few decades [6,12–20].

Nevertheless, the RNA sequence to SS GP map also suffers from the exponential growth of the size of genetic space, which has limited prior comprehensive GP studies to fairly small lengths, making direct comparison with evolutionary outcomes difficult. Here, we show that many detailed properties of this GP map can, in fact, be calculated, even for lengths as long as $L = 126$, where the set of all possible strands would weigh more than the mass of the visible universe (see Methods). One reason this is possible is because the map is highly biased towards a small fraction of phenotypes that take up the majority of genotypes. This bias means that sampling over genotypes (which we will call G-sampling) generates significantly different outcomes from sampling over the morphospace [3] of all shapes (phenotypes; which we will call P-sampling). The existence of such highly peaked distributions in the mapping from genotypes to phenotypes also explains, in a way that is reminiscent of statistical mechanics, why sampling a relatively small number of genotypes is enough to determine certain key properties of RNA SS.

Perhaps most strikingly, we find that the distributions of several properties of natural ncRNA taken from the function RNA database (fRNAdb) [21], including the number of stems, the mutational robustness and the number of genotypes per SS phenotype, are very similar to what we obtain from random sampling over genotypes, and significantly different from uniform sampling over phenotypes. This

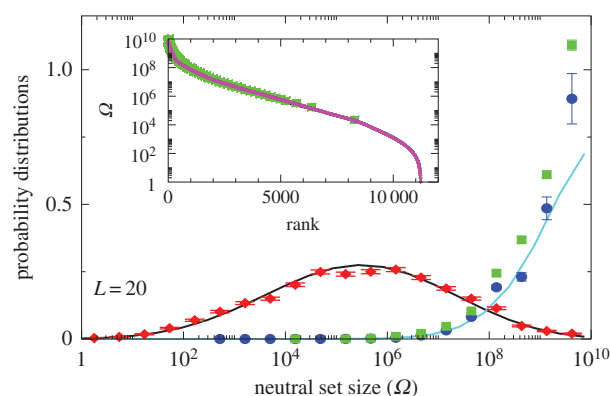


Figure 2. Comparison of P-sampled and G-sampled distributions to natural data for $L = 20$ RNA. The P-sampled $P_P(\Omega)$ (red diamonds) measures the probability distribution for a phenotype to have a given NS size Ω . It differs markedly from G-sampled $P_G(\Omega)$ (blue circles), generated by random sampling over genotypes. Error bars arise from binning data. The black and cyan lines are theoretical approximations to $P_P(\Omega)$ and $P_G(\Omega)$, respectively (see Methods). The probability distribution of Ω for the SSs all 7327 (non-trivial) $L = 20$ sequences for *Drosophila melanogaster* from the fRNAdb database [21] (green squares) is much closer to the G-sampled $P_G(\Omega)$ than to the P-sampled $P_P(\Omega)$. Inset: all 11 218 SS phenotypes (purple triangles) ranked by NS size Ω . There is strong bias, just 5% of phenotypes take up 58% of all genotypes. The 7327 natural data points (green squares) are clustered at lower rank (larger Ω). (Online version in colour.)

result does not mean that natural selection can be ignored, but rather, as we will argue below, that the mapping strongly prescribes which parts of morphospace are presented to natural selection as potential variation. Variation can be selected only if it arrives [22,23].

2. Results

2.1. P-sampling over phenotypes differs significantly from G-sampling over genotypes

We first analyse an exhaustive enumeration of all sequences for $L = 20$ RNA, the largest system for which this has so far been accomplished. The $4^{20} \approx 10^{12}$ sequences were folded with the Vienna Package [10], and map to $N_P = 11\,218$ unique bonded SSs and one trivial structure with no bonds, as their free-energy minima [20]. The set of all sequences that map onto a given SS is called the neutral set (NS), and we use Ω to denote the NS size (the number of sequences in the NS). This mapping exhibits strong phenotype bias [6,12,13,17–20]. For example, Ω varies by over 10 orders of magnitude for $L = 20$ (figure 2).

Next, we introduce the concept of P-sampling, that is, uniformly sampling over the set of possible phenotypes (the morphospace), and define $P_P(\Omega)$ as the probability distribution that a randomly chosen phenotype has NS size Ω . We calculate distributions for fixed L and bin data uniformly in $S = \log_{10}(\Omega)$, but write $P_P(\Omega)$ and $P_G(\Omega)$ for simplicity. $P_P(\Omega)$ has a maximum when Ω is about half of the maximum value of the exponent $\log(\Omega) \equiv U \approx 10$ (figure 2). We ignore the trivial structure with no bonds for which $\Omega \equiv 10^T$. For $L = 20$, $T \approx 11.56$; for larger L , $T/U \rightarrow 1$ and the probability of finding the trivial structure tends to zero (see Methods).

Instead of P-sampling, one could also sample uniformly over sequences (genotypes), which we refer to as G-sampling,

giving $P_G(\Omega) \propto \Omega P_P(\Omega)$, which highlights structures from the large Ω tail of $P_P(\Omega)$, as can be seen in figure 2.

Novel variation does not arise by uniform random sampling in the morphospace of all physically permissible phenotypes (P-sampling), but instead by processes such as mutation that change genotypes. While evolutionary dynamics do not proceed by simple uniform G-sampling either, recent detailed population genetics calculations for RNA [20] have shown that the rate at which novel variation (a particular SS) arises in an evolving population is, to first order, directly proportional to the NS size of the SS phenotype, and so also to $P_G(\Omega)$. This proportionality holds for a wide range of mutation rates and population sizes. It was also demonstrated explicitly that strong phenotype bias can overcome fitness differences in an RNA system [20]. Bias should therefore affect outcomes for multiple evolutionary scenarios. Indeed, important prior studies have suggested that natural RNA could have larger than average Ω [17,18]. We took every $L = 20$ sequence in the fRNAdb database [21] for *Drosophila melanogaster* (see Methods and electronic supplementary material) and calculated the NS size Ω of its associated SS using the neutral set size estimator (NSSE) from [18]. Not only are SSs with larger Ω overrepresented, but the entire natural distribution is remarkably close to the genotype sampled $P_G(\Omega)$ (figure 2).

2.2. Sampling for lengths up to $L = 126$

While these results for $L = 20$ are suggestive, distributions for larger length RNA are needed to ensure that we are not just observing database biases or artefacts of the short length. Because the number of sequences grows exponentially with increasing L , exhaustive enumeration is not an option for lengths much larger than $L = 20$. Instead, we estimate the NS size distributions by randomly sampling genotypes, folding them into an SS, and then measuring their NS size with the NSSE (Methods). This process naturally generates $P_G(\Omega)$; $P_P(\Omega)$ can be backed out by dividing by Ω and normalizing. However, it is hard to sample SSs with small Ω , so $P_P(\Omega)$ is only partially determined. To make progress, we use a simple analytical ansatz based on a log-binomial approximation to the distributions (see Methods), which works well for both $P_G(\Omega)$ and $P_P(\Omega)$ for $L = 20$ (figure 2). We compare this approximation with sampled data for $L = 20$ up to $L = 126$ (figure 3 and electronic supplementary material, figures S1 and S2). In each case, the analytic fit to $P_G(\Omega)$ is excellent, and the fit to $P_P(\Omega)$ works well, giving confidence that this form provides a reasonable approximation to the full $P_P(\Omega)$.

2.3. Entropy and the bias ratio

The $P_G(\Omega)$ distributions can be further quantified by the Shannon entropy $H = -\sum_{k=1}^{N_P} P(p_k) \log_2(P(p_k))$, where $P(p_k)$ is the probability of choosing one of the N_P possible phenotypes p_k by G-sampling. The exponential of the entropy, 2^H , is used in statistical physics and information theory [24] as a measure of the effective number of states. Because $P(p_k) = \Omega(p_k)/4^L$, it follows straightforwardly that

$$2^H = \frac{4^L}{10^{\bar{S}_G}}, \quad (2.1)$$

where \bar{S}_G is a G-sampled average of $S = \log(\Omega)$. Thus to measure the ‘effective number’ of phenotypes that take up the majority of genotypes, one only needs to determine \bar{S}_G . This

can be done rapidly, because the G-sampled distributions are sharply peaked in a manner reminiscent of statistical mechanics [24].

As an example of how the highly peaked nature of these distributions facilitates the calculation of averages, consider the case of $L = 126$. We find that $\bar{S}_G \approx 51.5$ with standard deviation (s.d.) $\bar{\sigma}_G \approx 2.1$, so that 95% of SS structures found by G-sampling have S in range 51.5 ± 4.2 , which is very narrow compared with the full range $[0, 64.5]$ and comparatively far (about 6 s.d.) from the phenotype averaged $\bar{S}_P \approx 38.6$. Rather strikingly, this implies that, even though the set of all possible $L = 126$ nucleotide sequences would weigh more than the mass of the observable universe (see Methods), sampling the NS size for just 10 randomly chosen $L = 126$ structures is enough to fix the exponent \bar{S}_G to within about 1.3% relative accuracy, which allows us to determine, for example, the number of relevant states 2^H via equation (2.1).

To further quantify the bias, we introduce the *bias ratio* $\beta \in (0, 1]$ as

$$\beta = \frac{2^H}{N_P} \approx 0.25 \times 0.91^L, \quad (2.2)$$

which can be interpreted as the ratio of the effective number of phenotypes that take up most of the sequences to the total number of phenotypes N_P . For example, if all Ω are equal, then $2^H = 2^{\log_2(N_P)} = N_P$, $\beta = 1$, and bias is weak. On the other hand, if just one phenotype accounts for nearly all the genotypes, then $2^H \approx 2^0 = 1$, $\beta \rightarrow 0$ and bias is very strong.

The number of phenotypes N_P can be estimated from $P_P(\Omega)$; $N_P \approx 0.13 \times 1.76^L$ fits the data well (see Methods). From this, we find $\beta \approx 0.25 \times 0.91^L$, which shrinks with increasing L . Typically, we find that about 75% of genotypes map to the ‘effective number’ $\beta N_P = 2^H$ of structures. Nevertheless, $\beta N_P = 2^H \approx 0.033 \times 1.60^L$, and continues to grow exponentially. For $L = 55$, $\beta \approx 0.001$ and $\beta N_P \approx 4 \times 10^9$ of the $N_P \approx 4 \times 10^{12}$ phenotypes take up about 75% of the genotypes. For $L = 126$, $\beta \approx 2 \times 10^{-6}$ and $\beta N_P \approx 1 \times 10^{24}$. Even though bias greatly reduces the effective number of phenotypes accessible to mutations, their number continues to grow exponentially, and so can still be extremely large for this system.

2.4. Comparison with fRNAdb database of functional non-coding RNA

To explore how this bias plays out in nature, we took, for lengths ranging from $L = 20$ to 126, every sequence in the fRNAdb database [21] and calculated the NS size Ω of its associated SS. The correspondence between $P_G(\Omega)$ from random sampling and the distribution of Ω from the database is striking (figure 3 and electronic supplementary material, figures S1 and S2). Not only are these natural RNAs mainly drawn from the minuscule fraction β of SSs with large Ω , but their distribution is remarkably close to what one would obtain from randomly sampling genotypes. Rank plots for $L = 55$ (figure 4) further illustrate how natural ncRNAs are constrained mainly to the set of $\beta N_P = 2^H$ ‘relevant structures’.

These results are, at first glance, surprising, because many ncRNAs have well-defined functional roles and so will have been subject to natural selection for which there exists extensive evidence [7]. Some insight may be gleaned from an exchange that occurred not long after the start of the molecular

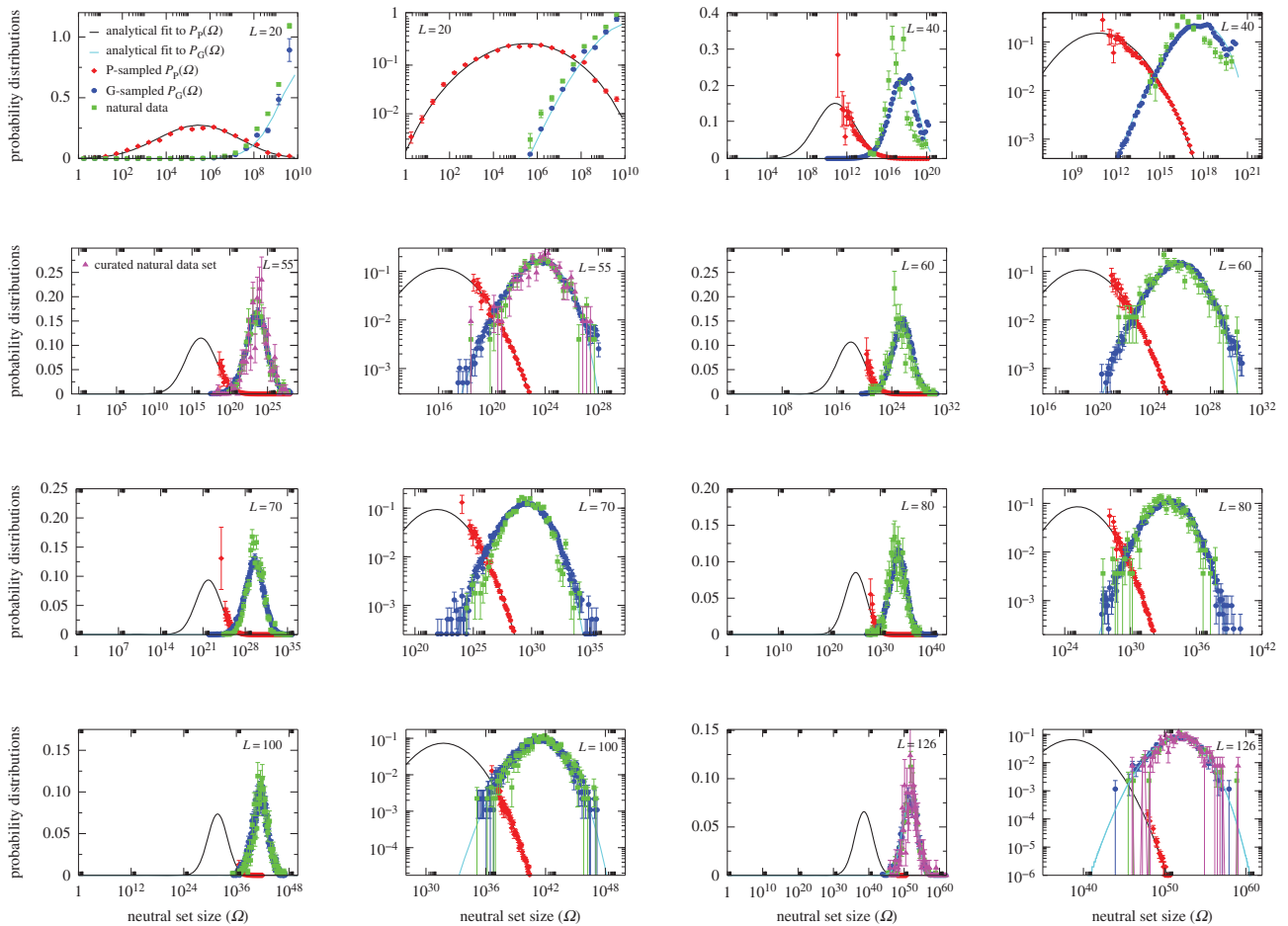


Figure 3. Neutral set size distributions illustrate how bias constrains natural RNA secondary structures. Sampled distributions $P_P(\Omega)$ (red diamonds) and $P_G(\Omega)$ (blue circles) and analytic approximations to $P_P(\Omega)$ (black lines) and $P_G(\Omega)$ (cyan lines) are compared with ncRNA from the fRNAdb database [21] (green squares). For each length (denoted in the top corners), the data are shown both on a lin–log and log–log graph to highlight different parts of the distributions. The natural data are remarkably close to the G-sampled $P_G(\Omega)$, but quite far from the P-sampled $P_P(\Omega)$. The number of natural structures plotted are 7327 for $L = 20$ (which is repeated from figure 2 for clarity of comparison); 658 for $L = 40$; 472 for $L = 50$; 350 for $L = 60$; 2263 for $L = 70$; 553 for $L = 80$; 731 for $L = 90$; 891 for $L = 100$; and 291 for $L = 126$. In each case, all structures in the fRNAdb [21] database are used, except for $L = 20$, where only structures for *Drosophila melanogaster* are used, and for $L = 55$ and $L = 126$, where we also plot a smaller curated dataset of 213 and 172 structures, respectively (magenta triangles), where the SS is known to be important (see electronic supplementary material). Smaller numbers of data lead to larger binning error bars, but the curated sets are clearly very similar to the full set of fRNAdb structures. (Online version in colour.)

revolution in biology. Frank Salisbury [4] expressed doubt that evolutionary search could work, because genetic spaces are typically exponentially large, and, he claimed, are probably sparsely populated with functional phenotypes. In a famous response [5], John Maynard Smith pointed out that evolution is aided by neutral mutations, because these imply many fewer phenotypes than genotypes and also connect genotypes into neutral networks that can be explored by evolving populations. Phenotype bias may further facilitate evolutionary search by limiting potential outcomes. But, taken together, this still does not counter the heart of Salisbury's objection, namely that functional phenotypes are extremely rare and so hard to find.

Here, we instead suggest that the reason the 'null model' $P_G(\Omega)$ predicts what is found in nature so accurately is that 'good enough' SSs are relatively easy to find. Of course, the full functional ncRNA phenotype typically has a smaller NS size than the SS does, because it requires further constraints on the sequence to achieve functionality [8]. We postulate that while natural selection creates function by acting on such sequence constraints, it also automatically draws from a palette of easily accessible SS variation that is strongly pre-sculpted

by the mapping (see [14] for a similar suggestion). Another possibility might be that the correspondence with $P_G(\Omega)$ simply means that the SS has no adaptive value. However, this scenario is unlikely, because it is well established that the SS is an important determinant of the three-dimensional structure [6,8,9], which, in turn, helps determine function. Moreover, the correlation with $P_G(\Omega)$ remains strong when we curate the database for structures where SSs is known to be important. Altogether, this suggests that RNA SSs that facilitate biological function are—*contra* Salisbury—not rare, at least not within the set of 'relevant structures' most easily accessible to random mutations.

Salisbury's arguments are also undermined by *in vitro* evolution experiments that selected random RNA strands for self-cleaving catalytic activity and found that the hammerhead ribozyme repeatedly emerged, suggesting convergent evolution [25]. The hammerhead ribozyme appears so frequently in all three kingdoms of life that it has recently been termed ubiquitous [26]. We examined the 13 hammerhead ribozymes from the natural dataset [21] for $L = 55$, finding that $S = \log(\Omega) \approx 23.87 \pm 0.64$, very close to the peak of the G-sampled distribution at $\bar{S}_G \approx 23.4$ (see $L = 55$ panel in figures 3 and 4). We postulate that evolutionary

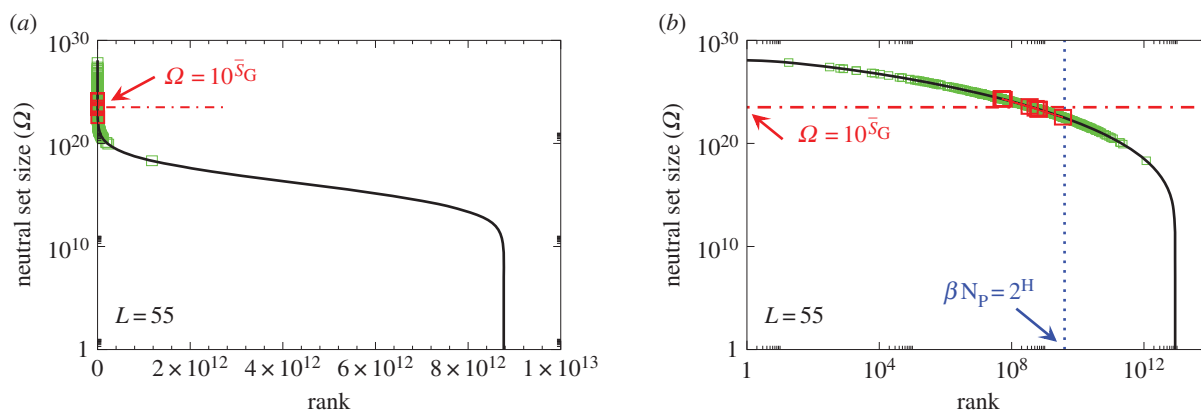


Figure 4. Ω versus rank plot for all $N_p \approx 4 \times 10^{12}$ possible SS structures for $L = 55$ shown as (a) log–lin and (b) log–log plots. The black line is from the analytical approximation and the green squares denote the natural data. The horizontal red dotted-dashed line is $\Omega = 10^{23} \approx 3 \times 10^{23}$, near the peak of the $P_G(\Omega)$ distribution for $L = 55$ in figure 3. The 13 $L = 55$ hammerhead ribozyme structures from the fRNA database (red squares) are clustered near this peak. The vertical blue dotted line in (b) denotes $\beta N_p = 2^9 \approx 4 \times 10^9$, the ‘effective’ number of SSSs. This set of $\beta \approx 0.1\%$ of all structures captures the majority ($\approx 75\%$) of natural structures. (Online version in colour.)

convergence is observed in these experiments and in nature, not so much because the hammerhead ribozyme is fitter than other possible self-cleaving enzymes but, rather, because it is particularly easy to find. There may even be many other self-cleaving ribozymes among the 99.9% of $L = 55$ structures that evolution is unlikely to search through [25]. It would in fact be interesting to devise artificial methods to search for such undiscovered ribozymes [27,28].

2.5. Distribution of stacks

Are the 2^H relevant structures different from the whole set of N_p possible structures? Our arguments above suggest that this should be the case for properties that correlate with NS size. Previous studies (typically on much smaller datasets) have found that structural features (e.g. distributions of stack and loop sizes) of the natural and random SSSs are quite similar [13], and that natural and random rRNA share strong similarities in the sequence nucleotide composition of SS motifs such as stems, loops and bulges [16], although natural RNAs are more stable than random RNAs [14]. Indeed, we find that the natural RNAs have slightly more bonds than in G-sampled structures (see electronic supplementary material, figure S3c,d).

We find that Ω correlates negatively with the number of stacks (i.e. sets of contiguous base-pairs) K (see electronic supplementary material, figure S6). The natural distribution of stacks closely follows the G-sampled distribution, but differs markedly from the P-sampled distribution (figure 4a). For example, the hammerhead ribozyme has three stems, close to the most likely number by random G-sampling for $L = 55$, but much less than the P-sampled average of ≈ 10 . Bias means that it will be difficult for evolution to find $L = 55$ structures with a large number of stacks, again raising the question of what kind of functionality is possible in principle that cannot be reached by evolution because of such phenotype bias constraints.

As an independent check on our stack predictions, we also obtained 214 natural experimentally determined SSSs from the STRAND RNA database [29], and plot the experimentally determined number of stacks versus length L in figure 5b. The majority of experimentally determined structures have numbers of stacks that are within 1 s.d. of the mean

calculated from G-sampling, as one would expect if our theoretical predictions were accurate. By contrast, the experimentally determined number of stacks differs significantly from P-sampling estimates.

2.6. Distribution of mutational robustness

Interestingly, the bias towards larger Ω also leads to structures with larger mutational robustness [6,12], and again natural data closely follow G-sampled distributions (figure 6). Larger robustness is considered to be advantageous [6], so that, in this important way, phenotype bias facilitates evolution. It is also interesting to note just how large the mutational robustness is for these data. For $L = 55$, there are of the order of $N_p \approx 8 \times 10^{12}$ phenotypes, so that the mean probability of finding a phenotype by randomly picking a genotype is of the order of 1×10^{-13} . Instead, for both the G-sampled and the P-sampled structures, the probability of a nearest neighbour generating the same phenotypes is of the order of 10^{12} times higher than random chance. As recently emphasized in [31], this large difference arises from genetic correlations, and typically lifts the robustness well over the minimal threshold $\delta = 1/3 L$ [5] ($\delta \approx 0.0061$ for $L = 55$) needed to generate large connected neutral networks.

3. Discussion

By solving properties of the GP map from sequences to RNA SSSs for strands up to $L = 126$ nucleotides in length, we show explicitly that the vast majority of sequences map to an exponentially small fraction of all possible phenotypes (a summary of scaling forms for key properties of RNA can be found in table 1). One consequence of this strong bias is that only an exponentially small proportion of the morphospace of possible structures can ever be presented to natural selection. Even if one could re-run the tape of life over again multiple times, many structures that are physically feasible and potentially biochemically functional are extremely unlikely to appear simply because they are inaccessible to evolutionary search.

While the existence of bias in the RNA GP map has been known for quite some time [6,12–20], and studies using smaller amounts of natural data have suggested that aspects of the mapping are reflected in nature [13,14,16–18], our ability to

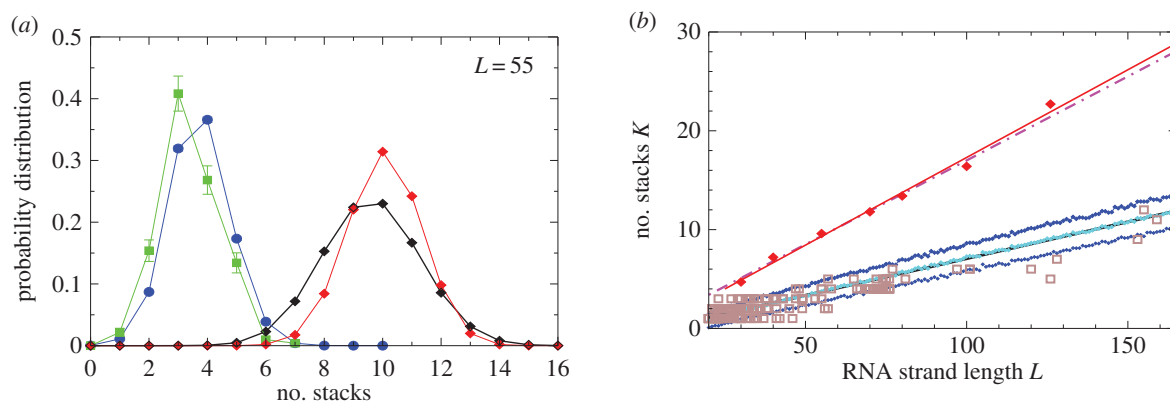


Figure 5. Natural stack distributions for ncRNA correlate with G-sampling but not with P-sampling. (a) Distribution of the number of stacks for $L = 55$ via G-sampling (blue circles) and P-sampling (black diamonds). Natural data (green squares) are remarkably close to the G-sampled distribution, and are drawn from a tiny fraction of the full morphospace of structures (the P-sampled distribution) with small numbers of stacks. The red diamonds are from a combinatorial estimate of stack number [30] that helps corroborate our P-sampling result. (b) Comparison with experimentally measured structures: the brown squares denote the number of stacks experimentally determined for 214 non-pseudo-knotted structures with $L > 20$ which were taken from the RNA STRAND database [29]. The cyan diamonds show \bar{K}_G , the mean number of stacks calculated by G-sampling with the Vienna package [10], which can be accurately fitted with $\bar{K}_G = 0.074 L - 0.377$. The blue diamonds show the G-sampled number of stacks 1 s.d. above or below the mean. The natural data (brown squares) from the RNA STRAND database are consistent with the G-sampled theoretical data. By contrast these natural data are far away from the expected number of stacks from P-sampling, shown here for an estimate that uses the linear relationship between the P-sampled distribution of $\log(\Omega)$ and K (electronic supplementary material, figure S6) to infer \bar{K}_P (red diamonds), which are well described by a linear fit $\bar{K}_P = 0.177 L - 0.443$ (solid red line). Independent estimates of \bar{K}_P come from [30], and include an asymptotic measure $\bar{K}_P \approx 0.1717 L$ (dashed-dotted line) (see electronic supplementary material). The close agreement between the two independent methods for estimating the mean number of stacks from P-sampling gives us further confidence in our fits to the full P-sampled distribution. (Online version in colour.)

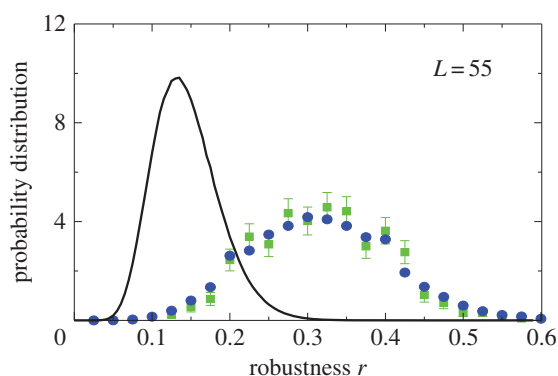


Figure 6. Natural robustness distributions for ncRNA correlate with G-sampling but not with P-sampling. The distribution of robustness, defined as the fraction of mutations r that retain the same SS phenotype, is given for $L = 55$ via P-sampling (black line) and G-sampling (blue circles). Natural data (green squares) are remarkably close to the G-sampled distribution, and are considerably more robust than the average of structures in the full morphospace. We also note that most phenotypes have a robustness that is above the threshold ($r > \delta \approx 0.0061$) needed for the formation of connected neutral networks [31]. (Online version in colour.)

calculate GP map properties for much larger L allows us to make a comprehensive comparison with the fRNAdb database for functional ncRNA [21]. While one might expect that, owing to bias, large Ω structures are relatively more plentiful in nature [17,18], perhaps the most surprising result of this study is just how closely the natural data follow the prediction of uniform G-sampling over genotypes. We find that the distribution of NS size Ω , the number of stacks S and the mutational robustness r of naturally occurring ncRNA all closely follow the G-sampled distributions, and deviate significantly from the P-sampled distributions. The distribution of bonds also deviates strongly from the P-sampled distribution, but, in contrast to the other distributions above,

it also exhibits a small deviation from the G-sampled distribution, with natural RNA having slightly more bonds [14] (see electronic supplementary material, figure S3c,d).

Why does it appear as if we can virtually ignore natural selection with G-sampling? We postulate that, even though the number of relevant structures remains extremely large, SSs that are good enough for function are nonetheless abundant, most likely because small differences between them do not matter that much. Instead, some broader coarse-grained structural features are likely to be sufficient for many functional roles. We suggest that natural selection works on this pre-sculpted variation mainly by further refining parts of the sequence. Some evidence for this picture can be gleaned by the fact that natural structures have slightly more bonds than G-sampled RNA structures do, suggesting selection for greater thermal stability [14].

Interestingly, the fact that many properties of ncRNA SSs so closely follow the G-sampled distribution is direct evidence against claims that the hyper-astronomically large size of genotype space makes functional structures virtually impossible to find. Such assertions were perhaps most famously made by Salisbury [4] over 45 years ago. Despite an illuminating response by Maynard Smith [5], and much experimental evidence to the contrary, such 'arguments from large numbers' remain a popular trope in anti-evolutionary polemics today.

The ease with which we can calculate distributions of RNA properties suggests an analogy to the concept of ergodicity in statistical physics, which means that ensemble averages are equivalent to time-averages. When ergodicity (approximately) holds, then it is often true in practice that sampling (via computer simulation for example) a relatively small number of 'typical' states (small compared with the total number of possible microstates) is sufficient to accurately calculate ensemble averages of key properties. Something akin to ergodicity may be operating in our case, because statistical averaging via G-sampling is close to the snapshot of many trajectories over

Table 1. Large L scaling of some key quantities for RNA SSs. For the number of sequences Ω for the NS of the largest non-trivial structure, defined as $\Omega = 10^U$, we find $U = (0.514 \pm 0.009)L - (0.20 \pm 0.5)$, whereas for the trivial structure, with $\Omega = 10^T$, we find $T = (0.5166 \pm 0.0009)L + (1.33 \pm 0.06)$ so that U becomes relatively more close to T as L increases. The G-sampled mean of $S = \log(\Omega)$ scales as $\bar{S}_G = (0.399 \pm 0.0014)L + (1.48 \pm 0.09)$. For large L , $\bar{S}_G \approx 0.78U$, whereas for P-sampling $\bar{S}_P \approx 0.63U$ so that $\bar{S}_P/\bar{S}_G \approx 0.8$. Similarly, the standard deviations of $\log(\Omega)$ can be directly calculated, and in the large L limit tend to $\bar{\sigma}_P/U \approx 0.37/\sqrt{L}$ and $\bar{\sigma}_G/U \approx 0.31/\sqrt{L}$. This explains analytically what can be observed qualitatively in figure 3 and the electronic supplementary material, figures S1 and S2: the $P_G(\Omega)$ distribution is slightly narrower than the $P_P(\Omega)$ distribution. As L increases both distributions become more sharply peaked relative to the total range $[0, U]$ and $P_G(\Omega)$ highlights SS phenotypes that are deeper into the tails of the $P_P(\Omega)$ distribution (and vice versa).

quantity	large L scaling form
total number of genotypes	$N_G = 4^L$
total number of SS phenotypes	$N_P \approx 0.13 \times 1.76^L$
mean Ω	$4^L/N_P \approx 7.7 \times 2.27^L$
largest non-trivial Ω	$10^U \approx 0.7 \times 3.27^L$
Ω for the trivial structure	$10^T \approx 21.4 \times 3.29^L$
probability to sample trivial structure	$P_T \approx 21.4 \times 0.82^L$
Ω near peak for phenotype sampling	$10^{\bar{S}_P} \approx 0.01 \times 2.1^L$
Ω near peak for genotype sampling	$10^{\bar{S}_G} \approx 30.2 \times 2.5^L$
Shannon entropy of distribution	$H \approx 0.675 L - 4.92$
'effective number' of SS phenotypes	$\beta N_P \approx 0.032 \times 1.60^L$
bias parameter	$\beta = 2^H/N_P \approx 0.25 \times 0.91^L$
G-sampled mean number of stacks	$\bar{K}_G = 0.074 L - 0.377$
P-sampled mean number of stacks	$\bar{K}_P = 0.177 L - 0.443$

time that we observe in the database, the latter being analogous to a time-average. On the other hand, evolution is not an equilibrium process. Indeed, the exponential bias in the GP map suggests that evolutionary waiting times for more rare phenotypes to appear also grows exponentially [20]. Thus, the number of novel phenotypic possibilities will continue to (slowly) increase with time. With large waiting times, contingency is more likely to play an important role. So if a certain evolutionary change is predicated on one of these rare phenotype fixing, then the process of waiting for these innovations may be non-ergodic.

The suggestion that biases in development or other internal processes could strongly affect evolutionary outcomes has traditionally been highly contentious (see [32,33] for an overview). For example, in a recent exchange, entitled: 'Does evolutionary theory need a rethink?', Laland *et al.* [34] argue in favour of the thesis by, among other things, advocating for the importance of developmental bias. In their rejoinder, Wray *et al.* [35] write: 'Lack of evidence also makes it difficult to evaluate the role that developmental bias may have in the evolution (or lack of evolution) of adaptive traits', and call for new evidence: 'The best way to elevate the prominence of

genuinely interesting phenomena such as... developmental bias... is to strengthen the evidence for their importance'. While the RNA system we study here is much simpler than a typical developmental system, that vice is also a virtue, because it allows us to make detailed calculations of the whole morphospace that can then be closely compared with natural data. Thus, RNA SS provides perhaps the clearest and most unambiguous evidence for the importance of bias in shaping evolutionary outcomes.

Bias in the GP map constrains outcomes and so naturally suggests one mechanism for homoplasy (where similar biological forms evolve independently) [36]. The causes of homoplasy are sometimes elaborated in the context of the difference between parallel evolution, where homoplasy is thought to occur because two organisms share a common genetic heritage, and convergence proper, where the same solution is found by different genetic means, and where the primary causal force is usually attributed to selection. While this binary distinction may be too simplistic (see [36–39] for some recent discussion), very roughly, parallel evolution is thought to be considerably weaker evidence than true convergent evolution is for the idea that re-running the tape of life would generate similar outcomes. Interestingly, the bias in the GP map discussed here does not fit into this twofold demarcation at all. Re-run the tape of life, and as long as RNA graces the replay, so will a very similar suite of molecular shapes. But the reason for this repetition is not a contingent common genetic history, nor the Allmacht of selection [40], but rather a different kind of 'deep structure in biology' [41].

Our ability to make detailed predictions about evolutionary outcomes as well as counterfactuals for RNA may also shed light on Mayr's famous distinction between proximate and ultimate causes in biology [42], which has been the subject of much recent debate in the literature [43]. This distinction has historically been used to argue against the role of developmental bias in determining evolutionary outcomes [33,44]. While, as also mentioned above, the RNA GP map is much simpler than a typical developmental system, it is instructive to consider how the bias described in this paper plays into Mayr's ultimate–proximate distinction. For example, what is the cause of convergent evolution of the hammerhead ribozyme [25]? The ultimate cause for a self-cleaving ribozyme emerging in populations or in *in vitro* experiments is surely natural selection for self-cleaving catalytic activity. But *why* is a three-stack structure repeatedly found, and not say a 10-stack structure, even though the latter are much more common in the morphospace? The cause here is not natural selection *per se* because it is unlikely that an efficient 10-stack ribozyme is biophysically or biochemically impossible to make. Instead, the explanatory force [44,45] for the 'why' question is mainly carried by a proximate GP mapping constraint, namely that the frequency with which 10-stack structures arise as potential variation is many orders of magnitude lower than the frequency with which three-stack structures do. Because the fittest can only be selected and survive if they arrive in the first place [22,23], the evolutionary mechanism that leads to convergence here might be better termed *the arrival of the frequent* [20].

We also note that the mapping constraint described here differs from classical physical constraints, which would act on the whole morphospace, and from phyletic constraints, which are contingent on evolutionary histories [46]. This mapping constraint has some resemblance to classical morphogenetic constraints which also bias the arrival of variation [47]. But it

also differs, because the latter are conceptualized at the level of phenotypes and developmental processes, and may have been shaped by prior selection, whereas the former constraint is a fundamental property of the mapping from genotypes to phenotypes and was not selected for (except perhaps at the origin of life itself).

Finally, strong phenotype bias is also found in model GP maps for protein tertiary [48,49] and quaternary structure [50], gene regulatory networks [51,52] and development [53], suggesting that some of the results discussed in this paper for RNA may hold more widely in biology. While all these GP maps only capture a tiny fraction of the full biological complexity of an organism, if the bias is also (exponentially) strong, then its effects on the rate with which novel variation appears are likely to persist even when further biological details are included. Although more evidence needs to be gathered before making firm pronouncements, it may well be that we need to let go of the commonly held expectation that variation is isotropic in phenotype space or morphospace [54]. Or to put it more provocatively, perhaps our null models should by default assume that variation is highly anisotropic and biased towards certain outcomes over others, unless there is direct evidence to the contrary.

4. Methods

4.1. Folding RNA structures

To map a sequence to an SS, we use the Vienna package [10] with all parameters set to their default values (e.g. the temperature $T = 37^\circ\text{C}$). This software employs dynamic programming techniques to efficiently fold sequences based on thermodynamic rules. Methods of this type are widely used, have been extensively tested and are thought to be relatively accurate. For example, a related method [55] was recently shown to correctly predict $73 \pm 9\%$ of canonical base pairs for a database of known RNA structures up to lengths of 700 nucleotides. Generally, these methods are expected to work better for shorter strands [56], and should work well for the lengths we explore. However, they typically cannot correctly predict pseudo-knots. Knowledge-based methods that also take input from known structures and other information may be more accurate for predicting the structures of individual sequences [57], but such methods could introduce biases for GP maps because they take input from natural structures. Thermodynamically based methods such as the Vienna package are therefore probably better suited for working out global properties of the entire GP map, including structures that have not (yet) been found in nature. For these reasons, this package has been most frequently used in studies of full GP maps [6,12–20]. To check our Vienna package results, we compared RNA structure package [55] calculations for the G-distributions of stacks and bonds (electronic supplementary material, figure S3), finding very similar distributions. We also compared the two packages for other motifs like bulges, loops and junctions, finding again very similar predictions (graphs are not shown).

4.2. Exponential growth of the number of strands with strand length L

To illustrate how rapidly the number of sequences grows with length consider the following: there are L^4 nucleotides in the set of all possible sequences of length L . The mean mass of a single RNA nucleotide is about 5×10^{-23} kg, so that, for example, the set of all $L = 55$ strands weighs about 3.8×10^{10} kg, the set of all $L = 79$ strands weighs about 1.5×10^{25} kg or about 2.6 times

the mass of our Earth, whereas the set of all $L = 126$ RNA strands would have an almost unimaginably large mass of about 5×10^{53} kg, more than the mass of the observable Universe, which is estimated to be about 10^{53} kg.

4.3. Generating distributions $P_G(\Omega)$ and $P_P(\Omega)$ by sampling

We used a standard (PYTHON) random number generator to create sets of random sequences. For each sequence, we used the Vienna package to find the lowest free energy SS. To determine Ω for each SS, we used the NSSE described in [18], which employs sampling techniques together with the inverse fold algorithm from the Vienna package. We used default settings except for the total number of measurements (set with the `-m` option), which we set to 1 instead of the default 10 for the sake of speed. We checked that this has a negligible effect (typically less than 1%) on the accuracy of our distributions. We also checked the NSSE against the full enumeration for $L = 20$, finding an agreement of $R^2 = 0.97$ for structures with Ω larger than the average; it performs slightly less well for rare structures. For longer lengths, the number of samples were: 10^5 for $L = 30$, 3×10^5 for $L = 40$, 20 000 for $L = 35-80$, 5000 for $L = 85-100$ and 1000 for $L = 126$. Sequences that generate the trivial structure are discarded. A small fraction of sequences (which increases with increasing length) were also discarded owing to the inverse folding package failing to converge (see electronic supplementary material).

To generate the $P_G(\Omega)$ distribution, we partition the support of the distribution into bins which are uniform on an $S = \log_{10}(\Omega)$ scale. We then determine the probability mass $P_G(\Omega)$ in each bin. Error bars are simply statistical: there is a trade-off between making smaller bins to give a greater resolution and minimizing statistical errors that increase when there are fewer measurements per bin. Bins with too few sampled points were typically not included in the graphs to avoid large error bars. $P_P(\Omega)$ is generated from the sampled data by dividing through by Ω (measured at the midpoint of the bin). $P_P(\Omega)$ is normalized with the N_P calculated from the analytic approximation to $P_P(\Omega)$ (see below).

4.4. Analytical fit to neutral set size distributions

For analytic fits, we make a simple log-binomial ansatz:

$$P_P(\Omega) = \binom{N}{q} (p_P)^q (1 - p_P)^{N-q}, \quad (4.1)$$

where $q = \log(\Omega)N/U$, 10^U is the largest non-trivial Ω , and N and p_P are parameters that are fitted to measured distributions. In other words, the probability that a P -sampled SS is found with $S = \log(\Omega)$ is distributed binomially. By definition $P_G(\Omega) \propto \Omega P_P(\Omega)$. Taking normalization into account is enough to show that $P_G(\Omega)$ has the same binomial form as equation (4.1), but with parameter p_P replaced by $p_G = (p_P 10^{U/N}) / (1 - p_P + p_P 10^{U/N})$. Fixing the parameters N , U and either p_G or p_P thus fixes both distributions.

For $L = 20$, equation (4.1) with $U = 10$, $N = 8.0$ and $p_P = 0.55$ describes the exact $P_P(\Omega)$ from full enumeration very well, as can be seen in figure 1b. The related approximation for $P_G(\Omega)$ performs slightly less well for G-sampled data for $L = 20$, but it still captures the main qualitative features.

For larger L , we determine the parameters as follows: first, we estimate U from the largest non-trivial NS size found. This method inevitably provides a lower bound U' on the true maximum U . However, given the rather sharp upper bound generated by the binomial fit, we expect that the relative errors in U are quite small. For example, for $L = 60$, we used 20 000 samples to determine $U' = 30.56$. From the binomial form, we estimate that U' is within an error $\delta U = 0.45$ of the true U with 90% probability. Next, we calculate \bar{S}_G , the G-sampled average of $S = \log(\Omega)$, as

well as the standard deviation $\bar{\sigma}_G$ in S from G -sampled data. We can then determine the parameter $p_G = \bar{S}_G/U$, derived by taking the mean of q through equation (4.1). The parameter N can subsequently be extracted from the measured G -sampled standard deviation: $\bar{\sigma}_G = \sqrt{N p_G(1-p_G)U/N}$. The P -sampled standard deviation is given by $\bar{\sigma}_P = \sqrt{N p_P(1-p_P)U/N}$. Because $p_P = \bar{S}_P/U < p_G$, so also $\bar{\sigma}_P < \bar{\sigma}_G$. In this way, we obtained the binomial fits shown in figure 3 and the electronic supplementary material, figures S1 and S2. The close agreement between the sampled data and our fits for lengths up to $L = 126$ suggests that this procedure is fairly robust.

It is harder to find structures with small Ω , so that the $P_P(\Omega)$ can only be partially sampled, especially for larger L . However, given how well our simple ansatz works for predicting $P_G(\Omega)$, and given that for $L = 20$ RNA the binomial form works so well for the full range of structures, we expect the full $P_P(\Omega)$ to be at least similar if not very close to equation (4.1). Further evidence for this form can also be extracted from combinatorial arguments for the distributions of stacks [30], which are correlated with $\log(\Omega)$ (see below).

4.5. Rank plots

Analytic rank-plot functions are the cumulative density function of $P_P(\Omega)$. For $L = 20$, all structures are known, so a rank plot can be directly made. For $L > 20$, the measured Ω of natural SSs was used to align points to a rank function calculated from the analytic binomial fit to $P_P(\Omega)$.

4.6. Scaling forms as a function of L

We further used the lengths $L = 30-100$ to extract scaling forms for several properties as a function of L . Linear fits in L are shown for U , T , S_G and N in the electronic supplementary material, figure S5. T is close to U , so that the relative difference between T and U decreases as L increases (electronic supplementary material, figure S5). A summary of the scaling forms for different properties in the large L limit can be found in table 1.

4.7. Probability P_T to find the trivial structure

The probability of finding the trivial structure, P_T , decreases exponentially with increasing L :

$$P_T = \frac{10^T}{4^L} \approx 21.4 \times 0.82^L. \quad (4.2)$$

For example, for $L = 20$, we can directly measure $P_T \approx 33\%$, whereas for $L = 55$ it has already dropped down to a mere 0.04%. This rapid decrease in P_T justifies our decision to ignore the trivial structure in our fitting to $P_G(S)$ and $P_P(S)$.

4.8. The number of SS, N_P , as a function of L

For $L = 11-20$, we used full enumerations [20] to calculate N_P . For longer L , we used our analytic fit as follows: the mean (including trivial structure) of Ω is $4^L/N_P$, so that $(1-P_T)4^L/N_P = \sum_{S=0}^U P_P(\Omega)\Omega$. Given the binomial form of $P_P(\Omega)$, the sum can be carried out analytically, from which it follows that $N_P = (1-P_T)4^L/(1-p_P + p_P 10^{U/N})^N$. At each L , we evaluated N_P, P_T, p_P and U , and used this to evaluate N_P . A simple linear form,

$$N_P \approx (0.13 \pm 0.04) \times (1.760 \pm 0.007)^L, \quad (4.3)$$

provides a good fit to the data (electronic supplementary material, figure S5).

4.9. Scaling forms for the Shannon entropy H and bias parameter β as a function of L

From the expression for the entropy derived above, it directly follows that $H = 2L - \log_2(10)\bar{S}_G$. Using the fit derived above for S_G , the entropy grows with L as $H \approx 0.675L - 4.92$ and the effective number of states scales as $2^H = 4^L/10^{\bar{S}_G} \approx 0.033 \times 1.60^L$ (this ignores the trivial structure, but the effect is very small for larger L). Note that one only has to find \bar{S}_G , which is easily obtainable, to determine this important quantity. By combining with equation (4.3), we find that the bias parameter scales as $\beta = 2^H/N_P \approx 0.26 \times 0.91^L$.

4.10. Sampling natural RNA from the fRNA database

To generate the distributions of functional ncRNA, we took all available sequences for each length studied (ranging from $L = 40$ to 126) from the non-coding functional RNA database (fRNAdb [21]). For $L = 20$, we took data from *Drosophila melanogaster* only, but this made up 77% of all $L = 20$ SSs in the database. For each sequence, we found the SS and used the NSSE to estimate its Ω . A small fraction of the natural RNA sequences contained non-standard nucleotide letters, e.g. N or R; such sequences were ignored, because the standard packages cannot treat them. Similarly, a small fraction of sequences were also discarded owing to the NSSE failing to converge (see electronic supplementary material). We checked that there were no repeated sequences in the database of natural RNA. Finally, in the electronic supplementary material, we provide a breakdown of the identity of the structures for $L = 20, 55, 70$ and 126 and also take curated subsets of the data to emphasize structures where the SS is known to be important. In figure 3 and the electronic supplementary material, figures S1 and S2, we show for $L = 55$ and $L = 126$ that the close correlation with $P_G(\Omega)$ remains when data are curated.

4.11. Checking for codon bias

Genetic mutations are random in the sense that they do not arise to benefit an organism. Nevertheless, it is well known that in other ways mutations are not uniformly random [58]. One example (among several) is that transitions (pyrimidines \leftrightarrow pyrimidines or purines \leftrightarrow purines) are more frequent than transversions (purines \leftrightarrow pyrimidines) [58]. For many of these biases, mutations can still effectively sample the whole space uniformly without preferring certain genotypes over others. Nevertheless, there are biases that lead, for example, to an excess of GC over AT base pairs, or vice versa in DNA [58]. To test for the effect of strong bias of this type, we generated Ω distributions with 30% GC (AU bias) and 70% GC (GC bias) content for different lengths. The overall effect becomes less pronounced for longer strands (electronic supplementary material, figure S4). Because natural DNA (and by extension RNA) can show biases for both larger and smaller GC content, we argue that this can to first order be ignored when comparing with natural datasets across many species, although the effects may be observable on datasets with large content bias.

4.12. Calculating P-distribution of stacks

We used a linear relationship between $\log(\Omega)$ and K (electronic supplementary material, figure S6) to transform $P_P(\Omega)$ into an estimated P -sampled distribution of stacks, as shown in figure 4a, and used this to obtain the P -sampled average \bar{K}_P . We also adapted analytic results from [30] based on combinatorics (see the electronic supplementary material) to calculate estimates for the P -sampled distribution of K and for \bar{K}_P . The close agreement shown in figure 4a,b, between the two independent methods for estimating the number of stacks from

P-sampling increases our confidence in our binomial fits to the full P-sampled distribution.

4.13. Calculating the robustness distribution

A strong positive correlation between mutational robustness and Ω is well established in the literature [6]. For figure 6, the robustness for G-sampled data (taken from 10^5 random sequences) and natural data (taken from the 504 natural sequences for $L = 55$) was calculated by folding all $3L$ strands within one point mutation and calculating the fraction r that generates the same SS. Error bars come from our binning procedure. For the P-sampled distribution, such a sequence robustness cannot be defined, so instead the robustness per sequence was first averaged over a set of sequences for each SS to estimate a mean robustness (phenotype robustness) per SS. We next generated a cubic fit to the mean r per SS versus

$\log \Omega$. The fit was constrained to have $r = 0$ at $\Omega = 1$ and had $R^2 = 0.93$. This was then combined with $P_P(\Omega)$ to generate an estimate of the r distribution for P-sampled data. Because we do not have many data points at small Ω , the fit is partially an extrapolation. Further error may arise from the partial sampling of the phenotype robustness. Nevertheless, we do not expect this procedure to lead to a significant difference in the qualitative comparison between P- and G-sampled data for robustness, given that the two are so different.

Competing interests. We declare we have no competing interests.

Funding. K.D. acknowledges funding from the EPSRC/EP/G03706X/1.

Acknowledgements. We acknowledge discussions with S.E. Ahnert, F.Q. Camargo, J.P.K. Doye, J.L. England, T.M.A. Fink, W. Haerty, I.G. Johnston, D.C. Lahti, A.C. Love, O. Martin, T.C.B. McLeish, T.E. Ouldridge, C.P. Ponting, F. Randisi and P. Sulc.

References

- Gould SJ. 1989 *Wonderful life: the Burgess Shale and the nature of history*. New York, NY: W. W. Norton & Co.
- Morris S. 2003 *Life's solution: inevitable humans in a lonely universe*. Cambridge, UK: Cambridge University Press.
- McGhee GR. 2006 *The geometry of evolution: adaptive landscapes and theoretical morphospaces*. Cambridge, UK: Cambridge University Press.
- Salisbury FB. 1969 Natural selection and the complexity of the gene. *Nature* **224**, 342–343. (doi:10.1038/224342a0)
- Maynard Smith J. 1970 Natural selection and the concept of a protein space. *Nature* **225**, 563–564. (doi:10.1038/225563a0)
- Wagner A. 2005 *Robustness and evolvability in living systems*. Princeton, NJ: Princeton University Press.
- Mattick JS, Makunin IV. 2006 Non-coding RNA. *Hum. Mol. Genet.* **15**(Suppl. 1), R17–R29. (doi:10.1093/hmg/ddl046)
- Carothers J, Oestreich S, Davis J, Szostak J. 2004 Informational complexity and functional activity of RNA structures. *J. Am. Chem. Soc.* **126**, 5130–5137. (doi:10.1021/ja031504a)
- Bailor MH, Mustoe III AM, Brooks CL, Al-Hashimi HM. 2011 Topological constraints: using RNA secondary structure to model 3d conformation, folding pathways, and dynamic adaptation. *Curr. Opin. Struct. Biol.* **21**, 296–305. (doi:10.1016/j.sbi.2011.03.009)
- Hofacker I, Fontana W, Stadler P, Bonhoeffer L, Tacker M, Schuster P. 1994 Fast folding and comparison of RNA secondary structures. *Chem. Mon.* **125**, 167–188. (doi:10.1007/BF00818163)
- Zuker M, Mathews D, Turner D. 1999 Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. *RNA Biochem. Biotechnol.* **70**, 11–44. (doi:10.1007/978-94-011-4485-8_2)
- Schuster P, Fontana W, Stadler P, Hofacker I. 1994 From sequences to shapes and back: a case study in RNA secondary structures. *Proc. R. Soc. Lond. B* **255**, 279–284. (doi:10.1098/rspb.1994.0040)
- Fontana W, Konings DA, Stadler PF, Schuster P. 1993 Statistics of RNA secondary structures. *Biopolymers* **33**, 1389–1404. (doi:10.1002/bip.360330909)
- Schultes EA, Spasic A, Mohanty U, Bartel DP. 2005 Compact and ordered collapse of randomly generated RNA sequences. *Nat. Struct. Mol. Biol.* **12**, 1130–1136. (doi:10.1038/nsmb1014)
- Schuster P. 2006 Prediction of RNA secondary structures: from theory to models and real molecules. *Rep. Prog. Phys.* **69**, 1419. (doi:10.1088/0034-4885/69/5/R04)
- Smit S, Yarus M, Knight R. 2006 Natural selection is not required to explain universal compositional patterns in rRNA secondary structure categories. *RNA* **12**, 1–14. (doi:10.1261/rna.2183806)
- 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**, e1000110. (doi:10.1371/journal.pcbi.1000110)
- 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**, 464. (doi:10.1186/1471-2105-9-464)
- 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**, e26324. (doi:10.1371/journal.pone.0026324)
- 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**, e86635. (doi:10.1371/journal.pone.0086635)
- Kin T *et al.* 2007 FRNadb: a platform for mining/annotating functional RNA candidates from non-coding RNA sequences. *Nucleic Acids Res.* **35**(Suppl. 1), D145–D148. (doi:10.1093/nar/gkl837)
- de Vries H. 1904 *Species and varieties, their origin by mutation*. Amsterdam, The Netherlands: The Open Court Publishing Company.
- Wagner A. 2014 *Arrival of the fittest: solving evolution's greatest puzzle*. New York, NY: Penguin.
- Mezard M, Montanari A. 2009 *Information, physics, and computation*. New York, NY: Oxford University Press.
- Salehi-Ashtiani K, Szostak J. 2001 *In vitro* evolution suggests multiple origins for the hammerhead ribozyme. *Nature* **414**, 82–83. (doi:10.1038/35102081)
- Hammann C, Luptak A, Perreault J, de la Peña M. 2012 The ubiquitous hammerhead ribozyme. *RNA* **18**, 871–885. (doi:10.1261/rna.031401.111)
- Gan HH, Pasquali S, Schlick T. 2003 Exploring the repertoire of RNA secondary motifs using graph theory; implications for RNA design. *Nucleic Acids Res.* **31**, 2926–2943. (doi:10.1093/nar/gkg365)
- Athavale SS, Spicer B, Chen IA. 2014 Experimental fitness landscapes to understand the molecular evolution of RNA-based life. *Curr. Opin. Chem. Biol.* **22**, 35–39. (doi:10.1016/j.cbpa.2014.09.008)
- Andronescu M, Bereg V, Hoos HH, Condon A. 2008 RNA STRAND: the RNA secondary structure and statistical analysis database. *BMC Bioinformatics* **9**, 340. (doi:10.1186/1471-2105-9-340)
- Hofacker IL, Schuster P, Stadler PF. 1998 Combinatorics of RNA secondary structures. *Discrete Appl. Math.* **88**, 207–237. (doi:10.1016/S0166-218X(98)00073-0)
- Greenbury SF, Ahnert SE, Louis AA. 2015 Genetic correlations greatly increase mutational robustness and can both reduce and enhance evolvability. (<http://arxiv.org/abs/1505.07821>)
- Yampolsky L, Stoltzfus A. 2001 Bias in the introduction of variation as an orienting factor in evolution. *Evol. Dev.* **3**, 73–83. (doi:10.1046/j.1525-142x.2001.003002073.x)
- Amundson R. 2005 *The changing role of the embryo in evolutionary thought: roots of evo-devo*. Cambridge, UK: Cambridge University Press.
- Laland K, Uller T, Feldman M, Sterelny K, Müller GB, Moczek A, Jablonka E, Odling-Smee J. 2014 Does evolutionary theory need a rethink? Yes, urgently. *Nature* **514**, 161–164. (doi:10.1038/514161a)
- Wray GA, Hoekstra HE, Futuyma D, Lenski R, Mackay T, Schluter D, Strassman J. 2014 Does evolutionary

- theory need a rethink? No, all is well. *Nature* **514**, 161–164.
36. Losos JB. 2011 Convergence, adaptation, and constraint. *Evolution* **65**, 1827–1840. (doi:10.1111/j.1558-5646.2011.01289.x)
 37. Scotland RW. 2011 What is parallelism? *Evol. Dev.* **13**, 214–227. (doi:10.1111/j.1525-142X.2011.00471.x)
 38. Pearce T. 2011 Convergence and parallelism in evolution: a neo-Gouldian account. *Br. J. Philos. Sci.* **63**, 429–448. (doi:10.1093/bjps/axr046)
 39. Powell R. 2012 Convergent evolution and the limits of natural selection. *Eur. J. Philos. Sci.* **2**, 355–373. (doi:10.1007/s13194-012-0047-9)
 40. Weismann A. 1893 The all-sufficiency of natural selection: a reply to Herbert Spencer. *Contemp. Rev.* **64**, 309–338.
 41. Morris SC. 2008 *The deep structure of biology: is convergence sufficiently ubiquitous to give a directional signal*. West Conshohocken, PA: Templeton Foundation Press.
 42. Mayr E. 1961 Cause and effect in biology. *Science (New York, NY)* **134**, 1501–1506. (doi:10.1126/science.134.3489.1501)
 43. 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**, 1512–1516. (doi:10.1126/science.1210879)
 44. Scholl R, Pigliucci M. 2014 The proximate–ultimate distinction and evolutionary developmental biology: causal irrelevance versus explanatory abstraction. *Biol. Philos.* **30**, 653–670. (doi:10.1007/s10539-014-9427-1)
 45. Schwenk K, Wagner GP. 2004 The relativism of constraints on phenotypic evolution. In *Phenotypic integration: studying the ecology and evolution of complex phenotypes* (eds M Pigliucci, K Preston), pp. 390–408. Oxford, UK: Oxford University Press.
 46. Gould S, Lewontin R. 1979 The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc. R. Soc. Lond. B* **205**, 581–598. (doi:10.1098/rspb.1979.0086)
 47. Gilbert SF. 2000 *Developmental biology*. Sunderland, MA: Sinauer Associates, Inc.
 48. Li H, Helling R, Tang C, Wingreen N. 1996 Emergence of preferred structures in a simple model of protein folding. *Science* **273**, 666–669. (doi:10.1126/science.273.5275.666)
 49. England J, Shakhnovich E. 2003 Structural determinant of protein designability. *Phys. Rev. Lett.* **90**, 218101. (doi:10.1103/PhysRevLett.90.218101)
 50. 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**, 20140249. (doi:10.1098/rsif.2014.0249)
 51. Nochomovitz Y, Li H. 2006 Highly designable phenotypes and mutational buffers emerge from a systematic mapping between network topology and dynamic output. *Proc. Natl Acad. Sci. USA* **103**, 4180–4185. (doi:10.1073/pnas.0507032103)
 52. Raman K, Wagner A. 2011 Evolvability and robustness in a complex signalling circuit. *Mol. Biosyst.* **7**, 1081–1092.
 53. Borenstein E, Krakauer D. 2008 An end to endless forms: epistasis, phenotype distribution bias, and nonuniform evolution. *PLoS Comput. Biol.* **4**, e1000202. (doi:10.1371/journal.pcbi.1000202)
 54. Gerber S. 2014 Not all roads can be taken: development induces anisotropic accessibility in morphospace. *Evol. Dev.* **16**, 373–381. (doi:10.1111/ede.12098)
 55. 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. USA* **101**, 7287–7292. (doi:10.1073/pnas.0401799101)
 56. Reuter JS, Mathews DH. 2010 Rnastructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **11**, 129. (doi:10.1186/1471-2105-11-129)
 57. Gardner PP, Giegerich R. 2004 A comprehensive comparison of comparative RNA structure prediction approaches. *BMC Bioinformatics* **5**, 140. (doi:10.1186/1471-2105-5-140)
 58. Li W, Graur D. 1999 *Fundamentals of molecular evolution*. Sunderland, MA: Sinauer.