Genetic robustness of let-7 miRNA sequence–structure pairs

QIJUN HE,¹ FENIX W. HUANG,¹ CHRISTOPHER BARRETT,^{1,2} and CHRISTIAN M. REIDYS^{1,3}

¹Biocomplexity Institute and Initiative, ²Department of Computer Science, ³Department of Mathematics, University of Virginia, Charlottesville, Virginia 22904, USA

ABSTRACT

Genetic robustness, the preservation of evolved phenotypes against genotypic mutations, is one of the central concepts in evolution. In recent years a large body of work has focused on the origins, mechanisms, and consequences of robustness in a wide range of biological systems. In particular, research on ncRNAs studied the ability of sequences to maintain folded structures against single-point mutations. In these studies, the structure is merely a reference. However, recent work revealed evidence that structure itself contributes to the genetic robustness of ncRNAs. We follow this line of thought and consider sequence–structure pairs as the unit of evolution and introduce the spectrum of extended mutational robustness (EMR spectrum) as a measurement of genetic robustness. Our analysis of the miRNA let-7 family captures key features of structure-modulated evolution and facilitates the study of robustness against multiple-point mutations.

Keywords: genetic robustness; let-7 miRNA; sequence–structure pairs; multiple-point mutation

INTRODUCTION

Genetic robustness can be characterized in terms of the variation of phenotype distribution induced by genotypic change (Schlichting and Pigliucci 1998; de Visser et al. 2003; Gu et al. 2003) and concerns the insensitivity of a phenotype to genetic changes. Mutational robustness has been studied in the context of noncoding RNA (ncRNA) (Borenstein and Ruppín 2006; Rodrigo and Fares 2012). RNA consists of a single strand of nucleotides (A, C, G, U) that can fold and bond to itself through base-pairing. ncRNAs are known to function in aptamer binding as riboswitches, in chemical catalysis as ribozymes, and in RNA splicing such as spliceosome (Breaker and Joyce 1994; Breaker 1996; Serganov and Patel 2007; Darnell 2011). Most importantly, it is the folded structure that is underlying all these mechanisms, allowing for the interaction with and subsequent modification of other biological molecules.

The self-folding of RNA makes it an ideal object to study genotype–phenotype relations. The well-established energy-based prediction of RNA secondary structure, a two-dimensional coarse grain of the real three-dimensional structure, makes such studies feasible (Waterman 1978;

Zuker and Stiegler 1981; Hofacker et al. 1994; Mathews et al. 1999).

Structures are an important determinant of the function of ncRNAs, whence the robustness of an RNA can be characterized in terms of the variation of secondary structure distribution (i.e., the stability of a secondary structure in the face of genetic sequence changes). Structural robustness of ncRNAs is considered to be a key component of the fitness of the molecule, and much research has been conducted to identify the footprints of natural selection on secondary structures of sncRNAs (Borenstein and Ruppín 2006; Rodrigo and Fares 2012). Borenstein and Ruppín (2006) define neutrality of an RNA sequence $\sigma = a_1 a_2 \dots a_n$ by $\eta(\sigma) = 1 - (d)/n$, where (d) denotes the average, taken over all $3n$ single-point mutants of σ , of the base pair distance d between the minimum free energy (MFE) structure, S_0 , of σ and the MFE structures of single-point mutants. The RNA sequence, σ , is then defined to be robust if $\eta(\sigma)$ is greater than the average neutrality of 1000 control sequences generated by the program RNAinverse (Lorenz et al. 2011), which fold into the same target structure, S_0 . The main finding of Borenstein and

Corresponding author: cmr3hk@virginia.edu

Article is online at <http://www.majournal.org/cgi/doi/10.1261/rna.065763.118>.

© 2019 He et al. This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://rnajournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Ruppin (2006) is that precursor miRNAs (pre-miRNA) exhibit a significantly higher level of mutational robustness than random RNA sequences, having the same structure. Subsequently, Rodrigo and Fares (2012) undertook a similar analysis for bacterial small RNAs. Their main finding was that, surprisingly, bacterial sncRNAs are not significantly more robust when compared with 1000 sequences having the same structure, as computed by RNAinverse. Rodrigo and Fares (2012) based their findings on the notion of ensemble *diversity* defined earlier in Gruber et al. (2008).

In the above-mentioned papers, robustness is defined by taking the average of all $3n$ single-point mutants, the underlying assumption being that all $3n$ single-point mutations are equally likely to occur and, in addition, taking exclusively single-point mutations into account. Incorporation of multiple-point mutations poses obvious difficulties, because the number of sequences that need to be taken into consideration will grow exponentially.

The neutral theory of Kimura (1983) stipulates that evolution is achieved by neutral mutations (i.e., by mutations that are necessarily compatible with the underlying structure). This is in accordance with recent findings showing that secondary structures have a genuine influence on selection in RNA genes. In Mimouni et al. (2008), the authors classify stem positions into structural classes and validate that they are under different selective constraints. In Price et al. (2011), the authors observe neutral evolution in *Drosophila* miRNA and evolution increasing the thermal dynamic stability of the RNA. Pollard et al. (2006b) and Beniaminov et al. (2008) study Human Accelerated Regions (HARs) in brains of primates—that is, noncoding RNAs with an accelerated rate of nucleotide substitutions along the lineage between human and chimpanzee. Beniaminov et al. (2008) concludes that this increased rate of nucleotide substitutions is of central importance for the evolution of the human brain. The most divergent of these regions, HAR1, has been biochemically confirmed to fold into distinct RNA secondary structures in human and chimpanzee. Interestingly, the mutations in the human HAR1 sequence, compared to the chimpanzee sequence, stabilize their respective RNA structure (Pollard et al. 2006a), suggesting a shape modulated evolution.

Recently, Barrett et al. (2017) proposed a framework considering RNA sequences and their RNA secondary structures simultaneously. This gives rise to an information-theoretic framework for RNA sequence–structure pairs. In particular, the authors studied the “dual Boltzmann distribution”—that is, the Boltzmann distribution of sequences with respect to a fixed structure. The authors develop a Boltzmann sampler of sequences with respect to a given structure and study the “inverse folding rate” (IFR) of the sampled sequences—that is, the proportion of the sampled sequences whose MFE structure equals to the given structure. Barrett et al. (2017) reports that nat-

ural structures have higher IFR than random structures, suggesting that natural structures have higher intrinsic robustness than random structures.

This suggests an approach to consider sequences and structures—as pairs—as the unit of evolution, instead of just sequences in isolation. In this article, we generalize the notion of inverse folding rate (IFR) and introduce a novel profile, the *extended* mutational robustness spectrum (EMR spectrum), of a sequence–structure pair. By construction, this spectrum entails both sequence and structure information. The key idea is that mutations will be biased by the underlying thermodynamic energy of the respective structure, instead of being equally likely to occur.

In this article, we study the *let-7* family miRNA, first discovered in *Caenorhabditis elegans* and functionally conserved along the evolutionary spectrum, from worm to human (Tanzer and Stadler 2004). First, *let-7* miRNAs are a class of RNAs of particular interest: They are small endogenous, noncoding RNAs that regulate gene expression at the translation level, playing important roles in regulating cell proliferation and differentiation during development in different species. *let-7* is widely viewed as a tumor suppressor miRNA. In particular, deregulation of *let-7* family miRNAs has been observed in a variety of cancer types (Takamizawa et al. 2004; Dahiya et al. 2008; O'Hara et al. 2009). The short, mature miRNAs (~22 nt), that directly regulate target gene expression, originate from longer RNA precursor molecules that fold into a stem–loop hairpin structure. The secondary structure of *let-7* precursor miRNAs plays a crucial role in the gene maturation process (Lee et al. 1993; Belter et al. 2014). The stem–loop structure has been under evolutionary pressures to be conserved. Second, the RNA secondary structure of *let-7* miRNAs can be derived by secondary structure prediction algorithms as in Lorenz et al. (2011), rendering it to be particularly suited for computational studies. Folding algorithms using the nearest neighbor thermodynamic model (NNTM) have high accuracy for short sequences like *let-7* precursor miRNAs (typically <150 nt) (Doshi et al. 2004; Mathews et al. 2004). As a result, the predicted structure is of biological relevance and can be considered as a meaningful representation of the phenotype. Furthermore, prediction algorithms allow us to analyze the structural, mutational robustness by comparing the predicted structure of original sequence with the predicted structure of the mutants. These features make the *let-7* gene family a particularly suitable test bed for studying the evolution of genetic robustness.

We organize our study as follows: First, we shall extend the analysis of Borenstein and Ruppin (2006) to EMR spectra and observe that most native sequence–structure pairs have a higher EMR spectrum than sequence–structure pairs obtained by the inverse folding algorithm. We shall investigate different aspects of the robustness of EMR spectra of native sequence–structure pairs and show that

these are distinctively more robust against multiple-point mutations. Second, we conduct cross-species comparisons of native sequence–structure pairs, observing that higher metazoan species have higher EMR spectra. Our analysis suggests that EMR spectra are being increased in the course of shape-modulated evolution.

RESULTS

We study a novel profile—that is, the EMR spectrum ($EMR_{\sigma,k}^S$ -spectrum; see Materials and Methods) of sequence–structure pairs (σ, S) . The $EMR_{\sigma,k}^S$ -spectrum is the family of mutational robustness indexed by the Hamming distance, k . We shall analyze the $EMR_{\sigma,k}^S$ -spectrum, for $1 \leq k \leq 20$. $EMR_{\sigma,k}^S$ is the fraction of k -point mutants of σ that fold back to S and thus characterizes the mutational robustness of sequence–structure pair (σ, S) . $EMR_{\sigma,k}^S$ is computed by Boltzmann sampling. We study mutational robustness of native let-7 sequence–structure pairs in two ways: first by comparison with pairs obtained by inverse folding and second by cross-species comparisons of the entire let-7 gene family.

Robustness of native let-7 sequence–structure pairs

Borenstein and Rupp (2006) report that miRNA sequences exhibit a high level of neutrality in comparison with random sequences folding into a similar structure, suggesting that native miRNA sequences are more robust against single-point mutations. These results motivate a further analysis of native and random sequences, in particular whether or not this neutrality is confined to a local neighborhood. To this end we uniformly select 50 of the 401 native sequence–structure pairs of the let-7 miRNA family (see Supplemental Figs. 2 and 3). For each such native pair $(\sigma,$

FIGURE 1. The $EMR_{\sigma,k}^S$ spectrum, where the x-axis denotes the Hamming distance and the y-axis the inverse folding rate. The $EMR_{\sigma,k}^S$ spectrum of the api-let-7 (*Acyrtosiphon pisum*) gene (yellow) and the spectrum of means of the corresponding control set, $\mu(EMR_{\sigma^*,k}^S)$ (blue). The error bar denotes the standard deviation of the control set of $\{EMR_{\sigma^*,k}^S\}$ for each k . Because $EMR_{\sigma,k}^S > \mu(EMR_{\sigma^*,k}^S)$, api-let-7 is k -robust, for all $1 \leq k \leq 20$. $EMR_{\sigma,k}^S$ decreases slower than $\mu(EMR_{\sigma^*,k}^S)$, as k increases.

TABLE 1 k -robustness of native sequence–structure pairs

Hamming distance	k -robust	Significantly k -robust
$k = 1$	96%	8%
$k = 2$	96%	28%
$k = 3$	94%	38%
$k = 4$	92%	38%
$k = 5$	94%	38%
$k = 10$	92%	48%
$k = 15$	94%	56%
$k = 20$	96%	58%

The second column shows the percentage of k -robust ($EMR_{\sigma,k}^S > \mu(EMR_{\sigma^*,k}^S)$) genes. The third column shows the percentage of the significantly k -robust genes for $P < 0.05$. The P -values denote the probability of observing $EMR_{\sigma,k}^S > EMR_{\sigma^*,k}^S$ by chance and are calculated via Z-tests.

S), we derive 100 random sequences, σ^* , whose MFE structure is identical to S , as a control set. These control sequences are computed using the inverse folding algorithm presented in (Levin et al. (2012), Garcia-Martin et al. (2016), and Barrett et al. (2017)). Given structure S , we compute the partition function of all sequences with respect to S . We then Boltzmann-sample sequences and filter (by rejection) those sequences, which fold into S . We then compute the $EMR_{\sigma^*,k}^S$ -spectra for the inverse folding solutions and compare, for each native sequence–structure pair (σ, S) , the $EMR_{\sigma,k}^S$ -spectrum with the $\mu(EMR_{\sigma^*,k}^S)$ -spectrum—that is, the mean of all 100 spectra, $EMR_{\sigma^*,k}^S$, taken at each respective k . We call (σ, S) k -robust, if $EMR_{\sigma,k}^S > \mu(EMR_{\sigma^*,k}^S)$. We then check the significance of k -robustness via Z-tests, assuming $\{EMR_{\sigma^*,k}^S\}$ is normally distributed.⁴ We call (σ, S) significantly k -robust, if $P < 0.05$. In Figure 1 we depict the $EMR_{\sigma,k}^S$ -spectrum of api-let-7 (*Acyrtosiphon pisum*) sequence–structure pair to illustrate k -robustness. In this example, api-let-7 is k -robust for $1 \leq k \leq 20$. Furthermore, api-let-7 is significantly k -robust for $5 \leq k \leq 20$.

In Table 1 we summarize the data on k -robustness of the 50 let-7 sequence–structure pairs included in our study. We observe that 96% of the selected let-7 miRNAs are 1-robust. This observation is consistent with Borenstein and Rupp (2006), providing further evidence that native let-7 sequence–structure pairs are robust against single-point mutations. However, the mutational robustness of native let-7 sequence–structure pairs is not restricted to single-point mutations: For all $1 \leq k \leq 20$, we observe that k -robustness holds for >90% of the selected sequence–structure pairs.

This suggests that the mutational robustness of let-7 sequence–structure pairs is not a local phenomenon. Furthermore, the percentage of significantly k -robust genes

⁴At $k = 5$, 31 out of the 50 control groups $\{EMR_{\sigma^*,k}^S\}$ have $P > 0.1$ via Shapiro–Wilk tests.

FIGURE 2. The distribution of $EMR_{\sigma,k}^S$ ranks, r_k , of the native sequence–structure pairs $EMR_{\sigma,k}^S$. The x-axis denotes ranking and the y-axis frequency, $k = 1$ (left), $k = 5$ (center), and $k = 20$ (right). For each native sequence–structure pair, (σ, S) , the $EMR_{\sigma,k}^S$ is ranked among 100 $EMR_{\sigma^*,k}^S$ values, where σ^* is a random sequence whose MFE structure is S .

increases as k increases (see Table 1). In Figure 1, we display the $EMR_{\sigma,k}^S$ spectrum of the *api-let-7* sequence–structure pair, illustrating the aforementioned phenomenon. The *api-let-7* sequence–structure pair is 1-robust but not significantly 1-robust. However, for $5 \leq k \leq 20$, the *api-let-7* sequence–structure pair is significantly k -robust.

We proceed by considering for each native sequence–structure pair, r_k , its $EMR_{\sigma,k}^S$ rank among the $EMR_{\sigma^*,k}^S$ values, respectively. Figure 2 present the distribution r_k , for $k = 1, 5, 20$. Examination of the rank distribution shows that native pairs exhibit a propensity toward high ranks for each k , supporting the observation of k -robustness. As k increases, the propensity toward high ranks becomes more and more pronounced. In the case of $k = 1$, only four out of 50 native sequence–structure pairs have $r_1 = 1$. For $k = 20$, however, this quantity increases to 23. This is consistent with the increasing percentage of significantly k -robust genes as k increases, further demonstrating that native *let-7* sequence–structure pairs exhibit mutational robustness against multiple-point mutations.

By comparing specific native sequence–structure pairs with the respective control pairs (obtained by inverse folding), we observe that native sequence–structure pairs exhibit in general higher $EMR_{\sigma,k}^S$ values. In this analysis, the sequence–structure pair (σ, S) is fixed and we contrast native pairs with pairs obtained via random inverse folded sequences.

We can augment this analysis by considering the ensemble of spectra of native pairs versus the ensemble of all inverse folded sequence–structure pairs. In that, for any fixed k , we can integrate the information of all native pairs contrasting this with the integrated information of the inverse folded pairs. Figure 3 displays these two distributions: $EMR_{\sigma,k}^S$ of native pairs and $\mu(EMR_{\sigma^*,k}^S)$ of the inverse folded pairs for $k = 1, 5, 20$. For each k , we not only observe that the mean of the $EMR_{\sigma,k}^S$ is greater than that of the terms $\mu(EMR_{\sigma^*,k}^S)$, but also that the distributions of $EMR_{\sigma,k}^S$ and $\mu(EMR_{\sigma^*,k}^S)$ are distinctively different. Furthermore, the difference between the two distributions for each k is statistically significant (see Table 2, where the P -value is calculated by two-tailed Wilcoxon signed-rank tests for paired data).

The increase of significantly k -robust native sequence–structure pairs for increasing k , as well as the tendency of native $EMR_{\sigma,k}^S$ to assume high ranks, for increasing k , suggest that k -robustness of native sequence–structure pairs, $2 \leq k \leq 20$, is not a byproduct of their 1-robustness. In other words, the high $EMR_{\sigma,k}^S$ of native sequence–structure pairs, $2 \leq k \leq 20$, is not solely obtained by optimizing their $EMR_{\sigma,1}^S$.

To further demonstrate this point, for each native sequence–structure pair (σ, S) , we collect the random sequence σ' exhibiting the highest $EMR_{\sigma',1}^S$ among the $\{EMR_{\sigma^*,1}^S\}$. A comparative analysis of the $EMR_{\sigma',k}^S$ values with the corresponding values of the native $EMR_{\sigma,k}^S$ shows

FIGURE 3. The distribution of $EMR_{\sigma,k}^S$ of native pairs (blue) and $\mu(EMR_{\sigma^*,k}^S)$ of the control sets (yellow). The x-axis denotes the inverse folding rate and the y-axis denotes the frequency, $k = 1$ (left), $k = 5$ (center), and $k = 20$ (right).

TABLE 2 Distinct distribution of $EMR_{\sigma,k}^S$ and $\mu(EMR_{\sigma,k}^S)$

Hamming distance	Mean $EMR_{\sigma,k}^S$	Mean $\mu(EMR_{\sigma,k}^S)$	P-value
$k = 1$	0.9118	0.7904	1.09×10^{-9}
$k = 5$	0.5952	0.4220	3.57×10^{-9}
$k = 20$	0.2483	0.1387	2.23×10^{-9}

The second and third columns display the mean value of $EMR_{\sigma,k}^S$ and $\mu(EMR_{\sigma,k}^S)$, respectively, for each k . P-values denote the probability of observing more extreme differences between $\{EMR_{\sigma,k}^S\}$ and $\{\mu(EMR_{\sigma,k}^S)\}$ at random, assuming two samples are drawn from the same distribution. The P-value is calculated by a two-tailed Wilcoxon signed-rank test for paired data.

that for $k = 1$, 46 out of the 50 σ' sequences exhibit an $EMR_{\sigma',1}^S$ value exceeding that of the corresponding $EMR_{\sigma,1}^S$ value. For $k = 20$, in contrast, that number⁵ drops to 16.

In analogy to our previous comparative analysis of $EMR_{\sigma,k}^S$ versus $EMR_{\sigma',k}^S$ rankings, we proceed by computing for each respective σ' the $EMR_{\sigma',20}^S$ rank among $EMR_{\sigma,k}^S$ values. Integrating over all σ' sequences, Figure 4 presents the distribution of the $EMR_{\sigma',20}^S$ ranks, as well as the previously derived (see Fig. 2) distribution of the native $EMR_{\sigma,20}^S$ ranks. We observe that the distributions of the $EMR_{\sigma',20}^S$ ranks and $EMR_{\sigma,20}^S$ ranks are distinctively different. $EMR_{\sigma',20}^S$ exhibits a propensity toward higher ranks compared to $EMR_{\sigma,20}^S$. These findings suggest that there exist factors beyond 1-robustness that contribute to the k -robustness of native sequence–structure pairs.

We finally study the role of the free energy for k -robustness. Bonnet et al. (2004) reports that miRNA exhibits increased thermodynamic stability and furthermore that mutations in the HAR further stabilize the structure (Pollard et al. 2006a). This gives rise to the question whether increased k -robustness is a result of the increased thermodynamic stability of miRNAs.

To test this, we select a native sequence–structure pair from the miRNA let-7 family and generate two sets of sequences— Σ_{high} and Σ_{low} —each consisting of 30 sequences that fold into S , having higher and lower energy than the native pair, respectively. It turns out, that sequences generated by RNAinverse (Lorenz et al. 2011) tend to have higher energy than the native pair, whereas sequences generated by the dual Boltzmann sampler, filtered by rejection to fold into S , tend to have lower energy. We compute the mean and the standard deviation of the spectra of the sequences of Σ_{low} and Σ_{high} , respectively, and integrate our findings in Figure 5.

Figure 5 shows that the $EMR_{\sigma,k}^S$ -spectrum of the native pair is distinctively higher than the ones derived from Σ_{low} and Σ_{high} . We then perform a correlation analysis be-

tween the $EMR_{\sigma,k}^S$ -spectrum and energy. Let $e(\sigma, S)$ denote the free energy of (σ, S) . For the 61 pairs mentioned above, the Spearman's rank correlation coefficient between $EMR_{\sigma,1}^S$ and $e(\sigma, S)$ is -0.16 . For $EMR_{\sigma,20}^S$ and $e(\sigma, S)$, the correlation coefficient is -0.10 . Both quantities exhibit practically no correlation. We have confirmed this result for 10 additional native sequence–structure pairs. The findings suggest that thermodynamic stability is not the sole factor—native pairs have been selected for. The mutational robustness of native pairs is not a byproduct of evolving toward thermodynamic stability.

Robustness of metazoan species

Freilich et al. (2010) studies genetic robustness of networks of bacterial genes, observing variations among species in their level of genetic robustness, reflecting adaptations to different ecological niches and lifestyles. The genetic robustness of a network refers to its ability to buffer mutations via the existence of alternative pathways. The species-specific variations raise the question whether such variations can also be observed for the $EMR_{\sigma,k}^S$ -spectra across different animal species. Are $EMR_{\sigma,k}^S$ -spectra to some extent a reflection of phylogenetic relationships?

Herein, we perform an evolutionary analysis of the $EMR_{\sigma,k}^S$ -spectra of let-7 miRNA sequence–structure pairs across different animal species. We first perform our analysis on six selected taxa: *Homo sapiens*, *Pan paniscus*, *Anolis carolinensis*, *Ciona*, *Drosophila*, and *Chromadorea*. There are 12, 12, 11, 11, 14, and seven let-7 genes found in the miRBase within each taxon, respectively. For almost all $1 \leq k \leq 20$, we observe the mean $EMR_{\sigma,k}^S$ value within each taxa to decrease from *H. sapiens* to *Chromadorea* (see Figure 6 and Table 3). The only exceptions are the $EMR_{\sigma,k}^S$ values of *P. paniscus*, which become slightly larger than those of *H. sapiens*, for $14 \leq k \leq 20$. We observe that “higher” metazoan species, such as *H. sapiens*, *P. paniscus*, *A. carolinensis*, and *Ciona*, all of which, being Chordata, exhibit larger $EMR_{\sigma,k}^S$ values than “lower” metazoan species, such as *Drosophila* and *Chromadorea*, all of which being Ecdysozoa. In addition, the difference of $EMR_{\sigma,k}^S$ values among evolutionary closely related species, such as *H. sapiens*, *P. paniscus*, and *A. carolinensis*, are small. However, comparing taxa less related in the phylogenetic tree of life, the difference becomes significant. Statistical tests (two-tailed Wilcoxon rank-sum test) are conducted to quantify the statistical significance of this difference within the let-7 genes $EMR_{\sigma,k}^S$ value distribution among these six taxa (see Figure 7), for $k = 1, 5, 20$.

The results of statistical tests at $k = 1, 5, 20$ demonstrate that $EMR_{\sigma,k}^S$ distributions of the six taxa reflect the complexity of the organisms and their phylogenetic relations. *H. sapiens*, *P. paniscus*, and *A. carolinensis* almost always exhibit significantly higher $EMR_{\sigma,k}^S$ values than *Drosophila* and *Chromadorea*, for $k = 1, 5, 20$. On the other hand,

⁵For these “locally selected” σ' -sequences.

FIGURE 4. The distribution of $EMR_{\sigma,20}^S$ ranks (yellow) and $EMR_{\sigma',20}^S$ ranks (blue). The x-axis denotes ranking and the y-axis frequency. For each native sequence–structure pair (σ, S) , σ is the corresponding native sequence, and σ' is the random sequence with the highest $EMR_{\sigma,1}^S$. The $EMR_{\sigma,20}^S$ and the $EMR_{\sigma',20}^S$ are ranked among 100 $EMR_{\sigma^*,20}^S$ values, where σ^* is a random sequence whose MFE structure is S .

differences among *H. sapiens*, *P. paniscus*, and *A. carolinensis* are statistically insignificant for $k=1, 5, 20$. However, we observe, for $k=20$ insignificant differences between *H. sapiens*, *P. paniscus*, and *A. carolinensis* with *Drosophila*, though the P -values are very close to 0.05. A more distinguished variation is observed comparing *Ciona* with *Drosophila*. In the case of $k=1$, the difference is close to being significant with $p=0.0554$, but for $k=20$, the difference is insignificant with $p=0.5417$.

We proceed by conducting the above analysis for all 401 *let-7* sequence–structure pairs of the miR-Base. *let-7* genes have been found in 82 metazoan species. Multiple homologous miRNA genes are commonly observed in vertebrates, whereas single miRNA *let-7* genes were more common in other animal species. The phylogenetic tree of all 82 species is constructed using the “Interactive Tree Of Life” (iTOL) (Letunic and Bork 2006, 2016), based on NCBI taxonomy (Federhen 2011). In addition, iTOL determines taxonomic classes of all internal nodes. Given the phylogenetic tree, we not only compute the mean $EMR_{\sigma,k}^S$ values of the *let-7* sequence–structure pairs within each specific species, but also the mean $EMR_{\sigma,k}^S$ values within taxa, corresponding to the internal nodes of the

phylogenetic tree. Figure 8 depicts a subtree of the phylogenetic tree that includes the major taxa as well as the mean $EMR_{\sigma,k}^S$ values within each taxon, for $k=1, 5, 20$. The complete phylogenetic tree is presented in Supplemental Figures 4–6.

For *let-7* miRNAs, we observe that higher metazoan species typically exhibit higher $EMR_{\sigma,k}^S$ values as well as significant differences in the $EMR_{\sigma,k}^S$ distribution across taxa. For instance, $EMR_{\sigma,k}^S$ distributions of Deuterostomia and Protostomia are found to be significantly different ($P=0.00016, 0.00058, 0.00033$, for $k=1, 5, 20$, respectively, calculated by two-tailed Wilcoxon rank-sum tests).

DISCUSSION

In this article, we augment the analysis of genetic sequences by incorporating structural information. The information represented by structures is distinctively different from that represented by sequences. Structures encode relations between pairs of *loci*. Such a relation can be realized in multiple ways—that is, by the bond between *loci* i and j by AU, UA, GU, UG, CG, and GC. In that, one can expect a meaningful enhancement of the sequence information.

FIGURE 5. Spectra of a native sequence–structure pair versus those of inverse folding solutions having lower and higher energies. We display the spectrum of the native, *let-7* miRNA pair (yellow), the spectrum of the mean $EMR_{\sigma,k}^S$ of inverse fold solutions having higher (red) and lower (blue) energy, respectively, and their standard deviations. The x-axis denotes the Hamming distance and the y-axis the inverse folding rate.

FIGURE 6. Mean $EMR_{\sigma,k}^S$ within each of the six taxa, from left to right: *Homo sapiens*, *Pan paniscus*, *Anolis carolinensis*, *Ciona*, *Drosophila*, and *Chromadorea*. The x-axis denotes the Hamming distance and the y-axis the inverse folding rate. The relative ordering of $EMR_{\sigma,k}^S$ values among taxa is consistent for each k (from high to low: *H. sapiens*, *P. paniscus*, *A. carolinensis*, *Ciona*, *Drosophila*, and *Chromadorea*, the only exception being $EMR_{\sigma,k}^S$ values of *P. paniscus* becoming slightly larger than those of *H. sapiens*, for $14 \leq k \leq 20$).

Incorporating structural information (i.e., considering sequences and structures combined) provides new ways to analyze genetic material. We investigate mutational robustness of let-7 miRNA, making use of this perspective, and introduce the $EMR_{\sigma,k}^S$ -spectrum, a novel observable, that quantifies mutational robustness. The $EMR_{\sigma,k}^S$ -spectrum depends on both the reference sequence and structure and allows us to delocalize the study of mutational robustness beyond single-point mutations.

Our study provides evidence for direct evolution of increased robustness in let-7 miRNAs, by comparing native let-7 miRNA sequence–structure pairs with the control pairs obtained by inverse folding algorithms. It provides evidence that mutational robustness is not local (i.e., it cannot be deduced from or restricted to single-point mutations). On the contrary, robustness effects become more pronounced in higher Hamming distances: The percentage of significantly k -robust, native let-7 sequence–structure pairs increases as the Hamming distance k increases. By conducting a comparative analysis between native sequences and random sequences with high robustness against single-point mutations, we provide evidence that there exist additional factors that contribute to mutational robustness against multiple-point mutations. The spectrum itself contains information that cannot be inferred from a local analysis.

The pronounced mutational robustness of native let-7 sequence–structure pairs against multiple-point mutations might play a role in genetic robustness on a population level. The presence of native sequences in a population al-

lows for significant sequence variation within a species while still preserving the phenotype. This would suggest that evolution is not identifying sequences that are locally robust but robust within entire regions of sequence space.

An evolutionary analysis of the $EMR_{\sigma,k}^S$ -spectrum of let-7 miRNAs shows that the $EMR_{\sigma,k}^S$ spectrum resembles phylogenetic relationships. Statistical tests exhibit that closely related species tend to have similar $EMR_{\sigma,k}^S$ -spectra, whereas distant species tend to exhibit distinct $EMR_{\sigma,k}^S$ -spectra. In general, higher-level animal species have higher $EMR_{\sigma,k}^S$ values than lower-level animal species. The increased mutational robustness in higher-level animals appears to reflect the complexity of the organisms. This phenomenon is likely related to the particular role of let-7. let-7 plays an important role in the differentiation of multiple cell types

across multiple organisms (Boyerinas et al. 2010). Higher-level animal species are subject to higher selection pressures because of their increased diversity of specialized cell types (McShea 1996).

The let-7 family showcases the usability of the $EMR_{\sigma,k}^S$ -spectrum. A natural question is whether other noncoding RNAs exhibit mutational robustness, in particular, against multiple-point mutations. In fact, our method can be applied to any noncoding RNA with one conserved, stable secondary structure. We conduct a preliminary study on a HIV-1 dimerization initiation site (DIS). The DIS, which is an essential part of the dimer linkage structure, adopts a stem–loop structure (Paillart et al. 1996). We study the mutational robustness of five randomly selected HIV-1 DIS sequences in Rfam (Griffiths-Jones et al. 2005) (see Supplemental Fig. 7 for the list of sequences) by

TABLE 3. Mean $EMR_{\sigma,k}^S$ within each taxa at each k

	$k=1$	$k=5$	$k=20$
<i>Homo sapiens</i>	0.9288	0.6632	0.3356
<i>Pan paniscus</i>	0.9234	0.6404	0.3402
<i>Anolis carolinensis</i>	0.9211	0.6412	0.2949
<i>Ciona</i>	0.8874	0.5509	0.1971
<i>Drosophila</i>	0.8524	0.4889	0.1802
<i>Chromadorea</i>	0.7288	0.2943	0.0642

The table displays the mean value of $EMR_{\sigma,k}^S$ of the let-7 sequence–structure pairs within each taxa, at $k=1, 5, 20$.

FIGURE 7. Statistical significance (P -value) of the difference within the *let-7* genes $EMR_{\sigma,k}^S$ distribution between taxa, for $k=1$ (left), $k=5$ (middle), and $k=20$ (right). P -values denote the probability of observing a larger difference in the $EMR_{\sigma,k}^S$ distribution between two corresponding taxa at random, assuming they are drawn from the same probability distribution, computed by a two-tailed Wilcoxon rank-sum test. Differences between two taxa that are statistically significant ($P < 0.05$) are highlighted, and phylogenetic relations among the six taxa are displayed.

comparing the $EMR_{\sigma,k}^S$ -spectrum of the native sequence–structure pair with that of the inverse folded sequence–structure pairs. We observe that four out of five native pairs are k -robust, but not significantly k -robust, for all $1 \leq k \leq 20$. The only exception is AJ288982: It is not k -robust for $1 \leq k \leq 4$, but then becomes k -robust for $5 \leq k \leq 20$ (see Fig. 9). Interestingly, for all five native pairs, even though they are not significantly k -robust, the Z-score of $EMR_{\sigma,k}^S$ for each native pair increases as k goes from 1 to 10. The increase in Z-score is consistent with the pronounced mutational robustness of native *let-7* sequence–structure pairs against multiple-point mutations. Such a phenomenon might be a result of “viral quasispecies effects” (Wilke et al. 2001). The mutational robustness against multiple-point mutations facilitates the existence of a broad and highly connected region in the neutral network around the native sequence (Reidys et al. 1997). Such a region might allow a quasispecies to evolve more efficiently in the presence of selection pressure.

Our framework is applicable to RNA with one stable secondary structure. In reality, there exist many functional noncoding RNAs with multiple stable secondary structures, such as riboswitches. To extend our framework to study the mutational robustness of these RNAs, one potential challenge is to generate k -mutants that are compatible with multiple secondary structures. A recent approach for positive RNA design, “RNARedPrint” (Hammer et al. 2019), might facilitate such an extension. RNARedPrint supports generic Boltzmann-weighted sampling, which enables the positive design of RNA sequences with specific free energies (for each of the multiple target structures) and GC content. The dynamic programming routine in RNARedPrint can be modified to incorporate Hamming distance filtration. Such a modification enables us to efficiently Boltzmann-sample k -mutants that are compatible with multiple secondary structures.

The role of structure in the context of the $EMR_{\sigma,k}^S$ -spectrum differs substantially from the role of neutrality in

FIGURE 8. Phylogenetic tree of the major taxa and mean $EMR_{\sigma,k}^S$ of the *let-7* sequence–structure pairs within each taxon, for $k=1$ (right), $k=5$ (center), and $k=20$ (left). Mean $EMR_{\sigma,k}^S$ value for $k=1$ (right), $k=5$ (center), and $k=20$ (left).

FIGURE 9. The $EMR_{\sigma,k}^S$ -spectrum. The x-axis denotes the Hamming distance and the y-axis the inverse folding rate. The $EMR_{\sigma,k}^S$ -spectrum of the AJ288982 HIV1-DIS gene is shown in yellow, and the spectrum of means of the corresponding control set, $\mu(EMR_{\sigma,k}^S)$ is shown in blue. The error bar denotes the standard deviation of the control set of $\{EMR_{\sigma,k}^S\}$ for each k . AJ288982 is not k -robust for $1 \leq k \leq 4$, but then becomes k -robust for $5 \leq k \leq 20$.

Borenstein and Rupp (2006): Structure is not merely used to measure the effect of the mutation but also affects how sequences mutate. This constitutes effectively a feedback loop between sequence and structure and is arguably the driving factor enhancing the signal in native sequence–structure pairs.

We use a discrete metric for measuring the structural change induced by mutation in the definition of the $EMR_{\sigma,k}^S$ -spectrum, instead of the base pair distance. This metric produces stable data: The standard error of $EMR_{\sigma,k}^S$ by sampling 5000 times is small, ± 0.0065 , compared to the difference of the $EMR_{\sigma,k}^S$ -spectra between native and control sequence–structure pairs and the difference in the $EMR_{\sigma,k}^S$ -spectra across species, respectively. This allows us to analyze efficiently mutational robustness against multiple-point mutations by Boltzmann sampling, obsoleting the need for large sample sizes.

In our analysis, we use the MFE structures, predicted by the RNA minimum free-energy folding algorithm, as phenotypes. The study can be enhanced by passing from MFE structures to partition functions of structures with respect to a fixed sequence (McCaskill 1990). One can envision an analysis that considers both the dual partition function, considered here, and the partition function of structures.

Traditionally, the “information” of a sequence is being identified with its actual sequence of nucleotides. This perspective results in using sequence alignments to quantify sequence similarities, the underlying assumption being that similar sequences should exhibit similar, biological properties. The sequence, however, does not represent the entire picture. Huynen et al. (1996) study the rugged-

ness of genotype to phenotype maps and show that similar sequences can exhibit distinctly different phenotypes. Furthermore, in the context of sequence design, sequences that fold into the same structure are considered to be equal (Busch and Backofen 2006; Lorenz et al. 2011). The presence of neutral networks (Grüner et al. 1996a,b; Reidys et al. 1997) of RNA secondary structures shows that there are complementary sequences that fold into the same structure. In other words, neither does sequence similarity necessarily imply phenotypic similarity nor does sequence dissimilarity imply phenotypic dissimilarity. This motivates one to augment the sequence information by incorporating structural information.

The analysis of $EMR_{\sigma,k}^S$ -spectra represents such an augmentation: The let-7 sequence–structure pairs exhibit a distinctive difference between native and random pairs.

As a result, integrating the information of sequences and structures facilitates at minimum the extension of the local analysis conducted in Borenstein and Rupp (2006) and may as well, as a novel paradigm alone, lead to further biological insights.

MATERIALS AND METHODS

EMR spectrum: a sequence–structure pair profile

In this section we introduce the technical details of our framework, starting with Borenstein and Rupp’s definition of neutrality (Borenstein and Rupp 2006), as the average of all $3n$ single-point mutants. By construction, all mutations are equivalent, and structure has no influence on the mutation rate. However, as suggested in Pollard et al. (2006a), Mimouni et al. (2008), and Price et al. (2011), structure genuinely affects mutation rate, and mutations in turn further increase the thermal dynamic stability of the structure. We consider here mutations with respect to a fixed structure and consequently deal exclusively with compatible mutations. Furthermore, the idea of the following is to favor energetically beneficial mutations, while penalizing detrimental mutations. To adequately quantify the above, we revisit some of the basic concepts of the thermodynamic model of RNA secondary structure.

The free energy η can be considered as a result of pairing sequences and structures as follows:

$$\eta: Q_4^n \times S_n \rightarrow \mathbb{R} \cup +\infty,$$

where Q_4^n and S_n denote the space of sequences, σ , and the space of secondary structures, S , respectively, and $\eta(\sigma, S)$ is the energy of S on σ . This mapping is computed as the sum of the energy

contributions of individual base pairs (Nussinov et al. 1978). A more elaborate model (Mathews et al. 1999; Turner and Mathews 2010) evaluates the total free energy to be the sum of the energies of loops involving multiple base pairs. We remark that, if a sequence σ is not compatible with a structure S , then $\eta(\sigma, S) = +\infty$. Then MFE-folding is a map from sequences to distinguished structures (i.e., the MFE structures):

$$\text{mfe} : Q_4^n \rightarrow S_n, \quad \sigma \mapsto \underset{S \in S_n}{\operatorname{argmin}} \eta(\sigma, S).$$

To incorporate the thermodynamic information into the probability of mutations, we introduce the notion of the Boltzmann distribution of k -point mutants of a sequence–structure pair. To this end, we consider the partition function of k -point mutants of a sequence–structure pair—namely, the partition function of all sequences that are at Hamming distance k to the given sequence with respect to the given structure.

Definition 1. Let σ be a sequence of n nucleotides and let S be its associated structure (i.e., its MFE or native structure). Then the partition function of k -point mutants of σ with respect to S is given by

$$Q_{\sigma,k}^S = \sum_{\sigma', h(\sigma, \sigma')=k} e^{-\frac{\eta(\sigma', S)}{RT}}, \quad (1)$$

where h is the Hamming distance, $\eta(\sigma', S)$ is the energy of S on σ' , R is the universal gas constant, and T is the temperature.

Equation (1) represents the “dual” of McCaskill’s partition function (McCaskill 1990) with an additional Hamming distance filtration (Barrett et al. 2018). Given this partition function, we are in position to introduce the *Boltzmann distribution* of k -point mutants of σ with respect to S : The probability of a specific k -point mutant σ^* of σ , $h(\sigma^*, \sigma) = k$, with respect to S , is

$$Pr_{\sigma,k}^S(\sigma^*) = \frac{e^{-\frac{\eta(\sigma^*, S)}{RT}}}{Q_{\sigma,k}^S}.$$

This expression allows us to consider Boltzmann weighted mutations, taking into account the free energy, when realizing S .

To quantify mutational robustness (i.e., the ability to maintain the structure), we consider all k -point mutants, that fold again into S . We call the fraction of k -point mutants, σ^* , that fold into S , the *extended mutational robustness* (EMR) of the sequence–structure pair (σ, S) at k —that is, we have

$$EMR_{\sigma,k}^S = \sum_{\text{mfe}(\sigma^*)=S} Pr_{\sigma,k}^S(\sigma^*).$$

$EMR_{\sigma,k}^S$ thus quantifies the mutational robustness of a sequence–structure pair (σ, S) , with respect to Boltzmann weighted k -point mutations, taking into account the energy of the sequence, when assuming the structure S .

$EMR_{\sigma,1}^S$ can be viewed as a variation of Borenstein and Rupp’s definition of neutrality. Instead of a uniform distribution of all $3n$ single-point mutants, a Boltzmann-weighted distribution for these mutants is used. In contrast to using base pair distance as the metric on structure space, we restrict ourselves to a discrete metric.

The $EMR_{\sigma,k}^S$ -spectrum of a sequence–structure pair (σ, S) (i.e., the collection of $EMR_{\sigma,k}^S$ for varying k) allows us to study

multiple-point mutations instead of confining the analysis to single-point mutations. In this study, we shall analyze the $EMR_{\sigma,k}^S$ -spectrum of sequence–structure pairs of Hamming distances $1 \leq k \leq 20$.

The $EMR_{\sigma,k}^S$ -spectrum can be viewed as the conditional probability of the IFR in Barrett et al. (2017), conditional to specific Hamming distances. As a result, we have

$$IFR^S = \sum_{k=0}^n Pr_{\sigma}^S(k) EMR_{\sigma,k}^S,$$

$$\text{Where } Pr_{\sigma}^S(k) = \frac{Q_{\sigma,k}^S}{Q^S}, \quad Q^S = \sum_{\sigma} e^{-\frac{\eta(\sigma, S)}{RT}}.$$

Computing the EMR spectrum via Boltzmann sampling

Computing $EMR_{\sigma,k}^S$ strictly requires folding all σ' that are S -compatible, such that $h(\sigma, \sigma') = k$. Although this task is impractical for large k , $EMR_{\sigma,k}^S$ can be efficiently computed by means of Boltzmann sampling. This is conducted here via the Hamming distance restricted dual Sampler (HamSampler) (Barrett et al. 2018), which facilitates the approximation of $EMR_{\sigma,k}^S$ for any given sequence–structure pair (σ, S) and Hamming distance k . HamSampler is available at <https://github.com/FenixHuang667/HamSampler>. HamSampler takes as input (σ, S) and k and outputs sequences σ^* having Hamming distance k with probability

$$Pr_{\sigma,k}^S(\sigma^*) = \frac{e^{-\frac{\eta(\sigma^*, S)}{RT}}}{Q_{\sigma,k}^S}.$$

We then have

$$EMR_{\sigma,k}^S \approx \frac{\# \text{ of sequences folding into } S}{\# \text{ of sampled sequences}}.$$

In this article, $EMR_{\sigma,k}^S$ is calculated by sampling 5000 sequences and computing their MFE structures. The standard error of measuring $EMR_{\sigma,k}^S$ by sampling 5000 times is ± 0.0065 (derived by computing the $EMR_{\sigma,5}^S$ of an aae-let-7 sequence–structure pair 100 times). By computing the $EMR_{\sigma,k}^S$ for each $1 \leq k \leq 20$, we obtain the $EMR_{\sigma,5}^S$ -spectrum of the sequence–structure pairs (σ, S) .

miRNA data

We consider 401 miRNA precursor sequences of the let-7 gene family, obtained from the miRBase database (Kozomara and Griffiths-Jones 2013). These originate from 82 different animal species. In vertebrates and urochordates, multiple homologous miRNA genes are commonly observed, whereas single miRNA let-7 genes were more common in other animal species. We provide in the [Supplemental Materials](#) (SM) the numbers of the respective let-7 genes (see [Supplemental Fig. 1](#)). The secondary structures of these sequences are derived using a MFE folding algorithm, using the Turner energy model (Turner and Mathews 2010). We compute the $EMR_{\sigma,k}^S$ -spectra, for $1 \leq k \leq 20$, for all 401 let-7 sequence–structure pairs.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We very much appreciate the constructive comments and suggestions from the anonymous reviewers. Special thanks to Stanley Hefta for his input on this manuscript. We gratefully acknowledge the help of Kevin Shinpaugh and the computational support team at BI, Mia Shu, Thomas Li, Henning Mortveit, Madhav Marathe, and Reza Rezazadegan, for discussions. This research is partially funded by Thermo Fisher, and C.M.R. is a Thermo Fisher Scientific Fellow in Advanced Systems for Information Biology and acknowledges their support of this work.

Received January 20, 2018; accepted August 20, 2019.

REFERENCES

- Barrett C, Huang FW, Reidys CM. 2017. Sequence–structure relations of biopolymers. *Bioinformatics* **33**: 382–389. doi:10.1093/bioinformatics/btw621
- Barrett C, He Q, Huang FW, Reidys CM. 2018. An efficient dual sampling algorithm with Hamming distance filtration. *J Comput Biol* **25**: 1179–1192. doi:10.1089/cmb.2018.0075
- Belter A, Gudanis D, Rolle K, Piwecka M, Gdaniec Z, Naskręć-Barciszewska MZ, Barciszewski J. 2014. Mature miRNAs form secondary structure, which suggests their function beyond risc. *PLoS One* **9**: e113848. doi:10.1371/journal.pone.0113848
- Beniaminov A, Westhof E, Krol A. 2008. Distinctive structures between chimpanzee and human in a brain noncoding RNA. *RNA* **14**: 1270–1275. doi:10.1261/ma.1054608
- Bonnet E, Wuyts J, Rouzé P, Van de Peer Y. 2004. Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. *Bioinformatics* **20**: 2911–2917. doi:10.1093/bioinformatics/bth374
- Borenstein E, Ruppín E. 2006. Direct evolution of genetic robustness in microRNA. *Proc Natl Acad Sci* **103**: 6593–6598. doi:10.1073/pnas.0510600103
- Boyerinas B, Park S-M, Hau A, Murmann AE, Peter ME. 2010. The role of let-7 in cell differentiation and cancer. *Endocr Relat Cancer* **17**: F19–F36. doi:10.1677/ERC-09-0184
- Breaker RR. 1996. Are engineered proteins getting competition from RNA? *Curr Opin Biotechnol* **7**: 442–448. doi:10.1016/S0958-1669(96)80122-4
- Breaker RR, Joyce GF. 1994. Inventing and improving ribozyme function: rational design versus iterative selection methods. *Trends Biotechnol* **12**: 268–275. doi:10.1016/0167-7799(94)90138-4
- Busch A, Backofen R. 2006. INFO-RNA: a fast approach to inverse RNA folding. *Bioinformatics* **22**: 1823–1831. doi:10.1093/bioinformatics/btl194
- Dahiya N, Sherman-Baust CA, Wang T-L, Davidson B, Shih I-M, Zhang Y, Wood W III, Becker KG, Morin PJ. 2008. MicroRNA expression and identification of putative miRNA targets in ovarian cancer. *PLoS One* **3**: e2436. doi:10.1371/journal.pone.0002436
- Darnell JE. 2011. *RNA: life's indispensable molecule*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- de Visser JAGM, Hermisson J, Wagner GP, Meyers LA, Bagheri-Chaichian H, Blanchard JL, Chao L, Cheverud JM, Elena SF, Fontana W, et al. 2003. Perspective: evolution and detection of genetic robustness. *Evolution* **57**: 1959–1972. doi:10.1554/02-750R
- Doshi KJ, Cannone JJ, Cobaugh CW, Gutell RR. 2004. Evaluation of the suitability of free-energy minimization using nearest-neighbor energy parameters for RNA secondary structure prediction. *BMC Bioinformatics* **5**: 105. doi:10.1186/1471-2105-5-105
- Federhen S. 2011. The NCBI taxonomy database. *Nucleic Acids Res* **40**: D136–D143. doi:10.1093/nar/gkr1178
- Freilich S, Kreimer A, Borenstein E, Gophna U, Sharan R, Ruppín E. 2010. Decoupling environment-dependent and independent genetic robustness across bacterial species. *PLoS Comput Biol* **6**: e1000690. doi:10.1371/journal.pcbi.1000690
- Garcia-Martin JA, Bayegan AH, Dotu I, Clote P. 2016. RNAduelPF: software to compute the dual partition function with sample applications in molecular evolution theory. *BMC Bioinformatics* **17**: 424. doi:10.1186/s12859-016-1280-6
- Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, Bateman A. 2005. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res* **33**: D121–D124. doi:10.1093/nar/gki081
- Gruber AR, Bernhart SH, Hofacker IL, Washietl S. 2008. Strategies for measuring evolutionary conservation of RNA secondary structures. *BMC Bioinformatics* **9**: 122. doi:10.1186/1471-2105-9-122
- Grüner R, Giegerich R, Strothmann D, Reidys CM, Weber J, Hofacker IL, Stadler PF, Schuster P. 1996a. Analysis of RNA sequence structure maps by exhaustive enumeration I. structures of neutral networks and shape space covering. *Chem Mon* **127**: 355–374. doi:10.1007/BF00810881
- Grüner R, Giegerich R, Strothmann D, Reidys CM, Weber J, Hofacker IL, Stadler PF, Schuster P. 1996b. Analysis of RNA sequence structure maps by exhaustive enumeration II. structures of neutral networks and shape space covering. *Chem Mon* **127**: 375–389. doi:10.1007/BF00810882
- Gu Z, Steinmetz LM, Gu X, Scharfe C, Davis RW, Li W-H. 2003. Role of duplicate genes in genetic robustness against null mutations. *Nature* **421**: 63. doi:10.1038/nature01198
- Hammer S, Wang W, Will S, Ponty Y. 2019. Fixed-parameter tractable sampling for RNA design with multiple target structures. *BMC Bioinformatics* **20**: 209. doi:10.1186/s12859-019-2784-7
- Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, Schuster P. 1994. Fast folding and comparison of RNA secondary structures. *Monatsh Chem* **125**: 167–188. doi:10.1007/BF00818163
- Huynen MA, Stadler PF, Fontana W. 1996. Smoothness within ruggedness: the role of neutrality in adaptation. *Proc Natl Acad Sci* **93**: 397–401. doi:10.1073/pnas.93.1.397
- Kimura M. 1983. *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge.
- Kozomara A, Griffiths-Jones S. 2013. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* **42**: D68–D73. doi:10.1093/nar/gkt1181
- Lee RC, Feinbaum RL, Ambros V. 1993. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**: 843–854. doi:10.1016/0092-8674(93)90529-Y
- Letunic I, Bork P. 2006. Interactive tree of life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**: 127–128. doi:10.1093/bioinformatics/btl529
- Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* **44**: W242–W245. doi:10.1093/nar/gkw290
- Levin A, Lis M, Ponty Y, O'Donnell CW, Devadas S, Berger B, Waldspühl J. 2012. A global sampling approach to designing and reengineering RNA secondary structures. *Nucleic Acids Res* **40**: 10041–10052. doi:10.1093/nar/gks768
- Lorenz R, Bernhart SH, Höner Zu Siederdisen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA package 2.0. *Algorithms Mol Biol* **6**: 26. doi:10.1186/1748-7188-6-26
- Mathews D, Sabina J, Zuker M, Turner D. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* **288**: 911–940. doi:10.1006/jmbi.1999.2700

- Mathews D, Disney M, Childs J, Schroeder S, Zuker M, Turner D. 2004. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci* **101**: 7287–7292. doi:10.1073/pnas.0401799101
- McCaskill JS. 1990. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers* **29**: 1105–1119. doi:10.1002/bip.360290621
- McShea DW. 1996. Perspective metazoan complexity and evolution: is there a trend? *Evolution* **50**: 477–492. doi:10.1111/j.1558-5646.1996.tb03861.x
- Mimouni NK, Lyngsø RB, Griffiths-Jones S, Hein J. 2008. An analysis of structural influences on selection in RNA genes. *Mol Biol Evol* **26**: 209–216. doi:10.1093/molbev/msn240
- Nussinov R, Piecchnik G, Griggs JR, Kleitman DJ. 1978. Algorithms for loop matching. *SIAM J Appl Math* **35**: 68–82. doi:10.1137/0135006
- O'Hara AJ, Wang L, Dezube BJ, Harrington WJ, Damanian B, Dittmer DP. 2009. Tumor suppressor microRNAs are underrepresented in primary effusion lymphoma and Kaposi sarcoma. *Blood* **113**: 5938–5941. doi:10.1182/blood-2008-09-179168
- Paillart J-C, Skripkin E, Ehresmann B, Ehresmann C, Marquet R. 1996. A loop-loop “kissing” complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *Proc Natl Acad Sci* **93**: 5572–5577. doi:10.1073/pnas.93.11.5572
- Pollard KS, Salama SR, King B, Kern AD, Dreszer T, Katzman S, Siepel A, Pedersen JS, Bejerano G, Baertsch R, et al. 2006a. Forces shaping the fastest evolving regions in the human genome. *PLoS Genet* **2**: e168. doi:10.1371/journal.pgen.0020168
- Pollard KS, Salama SR, Lambert N, Lambot M-A, Coppens S, Pedersen JS, Katzman S, King B, Onodera C, Siepel A, et al. 2006b. An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* **443**: 167–172. doi:10.1038/nature05113
- Price N, Cartwright RA, Sabath N, Graur D, Azevedo RB. 2011. Neutral evolution of robustness in *Drosophila* microRNA precursors. *Mol Biol Evol* **28**: 2115–2123. doi:10.1093/molbev/msr029
- Reidys CM, Stadler PF, Schuster P. 1997. Generic properties of combinatorial maps and neutral networks of RNA secondary structures. *Bull Math Biol* **59**: 339–397. doi:10.1007/BF02462007
- Rodrigo G, Fares MA. 2012. Describing the structural robustness landscape of bacterial small RNAs. *BMC Evol Biol* **12**: 52. doi:10.1186/1471-2148-12-52
- Schlichting CD, Pigliucci M. 1998. *Phenotypic evolution: a reaction norm perspective*. Sinauer Associates, Sunderland, Massachusetts.
- Serganov A, Patel DJ. 2007. Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat Rev Genet* **8**: 776. doi:10.1038/nrg2172
- Takamizawa J, Konishi H, Yanagisawa K, Tomida S, Osada H, Endoh H, Harano T, Yatabe Y, Nagino M, Nimura Y, et al. 2004. Reduced expression of the *let-7* microRNAs in human lung cancers in association with shortened postoperative survival. *Cancer Res* **64**: 3753–3756. doi:10.1158/0008-5472.CAN-04-0637
- Tanzer A, Stadler PF. 2004. Molecular evolution of a microRNA cluster. *J Mol Biol* **339**: 327–335. doi:10.1016/j.jmb.2004.03.065
- Turner D, Mathews DH. 2010. NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Res* **38**: 280–282. doi:10.1093/nar/gkp892
- Waterman MS. 1978. Secondary structure of single-stranded nucleic acids. *Adv Math* **1**: 167–212.
- Wilke CO, Wang JL, Ofria C, Lenski RE, Adami C. 2001. Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature* **412**: 331. doi:10.1038/35085569
- Zuker M, Stiegler P. 1981. Optimal computer folding of larger RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res* **9**: 133–148. doi:10.1093/nar/9.1.133

RNA

A PUBLICATION OF THE RNA SOCIETY

Genetic robustness of let-7 miRNA sequence–structure pairs

Qijun He, Fenix W. Huang, Christopher Barrett, et al.

RNA 2019 25: 1592-1603 originally published online September 23, 2019

Access the most recent version at doi:[10.1261/rna.065763.118](https://doi.org/10.1261/rna.065763.118)

Supplemental Material

<http://rnajournal.cshlp.org/content/suppl/2019/09/23/rna.065763.118.DC1>

References

This article cites 50 articles, 7 of which can be accessed free at:
<http://rnajournal.cshlp.org/content/25/12/1592.full.html#ref-list-1>

Creative Commons License

This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://rnajournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *RNA* go to:

<http://rnajournal.cshlp.org/subscriptions>