September 2021

Phenotype bias determines how natural RNA structures occupy the morphospace of all possible shapes

Fatme Ghaddar
gust0016134@gust.edu.kw

Dr Kamaludin Dingle
Gulf University for Science and Technology, Kuwait, dingle.k@gust.edu.kw

Dr Petr Sulc
Arizona State University, USA

Prof Ard A. Louis
University of Oxford, UK

Preprint on https://www.biorxiv.org/content/10.1101/2020.12.03.410605v1

Phenotype bias determines how natural RNA structures occupy the morphospace of all possible shapes

Fatme Ghaddar

Faculty supervisor: Dr Kamaludin Dingle (Gulf University for Science and Technology, Kuwait)

Collaborators: Prof Ard Louis (Oxford University, UK), Dr Petr Sulc (Arizona State University, USA)

Introduction

One of the long-standing controversies in evolutionary biology is the relative prominence of developmental bias versus natural selection.
In this paper, we demonstrate quantitatively that developmental bias is the primary explanation for the occupation of the morphospace of RNA secondary structure (SS) shapes. Coarse-grained SS classes methods are defined using RNAshapes method, where we measure the frequencies that non-coding RNA SS shapes appear in nature.

Key findings

1. Only the most frequent RNA structures appear in nature.
2. These frequencies are accurately predicted by the likelihood that structures appear upon uniform random sampling of sequences, which tightly constrains evolutionary dynamics to only act within a reduced subset of structures that are easy to "find".

Methods

Folding RNA:
The Vienna Package is used to get structures from folding sequences [1]. The G-sampling had random samples of lengths 5×10^6 for L = 40 and L = 55, and 10^5 for L = 70, 85, 100, 126. Moreover, fRNAdb sequences [2] were folded and used after discarding a very small fraction of sequences that had non-standard letters such as 'N' or 'R'.
Abstract shapes:
RNA SS are abstracted into a standard dot-bracket notation to obtain coarse grained abstract shapes. This was done using the RNAshapes tool at https://bibiserv.cebitec.uni-bielefeld.de/rnashapes.

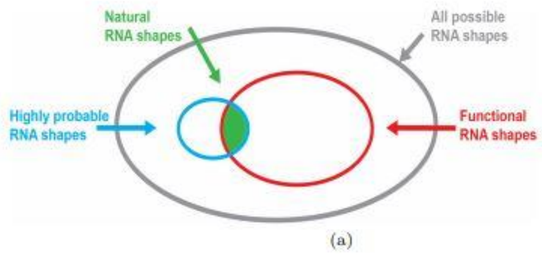

Figure 1. Conceptual diagram of the RNA SS shape morphospace

Results

1. Only a tiny fraction of the RNA SS phenotypic variation that is possibly available is utilized by Nature.
2. Remarkably, only a small numbers of sequences are needed to recover the full set of abstract shapes in the fRNAdb database.
3. G-sampling remarkably predicts the frequency with which structures are found in nature.

Data Analysis

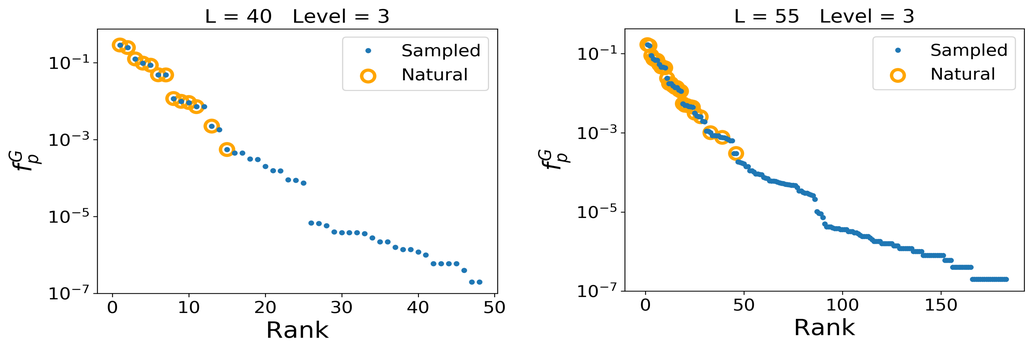

Figure 2. Nature selects highly frequent structures: The frequency f_p^G (blue dots) of each abstract shape, calculated by random sampling of sequences (G-sampling), is plotted versus the rank. Yellow circles highlight which of the randomly generated shapes were also found in the fRNAdb.

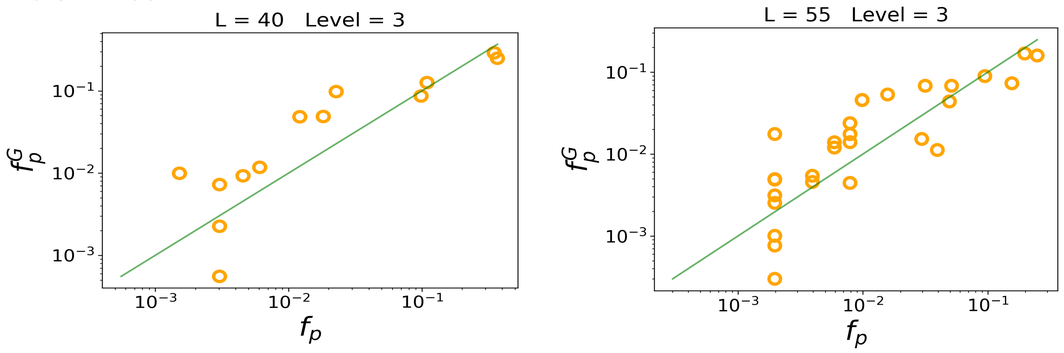

Figure. 3. The frequency of shapes in the nature correlates with the frequency of shapes from random sampling: Yellow circles denote the frequencies f_p of natural RNA from the fRNAdb [2]. The green line denotes x = y, i.e natural and sampled frequencies coincide.

Conclusions

In conclusion, the RNA sequence to SS map describes a pared down case of development, this simplicity is a strength as it gives us the ability explore counterfactual questions [4] such as: what kind of phenotypic variation did not appear due to phenotypic bias. According to our knowledge thus far, this system provides the cleanest evidence yet for developmental bias, hence strongly affecting evolutionary outcomes. A strong phenotype bias is also shown on many other GP maps [5,6]. An important future work question will be whether there is a universal structure to this phenotype bias and whether it also has such a clear effect on evolutionary outcomes in other biological systems.

Acknowledgements

We thank David McCandlish for discussions, and Michael Matthies for help with Fig 1(a).

References

[1].R. Lorenz, S. H. Bernhart, C. H. Zu Siederdissen, H. Tafer, C. Flamm, P. F. Stadler, and I. L. Hofacker, Algorithms for molecular biology 6, 26 (2011).
[2].T. Mituyama, K. Yamada, E. Hattori, H. Okida, Y. Ono, G. Terai, A. Yoshizawa, T. Komori, and K. Asai, Nucleic Acids Research 37, D89 (2009).
[3]. S. Janssen and R. Giegerich, Bioinformatics 31, 423 (2015).
[4]. A. A. Louis, Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences 58, 107 (2016).
[5] S. Manrubia, J. A. Cuesta, J. Aguirre, S. E. Ahnert, L. Altenberg, A. V. Cano, P. Catal´an, R. Diaz-Uriarte, S. F. Elena, J. A. Garc´ıa-Mart´ın, et al., arXiv preprint arXiv:2002.00363 (2020).
[6] K. Dingle, C. Q. Camargo, and A. A. Louis, Nature communications 9, 761 (2018).