

Software

Open Access

EvoRSR: an integrated system for exploring evolution of RNA structural robustness

Wenjie Shu^{†1,2}, Ming Ni^{†1,3}, Xiaochen Bo^{*1}, Zhiqiang Zheng² and Shengqi Wang^{*1}

Address: ¹Beijing Institute of Radiation Medicine, Beijing 100850, PR China, ²College of Electro-Mechanic and Automation, National University of Defense Technology, Changsha, Hunan 410073, PR China and ³Center for Bioinformatics, National Laboratory of Protein Engineering and Plant Genetic Engineering, College of Life Sciences, Peking University, Beijing 100871, PR China

Email: Wenjie Shu - shuwj@bmi.ac.cn; Ming Ni - nim@mail.cbi.pku.edu.cn; Xiaochen Bo* - boxc@bmi.ac.cn; Zhiqiang Zheng - xzyzheng@sohu.com; Shengqi Wang* - sqwang@bmi.ac.cn

* Corresponding authors †Equal contributors

Published: 13 August 2009

Received: 17 April 2009

BMC Bioinformatics 2009, 10:249 doi:10.1186/1471-2105-10-249

Accepted: 13 August 2009

This article is available from: <http://www.biomedcentral.com/1471-2105/10/249>

© 2009 Shu et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Robustness, maintaining a constant phenotype despite perturbations, is a fundamental property of biological systems that is incorporated at various levels of biological complexity. Although robustness has been frequently observed in nature, its evolutionary origin remains unknown. Current hypotheses suggest that robustness originated as a direct consequence of natural selection, as an intrinsic property of adaptations, or as a congruent correlate of environment robustness. To elucidate the evolutionary origins of robustness, a convenient computational package is strongly needed.

Results: In this study, we developed the open-source integrated system EvoRSR (Evolution of RNA Structural Robustness) to explore the evolution of robustness based on biologically important landscapes induced by RNA folding. EvoRSR is object-oriented, modular, and freely available at <http://biotech.bmi.ac.cn/EvoRSR> under the GNU/GPL license. We present an overview of EvoRSR package and illustrate its features with the miRNA gene *cel-mir-357*.

Conclusion: EvoRSR is a novel and flexible package for exploring the evolution of robustness. Accordingly, EvoRSR can be used for future studies to investigate the evolution and origin of robustness and to address other common questions about robustness. While the current EvoRSR environment is a versatile analysis framework, future versions can include features to enhance evolutionary studies of robustness.

Background

Robustness is a fundamental and ubiquitous phenomenon in biological systems, in which phenotypes are resistant to change in the presence of various perturbations. When these perturbations are inherited, such as genetic mutations, the phenomenon is known as genetic robustness. Alternatively, when the perturbations are due to

environmental factors, the phenomenon is called environmental robustness [1]. Both types of robustness appear at various levels of biological organization, affecting gene expression, protein folding, metabolic flux, physiological homeostasis, development, and organism fitness [2]. Biologists' long-standing interest in robustness has roots in Fisher's work on dominance [3-5] and Wadding-

ton's developmental canalization research [6,7]. Despite being found throughout nature, the evolutionary origins of robustness remain unclear. Current competing explanations for the origins of robustness include that it evolves as a direct consequence of natural selection, as an intrinsic property of adaptations, or as congruent correlate of environment robustness. Additionally, it is unknown how robustness evolves and how the robustness varies along the Hamming distance from the WT sequence.

Addressing these questions requires a convenient computational package that will fully elucidate the evolutionary origins of robustness. A good example to study for clarifying the origins of robustness is RNA folding from sequences into secondary structures. RNA folding provides a convenient biophysical model of a genotype-phenotype map that has been used in studies for robustness, evolvability, and epistasis. In such studies, RNA folding can be precisely defined and statistically measured, revealing simultaneous and non-independent effects of natural selection [8,9]. These studies have focused on the robustness of RNA folding in viruses [10-12], viroids [13,14], and microRNAs [15-18].

Given a quantitative measure of structural robustness [15,18,19], we developed an integrated system named EvoRSR (Evolution of RNA Structural Robustness) to explore the evolution of robustness based on important landscapes induced by RNA folding. EvoRSR is object-oriented, modular in design and freely available at <http://biotech.bmi.ac.cn/EvoRSR> under the GNU/GPL license. This open-source package inspects the evolution and origin of robustness through sampling genotype (sequence) space at each Hamming distance from the WT sequence. Here, we describe the EvoRSR package and analyze the miRNA gene *cel-mir-357* to illustrate how EvoRSR works.

Implementation

Mechanism and workflow of EvoRSR

Figure 1 illustrates the mechanism of EvoRSR. EvoRSR studies the evolution of robustness based on landscapes that result from mapping micro-configurations to scalar or nonscalar entities. Here, the micro-configurations are sequences of nucleotides. The scalar properties include free-energy of secondary structure and neutrality. Free-energy of secondary structure describes the thermodynamic stability of RNA secondary structure (conferring environmental robustness) [15,16,19]. Neutrality (see Figure 1a) quantitatively measures the genetic robustness of RNA secondary structure [15,16,18,19]. Based on these two scalar properties, we defined the free-energy landscape and neutrality landscape, respectively. The nonscalar structure landscape is generated from the RNA secondary structure. Based on these three landscapes, EvoRSR investigates the evolution of robustness in the phenotype space by sampling on genotype (sequence)

space at each Hamming distance from the WT RNA sequence (see Figure 1b).

The EvoRSR package is a free package written in C, which runs in a command-line mode within a Linux/Unix environment. The Vienna RNA package [20] is required to run the program. Detailed installation instructions for EvoRSR are provided on its web site. Currently, three programs are included in this package. Figure 2 shows the workflow of EvoRSR.

Evaluation of genetic and environmental robustness

Formally, the neutrality η of an RNA sequence with length l is defined as

$$\eta = \left\langle \frac{l-d}{l} \right\rangle$$

where d is the base-pair distance between the secondary structures of the WT sequence and its mutant, averaged over all $3 \times l$ one-mutant neighbors. d is calculated by RNADISTANCE in Vienna RNA package [21]. Thus, η represents the average fraction of the structure that remains unchanged after a mutation occurs. The free-energy, dG , is quantitatively measures the thermodynamic stability (which confers environmental robustness) of a WT RNA sequence [15-17,19]. dG is calculated as the minimum free-energy of secondary structure obtained by RNAFOLD in Vienna RNA package [21]. In the EvoRSR package, *Evoneu* is applied to calculate the η s and dG s of the sequences in a Fasta file (see Figure 2).

Because RNA molecules may function in dynamical, structural reconfigurations [22,23], an RNA molecule is better described by an ensemble of secondary structures, that have free energies close to the minimum of free-energy. In this case, we revise the quantitative definition of genetic and environmental robustness. The base-pair distance d in equation (1) is replaced by the general multi-structure distance between the ensemble of secondary structures of the WT sequence and its mutant [24], and the minimum free-energy dG is replaced by the ensemble free-energy.

Landscape and its density surfaces

For each WT RNA sequence, we employ a Monte Carlo method to sample sequences in the genotype space at each Hamming distance from the WT RNA sequence. The set of total sampling sequences is denoted by S , which can be divided into subsets S_i , $i = 1, 2, \dots, l$ that represent the set of sampling sequences within a Hamming distance of i from the WT sequence. All the subsets have an identical size, ($|S_i| = N$, $i = 1, 2, \dots, l$).

As a generic tool for the study statistical properties of landscapes, we propose the use of a two-dimensional probability density surface [25,26]. A density surface $P(t|h)$ is

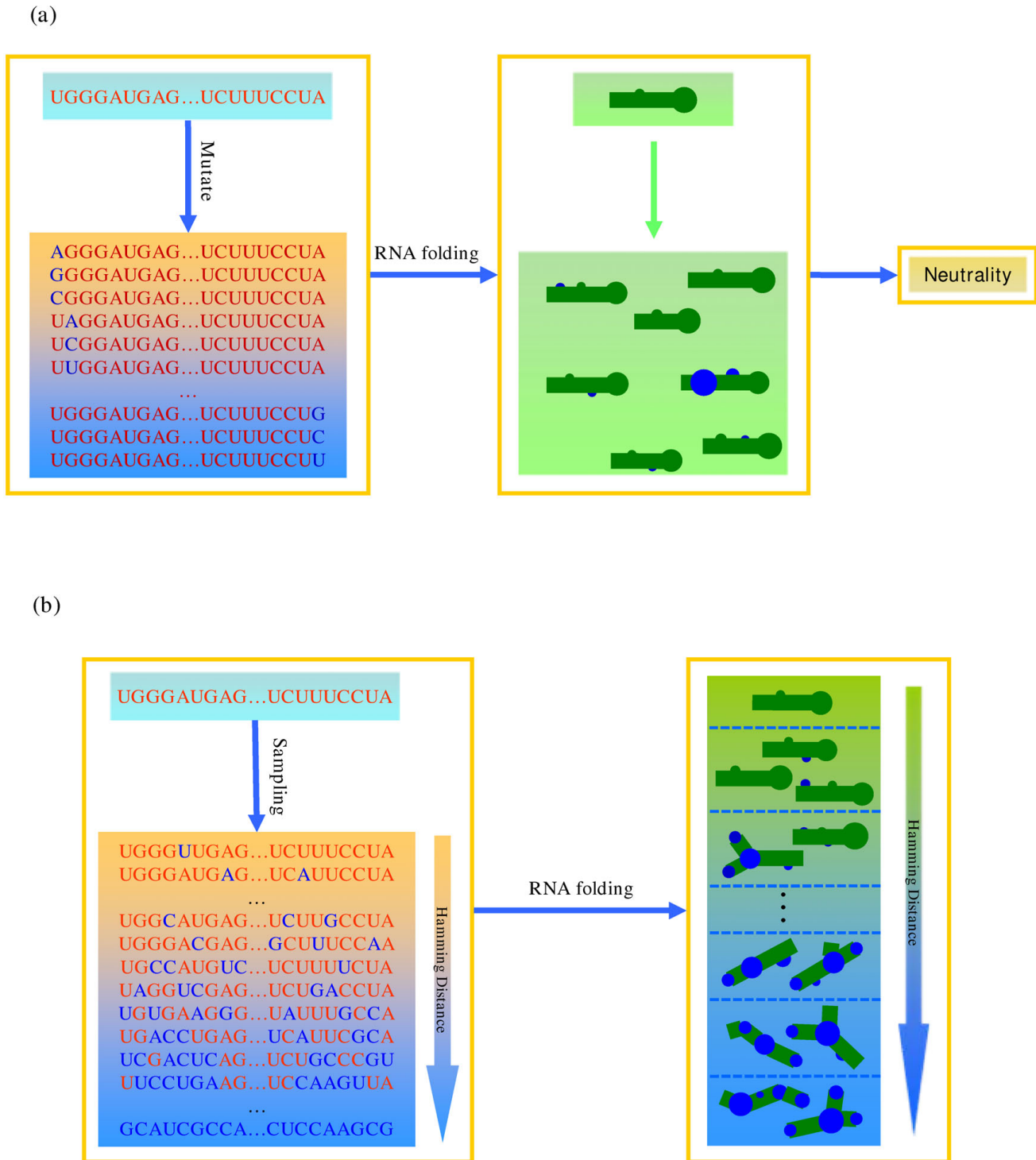


Figure 1
The mechanism of EvoRSR. (a) Evaluation of genetic robustness. (b) Sampling and folding on genotype space at each Hamming distance.

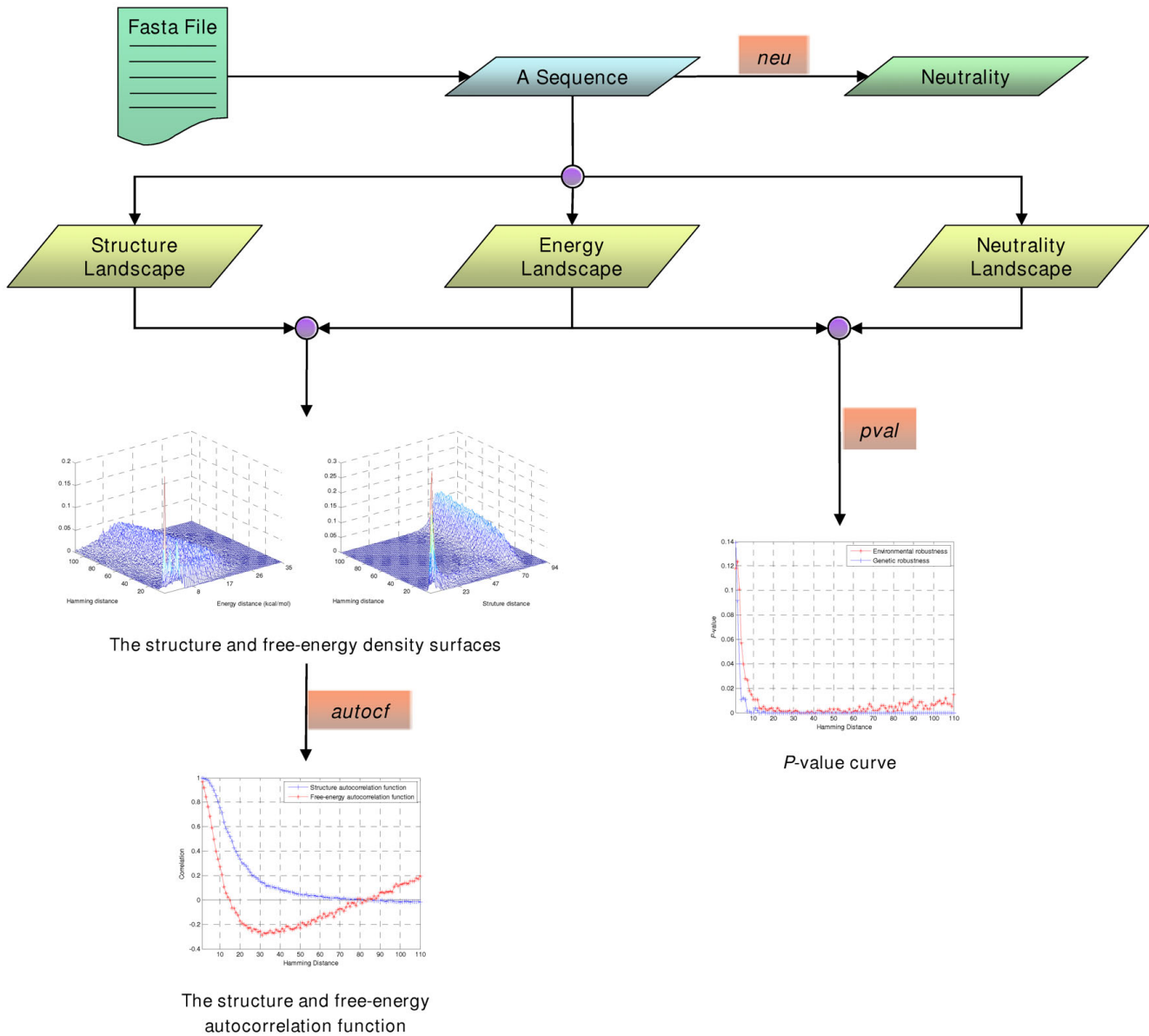


Figure 2
The workflow of EvoRSR.

the conditional probability that given two sequences Hamming distance h from each other, the two configurations have either a base-pair distance t or a free-energy difference t . The density surface describes how the distributions of free-energy values and configuration differences change along the Hamming distance from the WT sequence. Furthermore, the density surface condenses statistical aspects of the correlation between sequences and structures and provides a tool to derive and calculate local and global properties of sequence-structure relations.

Autocorrelation function and correlation length

Landscape can be characterized statistically by autocorrelation functions [27,28], which can be expressed in terms of mean squared distance:

$$\rho(h) = 1 - \frac{\langle d^2(h) \rangle}{\langle d^2 \rangle}$$

$\langle d^2 \rangle$ is the mean squared distance sampled over the entire sequence space, and $\langle d_t^2(h) \rangle$ is the conditional mean squared distance. Autocorrelation functions of base-pair distances $\rho(h)$ are approximated by an exponential fit to calculate a correlation length ℓ for secondary structures in sequence space:

$$\ln \rho(\ell) = -1$$

The correlation length increases roughly with the sequence length l [25]. Autocorrelation functions and correlation lengths of structures characterize the sequence-structure relation by a single function or a single value, respectively. They provide a useful measure for the sensitivity of RNA structures against point mutations. In the EvoRSR package, they are computed by the program *Evoautocf* (see Figure 2).

P-value curve of robustness

For each WT RNA sequence, EvoRSR measures the neutrality of the WT sequence, η_{WT} , and evaluates the neutralities η_i^j , $i = 1, 2, \dots, l$, $j = 1, 2, \dots, N$ of the corresponding sampling sequences in S_i , $i = 1, 2, \dots, l$. To evaluate the level of the increased neutrality for each WT sequence at each Hamming distance separately, the rank of the neutrality of WT sequence, r_i , $i = 1, 2, \dots, l$, among the neutralities of the sampling sequences in S_i , $i = 1, 2, \dots, l$, is calculated. This order statistics measure has no requirements on the nature of the neutrality value distribution. The significance level of robustness of WT sequence at each Hamming distance is then defined as the *P*-value curve $p_i = \frac{r_i}{N+1}$, $i = 1, 2, \dots, l$, which estimates the probability of observing an equal or higher neutrality value by chance at each Hamming distance. The same analysis applies to the environmental robustness, in which the neutrality of a WT RNA sequence is replaced by its free-energy, dG . The significance analysis process is realized by the program *Evoeval* in the EvoRSR package (see Figure 2).

Results and discussion

To illustrate how EvoRSR can be used to study the evolution of robustness, we analyzed the *C. elegans* miRNA *mir-357* (see Figure 2). The detail results are presented as Additional file available on the website of EvoRSR [see Additional file 1]. Our result indicates that along the Hamming distance from the WT sequence the genetic and environmental robustness of miRNA gene *cel-mir-357* vary in a consistent way, and the sub-optimal structures may have little effect on our conclusions [see Additional file 1].

Robustness reduces an organism's susceptibility to genetic and environmental perturbations. To understand the evolutionary origins of robustness, we needed to know how phenotype and genotype are related, and how the genotype-phenotype map interacts with evolution. We developed a convenient computational package EvoRSR to fully elucidate the evolutionary mechanisms of the genetic robustness in RNA structure. EvoRSR can investigate the statistical details of RNA structure and the free-energy landscapes, providing the corresponding autocorrelation function and correlation length. Based on these landscapes, EvoRSR explored the evolution of genetic robustness along the Hamming distance from the WT sequence. By providing the *P*-value curves of both genetic and environmental robustness, EvoRSR presents a scenario of how, and how fast, significant levels of robustness vary along the Hamming distance from the WT sequence. Additionally, EvoRSR helped examine the statistical relationship between genetic and environmental robustness along the Hamming distance from the WT sequence.

EvoRSR is a novel and flexible package for exploring the evolution of genetic robustness. EvoRSR was used to study the robustness of RNA secondary structures, providing a promising framework to examine central issues concerning the evolution of robustness [15,16]. Recently, we examined the neutrality of the structural element in 1,082 native miRNA genes from six species and demonstrated that the structural elements within native miRNA genes exhibited a significantly higher level of genetic robustness [18]. An examination of miRNAs of several eukaryotic species revealed that the stem-loop structures of miRNA genes exhibits a significantly higher level of genetic robustness compared to randomly reshuffled pseudo miRNAs [15,16]. This finding indicated that the excess robustness of miRNAs goes beyond the intrinsic robustness of the stem-loop structure. Our results indicate that the increased genetic robustness of miRNAs may result from congruent evolution for environment robustness [16]. However, Borenstein and Ruppin suggested that the excess robustness of miRNA stem-loops results from direct evolutionary pressure for increased robustness [15]. Furthermore, these studies do not solve how both genetic as well as environmental robustness evolve or how environmental and genetic robustness correlate with each other along the evolutionary path from the WT sequence. EvoRSR will elucidate the evolutionary mechanisms of genetic robustness.

While the EvoRSR environment is a versatile analysis framework already in the present version, there have many options for further enhancement. The mechanisms underlying robustness are diverse, ranging from thermodynamic stability at the RNA and protein level to behavior

at the organismal level [2]. The increased neutrality and thermodynamic stability of RNAs examined by EvoRSR can be conceived as first-order robustness, based only on RNA folding map that that assigns each sequence to a minimum-free-energy structure. The simplicity of this form of robustness, the full tractability of RNA secondary structure, and the complete control of reference background facilitate the exploration of its evolutionary origins. Protein structures, a step up in complexity, may possess similar features to test the evolution of robustness. With a better understanding of protein folding and more accurate prediction algorithms [29], our methodology can be applied to the evolution of robustness in protein structures. Based on the understanding of the first-order robustness, we can further explore the evolution of higher-level robustness.

Conclusion

In this study, we developed the open-source integrated system EvoRSR (Evolution of RNA Structural Robustness) to explore the evolution of robustness based on biologically important landscapes induced by RNA folding. EvoRSR is object-oriented, modular, and freely available at <http://biotech.bmi.ac.cn/EvoRSR> under the GNU/GPL license. EvoRSR can be used for future studies to investigate the evolution and origin of robustness and to address other common questions about robustness. While the current EvoRSR environment is a versatile analysis framework, future versions can include features to enhance evolutionary studies of robustness.

Availability and requirements

Project name: EvoRSR (Evolution of RNA Structural Robustness)

Project home page: <http://biotech.bmi.ac.cn/EvoRSR>

Operating system(s): Linux, UNIX (no GUI)

Programming language: C++ and Perl

Other requirements: Vienna RNA package

License: GNU/GPL license

Restrictions to use by non-academics: None

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

WS and MN wrote the programs, analyzed the results. WS drafted the manuscript. XB and ZZ helped in analysis and discussion, gave useful comments. SW and XB guided the

project. All authors read and approved the final manuscript.

Additional material

Additional file 1

The results of cel-mir-357. The C. elegans microRNA (miRNA) mir-357 (cel-mir-357) is analyzed as example to illustrate how EvoRSR can be helpful for studying the evolution of robustness.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-10-249-S1.pdf>]

Acknowledgements

The authors would like to thank the Super Biomed Computation Center at Beijing Institute of Health Administration and Medicine Information for providing computing resources. This work is supported by grants from the National High Technology Research and Development Program of China (No. 2007AA02Z311 and No. 2006AA02Z304) and grants from the National Nature Science Foundation of China (No. 30700139 and No. 30600120).

References

1. Wagner GP, Booth G, Bagheri-Chaichian H: **A population genetic theory of canalization.** *Evolution* 1997, **v51**:329-347.
2. de Visser JA, Hermisson J, Wagner GP, Ancel ML, Bagheri-Chaichian H, Blanchard JL, Chao L, Cheverud JM, Elena SF, Fontana W, Gibson G, Hansen TF, Krakauer D, Lewontin RC, Ofria C, Rice SH, von Dassow G, Wagner A, Whitlock MC: **Perspective: Evolution and detection of genetic robustness.** *Evolution Int J Org Evolution* 2003, **57**:1959-1972.
3. Fisher RA: **The possible modifications of the response of the wild type to recurrent mutations.** *Amer Nat* 1928, **62**:115-116.
4. Fisher RA: **Two further notes on the origin of dominance.** *Amer Nat* 1928, **62**:571-574.
5. Fisher RA: **The evolution of dominance.** *Biological reviews* 1931, **6**:345-368.
6. Waddington CH: **The genetic assimilation of an acquired character.** *Evolution* 1953, **7**:118-126.
7. Waddington CH: *The strategy of the genes* New York:MacMillan; 1957.
8. Fontana W, Schuster P: **Shaping space: the possible and the attainable in RNA genotype-phenotype mapping.** *J Theor Biol* 1998, **194**:491-515.
9. Schuster P, Fontana W, Stadler PF, Hofacker IL: **From sequences to shapes and back: a case study in RNA secondary structures.** *Proc Biol Sci* 1994, **255**:279-284.
10. Wagner A, Stadler PF: **Viral RNA and evolved mutational robustness.** *J Exp Zool* 1999, **285**:119-127.
11. Elena SF, Carrasco P, Daros JA, Sanjuan R: **Mechanisms of genetic robustness in RNA viruses.** *EMBO Rep* 2006, **7**:168-173.
12. Montville R, Froissart R, Remold SK, Tenaillon O, Turner PE: **Evolution of mutational robustness in an RNA virus.** *PLoS Biol* 2005, **3**:e381.
13. Sanjuan R, Forment J, Elena SF: **In silico predicted robustness of viroids RNA secondary structures. I. The effect of single mutations.** *Mol Biol Evol* 2006, **23**:1427-1436.
14. Sanjuan R, Forment J, Elena SF: **In Silico Predicted Robustness of Viroids RNA Secondary Structures. II. Interaction Between Mutation Pairs.** *Mol Biol Evol* 2006, **23**:2123-2130.
15. Borenstein E, Ruppin E: **Direct evolution of genetic robustness in microRNA.** *Proc Natl Acad Sci USA* 2006, **103**:6593-6598.
16. Shu W, Bo X, Ni M, Zheng Z, Wang S: **In silico genetic robustness analysis of microRNA secondary structures: potential evidence of congruent evolution in micro RNA.** *BMC Evol Biol* 2007, **7**:223.
17. Bonnet E, Wuys J, Rouze P, Van de PY: **Evidence that microRNA precursors, unlike other non-coding RNAs, have lower fold-**

- ing free energies than random sequences. *Bioinformatics* 2004, **20**:2911-2917.
18. Shu W, Ni M, Bo X, Zheng Z, Wang S: **In Silico Genetic Robustness Analysis of Secondary Structural Elements in the miRNA Gene.** *J Mol Evol* 2008, **67**:560-569.
 19. Shu W, Bo X, Zheng Z, Wang S: **RSRE: RNA structural robustness evaluator.** *Nucleic Acids Res* 2007, **35**:W314-W319.
 20. Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, Schuster P: **Fast folding and comparison of RNA secondary structures.** *Monatshefte für Chemie/Chemical Monthly* 1994, **125**:167-188.
 21. Hofacker IL: **Vienna RNA secondary structure server.** *Nucleic Acids Res* 2003, **31**:3429-3431.
 22. Hou T, Chen K, McLaughlin WA, Lu B, Wang W: **Computational analysis and prediction of the binding motif and protein interacting partners of the Abl SH3 domain.** *PLoS Comput Biol* 2006, **2**:e1.
 23. Rischel C, Spiedel D, Ridge JP, Jones MR, Breton J, Lambry JC, Martin JL, Vos MH: **Low frequency vibrational modes in proteins: changes induced by point-mutations in the protein-cofactor matrix of bacterial reaction centers.** *Proc Natl Acad Sci USA* 1998, **95**:12306-12311.
 24. Bonhoeffer S, McCaskill JS, Stadler PF, Schuster P: **RNA multi-structure landscapes. A study based on temperature dependent partition functions.** *Eur Biophys J* 1993, **22**:13-24.
 25. Fontana W, Konings DA, Stadler PF, Schuster P: **Statistics of RNA secondary structures.** *Biopolymers* 1993, **33**:1389-1404.
 26. Fontana W, Stadler PF, Bornberg-Bauer EG, Griesmacher T, Hofacker IL, Tacker M, Tarazona P, Weinberger ED, Schuster P: **RNA folding and combinatory landscapes.** *PHYSICAL REVIEW E STATISTICAL PHYSICS, PLASMAS, FLUIDS, AND RELATED INTERDISCIPLINARY TOPICS* 1993, **47**:2083-2099.
 27. Eigen M, McCaskill JS, Schuster P: **The molecular quasi-species.** *Adv Chem Phys* 1989, **75**:149-263.
 28. Weinberger ED: **Correlated and uncorrelated fitness landscapes and how to tell the difference.** *Biol Cybern* 1990, **63**:325-336.
 29. Baker D: **A surprising simplicity to protein folding.** *Nature* 2000, **405**:39-42.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

