Direct evolution of genetic robustness in microRNA

Elhanan Borenstein*† and Eytan Ruppin*†‡

*School of Computer Science and ‡School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

Edited by Tomoko Ohta, National Institute of Genetics, Mishima, Japan, and approved March 1, 2006 (received for review December 8, 2005)

Genetic robustness, the invariance of the phenotype in the face of genetic perturbations, can endow the organism with reduced susceptibility to mutations. A large body of work in recent years has focused on the origins, mechanisms, and consequences of robustness in a wide range of biological systems. Despite the apparent prevalence of mutational robustness in nature, however, its evolutionary origins are still unclear. Does robustness evolve directly by natural selection or is it merely a correlated byproduct of other phenotypic traits? By examining microRNA (miRNA) genes of several eukaryotic species, we show that the structure of miRNA precursor stem–loops exhibits a significantly high level of mutational robustness in comparison with random RNA sequences with similar stem–loop structures. Hence, this excess robustness of miRNA goes beyond the intrinsic robustness of the stem–loop hairpin structure. Furthermore, we show that it is not the byproduct of a base composition bias or of thermodynamic stability. These findings suggest that the excess robustness of miRNA stem–loops is the result of direct evolutionary pressure toward increased robustness. We further demonstrate that this adaptive robustness evolves to compensate for structures with low intrinsic robustness.

Genetic robustness, the preservation of an optimal phenotype facing genetic mutations, is a fundamental property of living systems incorporated at various levels of biological complexes (1). Protein tolerance to amino acid substitutions (2), gene dispensability in yeast (3), and the error tolerance of complex biological networks (4) are just a few examples. Yet, despite the plethora of observations of genetic robustness, its evolutionary origins are less obvious. In a recent review of the debate concerning the evolutionary origins of robustness, de Visser *et al.* (1) have grouped the theories addressing the evolution of genetic robustness into three main classes: (*i*) adaptive, wherein robustness is favored by natural selection and hence evolves directly; (*ii*) intrinsic, suggesting that robustness is a nonadaptive correlated byproduct of some other phenotypic traits; and (*iii*) congruent, conjecturing a correlation between environmental and genetic robustness and in which genetic robustness is the side effect of the evolution toward environmental robustness. To date, however, the extent to which each of these evolutionary forces contributes to the evolution of robustness remains unresolved, which is partly because providing evidence for the evolution of adaptive robustness, i.e., robustness that is the direct product of the evolutionary process and not a correlated outcome of other phenotypic properties, is a challenging task.

Addressing this challenge, several studies in recent years have resorted to theoretical analyses and computer simulations, examining the evolutionary origins of genetic robustness. These studies examine a wide range of mechanisms and aspects of robustness, including redundancy and fragility (5), duplicate genes and alternative pathways (6, 7), complex networks structure (4), and dominance (8). Some of these studies have focused on increased neutrality (i.e., a higher than chance frequency of one mutant neighbors that preserve the phenotype) as a simple manifestation of genetic robustness. Although increased neutrality may lack some of the complex features characterizing other mechanisms of robustness, it forms an ideal model for studying the evolutionary origins, properties, and effects of robustness (see also *Discussion*). By using mathematical modeling, it has been shown that populations percolating across wide neutral networks (9) tend to concentrate on highly neutral regions of the genotypic space (and hence, give rise to mutational robustness) as a consequence of neutral evolution dynamics (10). This model can account for the direct evolution of robustness even without explicit selection pressures favoring robust individuals. The emergence of increased neutrality through evolution was also demonstrated in computer simulations (11, 12).

Yet, evidence for the evolutionary origins of mutational robustness in natural biological systems has been difficult to obtain. In a recent study, Montville *et al.* (13) have demonstrated an increased mutational robustness in RNA viruses that were experimentally evolved in low vs. high coinfection, suggesting that coinfection and complementation can serve as an alternative mechanism for buffering deleterious mutations. However, it is more likely that the observed difference in this study stems from the loss of mutational robustness due to weakened selection pressures rather than a direct evolution of robustness. Another approach to inferring the evolution of robustness is the use of a plausible background model, allowing evaluation of the significance of any excess robustness found in the WT (1). One important effort to provide such evidence has focused on the mutational robustness of conserved secondary structure RNA elements in a viral genome by using a random sample of nonconserved elements from the same genome as a reference set (14). It was shown that conserved elements are consistently more robust than nonconserved elements. Other studies, focusing on the thermodynamic stability of noncoding RNA secondary structures (which confers environmental robustness) have used sequence shuffling to generate a reference set (e.g., refs. 15–18). These studies provide valuable support for the evolution of robustness in natural systems, however, because the reference sets used in these studies do not preserve the structural phenotype of the native elements, and because certain secondary structures are inherently more robust than others, it cannot be determined whether the observed increase in robustness evolved independently or is a correlated side effect of the evolution of specific, functionally important, structures. To provide evidence that the genetic robustness of the WT is the outcome of direct evolution toward robustness, it is essential to demonstrate that the WT is significantly more robust than other, phenotypically similar, genetic configurations. Such a background model can be constructed for microRNAs (miRNAs), allowing to carefully control for the effects of secondary structure evolution.

We study the genetic robustness of miRNAs, small endogenous noncoding RNAs that regulate the expression of protein-coding genes in animal and plants via miRNA cleavage or translational repression (19–22). The short mature miRNAs (≈22 nt) originate from longer RNA precursor molecules that fold into a stem–loop hairpin structure (23, 24). The secondary structure of miRNA stem–loops embodies a combination of three features that makes it a particularly suitable test bed for studying the evolution of genetic robustness: (*i*) Serving a crucial

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviation: miRNAs, microRNAs.

†To whom correspondence may be addressed. E-mail: borens@post.tau.ac.il or ruppin@post.tau.ac.il.

EVOLUTION

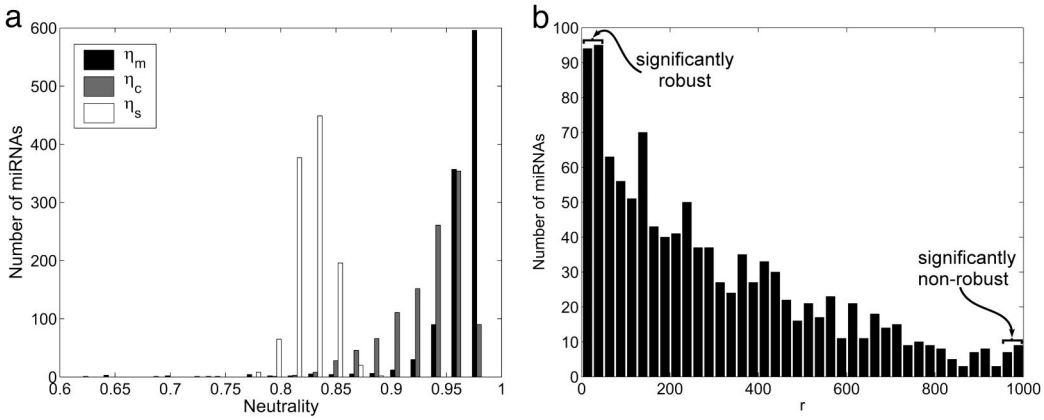

Fig. 1. Neutrality and rank distributions for all 1,120 miRNAs genes included in the analysis. (*a*) Neutrality value distribution for native stem–loops (η_m) and for the corresponding inversely folded stem–loops (η_c). On average, the neutrality of a native stem–loop is 0.96 ± 0.03, whereas that of the inversely folded sequence is 0.93 ± 0.03. Evidently, native miRNA stem–loops possess higher neutrality values than random sequences with identical structures ($p < 10^{-137}$; Wilcoxon's signed rank test for paired data). The neutrality distribution of shuffled sequences (η_s) is also illustrated, providing an estimate for the baseline neutrality level (average neutrality 0.83 ± 0.02). (*b*) The distribution of miRNA neutrality ranks, *r*. Each miRNA stem–loop neutrality value is ranked among the neutrality value of 1,000 matching inversely folded stem–loops. High ranking miRNA genes ($r \leq 50$) correspond to significantly robust stem–loops.

role in the miRNA gene maturation process (23), the stem–loop structure has been under evolutionary pressures to conserve its structure. Such stabilizing pressures favor robust configurations and may have led to the evolution of robustness. (*ii*) Unlike most genotype-phenotype mappings, the RNA secondary structure is fully tractable through secondary structure prediction algorithms (e.g., ref. 25). These algorithms provide a simple measure of the structural robustness of a given miRNA stem–loop sequence by comparing the predicted structure of this sequence with the predicted structure of all its one-mutant neighbors. And (*iii*) by using inverse folding algorithms (26), a reference set of random RNA sequences with the same structure as the native precursor sequence can be produced, providing a natural background model to evaluate the level of genetic robustness expected by chance (see also refs. 11, 27, and 28). Comparing the structural neutrality of an evolved miRNA stem–loop with that of random sequences that perfectly preserve the stem–loop structural phenotype excludes from the analysis the intrinsic robustness associated with the (potentially more robust) hairpin structure and can confirm that the increase robustness is the product of direct evolution toward robustness.

Results

Examining the neutrality of all miRNA stem–loops included in the analysis, 1,016 of 1,120 (90.7%) were found to be robust ($\eta_m > \eta_c$; i.e., they were more robust than random RNA sequences having an identical secondary structure; see *Materials and Methods*). The increased neutrality of the native miRNA stem–loops was also evident from the different distributions of neutrality values in native vs. inversely folded stem–loops (Fig. 1*a*). Although the difference between the mean neutralities of native and inversely folded sequences was relatively small (0.96 vs. 0.93 respectively), the two distributions were significantly different ($p < 10^{-137}$; Wilcoxon's signed rank test for paired data), corroborating the hypothesis that evolved miRNA stem–loops are more robust than expected by chance. Examining the robustness of miRNA genes within species (Table 1) provided a similar picture for each species separately. The level of the increased neutrality of each miRNA was evaluated by its neutrality rank, *r* (i.e., the rank of the native stem–loop neutrality value among the neutrality values of 1,000 inversely folded sequences; see *Materials and Methods*). Examining the distribution of ranks across all miRNA genes, a marked propensity toward high ranks (and hence, increased neutrality) was demonstrated (Fig. 1*b*). A total of 185 genes were found to be significantly robust ($r \leq 50$), whereas only 15 genes are significantly nonrobust ($r \geq 951$). The high number of significantly robust genes and low number of significantly nonrobust genes further support the robustness of evolved miRNA stem–loops ($p < 10^{-45}$ and $p < 10^{-10}$, respectively). See also Table 2 for a list of the most robust genes.

Considering the slow, second-order dynamics leading to the evolution of adaptive robustness (14), we quantified the ratio between the adaptive robustness of miRNAs (measured by the difference between the neutrality of miRNA stem–loops and inversely folded stem–loops) and its intrinsic robustness (the

Table 1. Robustness analysis of miRNA stem–loop structures within each taxa

Species	No. of miRNAs	No. of robust miRNAs (%)	*p*	$\overline{\eta_m}$	$\overline{\eta_c}$
C. elegans	115	105 (91.3)	3.5×10^{-15}	0.960	0.922
C. briggsae	76	72 (94.7)	7.9×10^{-12}	0.967	0.928
D. melanogaster	78	73 (93.6)	3.3×10^{-11}	0.962	0.930
D. pseudoobscura	73	67 (91.8)	3.2×10^{-11}	0.965	0.932
D. rerio	30	25 (83.3)	1.1×10^{-03}	0.947	0.928
G. gallus	120	115 (95.8)	6.4×10^{-20}	0.968	0.941
M. musculus	223	198 (88.8)	2.2×10^{-27}	0.963	0.939
R. norvegicus	185	167 (90.3)	1.8×10^{-22}	0.960	0.937
H. sapiens	220	194 (88.2)	7.0×10^{-29}	0.963	0.938

p values denote the probability of observing the increased neutrality found in each species by chance and are calculated with the Wilcoxon signed rank test for paired data.

Table 2. A list of the most significantly robust miRNA genes

miRNA genes	η_m	r	p
cel-mir-357	0.989	1	0.001
cel-mir-86	0.984	1	0.001
cel-mir-56	0.977	1	0.001
cbr-mir-239a	0.985	2	0.002
dre-mir-181b-2	0.985	2	0.002
dme-mir-125	0.978	2	0.002
dps-mir-13b-2	0.982	3	0.003
dps-mir-9b	0.986	3	0.003
gga-mir-181b-2	0.987	3	0.003
rno-mir-31	0.979	3	0.003
cbr-mir-250	0.979	3	0.003
cel-mir-252	0.981	3	0.003
dre-mir-213	0.979	4	0.004
dps-mir-125	0.978	4	0.004
cbr-mir-34	0.972	4	0.004
mmu-mir-24-1	0.981	4	0.004
dps-mir-276b	0.976	5	0.005
rno-mir-213	0.981	5	0.005
cbr-mir-252	0.978	5	0.005
hsa-mir-132	0.984	5	0.005
hsa-mir-213	0.978	5	0.005
hsa-mir-147	0.976	5	0.005
mmu-mir-126	0.983	5	0.005
cel-mir-358	0.978	6	0.006
mmu-mir-181b-1	0.987	6	0.006
dme-mir-275	0.986	6	0.006
cel-mir-253	0.980	6	0.006
cel-mir-34	0.974	6	0.006

See Data Set 1, which is published as supporting information on the PNAS web site, for the full list.

robustness associated with the specific secondary structure of miRNAs). To this end, we measured the neutrality of shuffled miRNA sequences (see *Materials and Methods*). The difference between the neutrality of this set and the average neutrality of the inversely folded sequences provided an estimate for the neutrality arising from the miRNA-specific structure. The values obtained were 0.03 ± 0.04 for the adaptive robustness (i.e., $\eta_m - \eta_c$) and 0.11 ± 0.04 for the intrinsic robustness ($\eta_c - \eta_s$). Hence, although most of the neutrality observed in miRNAs could be attributed to the intrinsic robustness of hairpin structures, the adaptive robustness was still of the same order of magnitude. Interestingly, the adaptive robustness ($\eta_m - \eta_c$) was negatively correlated (Pearson's correlation coefficient, -0.573; $p < 10^{-99}$) with the structure intrinsic robustness ($\eta_c - \eta_s$), suggesting that increased selective pressure for robustness is exerted by evolution to compensate for miRNAs with relatively low structural intrinsic robustness.

To demonstrate the difference between robust miRNAs and the corresponding inversely folded stem–loops, the one mutant neighborhood of a specific significantly robust miRNA gene (cel-mir-357) was examined in detail (Fig. 2). Evidently, the secondary structures' repertoire that can be found in the genetic neighborhood of the native stem–loop consisted entirely of hairpin structures with few differences from the original hairpin structure (Fig. 2*a*). The genetic neighborhood of a corresponding inversely folded stem–loop (having an identical secondary structure) included a variety of numerous structures noticeably different from the original stem–loop (Fig. 2*b*). Considering the crucial role of the hairpin structure in the miRNA gene biogenesis, the different composition of the two secondary structure ensembles exemplifies the adaptive value of the capacity to buffer single-point mutations.

To further corroborate that the observed excess robustness indeed stems from direct evolution toward robustness and not from other origins, two additional alternative hypotheses were examined. First, because the mononucleotide and dinucleotide frequencies of an RNA sequence (which are not preserved in the reference set) affect the physical stability of the secondary structure (29), it is important to verify that the increased neutrality observed is not a byproduct of a bias in the base composition of the native stem–loops compared with the sequences generated by the inverse folding algorithm. To this end, we examined a subset of miRNA genes from four taxa, comparing the neutrality of the miRNA stem–loop with the neutrality of a new reference set that preserves not only the secondary structure but also the dinucleotide base composition (see *Materials and Methods*). It was shown (Fig. 3*a*) that the native sequences were significantly robust also in comparison to this reference set ($p < 10^{-30}$; Wilcoxon's signed rank test for paired data). An additional set of experiments, using an indirect (but computationally more tractable) approach to examine the potential effect of base composition on neutrality, was applied to all miRNAs in our analysis, further corroborating that the observed robustness of miRNA cannot be attributed to miRNAs' base composition bias (see *Supporting Text*, which is published as supporting information on the PNAS web site).

A second alternative hypothesis may posit that the increased mutational robustness described above arises from the increased thermodynamic stability of miRNAs, reported recently by Bonnet *et al.* (18), due to the correlation between the thermodynamic stability of RNA secondary structure and its neutrality (28). Accordingly, genetic robustness has evolved as a correlated side effect of environmental robustness. Such congruent evolution of robustness was demonstrated in a computer simulation of RNA evolution (11), wherein genetic canalization occurs concurrently with environmental canalization. However, reexamining the thermodynamic stability of miRNAs in an analogous manner to Bonnet *et al.* (18) but using a background model based on inversely folded sequences rather than the shuffled sequences used in the original study, we found that most of the statistical effect originally demonstrated vanishes (Fig. 3*b* and compare with figure 2 in ref. 18). Evidently, the significantly low folding energies originally found can be attributed mostly to the overall stability of the hairpin structure and are not a specific feature of miRNA precursors. Although the genetic robustness of miRNAs was still weakly correlated with their thermodynamic stability under this background model (Pearson's correlation coefficient, 0.24), most of the significantly robust genes (86%) did not exhibit significant thermodynamic stability.

Finally, to study the relationship between the functional importance of miRNA genes and robustness, we used miRNA family annotation to detect robustness-enriched miRNA families (see *Materials and Methods*). Families *mir-7*, *mir-135*, and *mir-10* were found to be robustness-enriched ($p < 0.003$, $p < 0.009$, and $p < 0.003$, respectively). Unfortunately, because of the current lack of information concerning miRNA functionality, we could not verify that these miRNA families were indeed participating in essential processes.

Discussion

Focusing on the mutational robustness of the miRNA stem–loop structure and addressing the longstanding debate concerning the evolutionary forces underlying genetic robustness, this study provides evidence for direct evolution of increased robustness in miRNAs. Although it is difficult to completely rule out alternative hypotheses, the carefully controlled background model that preserves the most relevant phenotype (namely, the secondary structure) and the statistical analysis of the results strongly support the hypothesis of direct evolution of robustness. Although it was shown that most of the mutational robustness of miRNA stem–loops results from the intrinsic mutational stability of hairpin structures,

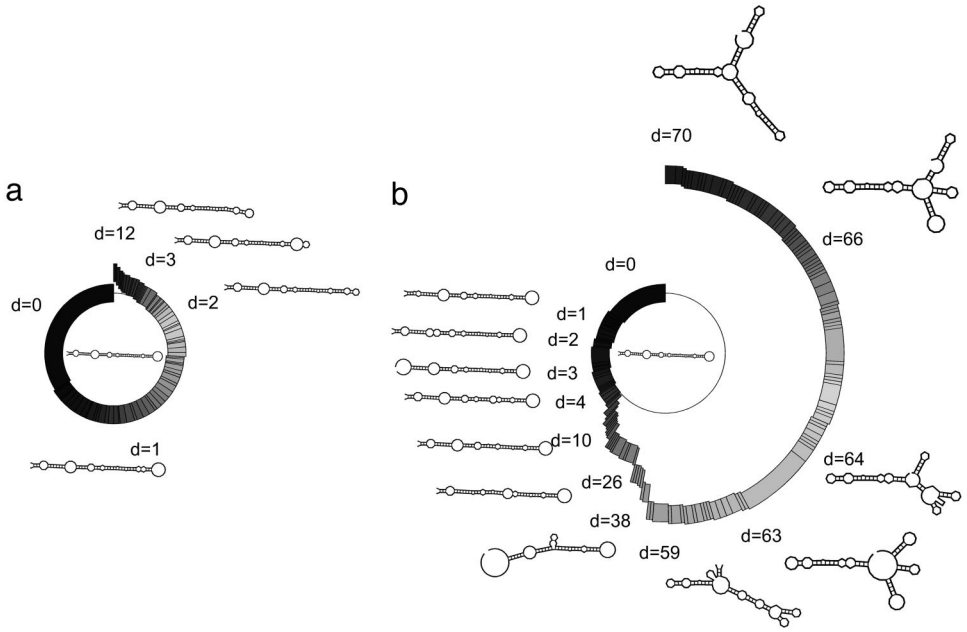

Fig. 2. The one-mutant neighborhood structure composition for a robust native stem–loop (cel-MIR-357) and for a corresponding inversely folded reference stem–loop. The entire collection of structures that compose the one-mutant neighborhood is represented on the circle's circumference. Each gray arc segment corresponds to one of the structures that can be found in the collection. The length of the arc segment represents the frequency of this structure among the $3L$ ($L = 110$, in this case) mutants. The segment's distance from the thin black circle corresponds to the structure's base pair distance, d, from the original structure (illustrated in the center of the circle). Some of the structures in the collection are also illustrated next to the appropriate distance value. (*a*) The one-mutant neighborhood of cel-MIR-357 includes 80 different structures. Three hundred and twenty-six of 330 mutants have a base-pair distance <10. The average distance is 1.248. (*b*) The one-mutant neighborhood of an identical, inversely folded, reference stem–loop includes 146 different structures. Only 123 mutants have a base-pair distance <10. The average distance is 37.076.

the significant excess neutrality found in comparison to random, inversely folded, identical hairpin structures can only be the product of direct evolution toward robustness. Furthermore, the negative correlation between intrinsic and adaptive robustness found in our analysis exemplifies intriguing evolutionary dynamics wherein adaptive robustness evolves to compensate for stem–loop structures with low intrinsic robustness.

One potential caveat of our findings stems from the assumption that each RNA molecule maps into a fixed structure predicted by the RNA minimum free-energy folding algorithm. This simple-minded static representation of molecules is necessary for the computational derivation of neutrality values in our analysis. In reality, however, molecules may confer function by a dynamical, local reconfiguration of structure (e.g., refs. 30 and 31). The excess robustness observed in this study may be, at least in part, the outcome of an evolutionary pressure toward a functioning set of vibrational modes. We believe that such molecule vibrational dynamics pose an exciting and profound challenge to further research concerning the detection, evolution, and definition of mutational robustness.

Many miRNA (and other, functional noncoding RNAs) prediction algorithms are partially based on detecting conserved second-

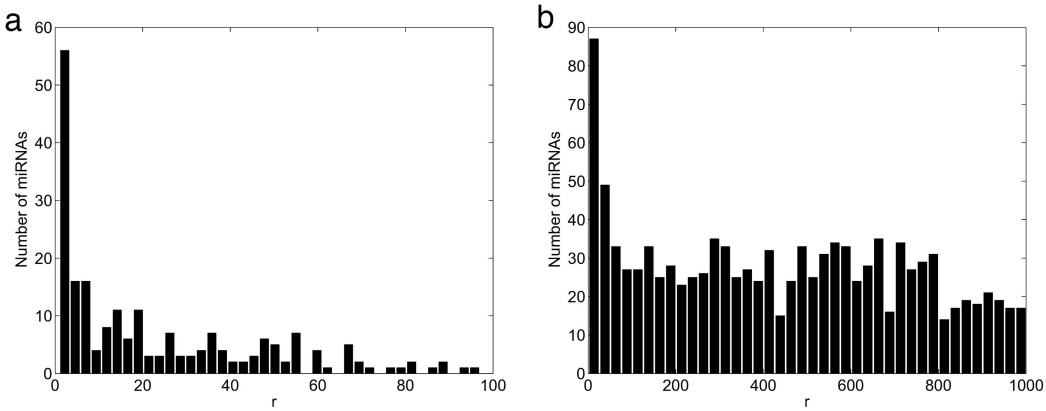

Fig. 3. Ruling out alternative hypotheses for the origins of genetic robustness. (*a*) The distribution of the neutrality rank, r, within a reference set that preserves also the dinucleotide base composition. As before, high ranks correspond to significantly robust genes. Of 211 genes that have been analyzed here, 194 are robust ($\eta_m > \eta_c$) and 72 are significantly robust ($r \leq 50$), indicating that the observed robustness of miRNAs cannot be attributed to base composition bias. (*b*) The distribution of the rank, r, for thermodynamic stability (as manifested by a lower than chance minimum folding energy) under the inversely folded background model. In contrast to the results presented by Bonnet *et al.* (18) where $>90\%$ of the genes have significantly increase thermodynamic stability ($P < 0.05$), applying a background model that preserves the secondary structure results in only 12% of all miRNA genes with significantly increased thermodynamic stability levels.

ary structure regions across multiple aligned genomes (e.g., see refs. 32–37). Significantly increased neutrality may also facilitate improved miRNA prediction on a single genome. Although the signal produced by increased neutrality calculation is probably not sufficient by itself to predict miRNA stem–loops, it can serve well as a complementary method to filter out random hairpin structures.

Finally, as miRNA stem–loop's secondary structure embodies many of the properties controlling molecular evolution, it forms a promising framework for studying central issues concerning the evolution of robustness. To some extent, the increased neutrality of miRNAs examined in this study can be conceived as first-order robustness, based only on the gene sequence and the physical properties that map sequence to structure. The simplicity of this form of robustness and the full tractability of RNA secondary structure facilitate the analysis of its evolutionary origins. Protein structures, being the next step up in complexity, may possess a similar tendency for sequence-based increased neutrality as well as additional mechanisms (e.g., molecular chaperones) contributing to their functional robustness. The effect of these mechanisms on the selection pressures toward sequence-based robustness is especially interesting. Once a better understanding and accurate prediction algorithms for protein folding are available, our methodology for robustness analysis can be applied to confirm that the observed protein-structure tolerance to amino acid substitutions (2) is higher than expected by chance. Other forms of robustness may be the product of second-order mechanisms, actively promoting robustness through alternative developmental pathways, error-repair, or phenotypic plasticity. The direct selection pressures toward increased robustness exemplified in this study may account, at least in part, for the abundance of biologically robust systems and the rich variety of mechanisms that underlie them.

Materials and Methods

miRNA Data and RNA Folding. miRNA precursor sequences were downloaded from the miRNA registry database, release 5.1 (38). All available miRNA genes of nine Metazoa species (*Caenorhabditis elegans*, *Caenorhabditis briggsae*, *Drosophila melanogaster*, *Drosophila pseudoobscura*, *Danio rerio*, *Gallus gallus*, *Mus musculus*, *Rattus norvegicus*, and *Homo sapiens*) were used. A small number of genes (9) were excluded from the analysis because of repeated failures of the inverse folding algorithm to produce random sequences with the same structure. Overall, a total of 1,120 miRNAs from the above taxa are included in the analysis. miRNA families annotation is based on the Rfam database (39). Secondary structures were determined with Zuker's minimum free-energy algorithm (25), implemented by the VIENNA RNA package, version 1.4 (26), which also supports sequence generation through inverse folding. A computer cluster of ≈100 Linux machines was used to perform the intensive computation.

Neutrality Calculation. The neutrality, η, of an RNA sequence of length, L, is defined as $\eta = \langle (L - d)/L \rangle$, where d is the base-pair distance between the secondary structure of the original sequence and the secondary structure of the mutant, averaged over all $3L$ one-mutant neighbors. The base-pair distance measures the dissimilarity between two structures as the number of closed pairs that are present in only one of the two structures (i.e., the number of pairs that should be opened or closed to transform one structure to the other) (28). η thus represents the average fraction of the structure that remains unchanged after a mutation occurs. An alternative and more strict definition of neutrality, one that does not assume any structure distance metric, has also been examined, wherein η is simply defined as the fraction of the $3L$ one-mutant neighbors that perfectly preserves the original structure. Applying this measure of neutrality yields qualitatively similar results.

Robustness and Significance Analysis. For each miRNA, we measure the neutrality of the native stem–loop, η_m, and evaluate the neutrality of this stem–loop structure expected by chance, η_c, as the average neutrality of 1,000 "random" sequences that fold into an identical stem–loop structure (generated by the inverse folding algorithm). Note that although all of the miRNA precursors fold into a hairpin shape structure, the exact secondary structure is specific to each miRNA. An inversely folded random reference set is thus generated separately for each miRNA to perfectly retain the structural phenotype of the control. A miRNA gene is defined as robust if $\eta_m > \eta_c$. To determine whether the increased robustness observed for a group of genes (e.g., a specific species or miRNA family) is statistically significant, the magnitude and sign of the differences between the paired values η_m and η_c obtained for these genes are analyzed by the nonparametric Wilcoxon signed rank test for paired data. To further evaluate the level of the increased neutrality for each miRNA separately, the rank of the native stem–loop neutrality, r, among the neutrality of the 1,000 inversely folded sequences is also calculated. Such an order statistics measure does not entail any assumption on the nature of neutrality value distribution. The significance level of each miRNA robustness is then defined as $p = \frac{r}{1,001}$, providing a good estimate for the probability of observing an equal or higher neutrality value by chance (see also ref. 18). miRNAs for which $r \leq 50$ (corresponding to $p < 0.05$) are labeled as significantly robust. The number of significantly robust genes is evaluated in comparison with a null hypothesis of uniformly distributed ranks through a binomial cumulative distribution function.

miRNA Hairpin Structure Intrinsic Neutrality. To quantify the level of the miRNA neutrality that stems intrinsically from the miRNA hairpin structure (as opposed to neutrality resulting from direct selection toward robustness measured above), an additional reference set is produced for each stem–loop. This set consists of 100 shuffled sequences produced by the dinucleotide shuffle algorithm (15), maintaining the miRNA stem–loop mononucleotide and dinucleotide frequencies. The average neutrality of this set is used as a baseline neutrality value to evaluate the level of excess neutrality associated with the miRNAs hairpin structures.

Testing for Base Composition Bias Effects. To rule out the effect of base composition bias on neutrality, we compare the neutrality of the native stem–loops with that of inversely folded stem–loops that also maintain a similar base composition. To this end, inversely folded sequences are produced and filtered, leaving only those sequences with a similar dinucleotide base composition. The distance between the base composition of the original sequence and that of the inversely folded sequence is measured by Jensen–Shannon divergence (40), a symmetric version of Kullback–Leibler distance (41). Only inversely folded sequences for which Jensen–Shannon divergence <0.01 are included in this reference set. Because of the high computational costs associated with the filtering process, we have applied this method to a subset containing 211 miRNAs from four taxa (*Caenorhabditis elegans*, *Caenorhabditis briggsae*, *Drosophila melanogaster*, and *Homo sapiens*), generating a reference set of 100 sequences for each gene.

Robustness–Enriched miRNA Families. For each miRNA family including ≥10 genes in our analysis, the average rank of the family members is compared with the average rank obtained for 100,000 random groups comprising of the same number of genes selected uniformly from the database. miRNA families for which <0.01 of the random groups obtained a higher average rank are defined as robustness enriched families.

We thank David Krakauer, Ranit Aharonov, Roded Sharan, Mark Feldman, and Tal Pupko for their comments. E.B. is supported by the Yeshaya Horowitz Association through the Center for Complexity Science and E.R. is supported by the Tauber Fund and the Center for Complexity Science.

1. de Visser, J. A. G. M., Hermisson, J., Wagner, G. P., Ancel Meyers, L., Bagheri-Chaichian, H., Blanchard, J. L., Chao, L., Cheverud, J. M., Elena, S. F., Fontana, W., *et al.* (2003) *Evolution* **57,** 1959–1972.
2. Bowie, J. U., Reidhaar-Olson, J. F., Lim, W. A. & Sauer, R. T. (1990) *Science* **247,** 1306–1310.
3. Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B., *et al.* (2002) *Nature* **418,** 387–391.
4. Albert, R., Jeong, H. & Barabasi, A. L. (2000) *Nature* **406,** 378–382.
5. Krakauer, D. C. & Plotkin, J. B. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 1405–1409.
6. Wagner, A. (200) *Nature Genet.* **24,** 355–361.
7. Gu, Z., Steinmetz, L. M., Gu, X., Scharfe, C., Davis, R. W. & Li, W. H. (2003) *Nature* **421,** 63–66.
8. Bourguet, D. (1999) *Heredity* **83,** 1–4.
9. Schuster, P., Fontana, W., Stadler, P. F. & Hofacker, I. L. (1994) *Proc. R. Soc. London Ser. B* **255,** 279–284.
10. van Nimwegen, E., Crutchfield, J. P. & Huynen, M. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 9716–9720.
11. Ancel, L. W. & Fontana, W. (2000) *J Exp. Zool.* **288,** 242–283.
12. Wilke, C. O., Wang, J., Ofria, C., Adami, C. & Lenski, R. E. (2001) *Nature* **412,** 331–333.
13. Montville, R., Froissart, R., Remold, S. K., Tenaillon, O. & Turner, P. E. (2005) *PLoS Biology* **3,** 1939–1945.
14. Wagner, A. & Stadler, P. F. (1999) *J Exp. Zool.* **285,** 119–127.
15. Workman, C. & Krogh, A. (1999) *Nucleic Acids Res.* **27,** 4816–4822.
16. Rivas, E. & Eddy, S. R. (2000) *Bioinformatics* **16,** 583–605.
17. Katz, L. & Burge, C. B. (2003) *Genome Res.* **13,** 2042–2051.
18. Bonnet, E., Wuyts, J., Rouze, P. & Van de Peer, Y. (2004) *Bioinformatics* **20,** 2911–2917.
19. Lagos-Quintana, M., Rauhut, R., Lendeckel, W. & Tuschl, T. (2001) *Science* **294,** 853–858.
20. Lau, N. C., Lim, L. P., Weinstein, E. G. & Bartel, D. P. (2001) *Science* **294,** 858–862.
21. Lee, R. C. & Ambros, V. (2001) *Science* **294,** 862–864.
22. Bartel, D. P. (2004) *Cell* **116,** 281–297.
23. Lee, R. C., Feinbaum, R. L. & Ambros, V. (1993) *Cell* **75,** 843–854.
24. Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., *et al.* (2003) *Nature* **425,** 415–419.
25. Zuker, M. & Stiegler, P. (1981) *Nucleic Acids Res.* **9,** 133–148.
26. Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, S., Tacker, M. & Schuster, P. (1994) *Monatsh. Chem.* **125,** 167–188.
27. Huynen, M. A., Stadler, P. F. & Fontana, W. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 397–401.
28. Wuchty, S., Fontana, W., Hofacker, I. L. & Schuster, P. (1999) *Biopolymers* **49,** 145–165.
29. Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T. & Turner, D. H. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 9373–9377.
30. Hou, T., Chen, K., McLaughlin, W. A., Lu, B. & Wang, W. (2006) *PLoS Comput. Biol.* **2,** 46–55.
31. Rischel, C., Spiedel, D., Ridge, J. P., Jones, M. R., Breton, J., Lambry, J. C., Martin, J. L. & Vos, M. H. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 12306–12311.
32. Rivas, E. & Eddy, S. R. (2001) *BMC Bioinformatics* **2,** 8.
33. Rivas, E., Klein, R. J., Jones, T. A. & Eddy, S. R. (2001) *Curr. Biol.* **11,** 1369–1373.
34. McCutcheon, J. P. & Eddy, S. R. (2003) *Nucleic Acids Res.* **31,** 4119–4128.
35. di Bernardo, D., Down, T. & Hubbard, T. (2003) *Bioinformatics* **19,** 1606–1611.
36. Pavesi, G., Mauri, G., Stefani, M. & Pesole, G. (2004) *Nucleic Acids Res.* **32,** 3258–3269.
37. Washietl, S., Hofacker, I. L. & Stadler, P. F. (2005) *Proc. Natl. Acad. Sci. USA* **102,** 2454–2459.
38. Griffiths-Jones, S. (2004) *Nucleic Acids Res.* **32,** D109–D111.
39. Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A. & Eddy, S. R. (2003) *Nucleic Acids Res.* **31,** 439–441.
40. Kullback, S. & Leibler, R. A. (1951) *Ann. Math. Stat.* **22,** 79–86.
41. Lin, J. (1991) *IEEE Trans. Info. Theory* **37,** 145–151.