Mutational Robustness Accelerates the Origin of Novel RNA Phenotypes through Phenotypic Plasticity

Andreas Wagner[†‡§*]

[†]Institute of Evolutionary Biology and Environmental Sciences, University of Zurich, CH-8057 Zurich, Switzerland; [‡]The Swiss Institute of Bioinformatics, Bioinformatics, Quartier Sorge, Batiment Genopode, 1015 Lausanne, Switzerland; and [§]The Santa Fe Institute, Santa Fe, NM

ABSTRACT   Novel phenotypes can originate either through mutations in existing genotypes or through phenotypic plasticity, the ability of one genotype to form multiple phenotypes. From molecules to organisms, plasticity is a ubiquitous feature of life, and a potential source of exaptations, adaptive traits that originated for nonadaptive reasons. Another ubiquitous feature is robustness to mutations, although it is unknown whether such robustness helps or hinders the origin of new phenotypes through plasticity. RNA is ideal to address this question, because it shows extensive plasticity in its secondary structure phenotypes, a consequence of their continual folding and unfolding, and these phenotypes have important biological functions. Moreover, RNA is to some extent robust to mutations. This robustness structures RNA genotype space into myriad connected networks of genotypes with the same phenotype, and it influences the dynamics of evolving populations on a genotype network. In this study I show that both effects help accelerate the exploration of novel phenotypes through plasticity. My observations are based on many RNA molecules sampled at random from RNA sequence space, and on 30 biological RNA molecules. They are thus not only a generic feature of RNA sequence space but are relevant for the molecular evolution of biological RNA.

INTRODUCTION

Evolutionary adaptations and innovations are new traits (phenotypes) that help organisms survive and reproduce. They can have two principal evolutionary origins. One of them requires DNA mutations, changes in a genotype that bring forth a new phenotype. The other relies on phenotypic plasticity. Phenotypic plasticity is the ability of a biological system with a given genotype to adopt multiple phenotypes. Such plasticity is widespread (1) and exists in whole organisms, from amphibians that undergo metamorphosis, to casts in social insects, and plants with changing leaf shapes, all the way down to molecules, where the shapes of protein and RNA molecules fluctuate incessantly between different conformations. What causes such plasticity is the environment, broadly defined, which includes the nutrients that an organism is exposed to—food, for example, helps determine casts in social insects—to the molecules inside a cell, whose thermal motion drive any one protein's continual shape change. Because phenotypic plasticity is so ubiquitous, the environment and its change play a critical role in the origin of novel phenotypes. Although one could thus argue that the environment's role is even more important than that of the genotype, the two roles are inseparably intertwined (1). Absent the right genotype, no environment will allow an organism to bring forth certain phenotypes, and vice versa.

The phenotypic plasticity of molecules plays important roles in cellular life. A prominent example involves RNA molecules called riboswitches, which regulate the biosynthesis of small molecules, such as vitamins and amino acids (2,3). In addition to conformational motions, which other RNA molecules also undergo, these molecules can switch between alternative secondary structures—planar shapes caused by internal base pairing—through the binding of a small regulatory molecule. One of these conformations may allow transcription of a gene or translation of its mRNA, whereas the other conformation prevents it. Another example comes from enzymes that are catalytically promiscuous. In addition to their primary reaction, such enzymes catalyze several side reactions. A case in point is the cytochrome P450 protein family, whose members can hydroxylate many different chemicals; chymotrypsin, a digestive enzyme that can cleave many kinds of proteins; or bovine carbonic anhydrase, which interconverts carbon dioxide and bicarbonate ions but can also cleave highly toxic organophosphates (4–6). In the right environment, any one side reaction may become important to the survival of an organism and may thus become the target of natural selection to improve its efficiency through genetic change. Phenotypic plasticity is thus also a source of *exaptations* or *preadaptations*—adaptive traits whose original role in the life of an organism was different from their current role (7). Examples of such exaptations range from the macroscopic—the feathers of birds, the wings of (flightless) penguins—to the molecular, such as a subunit of the lactose synthetase that synthesizes the lactose in mammalian milk but is derived from lysozyme, a protein that kills bacteria (8,9).

Mutational robustness is the ability of a biological system's phenotype to persist in the face of DNA mutations. Biological systems on all levels of organization, from whole organisms to molecules, are to some extent robust to

CrossMark

mutations. One would think that mutational robustness, by its very definition, could only hinder the ability of mutations to bring forth novel adaptive phenotypes. However, with possible exceptions (10), this is not generally the case, especially for molecular phenotypes (11–16). For example, in laboratory evolution experiments, robust proteins evolve new catalytic activities more easily (14), and proteins with robust folds have evolved a greater diversity of catalytic functions than other proteins (16). Robustness is both a necessary and sufficient condition for the existence of genotype networks (17) or neutral networks (18,19), vast connected sets of genotypes with the same phenotypes. Such networks are ultimately the reason why robustness can facilitate the origins of new phenotypes through mutations, as has been discussed elsewhere (20).

It is much better understood how robustness can affect the origin of new phenotypes through *mutation* than through *plasticity* (10,21). In this study I investigate whether mutational robustness facilitates or hinders this origin, using RNA molecules and their secondary structure phenotypes. RNA is ideal for this purpose because, first, it regulates many cellular processes through molecules such as small interfering RNAs, guide RNAs involved in RNA editing, and small nuclear RNAS involved in splicing (22–26). Second, in many biological RNA molecules, the secondary structure is itself a functionally important phenotype and thus worthy of study. Many examples come from RNA viruses, whose life cycle is regulated by RNA secondary structure elements (27–31), and secondary structures in mRNA molecules that regulate half-life and translation (32–35). Third, efficient algorithms exist to predict the *native* or minimum-free energy secondary structure phenotype P_{mfe}, of an RNA molecule (36), and the spectrum of *alternative* phenotypes that such a molecule can adopt through thermal motion (37). The latter indicates an RNA molecule's phenotypic plasticity, a central focus of this paper. With such algorithms, one is not restricted to study the relationship between robustness and plasticity for one or few biological RNA molecules, but one can sample thousands of sequences from RNA sequence space, examine their phenotypic plasticity, and thus arrive at conclusions that hold generically for most RNAs.

The mutational robustness of an RNA genotype is the fraction of the molecule's neighbors—molecules that differ from it in a single nucleotide—with the same *native* phenotype P_{mfe} as the molecule itself. Most RNA molecules are to some extent robust, and this robustness has two consequences. The first, *structural* consequence, which I have already mentioned, is that RNA genotypes with the same native phenotype are organized into genotype networks (17–19). The second, *dynamic* consequence affects how populations evolve on an RNA genotype network (11). In this study I show that both consequences of robustness facilitate the origin of novel phenotypes through phenotypic plasticity for a broad range of RNA molecules.

METHODS

Folding, inverse folding, and thermodynamic ensemble calculations

Mediated by thermal motion, any one RNA molecule continually changes its secondary structure phenotype P through formation and dissociation of internal base pairs. In thermodynamic equilibrium, the molecule will spend a fraction of time $\exp(-\Delta G_P/kT)/Z$ in phenotype P, where ΔG_P is the free energy of the phenotype, k is the Boltzmann constant, T is the absolute temperature ($kT \approx 0.62$ kcal/mol at 310.15 Kelvin or 37°C), and Z is the partition function, the normalization factor $\sum_P \exp(-\Delta G_P/kT)$, where summation is over all possible secondary structures P. Of special importance is the phenotype P_{mfe} with the smallest or minimum free energy (mfe) among all secondary structure, because the molecule spends the relative majority of its time in this native phenotype. To compute the mfe secondary structure of an RNA molecule, I used the routine fold from the Vienna RNA package (http://www.tbi.univie.ac.at/~ivo/RNA/, (36)). To determine arbitrary sequences that fold into a given structure, I used the program *inverse_fold* from the same package. Specifically, this program starts with a randomly chosen sequence and performs a guided random walk through sequence space to find a sequence with the desired structure (36). *Inverse_fold* effectively samples the space of sequences folding into a given structure at random or nearly so (18,39). To compute all secondary structures within a given energy interval E above the mfe of a given sequence, I used the routine subopt (37,40). In all routines I used default parameters, in particular a folding temperature of 37°C. To compute the structural diversity of all sequences in the plastic repertoire of a given molecule, I computed the pairwise Hamming distance of the dot-parenthesis representation of the structures, which increases by one for every base that is paired in one structure but unpaired in the other.

Choice of structures

I used two classes of RNA sequences and structures in this analysis. The first are *biological* molecules (Table S1), all of which are taken from the functional RNA database fRNAdb (http://www.ncrna.org/frnadb, (41)), and have been used in a previous study to estimate genotype network sizes (42). Their choice is arbitrary, except that I chose only molecules of length up to $L = 43$ to make the computational analysis herein feasible. I also used *random* RNA molecules of varying length L, which I generated as follows. I first created a (pseudo)random nucleotide string of length L, in which each of the four nucleotides occurs with equal likelihood at each position. I then computed the mfe structure of this molecule. If the molecule folded into some secondary structure, I kept it for further analysis, otherwise I created another random nucleotide string and repeated the procedure until I had found a string that folded. This iterative procedure is necessary, because a substantial fraction of short random RNA strings do not fold.

Neighborhoods in sequence space

Consider an RNA sequence (genotype) G and its mfe phenotype P. The k-mutant neighborhood, or k-neighborhood of G comprises all sequences that differ in no more than k nucleotides from G. I define the *neutral k-neighborhood* as comprising all sequences that differ in no more than k nucleotides from G, *and* that have the same native phenotype P as G. I studied two different aspects of these neighborhoods. First, I studied, for neutral neighborhoods up to $k = 2$, the number of different RNA secondary structures that occur in the plastic repertoire of sequences in the neighborhood. To this end, I determined all k-mutant neutral neighbors of G, as well as the phenotypes in the plastic repertoire of each neighbor. I then counted all unique phenotypes in the union of all these plastic repertoires. That is, if

two phenotypes occurred in the plastic repertoire of two sequences, I counted them only once. Second, I studied for neighborhoods up to $k = 3$, the number of different native (mfe) phenotypes that occur in this neighborhood. To this end I created computationally all sequences in the k-neighborhood of G, determined their native phenotype, and counted the number of unique phenotypes in the resulting set of phenotypes. Analyses of larger neighborhoods were computationally infeasible, because neighborhood size increases exponentially with k.

Random RNAs with varying genotype set sizes

The total number of RNA sequences with a given native phenotype is the neutral set or genotype set of this phenotype (18,43). It is straightforward to generate a sample of phenotypes with different genotype set sizes, but much more difficult to estimate the genotype set size of a given phenotype (42). If genotype set sizes are to be estimated for many different structures, the following approach based on random sequences is computationally feasible. First, generate a very large sample ($>10^6$) of random RNA sequences and determine their mfe structure. For any one structure, count the number of times M that it occurs in this sample. For structures that occur multiple times in the sample, M divided by the genotype space size, is an estimate of the genotype set size of the structure, whose standard error is inversely proportional to the square root of M. This approach works best for sequences of a modest length (11), because sequences much shorter than $L = 30$ nucleotides adopt too limited a repertoire of structures, and in much longer sequences even a large random sample leads only to unique structures. For instance, a sample of 10^6 random sequences of length $L = 75$ typically contains only structures that occur merely once in the sample. Sequences of modest length ($L = 30$) occupy a middle ground of being structurally diverse, yet being tractable with realistic sample sizes.

Population simulation and random walks on genotype networks

I simulated the evolution of populations on a genotype network through cycles of mutation and "soft selection" as follows. For populations of size $N > 1$, all members of a population initially had the same genotype, which forms some native phenotype P. I subjected each individual to exactly one point mutation, that is, I replaced one nucleotide at a randomly chosen sequence position with a randomly chosen nucleotide. After this mutation step, I determined the phenotype of each mutant sequence and considered those sequences that had changed their native phenotype inviable. I then sampled from the remaining viable sequences exactly N sequences with replacement to form a population for the next generation.

For populations of $N = 1$ individuals, this process reduces to a random walk on a genotype network. This random walk starts from a single sequence with a given phenotype P, mutates the sequence as described, and mutates the mutant sequence's phenotype. If this phenotype is different from P, the mutation is rejected, another mutation is introduced, and so on, until a new genotype is found that has the phenotype P. Once such a sequence has been found, a single step of the random walk is considered completed.

Statistical analyses

All statistical analyses were performed in Statistica (StatSoft Inc., Tulsa, OK, http://www.statsoft.com/). Box plots include information about the nonoutlier range, defined as follows. If l designates the 25th percentile and u the 75th percentile of a distribution, then the nonoutlier range (whiskers in box plots) span the interval $(u + (u - l), l - (u - l))$. Where shown, the extremes of the distribution comprise data points that lie outside the interval $(u + 3(u - l), l - 3(u - l))$.

RESULTS

Genotype networks greatly increase the number of phenotypes accessible through phenotypic plasticity

I define plasticity as the number of all secondary structures that an RNA molecule can adopt within a specific energy interval E above the minimum free energy structure. I also refer to this set as the molecule's *plastic repertoire* (10). E is conveniently given in units of kT, the product of the Boltzmann constant k and the absolute temperature T (see Methods). I then explore energy intervals E between one and five kT units. Much larger energy intervals are prohibitive, because the number of alternative secondary structures increases rapidly with free energy. Moreover, the phenotypes found far above the minimum free energy are also of little biological relevance, because molecules spend vanishing amounts of time in them. As a point of reference, a change in free energy of 3 kT corresponds to ~1.8 kcal/mol, which is the energy involved in the stacking interaction of one adjacent G-C/C-G base pairs, or that of two adjacent A-U/U-A base pairs (44).

For my analysis, two different kinds of information are useful. First, many RNA molecules sampled at random from RNA genotype space can provide information about properties of *typical* RNA molecules. (Exhaustive analysis is infeasible for all but the shortest sequences, because sequence space contains 4^L sequences of length L.) Second, analysis of biological RNA molecules isolated from some organism can show whether such molecules are unusual, in that their properties deviate from those of randomly sampled molecules. In a preliminary analysis, I compared the plasticity of 30 different biological RNA molecules with that of randomly sampled RNA molecules of comparable length. They are similar (see Figs. S1–S4 in the Supporting Material), which suggests that the generic properties of RNA genotype space revealed by random sampling are relevant for the evolution of biological molecules. I juxtapose observations from both kinds of analysis in the next sections.

The more alternative phenotypes a molecule can explore—either through mutation or through plasticity—the greater the chances that it stumbles upon the few new mutants that are not deleterious (45) and that may constitute adaptations. The genotype networks that robustness brings forth could aid such exploration, *if* different genotypes with the same native phenotype produce different alternative secondary structures through plasticity. I now address whether this is the case.

The left-most black bar in Fig. 1 shows the total number of unique (different) alternative phenotypes that can be found in a $E = 3kT$ energy interval above the native phenotype of a guide RNA G ($L = 40$, Table S1, also shown in Fig. S1 a, inset). This number is equal to 65.7 ($\pm$ 27.5 standard errors of the mean [s.e.m.]), averaged over 10
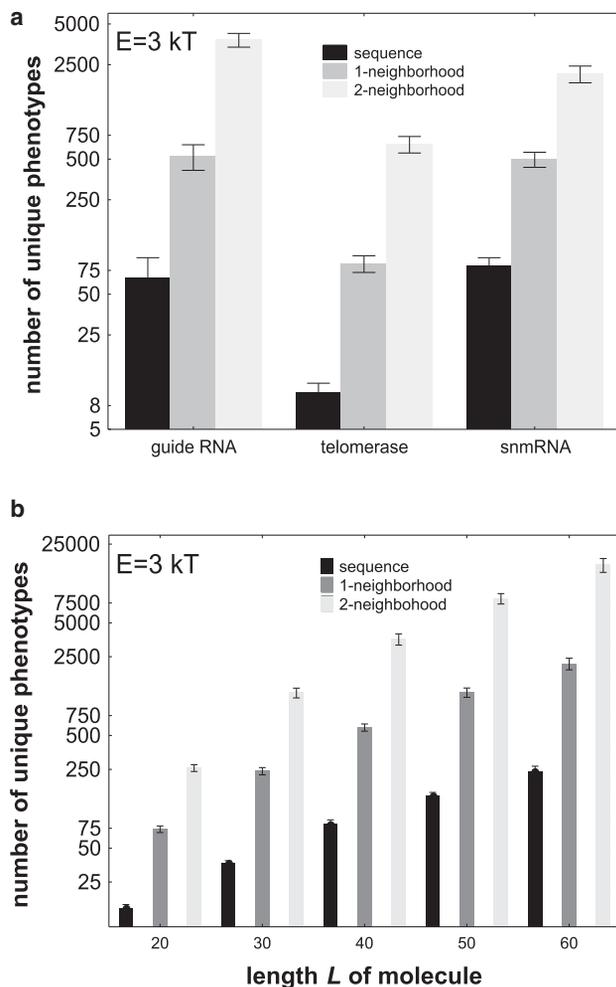
FIGURE 1 Even small neutral neighborhoods of a sequence render many novel phenotypes accessible through plasticity. (*a*) Each group of three bars corresponds to data based on one biological RNA molecule indicated on the horizontal axis. The molecules (Table S1) are part of a guide RNA ($L =$ 40nt; Genbank Acc. No. L25590), a telomerase fragment ($L =$ 33nt; Acc. No. AF061109), and a snmRNA ($L =$ 35nt; Acc. No. AJ430256, all from fRNAdb (41)). Predicted native phenotypes P for these molecules are shown in Table S1. The black bar in each group of bars indicates the mean number of *unique* phenotypes, i.e., phenotypes different from one another, in the plastic repertoire P_0 of an (inversely folded) sequence with the same native phenotype P as the biological molecule. The height of the medium gray (middle) bar indicates the number of unique phenotypes in the union of P_0 and P_1, where P_1 is the number of all different phenotypes in those 1-neighbors of the sequence that have the same native phenotype P as the sequence itself. Light gray bars indicate, analogously, the number of unique phenotypes in the union of P_0, P_1, and P_2, where P_2 denotes the plastic repertoire of all 2-neighbors of the sequence with the same phenotype P as the sequence itself. The height of each bar indicates an average over 10 inversely folded sequences. (*b*) As in (*a*), but for random RNA sequences of varying length (horizontal axis). For each L with the exception of $L = 60$ nucleotides, data are based on $n = 20$ different RNA phenotypes of randomly chosen RNA sequences, and on five inversely folded sequences per phenotype. For length 60, $n = 18$. The height of each bar indicates an average over the $5 \times 20 = 100$ (or $5 \times 18 = 90$) inversely folded sequences. Whiskers indicate one standard error of the mean.

genotypes obtained through inverse folding of P. These molecules have on average 52 1-mutant neighbors with the same mfe structure P. The left-most medium gray bar in Fig. 1 *a*—note the logarithmic vertical axis—shows the number of unique alternative phenotypes that these 1-neighbors can form through plasticity. Here, "unique" means that if a phenotype occurs in the plastic repertoire of two neighbors, it is counted only once. The mean number of unique phenotypes is 525.7 ($\pm$ 112.8 s.e.m.), again averaged over 10 inversely folded mfe sequences with phenotype P and their 1-neighbors. Taken together, the plastic repertoire of molecules in the 1-neighborhood is about eight times greater than the number of phenotypes found in the plastic repertoire of the genotype G itself. Thus, the mere exploration of all 1-mutant neighbors with phenotype P leads to a large increase in the number of different phenotypes that become accessible through plasticity. This increase is even larger if one includes not only 1-neighbors, but also 2-neighbors of G with phenotype P, which renders 3814.4 ($\pm$ 446.1 s.e.m.) (see Fig. 1 *a*, left-most light gray bar) alternative phenotypes accessible through plasticity, more than 58 times as many as in the plastic repertoire of G itself. (The exploration of larger neighborhoods is computationally prohibitive.) Fig. 1 *a* also shows similar results for two further RNA molecules, and Fig. S5 summarizes analogous results for all 30 biological RNA molecules considered in this study. For example, at $E = 3kT$, the number of phenotypes that are accessible through plasticity from a sequence itself, from its neutral 1-neighbors, and from its neutral 1- and 2-neighbors averages to 62.1 ($\pm$ 6.2), 441.7 ($\pm$ 49.5), and 2396 ($\pm$ 319.8) for the 30 biological RNA molecules. Taken together, these observations show that even small neighborhoods of a genotype network can render many novel phenotypes accessible through phenotypic plasticity. More generally, mutational robustness greatly facilitates the exploration of new phenotypes through phenotypic plasticity.

A related question is whether the number of novel phenotypes accessible through plasticity is smaller or greater than those accessible through mutation. For example, one can ask whether the number of novel phenotypes found in the plastic repertoire of a sequence is greater than the number of new phenotypes that can be produced from this sequence through single-point mutations as novel native phenotypes. The answer must depend on the energy interval one considers. It turns out that only for $E = 1kT$ is the number of novel phenotypes accessible through plasticity smaller than those accessible through mutation (Fig. S6 *a*). For $E = 3kT$ and greater, more novel phenotypes are accessible through plasticity (Fig. S6 *a*). Larger neighborhoods lead to the same answers. Specifically, at $E = 3kT$ and higher, more phenotypes are accessible from the neutral 1-neighbors of a sequence through plasticity than through mutations of these 1-neighbors, and analogously for neutral 2-neighbors (Fig. S6 *b* and *c*).

Random RNA sequences give rise to the same observations as biological RNA sequences, as shown in Fig. 1 b for $E = 3kT$ (and in Fig. S7 for different energy intervals E). Note the logarithmic scale on the vertical axis, which indicates that the number of unique phenotypes accessible through plasticity increases approximately exponentially, both with the length of the molecule, and with the size of the neighborhood considered. For example, for RNA molecules of length 40, which lies within the length range of the biological molecules considered in this study, the number of phenotypes that are accessible through plasticity from a sequence itself, from its neutral 1-neighbors, and from its neutral 1- and 2-neighbors averages to 80.0 ($\pm$ 8.1), 589.1 ($\pm$ 43.1), and 3577.3 ($\pm$ 410.9) phenotypes (Fig. 1 b).

In sum, in both biological and random RNA molecules, mutational robustness leads to a massive increase in the number of phenotypes that can be explored through plasticity, even if one just considers the neutral 1-mutant and 2-mutant neighborhoods of genotype G.

The plastic repertoires of distant sequences are very different

Because genotype network sizes are generally very large—for the biological sequences examined in this study, they can exceed 10^{20} genotypes (Table S1, 42)—it is not feasible to examine the total number of phenotypes that all of these genotypes can form through plasticity. However, other approaches can help to understand the phenotypic diversity of genotypes with the same native phenotype. In the first of them, one examines pairs of genotypes G_1 and G_2 drawn at random from the same genotype network, i.e., they have the same native phenotype P. One asks whether the set of alternative phenotypes P_1 and P_2 in the plastic repertoires of each genotype overlap to a great extent. If so, the observation that the plastic repertoire increases greatly when one examines not only a sequence but also its neutral neighbors (Fig. 1) would not extend to genotypes distant from this sequence. Specifically, one can examine the fraction U of unique phenotypes, i.e., phenotypes that occur in P_2 but not in P_1, as indicated by the gray region in the inset of Fig. 2 a. The data in Fig. 2 a is a histogram of U based on 500 sequence pairs obtained through inverse folding of a guide RNA secondary structure (Table S1). Regardless of the energy interval chosen, for the vast majority of sequence pairs, U is greater than 0.8. This means that more than 80 percent of structures that occur in the plastic repertoire of one sequence do not also occur in the repertoire of the other sequence. Fig. S8 shows analogous histograms for a telomerase fragment ($L = 33$) and a snmRNA ($L = 35$; Table S1), and Fig. S9 summarizes data on U for all 30 biological RNA sequences. Fig. 2 b shows that much the same holds for randomly sampled RNA molecules, where the median value of U always exceeds $U = 0.79$. Taken together, these obser-
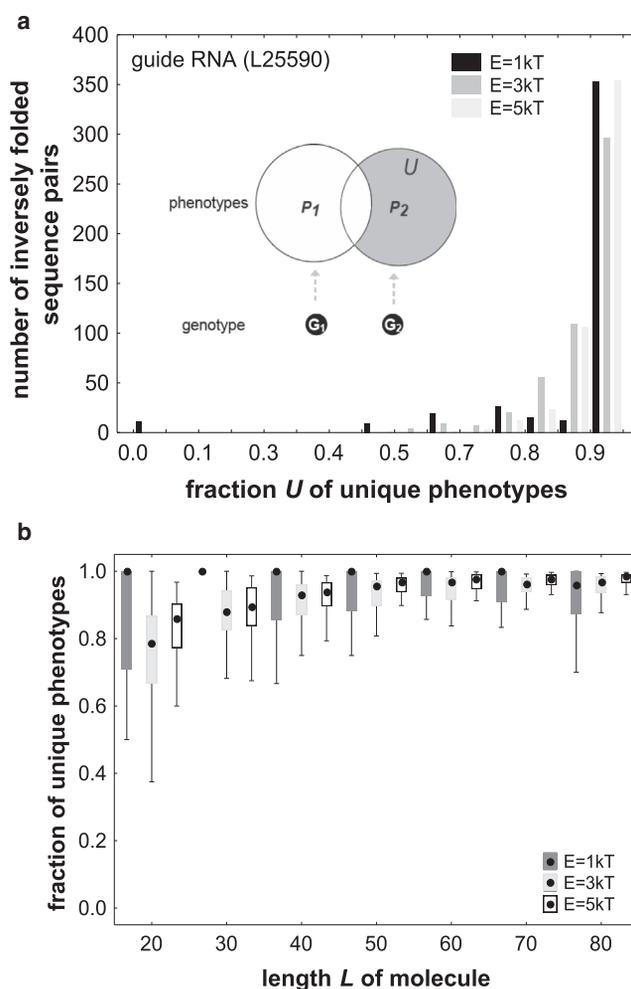
FIGURE 2 Different genotypes with the same native phenotype differ greatly in the phenotypes of their plastic repertoire. The figure shows data about the plastic repertoires of two genotypes G_1 and G_2 with the same native phenotype P. These plastic repertoires are indicated by circles in the inset of panel (a). Specifically, the figure displays the fraction U of phenotypes that are in the plastic repertoire P_2, of G_2, but not in P_1, of G_1, as indicated by the gray shaded region of the inset. (a) the distribution of U for 500 inversely folded sequence pairs with a predicted mfe structure identical to that of the guide RNA ($L = 40$nt; Genbank Acc. No. L25590) from Table S1. The distribution is shown for three different energy intervals E, as indicated in the right inset. Note that for most sequence pairs, U is close to 1, meaning that most phenotypes in the plastic repertoire of one sequence are unique. (b) Box plot of U for random sequences of different lengths (horizontal axis) and different energy intervals (inset). For each length value, data are based on 10 random RNA sequences that fold into a secondary structure phenotype, and for 5 inversely folded sequence pairs for each phenotype, i.e., for 50 sequence pairs in total. Black dots indicate medians, boxes span the 25th to 75th percentile of the distribution, and whiskers indicate the nonoutlier range (see Methods). A median of $U = 1$ indicates that for more than half of all sequence pairs, the plastic repertoires do not share a single structure.

vations show that in both biological and random RNA sequences, the plastic repertoires of two distant sequences with the same native phenotype are generally very different.

The number of unique phenotypes accessible through plasticity increases steadily during long random walks on a genotype network

The previous analysis focused on pairs of genotypes with the same phenotype *P*, and showed that their plastic repertoires are typically very different. This might no longer hold if one compares not just two but multiple genotypes with the same *P*. That is, most phenotypes accessible through the plastic repertoire of one of these genotypes might occur as part of the plastic repertoire of other genotypes, and would thus no longer be new. My next analysis shows that this is not the case, at least unless the number of genotypes becomes very large. To this end, I analyzed long random walks on a given genotype network. Each such random walk begins with a genotype *G* that has some phenotype *P*. Each step in the random walk consists of a point mutation, that is, a change in a randomly chosen nucleotide of this genotype that is required to preserve the phenotype *P*. (If any one mutation changes *P*, the mutation is rejected, and another mutation is chosen, until one is found that preserves *P*.) After each step of this random walk, I identified all phenotypes in the plastic repertoire of the random walker. I then determined for each of these phenotypes whether it was unique, that is, whether it had not already occurred in the plastic repertoire of the random walker during any one of its previous steps. If a phenotype was unique, I added it to a list of unique phenotypes. The length of this list represents the cumulative number of unique phenotypes accessible through plasticity, a number that can only increase, never decrease during the random walk. I asked how long this cumulative number of phenotypes increases. If it stops to increase after a few steps, then the total (cumulative) plastic repertoire of multiple pairs of genotypes *G* with the same *P* could be small.

Fig. 3 *a* shows the answer for three biological RNA molecules and for the energy interval $E = 3kT$. The figure shows the cumulative number of phenotypes accessible through plasticity as a function of the number of steps (mutations) in the random walk. For any one of the three molecules, this number increases steadily, and shows no signs of saturation, at least for the 1000 steps shown here. Fig. S10 shows a very similar pattern for 15 biological RNA molecules (including these three), and for all three energy intervals. Although the rate at which new accessible phenotypes accumulate differs substantially among molecules, they accumulate approximately linearly for each of the molecules.

To find out whether this is an unusual property that holds only for biological RNA molecules, I carried out multiple random walks starting from random RNA genotypes, each of which had a different phenotype *P*. Fig. 3 *b* shows, as an example, the results of 12 such random walks, three each for molecules of length 20, 40, 60, and 80. Again, novel accessible phenotypes accumulate at a roughly constant rate without any apparent saturating behavior. The
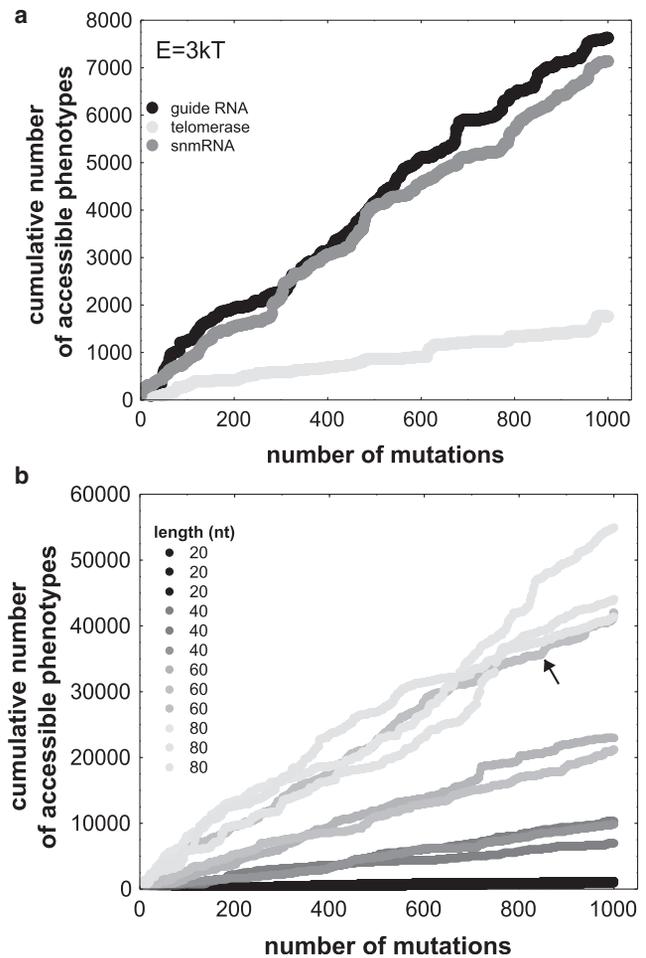
FIGURE 3 New accessible phenotypes accumulate at a nondecreasing rate. The horizontal axes show the number of steps along a random walk of 1000 steps on a genotype network. The vertical axes show the cumulative number of phenotypes that become accessible through phenotypic plasticity during the random walk. (*a*) Cumulative accessible phenotypes during one random walk each for three biological molecules that are part of a guide RNA (Genbank Acc. No. L25590), a telomerase fragment (Acc. No. AF061109), and a snmRNA (Acc. No. AJ430256, Table S1), all taken from fRNAdb (http://www.ncrna.org/frnadb, (41)). (*b*) Cumulative accessible phenotypes during 12 random walks for random starting genotypes with an arbitrary phenotype. Specifically, data are shown for starting genotypes in four different length categories (see legend), and for three random starting genotypes (with three different phenotypes) in each category. All data are shown for the energy interval $E = 3kT$.

rate is generally greater for longer molecules, which is unsurprising, because such molecules have a greater plastic repertoire (Fig. S1 *c*). However, substantial rate variation can exist among molecules with the same length. The arrow in Fig. 3 *b* highlights an example, a sequence of length 60 for which accessible phenotypes accumulate at a faster rate characteristic of longer sequences. Overall, the biological molecules accumulate accessible phenotypes at a rate comparable with those of random molecules. For example, for the energy interval $E = 3kT$, the mean cumulative number of accessible phenotypes after 1000 steps is 3342.5 for

random walks starting from 10 random sequences of length 30 nucleotides, and 4171.7 for the 8 biological sequences of the same length (Table S1). If one were to assume that these 8 biological sequences are representative of all biological sequences, then this difference would not be not statistically significant (t-test: $p = 0.25$; Mann-Whitney U-test: $p = 0.41$; $n = 10,7$).

Evolving populations with robust phenotypes can access more phenotypic variation through plasticity

The preceding sections focused on *structural* aspects of mutational robustness, i.e., its role in structuring genotype space, whereas this section focuses on the role that robustness can play in the evolutionary *dynamics* of a population. To appreciate this role, it is necessary to make a distinction between two notions of mutational robustness. Thus far, I have referred to mutational robustness as the fraction of neighbors with the same native phenotype as a given RNA genotype G. This is the robustness of a genotype G, or *genotypic robustness*. A complementary notion of robustness, based on *all* genotypes with a given native phenotype P, can be calculated by averaging the genotypic robustness of all these genotypes. Because this measure of robustness is specific to a phenotype, I call it *phenotypic robustness* (11). The robustness of a phenotype P increases with the number of RNA genotypes that form this phenotype (11). In other words, in large RNA genotype sets or genotype networks, individual RNA molecules are on average more robust. For this reason, one can use the number of RNA genotypes that form a phenotype as a proxy for its phenotypic robustness. I express this number as f, the fraction of all 4^L genotypes in genotype space that fold into a given phenotype.

The higher the mutational robustness of an RNA genotype, the lower the number of alternative phenotypes in its plastic repertoire (10). By extrapolating this simple relationship to phenotypic robustness, one might conclude that phenotypic robustness might likewise reduce the number of new phenotypes accessible through plasticity. But for populations evolving on a given genotype network this is not necessarily so. Consider a population of initially identical RNA molecules that is subject to multiple cycles ("generations") of stabilizing selection, which maintains the phenotype P of these molecules, and point mutations in individual nucleotides. As such a population spreads through the genotype network of P, two conflicting factors will affect the number of novel phenotypes it can explore through plasticity. First, the more robust the phenotype P of this population is, the fewer alternative phenotypes will occur in the plastic repertoire of each of the population's individuals. Second, the more robust P is, the less likely it is that any one mutation affecting a member of the population has a deleterious effect. This means that the population

will, on average, experience fewer of the deleterious mutations that selection eliminates, and it will thus spread faster through P's genotype network. In other words, at any one time, its members will have more diverse genotypes. If more diverse genotypes have a greater plastic repertoire, then the population with a robust phenotype may actually be able to access more phenotypic diversity through plasticity. In sum, although high phenotypic robustness implies that each of a population's members has a reduced plastic repertoire, the population as a whole will be more diverse and may thus have a greater plastic repertoire. It is not clear which of these factors dominates in its influence.

To find out, I studied a large sample of 10^6 random RNA genotypes of length $L = 30$ whose phenotypes vary broadly in their robustness, as reflected in the fraction f of genotype space that form them, which spans more than three orders of magnitude (11). Such sequences strike a balance between forming many phenotypes and being short enough that their phenotypic robustness can be estimated for many of them (see Methods). I established a population of 100 initially identical RNA genotypes that form a native phenotype P_{mfe} with a given robustness. I then subjected this population to repeated cycles ("generations") of mutation, at a rate of $\mu = 1$ nucleotide per cycle per molecule, and selection, where only those molecules survived in which mutation had not altered the native phenotype P. During each such generation, I determined the total number of unique phenotypes accessible to at least one member of the population via plasticity. Fig. 4 *a* shows this number for two phenotypes of high and low f, and for $E = 3kT$ (see Fig. S11 for other energy intervals). The figure demonstrates that the population with a highly robust phenotype can access many more novel phenotypes through plasticity. Fig. 4 *b* and *c* demonstrate that high phenotypic robustness is associated with greater genotypic diversity. Fig. 4 *d* shows data not just for two phenotypes, but for 3.8×10^4 phenotypes with broadly varying robustness (f) and three replicate populations per phenotype ($E = 3kT$, see Fig. S12 for $E = 1kT$ and $E = 5kT$). Specifically, the figure shows the total number of accessible phenotypes at generation 10 for three replicate populations per phenotype. Analogous simulations starting from each of the 30 biological molecules in this study also show that the number of accessible phenotypes at generation 10 is higher for molecules whose phenotype has a high f. Despite the few molecules used in this analysis, this association is significant for two out of the three energy intervals ($E = 5kT$: Spearman's $r = 0.57$, $p = 0.0018$; $E = 3kT$: Spearman's $r = 0.47$, $p = 0.013$; $E = 1kT$: Spearman's $r = 0.32$, $p = 0.095$). Taken together, these observations show that populations with a highly robust phenotype are more diverse and have access to more novel phenotypes through plasticity.

If one counts the cumulative number of accessible phenotypes, that is, the number of phenotypes accessible since generation one, the differences between the robust and

FIGURE 4 Phenotypic robustness entails access to more novel phenotypes through plasticity. (*a*)–(*c*): The horizontal axes show time in generations (mutation-selection cycles). The vertical axes show (*a*) the total number of unique phenotypes accessible to at least one individual in an entire population through plasticity, (*b*) the total number of sequences in the population, and (*c*) the pairwise nucleotide diversity, i.e., the number of nucleotides differing between sequence pairs, averaged over all pairs of sequences in the population. All data are based on populations of size $N = 100$ individuals, with a mutation rate of $\mu = 1$ nucleotides per generation, where selection was required to preserve the native phenotype of each individual. Each population started out with identical genotypes of $L = 30$ nucleotides that folded into the same native phenotype. The fraction of genotype space occupied by sequences folding into that phenotype is given by the value of f in the inset. The number of accessible phenotypes is shown for the energy interval $E = 3\ kT$. Circles indicate means over five replicate populations starting from the same genotype. Whiskers indicate one standard error of the mean. (*d*) Association between phenotypic robustness ($\log_{10}(f)$, horizontal axis) and phenotypes accessible through plasticity (vertical axis), based on more than 38,000 RNA phenotypes of $L = 30$, and on at least three replicate evolving populations per phenotype, where each population had $N = 100$ members and a mutation rate of $\mu = 1$ nucleotides per generation.

less robust populations become even more pronounced (Fig. S13). For example, at $E = 3kT$, and after merely 10 generations, the population with a robust phenotype could access more than 2400 novel phenotypes through plasticity, whereas the population with the less robust phenotype could access fewer than 450 novel phenotypes (Fig. S13 *b*). I also note that all these observations were made for populations where the population genetic parameter $N\mu$ was much greater than one. Such populations are polymorphic most of the time (46), but analogous observations hold for populations that are monomorphic most of the time ($N\mu \ll 1$). Fig. S14 shows an example.

DISCUSSION

Mutational robustness affects the exploration of novel phenotypes by evolving RNA molecules in two complementary ways. First, it structures genotype space into myriad interwoven networks of molecules with the same native phenotype. Evolving RNA molecules can explore these networks while preserving their native phenotype, which facilitates exploration of new phenotypes through mutation (10,18,47,48). In this study I showed that the same holds for exploration of new phenotypes through phenotypic plasticity, and for similar reasons. Specifically, small neutral neighborhoods of a molecule have a much greater plastic repertoire than the molecule itself, randomly sampled molecules from the same genotype network have very different plastic repertoires, and molecules that walk

randomly along such a network gain access to ever-increasing numbers of novel phenotypes in their plastic repertoire.

Second, robustness influences the evolutionary dynamics of RNA populations. Specifically, molecules evolving on large genotype networks—those of more robust phenotypes—diversify faster, and thus gain access to more novel phenotypes in their plastic repertoire. This accelerating effect of robustness mirrors previous observations about its role in facilitating access of new phenotypes through mutation (11). A similar accelerating role has been observed in models of gene regulation circuits (49).

Biological macromolecules are important sources of exaptations, adaptive traits whose current role is different from the reasons for their origins (7). A special case are exaptations that do have nonadaptive origins. Examples include promiscuous proteins, which catalyze not only one main chemical reaction, but also multiple side reactions on alternative substrates, a byproduct of their conformational flexibility (6,50,51). Natural or artificial selection in the laboratory can increase an enzyme's activity on such alternative substrates, which can thus become exaptations with nonadaptive origins (14,52,53). The alternative conformations of RNA molecules are also potential sources of exaptations. This "exaptive potential" is amplified by mutational robustness, because both increase the number of alternative phenotypes that RNA molecules can explore. Moreover, an intriguing connection—recently also demonstrated in metabolism (54)—exists between complexity

and this exaptive potential: more complex, i.e., longer molecules generally have a greater plastic repertoire (Fig. S1 c), and their exaptive potential is therefore also greater.

It is sometimes argued that mutations are less important than plasticity in creating new phenotypes. Systems such as RNA provide a quantitative way to address this issue. One can, for example, ask how large the set of genotypes is that have a specific phenotype in their plastic repertoire. If plasticity was all-important for the accessibility of new phenotypes, then this set of genotypes should comprise all or most of genotype space for any one phenotype. This is clearly not the case for RNA. For example, for 70 nucleotide-long random RNA molecules, the set of genotypes that have a given phenotype in their plastic repertoire at $E = 5kT$ above the mfe is more than 4000 times larger than the set of genotypes that have this phenotype as their native phenotype (Fig. S15 b), but they still only comprise a vanishing fraction 10^{-10} of genotype space (Fig. S15 c). Moreover, the plastic repertoire of the longest molecules considered here at $E = 5kT$—where the fraction of time spent by a molecule in one such phenotype may be as small as 10^{-5}—comprise only a few thousand phenotypes (Fig. S1 c), a vanishing fraction of the total number of RNA phenotypes, which scales exponentially with sequence length L at $\sim 1.8^L$ (18). Thus, to reach most RNA phenotypes, mutational alterations of sequences are clearly necessary. But whereas plasticity is no more important than mutation from the *global* perspective of an entire genotype space, plasticity can provide a *local* advantage in allowing access to new phenotypes, e.g., by making more phenotypes accessible from a given sequence and its neighbors than mutations (Fig. S6).

A potential limitation of this work is that not all of a genotype network's parts are equally accessible to an evolving population, because sets of genotypes with the same native phenotype form multiple disjoint components or subnetworks (19,55). However, this is not likely to affect my main conclusions. First, the results of both random walkers and evolving populations studied are based on a single component of a genotype network. Second, molecular evolution data suggest that so-called compensatory mutations, which can bridge the gaps between such subnetworks, are frequent in nature (56–59), such that evolving populations can explore more than one subnetwork. A further limitation is the tacit assumption that accessible phenotypes are those that occur in a molecule's plastic repertoire in thermodynamic equilibrium, which neglects phenomena such as kinetic folding (60). The incorporation of such nonequilibrium scenarios is computationally demanding, and thus needs to be deferred to future work.

A genotype network may have regions where robustness can be so high that the accessibility of novel phenotypes is reduced, which can lead to entrapment of populations in such regions (10,61). The observations reported in this study do not show any signs of such entrapment, even though they are based on many different RNA phenotypes. However, entrapment may well occur in the evolution of some biological RNA molecules, and to examine where it occurs remains another task for future work.

Because my observations are based on many molecules sampled at random from genotype space, and because they also apply to 30 biological RNA molecules, they are generic properties of sequence space and relevant for biological molecules. Conversely, this means that I characterized *typical* properties of RNA genotype space, and exceptions may exist. For example, some RNA molecules may show greater structural variability in response to mutations than through plasticity. The incidence and biological significance of such exceptions remains to be elucidated.

The plasticity of the biological molecules in this study is similar to those of random molecules with similar length, suggesting that natural selection has neither favored increased plasticity nor reduced plasticity in them. One might think that is surprising. For example, it is sometimes argued that being plastic is advantageous in adaptive evolution, because it provides more rapid access to novel phenotypes than mutations. The same plasticity can be a hindrance if the biological role of a molecule requires that it spends the vast majority of its time in its native phenotype (62). To identify biological molecules that show signatures of selection for or against plasticity is yet another question for future work.

SUPPORTING MATERIAL

Fifteen figures, three tables, and References (63,64) are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(14)00059-9.

REFERENCES

1. West-Eberhard, M. J. 2003. Developmental Plasticity and Evolution. Oxford University Press, Oxford, UK.

2. Mandal, M., and R. R. Breaker. 2004. Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.* 5:451–463.

3. Vitreschak, A. G., D. A. Rodionov, …, M. S. Gelfand. 2004. Riboswitches: the oldest mechanism for the regulation of gene expression? *Trends Genet.* 20:44–50.

4. Tokuriki, N., and D. S. Tawfik. 2009. Protein dynamism and evolvability. *Science.* 324:203–207.

5. Muralidhara, B. K., L. Sun, …, J. R. Halpert. 2008. Thermodynamic fidelity of the mammalian cytochrome P450 2B4 active site in binding substrates and inhibitors. *J. Mol. Biol.* 377:232–245.

6. O'Brien, P. J., and D. Herschlag. 1999. Catalytic promiscuity and the evolution of new enzymatic activities. *Chem. Biol.* 6:R91–R105.

7. Gould, S., and E. Vrba. 1982. Exaptation—a missing term in the science of form. *Paleobiology.* 8:4–15.

8. McKenzie, H. A., and F. H. White, Jr. 1991. Lysozyme and alpha-lact-albumin: structure, function, and interrelationships. *Adv. Protein Chem.* 41:173–315.

9. Gerhart, J., and M. Kirschner. 1998. Cells, Embryos, and Evolution. Blackwell, Boston.

10. Ancel, L. W., and W. Fontana. 2000. Plasticity, evolvability, and modularity in RNA. *J. Exper. Zoo./Molec. Dev. Evol.* 288:242–283.

11. Wagner, A. 2008. Robustness and evolvability: a paradox resolved. *Proc. Biol. Sci.* 275:91–100.

12. Bloom, J. D., P. A. Romero, …, F. H. Arnold. 2007. Neutral genetic drift can alter promiscuous protein functions, potentially aiding functional evolution. *Biol. Direct.* 2:17.

13. Bloom, J. D., L. I. Gong, and D. Baltimore. 2010. Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. *Science.* 328:1272–1275.

14. Bloom, J. D., S. T. Labthavikul, …, F. H. Arnold. 2006. Protein stability promotes evolvability. *Proc. Natl. Acad. Sci. USA.* 103:5869–5874.

15. Bershtein, S., K. Goldin, and D. S. Tawfik. 2008. Intense neutral drifts yield robust and evolvable consensus proteins. *J. Mol. Biol.* 379:1029–1044.

16. Ferrada, E., and A. Wagner. 2008. Protein robustness promotes evolutionary innovations on large evolutionary time-scales. *Proc. Biol. Sci.* 275:1595–1602.

17. Wagner, A. 2011. The molecular origins of evolutionary innovations. *Trends Genet.* 27:397–410.

18. Schuster, P., W. Fontana, …, I. L. Hofacker. 1994. From sequences to shapes and back: a case study in RNA secondary structures. *Proc. Biol. Sci.* 255:279–284.

19. Reidys, C., P. F. Stadler, and P. Schuster. 1997. Generic properties of combinatory maps: neutral networks of RNA secondary structures. *Bull. Math. Biol.* 59:339–397.

20. Wagner, A. 2012. The role of robustness in phenotypic adaptation and innovation. *Proc. Biol. Sci.* 279:1249–1258.

21. Espinosa-Soto, C., O. C. Martin, and A. Wagner. 2011. Phenotypic plasticity can facilitate adaptive evolution in gene regulatory circuits. *BMC Evol. Biol.* 11:5.

22. Ponting, C. P., P. L. Oliver, and W. Reik. 2009. Evolution and functions of long noncoding RNAs. *Cell.* 136:629–641.

23. Storz, G., S. Altuvia, and K. M. Wassarman. 2005. An abundance of RNA regulators. *Annu. Rev. Biochem.* 74:199–217.

24. Waters, L. S., and G. Storz. 2009. Regulatory RNAs in bacteria. *Cell.* 136:615–628.

25. Bartel, D. P. 2009. MicroRNAs: target recognition and regulatory functions. *Cell.* 136:215–233.

26. Alberts, B., A. Johnson, …, P. Walter. 2008. Molecular Biology of the Cell. Garland Science, New York.

27. Dayton, E. T., D. A. Konings, …, A. I. Dayton. 1992. Extensive sequence-specific information throughout the CAR/RRE, the target sequence of the human immunodeficiency virus type 1 Rev protein. *J. Virol.* 66:1139–1151.

28. Baudin, F., R. Marquet, …, C. Ehresmann. 1993. Functional sites in the 5′ region of human immunodeficiency virus type 1 RNA form defined structural domains. *J. Mol. Biol.* 229:382–397.

29. Powell, D. M., M. J. Zhang, …, A. I. Dayton. 1995. Sequence specificity in the higher-order interaction of the Rev protein of HIV-1 with its target sequence, the RRE. *J. Acquir. Immune Defic. Syndr. Hum. Retroviral.* 10:317–323.

30. Jackson, R. J., and A. Kaminski. 1995. Internal initiation of translation in eukaryotes: the picornavirus paradigm and beyond. *RNA.* 1:985–1000.

31. Mandl, C. W., H. Holzmann, …, F. X. Heinz. 1998. Spontaneous and engineered deletions in the 3′ noncoding region of tick-borne encephalitis virus: construction of highly attenuated mutants of a flavivirus. *J. Virol.* 72:2132–2140.

32. Manzella, J. M., W. Rychlik, …, P. J. Blackshear. 1991. Insulin induction of ornithine decarboxylase. Importance of messenger RNA secondary structure and phosphorylation of eukaryotic initiation factor eIF-4B and factor eIF-4E. *J. Biol. Chem.* 266:2383–2389.

33. Van de Guchte, M., T. Van der Lende, …, G. Venema. 1991. A possible contribution of messenger-RNA secondary structure to translation initiation efficiency in *Lactococcus lactis. FEMS Microbiol. Lett.* 81:201–208.

34. Okamura-Ikeda, K., K. Fujiwara, and Y. Motokawa. 1992. Molecular cloning of a cDNA encoding chicken T-protein of the glycine cleavage system and expression of the functional protein in Escherichia coli. Effect of mRNA secondary structure in the translational initiation region on expression. *J. Biol. Chem.* 267:18284–18290.

35. Pati, S., D. DiSilvestre, and W. S. Brusilow. 1992. Regulation of the *Escherichia coli* uncH gene by mRNA secondary structure and translational coupling. *Mol. Microbiol.* 6:3559–3566.

36. Hofacker, I. L., W. Fontana, …, P. Schuster. 1994. Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.* 125:167–188.

37. Wuchty, S., W. Fontana, …, P. Schuster. 1999. Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers.* 49:145–165.

38. Reference deleted in proof.

39. Sumedha, O. C. M., O. C. Martin, and A. Wagner. 2007. New structural variation in evolutionary searches of RNA neutral networks. *Biosystems.* 90:475–485.

40. McCaskill, J. S. 1990. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers.* 29:1105–1119.

41. Kin, T., K. Yamada, …, K. Asai. 2007. fRNAdb: a platform for mining/annotating functional RNA candidates from non-coding RNA sequences. *Nucleic Acids Res.* 35 (Database issue):D145–D148.

42. Jörg, T., O. C. Martin, and A. Wagner. 2008. Neutral network sizes of biological RNA molecules can be computed and are not atypically small. *BMC Bioinformatics.* 9:464.

43. Gruner, W., R. Giegerich, …, P. Schuster. 1996. Analysis of RNA sequence structure maps by exhaustive enumeration. I. Neutral networks. *Monatsh. Chem.* 127:355–374.

44. Freier, S. M., R. Kierzek, …, D. H. Turner. 1986. Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl. Acad. Sci. USA.* 83:9373–9377.

45. Freeman, S., and J. C. Herron. 2007. Evolution, 4th ed. Pearson, San Francisco, CA.

46. Kimura, M. 1983. The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge, UK.

47. Huynen, M. A., P. F. Stadler, and W. Fontana. 1996. Smoothness within ruggedness: the role of neutrality in adaptation. *Proc. Natl. Acad. Sci. USA.* 93:397–401.

48. Huynen, M. A. 1996. Exploring phenotype space through neutral evolution. *J. Mol. Evol.* 43:165–169.

49. Espinosa-Soto, C., O. C. Martin, and A. Wagner. 2011. Phenotypic robustness can increase phenotypic variability after nongenetic perturbations in gene regulatory circuits. *J. Evol. Biol.* 24:1284–1297.

50. Khersonsky, O., and D. S. Tawfik. 2010. Enzyme promiscuity: A mechanistic and evolutionary perspective. *In* Annual Review of Biochemistry, Vol. 79. R. D. Kornberg, C. R. H. Raetz, J. E. Rothman, and J. W. Thorner, editors. 471–505.

51. Sikosek, T., H.-S. Chan, and E. Bornberg-Bauer. 2012. Escape from Adaptive Conflict follows from weak functional trade-offs and mutational robustness. *Proc. Natl. Acad. Sci. USA.* 109:14888–14893.

52. Amitai, G., R. Gupta, and D. Tawfik. 2007. Latent evolutionary potentials under the neutral mutational drift of an enzyme. *HFSP Journal.* 1:67–78.

53. Aharoni, A., L. Gaidukov, …, D. S. Tawfik. 2005. The 'evolvability' of promiscuous protein functions. *Nat. Genet.* 37:73–76.

54. Barve, A., and A. Wagner. 2013. A latent capacity for evolutionary innovation through exaptation in metabolic systems. *Nature.* 500:203–206.

55. Schaper, S., I. G. Johnston, and A. A. Louis. 2012. Epistasis can lead to fragmented neutral spaces and contingency in evolution. *Proc. Biol. Sci.* 279:1777–1783.

56. Parsch, J., J. M. Braverman, and W. Stephan. 2000. Comparative sequence analysis and patterns of covariation in RNA secondary structures. *Genetics.* 154:909–921.

57. Kondrashov, A. S., S. Sunyaev, and F. A. Kondrashov. 2002. Dobzhansky-Muller incompatibilities in protein evolution. *Proc. Natl. Acad. Sci. USA.* 99:14878–14883.

58. Kern, A. D., and F. A. Kondrashov. 2004. Mechanisms and convergence of compensatory evolution in mammalian mitochondrial tRNAs. *Nat. Genet.* 36:1207–1212.

59. Meer, M. V., A. S. Kondrashov, …, F. A. Kondrashov. 2010. Compensatory evolution in mitochondrial tRNAs navigates valleys of low fitness. *Nature.* 464:279–282.

60. Tacker, M., P. Stadler, …, P. Schuster. 1996. Algorithm independent properties of RNA secondary structure predictions. *Eur. Biophys. J.* 25:115–130.

61. Manrubia, S. C., and J. A. Cuesta. 2013. Evolution on genotype networks leads to phenotypic entrapment. (http://arxiv.org/abs/1307.0968) (submitted).

62. Ancel, L. W. 2000. Undermining the Baldwin expediting effect: does phenotypic plasticity accelerate evolution? *Theor. Popul. Biol.* 58:307–319.

63. Riley, G. R., R. A. Corell, and K. Stuart. 1994. Multiple guide RNAs for identical editing of *Trypanosoma brucei* apocytochrome b mRNA have an unusual minicircle location and are developmentally regulated. *J. Biol. Chem.* 269:6101–6108.

64. Benson, D. A., M. Cavanaugh, …, E. W. Sayers. 2013. GenBank. *Nucleic Acids Res.* 41 (Database issue):D36–D42.

Supplementary Results to

Mutational robustness accelerates the origin of novel RNA phenotypes through phenotypic plasticity

Andreas Wagner

andreas.wagner@ieu.uh.ch

Biological and randomly sampled RNA molecules have similar plasticity.

To analyze the plasticity of biological RNA molecules, I studied 30 representatives from fRNAdb, a database of functional RNA molecules (1) isolated from different organisms. The database includes functionally diverse molecules, such as guide RNAs and telomerases (Table S1), as well as RNA molecules with unknown functions, such as small non-messenger RNAs (snmRNAs).

The inset of Figure S1a shows the predicted minimum free energy structure P_{mfe} of a biological RNA molecule, part of a guide RNA from the protozoan parasite *Trypanosoma brucei* (2) with genbank accession number L25590 (3). In energy intervals of E=1 kT, 3 kT, and 5 kT above the minimum free energy, one finds 5, 53, and 386 phenotypes, respectively (see also Table S2). To find out whether this is an unusually small or large number compared to random sequences with the same phenotype, I took advantage of the ability to create such sequences through inverse folding (4). Specifically, I created 1000 sequences that fold into the phenotype shown in Figure S1a, and evaluated the number of their alternative phenotypes in each of the above three energy intervals. The answer is shown in the black, dark, and light grey histograms in Figure S1a. Arrows indicate the plasticity of the biological sequence, i.e., their number of alternative phenotypes reported above. Note the logarithmic scale on the horizontal axis. I note that each arrow points to a bar that lies within the bulk of the respective distribution, indicating that the plasticity of this molecule is not unusual.

I next studied the average pairwise sequence divergence between the alternative phenotypes. To this end, I computed the pairwise Hamming distances between the dot-parentheses representations of two secondary structures, which increases by one for every base that is paired in one structure but unpaired in the other. I then averaged this distance over all pairs of phenotypes in the molecule's plastic repertoire. For the guide RNA molecule, this average pairwise distance is equal to 6.6, 9.1, and 10.7 for E=1, 3, and 5kT intervals above the mfe, respectively. The arrows in Figure S1b indicate these values in the context of a box plot of the average pairwise distances of alternative phenotypes found for 1000 inversely folded sequences. Again, the pairwise distances of the guide RNA are not unusual. Note also the following pattern: The number of alternative phenotypes increases approximately tenfold between each successive energy interval (Figure S1a), whereas their phenotypic diversity increases by 47 percent in the first step, and by less than 20 percent in the second step. In other words, as the total number of alternative phenotypes is increased, their diversity levels off.

Figures S2 shows an analogous analysis for two further biological RNA molecules, a telomerase and a snmRNA, with qualitatively identical results. Data on plasticity for all 30 molecules (Table S2), as well as for 100 inversely folded RNA sequences for each of them (Table S2 and Figure S3), as well as data on phenotypic diversity of the plastic repertoire for these 30 molecules (Table S3) and 100 inversely folded sequences for each (Table S3 an Figure S4) demonstrate that the biological sequences have similar plasticity to random sequences with the same phenotypes. Also, note that plasticity increases approximately tenfold, from 5 to 50 to 500 molecules, for each two-unit increase in E (Figure S3a through S3c), whereas phenotypic diversity increases only modestly from E=1kT to E=5kT. Taken together, these observations suggest that biological molecules with a given minimum free

energy phenotype are no more or less plastic than randomly sampled molecules with the same phenotype.

Figure S1c broadens this analysis to randomly sampled molecules with *any* phenotype, not just the 30 phenotypes of the biological RNA molecules studies above. For this analysis, I chose molecules of a given length (horizontal axis) at random from sequence space, until I had obtained 100 molecules with 100 different minimum free energy phenotypes. The figure indicates the plasticity of molecules thus obtained for different lengths and different energy intervals E. Note the logarithmic scale and that plasticity increases much faster than linearly with E. Two observations are germane. First, the observed plasticities are very similar to those of biological molecules (Table S2 and Figure S3, note that the lengths of the biological molecules range between L=30 and 40). Second, for both biological and random molecules, the range of plasticity is extremely broad, and some sequences are extremely plastic. For example, whereas random molecules of length L=80 at E=5 kT adopt 2648 alternative phenotypes, the 10th and 90th percentile lie at 671 and 11864 phenotypes, and one molecule can form 7.4×10^{5} alternative phenotypes. Figure S1d shows an analogous analysis for the average pairwise diversity of alternative phenotypes. As observed above, this diversity is within the range observed for biological phenotypes (Table S3 and Figure S4), and it increases only very modestly from E=3kT to E=5kT.

In sum, based on large samples of inversely folded phenotypes and randomly sampled genotypes, biological RNA sequences do not appear different in their plasticity from random sequences. Thus, the plasticity of random molecules can help us understand general features of RNA phenotypes that are also relevant for biological molecules.

Supplementary Figures to

Mutational robustness accelerates the origin of novel RNA phenotypes through phenotypic plasticity

Andreas Wagner

andreas.wagner@ieu.uh.ch
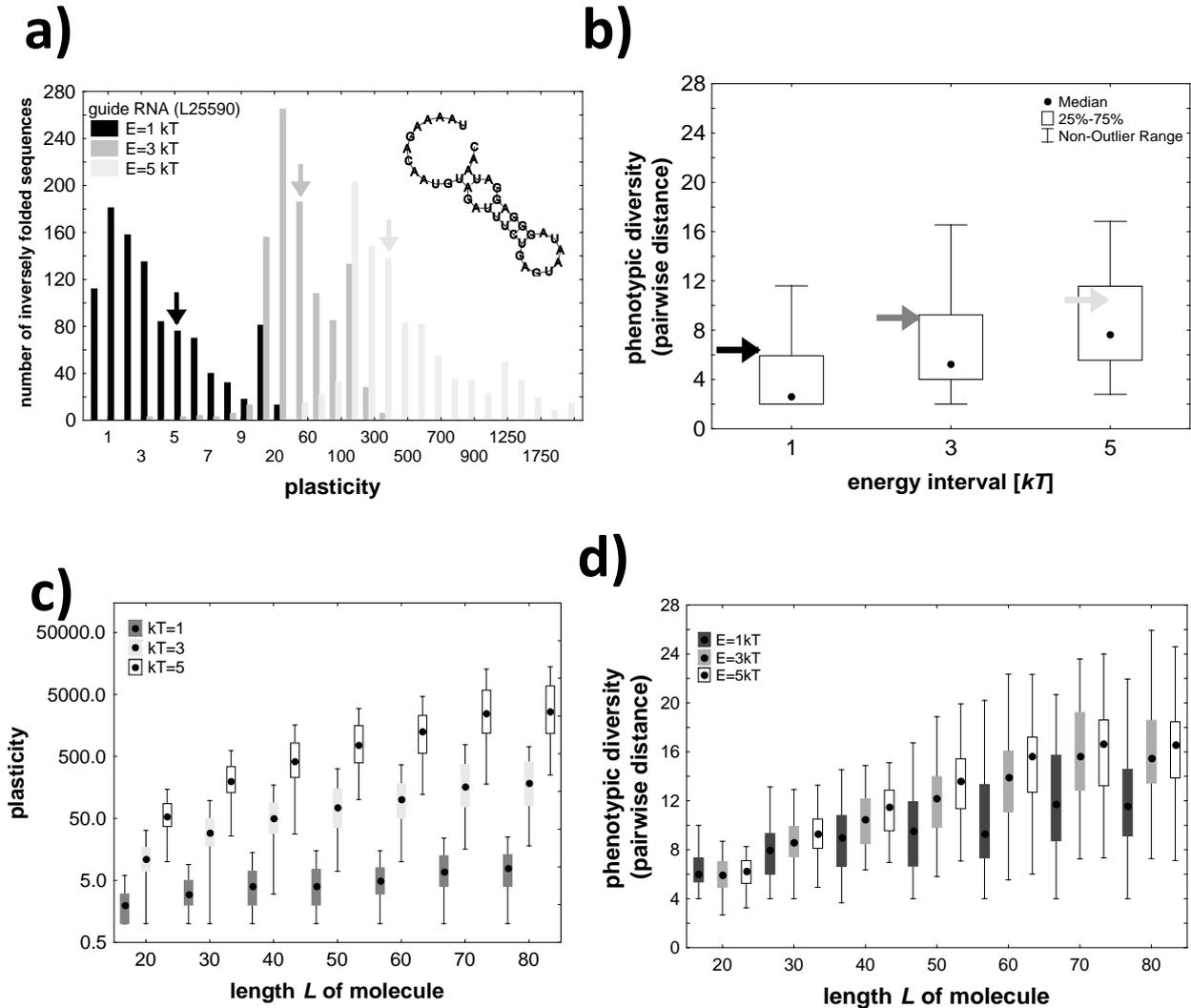
Figure S1. Biological and random RNA molecules have similar phenotypic plasticity. a)
Distribution of phenotypic plasticity, that is, the number of phenotypes (secondary structures) in a
given energy interval E above the mfe, as greyscale-coded in the left inset. Histograms are based on
1000 inversely folded sequences for the secondary structure shown in the right inset, which is the
predicted secondary structure (from the Vienna RNA web server) of part of a guide RNA (genbank
acc. no L25590, from fRNA db, 1). Arrows indicate the plasticity of this biological sequence. **b)** Box
plots of the average pairwise structural diversity (see methods) for all structures in the plastic
repertoire of 1000 inversely folded sequences with the same native phenotype as in a). Greyscale-
coded arrows indicate the structural diversity of the plastic repertoire of the sequence shown in a). **c)**
Box plot of phenotypic plasticity (vertical axis, note the logarithmic scale), and **d)** pairwise phenotypic
diversity, for random RNA sequences of a given length (horizontal axis), and different energy
intervals (grey scales, see inset). Data in each box of b) and d) are based on the plastic repertoires of
100 random RNA genotypes that adopt some mfe secondary structure. In box plots, black dots indicate
medians, boxes span the 25th to 75th percentile of the distribution, and whiskers indicate the non-outlier
range (see Methods).
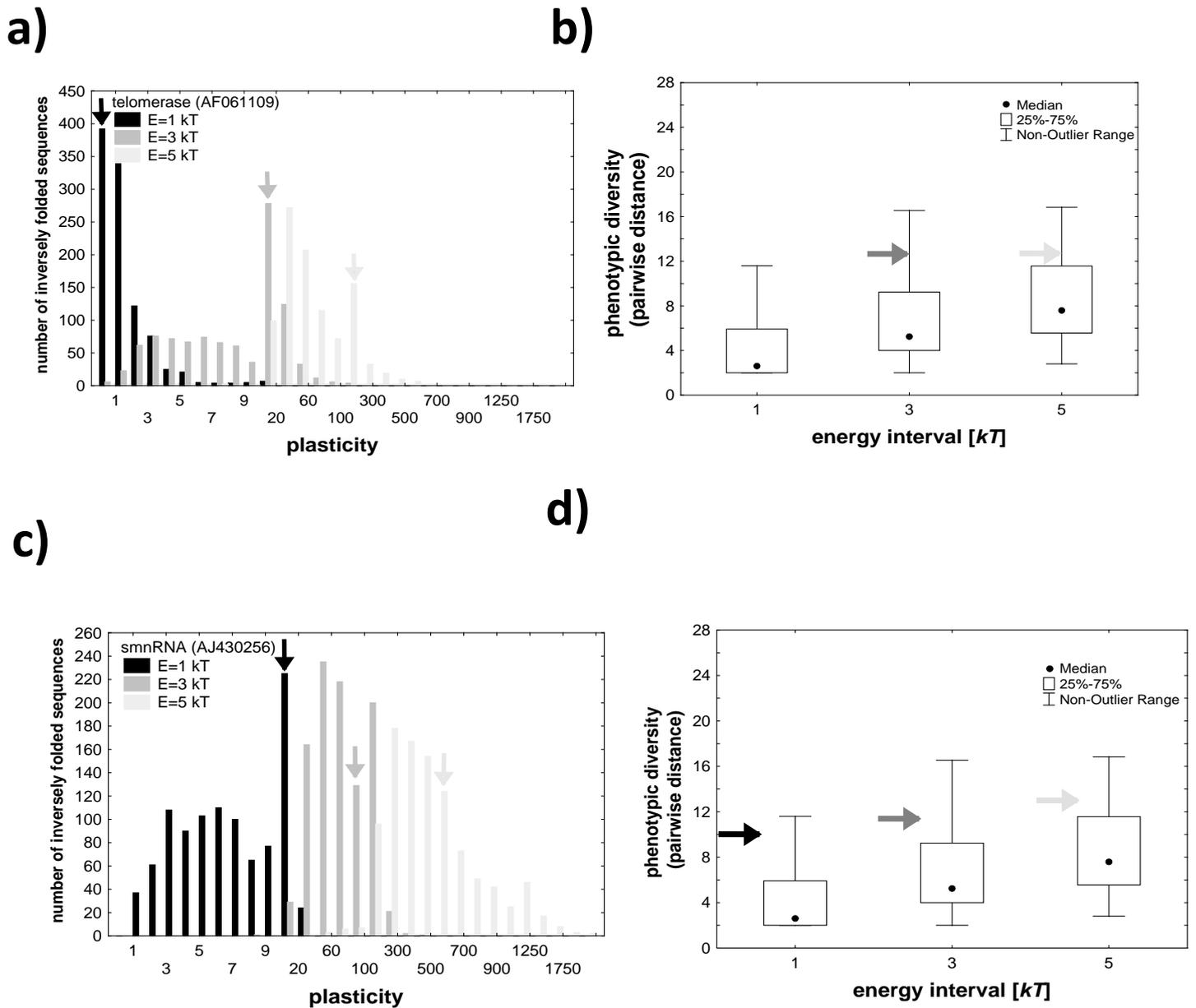
Figure S2: Phenotypic plasticity of two biological RNA molecules. a), c) Distribution of phenotypic plasticity, that is, the number of secondary structure phenotypes in a given energy interval *E* above the mfe, as greyscale-coded in the left inset. Histograms are based on 1000 inversely folded sequences for the secondary structure of two biological sequences that form part of a telomerase in a) (genbank acc. no. AF061109), and of a smnRNA in c) (genbank acc. no. AJ430256), both from fRNAdb (1). See Table S1 for predicted mfe secondary structures of these molecules. Arrows indicate the plasticity of the biological sequence at different energy intervals. **b)** and **d)** Box plots of the average pairwise phenotypic diversity (see methods) for all structures in the plastic repertoire of 1000 inversely folded sequences with the same secondary structure as the biological sequences in a) and c), respectively. Greyscale-coded arrows indicate the structural diversity of the plastic repertoire of the biological sequences from a) and c). An arrow is missing at the left-most box of panel b), because the plastic repertoire of the telomerase sequence contains no phenotype (aside from the native phenotype) at *E*=1*kT*. In box plots, black dots indicate medians, boxes span the 25th to 75th percentile of the distribution, and whiskers indicate the non-outlier range (see Methods).
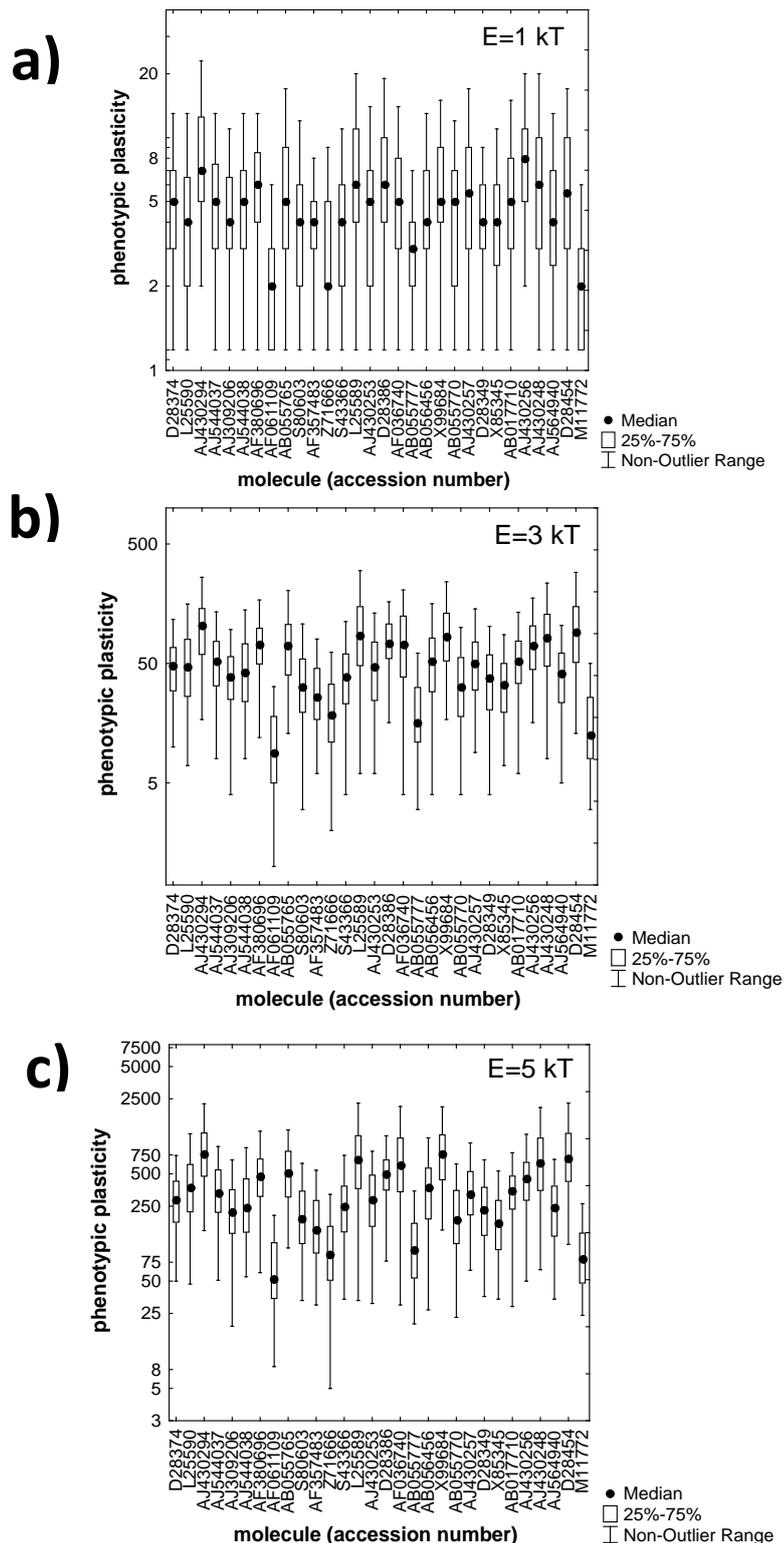
Figure S3. Phenotypic plasticity of 30 different biological RNA molecules. The horizontal axes show the genbank accession numbers of 30 biological RNA molecules from Table S1. The vertical axes show box plots of the distribution of phenotypic plasticity for each molecule. Specifically, data in each box are based on 100 inversely folded sequences of the same predicted secondary structure as the biological molecule. **a)** *E=1kT*, **b)** *E=3kT*, and **c)** *E=5kT*. Black dots indicate medians, boxes span the 25th to 75th percentile of the distribution, and whiskers indicate the non-outlier range (see Methods).
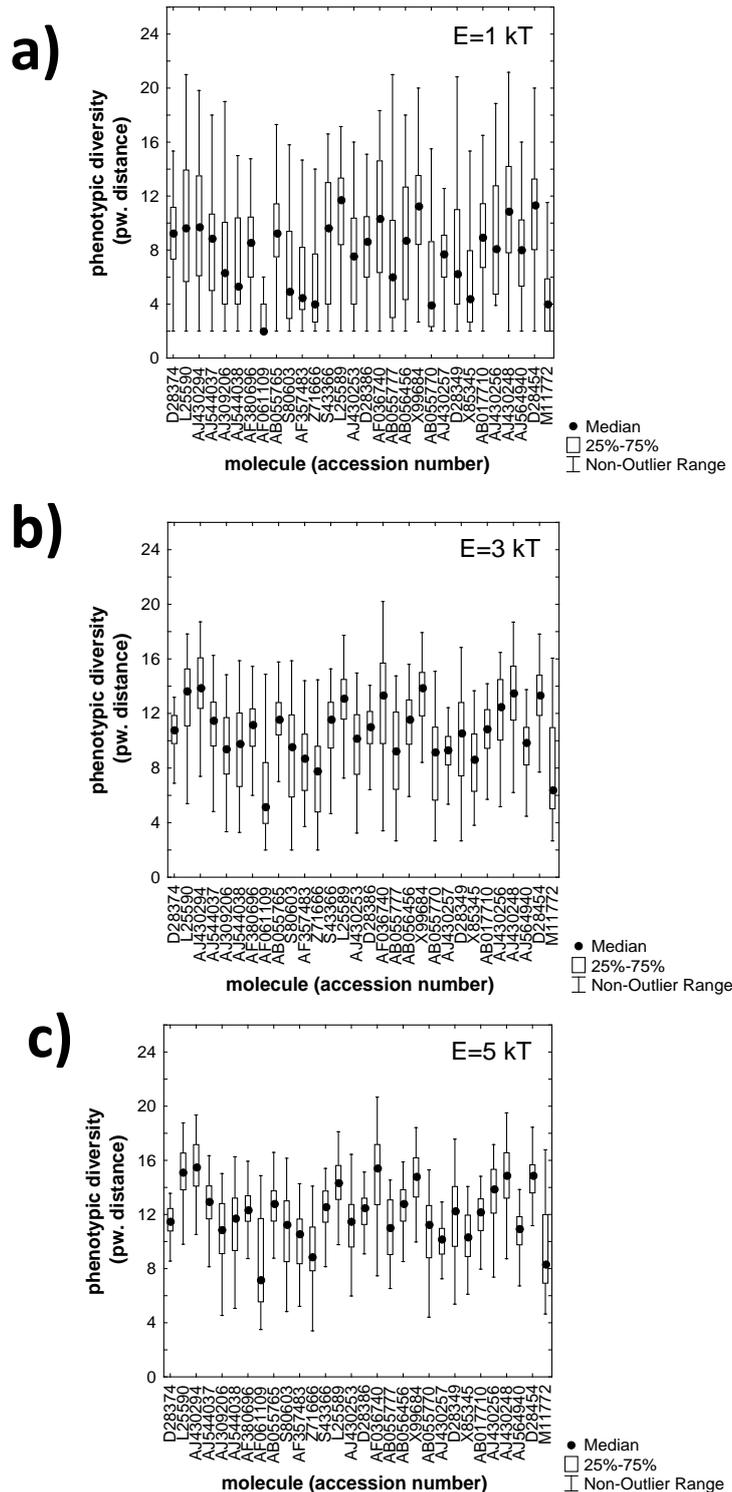
Figure S4. Phenotypic diversity of the plastic repertoire of 30 different biological RNA molecules. The horizontal axes show the genbank accession numbers of 30 biological RNA molecules from table S2. The vertical axes show box plots of the pairwise structural diversity of the plastic repertoire of each RNA for different energy intervals. **a)** $E=1kT$, **b)** $E=3kT$, and **c)** $E=5$kT. Data in each box is based on 100 inversely folded sequences of the same predicted secondary structure as the biological molecule, and on the average pairwise structural diversity of the plastic repertoire of each of these sequences. Black dots indicate medians, boxes span the 25th to 75th percentile of the distribution, and whiskers indicate the non-outlier range (see Methods).
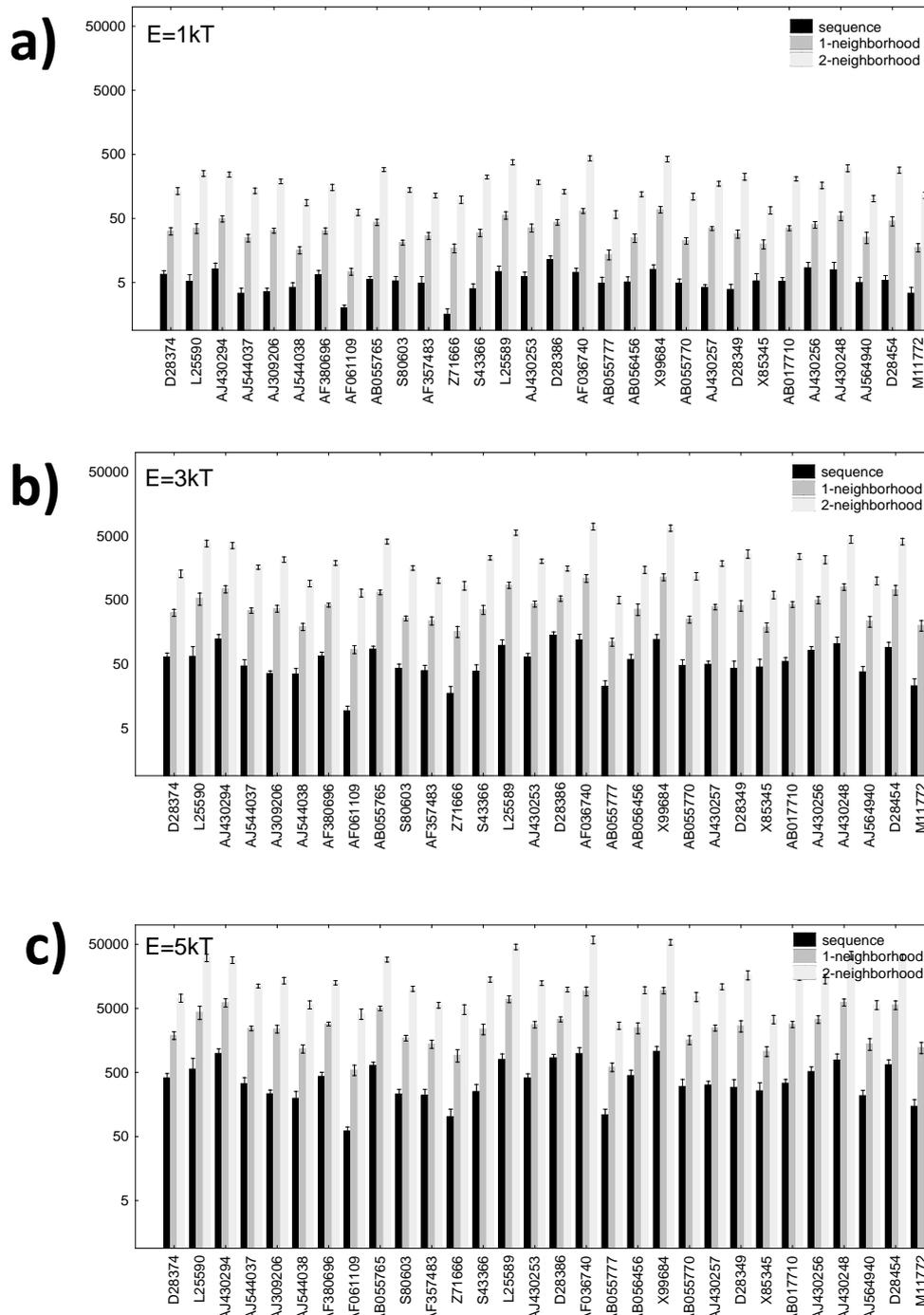
Figure S5. Even small neutral neighborhoods of a sequence render many novel phenotypes accessible through plasticity. The horizontal axes show genbank (5) accession numbers of 30 biological RNA sequences shown in Table S1. Above each accession number is a group of three bars. The black bar in each group of bars indicates the mean number of *unique* phenotypes, i.e., phenotypes different from one another, in the plastic repertoire P_0 of an (inversely folded) sequence with the same native phenotype P as the biological molecule. The height of the medium gray (middle) bar indicates the number of unique phenotypes in the union of P_0 and P_1, where P_1 is the number of all different phenotypes in those 1-neighbors of the sequence that have the same native phenotype P as the sequence itself. Light gray bars indicate, analogously, the number of unique phenotypes in the union of $P_0, P_1,$ and $P_2,$, where P_2 denotes the plastic repertoire of all 2-neighbors of the sequence with the same phenotype P as the sequence itself. Note the logarithmic scales on the vertical axes. Averages (bar heights) are computed over 10 inversely folded sequences for each secondary structure, and whiskers indicate one standard error of the mean. **a), b),** and **c)** show plastic repertoires for *E=1kT, E=3kT, and E=5kT*, respectively.
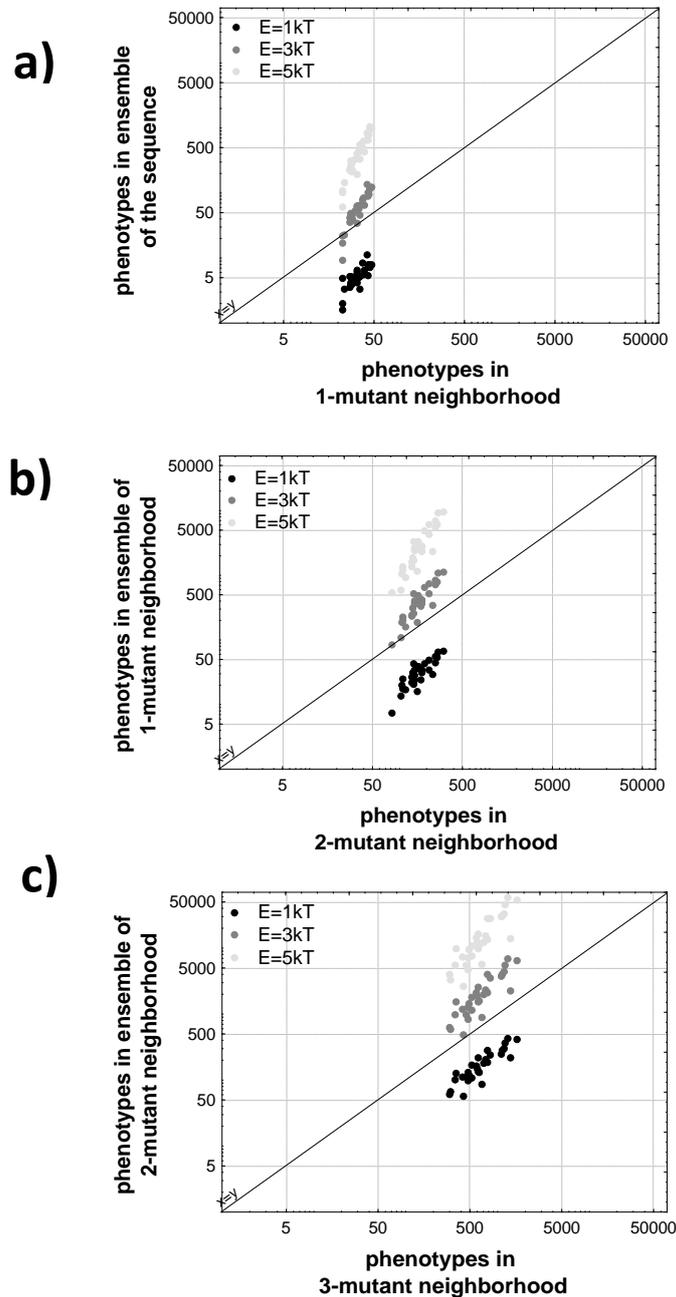
Figure S6: Phenotypic plasticity (vertical axes) can produce more novel phenotypes than mutation (horizontal axes) for energy intervals $E>1kT$. Data in these figures are based on the 30 biological RNA sequences of Table S1. Specifically, each data point is an average over 10 inversely folded sequences with the predicted native phenotype of one of the 30 RNA molecules. Different shades of gray indicate data for different energy intervals E (left inset). The solid horizontal line in each panel indicates the line of equality on the horizontal and vertical axes. Note the logarithmic scales on the vertical axes. **a)** Horizontal axis: average total number of unique phenotypes in the 1-mutant neighborhood of an (inversely folded) sequence G. Vertical axis: average total number of unique phenotypes in the plastic repertoire of the sequence G. **b)** Horizontal axis: average total number of unique phenotypes in the 2-mutant neighborhood of an (inversely folded) sequence G. Vertical axis: average total number of phenotypes in the plastic repertoire of all sequences in the neutral 1-mutant neighborhood of G. **c)** Horizontal axis: average total number of unique phenotypes in the 3-mutant neighborhood of an (inversely folded) sequence G. Vertical axis: average total number of unique phenotypes in the plastic repertoire of all sequences in the neutral 2-neighborhood of G.
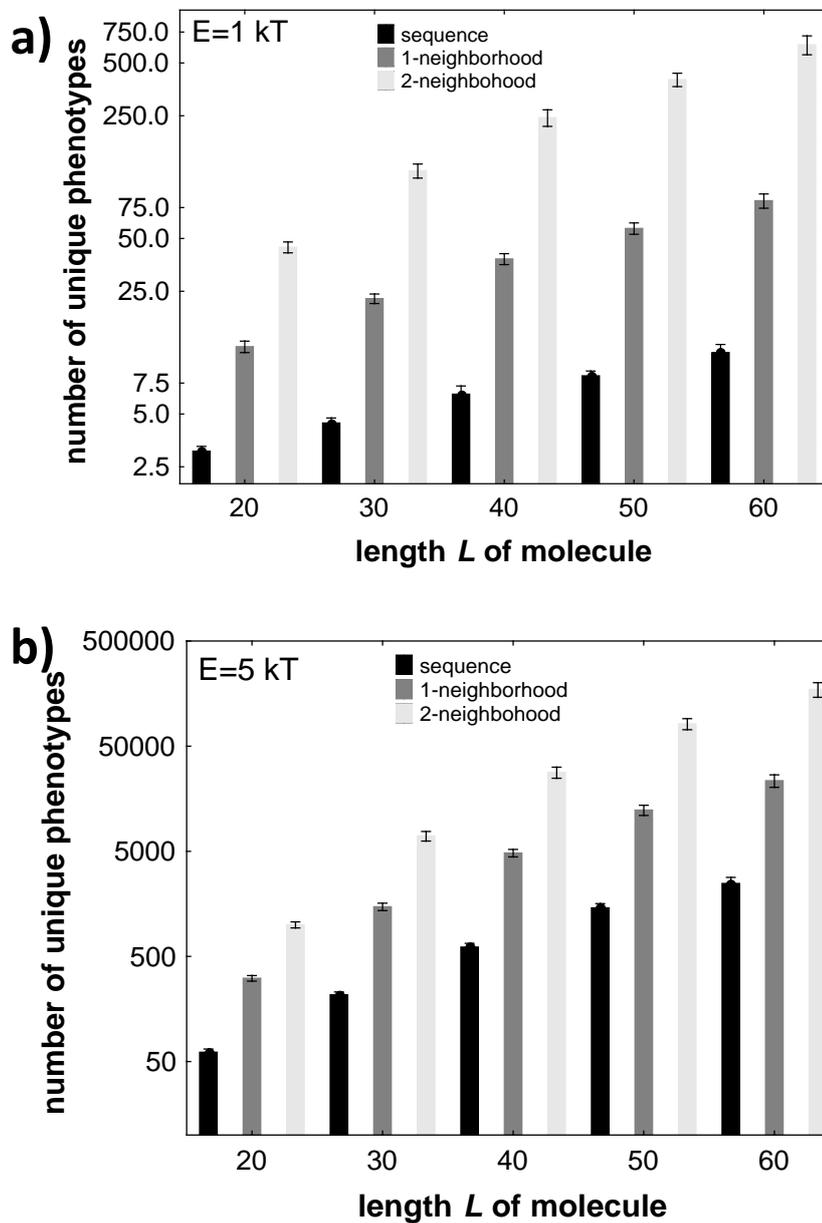
Figure S7: Even small neutral neighborhoods of a sequence render many novel phenotypes accessible through plasticity (random sequences in the energy interval *E=1kT* and *E=5kT*). The horizontal axis shows the lengths of the RNA molecules considered. For each length value with the exception of length *L*=60 nucleotides, data are based on *n*=20 different RNA phenotypes *P* of randomly chosen RNA sequences *G,* and on five inversely folded sequences per phenotype *P*. For length 60 *n*=18. Black bars indicate the mean number of *unique* phenotypes, i.e., phenotypes different from one another, in the plastic repertoire P_0 of an (inversely folded) sequence of a given length and a given random phenotype. Medium gray bars indicate the number of unique phenotypes in the union of P_0 and P_1, where P_1 is the number of all different phenotypes in those 1-neighbors of the sequence that have the same native phenotype *P* as the sequence itself. Light gray bars indicate, analogously, the number of unique phenotypes in the union of P_0, P_1, and P_2,, where P_2 denotes the plastic repertoire of all 2-neighbors of the sequence with the same phenotype *P* as the sequence itself. The height of each bar corresponds to an average over the 5x20=100 (or 5x18=90) inversely folded sequences considered in each length category. Whiskers indicate one standard error of the mean. **a)** *E=1kT* **b)** *E=5kT.*
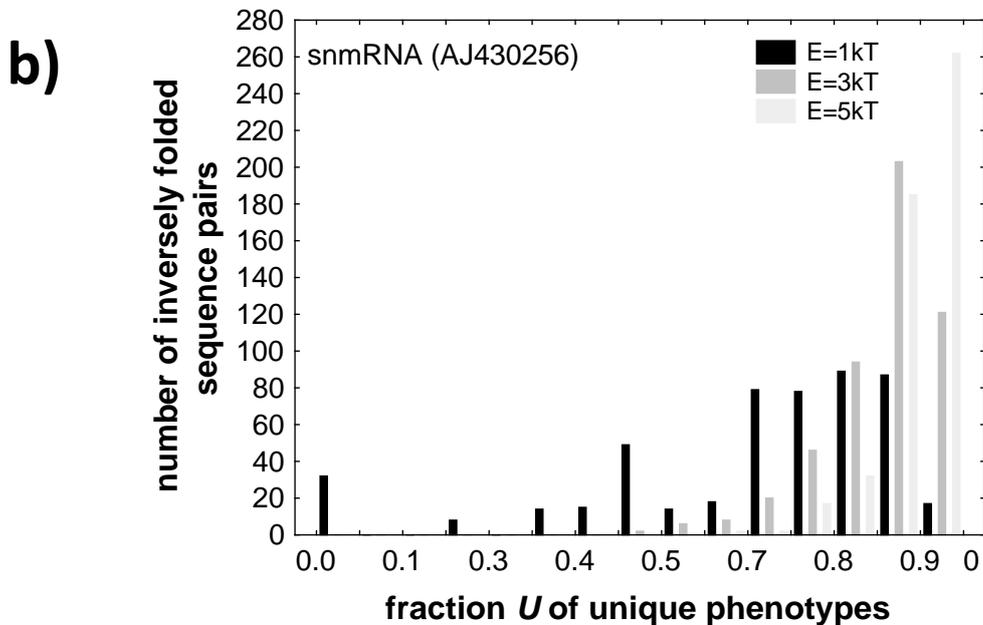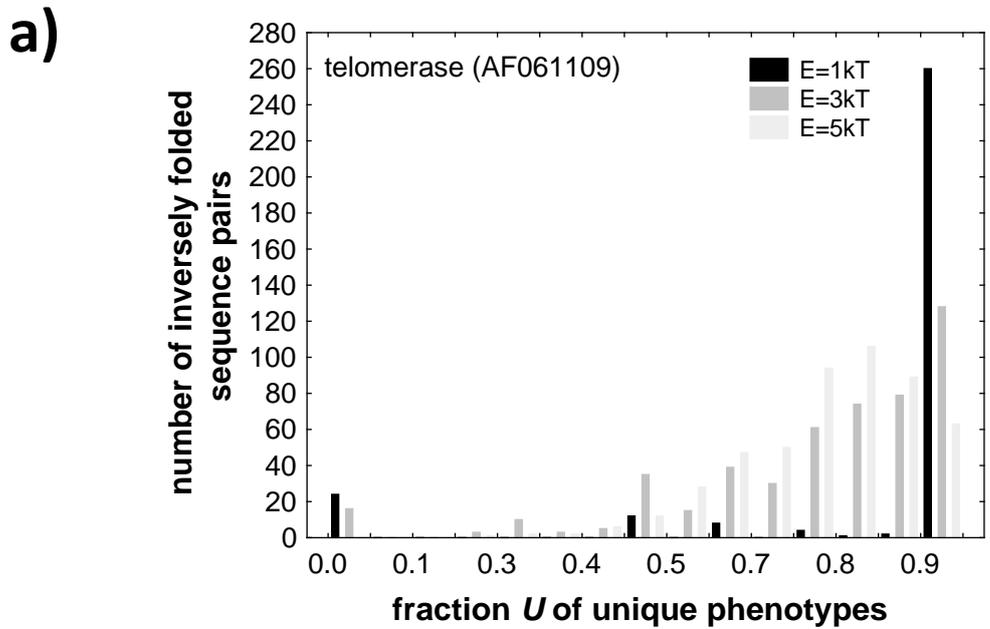
Figure S8. Different genotypes with the same native phenotype differ greatly in the phenotypes of their plastic repertoire. The figure shows data about the plastic repertoires of two genotypes G_1 and G_2 with the same native phenotype P. Specifically, the figure displays the fraction U of phenotypes that are in the plastic repertoire of G_2, but not of G_1 (as indicated by the gray shaded region of the inset of Figure 2a). The panels show the distribution of U for 500 inversely folded sequence pairs with a predicted mfe structure identical to that of **a)** a telomerase(genbank acc. no AF061109) and **b)** a snmRNA (genbank acc. no. AJ430256) from Table S1. Distributions are shown for three different energy intervals E, as indicated in the inset. Note that for most sequence pairs U is close to one, meaning that most phenotypes in the plastic repertoire of a sequence are unique.
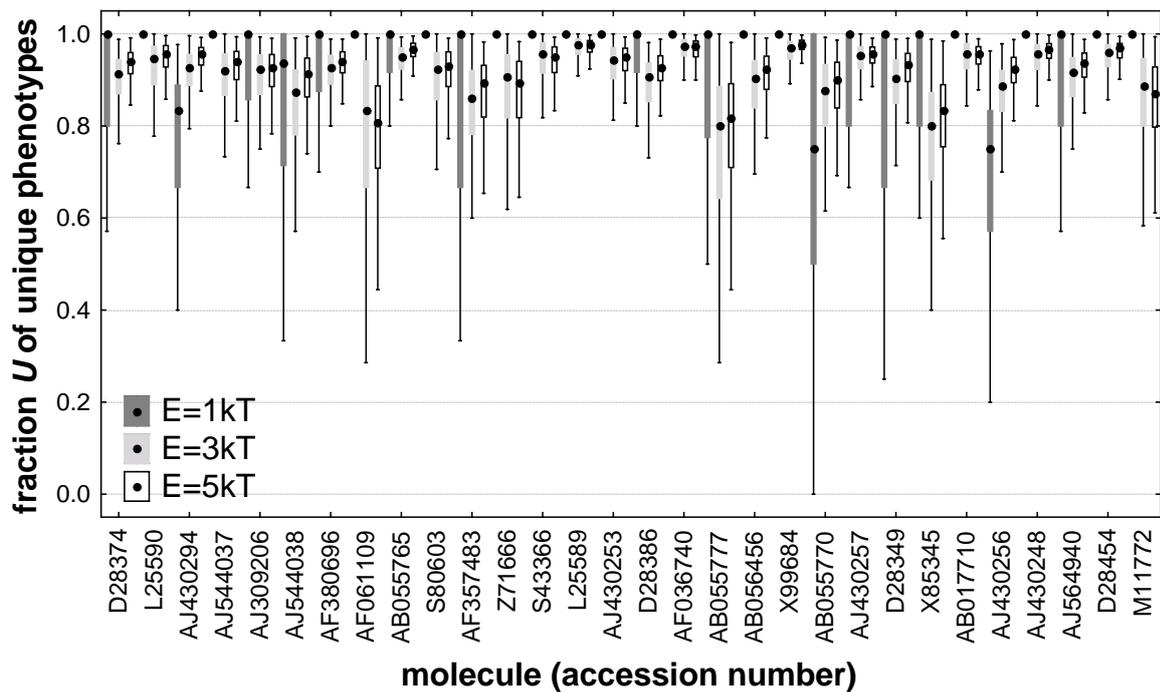
Figure S9. Different genotypes with the same native phenotype differ greatly in their plastic repertoire. The figure shows data about the plastic repertoires of two genotypes G_1 and G_2 with the same native phenotype P. Specifically, the figure displays the fraction U of phenotypes that are in the plastic repertoire of G_2, but not of G_1 (as indicated by the gray shaded region of the inset of Figure 2a). For each of 30 biological molecules from Table S1, whose genbank accession numbers are indicated on the horizontal axis, the figure shows a box plot of U based on 500 inversely folded sequence pairs with a predicted mfe structure identical to that of the biological molecule. The three boxes above each accession number indicate data for three different energy intervals E, as indicated in the inset. Black dots indicate medians, boxes span the 25th to 75th percentile of the distribution, and whiskers indicate the non-outlier range (see Methods). Note that U is generally close to one, meaning that most phenotypes in the plastic repertoire of a sequence are unique.
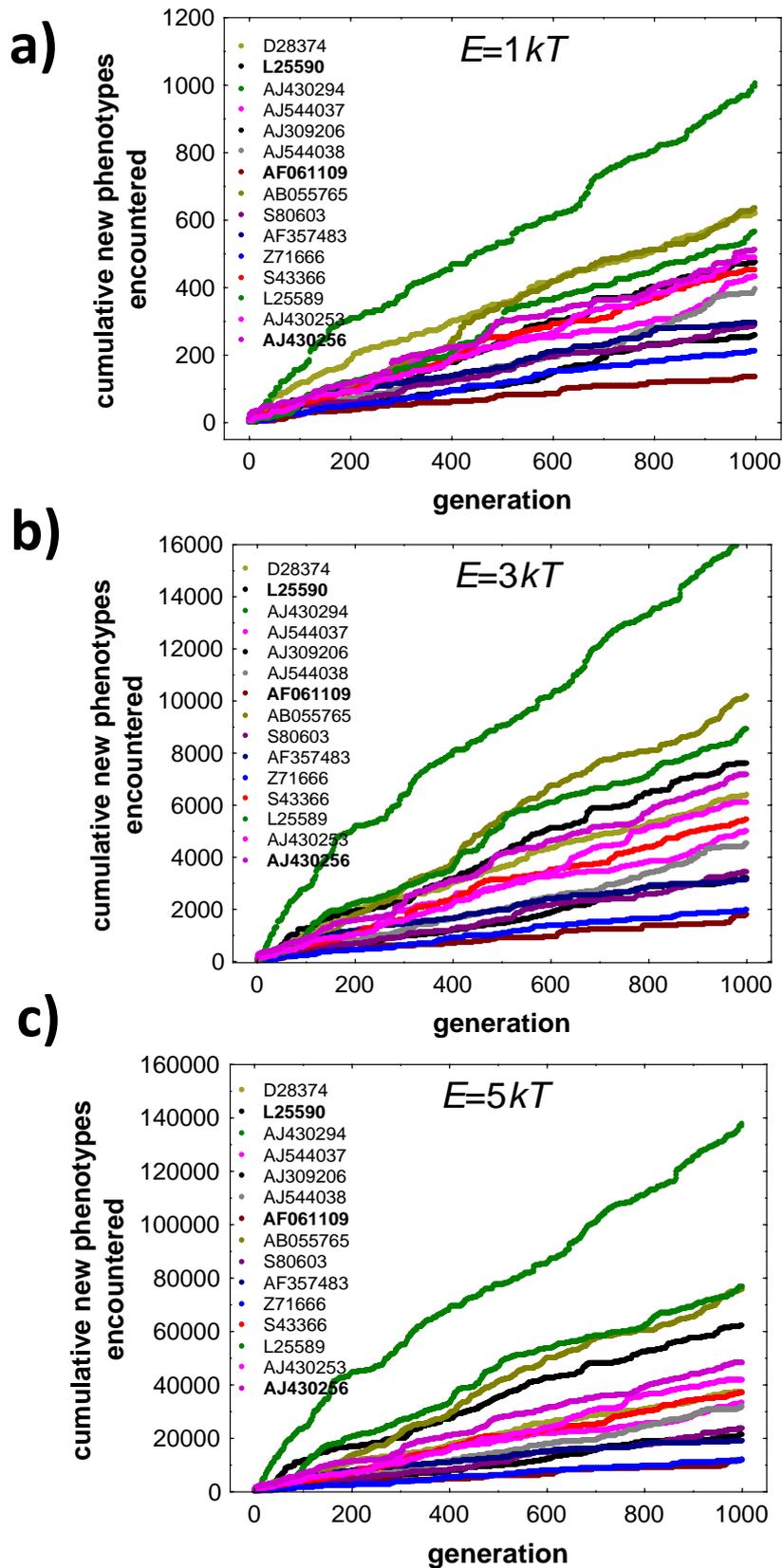
Figure S10: New accessible phenotypes accumulate at a non-decreasing rate for 15 biological RNA molecules. The horizontal axes show the number of steps along a random walk of 1000 steps on a genotype network. The vertical axes show the cumulative number of phenotypes that become accessible through phenotypic plasticity during the random walk. Shown are cumulative accessible phenotypes during one random walk each for 15 biological molecules (Table S1), including the three molecules used in Figure 4 (bold type in legend). Trajectories are shown only for 15 and not all 30 molecules because of space limitations. **a)** $E=1kT$, **b)** $E=3kT$, **c)** $E=5kT$.
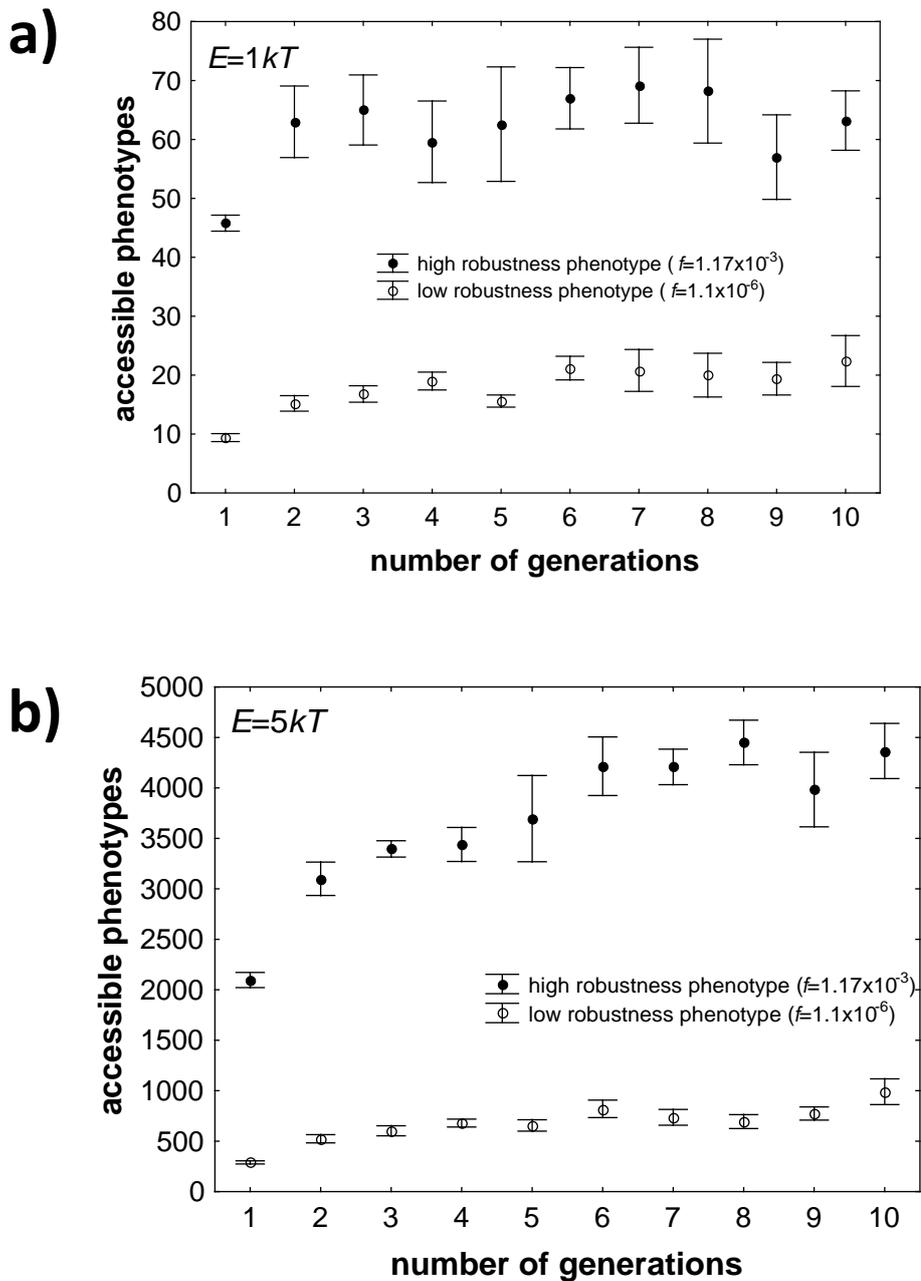
Figure S11: Phenotypic robustness allows evolving populations to access greater numbers of novel phenotypes through plasticity. The horizontal axes show time in generations (mutation-selection cycles). The vertical axes show the total number of unique phenotypes accessible through plasticity to at least one individual in an evolving population. **a)** $E=1kT$, **b)** $E=5kT$. All data are based on populations of size $N=100$ individuals, with a mutation rate of $\mu=1$ nucleotides per generation, where selection was required to preserve the native phenotype of each individual. Each population started out with identical genotypes of $L=30$ nucleotides that folded into the same native phenotype. The fraction of genotype space occupied by sequences folding into that phenotype is given by the value of f in the inset. Circles indicate means over five replicate populations starting from the same genotype. Whiskers indicate one standard error of the mean.
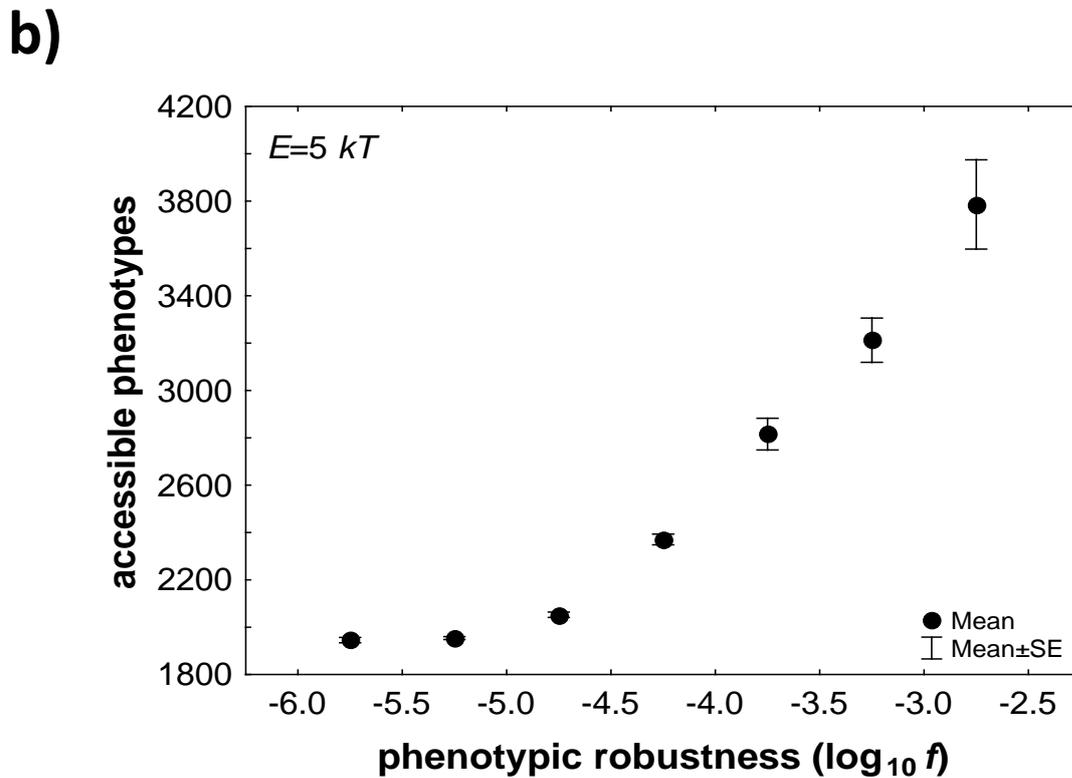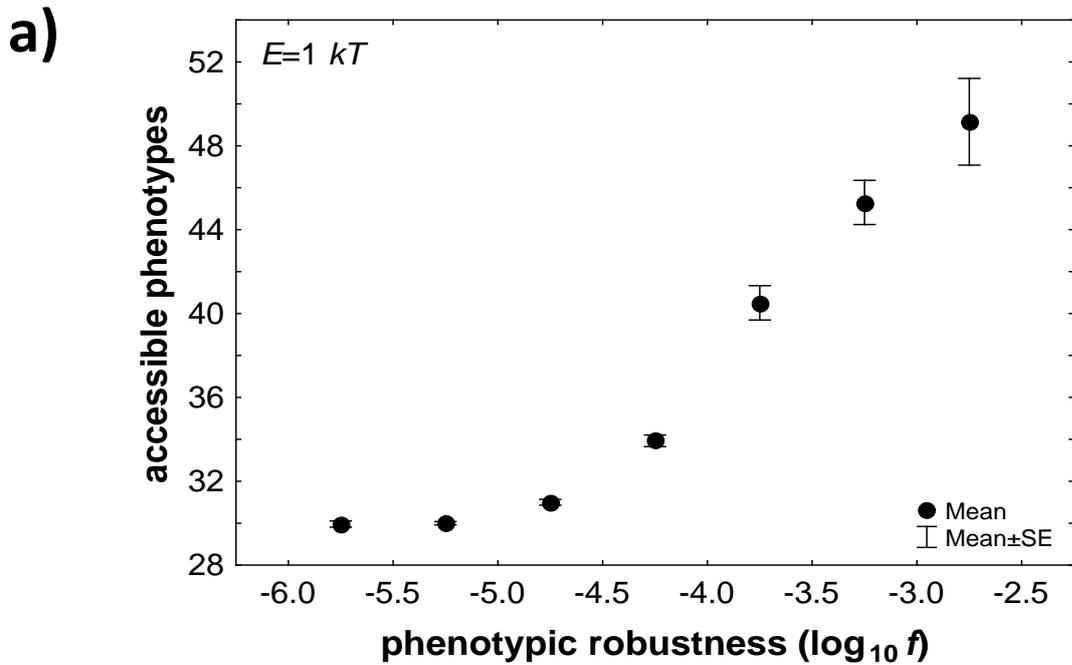
Figure S12: Phenotypic robustness allows evolving populations to access greater numbers of novel phenotypes through plasticity. The vertical axes show the total number of unique phenotypes accessible to at least one individual in a population through plasticity, where the population has evolved on the genotype network of a given phenotype for 10 generations. The horizontal axis shows phenotypic robustness ($\log_{10} f$), where f is the fraction of genotype space occupied by sequences folding into a given phenotype. **a)** $E=1kT$, **b)** $E=5kT$. Data are based on independent simulations for 3.8×10^4 different RNA phenotypes of $L=30$ nucleotides, and on at least three replicate populations per phenotype, where each population had $N=100$ members, and a mutation rate of $\mu=1$ nucleotides per generation. Each population started out with identical genotypes of length $L=30$ nucleotides.
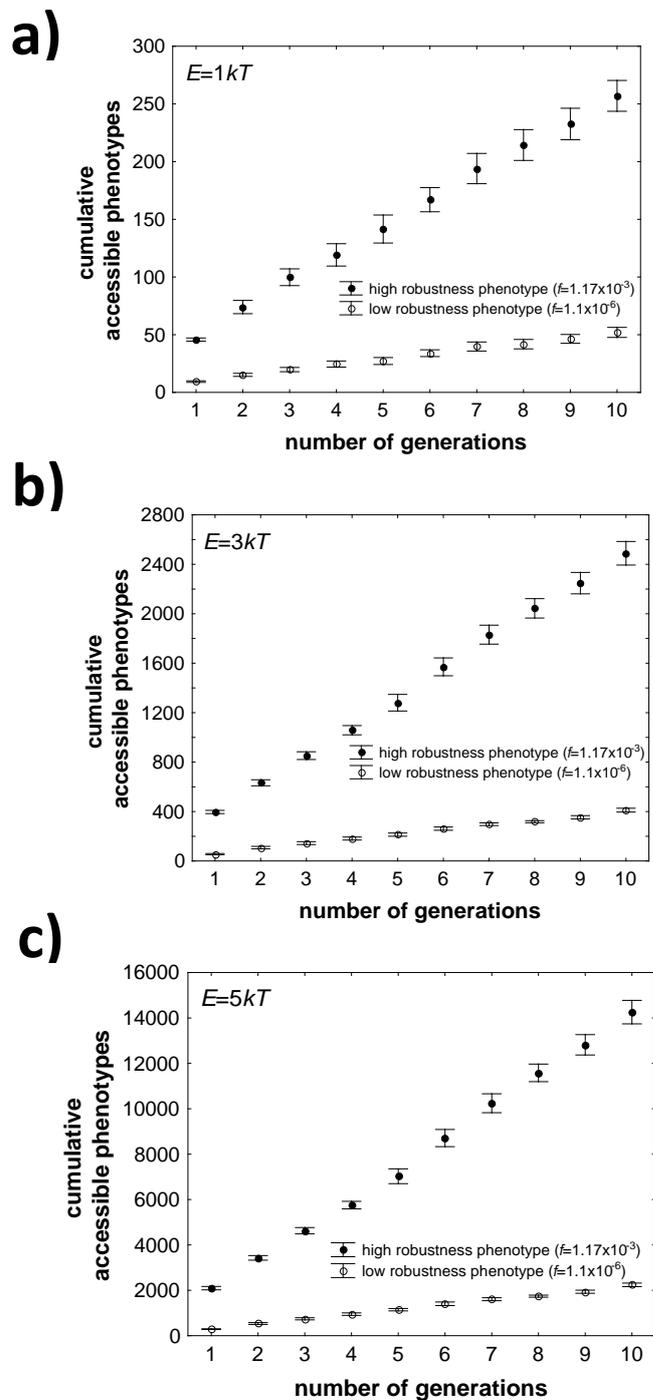
Figure S13: Phenotypic robustness entails access to greater cumulative numbers of novel phenotypes through plasticity. The horizontal axes show time in generations (mutation-selection cycles). The vertical axes show the total *cumulative* number of unique phenotypes accessible to a population through plasticity, i.e. the number of different accessible phenotypes since generation zero. **a)** *E=1kT*, **b)** *E=3kT*, **c)** *E=5kT*. All data are based on populations of size *N*=100 individuals, with a mutation rate of μ=1 nucleotides per generation, where selection was required to preserve the native phenotype of each individual. Each population started out with identical genotypes of *L*=30 nucleotides that folded into the same native phenotype. The fraction of genotype space occupied by sequences folding into that phenotype is given by the value of *f* in the inset. Circles indicate means over five replicate populations starting from the same genotype. Whiskers indicate one standard error of the mean.
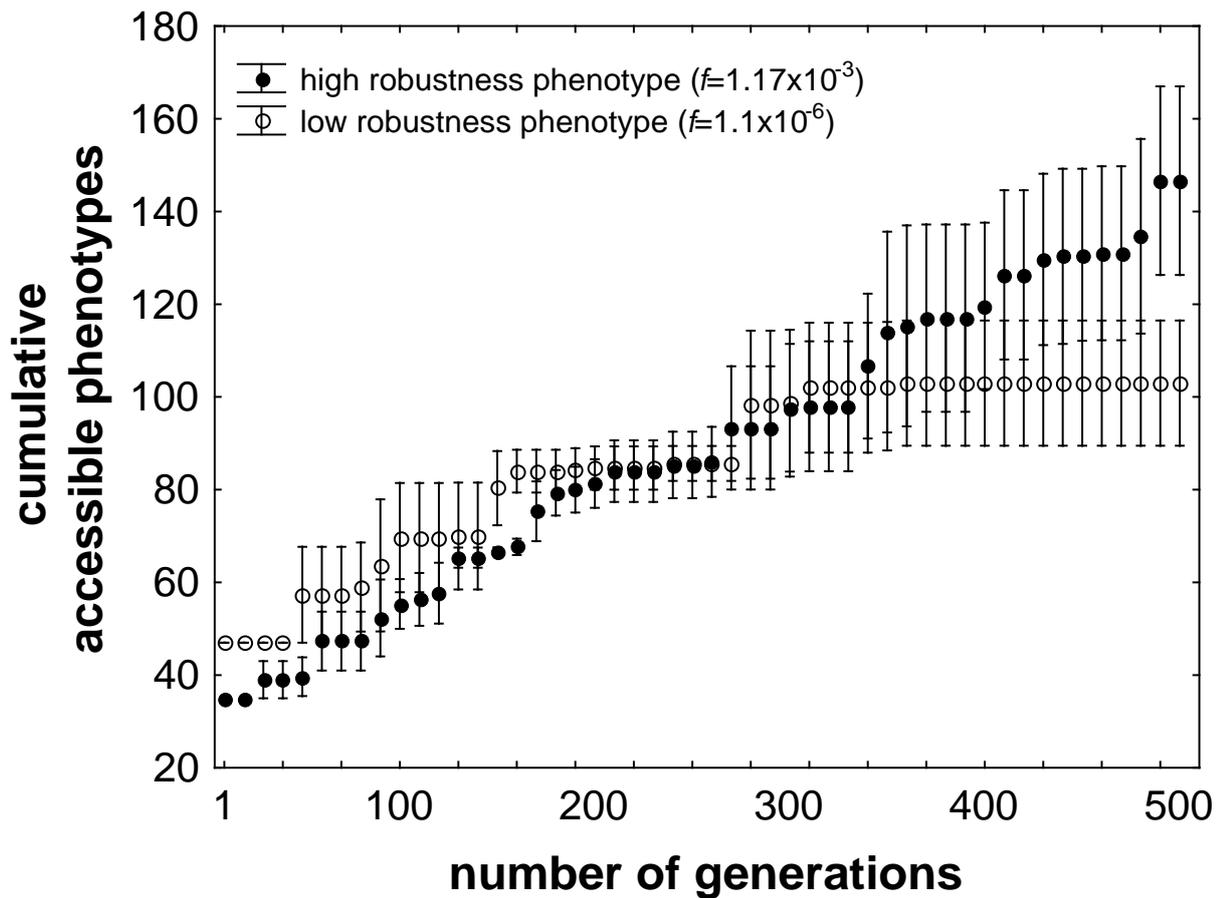
Figure S14: Phenotypic robustness entails access to greater numbers of novel phenotypes through plasticity also for populations with $N\mu \ll 1$. The horizontal axes show time in generations (mutation-selection cycles). The vertical axes show the total *cumulative* number of unique phenotypes accessible to a population through plasticity, i.e. the number of different accessible phenotypes since generation zero, at $E=3kT$. Data are based on populations of size $N=10$ individuals, with a mutation rate of $\mu=0.01$ nucleotides per generation. Each population started out with identical genotypes of length 30 nucleotides that folded into the same native phenotype. Data are shown for populations of two different phenotypes (closed and open circles), and for three replicate populations per phenotype. The fraction of genotype space occupied by sequences folding into each phenotype is given by the value of f in the inset. Circles indicate means over replicates, and whiskers indicate one standard error of the mean. Because populations like this are monomorphic most of the time, it is not informative to evaluate the number of phenotypes accessible to the population at any one time, because the population will usually consist only of a single individual, and the number of phenotypes accessible to that single individual will be dominated by the higher genotypic robustness of individuals with a highly robust phenotype. However, populations with a more robust phenotype will explore the genotype network of this phenotype more rapidly. Thus, over time, they can explore more novel phenotypes through plasticity. Notice that at generation 1, fewer phenotypes are accessible to the population with high phenotypic robustness, just as one might expect, because all individuals have identical genotypes and the number of accessible phenotypes is thus determined by their genotypic robustness. Over time, however, this population diversifies more rapidly, and thus gains access to greater number of phenotypes than the other population, as indicated by the crossing over of the traces of closed and open circles.
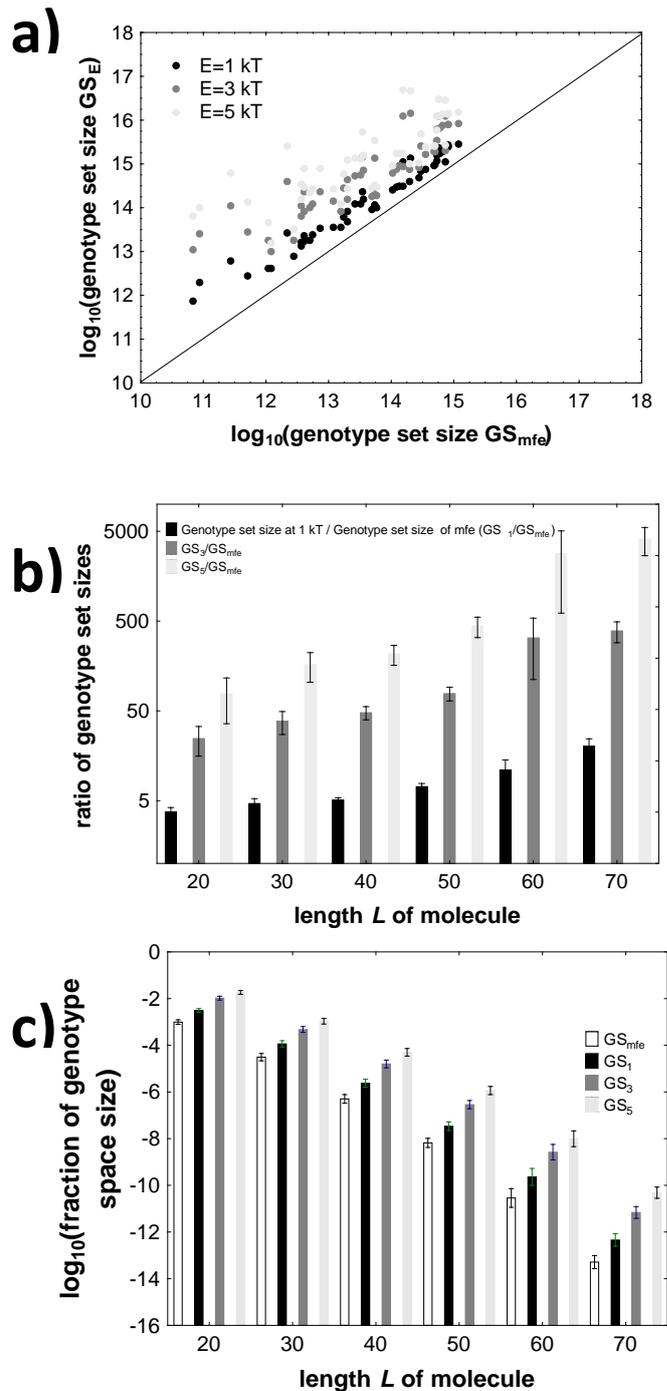
Figure S15: Genotype set sizes at different energy intervals E. For any one RNA phenotype *P*, let GS_{mfe} be the number of genotypes that adopt *P* as its minimum free energy structure. Similarly, let GS_E be the number of genotypes that contain *P* in its plastic repertoire at some energy interval *E* above the minimum free energy structure. **a)** The horizontal axis shows the decadic logarithm of GS_{mfe}, and the vertical axis shows the decadic logarithm of GS_E, for three different values of *E*, as indicated in the legend. The data are based on 50 random RNA genotypes of length 30 nucleotides, all of which form different minimum free energy phenotypes *P*. The solid line indicates equality of values on the horizontal and vertical axes. Note that genotype set sizes GS_E increase with increasing energy interval *E*. **b)** The ratio of GS_E/GS_{mfe} (vertical axis) for RNA molecules of various lengths (horizontal axis), and for different values of *E*, as indicated in the legend. Note the logarithmic scale on the vertical axis, which indicates that genotype set sizes increase approximately exponentially with increasing length of

a molecule and with increasing energy interval E. **c)** The vertical axis shows the logarithm of genotype set sizes GS_E expressed as a fraction of genotype space size 4^L, where L is the length of a molecule, i.e., it shows $\log_{10}(GS_E/4^L)$. The horizontal axis shows the length of a molecule L. Note that although genotype set size increases exponentially with length, it still occupies only a small fraction of genotype space, especially for longer molecules. Data in b) and c) are based on 50 random RNA molecules with different phenotypes for $20 \leq L \leq 50$, and for 20 random RNA molecules at $L=60, 70$. Bars and whiskers correspond to means and standard errors of the mean. GS_{mfe} and GS_E are estimated with an algorithm modified from (6), with three replicate estimates for each phenotype.

Supplementary Tables to

Mutational robustness accelerates the origin of novel RNA phenotypes through phenotypic plasticity

Andreas Wagner

andreas.wagner@ieu.uh.ch

Identifier	Organism	RNA type	Sequence	Secondary structure	Neutral set size
AF357483	Mus musculus	snmRNA	AAGCAAUUGUUUUACUUACAGUCUGGAGAA	...((((((.......)))))).)).....	1.50×10^{13}
AJ430257	Archaeoglobus fulgidus	snmRNA	CAUCCGUAUAGUAUGUGUAAGAAAGGGAGG	..(((...................))))..	5.76×10^{11}
Z71666	S. cerevisiae	snoRNA	AGGCGUGUAACAUUUAUUGGUUACAACAUG	.....((((((.......)))))).....	1.72×10^{11}
AB055777	Homo sapiens	noncoding transcript	CUCUUUUACCAAGGACCCGCCAACAUGGGC	.(((((....)))))((((......)))).	8.65×10^{12}
D28374	Homo sapiens	noncoding transcript	CCUUUUGGCUCUCUGACCAGCACCAUGGCG	((....))........(((......)))..	8.66×10^{12}
AJ564940	Homo sapiens	noncoding transcript	CGGAAAUCUGAAGGGAAAGUGGCAGGAAAG	......((((..(......))..))))....	3.70×10^{13}
X85345	Homo sapiens	noncoding transcript	ACAUCUCAUAGGUGCCUCUGACUGAGUUGC	.((.(((((((....))))..)))).)).	1.95×10^{12}
AF061109	Moneuplotes crassus	telomerase RNA	CCUCCAUUAUAUUCAAUUUGGGAAGGAUUGAAA	...........(((((((.....)))))))).	8.54×10^{15}
AJ430253	Archaeoglobus fulgidus	snmRNA	CAAACUUAACGAACGCUACGAGUUGAAAUCAGA	.....((((...........))))).......	9.02×10^{15}
S80603	Trypanosoma cruzi	gRNA	GAUAGAGAACGGAUGAGAAGGAUGUUUGUGAGA	........((((((.........))))))....	1.03×10^{16}
AB017710	Homo sapiens	snoRNA	UAAUCAAUGAUGAAACCUAUCCCGAAGCUGAUA	........((((.....))))..........	2.92×10^{16}
AB055770	Homo sapiens	noncoding transcript	CUUUCUGGUCUCGGCCGCAGAAGCGAGAUGACG	.....(.(((((..........))))).)..	5.34×10^{13}
AB056456	Homo sapiens	noncoding transcript	CUCCUCUCUCCGGUCCGUGCCUCCAAGAUGACA	.((.(((....(((....))).))).))..	1.27×10^{14}
D28386	Homo sapiens	noncoding transcript	CUUCCUUUCCAACUUAGACGCUGCAGAAUGGCU	...((.(((((.(......).)))..))).))..	1.87×10^{16}
S43366	Homo sapiens	noncoding transcript	GCCUUGAAGGGAGGUUUGUAGAAAACUGUAAGC	(((((.....)))))..................	1.43×10^{14}
AJ309206	Homo sapiens	noncoding transcript	CCCGAAACUCUGGAUUCCGGGGCCCUUCACAGC	.......((((((...))))))..........	1.45×10^{14}
AF380696	Leishmania tarentolae	gRNA	AAACAAACAGUGCACGGGGUCGGAUGUUAAACACA	.....(((..(.((...))).)))).....	1.11×10^{14}
AJ430256	Archaeoglobus fulgidus	snmRNA	GCGAGAGUUUAGGCGGCUUGUAAGAAAGGGAGGCU	((..........))(((((.(......).)))))	2.28×10^{16}
AJ544037	Drosophila melanogaster	snmRNA	GCCCGAAGAAGCCGAAAGCUUGGAUGGCUGGGAUG	.((((......(((....))))......)))...	5.37×10^{15}
AJ544038	Drosophila melanogaster	snmRNA	CAUUUGGAGAACGAAAGAAGACCAACAAGUGCGCG	(((((..(..(......)..).)))))))....	2.36×10^{14}
M11772	Homo sapiens	snRNA	ACCCAAUUUUUUGAGGCCUUGCUUUGGCAAGGCUA	.............(((((((....))))))).	1.28×10^{17}
AB055765	Homo sapiens	noncoding transcript	CUUCUUCCUUCUCGCCUAACGCUGCCAACAUGGUG	............(((.................))))	3.51×10^{17}
D28349	Homo sapiens	noncoding transcript	UUCCUUUUUUCUUUUUUCCGGCGUUCAAGAUGUCG	.................((((((...)))))))	9.73×10^{16}
AJ430248	Archaeoglobus fulgidus	snmRNA	GCCGAAUCGAUCAGAAGAGAACUGACAAGUUGAAAUCAGA	......(((((((.......))))..))).......	1.12×10^{18}
AJ430294	Archaeoglobus fulgidus	snmRNA	CAACGCUGCAACAAGAGUGGUUUUACCGUGAUCAACGUGA	...(((((........))))......(((((.....))).).	4.45×10^{15}
L25589	Trypanosoma brucei	gRNA	GACACAAUGAAGAAAUGCACUCUAAAAUGCACAGUGAUUA	..............((((........))))..........	1.17×10^{20}
L25590	Trypanosoma brucei	gRNA	UAAAAGACAAUGUAGAUUUCUGAGUAAUAGGGAGGAUAAC	...........((..(((((.......))))).)))..	9.13×10^{17}
D28454	Homo sapiens	noncoding transcript	CUUUCCGGCGGUGACGACCUACGCACACGAGAACAUGCCU	.......((((((.(.....))))).))...........	9.19×10^{15}
X99684	Homo sapiens	noncoding transcript	CAGGCCGCUGUCCUCCUGCUCCUACCCACCUAGUGUCAAG	.(((..((.........)).)))...............	7.28×10^{17}
AF036740	Schistosoma mansoni	ribozyme	AUCCAGCUCACGAGUCCCAAAUAGGACGAAACGCGUCCUCCAU	....................((((((.....))))))....	5.84×10^{19}

Table S1: The 30 biological RNA molecules used in this study. Column 1 ("Identifier") contains database identifiers from (http://www.ncrna.org/frnadb,1). Secondary structures are predicted minimum free energy structures using the fold routine of the Vienna RNA package package (http://www.tbi.univie.ac.at/~ivo/RNA/) (7) with default parameters. RNA molecules are ordered by increasing length, ranging from 30 to 43 nucleotides. Estimates of neutral set sizes are from (6).

Identifier	Organism	RNA type	Phenotypic plasticity			Median phenotypic plasticity		
			1kT	3kT	5kT	1kT	3kT	5kT
AF357483	Mus musculus	snmRNA	2	20	145	4.0	26.0	149.0
AJ430257	Archaeoglobus fulgidus	snmRNA	8	71	461	5.5	50.0	327.0
Z71666	S. cerevisiae	snoRNA	1	13	84	2.0	18.5	89.5
AB055777	Homo sapiens	noncoding transcript	2	34	216	3.0	16.0	97.0
D28374	Homo sapiens	noncoding transcript	7	59	346	5.0	47.5	289.5
AJ564940	Homo sapiens	noncoding transcript	2	11	68	4.0	41.5	244.0
X85345	Homo sapiens	noncoding transcript	6	40	199	4.0	33.5	175.0
AF061109	Moneuplotes crassus	telomerase RNA	1	19	150	2.0	9.0	52.0
AJ430253	Archaeoglobus fulgidus	snmRNA	3	46	339	5.0	46.5	286.0
S80603	Trypanosoma cruzi	gRNA	8	55	359	4.0	32.0	192.5
AB017710	Homo sapiens	snoRNA	6	61	421	5.0	52.0	343.5
AB055770	Homo sapiens	noncoding transcript	2	34	216	5.0	32.0	188.5
AB056456	Homo sapiens	noncoding transcript	1	27	158	4.0	52.0	373.5
D28386	Homo sapiens	noncoding transcript	7	54	396	6.0	74.0	499.0
S43366	Homo sapiens	noncoding transcript	4	41	260	4.0	38.5	247.5
AJ309206	Homo sapiens	noncoding transcript	2	23	132	4.0	38.5	220.5
AF380696	Leishmania tarentolae	gRNA	5	118	717	6.0	72.5	471.5
AJ430256	Archaeoglobus fulgidus	snmRNA	13	102	628	8.0	70.0	454.5
AJ544037	Drosophila melanogaster	snmRNA	10	76	363	5.0	52.5	329.0
AJ544038	Drosophila melanogaster	snmRNA	4	20	93	5.0	42.5	245.5
M11772	Homo sapiens	snRNA	6	14	89	2.0	12.5	81.5
AB055765	Homo sapiens	noncoding transcript	5	34	207	5.0	70.5	507.0
D28349	Homo sapiens	noncoding transcript	2	5	18	4.0	38.0	230.5
AJ430248	Archaeoglobus fulgidus	snmRNA	10	87	629	6.0	82.5	635.5
AJ430294	Archaeoglobus fulgidus	snmRNA	5	66	681	7.0	103.5	760.0
L25589	Trypanosoma brucei	gRNA	3	46	380	6.0	85.0	683.5
L25590	Trypanosoma brucei	gRNA	5	53	386	4.0	47.0	371.5
D28454	Homo sapiens	noncoding transcript	12	172	1046	5.5	90.5	695.5
X99684	Homo sapiens	noncoding transcript	14	144	1296	5.0	83.5	775.0
AF036740	Schistosoma mansoni	ribozyme	3	18	109	5.0	71.5	604.0

Table S2: Phenotypic plasticity for 30 biological RNA molecules used in this study, as well as median plasticity for 100 RNA molecules inversely folded from each of the 30 biological RNA phenotypes. Column 1 ("Identifier") contains database identifiers from (http://www.ncrna.org/frnadb, ref 1). Phenotypic plasticity (columns 4,5,6) refers to the number of secondary structures (including the mfe structure) found within a given energy interval E (in units of kT) above the mfe of the biological RNA sequence in column 1. Median phenotypic plasticity (columns 7,8,9) refers to the median number of secondary structures found within a given energy interval E above the mfe, where the median is taken over 100 inversely folded sequences with the same predicted secondary structure phenotype as the sequence in column 1. RNA molecules (rows) are ordered by increasing length, ranging from 30 to 43 nucleotides (see also Table S1).

Identifier	Organism	RNA type	Phenotypic diversity			Median phenotypic diversity		
			1kT	3kT	5kT	1kT	3kT	5kT
AF357483	Mus musculus	snmRNA	4	11.2	12.8	4.5	8.7	10.6
AJ430257	Archaeoglobus fulgidus	snmRNA	5.8	9.1	10.0	7.7	9.4	10.2
Z71666	S. cerevisiae	snoRNA	NA	4.3	7.5	4.0	7.8	8.9
AB055777	Homo sapiens	noncoding transcript	2	7.8	11.8	6.0	9.3	11.1
D28374	Homo sapiens	noncoding transcript	10.9	11.4	12.1	9.3	10.8	11.5
AJ564940	Homo sapiens	noncoding transcript	2	8.0	8.4	8.0	9.9	10.9
X85345	Homo sapiens	noncoding transcript	7.1	8.9	10.9	4.4	8.6	10.3
AF061109	Moneuplotes crassus	telomerase RNA	ND	14.0	12.7	2.0	5.2	7.2
AJ430253	Archaeoglobus fulgidus	snmRNA	8.7	9.5	11.1	7.5	10.2	11.5
S80603	Trypanosoma cruzi	gRNA	8.9	9.6	10.2	5.0	9.6	11.3
AB017710	Homo sapiens	snoRNA	10.9	10.7	11.7	9.0	10.8	12.2
AB055770	Homo sapiens	noncoding transcript	2	7.8	11.8	3.9	9.2	11.3
AB056456	Homo sapiens	noncoding transcript	ND	9.1	10.2	8.8	11.6	12.8
D28386	Homo sapiens	noncoding transcript	10.0	10.3	11.4	8.6	11.0	12.5
S43366	Homo sapiens	noncoding transcript	11.5	12.8	13.2	9.7	11.6	12.6
AJ309206	Homo sapiens	noncoding transcript	6	11.3	11.3	6.3	9.4	10.9
AF380696	Leishmania tarentolae	gRNA	8.2	11.6	11.9	8.6	11.2	12.3
AJ430256	Archaeoglobus fulgidus	snmRNA	10.0	11.9	12.8	8.1	12.5	13.9
AJ544037	Drosophila melanogaster	snmRNA	9.6	10.0	11.6	8.9	11.5	12.9
AJ544038	Drosophila melanogaster	snmRNA	2.7	8.4	10.3	5.3	9.8	11.7
M11772	Homo sapiens	snRNA	6.1	5.5	7.0	4.0	6.4	8.3
AB055765	Homo sapiens	noncoding transcript	6.2	8.0	9.2	9.3	11.6	12.8
D28349	Homo sapiens	noncoding transcript	2	3.2	5.3	6.3	10.6	12.3
AJ430248	Archaeoglobus fulgidus	snmRNA	13.0	14.8	15.2	10.9	13.5	14.9
AJ430294	Archaeoglobus fulgidus	snmRNA	6.8	10.5	16.6	9.8	13.9	15.5
L25589	Trypanosoma brucei	gRNA	6.7	11.1	13.1	11.7	13.1	14.4
L25590	Trypanosoma brucei	gRNA	6.6	9.1	10.7	9.7	13.6	15.1
D28454	Homo sapiens	noncoding transcript	14.0	14.4	14.9	11.3	13.4	14.9
X99684	Homo sapiens	noncoding transcript	12.9	14.7	15.5	11.3	13.9	14.8
AF036740	Schistosoma mansoni	ribozyme	5.3	6.9	9.6	10.3	13.3	15.4

Table S3: Phenotypic diversity of the plastic repertoire for 30 biological RNA molecules used in this study, as well as median diversity of the plastic repertoire for 100 RNA molecules inversely folded from each of the 30 biological RNA phenotypes. Column 1 ("Identifier") contains database identifiers from (http://www.ncrna.org/frnadb, 1). Phenotypic diversity (columns 4,5,6) refers to the average pairwise Hamming distance between the dot-parenthesis representations of all RNA secondary structures (including the mfe structure) found within a given energy interval E (in units of kT) above the mfe of the biological RNA sequence in column 1. Median phenotypic diversity (columns 7,8,9) refers to the Hamming distance among secondary structures found within a given energy interval E above the mfe, where the median is taken over 100 inversely folded sequences with the same predicted secondary structure phenotype as the sequence in column 1. 'ND' indicates that pairwise distance could not be determined, because only one secondary structure occurs in the given energy interval. The RNA molecules are ordered by increasing length, ranging from 30 to 43 nucleotides (see also Table S1)

Supporting References

1. Kin T, Yamada K, Terai G, Okida H, Yoshinari Y, Ono Y, Kojima A, Kimura Y, Komori T, and A. K. 2007. fRNAdb: a platform for mining/annotating functional RNA candidates from non-coding RNA sequences. Nucleic Acids Research (Database Issue) 35 D145-D148.

2. Riley, G. R., R. A. Corell, and K. Stuart. 1994. Multiple guide RNAs for identical editing of *Trypanosoma brucei* apocytochrome b mRNA have an unusual minicircle location and are developmentally regulated. Journal of Biological Chemistry 269:6101-6108.

3. Benson, D. A., M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and E. W. Sayers. 2013. GenBank. Nucleic Acids Research 41:D36-D42.

4. Schuster, P., W. Fontana, P. Stadler, and I. Hofacker. 1994. From sequences to shapes and back - a case-study in RNA secondary structures. Proceedings of the Royal Society of London Series B 255:279-284.

5. Freeman, S., and J. C. Herron. 2007. Evolution (4th ed). Pearson, San Francisco, CA.

6. Jörg, T., O. Martin, and A. Wagner. 2008. Neutral network sizes of biological RNA molecules can be computed and are atypically large. BMC Bioinformatics 9:464.

7. Hofacker, I. L., W. Fontana, P. F. Stadler, L. S. Bonhoeffer, M. Tacker, and P. Schuster. 1994. Fast folding and comparison of RNA secondary structures. Monatshefte für Chemie 125:167-188.