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 10^{th} and 90^{th} 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

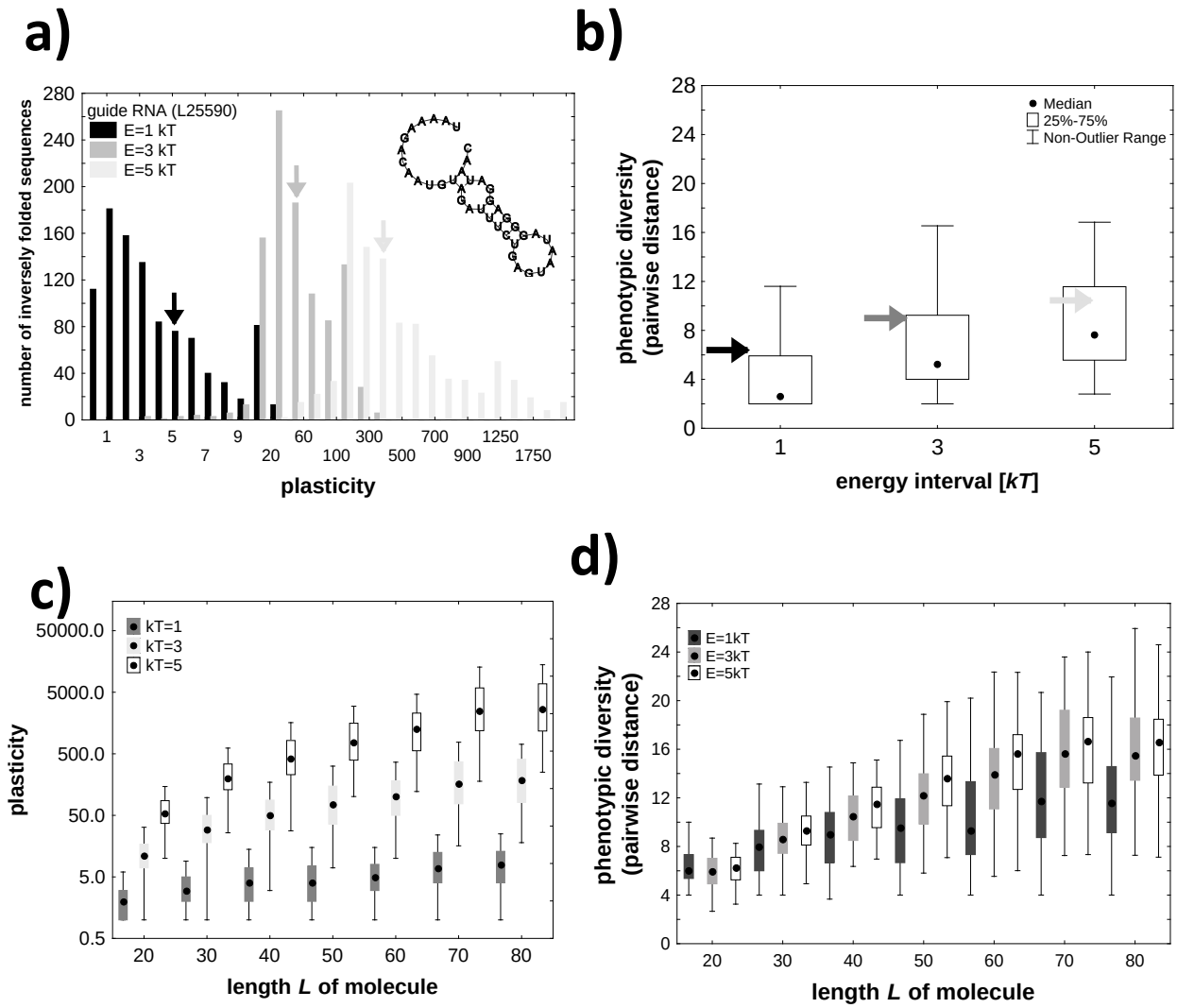

Figure S1. Biological and random RNA molecules have similar phenotypic plasticity. a) Distribution of phenotypic plasticity, that is, the number of phenotypes (secondary structures) in a given energy interval *E* above the mfe, as greyscale-coded in the left inset. Histograms are based on 1000 inversely 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).

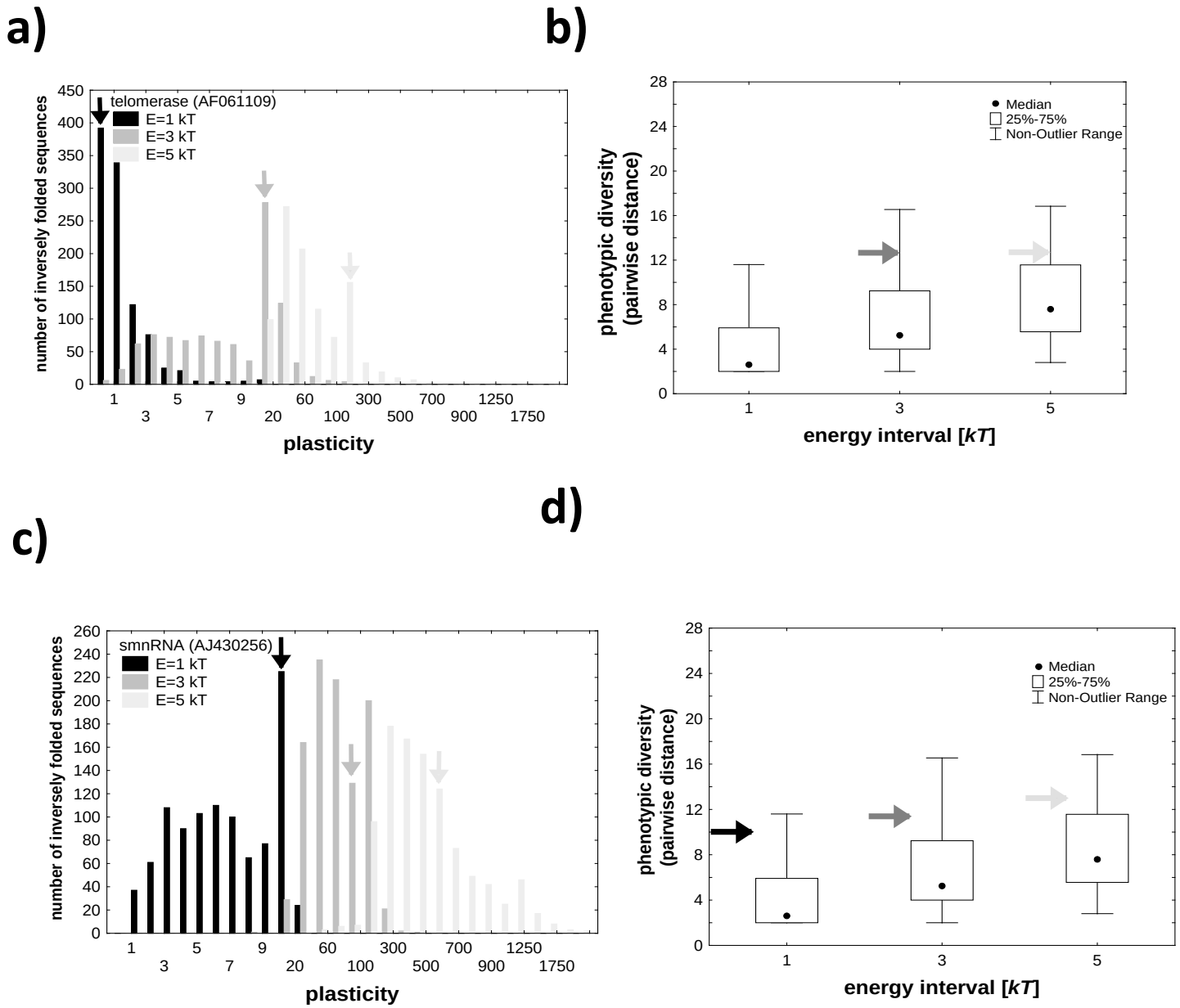

Figure S2: Phenotypic plasticity of two biological RNA molecules. a), c) Distribution of phenotypic plasticity, that is, the number of secondary structure phenotypes in a given energy interval *E* above the mfe, as greyscale-coded in the left inset. Histograms are based on 1000 inversely folded sequences 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=1kT. 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).

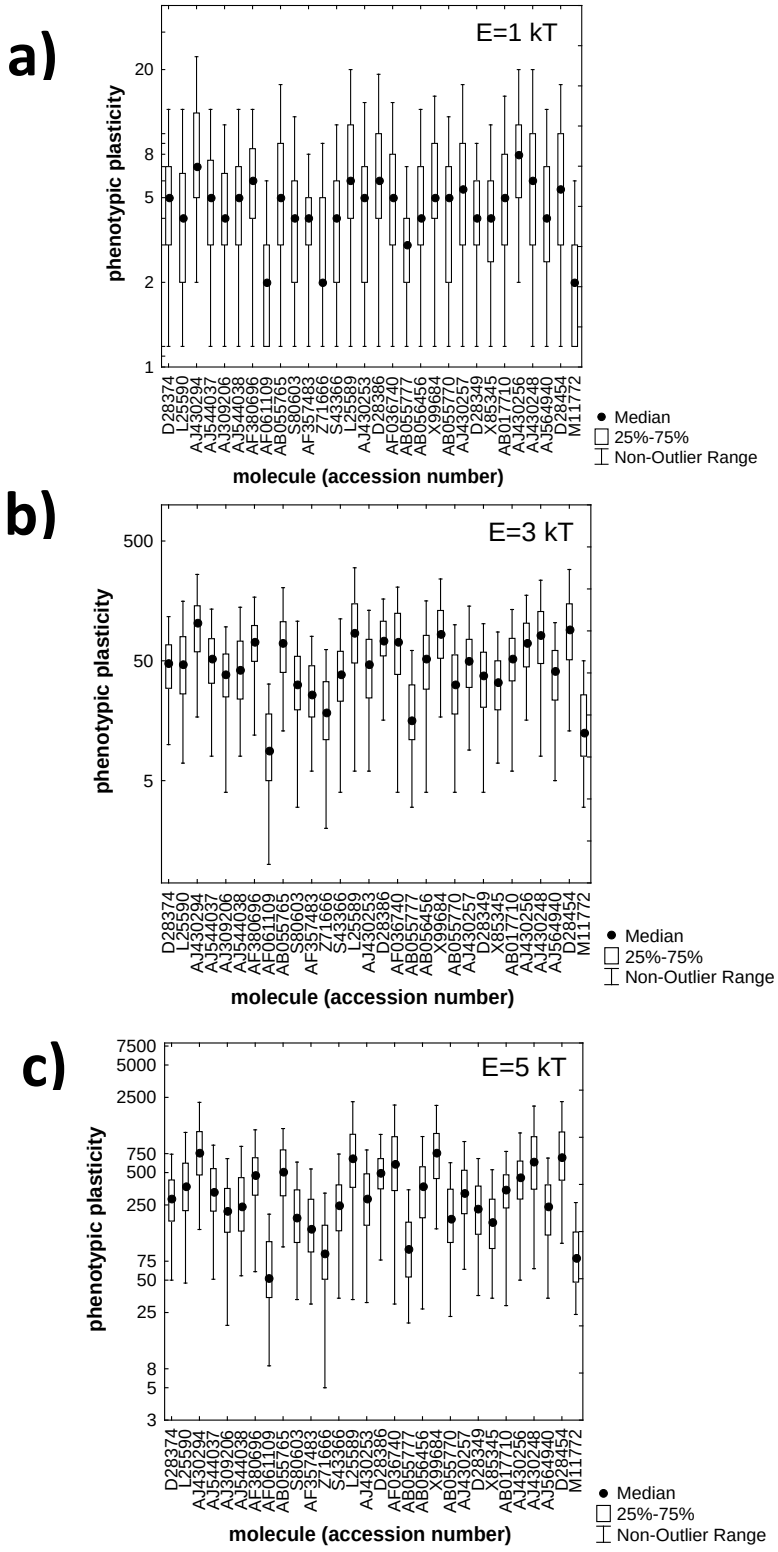

Figure S3. Phenotypic plasticity of 30 different biological RNA molecules. The horizontal axes show the genbank accession numbers of 30 biological RNA molecules from Table S1. The vertical axes show box plots of the distribution of phenotypic plasticity for each molecule. Specifically, data in each box 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 25^{th} to 75^{th} percentile of the distribution, and whiskers indicate the non-outlier range (see Methods).

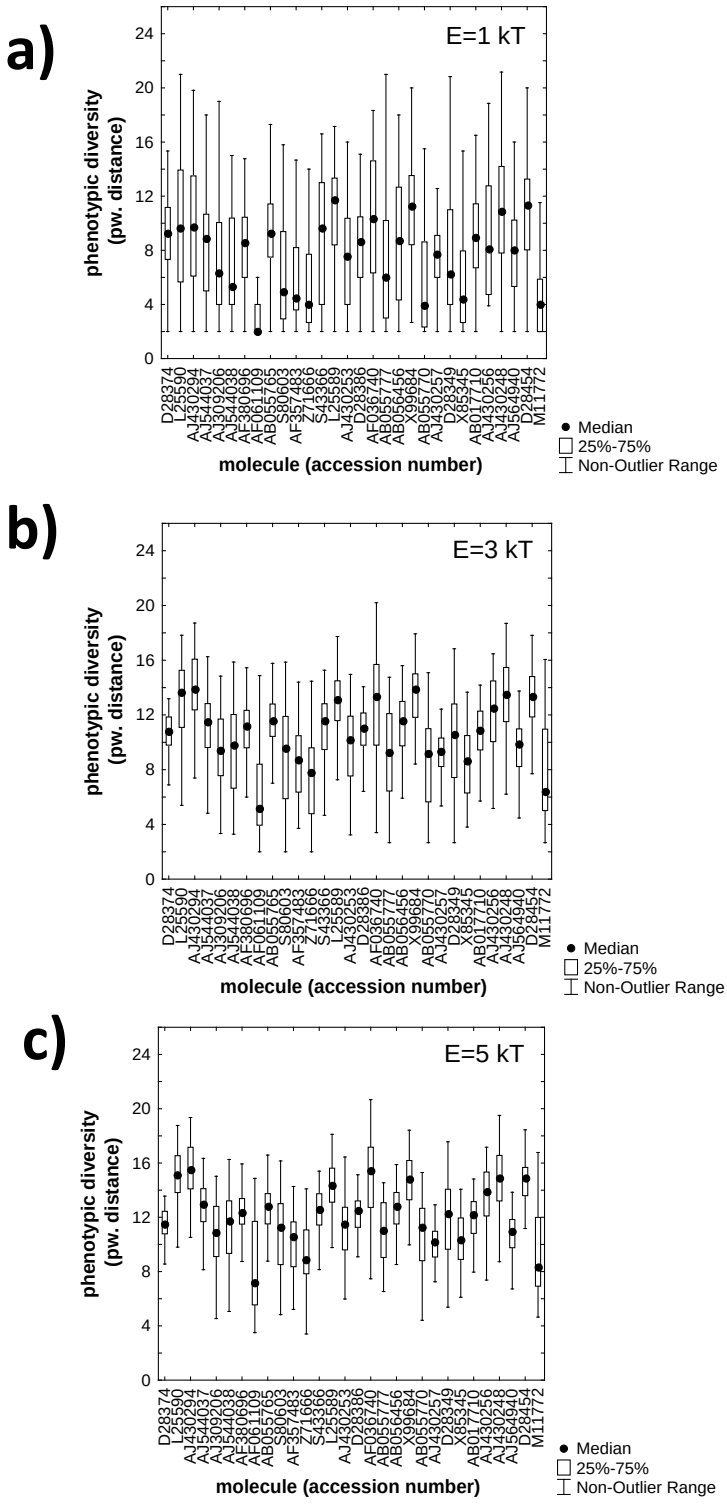

Figure S4. Phenotypic diversity of the plastic repertoire of 30 different biological RNA molecules. The horizontal axes show the genbank accession numbers of 30 biological RNA molecules from table S2. The vertical axes show box plots of the pairwise structural diversity of the plastic repertoire of each RNA for different energy intervals. **a)** *E=1kT*, **b)** *E*=3*kT*, and **c)** *E*=5kT. 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).

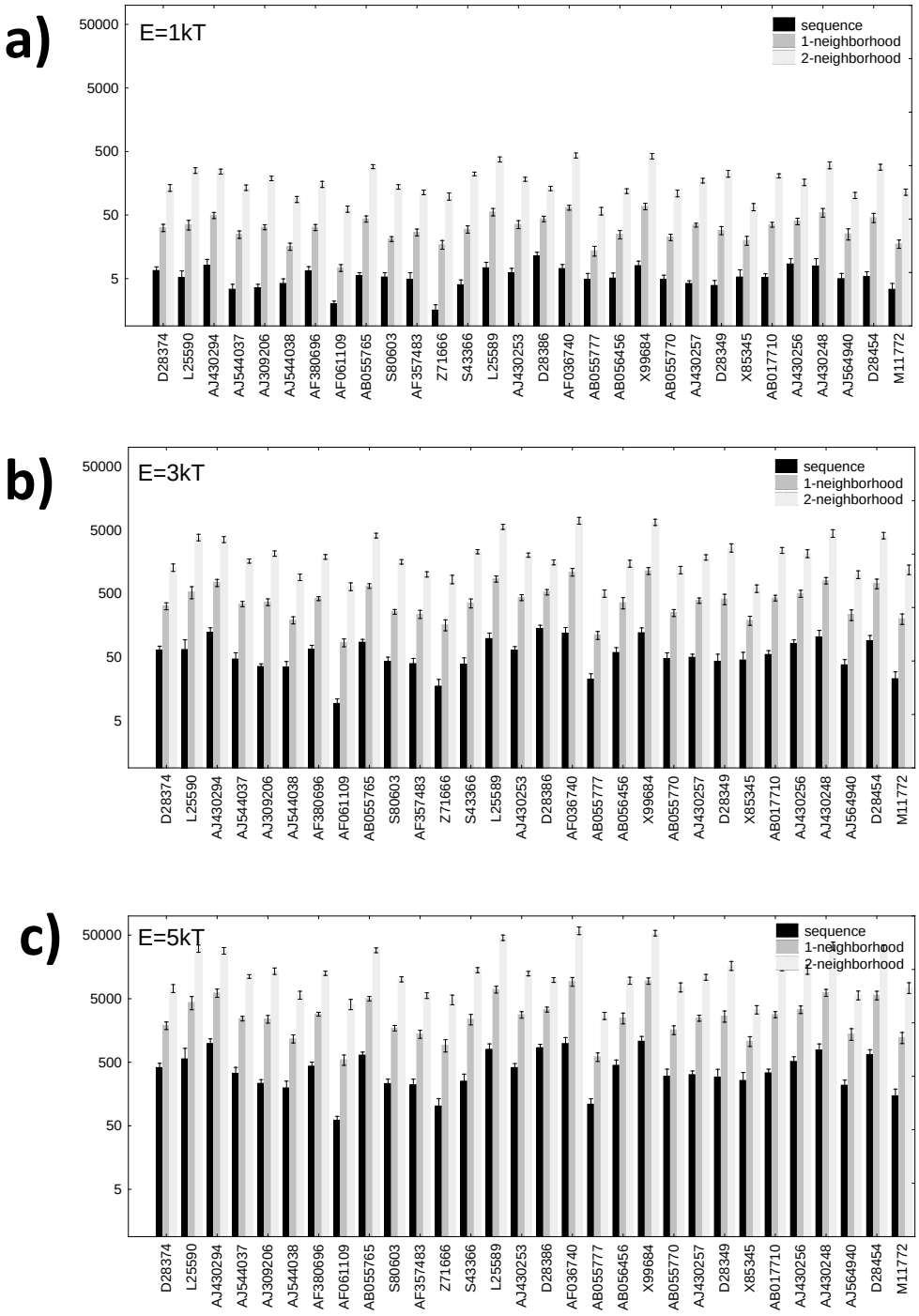

Figure S5. Even small neutral neighborhoods of a sequence render many novel phenotypes accessible through plasticity. The horizontal axes show genbank (5) accession numbers of 30 biological RNA sequences shown in Table S1. Above each accession number is a group of three bars. The black bar in each group of bars indicates the mean number of *unique* phenotypes, i.e., phenotypes different from one another, in the plastic repertoire $\boldsymbol{P_0}$ 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 $\boldsymbol{P_0}$ and $\boldsymbol{P_1}$, where $\boldsymbol{P_1}$ 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 $\boldsymbol{P_0}$, $\boldsymbol{P_1}$, and $\boldsymbol{P_2}$,, where $\boldsymbol{P_2}$ 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.

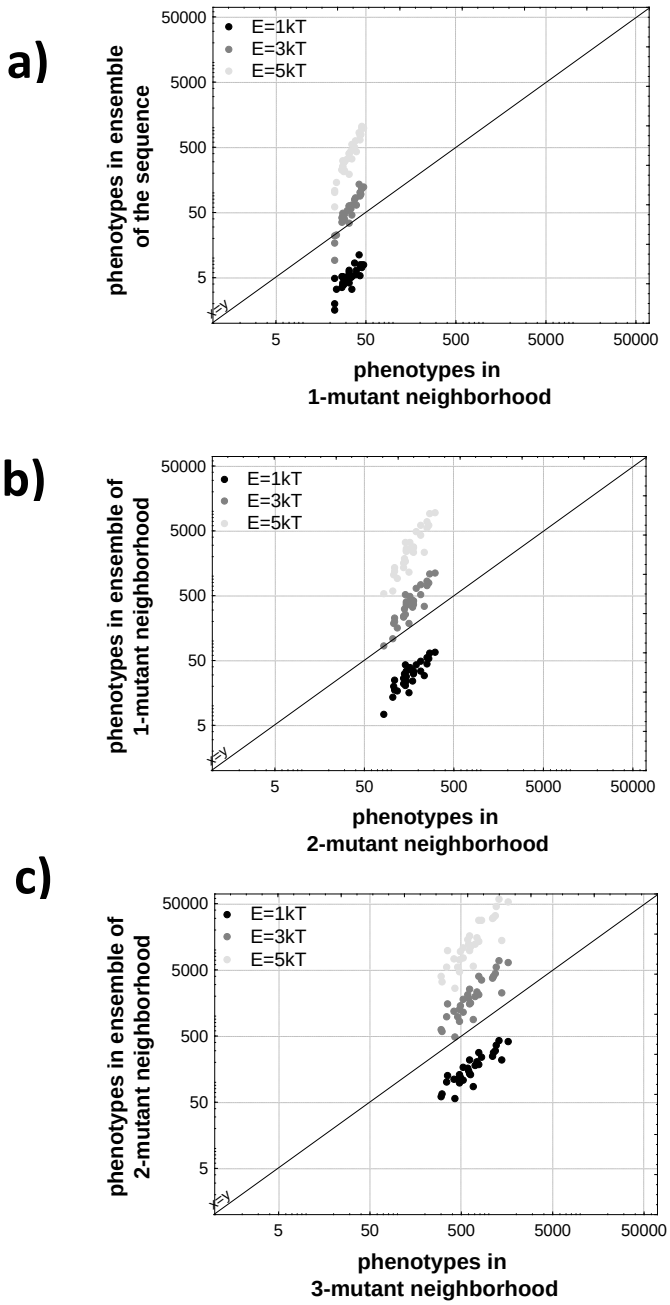

Figure S6: Phenotypic plasticity (vertical axes) can produce more novel phenotypes than mutation (horizontal axes) for energy intervals $E>1kT$. Data in these figures are based on the 30 biological RNA sequences of Table S1. Specifically, each data point is an average over 10 inversely folded sequences 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*.

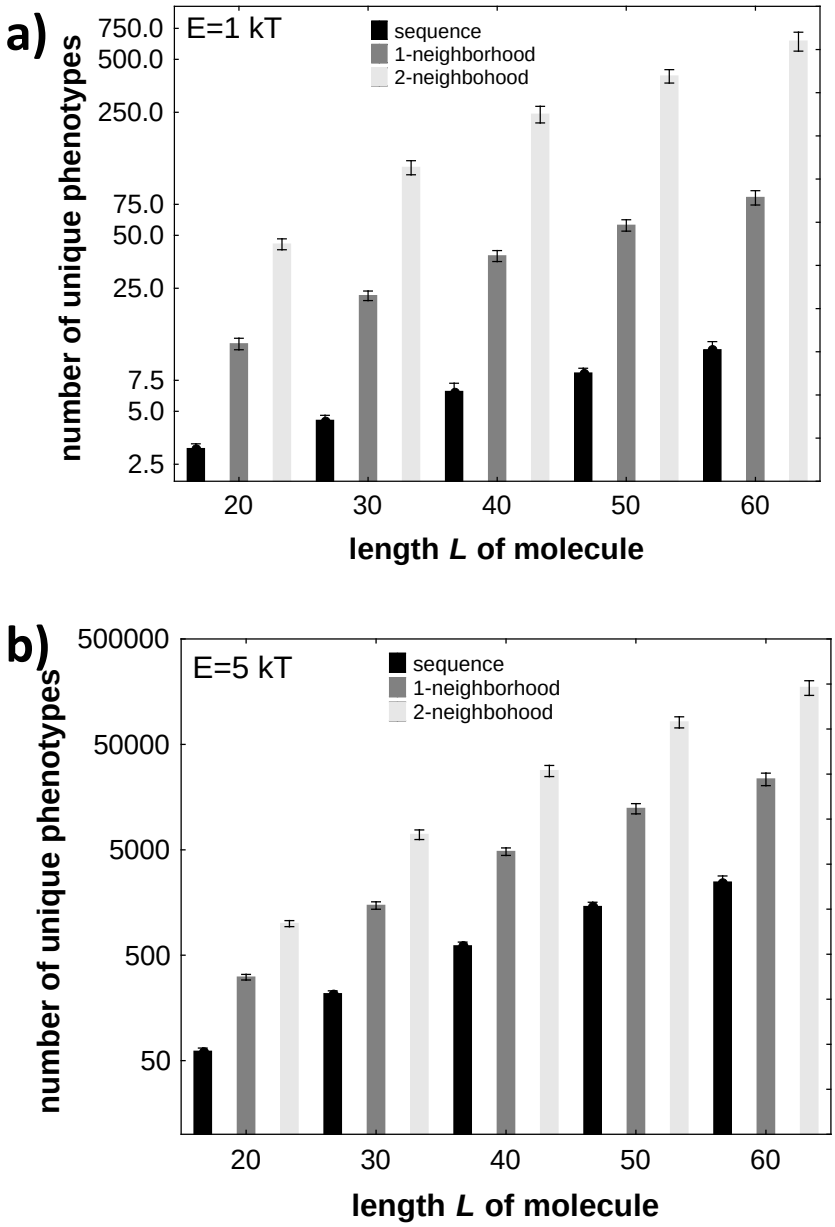

Figure S7: Even small neutral neighborhoods of a sequence render many novel phenotypes accessible through plasticity (random sequences in the energy interval E=1kT and E=5kT). The horizontal axis shows the lengths of the RNA molecules considered. For each length value with the exception of length L=60 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 $\boldsymbol{P_0}$ 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 $\boldsymbol{P_0}$ and $\boldsymbol{P_1}$, where $\boldsymbol{P_1}$ 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 $\boldsymbol{P_0}$, $\boldsymbol{P_1}$, and $\boldsymbol{P_2}$,, where $\boldsymbol{P_2}$ 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**.**

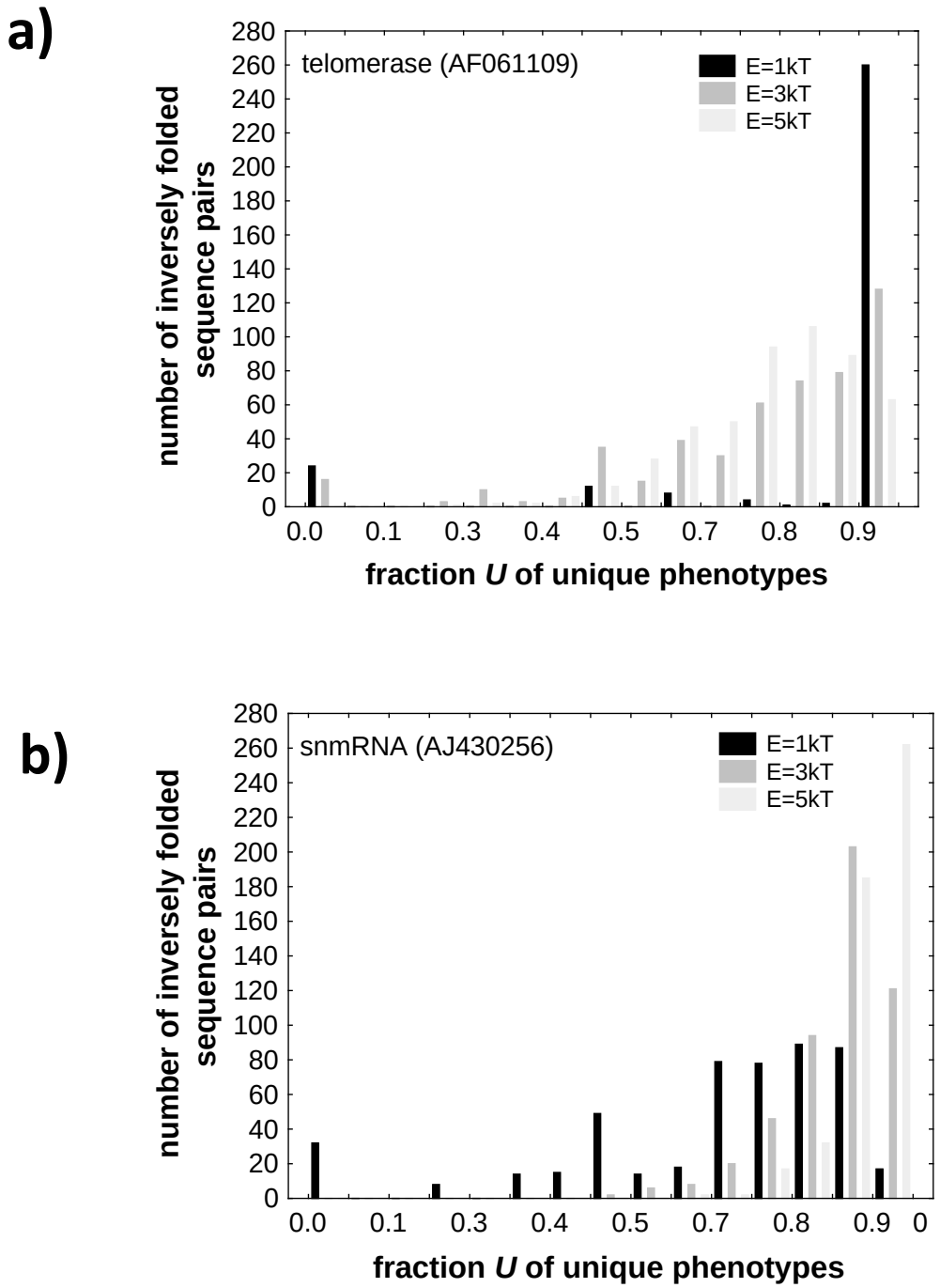

Figure S8. Different genotypes with the same native phenotype differ greatly in the phenotypes of their plastic repertoire. The figure shows data about the plastic repertoires of two genotypes G_1 and G_2 with the same native phenotype P. Specifically, the figure displays the fraction U of phenotypes 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.

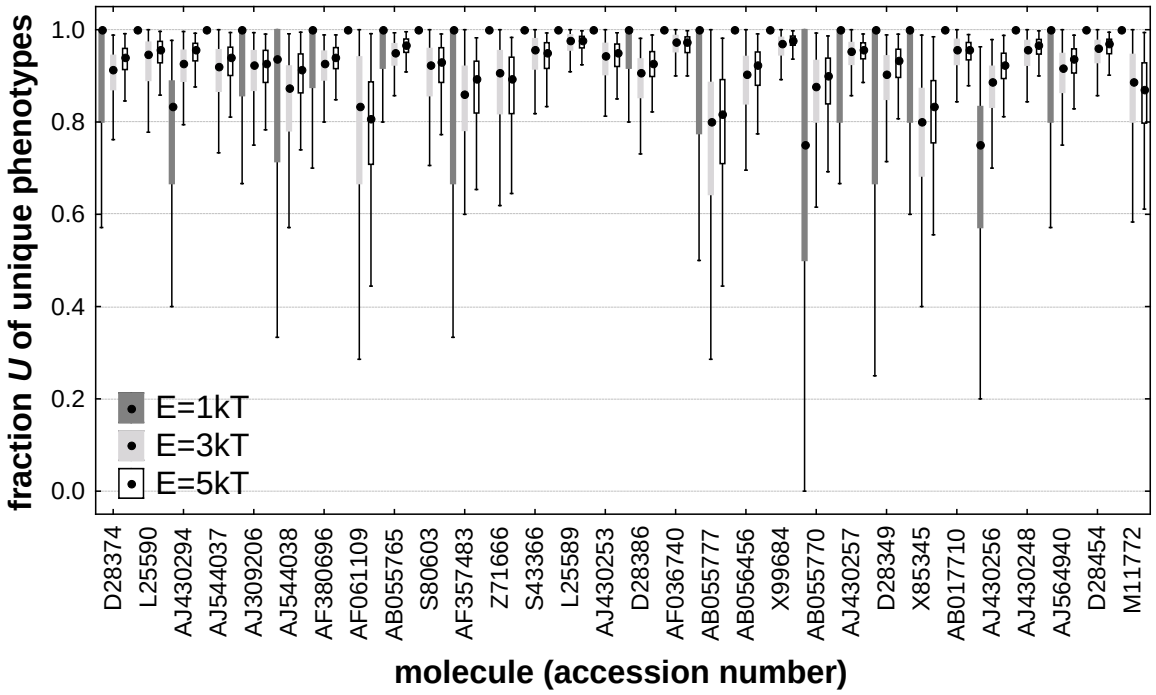

Figure S9. Different genotypes with the same native phenotype differ greatly in their plastic repertoire. The figure shows data about the plastic repertoires of two genotypes G_1 and G_2 with the same native phenotype P. Specifically, the figure displays the fraction U of phenotypes that are in the plastic 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 25^{th} to 75^{th} 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.

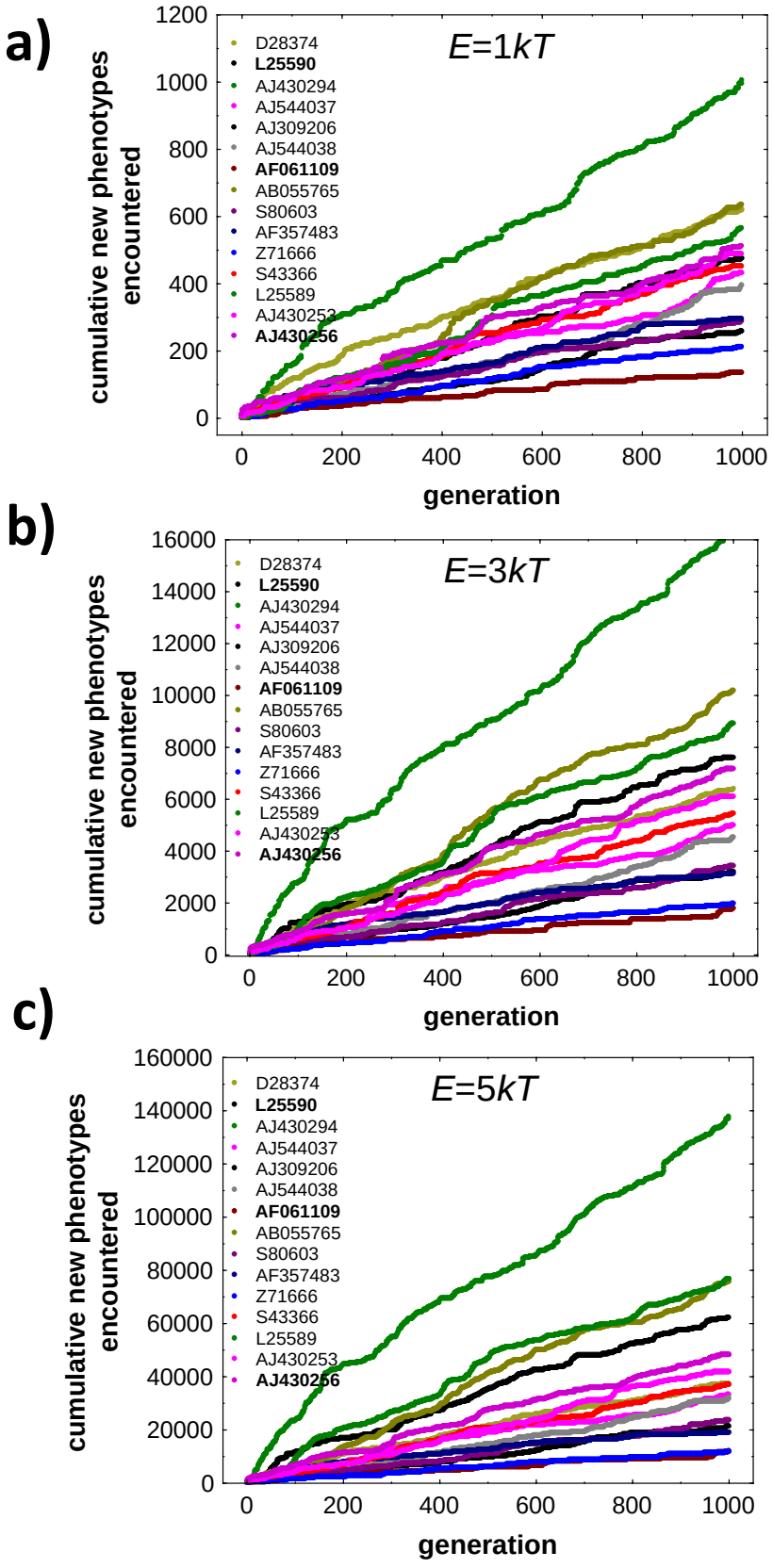

Figure S10: New accessible phenotypes accumulate at a non-decreasing rate for 15 biological RNA molecules. The horizontal axes show the number of steps along a random walk of 1000 steps on a genotype network. The vertical axes show the cumulative number of phenotypes that become accessible through phenotypic 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*.

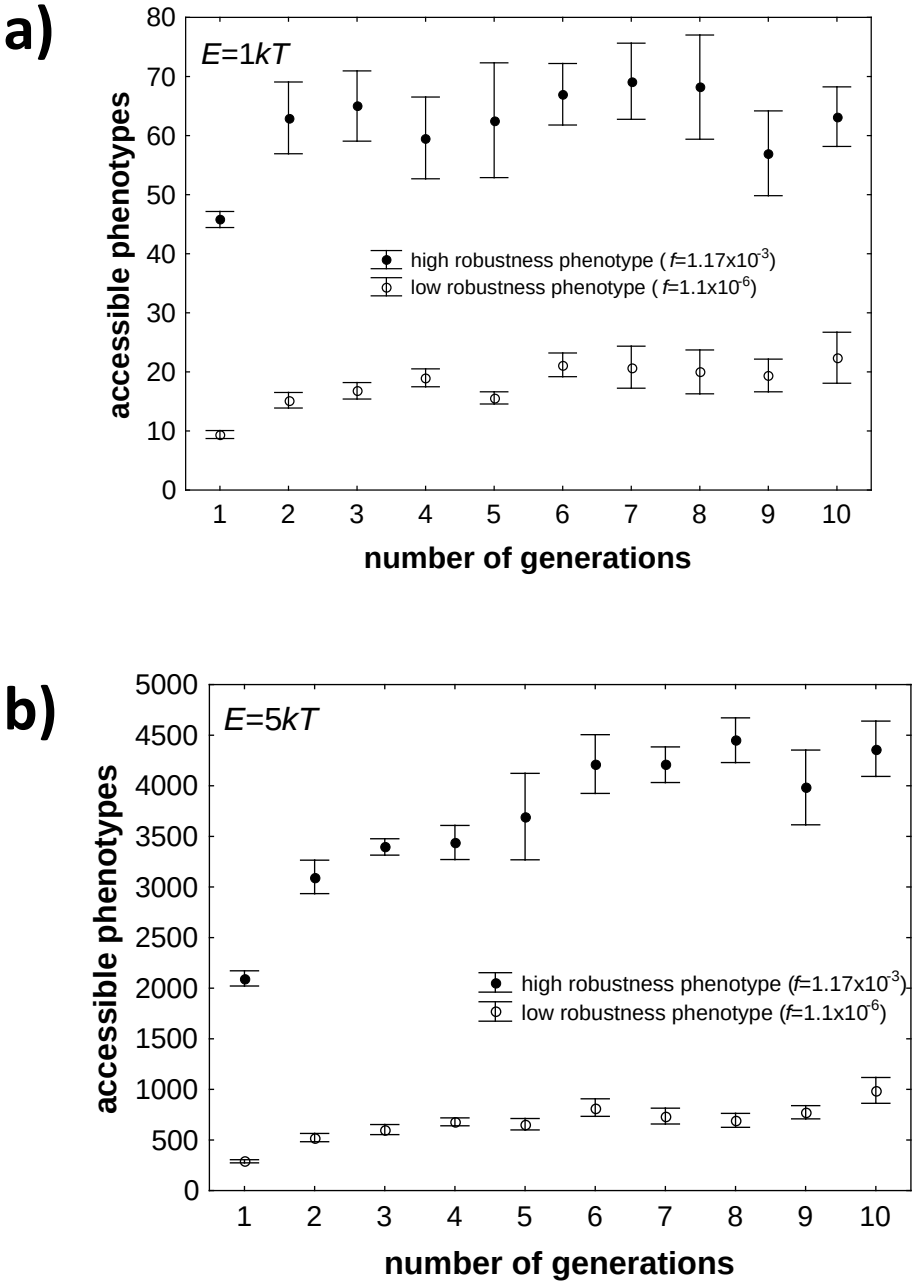

Figure S11: Phenotypic robustness allows evolving populations to access greater numbers of novel phenotypes through plasticity. The horizontal axes show time in generations (mutation-selection cycles). The vertical axes show the total number of unique phenotypes accessible through plasticity to at least 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 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.

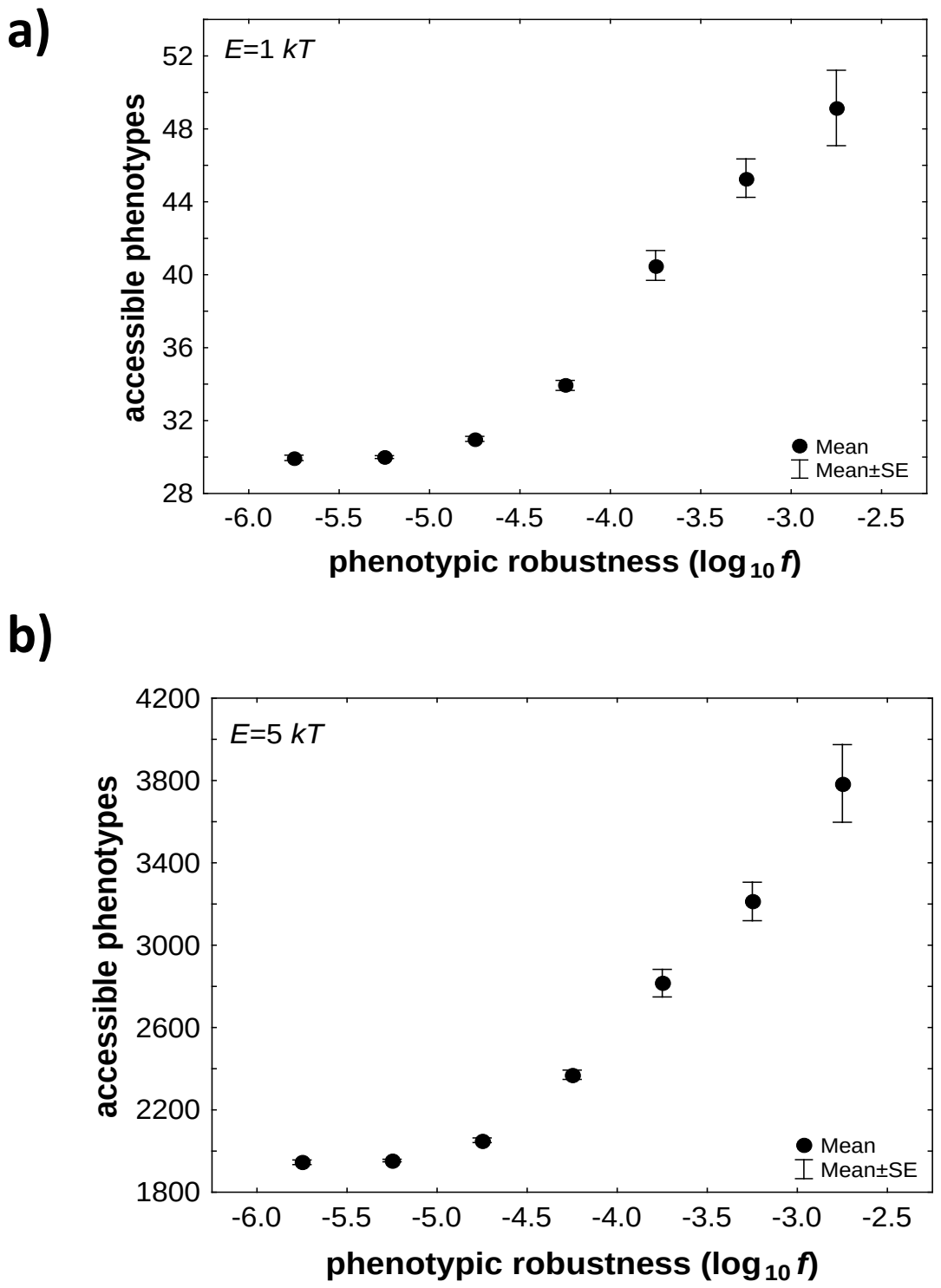

Figure S12: Phenotypic robustness allows evolving populations to access greater numbers of novel phenotypes through plasticity. The vertical axes show the total number of unique phenotypes accessible to at least one individual in a population through plasticity, where the population has evolved on the 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 μ=1 nucleotides per generation. Each population started out with identical genotypes of length L=30 nucleotides.

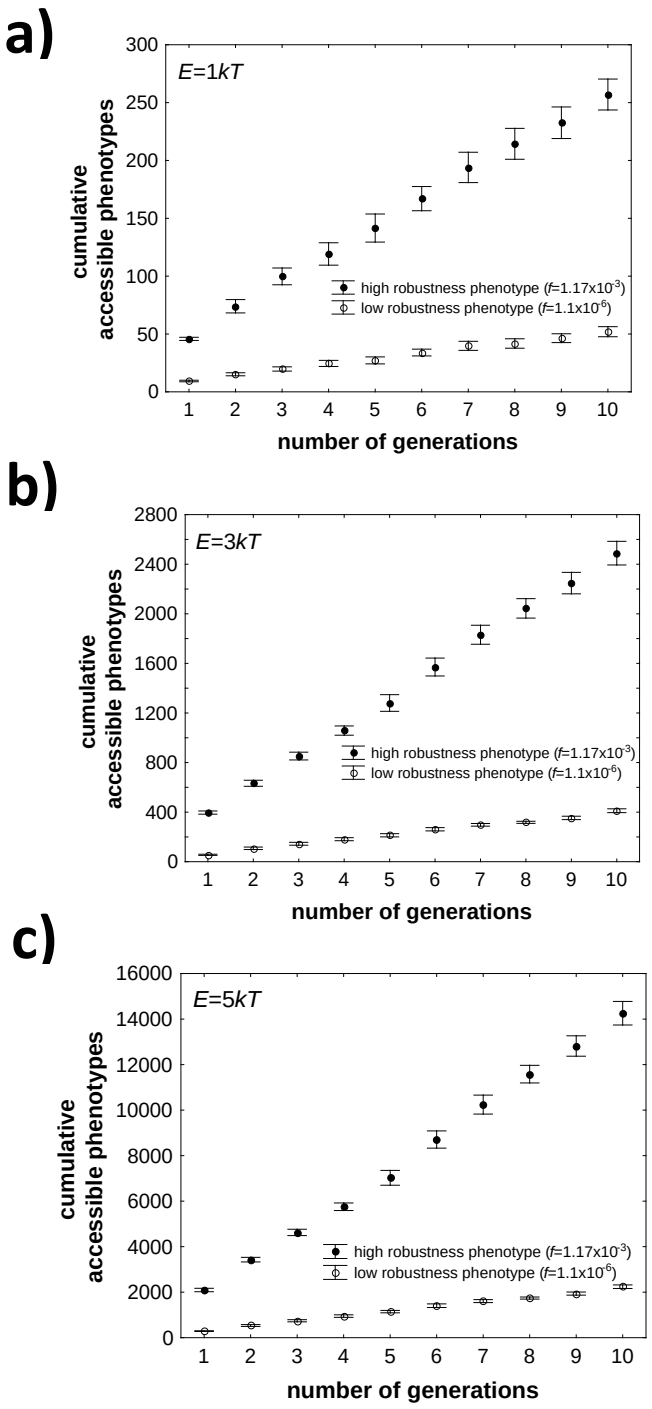

Figure S13: Phenotypic robustness entails access to greater cumulative numbers of novel phenotypes through plasticity. The horizontal axes show time in generations (mutation-selection cycles). The vertical axes show the total *cumulative* number of unique phenotypes accessible to a population through 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 $\mu=1$ 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.

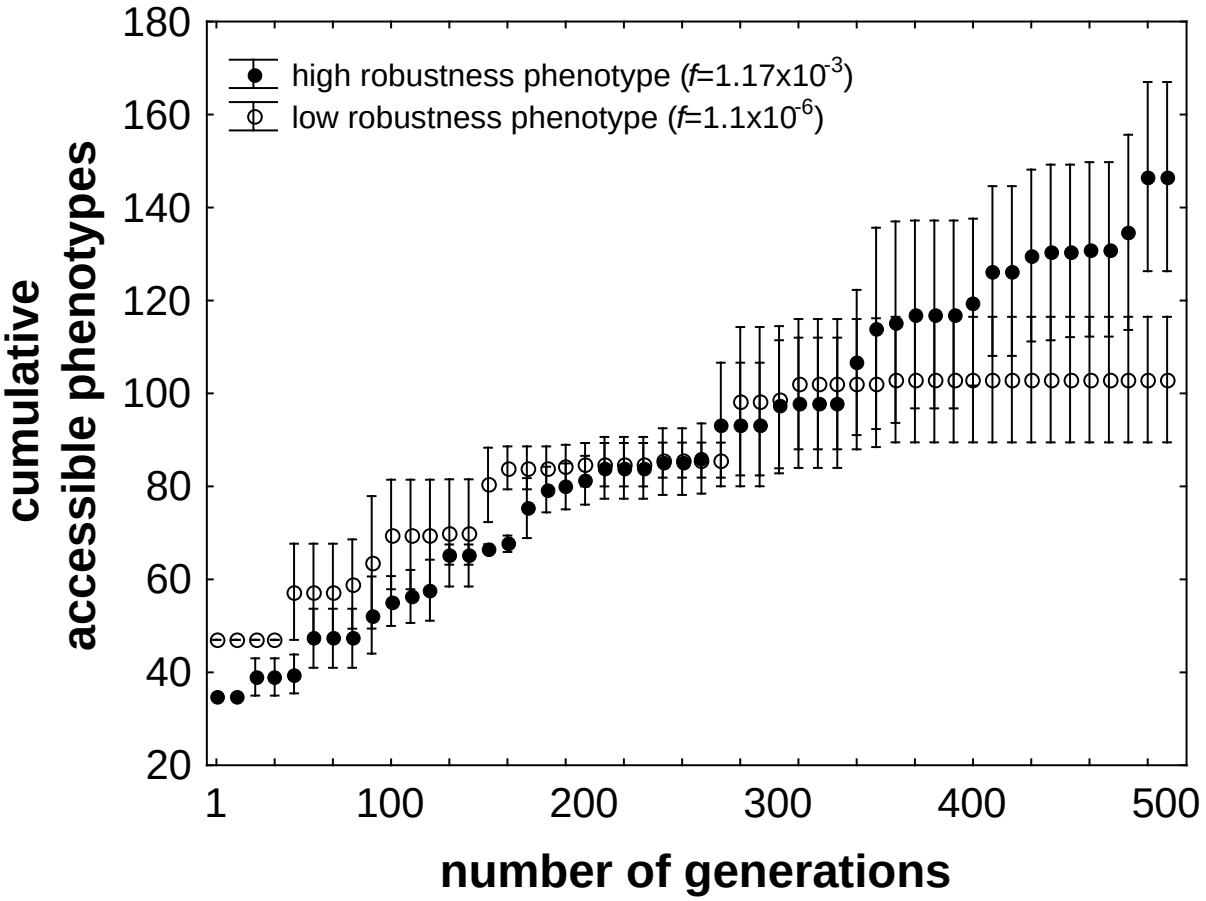

Figure S14: Phenotypic robustness entails access to greater numbers of novel phenotypes through plasticity also for populations with $N\mu$<<1. The horizontal axes show time in generations (mutation-selection cycles). The vertical axes show the total *cumulative* number of unique phenotypes accessible 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.

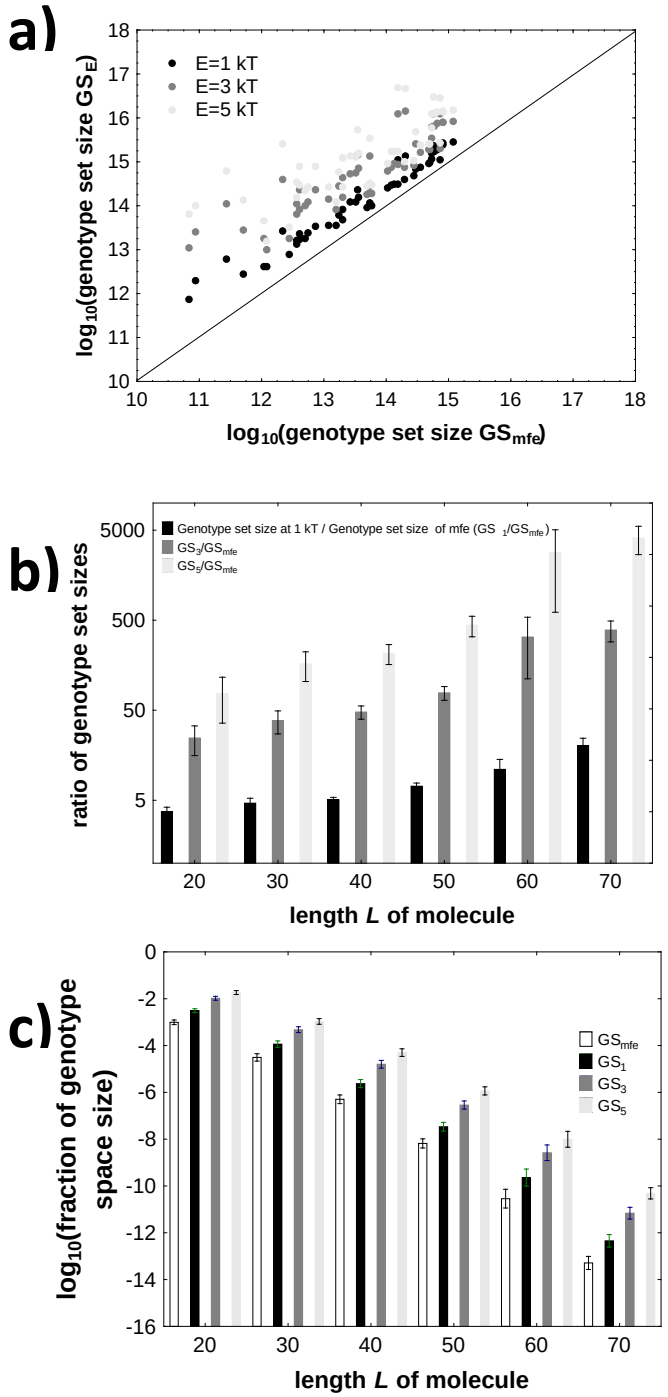

Figure S15: Genotype set sizes at different energy intervals E. For any one RNA phenotype *P*, let GS_{mfe} be the number of genotypes that adopt *P* as its minimum free energy structure. Similarly, let GS_E be the number of genotypes that contain *P* in its plastic repertoire at some energy interval *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.