Supplementary figures

Figure S1: **Fraction of neutral single-mutation neighbors.** For each of the **(a)** nine secondary structures of length 10 (Table 1) and **(b)** the four biological secondary structures (Table 2) (both depicted on horizontal axes), we selected 1,000 random sequences and determined the fraction of neighbors with a fitness difference smaller than *1/N* for a range of population sizes (legend). In these boxplots, each box encloses the second and third quartiles of the 1000 replicates, and the center line corresponds to the median. As expected, the fraction of neutral neighbors decreases with increasing population size.
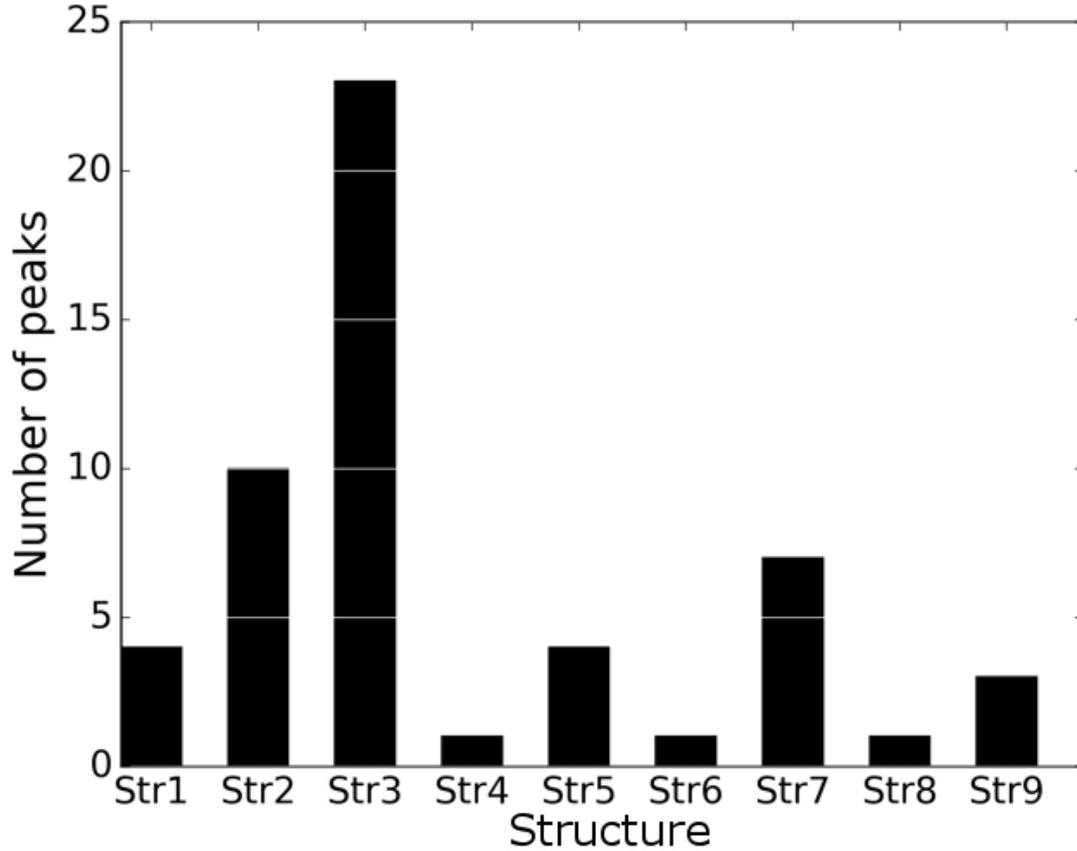
Figure S2: **Number of fitness peaks for different structures of sequences of length 10.** A peak corresponds to one or more nodes in a fitness landscape, whose neighbors all have lower fitness. (See Methods for peak identification).
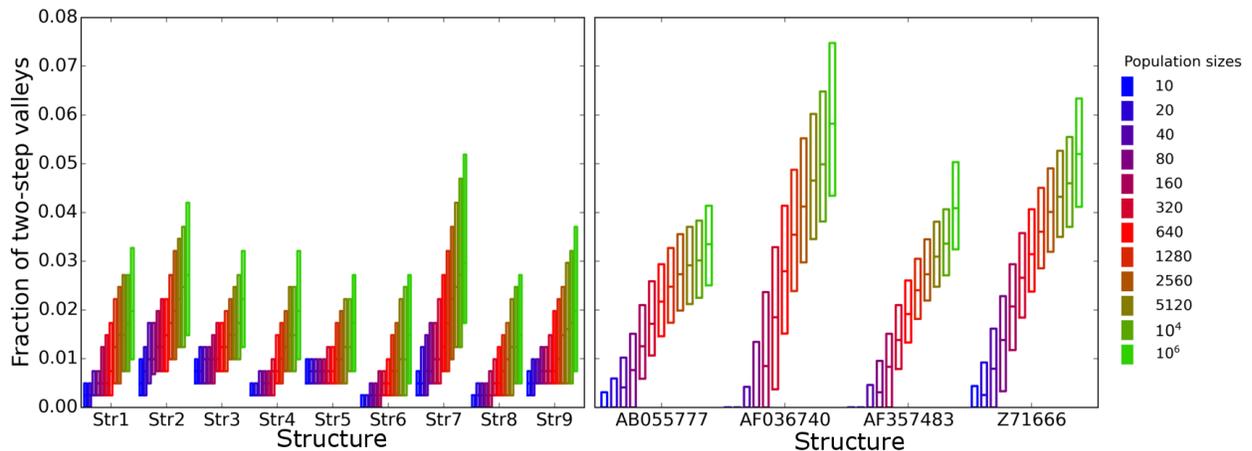
Figure S3: **Extent of reciprocal sign epistasis in fitness landscapes of biological and length 10 sequences.** The figure shows the estimated fraction of sequence quadruplets with sign epistasis, which is equivalent to the estimated fraction of fitness valleys caused by reciprocal sign epistasis in **(a)** genotype networks of sequences of length 10 (Table 1), **(b)** genotype networks of biological sequences

(Table 2). With increasing population size, the incidence of fitness valleys due to reciprocal sign epistasis increases. However, the overall fraction of such valleys is small (less than 0.06). See methods for the identification of reciprocal sign epistasis.
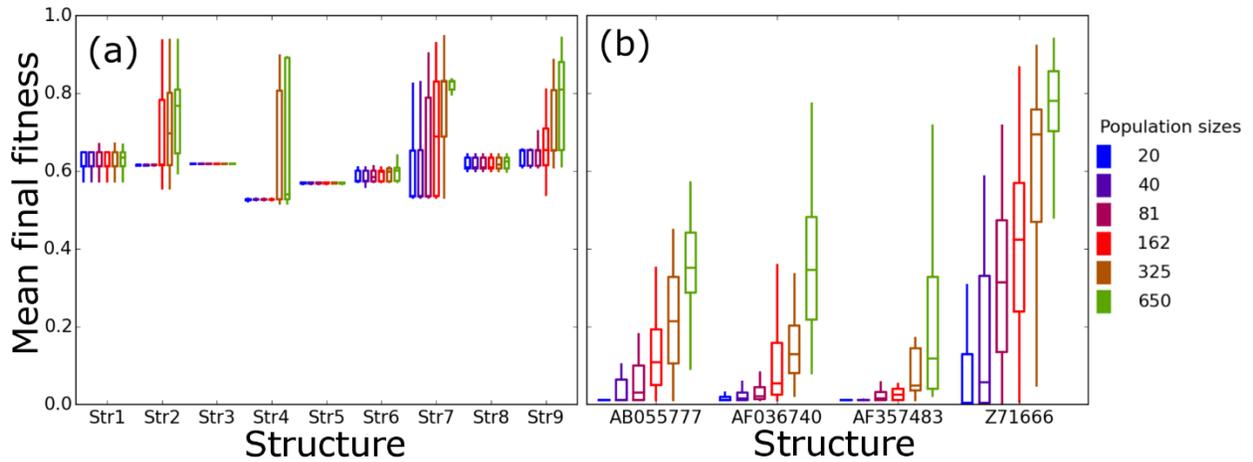
Figure S4: **Mean population fitness at the end of the simulations at constant μ=0.0001. (a)** all nine considered RNA structures of length 10 (Table 1), **(b)** biological sequences (Table 2). We randomly-selected a low-fitness sequence to initialize each simulation, and then simulated 800 generations of mutation and selection. We performed 50 replicate simulations for each population size at a fixed mutation rate of μ=0.0001 per sequence per generation (see Methods). Each box encloses the second and third quartiles of the 50 replicates, the center line corresponds to the median, and the whiskers depict the minimum and maximum values obtained from any replicate, excluding the outliers.
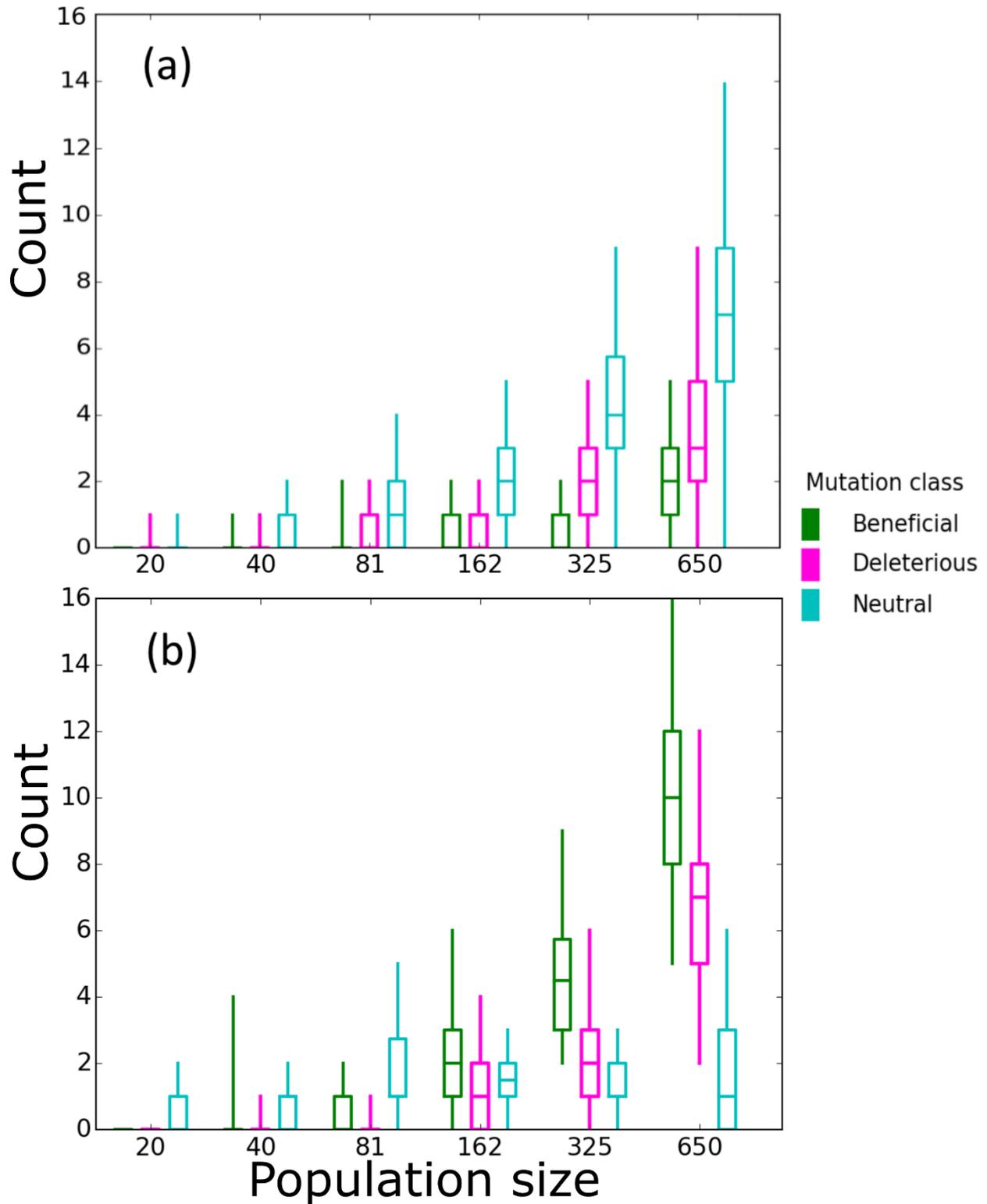
Figure S5: **Mean numbers of unique beneficial, deleterious, and neutral substitutions for RNA secondary structure 1 (Str1, Table 1) and AF036740 (Table 2) at μ=0.0001.** We randomly-selected a low-fitness sequence to initialize each simulation, and then simulated 800 generations of mutation and selection. Fifty replicates were simulated for mutation rate μ=0.0001 and a range of population sizes

(horizontal axes, see Methods). Data are based on the final generation of 50 replicate simulations. In these boxplots, we grouped the data by their fitness effects (beneficial, deleterious, neutral); each box encloses the second and third quartiles of the 50 replicates, the center line corresponds to the median, and the whiskers depict the minimum and maximum values obtained from any replicate, excluding outliers. Interestingly, we observed more deleterious than beneficial mutations in the **(a)** Str1 secondary structure simulations, and more beneficial than deleterious mutations in the **(b)** biological AF036740 secondary structure.
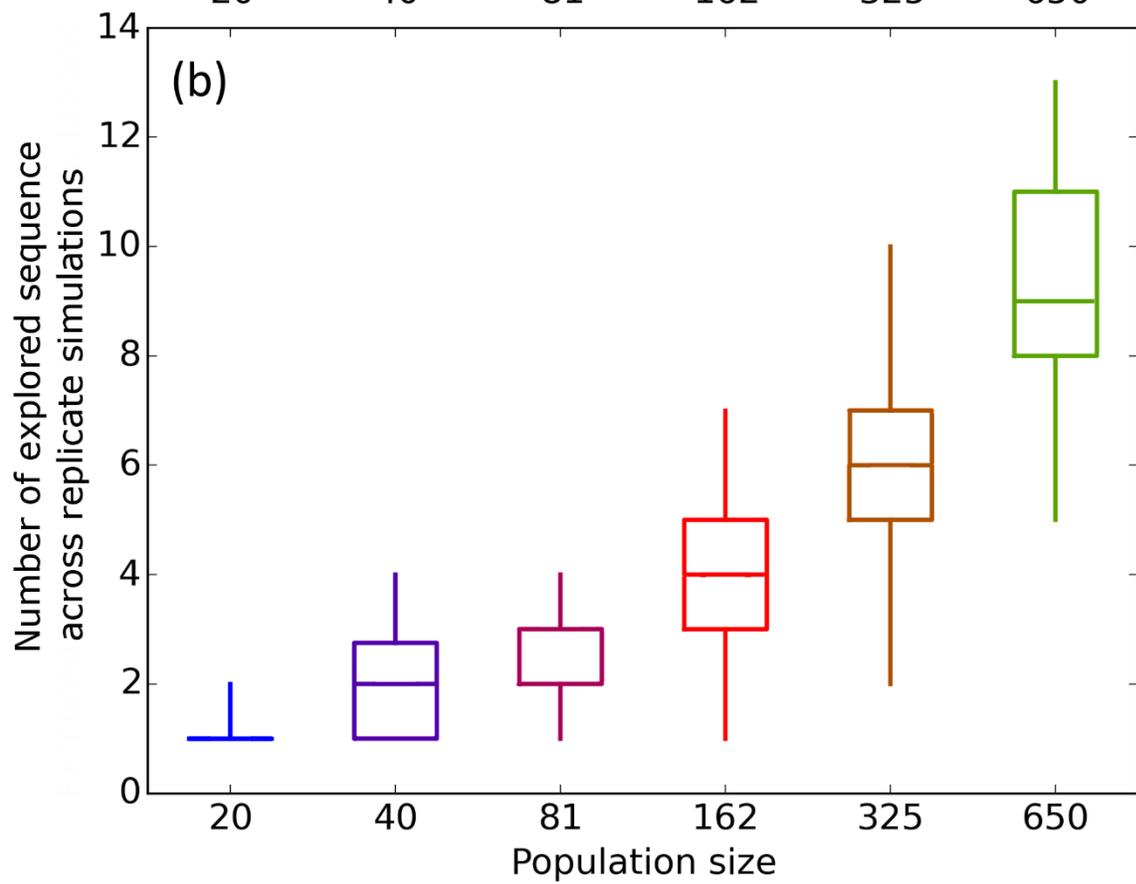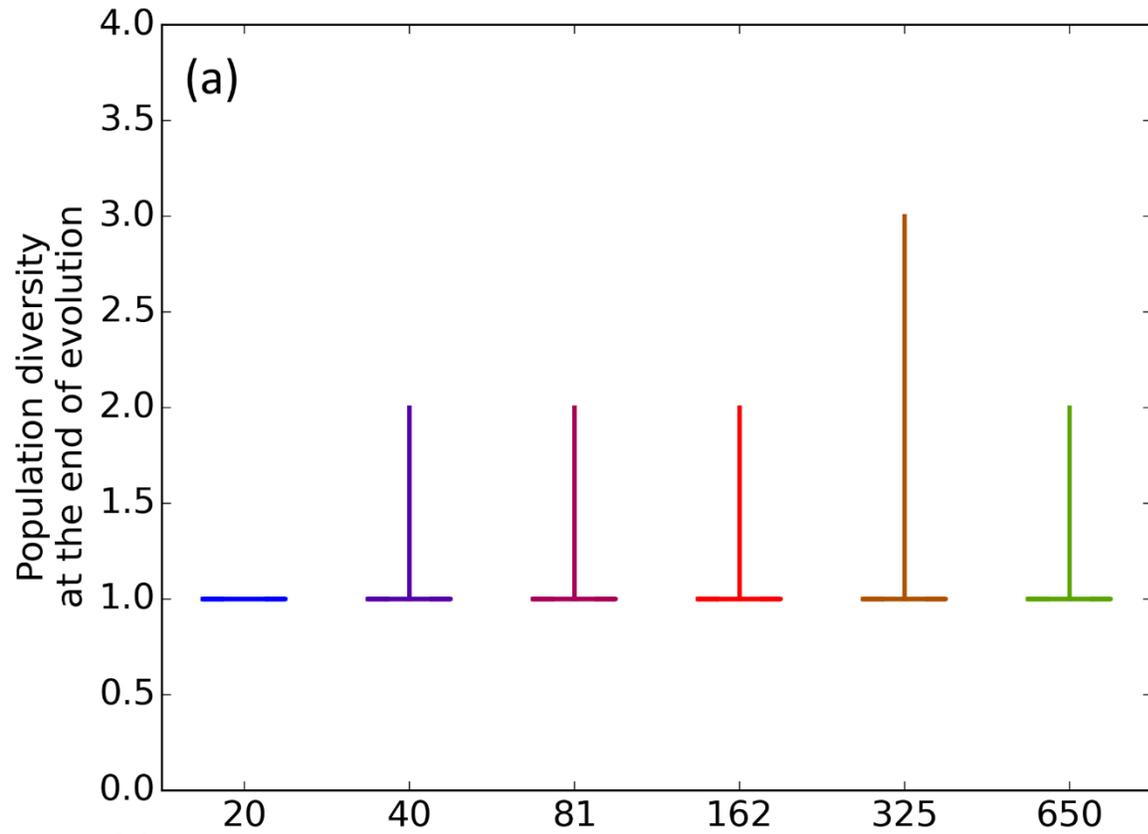
Figure S6: **Population diversity and sequence exploration in sequences with secondary structure 1 (Str1, Table 1) at μ=0.0001.** We randomly-selected a low-fitness sequence to initialize each simulation, and then simulated 800 generations of mutation and selection. We performed 50 replicate simulations for each population size (horizontal axes, see Methods). Boxplots show **(a)** final population diversity (number of unique sequences at generation 800), and **(b)** total unique sequences explored. Each box encloses the second and third quartiles of the 50 replicates, the center line corresponds to the median, and the whiskers depict the minimum and maximum values obtained from any replicate, excluding the outliers.
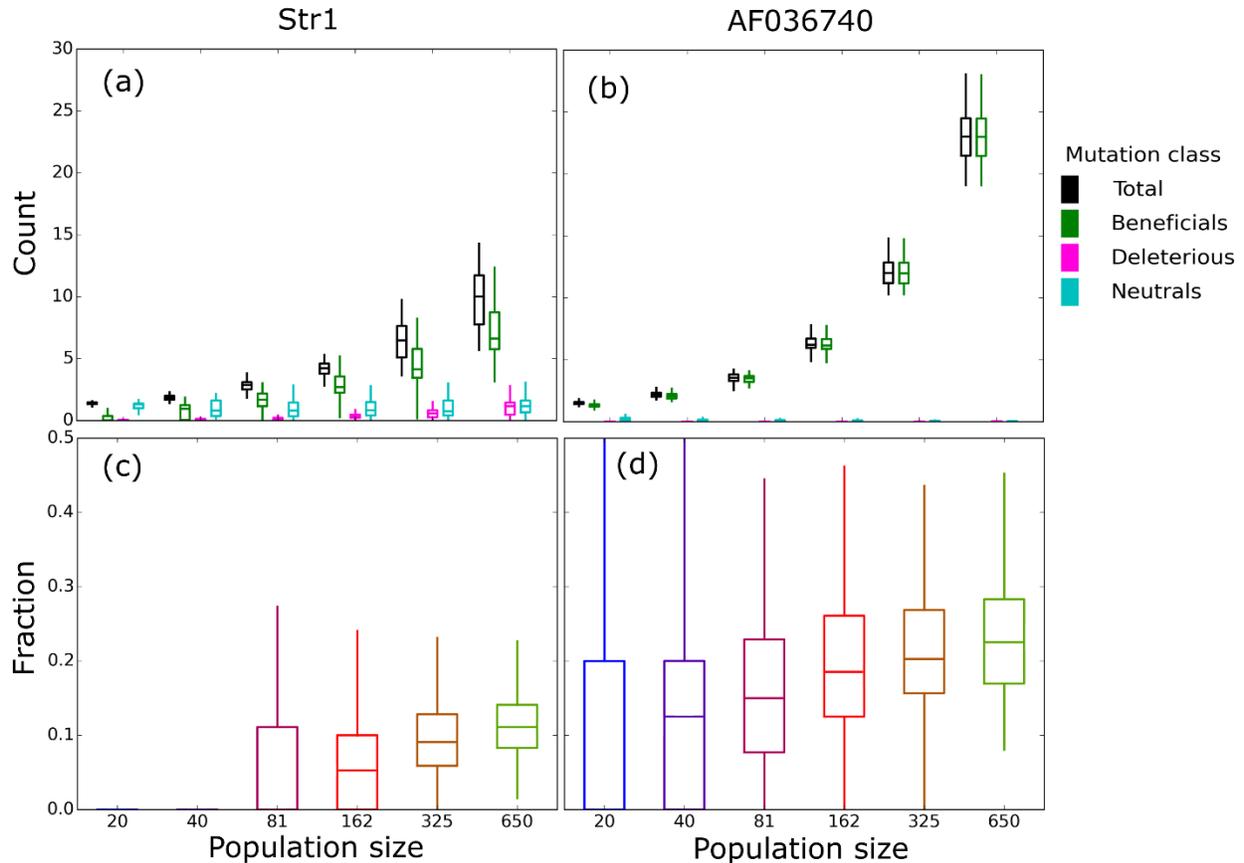
Figure S7: **Mean numbers of unique beneficial, deleterious, and neutral mutations per generation and fraction of beneficial mutations for RNA secondary structure 1 (Str1, Table 1) and AF036740 (Table 2) at μ=0.01.** We randomly-selected a low-fitness sequence to initialize each simulation, and then simulated 800 generations of mutation and selection. Fifty replicates were simulated for mutation rate μ=0.01 and a range of population sizes (horizontal axes, see Methods). Data are based on the final generation of 50 replicate simulations. Boxplots summarize mean numbers of unique beneficial, deleterious, and neutral mutations for **(a)** Str1 and **(b)** AF036740, and mean fraction of beneficial mutations for **(c)** Str1 and **(d)** AF036740. Each box encloses the second and third quartiles of the 50 replicates, the center line corresponds to the median, and the whiskers depict the minimum and maximum values obtained from any replicate, excluding the outliers.
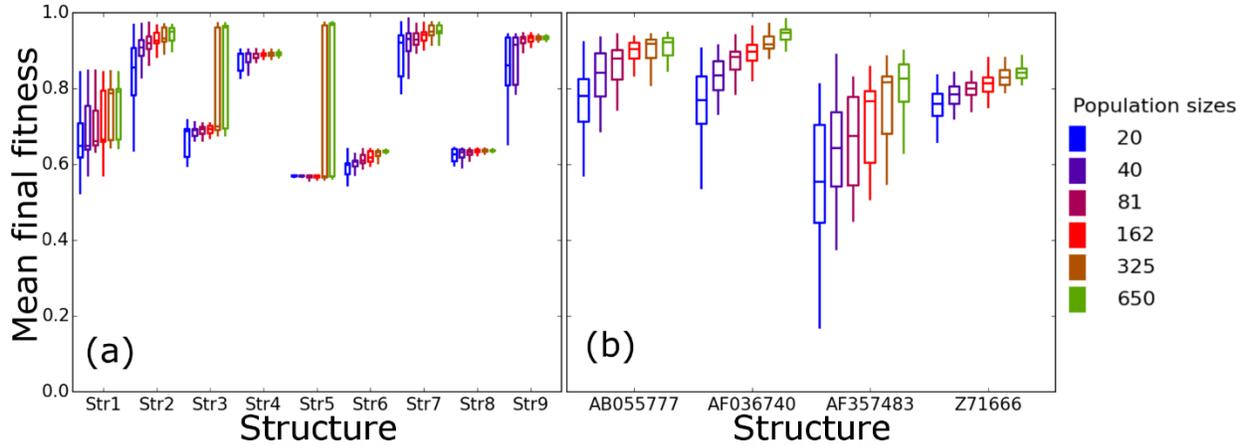
Figure S8: **Final mean population fitness after evolution of secondary structures of length 10 (Table 1) and biological RNA secondary structures (Table 2) at μ=0.01.** We randomly-selected a low-fitness sequence to initialize each simulation, and then simulated 800 generations of mutation and selection. We performed 50 replicate simulations for each structure (horizontal axes) and population size (legend, see Methods). Boxplots show mean final population fitness of all the replicates for **(a)** the structures of length 10 and **(b)** the biological RNA molecules. Each box encloses the second and third quartiles of the 50 replicates, the center line corresponds to the median, and the whiskers depict the minimum and maximum values obtained from any replicate, excluding outliers. In all structures, except Str4, the largest population (N=650) has a significantly higher final mean population fitness than the smallest population (N=20) (Mann-Whitney U test, multiple-testing correction according to FDR; Str1: p=1.01 × 10^{-7}; Str2: p=9.55 × 10^{-12}; Str3: p=8.00 × 10^{-9}; Str5: p=4.81 × 10^{-3}; Str6: p=8.42 × 10^{-14}; Str7: p=1.01 × 10^{-9}; Str8: p=2.00 × 10^{-2}; Str9: p=1.55 × 10^{-5}; AB055777: p=1.16 × 10^{-13}; AF036740: p=1.16 × 10^{-17}; AF357483: p=8.64 × 10^{-12}; Z71666: p=8.03 × 10^{-16}).
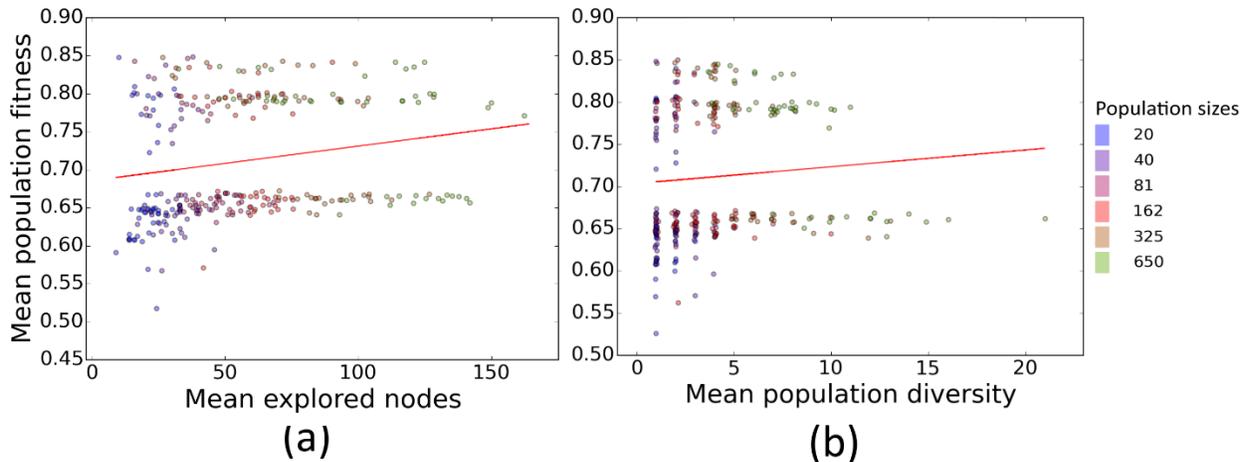
Figure S9: **Average final fitness is associated with the number of sequences explored.** Results are based on simulations with structure a (Str1, Table 1) at constant μ=0.01 and population size N. We randomly-selected a low-fitness sequence to initialize each simulation, and then simulated 800 generations of mutation and selection. Each data point represents the results of one simulation at a given population size (color legend). The vertical axes show mean final population fitness (i.e., fitness at generation 800). To better distinguish data points, a small amount of noise is added to each point. **(a)** Mean final fitness is

significantly associated with the total number of unique sequences that the population explored during 800 generations (horizontal axis; Pearson's r=0.19, p=0.00094). **(b)** However, mean final fitness is not significantly associated with the final population diversity, defined as the number of unique sequences at generation 800 (horizontal axis; Pearson's r = 0.079, p = 0.17). The bimodal distribution of mean final fitness evident in both panels is specific to the simulated structure, and not a consistent pattern across different structures.
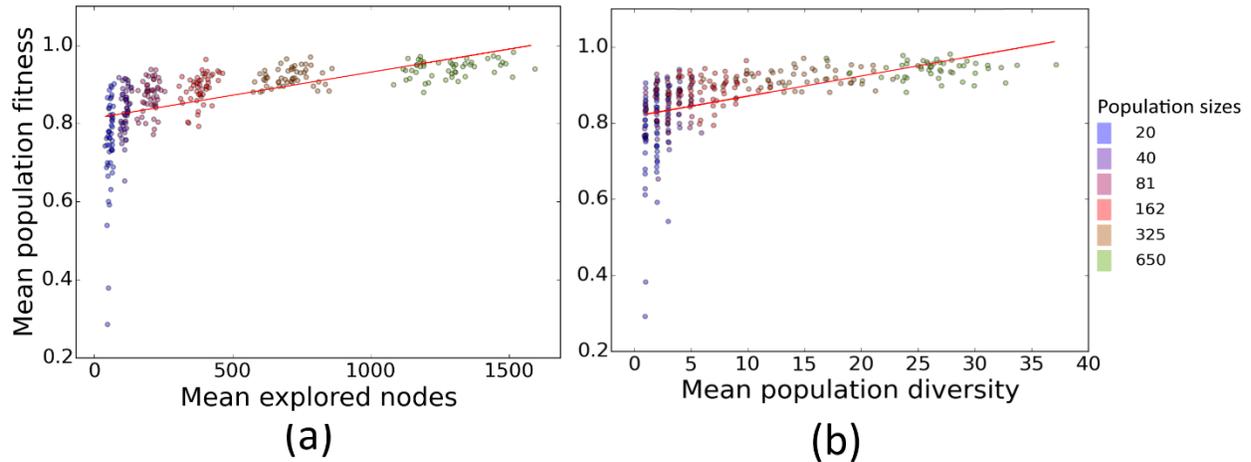
Figure S10: **Average final fitness is associated with the number of sequences explored and final population diversity for sequences with AF036740 RNA secondary structure.** Results are based on simulated evolution of structure of AF036740 of biological sequences (Table 2) at constant μ=0.01 and population size N. We randomly-selected a low-fitness sequence to initialize each simulation, and then simulated 800 generations of mutation and selection. Each data point represents the results of one simulation at a given population size (color legend). The vertical axes show mean final population fitness (i.e., fitness at generation 800). To better distinguish data points, a small amount of noise is added to each point. Mean final fitness is significantly associated with the **(a)** total number of unique sequences that the population explored during 800 generations (horizontal axis; Pearson's r=0.59, p=4.51 × 10^{-30}), and **(b)** the final population diversity, defined as the number of unique sequences at generation 800 (horizontal axis; Pearson's r=0.56, p=1.21 × 10^{-26}).
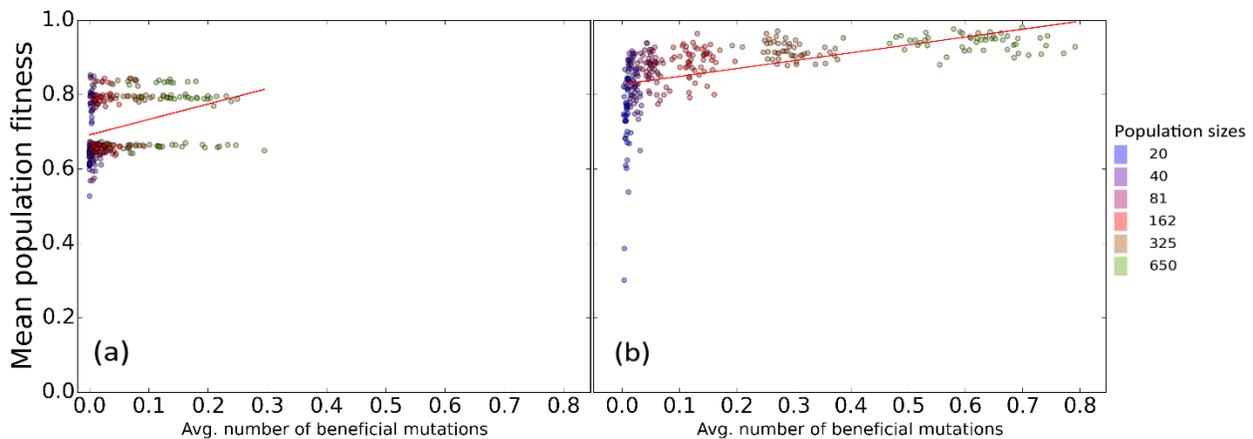
Figure S11: **Average final fitness is associated with average number of beneficial mutations at μ=0.01.** Results are based on sequences with **(a)** secondary structure 1 (Str1, Table 1) and **(b)** with structure of

AF036740 (Table 2). Correlations are significant for both samples (Mann-Whitney U test, p=8.5 × 10^{-9} and p=2.5 × 10^{-24} for (a) and (b), respectively.) We randomly-selected a low-fitness sequence to initialize each simulation, and then simulated 800 generations of mutation and selection. We simulated 50 replicate populations for each structure (horizontal axes) and population size (legend, see Methods). For better presentation of data, small noise is added to each data point. Colors of data points show which population size they represent (color legend).
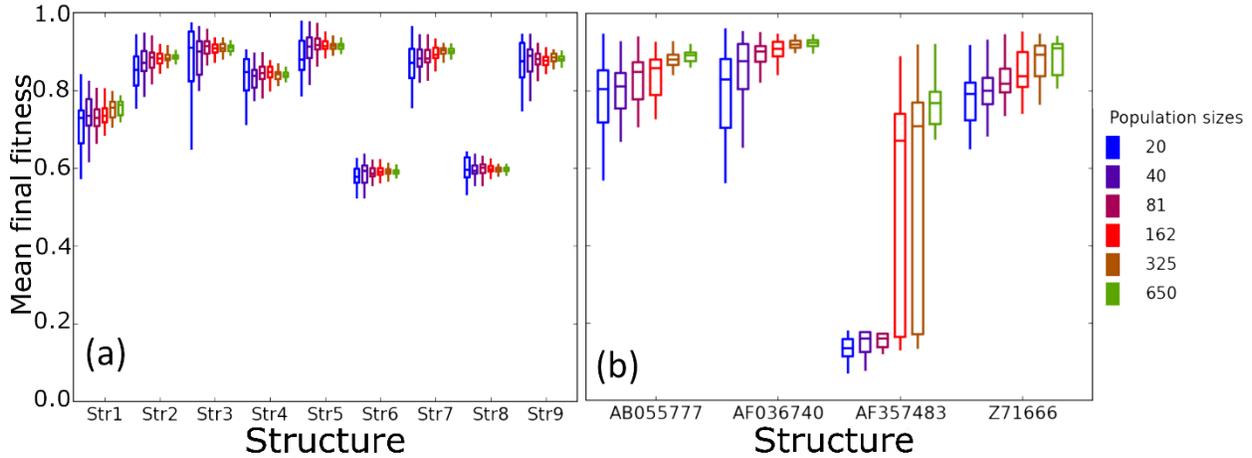
Figure S12: **Final mean population fitness after evolution of secondary structures of length 10 (Table 1) and biological RNA secondary structures (Table 2) at μ=0.1.** We randomly-selected a low-fitness sequence to initialize each simulation, and then simulated 800 generations of mutation and selection. We performed 50 replicate simulations for each structure (horizontal axes) and population size (legend, see Methods). Boxplots show the final mean population fitness of all the replicates for **(a)** the structures of length 10 and **(b)** the biological RNA molecules. Each box encloses the second and third quartiles of the 50 replicates, the center line corresponds to the median, and the whiskers depict the minimum and maximum values obtained from any replicate, excluding outliers. In all structures, except Str3, Str4, Str5, Str8 and Str9, the largest population (N=650) has a significantly higher final mean population fitness than the smallest population (N=20) (Mann-Whitney U test, multiple-testing correction according to FDR; Str1: p=4.00 × 10^-5; Str2: p=2.22 × 10^{-4}; Str6: p=1.81 × 10^{-2}; Str7: p=6.36 × 10^{-4}; AB055777: p=5.73 × 10^{-7}; AF036740: p=4.97 × 10^{-12}; AF357483: p=4.75 × 10^{-16}; Z71666: p=9.08 × 10^{-12}).
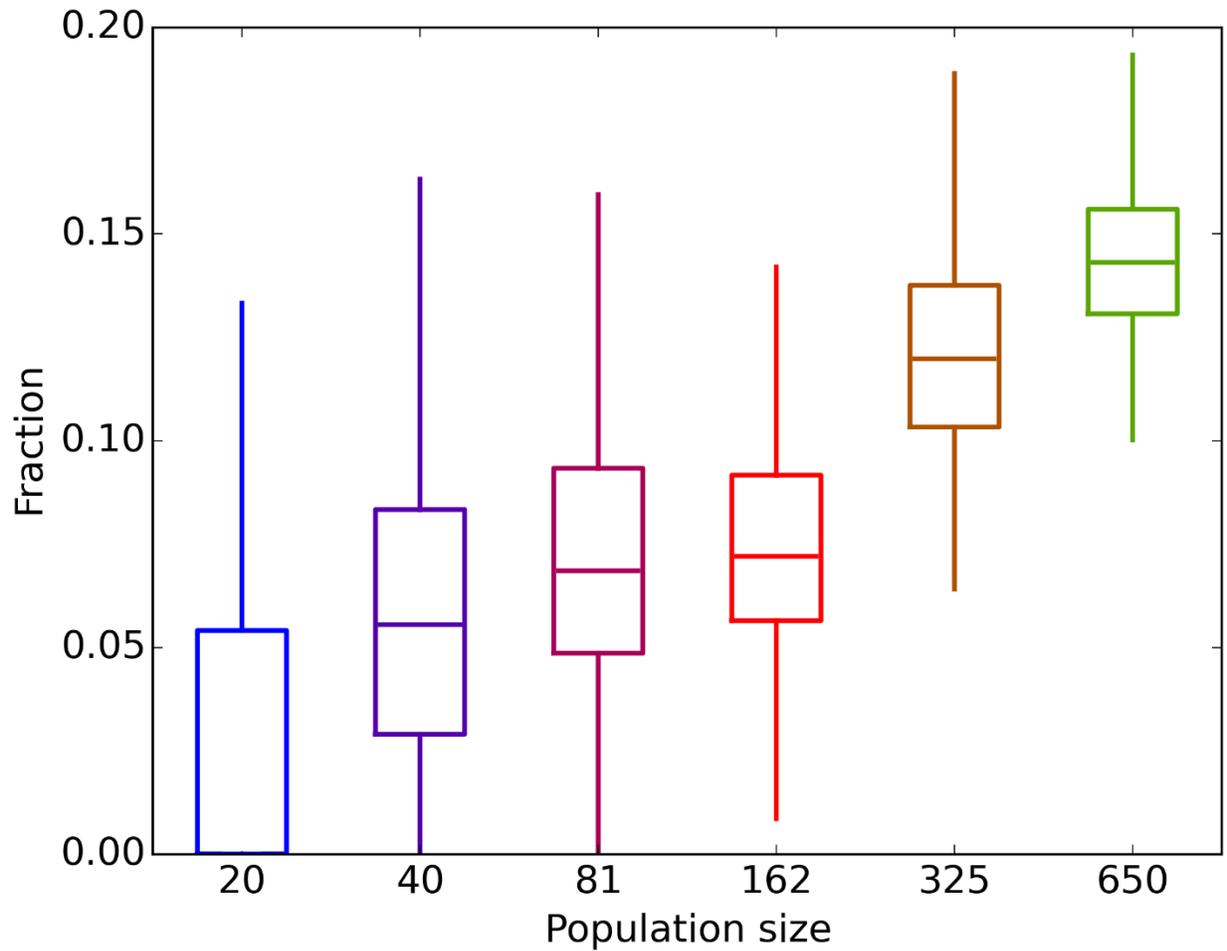
Figure S13: **Fraction of beneficial mutations for RNA secondary structure 1 (Str1, Table 1) at μ=0.1.** We randomly-selected a low-fitness sequence to initialize each simulation, and then simulated 800 generations of mutation and selection. Fifty replicates were simulated for mutation rate μ=0.1 and a range of population sizes (horizontal axes, see Methods). Data are based on the final generation of 50 replicate simulations. Boxplots summarize the mean fraction of beneficial mutations. Each box encloses the second and third quartiles of the 50 replicates, the center line corresponds to the median, and the whiskers depict the minimum and maximum values obtained from any replicate, excluding the outliers.

(a)

(b)
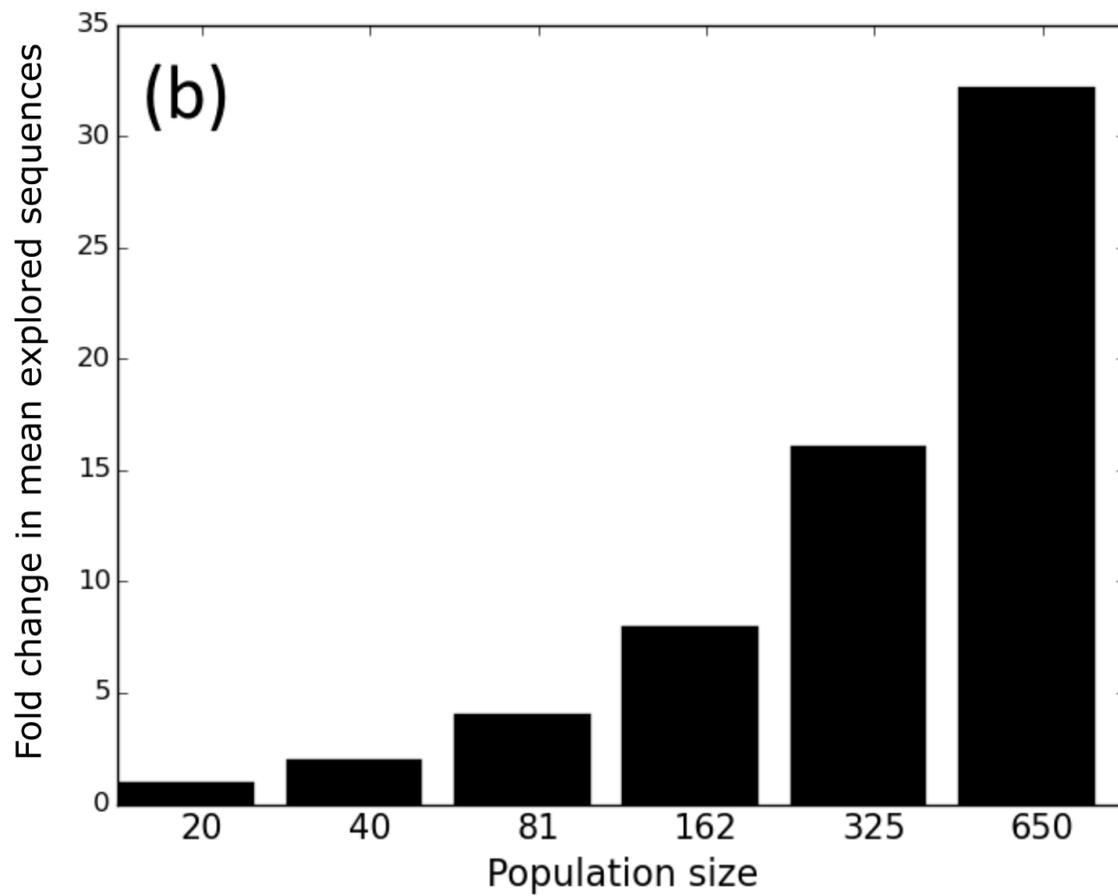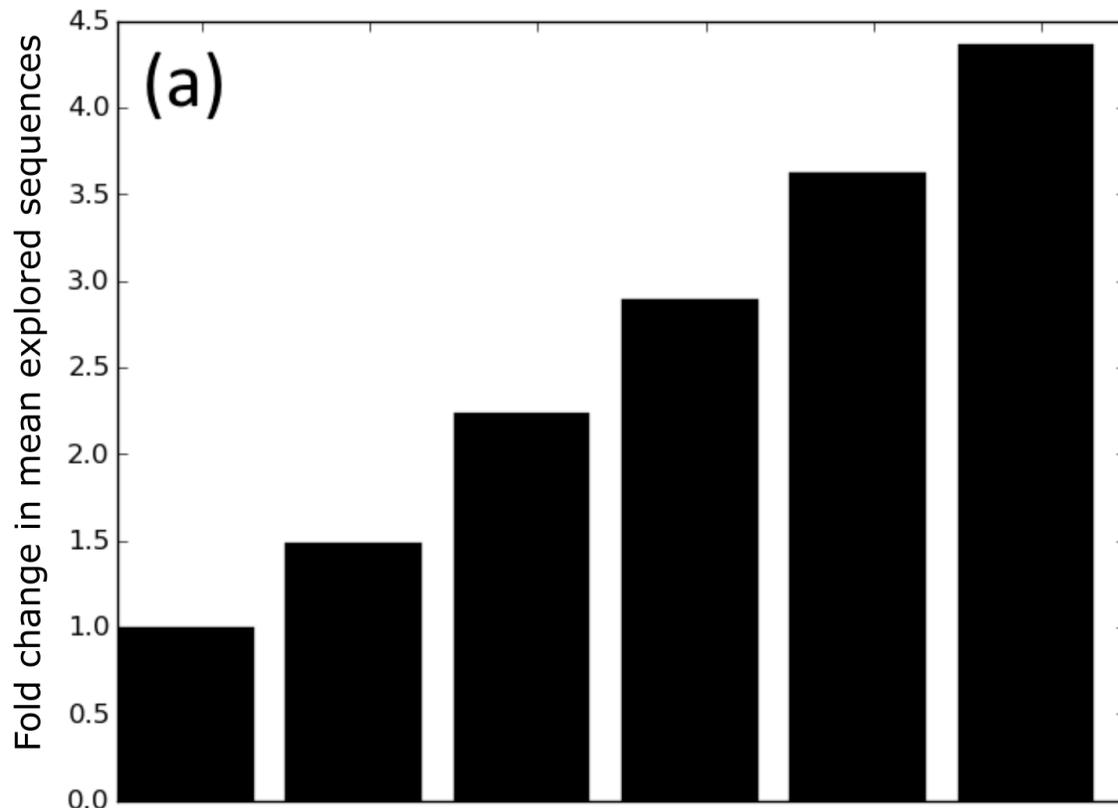
Population size

Figure S14: **Fold change in number of explored sequences for two different structures. (a)** Str1 (Table 1), **(b)** AF036740 (Table 2). The bars show by how many fold the mean number of explored sequences increases when population size increases, relative to populations of size *N*=20. The mutation rate is held constant at μ=0.1 for all populations.

Supplementary tables

Table S1: Associations between mean final fitness and population diversity at generation 800 in all 9 structures of length 10 (Table 1), when mutation rate μ=0.01.

Structure	Pearson's r	p-value
Str1	0.08	0.17
Str2	0.33	2.72E-09
Str3	0.53	2.63E-23
Str4	0.17	0.003
Str5	0.46	9.29E-17
Str6	0.49	1.36E-19
Str7	0.38	1.09E-11
Str8	0.28	4.44E-07
Str9	0.30	7.49E-08

Table S2: Associations between mean final fitness and explored sequences across generations in all 9 structures of length 10 (Table 1), when mutation rate μ=0.01.

Network	Pearson's r	p-value
Str1	0.19	9.36E-04
Str2	0.46	7.08E-17
Str3	0.69	9.38E-44
Str4	0.22	1.19E-04
Str5	0.68	1.74E-42
Str6	0.65	8.42E-37
Str7	0.46	8.32E-17
Str8	0.33	3.69E-09
Str9	0.41	1.42E-13