Spiraling complexity: a test of the snowball effect in a computational model of RNA folding

Ata Kalirad* **and Ricardo B. R. Azevedo***

*Department of Biology and Biochemistry, University of Houston,, Houston, Texas, United States of America

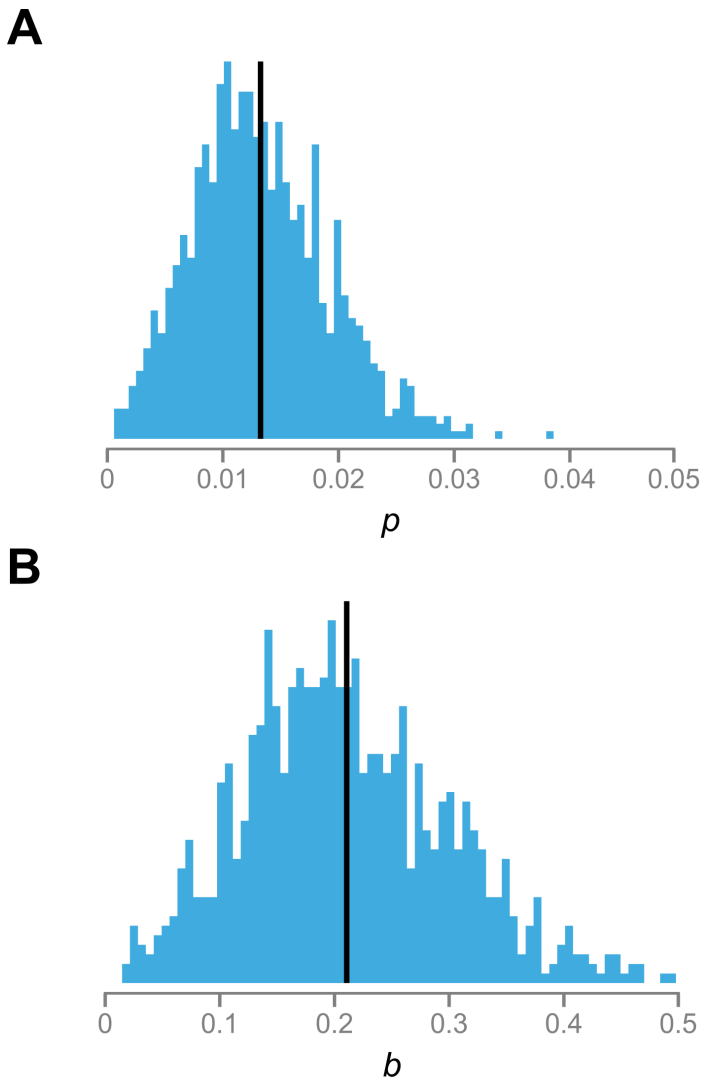

Figure S1 Distributions of the parameters of the snowball and linear models in the RNA folding simulations. (A) Probability, p, that a simple DMI appears in the snowball model. (B) Rate of accumulation, b, of simple DMIs in the linear model. For each of the 10^3 stochastic RNA folding simulations we estimated p and b by fitting the models in Equations 2 and 6, respectively, by the method of least squares. The solid black lines indicate the means of the distributions: $\bar{p} = 0.013$ and $\bar{b} = 0.211$.

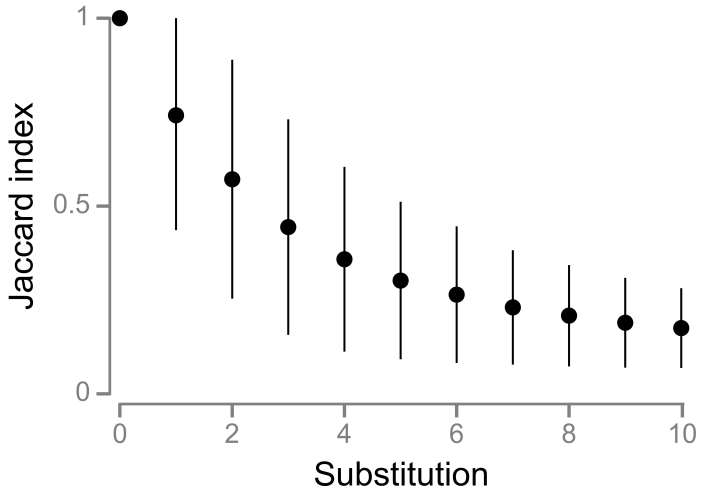

Figure S2 Networks of potential simple DMIs are not static in the RNA folding model. Jaccard index (Equation 5) of the DMI networks of each descendant lineage after k substitutions compared to its ancestor. Values are means of 2×10^3 DMI networks (10^3 simulations, 2 lineages per simulation). Based on the simulations summarized in Figure 5. Error bars show ± 1 standard deviation.

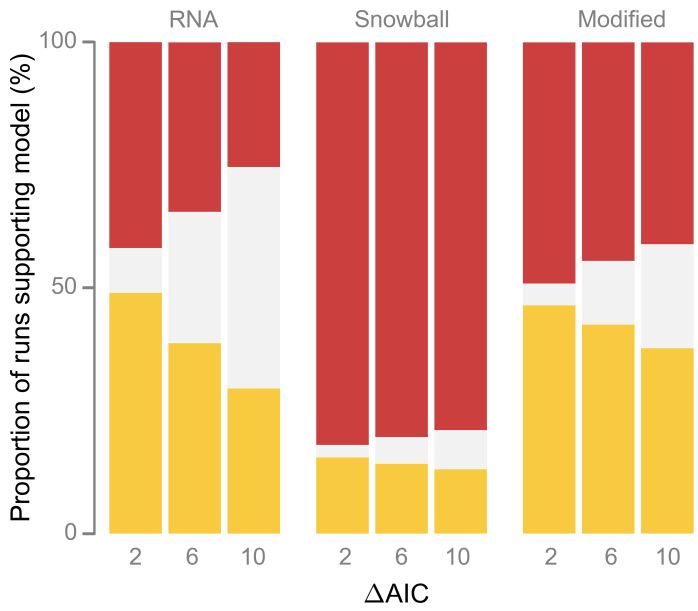

Figure S3 The RNA folding model behaves as expected under the modified Orr model. We fitted the snowball model and a linear model to each run from three kinds of simulations: simulations of the RNA folding model ("RNA"), direct simulations of the Orr model ("Snowball"), and direct simulations of the modified Orr model with values of p and q estimated directly from each RNA folding simulation ("Modified") (Figure 8). Red segments show the proportions of runs providing stronger support for the snowball model; yellow segments show the proportions of runs providing stronger support for the linear model; gray segments show the proportions of runs providing approximately equal support for both models. Each proportion is based on 10^3 stochastic simulations. The level of support for the two models was evaluated for three different ΔAIC thresholds.

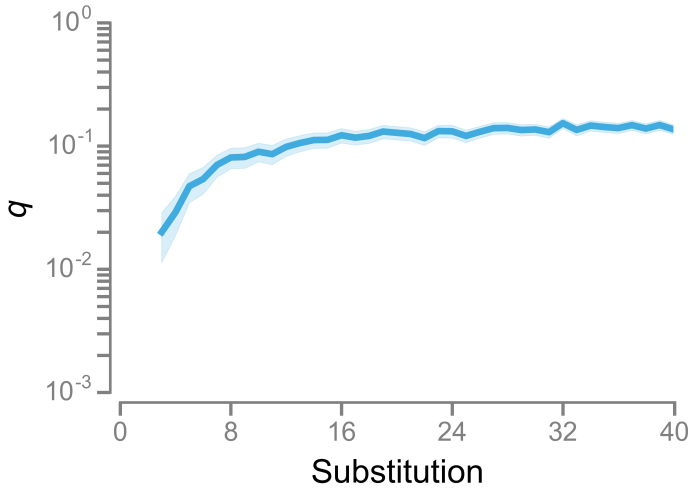

Figure S4 Evolution of the probability, q, that a simple DMI becomes complex. For each stochastic RNA folding simulation we measured q_k, the probability that a simple DMI present after k substitutions will become complex after the next substitution. Values are means of 10^3 simulations at each k. Based on the simulations summarized in Figure 5. Shaded regions indicate 95% CIs.

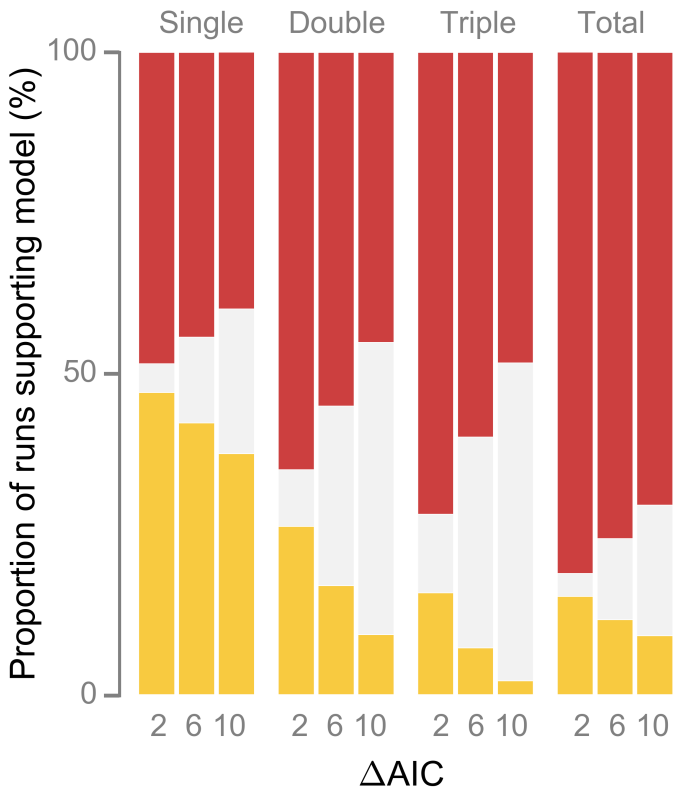

Figure S5 Complex DMIs snowball in the RNA folding model. We fitted two models to the evolutionary responses in the numbers of complex DMIs found through single, double, and triple introgressions (Figure 9): a linear model ($I_k = bk$) and a snowball model ($I_k = bk^2$). Red segments show the proportions of runs providing stronger support for the snowball model; yellow segments show the proportions of runs providing stronger support for the linear model; gray segments show the proportions of runs providing approximately equal support for both models. Each proportion is based on 10^3 stochastic simulations. The level of support for the two models was evaluated for three different ΔAIC thresholds.

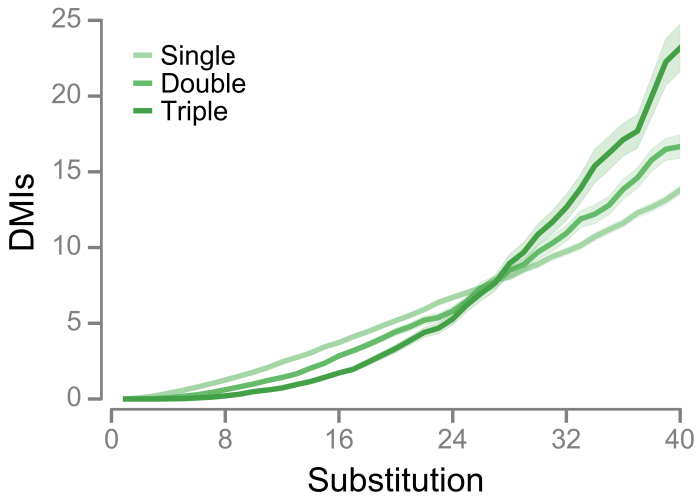

Figure S6 Allowing sites to undergo multiple substitutions does not affect the pattern of accumulation of DMIs inferred through single, double, and triple introgressions. The results are based on 10^3 RNA folding simulations for $\alpha = 12$. Shaded regions indicate 95% CIs.

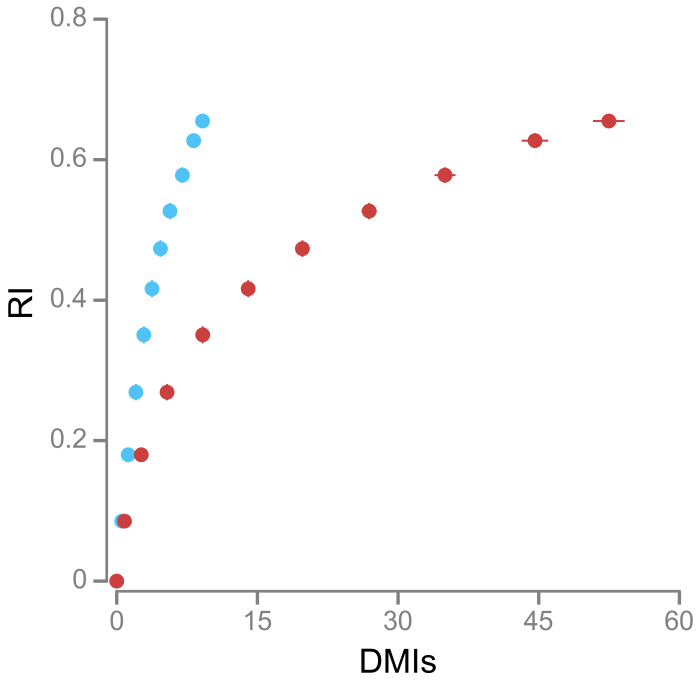

Figure S7 The numbers of simple (blue) and complex DMIs (red) are not linearly related to RI. Values are means of 10^3 simulations for $k = 0, 4, 8, \ldots, 40$. Error bars are 95% CIs.

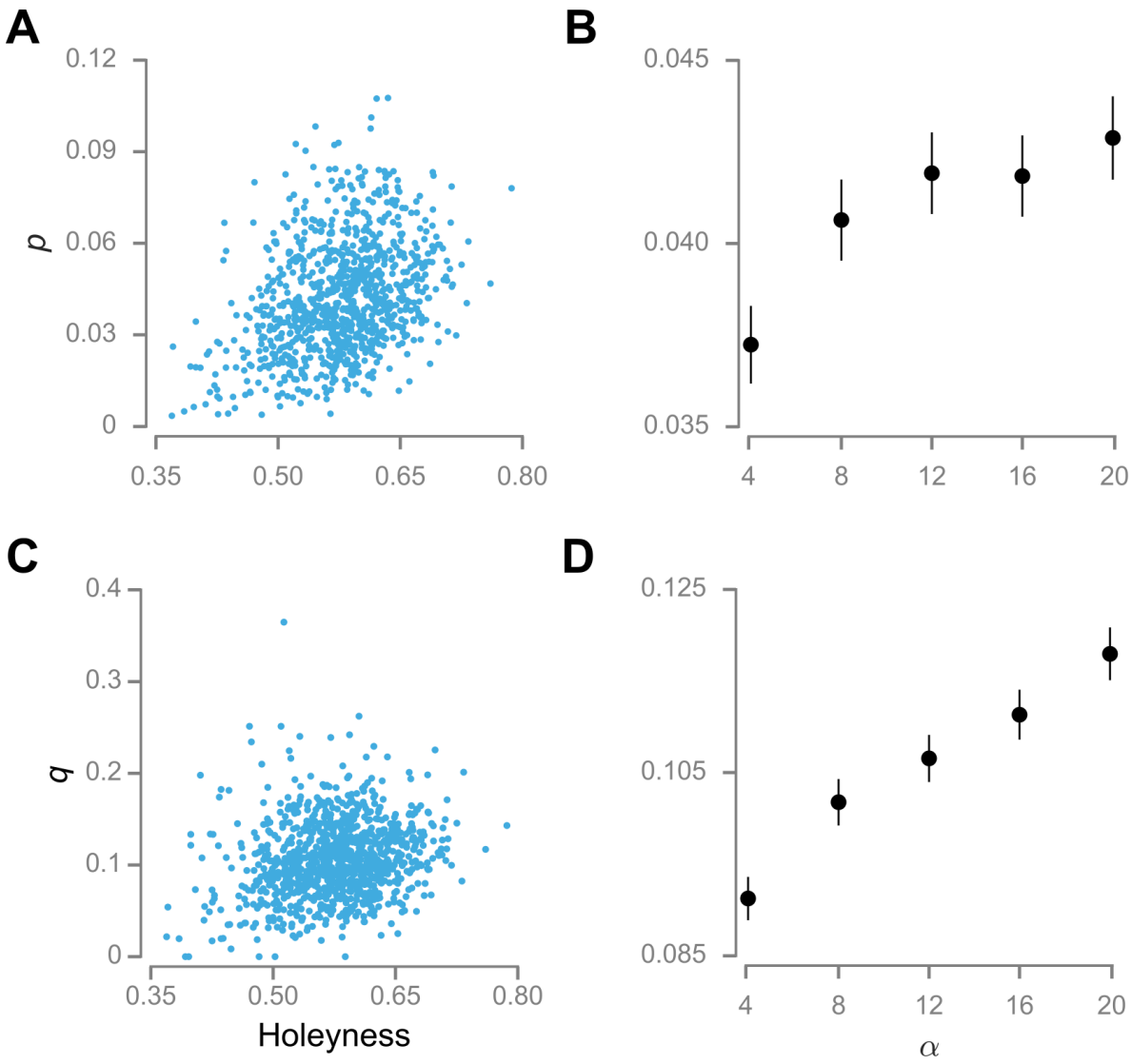

Figure S8 The fitness landscape influences the parameters of the melting snowball model. (A, C) Both parameters are positively related to the local holeyness of the fitness landscape. Values are individual estimates of p and q for each of 10^3 RNA folding simulations for $\alpha = 12$. (B, D) Both parameters are positively related to α. Values are means of 10^3 RNA folding simulations for each value of α. Error bars are 95% CIs.

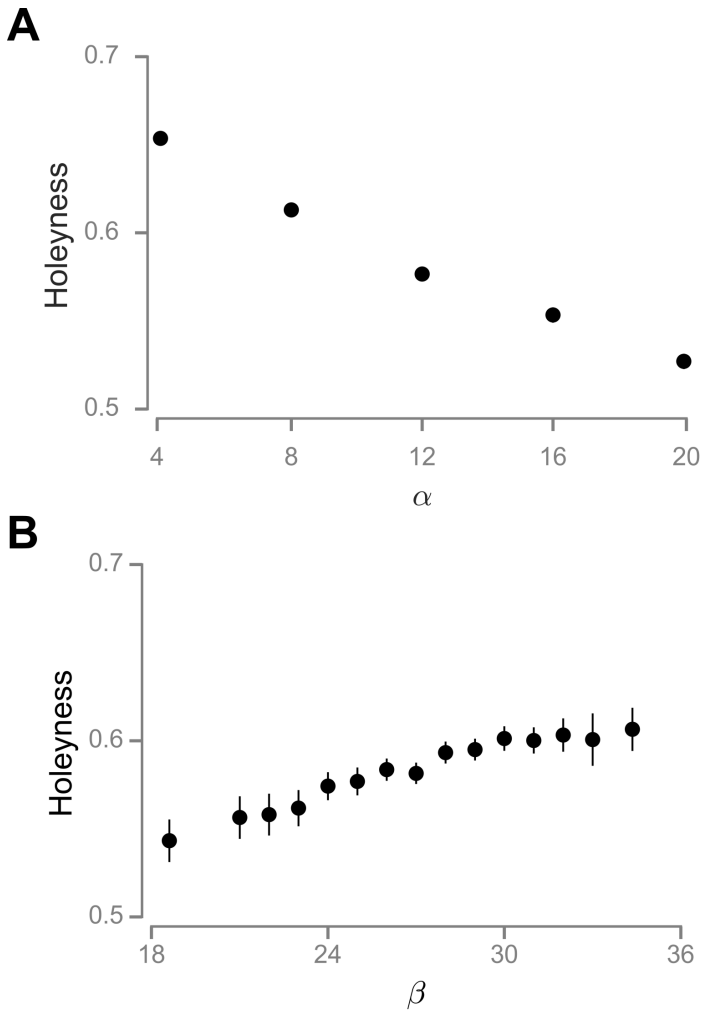

Figure S9 Holeyness decreases with the value of α (A) and increases with the number of base pairs, β, in the reference sequence (B). (A) Values are means of 10^3 RNA folding simulations for each value of α. (B) The holeyness data from the 5×10^3 simulations used in (A) were grouped by individual values of β. We pooled estimates for $\beta \leqslant 20$ and for $\beta \geqslant 34$. The resulting β groups have sample sizes ranging from 120 to 581. Error bars are 95% CIs. The error bars in (A) are covered by the points.

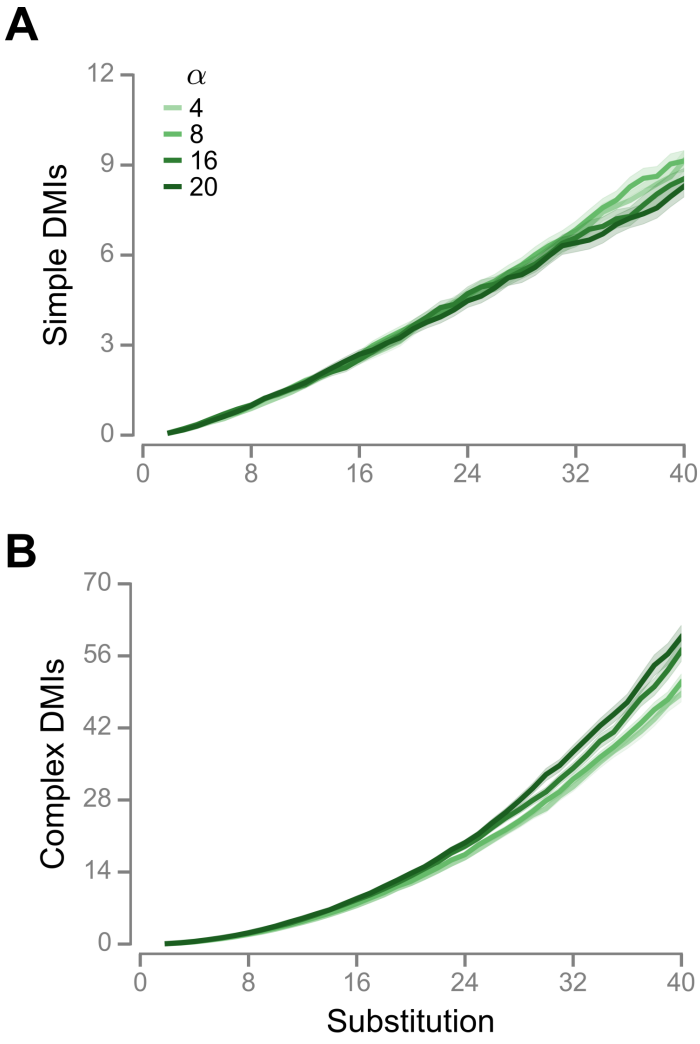

Figure S10 Simple DMIs accumulate more slowly (A) and complex DMIs accumulate faster (B) as α increases. The number of complex DMIs was calculated as in Figure 9B. Values are means of 10^3 RNA folding simulations for each value of α. Shaded regions indicate 95% CIs.

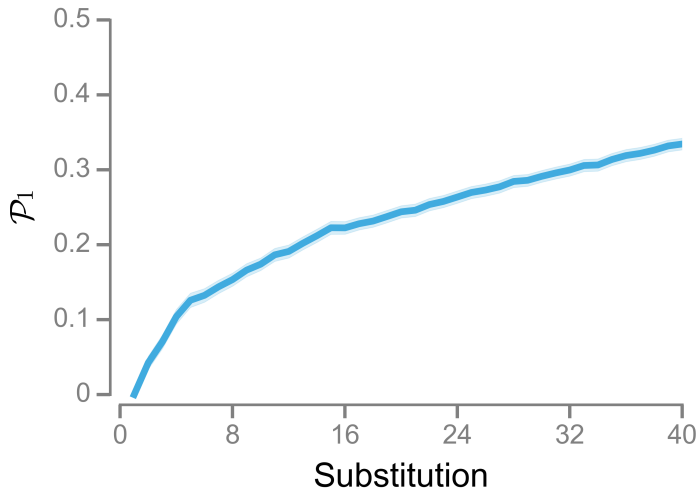

Figure S11 Evolution of the proportion of single introgressions involved in a DMI, $\mathcal{P}_1$, as populations diverge in the RNA folding model. Values are means of 10^3 stochastic simulations. Shaded region indicates 95% CIs.

Table S1 Properties of the 10^3 ancestors used in the simulations with $\alpha = 12$

Property	Mean	(Standard deviation)
Sequence		
GC content	0.49	(0.05)
Hamming distance from the reference sequence	56.16	(5.30)
Structure		
Minumum free energy (kcal mol^{-1})	–22.56	(5.70)
Number of base pairs	24.91	(4.16)
Proportion of inviable single mutants	0.57	(0.11)
Number of potential DMIs per site	8.99	(5.70)
Base pair distance from the reference sequence	11.29	(1.03)
Ensemble		
Base pair distance between pairs of sequences	49.51	(6.01)

Table S2 Estimates of parameters c_1 and c_2 of the model $I_k = c_1(k-i)^{c_2}$. The model was fitted to the average number of DMIs from 10^3 simulations with $\alpha = 12$ by nonlinear least squares regression.

Type of DMIs	Order (n)	c_1	(95% CI)	c_2	(95% CI)	R^2
Single introgressions $(i = 1)$[a]						
Simple DMIs[b]	2	0.074	(0.006)	1.31	(0.022)	0.998
Complex DMIs[c]	$\geqslant 3$	0.080	(0.003)	1.39	(0.010)	0.999
All DMIs	$\geqslant 2$	0.153	(0.008)	1.36	(0.014)	0.999
Double introgressions[c] $(i = 2)$[a]	$\geqslant 3$	0.024	(0.002)	1.78	(0.024)	0.999
Triple introgressions[c] $(i = 3)$[a]	$\geqslant 4$	0.0029	(0.0004)	2.45	(0.041)	0.999

[a] i is the number of introgressed alleles.
[b] Data in Figure 5A.
[c] Data in Figure 9A.

Table S3 Estimates of parameters c_1 and c_2 based on the simulations summarized in Figure S6. See Table S2 for more details.

Type of DMIs	Order (n)	c_1	(95% CI)	c_2	(95% CI)	R^2
Single introgressions $(i = 1)$	$\geqslant 2$	0.092	(0.005)	1.36	(0.016)	0.999
Double introgressions $(i = 2)$	$\geqslant 3$	0.020	(0.002)	1.85	(0.032)	0.999
Triple introgressions $(i = 3)$	$\geqslant 4$	0.0018	(0.0003)	2.59	(0.055)	0.998