Mutational robustness can facilitate adaptation Supplementary Information

Jeremy A. Draghi[1], Todd L. Parsons[1], Günter P. Wagner[2], Joshua B. Plotkin[1]

[1]Department of Biology, University of Pennsylvania, Philadelphia, PA, USA
[2]Department of Ecology & Evolutionary Biology, Yale University, New Haven, CT, USA

1 Robustness and adaptation time

In this section we study how robustness is related to adaptation time.

1.1 Model Description

As in the main text, we consider a population of N individuals under the infinite alleles Moran model. In each discrete time step, a randomly-chosen individual produces one offspring, which replaces a random individual. A mutation occurs with probability μ and produces a unique genotype. With probability q a mutation is neutral; q therefore quantifies robustness. Otherwise, the mutation is non-neutral and changes the phenotype to one of K phenotypes accessible from a given genotype. Each genotype has a specific set of K accessible phenotypes which constitute its phenotypic neighborhood; these K phenotypes are drawn uniformly from P possible alternatives. Genotypes have independent phenotypic neighborhoods, so the K accessible phenotypes are redrawn whenever a mutation occurs. The model is described by the five parameters K, P, μ, q, and N.

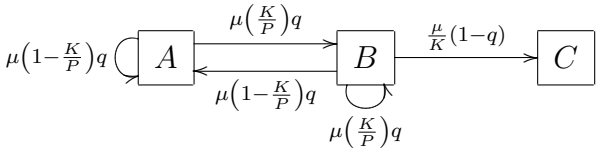
Starting from a population in steady state, we consider an environmental shift that assigns one of the P alternative phenotypes the highest fitness. Before the environmental shift, we assume that all P of the alternative phenotypes are inviable, such that only genotypes expressing the wild-type phenotype survive and reproduce. We consider a population evolving in this regime of stabilizing selection until it

reaches steady-state, as described in Section **??** below. Sometime after the population reaches steady-state, the selective environment shifts such that one of the P alternative phenotypes is no longer inviable, but is instead more fit than the wild-type. In this section we derive an analytic expression for the mean adaptation time – i.e the average amount of time elapsed before the newly beneficial phenotype first arises in the population.

Let $t = 0$ denote the time at which the alternative phenotype becomes more fit than the current phenotype – i.e. the time of the environmental shift. At this time we classify all genotypes in the population into three distinct groups. Genotypes that express the newly beneficial phenotype are called class C. Genotypes that express the wild-time phenotype are divided into two classes: those that can reach the optimal phenotype by a single point mutation, called class B or 'adaptable', and those that cannot reach the optimal phenotype by a single mutation, called class A. This simplification into three classes is equivalent to the infinite-allele model described in the main text because all genotypes within a class have identical mutations rates to other classes.

A mutation that arises in a genotype of class A produces a genotype of class B with probability $q\frac{K}{P}$, and it produces another genotype in class A with probability $q(1 - \frac{K}{P})$. The same holds for mutations from class B to class A. However, a mutation arising in a genotype of class B might alternatively produce a genotype of class C; this probability of this occurrence is given by the chance that the mutation is non-neutral, $(1 - q)$, times the probability that a non-neutral mutation will produce the distinguished beneficial phenotype out of K phenotypic neighbors, $\frac{1}{K}$. These mutation rates are summarized in Supplementary Figure 1.

We assume that the population dynamics follow a discrete-time Moran model with a total population size N: at each time step, a randomly-chosen individual produces one offspring that replaces any individual, including the parent, with equal probability. We assume that all individuals are equally likely to give birth, though our results remain unchanged if we assume one of class A or B is weakly selected.



Supplementary Figure 1 – Probability of mutation within and between genotype classes

We are interested in the first time to arrival of an individual of type C. Prior to the first arrival, the total number of individuals is held fixed at N, and so it suffices to keep track of the number of individuals in class B until the first arrival. When the first individual of class C arises, we assume that the process jumps to an absorbing "graveyard" state Δ, at which time we stop the process. We denote this Markov stopping time by τ_Δ. We wish to compute the expectation of τ_Δ in terms of the parameters K, P, μ, N, and q.

Let $X_N(t)$ denote the number of individuals in class B in the t^{th} time step for $t < \tau_\Delta$. For $t \geq \tau_\Delta$ we let $X_N(t) = \Delta$. Then $X_N(t)$ is a discrete-time Markov chain on the set $\mathcal{I} \stackrel{\text{def}}{=} \{0, \ldots, N\} \cup \{\Delta\}$. The transition probabilities of this chain,

$$Q_{i,j}^N = \mathbb{P}\left\{X_N(t+1) = j | X_N(t) = i\right\},$$

are given by

$$Q_{i,i+1}^N = (1-\mu)\left(\frac{i}{N}\right)\left(\frac{N-i}{N}\right) + \mu\left(\frac{K}{P}\right) q\left(\left(\frac{N-i}{N}\right)^2 + \left(\frac{i}{N}\right)\left(\frac{N-i}{N}\right)\right),$$

$$Q_{i,i-1}^N = (1-\mu)\left(\frac{i}{N}\right)\left(\frac{N-i}{N}\right) + \mu\left(1-\frac{K}{P}\right) q\left(\left(\frac{i}{N}\right)^2 + \left(\frac{i}{N}\right)\left(\frac{N-i}{N}\right)\right),$$

$$Q_{i,\Delta}^N = \frac{\mu}{K}(1-q)\left(\frac{i}{N}\right),$$

$$Q_{i,i}^N = 1 - Q_{i,i+1}^N - Q_{i,i-1}^N - Q_{i,\Delta}^N,$$

for $0 \leq i \leq N$. Furthermore, $Q_{\Delta,\Delta}^N = 1$, and $Q_{i,j}^N = 0$ for all other pairs $i, j \in \mathcal{I}$. We use the notation $\mathbf{Q}_N$ to denote the matrix with entries $Q_{i,j}^N$, the generator of the Markov chain $X_N(t)$.

1.2 Two continuous-time approximations

The discrete process $X_N(t)$ is too complicated to consider directly, so we will introduce two different limiting processes that asymptotically capture the essential behavior. We will use these continuous-time approximations to derive an analytic expression for the mean adaptation time. This approximation will be asymptotically accurate for large N.

We analyze two different time- and mass-rescaled processes:

$$\begin{aligned} Y_{1,N}(t) &= \eta_{1,N}^{-1} X_N\left(\lfloor \alpha_{1,N} t \rfloor\right), \\ Y_{2,N}(t) &= \eta_{2,N}^{-1} X_N\left(\lfloor \alpha_{2,N} t \rfloor\right) \end{aligned}$$

where we define the choices of $\alpha_{1,N} \gg 1$, $\eta_{1,N} \gg 1$, $\alpha_{2,N} \gg 1$, and $\eta_{2,N} \gg 1$ below. For each $n = 1, 2$ note that $Y_{n,N}$ is a continuous-time processes, whereas X_N is discrete. One unit of time for $Y_{n,N}(t)$ corresponds to $\alpha_{n,N}$ time units in the discrete process $X_N(t)$. By definition, $Y_{n,N}(t)$ is $O(1)$ when $X_N(\lfloor \alpha_{n,N} t \rfloor) \propto \eta_{n,N}$.

We show below that for small initial values $X_N(0)$, by choosing $\eta_{1,N} = N^{\frac{1}{2}}$ and $\alpha_{1,N} = N^{\frac{3}{2}}$ the process $Y_{1,N}(t)$ converges in distribution to a diffusion process with killing, $Y_1(t)$, on $\mathbb{R}_+^\Delta \stackrel{\text{def}}{=} [0, \infty) \cup \{\Delta\}$, with transition density function

$$p(t, z, y)\, dz = \mathbb{P}\{Y_1(t) \in [z, z + dz) | Y_1(0) = y\},$$

satisfying the forward Kolmogorov equation

$$\frac{\partial p}{\partial t} = \frac{\partial^2}{\partial z^2}\left[zp(t, z, y)\right] - \frac{\partial}{\partial z}\left[\frac{\beta K q}{P} p(t, z, y)\right] - \frac{\beta(1 - q)}{K} zp(t, z, y).$$

where $\beta = N\mu$. Thus, when the number of adaptable individuals is small, back mutations have negligible effect and the rate of mutation to the adaptable type is effectively constant. Changes in number are a result of three effects, encapsulated in the three terms: genetic drift, mutations from non-adapable to adaptable types and finally mutations from adaptable types to the target phenotype. Even when the specific form of the mutation rates is varied, as we will consider below, this asymptotic form, and the interpretation of its three terms, remains unchanged.

We will also show below that by choosing $\eta_{2,N} = N$ and $\alpha_{2,N} = N$ the process $Y_{2,N}(t)$ converges to a pure jump process that jumps into state Δ with probability one, after a random time which is exponentially distributed with rate

$$\frac{\beta(1 - q)}{K} Y_2(0).$$

In the following sections, we give a proof of these two limits. Readers who are less interested in the formal details may skip ahead to Section 1.3 for the derivation of the expected first arrival time.

1.2.1 Generators of Stochastic Processes

Let $Y(t)$ be some continuous-time stochastic process, and let $\mathbb{P}_y$ and $\mathbb{E}_y$ denote probability and expectation conditioned on $Y(0) = y$, respectively. We recall from (Karlin & Taylor 1981, Ethier & Kurtz 2005) that the infinitesimal generator A of $Y(t)$ is

$$Af(y) \stackrel{\text{def}}{=} \lim_{t \downarrow 0} \frac{\mathbb{E}_y\left[f(Y(t))\right] - f(y)}{t},$$

for all f for which the limit on the right exists. We refer to all such f as the domain of A, $\mathfrak{D}(A)$. For continuous time and continuous state processes, the generator plays a role analogous to the transition matrix, and may be used, in conjunction with restrictions on the domain, to uniquely characterize the process. Infinitesimal generators provide a convenient unifying framework within which to study Markov processes.

For example, for a diffusion process with killing $Z(t)$, with probability transition density $p(t, z, y)$, *i.e.*

$$\mathbb{P}_y \{Z(t) \in (a, b)\} = \int_a^b p(t, z, y)\, dz,$$

satisfying the Komolgorov forward equation

$$\frac{\partial p}{\partial t} = \frac{1}{2}\frac{\partial^2}{\partial z^2}\left[a(z)p(t, z, y)\right] - \frac{\partial}{\partial z}\left[b(z)p(t, z, y)\right] - k(z)p(t, z, y),$$

the corresponding generator is

$$Af(y) = \frac{1}{2}a(y)f''(y) + b(y)f'(y) - k(y)f(y).$$

The domain of A consists of all twice differentiable functions satisfying appropriate boundary conditions (absorbing, reflecting, etc.) and vanishing at the graveyard point Δ.

A continuous time Markov jump process, where the time to the next jump, starting from y, is exponentially distributed with rate $r(y)$, and $\mu(y, dz)$ is the probability that the process jumps from y to z, has generator

$$Af(y) = r(y) \int f(z) - f(y)\, \mu(y, dz).$$

For the two limits described above, we have,

$$A_1 f(y) = yf''(y) + \frac{\beta K q}{P} f'(y) - \frac{\beta(1-q)}{K} yf(y). \tag{1.1}$$

corresponding to $\eta_{n,N} = N^{\frac{1}{2}}$, for $\eta_{n,N} = N$, we get generator

$$\begin{aligned} A_2 f(y) &= \frac{\beta}{K}(1-q)y\left(f(\Delta) - f(y)\right), \\ &= -\frac{\beta}{K}(1-q)yf(y) \end{aligned} \tag{1.2}$$

for f continuous in $[0,1]$ (and vanishing at Δ). The latter generator describes a Markov jump process with jump rate $r(y) = \frac{\beta}{K}(1-q)y$ and jump distribution given by a Dirac point mass, $\mu(y, dz) = \delta_\Delta(dz)$, *i.e.* all jumps are to the graveyard state. A_2 uniquely characterizes the corresponding process. For A_1 however, we also need to determine appropriate boundary conditions, which we discuss below.

1.2.2 Domain and Boundary Conditions for A_1

In general, the infinitesimal generator does not uniquely specify a diffusion process. We must also characterize the domain to which we apply the generator by determining or imposing appropriate boundary conditions. We do so by following Feller's boundary classification (Feller 1954*a*,*b*) (see also (Karlin & Taylor 1981, Ethier & Kurtz 2005)).

For A_1, ∞ is always a natural boundary, and cannot be reached in finite time. According to Feller's classification scheme we consider the parameter $\nu = \frac{1}{2}\left(1 - \frac{\beta q K}{P}\right)$. If $\nu < 0$, then 0 is a regular boundary, and it is an entrance boundary if $\nu > 0$. In the former case, the process can enter or leave at 0, and we may specify any boundary condition from absorption to reflection. In the latter, the process can enter the interval $(0, \infty)$ in finite time if started from 0, but can never reach 0 when started from an interior point. Since our original finite Markov chain model allows for mutation away from 0, we will impose reflecting boundary conditions at 0 for $\nu < 0$.

Whether the boundary behavior at 0 is entrance or reflecting, the corresponding condition for a function f belonging to $\mathfrak{D}(A)$ is the same,

$$\lim_{y \downarrow 0} \frac{f'(y)}{s(y)} = 0.$$

where

$$s(y) = y^{2\nu - 1}.$$

The natural boundary at ∞ leads to the requirement

$$\lim_{y \to \infty} f(y) = 0.$$

We also require that Δ be an absorbing state, which implies that

$$f(\Delta) = 0$$

for all $f \in \mathfrak{D}(A_1)$. Lastly, we require that all functions in $\mathfrak{D}(A_1)$ be continuous on $[0, \infty)$ and be twice continuously differentiable on $(0, \infty)$. Together, these conditions specify the domain of A_1 and uniquely determine the process $Y_1(t)$.

1.2.3 Deriving the Limiting Generators

To prove convergence, Theorem 6.5 in Chapter 1 of (Ethier & Kurtz 2005) proves that it suffices to show that, for appropriate choices of $\alpha_{n,N}$ and $\eta_{n,N}$,

$$\lim_{N\to\infty}\sup_{y\in G_{n,N}}|\alpha_{n,N}(\mathbf{Q}_N-\mathbf{I})f(y)-A_nf(y)|=0$$

for $n=1,2$ and $f\in\mathfrak{D}(A_i)\cap C^3(0,\infty)$, where

$$G_{n,N}=\{\eta_{n,N}^{-1}j|j\in\mathbb{N},1\le j\le N\},$$

is the set of all possible values of $Y_{n,N}(t)$.

Now, for $y=\eta_{n,N}^{-1}i\in G_{n,N}$,

$$\begin{aligned}\alpha_{n,N}(\mathbf{Q}_N-\mathbf{I})f(\eta_{n,N}^{-1}i)&=\alpha_{n,N}\sum_{j\in\mathcal{I}}Q_{i,j}^Nf(\eta_{n,N}^{-1}j)-\alpha_{n,N}f(\eta_{n,N}^{-1}i)\\&=\alpha_{n,N}\sum_{j\neq i,\Delta}Q_{i,j}^N\left(f(\eta_{n,N}^{-1}j)-f(\eta_{n,N}^{-1}i)\right)-\alpha_{n,N}Q_{i,\Delta}^N\left(f(\Delta)-f(\eta_{n,N}^{-1}i)\right)\end{aligned}$$

Taylor expanding $f(y)$ about $y=\eta_{n,N}^{-1}i$, and recalling that $f(\Delta)=0$, yields

$$\begin{aligned}&=\alpha_{n,N}\sum_{j\neq i,\Delta}Q_{i,j}^N\Big(\eta_{n,N}^{-1}(j-i)f'(\eta_{n,N}^{-1}i)+\frac{1}{2}\eta_{n,N}^{-2}(j-i)^2f''(\eta_{n,N}^{-1}i)+\frac{1}{6}\eta_{n,N}^{-3}(j-i)^3f'''(\xi)\Big)\\&\qquad-\alpha_{n,N}Q_{i,\Delta}^Nf(\eta_{n,N}^{-1}i)\\&=\alpha_{n,N}\left(Q_{i,i+1}^N-Q_{i,i-1}^N\right)\eta_{n,N}^{-1}f'(\eta_{n,N}^{-1}i)+\frac{1}{2}\alpha_{n,N}\left(Q_{i,i+1}^N+Q_{i,i-1}^N\right)\eta_{n,N}^{-2}f''(\eta_{n,N}^{-1}i)\\&\qquad+\frac{1}{6}\alpha_{n,N}\left(Q_{i,i+1}^N-Q_{i,i-1}^N\right)\eta_{n,N}^{-3}f'''(\xi)-\alpha_{n,N}Q_{i,\Delta}^Nf(\eta_{n,N}^{-1}i)\end{aligned}$$

for some ξ between $\eta_{n,N}^{-1}i$ and $\eta_{n,N}^{-1}j$. Substituting $y=\eta_{n,N}^{-1}i$ and $\beta=N\mu$ into $Q_{i,j}^N$ and simplifying yields

$$\begin{aligned}&=\alpha_{n,N}\beta q\left(\frac{\eta_{n,N}^{-1}}{N}\frac{K}{P}-\frac{1}{N^2}y\right)f'(y)\\&+\frac{\alpha_{n,N}}{2}\left(2\frac{\eta_{n,N}^{-1}}{N}y-2\frac{1}{N^2}y^2-2\frac{\eta_{n,N}^{-1}}{N^2}\beta y+\frac{\eta_{n,N}^{-2}}{N}\beta q\left(\frac{K}{P}\right)+\frac{\eta_{n,N}^{-2}}{N^4}\beta q\left(1-2\frac{K}{P}\right)y\right)f''(y)\\&\qquad+\frac{\alpha_{n,N}}{6}\left(\frac{\eta_{n,N}^{-3}}{N}\frac{K}{P}-\frac{\eta_{n,N}^{-1}}{N^2}y\right)f'''(\xi)-\frac{\alpha_{n,N}}{\eta_{n,N}^{-1}N^2}\frac{\beta}{K}(1-q)yf(y)\end{aligned}$$

We look for scalings for which all terms are finite, so that the limit is well-posed, and for which the killing term is non-zero. For the former, we must have $\eta_{n,N} = O(N^{\frac{1}{2}})$ or smaller, while for the latter, we must have $\alpha_{n,N} = \eta_{n,N}^{-1} N^2$.

If we take $\alpha_{1,N} = N^{\frac{3}{2}}$ and $\eta_{1,N} = N^{\frac{1}{2}}$, then

$$\alpha_{1,N}(\mathbf{Q}_N - \mathbf{I})f(y) \to A_1 f(y)$$

as $N \to \infty$, giving the diffusion limit, while if we take $\eta_{2,N} = N$ and $\alpha_{2,N} = N$, then

$$\alpha_{2,N}(\mathbf{Q}_N - \mathbf{I})f(y) \to A_2 f(y)$$

Scalings with $N \gg \eta_N \gg N^{\frac{1}{2}}$ give rise to a generator identical to A_2,

$$Af(y) = -\frac{\beta}{K}(1-q)yf(y),$$

only the domain now consists of functions continuous on $[0, \infty)$ and vanishing at Δ. All other choices of α_N and η_N result in a generator that becomes unbounded or tends to 0 as $N \to \infty$.

1.3 Adaptation time from fixed initial condition

In this section we calculate the expected value of τ_Δ, the first time at which the newly beneficial phenotype arises, for each of our two limiting processes $Y_1(t)$ and $Y_2(t)$, conditioned on $Y_1(0) = y$ or $Y_2(0) = y$:

$$T_1(y) \stackrel{\text{def}}{=} \mathbb{E}_y^{Y_1}[\tau_\Delta] \quad \text{and} \quad T_2(y) \stackrel{\text{def}}{=} \mathbb{E}_y^{Y_2}[\tau_\Delta].$$

Once we account for the rescaling of time and mass, each of these may be used as an asymptotic approximation to the expected first arrival time starting from a population with i individuals of class B, in our original discrete model:

$$T^N(i) \stackrel{\text{def}}{=} \mathbb{E}_i^{X_N}[\tau_\Delta].$$

In particular, for $i \propto N^{\frac{1}{2}}$, we will use $Y_1(t)$ to approximate $X_N(t)$, so

$$T^N(i) \sim N^{\frac{3}{2}} T_1(N^{-\frac{1}{2}} i), \tag{1.3}$$

while for $i \propto N$, we will use $Y_2(t)$ to approximate $X_N(t)$, so

$$T^N(i) \sim N T_2(N^{-1} i). \tag{1.4}$$

In fact, as we show in Section 1.4, Eq.(1.3) is valid for all $1 \leq i \leq N$.

When we consider generator A_2 (Eq.(1.2)), the mean adaptation time is simply the mean of an exponential distribution:

$$T_2(y) = \frac{1}{\frac{\beta}{K}(1-q)y}. \tag{1.5}$$

On this timescale, the numbers of individuals in class B remains fixed before C arrives. Unfortunately, this simple expression becomes unbounded as $y \to 0$, because individuals in class C can only arise from B individuals. Moreover, the expression above cannot be integrated against the steady-state distribution (see below). Therefore, we must use a different time scaling in order to determine the expected arrival time starting from very small numbers in class B. We will use the diffusion approximation with generator (1.1).

Since Δ is the only absorbing state for our process, $T_1(y)$ is the expected first time to absorption for $Y_1(t)$ conditioned on starting from y, which can be written as a Dirichlet problem for the generator (Karlin & Taylor 1981),

$$A_1 T_1(y) = -1$$
$$T \in \mathfrak{D}(A_1)$$

i.e.

$$yT_1''(y) + \frac{\beta K q}{P} T_1'(y) - \frac{\beta(1-q)}{K} y T_1(y) = -1$$
$$\lim_{y \downarrow 0} y^{1-2\nu} T_1'(y) = 0$$
$$\lim_{y \to \infty} T_1(y) = 0.$$

This can be readily solved via the method of Green's functions (Karlin & Taylor 1981).

We first find solutions $u_0(y), u_\infty(y)$ to the homogeneous problem $A_1 u(y) = 0$, with $u_0(y)$ and $u_\infty(y)$ satisfying the boundary conditions at zero ($\frac{u_0'}{s}(0+) = 0$) and infinity ($u_\infty(\infty) = 0$), respectively:

$$u_0(y) = y^\nu I_{-\nu}(ay) \qquad \text{and} \qquad u_\infty(y) = y^\nu K_\nu(ay).$$

where $a = \sqrt{\frac{\beta(1-q)}{K}}$ and, as before, $\nu = \frac{1}{2}\left(1 - \frac{\beta q K}{P}\right)$. $I_\nu(z)$ and $K_\nu(z)$ are the modified Bessel functions (Abramowitz & Stegun 1965). The Green's function $G(y, \xi)$ is then given by

$$G(y,\xi) = \begin{cases} m(\xi) u_0(y) u_\infty(\xi) & \text{if } \xi \leq y \\ m(\xi) u_0(\xi) u_\infty(y) & \text{if } y \leq \xi. \end{cases}$$

where

$$m(y) = y^{-2\nu}.$$

Intuitively, $G(y,\xi)\,d\xi$ represents the expected time $Y_1(t)$ spends in $[\xi, \xi + d\xi)$, given that $Y_1(0) = y$. Thus the expected adaptation time is given by

$$\begin{aligned} T_1(y) = \int_0^\infty G(y,\xi)\,d\xi &= u_0(y)\int_y^\infty m(\xi)u_\infty(\xi)\,d\xi + u_\infty(y)\int_0^y m(\xi)u_0(\xi)\,d\xi \\ &= y^\nu I_{-\nu}(ay)\left[\frac{\sqrt{\pi}}{2a}\left(\frac{a}{2}\right)^\nu \Gamma\left(\tfrac{1}{2}-\nu\right)\right. \\ &\quad + \frac{1}{2}\frac{\pi}{\sin(\nu\pi)}\left(\frac{\left(\frac{a}{2}\right)^\nu y}{\Gamma(1+\nu)}{}_1F_2\left(\tfrac{1}{2};\tfrac{3}{2},1+\nu;\left(\tfrac{ay}{2}\right)^2\right)\right. \\ &\quad \left.\left. - \frac{\left(\frac{a}{2}\right)^{-\nu} y^{1-2\nu}}{(1-2\nu)\Gamma(1-\nu)}{}_1F_2\left(\tfrac{1}{2}-\nu;\tfrac{3}{2}-\nu,1-\nu;\left(\tfrac{ay}{2}\right)^2\right)\right)\right] \\ &\quad + K_\nu(ay)\frac{\left(\frac{a}{2}\right)^{-\nu} y^{1-\nu}}{(1-2\nu)\Gamma(1-\nu)}{}_1F_2\left(\tfrac{1}{2}-\nu;\tfrac{3}{2}-\nu,1-\nu;\left(\tfrac{ay}{2}\right)^2\right), \quad (1.6) \end{aligned}$$

where ${}_1F_2(a_1;b_1,b_2;z)$ is a generalized hypergeometric function (Slater 1966). This equation, while unwieldy, gives an analytic expression for the mean adaptation time, starting from a fixed frequency of B-type individuals.

Using asymptotic properties of the Bessel functions, it is possible to show that

$$T_1(y) = \frac{1}{\frac{\beta(1-q)}{K}y}\left(1 + O\left(\tfrac{1}{y}\right)\right), \quad (1.7)$$

in agreement with (1.5), the expression we previously obtained for large initial frequencies. Thus (1.6) is valid not only for small initial frequencies of $X_N(0)$, but in fact gives a uniform asymptotic estimate for all starting frequencies $X_N(0)$.

1.3.1 Approximation near $y = 0$

From Abramowitz & Stegun (1965), we have

$$I_\nu(z) \sim \frac{1}{\Gamma(1+\nu)}\left(\frac{1}{2}z\right)^\nu, \qquad 0 \le z \ll 1. \quad (1.8)$$

and

$$K_\nu(z) = \frac{1}{2}\frac{\pi}{\sin(\nu\pi)}\left(\frac{1}{\Gamma(1-\nu)}\left(\frac{1}{2}z\right)^{-\nu} - \frac{1}{\Gamma(1+\nu)}\left(\frac{1}{2}z\right)^\nu\right)\left(1 + O\left(\tfrac{1}{z}\right)\right) \quad (1.9)$$

Combining these asymptotics expressions with hypergeometric series expressions for the integrals $\int_y^\infty \xi^{-\nu} K_\nu(a\xi)\,d\xi$ and $\int_0^y \xi^{-\nu} I_{-\nu}(a\xi)\,d\xi$, and simplifying using the reflection formula for the gamma function (Abramowitz & Stegun 1965),

$$\Gamma(\nu)\Gamma(1-\nu) = \frac{\pi}{\sin(\nu\pi)},$$

we obtain an asymptotic expression for $T_1(y)$ for y very small,

$$T_1(y) = \frac{\sqrt{\pi}}{2a}\frac{\Gamma\left(\frac{1}{2}-\nu\right)}{\Gamma(1-\nu)} + \frac{y}{2\nu}\left(1 - \frac{1}{1-2\nu}\right) + o(y)$$

In particular, we obtain a simple expression for the expected first arrival time from a population consisting only of A-type individuals,

$$T_1(0+) = \frac{\sqrt{\pi}}{2a}\frac{\Gamma\left(\frac{1}{2}-\nu\right)}{\Gamma(1-\nu)}.$$

We may further simplify this using Gauss' duplication formula for the gamma function (Abramowitz & Stegun 1965),

$$\Gamma(2z) = \frac{2^{2z-1}}{\sqrt{\pi}}\Gamma(z)\Gamma\left(z+\tfrac{1}{2}\right),$$

to obtain

$$\begin{aligned} T_1(0+) &= \frac{2^{-2\nu-1}}{a}\frac{\Gamma\left(\frac{1}{2}-\nu\right)^2}{\Gamma(1-2\nu)} \\ &= \frac{2^{\frac{\beta qK}{P}-2}}{\sqrt{\frac{\beta(1-q)}{K}}}\frac{\Gamma\left(\frac{1}{2}\frac{\beta qK}{P}\right)^2}{\Gamma\left(\frac{\beta qK}{P}\right)}. \end{aligned} \tag{1.10}$$

1.4 Adaptation time from steady state

Although (1.6) gives an analytic expression for the mean adaptation time starting from a fixed initial number of class-B genotypes, $X_N(0)$, we are actually interested in the adaptation time starting from a population in steady state. Therefore, we must integrate (1.6) over the probability distribution for the frequency of class-B genotypes in steady state.

Prior to the environmental shift, we have assumed that all genotypes not expressing the wild-type, including those in class C, are inviable. Therefore, the relative frequencies of class A and class B genotypes follow a standard, neutral Moran model

with asymmetric mutations between two types. Genotypes of class A mutate to class B at rate $\beta q\frac{K}{P}$, and class B mutates to class A at rate $\beta q\left(1-\frac{K}{P}\right)$. Therefore the steady state frequency of class B individuals follows the well-known beta distribution (Ewens 2004) with probability density function

$$\mathbb{P}\left\{X_N(0)=x\right\}=g(x)=\frac{\Gamma(\beta q)}{\Gamma\left(\beta\left(1-\frac{K}{P}\right)q\right)\Gamma\left(\beta\left(\frac{K}{P}\right)q\right)}x^{\beta\left(1-\frac{K}{P}\right)q-1}(1-x)^{\beta\left(\frac{K}{P}\right)q-1}.$$

Thus, the expected adaptation time, in generations, starting from a population at steady state is

$$T=N^{\frac{1}{2}}\int_0^1 g(x)T_1(xN^{\frac{1}{2}})\;dx. \tag{1.11}$$

The integrand above is difficult to compute numerically for large arguments, x. Fortunately, for x large we have the asymptotic expansion given by Eq. 1.7. Therefore, in practice we numerically evaluate the expression above by dividing the integral into two regimes:

$$T\approx N^{\frac{1}{2}}\int_0^{3/\sqrt{N}} g(x)T_1(xN^{\frac{1}{2}})\;dx+\int_{3/\sqrt{N}}^1 g(x)T_2(x)\;dx$$

When $K\ll P$, almost all of the mass of the Dirichlet distribution is concentrated near $x=0$; *i.e.* with very high probability there are few or no adaptable types at time $t=0$. Thus, some degree of robustness is necessary for the arrival of sufficient adaptable types to survive drift and mutate to the target phenotype. As q increases from 0, the rate of arrival of adapted phenotypes increases, accelerating the arrival of the target type. However, when q approaches 1, the rate of mutation to the target type becomes exceedingly small, creating the observed non-monotonicity in the expected arrival time.

2 The relationship between robustness and phenotypic diversity

In this section we analyze how robustness, q, is related to the diversity of phenotypes produced by mutation, each generation, in a population at steady state. This calculation avoids the temporal complexity of waiting times and addresses a simple, general question: can populations that are more robust to mutation ever produce greater phenotypic diversity?

As before, we assume that one phenotype is fit and that the P alternative phenotypes are lethal. Mutations to a given genotype can produce only K phenotypes,

q is the proportion of mutations that are neutral, and a neutral mutation produces a novel genotype with a new set of K mutational neighbors. Diversity is measured as the expected number of distinct lethal phenotypes produced in a single generation, which we refer to as $D(q)$.

We wish to understand under what parameters a population with *some* level of robustness can produce more phenotypic diversity than a comparable population with no robustness in the genotype-phenotype map. In other words, we wish to understand the derivative $\frac{dD(q)}{dq}\Big|_{q=0}$. In this section we derive an analytic expression for this derivative in terms of the model parameters K, P, μ, and N.

As before we consider a haploid population of N asexuals which mutate at a rate μ and reproduce according to the Moran process. Let A denote the number of distinct neutral genotypes present in the population. We define

$$\beta_q = N\mu(1-q)$$

and

$$\beta_n = N\mu q \frac{1-\mu^{N+1}q^{N+1}}{1-\mu q}.$$

β_q denotes the expected number of non-neutral mutations that arise in the population, each generations. β_n is the neutral mutation rate appropriate to the infinite-alleles Moran model (Ewens 2004), which we use to determine the limiting number and distribution of neutral alleles.

Let $\alpha_{A,m}$ denote the expected number of novel phenotypes among m mutations arising in a population with A neutral types, given that each mutation gives rise to one of K equally-probable neighboring phenotypes. Then, the number of non-neutral mutations is Poisson distributed with mean β_q, and

$$D(q) = \sum_{m=0}^{\infty} \frac{\beta_q^m e^{-\beta_q}}{m!} \sum_{k=0}^{\infty} \alpha_{k,m} \mathbb{P}\{A=k\}. \tag{2.1}$$

If $q = 0$, there there is only one neutral genotype present, i.e. $A = 1$ with probability one. This genotype can produce up to K alternative phenotypes by mutation. The expected number of non-neutral mutations in a single generation is then $\beta_0 = N\mu$. We thus have

$$D(0) = \sum_{m=0}^{\infty} \frac{\beta_0^m e^{-\beta_0}}{m!} \alpha_{1,m} \tag{2.2}$$

We now turn to computing $\alpha_{1,m}$. Let $\xi_i = 1$ if the i^{th} mutation gives rise to a novel phenotype, and 0 otherwise. Then,

$$\xi = \sum_{i=1}^{m} \xi_i$$

is the number of distinct phenotypes, and

$$\mathbb{P}\{\xi_i = 1\} = \left(1 - \frac{1}{K}\right)^{i-1}.$$

Thus,

$$\alpha_{1,m} = \mathbb{E}\left[\xi\right] = \sum_{i=1}^{m} \mathbb{E}\left[\xi_i\right] = K\left(1 - \left(1 - \frac{1}{K}\right)^{m}\right)$$

We now turn to $D(q)$ for $q > 0$. Since we are only interested in the derivative $\left.\frac{dD(q)}{dq}\right|_{q=0}$, we need only determine $D(q)$ up to $o(q)$. To this end, we observe that

$$\beta_n = N\mu q\left(\sum_{n=0}^{N} \mu^n q^n\right) = \beta_0 q + O(q^2).$$

Next, we consider $\mathbb{P}\{A = k\}$. The distribution of neutral genotypes in the infinite-alleles Moran model may be obtained via Hoppe's urn (Ewens 2004). Briefly, we start with a population of size 1. At each time step, with probability $\frac{i-1}{i-1+\beta_n}$ a random individual produces a clone, while with probability $\frac{\beta_n}{i-1+\beta_n}$, we add an individual with a novel genotype. After N iterations, we have a population with N individuals, for which

$$\mathbb{P}\left\{A = k\right\} = \sum_{2 \le j_1 < j_2 < \ldots \le N} \prod_{i=1}^{k} \frac{\beta_n}{j_i - 1 + \beta_n} \prod_{\substack{2 \le i \le N \\ i \notin \{j_1, \ldots, j_k\}}} \frac{i-1}{i-1+\beta_n}.$$

Thus,

$$\begin{aligned} \mathbb{P}\left\{A = 1\right\} &= \prod_{i=2}^{N} \frac{i-1}{i-1+\beta_n} \\ &= 1 - \beta_0 q H_{N-1} + O(q^2), \end{aligned} \tag{2.3}$$

where H_n is the n^{th} partial Harmonic sum,

$$H_n \equiv \sum_{i=1}^{n} \frac{1}{i}.$$

Similarly,

$$\begin{aligned}\mathbb{P}\{A=2\} &= \sum_{j=2}^{N} \frac{\beta_n}{j-1+\beta_n} \prod_{i\neq j} \frac{i-1}{i-1+\beta_n} \\ &= \beta_0 q H_{N-1} + O(q^2),\end{aligned} \tag{2.4}$$

while $\mathbb{P}\{A=k\} = O(q^{k-1})$ for $k > 2$. Thus, in the limit $q \to 0$, we need only consider populations composed of one or two neutral genotypes.

We next turn to $\alpha_{2,m}$, the expected number of unique phenotypes given m mutations and $A = 2$. If p_i is the probability that there are $N - i$ individuals of the first genotype and i of the second, the expected number of unique phenotypes is:

$$\alpha_{2,m} = \sum_{i=1}^{N-1} p_i \gamma_{m,i}, \tag{2.5}$$

where $\gamma_{m,i}$ is the expected number of phenotypes, resulting from m mutations in a population with $(N-i)$ type 1 individuals and i type 2 individuals. Note that types 1 and 2 are equivalent (they both encode the wild-type phenotype) except that each has an independent set of K phenotypic neighbors.

We find $\gamma_{m,i}$ as before: let $\xi_{j,k}$ ($j = 1, 2$, $k = 1, \ldots, i$) be 1 if the k^{th} mutation to one of the type j individuals gives rise to a new phenotype. Without loss of generality, we first consider mutations in type 2. As above, the total number of mutations to type 1 is

$$\xi_1 = \sum_{k=1}^{i} \xi_{1,k}$$

and $\mathbb{E}[\xi_1] = \alpha_{1,i}$. Now $\xi_{2,k}$ is 1 if and only if the k^{th} mutation gives rise to different phenotype from all $k-1$ previous mutations to type 2 individuals, and moreover, that phenotype was not produced by a mutation to a type 1 individual. For the latter to be true, either the new phenotype is inaccessible to type 1, with probability $1 - \frac{K}{P}$, or the phenotype is adjacent to type 1, with probability $\frac{K}{P}$, but was not one of the ξ_1 phenotypes that arose. Thus,

$$\mathbb{P}\{\xi_{2,k} = 1\} = \left(1 - \frac{1}{K}\right)^{k-1} \left(\frac{K}{P}\frac{K-\xi_1}{K} + 1 - \frac{K}{P}\right)$$

Thus, if $\xi_2 = \sum_{k=1}^{m-i} \xi_{2,k}$,

$$\begin{aligned}\mathbb{E}[\xi_1 + \xi_2] &= \alpha_{1,i} + \alpha_{1,m-i}\left(\frac{K}{P}\frac{K-\alpha_{1,i}}{K} + 1 - \frac{K}{P}\right) \\ &= \alpha_{1,i} + \alpha_{1,m-i} - \frac{\alpha_{1,i}\alpha_{1,m-i}}{P},\end{aligned}$$

and, since the m mutations are binomially distributed among the two genotypes,

$$\gamma_{m,i} = \sum_{j=0}^{m} \binom{m}{j} \left(\frac{i}{N}\right)^{j} \left(1 - \frac{i}{N}\right)^{m-j} \left(\alpha_{1,j} + \alpha_{1,m-j} - \frac{\alpha_{1,j}\alpha_{1,m-j}}{P}\right). \tag{2.6}$$

Finally, the probability that there are i individuals with the first genotype, given that $A = 2$, can be determined via Hoppe's urn. At each step, a new individual is added. The first individual of the second genotype arrives at step $k+1$ with probability

$$\frac{\frac{\beta_n}{\beta_n + k}}{\sum_{k=1}^{N-i} \frac{1}{\beta_n + k}},$$

joining the k individuals of type 1 already present. Treating the k type 1's and the single type 2 as separate lineages, the probability of i individuals of type 2 in the total population of N is simply the proportion of partitions of N with one partition of size i (Joyce & Tavaré 1987),

$$\frac{\binom{N-i-1}{k-1}}{\binom{N-1}{k}}.$$

Summing over all k gives

$$p_i = \left(\sum_{k=1}^{N-i} \frac{1}{\beta_n + k}\right)^{-1} \sum_{k=1}^{N-i} \frac{1}{\beta_n + k} \frac{\binom{N-i-1}{k-1}}{\binom{N-1}{k}} \tag{2.7}$$

$$= \frac{1}{H_{N-1}} \sum_{k=1}^{N-i} \frac{(N-i-1)!(N-k-1)!}{(N-i-k)!(N-1)!} + O(q) \tag{2.8}$$

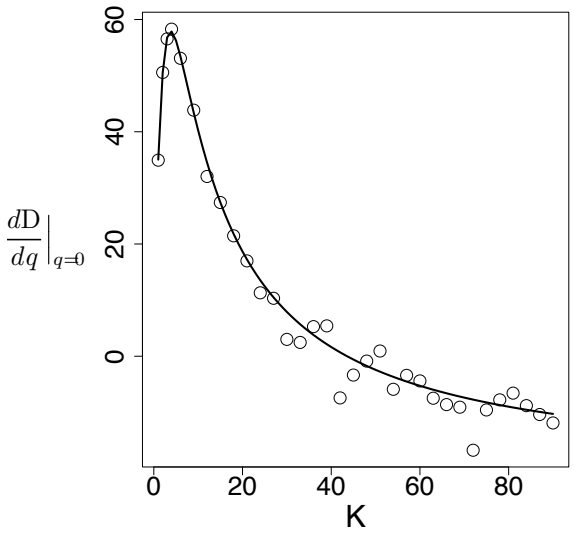
Combining (2.1), (2.3) and (2.4), we have,

$$\begin{aligned} D(q) &= \sum_{m=0}^{\infty} \left(\frac{\beta_q^m e^{-\beta_q}}{m!} \left(\mathbb{P}\left\{A = 1\right\} \alpha_{1,m} + \mathbb{P}\left\{A = 2\right\} \alpha_{2,m} \right) \right) + O(q^2) \\ &= \sum_{m=0}^{\infty} \left(\frac{\beta_0^m e^{-\beta_0}}{m!} \left((1 - \beta_0 q H_{N-1}) \alpha_{1,m} + \beta_0 q H_{N-1} \alpha_{2,m} \right) \right) + O(q^2) \end{aligned}$$

while subtracting (2.2), dividing by q, and taking $q \to 0$ yields

$$\frac{dD(q)}{dq}\Big|_{q=0} = \sum_{m=0}^{\infty} \frac{\beta_0^m e^{-\beta_0}}{m!} \left[\beta_0 H_{N-1} (\alpha_{2,m} - \alpha_{1,m}) - (m - \beta_0) \alpha_{1,m} \right]. \tag{2.9}$$

Supplementary Figure 2 plots Eq. (2.9) for $P = 90$, $N = 300$, and $\mu = 0.04$ for a range of K. Also plotted are simulation results for $\frac{\Delta D}{\Delta q}$ calculated at $\Delta q = 0.0001$, for 100 million replicates for each point.



Supplementary Figure 2 – Predictions using Eq. 2.9 (line) compared with the means of 100 million replicate simulations at q = 0.0001 (circles).

3 Simulation Methods

Monte Carlo simulations were performed in two main ways. For the data in Figure 2 & 3 in the main text and Supplementary Figures 4, matrices of transition probabilities were pre-calculated and used to simulate individual replicates of a Markov chain. Starting distributions for a two-allele Moran model were calculated using Eq. 3.58 in (Ewens 2004). These simulations were validated by comparison to individual-based simulations of populations of N genotypes. Similar individual-based simulations were also used to generate the data shown in Supplementary Figure **??**.

Code for all simulations was written in C or C++ and compiled using gcc 4.0.1. The GSL libraries (v.1.9) were used for pseudorandom number generation, special functions, and probability distributions.

4 Quantitative effects of N and μ on the optimal level of robustness

We used numerical methods to explore how the adaptation time from steady state, given by Eq. 1.11, and the mutational diversity, described above in Section 2, depend on the population parameters N and μ. The tables below show the approximate q

which minimizes the adaptation time, or maximizes the mutational diversity, for a given N and μ. Note that the effects of N and μ can be approximated by considering only their product, β, but that β affects our measurements of evolvability in qualitatively different ways: increasing β decreases the q which minimizes adaptation time, but increases the q which maximizes mutational diversity.

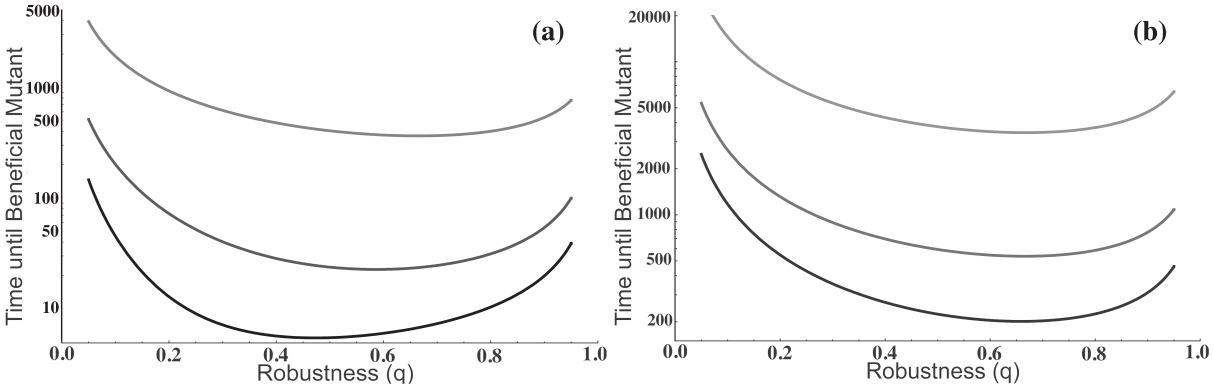
N	μ	β	q of minimum time	q of maximum diversity
5000	0.001	5	0.1	0.67
10000	0.0005	5	0.13	0.68
10000	0.001	10	0.17	0.66
10000	0.0015	15	0.19	0.64
15000	0.001	15	0.19	0.65
10000	0.003	30	0.2	0.58
30000	0.001	30	0.2	0.6

Table 1 – Changes in the q which minimizes adaptation time, and the q which maximizes mutational diversity, with population size and mutation rate.

5 Relaxation of model assumptions

We relax each of the four main assumptions underlying the model presented in the main text: (1) a single neutral mutation changes completely the phenotypes neighboring a genotype; (2) the number of phenotypes, K, in a genotype's mutational neighborhood is independent of its robustness, q; (3) K and q do not vary among the genotypes on a neutral network and; (4) alternative phenotypes are lethal (or adaptive).

To relax the first assumption we introduced a new parameter, f, defined as the fraction of the K phenotypes neighboring a genotype which are redrawn after a mutation. Our original model is therefore equivalent to this generalization when $f = 1$. Considering this new parameter in light of the analysis above, it is clear that the only effect of f is to scale the mutation rates between neutral genotypes that are or are not adaptable – i.e. class B and class A genotypes, respectively. The mutation rate from A to B is therefore $\mu q f \left(\frac{K}{P}\right)$. Since decreasing f is equivalent to increasing P, values of $f < 1$ actually broaden the range of parameters for which robustness and evolvability are positively correlated. This is illustrated in Supplementary Figure 3, which shows analytic predictions of adaptation times for a range of f.



Supplementary Figure 3 – The predicted mean time before the arrival of a beneficial mutation with varying f. $N = 10,000$, $\mu = 0.001$, $K = 5$; a) $P = 15$; b) $P = 100$. $f = 1$, bottom curves; $f = 0.5$, middle curves; and $f = 0.1$, top curves. Decreasing f increases adaptation time, but the shape of the relationship with robustness remains consistent.

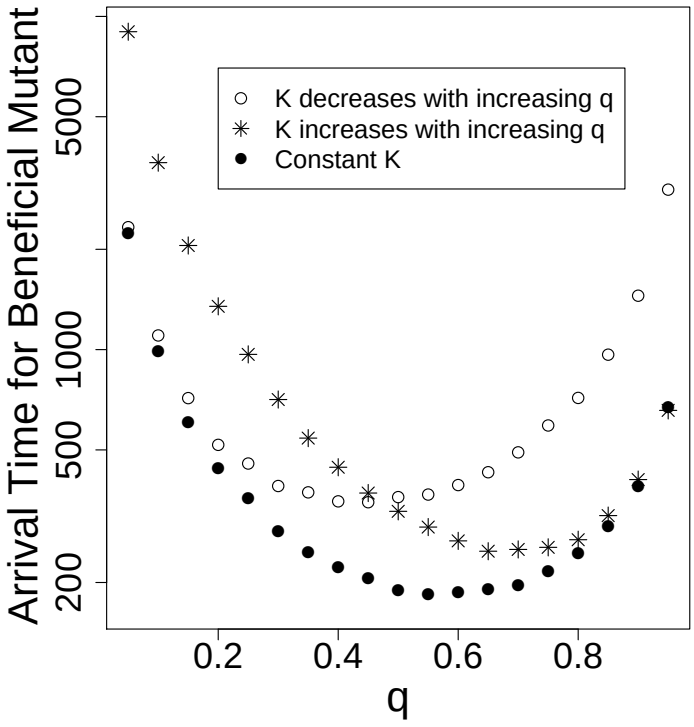
The second assumption is an implicit description of the topology of neutral networks. In our original model, the fraction of mutations that are neutral (q) was independent of the phenotypic diversity of non-neutral mutations (K). We relax this assumption by allowing these two quantities to be correlated. As q increases, the number of mutational neighbors with distinct phenotypes might be expected to decrease, and so the richness of those phenotypic mutants, K, might also be expected to decrease. Therefore, in relaxing the second assumption we explored a linear, negative relationship between q and an effective value of K, denoted $\hat{K}(q)$:

$$\hat{K}(q) = \lfloor K(1-q) \rfloor \tag{5.1}$$

This gives rise to an alternate limiting process

$$\frac{\partial p}{\partial t} = \frac{\partial^2}{\partial z^2}\left[zp(t,z,y)\right] - \frac{\partial}{\partial z}\left[\frac{\beta K q(1-q)}{P} p(t,z,y)\right] - \frac{\beta}{K} zp(t,z,y),$$

for which the expected arrival time, $T_1(y)$ still takes the form of Equation (1.6), given an appropriate choice of a and ν (see Section 1). This process also results in a non-monotonic relationship between the arrival time of the adapted phenotype and the robustness, q, but the mechanism is slightly different. Here, the rate of mutation from the adaptable types to the target phenotype is constant, while the rate of mutation to the adaptable class approaches 0 for q near 0 or 1, and therefore leads to a non-monotonic relationship between robustness and evolvability.



Supplementary Figure 4 – The mean time before arrival of a beneficial mutation, with either K constant or $\hat{K}$ correlated with q according to Eq. 5.1 or Eq. 5.2. $N = 10000$, $\mu = 0.001$, $P = 200$, and $K = 20$. Points are means of 10,000 replicate simulations.

Supplementary Figure 4 shows the results of simulations when K is negatively correlated with q according to Eq. 5.1. Comparing the open circles to the closed circles, which show numerical results for the same parameters and K independent of q, we note a quantitative difference for larger values of q. However, the non-monotonicity is preserved, suggesting that our qualitative conclusions are not sensitive even to a strong negative correlation between K and q.

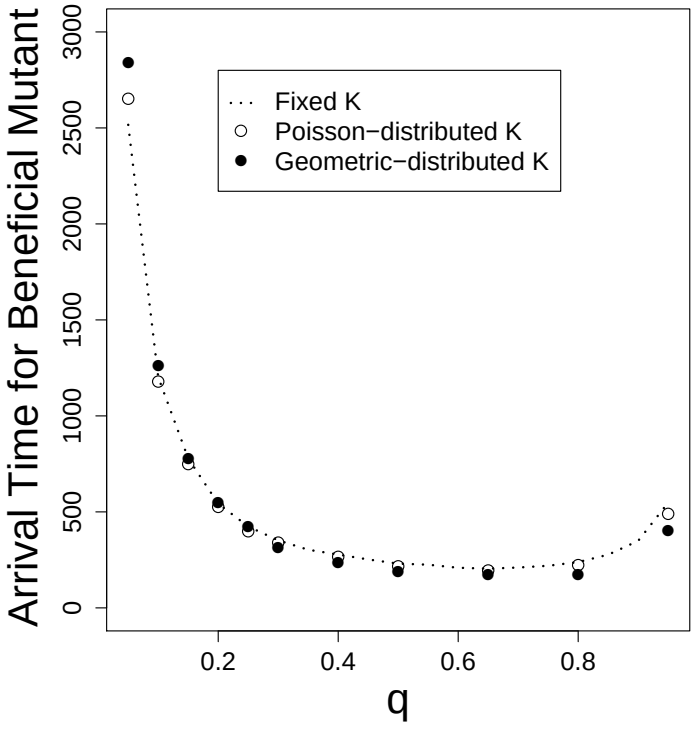
For the sake of completeness, we also explored a positive correlation between q and K:

$$\hat{K}(q) = \lfloor Kq \rfloor + 1 \tag{5.2}$$

Simulations in this regime produced qualitatively similar results; see Supplementary Figure 4.

To relax the third assumption, we performed simulations in which either K or q

were allowed to vary by genotype. (P is a property of an entire fitness landscape and so cannot vary with genotype). To vary K, we assigned each genotype in our neutral network a value of K drawn independently from a distribution. K is therefore effectively re-drawn following a neutral mutation. Supplementary Figure 5 shows the relationship between robustness, q, and mean adaptation time in three different situations: constant $K = 5$, K drawn from a Poisson distribution with mean 5, and K drawn from a shifted geometric distribution with mean 5 (distributions were truncated at $K = P$). Populations were allowed to evolve before the environmental shift until the number of adaptable individuals, and the distribution of K, had equilibrated. As the figure demonstrates, our results are remarkably robust even when we allow K to vary across the neutral network. In fact, our analytic prediction for the mean adaptation time is highly accurate, after replacing the fixed value of K, in the original model, with the mean of the distribution from which K is drawn, in the extended model.

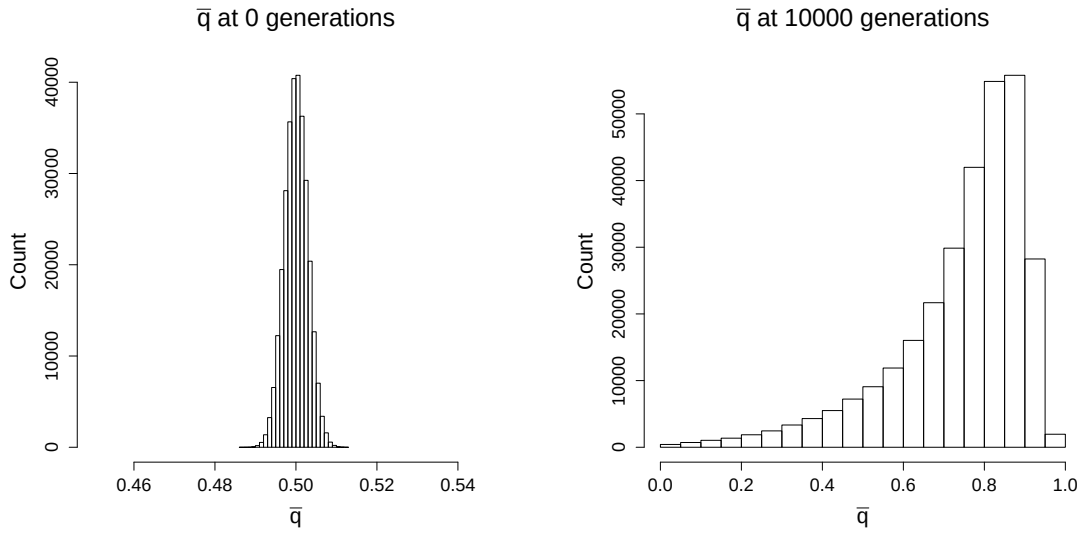


Supplementary Figure 5 – The mean time before arrival of a beneficial mutation, for constant K and variable K. $N = 10000$, $\mu = 0.001$, $P = 100$, and mean $K = 5$. Points show the means of at least 4,000 replicate simulations.

As before, this can be explained by considering the rate of mutation into the family of adaptable genotypes and the rate of mutation from the adapted types to the target phenotype. We omit the calculations here, but it can be shown that the rate of mutation into the adaptable class takes the form $\mu q \left(\frac{\mathbb{E}[K]}{P}\right)$, while the rate of mutation into the target phenotype is obtained by averaging $\frac{\beta(1-q)}{K}$ over all individuals in the population. As before, these two rates determine the expected arrival time of the target phenotype.

We also relaxed the third assumption by allowing q to vary among genotypes in a neutral network. To do so, we performed simulations in which an offspring inherits its parent's value of q, plus a random perturbation whenever a neutral mutation occurs. These perturbations are drawn from a Gaussian with standard deviation 0.1; if a perturbation would produce a q outside of $[0, 1]$, then q is left unaltered. This method of mutating q allows robustness to evolve. Prior theory suggests that a population should evolve towards elevated robustness (higher q) in a fixed environment, because more robust individuals have an effective selective advantage of order μ (van Nimwegen et al. (1999), Forster et al. (2006)). This is indeed what we observe: the mean q in a population, denoted $\overline{q}$, tends to increase over time in a fixed environment. Figure 6 shows this behavior by plotting the distribution of $\overline{q}$ across the ensemble of replicate simulations. At the beginning of a simulation, each individual is assigned a q from the uniform distribution, and $\overline{q}$ is therefore approximately normally distributed with a mean of 0.5. After each population evolves for 10,000 generations in a fixed environment, the ensemble distribution of $\overline{q}$ has shifted significantly towards larger values (Supplementary Figure 6, reflecting the evolution of robustness in a fixed environment.

After 10,000 generations, we assign one of the P alternative phenotypes the highest fitness, and we record the subsequent arrival time of the first such adaptive mutation. The relationship between $\overline{q}$ in a population at the time of this environmental shift and the subsequent time before the arrival of a beneficial mutation is summarized in Supplementary Figure 7. In this figure, the results of almost 300,000 replicates are binned according to $\overline{q}$ and plotted as grey points. The line illustrates analytical results results with fixed q, while black dots show comparable simulation results for fixed q. Although q may continue to evolve between the 10,000th generation and the eventual arrival of a beneficial mutant, we nonetheless find that $\overline{q}$ is an excellent predictor of adaptation time. In fact, our analytic formula for the mean adaptation time is highly accurate, after replacing the fixed value of q in the original model with observed mean $\overline{q}$ in the population. We also performed simulations in which we waited 100,000 generations before allowing beneficial mutations; these

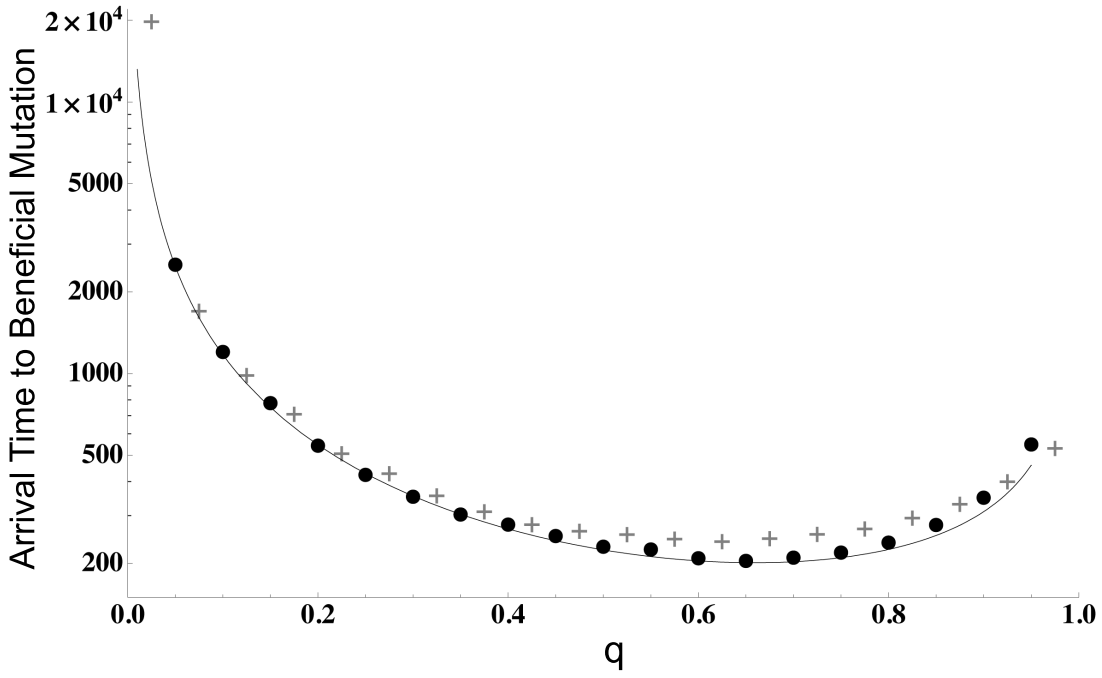


Supplementary Figure 6 – (a) Histograms of $\overline{q}$ among nearly 300,000 replicates at generation zero (left) and generation 10,000 (right). Parameters: $N = 10000$, $\mu = 0.001$, $P = 100$, and $K = 5$.

simulations produced very similar results. Thus, our results are robust with respect to significant variation in q across a neutral network and the associated evolution of q within a population.

For the sake of completeness, we also allowed f to vary across the neutral network. We assigned each genotype in our neutral network a value of f drawn independently from a distribution. f is therefore effectively re-drawn following a neutral mutation. Supplementary Figure 8 shows the relationship between robustness, q, and mean adaptation time in two different situations: constant $f = 0.5$, f drawn from a normal distribution with mean 0.5 and variance 0.1 (truncated at zero and one). Populations were allowed to evolve before the environmental shift until the number of adaptable individuals, and the distribution of f, had equilibrated. As the figure demonstrates, our results are robust even when we allow f to vary across the neutral network. In fact, our analytic prediction for the mean adaptation time is highly accurate, after replacing the fixed value of f, in the original model, with the mean of the distribution from which f is drawn, in the extended model.

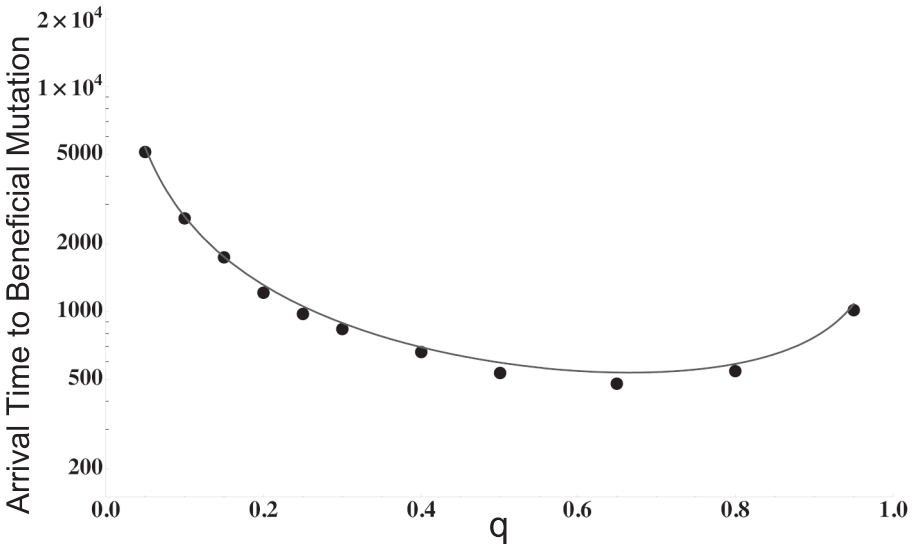
Finally, we considered the effects of relaxing the fourth assumption. In our original model all mutations were either neutral or lethal, prior to the environmental shift. We simulated an alternative model in which the previously-lethal phenotypes



Supplementary Figure 7 – Mean arrival time of the first beneficial mutant. The grey crosses depict means binned according to the value of $\overline{q}$ at the time of the environmental shift. The solid line shows comparable analytical results for fixed q, while the black circles show simulations for fixed q. $N = 10000$, $\mu = 0.001$, $P = 100$, and $K = 5$.

were assigned fitness $1 - s$, reflecting a disadvantage of size s compared the fitness of the focal phenotype. As before, individuals are chosen to reproduce in proportion to their fitnesses.

To describe mutation among the phenotypic classes in this more general model it is helpful to extend the notation used above. Genotypes of type C express the phenotype that will be beneficial after the environment changes, and they have fitness $1 - s$ prior to that shift. Genotypes of type B can produce type C mutants, and they can have fitness 1, which we will call B_{fit}, or $1 - s$, which we denote by B_{del}. Similarly, type A genotypes cannot produce type C, and they may either be A_{fit} or A_{del}. Any mutation, whether neutral, deleterious, or beneficial, causes the set of K neighbors to be redrawn. Therefore, A_{del}-types may or may not be able to mutate back to A_{fit}, depending on the phenotypic neighborhood of that type. Finally, we impose a necessary constraint on phenotypic neighbors: a mutant's phenotypic neighborhood *must* contain its parent's phenotype. So, if a type C genotype mutates



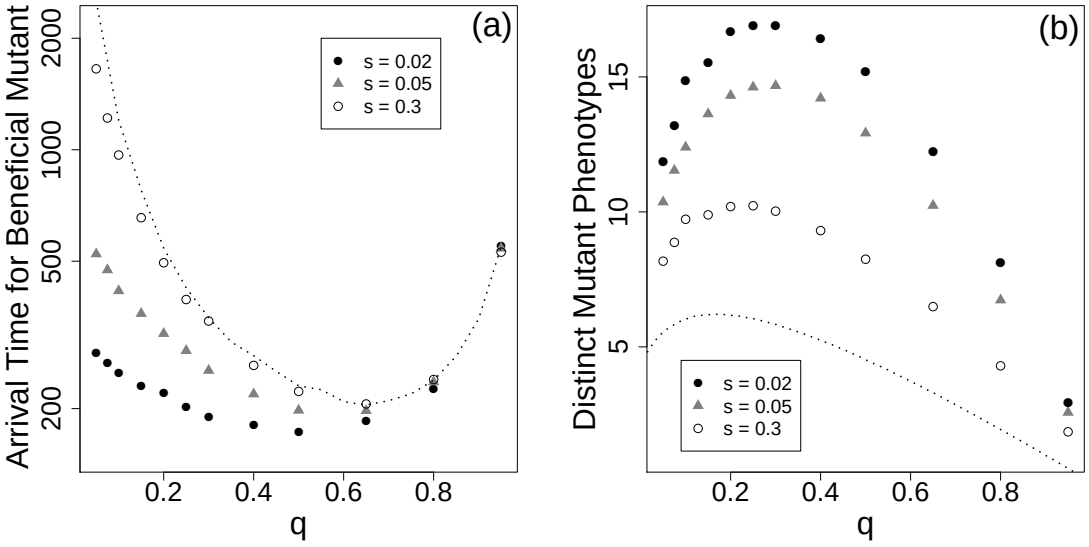
Supplementary Figure 8 – The mean time before arrival of a beneficial mutation, for constant f and variable f. $N = 10000$, $\mu = 0.001$, $P = 100$, and mean $f = 0.5$. Points show the means of at least 4000 replicate simulations.

to a deleterious phenotype, it must become B_{del}.

We used individual-based Moran simulations to explore the time elapsed until the first type-C individual arises. Each simulation began with an average proportion $\frac{K}{P}$ individuals of type B_{fit}, with the remaining of type A_{fit}. Populations evolved for 30,000 generations before the environmental shift, at which time type C individuals become beneficial. 30,000 generations was determined to be sufficient for equilibrium by observing the distributions of B_{fit}, B_{del}, and type C individuals across an ensemble of evolving populations. We recorded the arrival time of the beneficial type in terms of the number of generations after the environmental shift; if type C individuals were present at the environmental shift, this time was recorded as zero.

When alternative phenotypes are strongly deleterious ($s \approx 1$) we expect to see little difference between the results of the original, lethal-mutation model and the new model. However, when $s \approx 0$ we anticipate a monotonic, negative relationship between robustness and evolvability. When $s = 0$ non-neutral mutations have no cost, contribute to diversity in phenotypic neighborhoods, and generate type C individuals. Therefore, we ask how large must s be to match qualitatively the results of our original, lethal-mutation model.

Supplementary Figure 9(a) confirms that the relationship between robustness and evolvability is still non-monotonic when alternative phenotypes are moderately dele-



Supplementary Figure 9 – Adaptation time and phenotypic diversity when alternative phenotypes are deleterious, but not lethal. $N = 10,000$, $\mu = 0.001$, $P = 100$, and $K = 5$ for all simulations shown. The dotted lines displays the means for the original model with lethal alternative phenotypes. (a) Mean adaptation time for three values of s, the fitness penalty of alternative phenotypes. Each point is the mean of 10,000 replicate simulations. (b) Mean number of unique mutant phenotypes. Each point is the mean of 5000 simulations.

terious. Similarly, we also used Moran simulations to measure phenotypic diversity in steady-state populations when the alternative phenotypes were deleterious but not lethal. These results, shown in Supplementary Figure 9(b), demonstrate that robustness and phenotypic diversity can also exhibit a non-monotonic relationship when alternative phenotypes are moderately disadvantageous. While the assumption that other phenotypes are lethal is mathematically convenient in our analyses, it can be relaxed without changing our qualitative results.

6 RNA Simulations

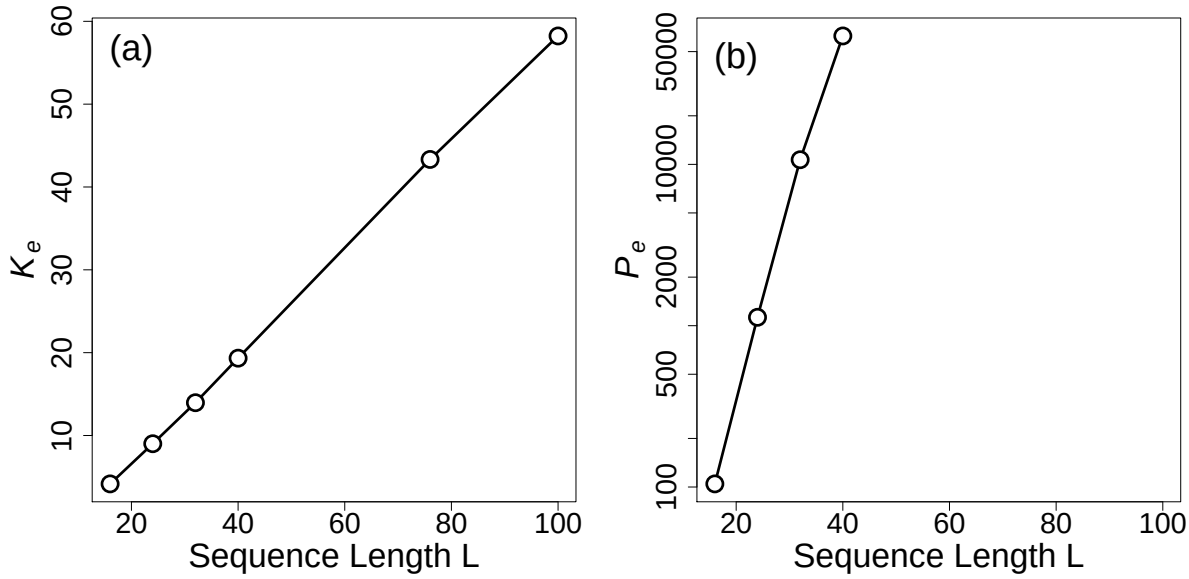
6.1 Measuring Epistasis in the RNA Landscape

We used the Vienna RNA package to calculate the minimum-free-energy secondary structures of RNA nucleotide sequences. The resulting genotype-phenotype map

was analyzed to estimate P, K, and f for a range of sequence lengths, L. While there are significant differences between the RNA landscapes and the type of abstract landscapes considered in our model, these estimated values provide some intuition about reasonable parameter values.

Our model assumes that out of P equally-likely phenotypes, K are equally-accessible from a given genotype. Since phenotypes are not evenly distributed in the RNA landscape, either globally or in the neighborhood of a single genotype, we must calculate some *effective* number of types for RNA. We note that, in our model, if we introduce two mutations into individuals of the same genotype, then the probability these mutation produce the same phenotype is $\frac{1}{K}$. Similarly, if the mutations occur in individuals with different genotypes, the probability of producing the same phenotype is $\frac{1}{P}$. Since these probabilities have an intuitive connection to the roles of K and P in our model, we define the effective K and P in the RNA landscape to yield the same probabilities. Additionally, some RNA sequences have the trivial, unfolded shape as their minimum free energy structure. We choose to regard this unfolded phenotype as inviable. We therefore define K_e as the inverse of the probability, p_{clone}, that two viable phenotypes produced by non-neutral mutations in clones of the *same* genotype, are the same. Similarly, we define P_e as the inverse of the probability, p_{random}, that two viable phenotypes produced by non-neutral mutations in individuals with random, viable genotypes, are the same. Both p_{clone} and p_{random} are measured by sampling random genotypes of length L with viable structures, and recording the phenotypes of random, non-neutral mutants of those genotypes; one hundred thousand genotypes were sampled for each L. K_e is therefore an average K across all genotypes in the landscape, while P_e is a lower bound on the number of accessible phenotypes; both are plotted in Supplementary Figure 10 for a range of L.

Estimating f is complicated by variation in K across the network. In keeping with the approach for K_e and P_e above, we relate f to the probability that two mutants have identical phenotypes. In our model, f determines the expected fraction of K neighbors that differ between two immediate neighbors on a neutral network, for P large. We therefore measure the probability, $p_{neighbor}$, that a viable, non-neutral mutant of genotype x has the same phenotype as a viable, non-neutral mutant produced from a neutral neighbor of x. $p_{neighbor}$ can be related to f, p_{clone}, and p_{random} as follows. First, note that we expect an adjacent pair of genotypes to have $(1-f)K$ phenotypic neighbors in common, and fK neighbors drawn from the remaining $P-(1-f)K$ possible phenotypes. A given phenotype can appear only once in a set of K neighbors, so if a phenotype falls among the shared $(1-f)$ portion of the neighborhoods, then it cannot fall among the unshared, redrawn f portion



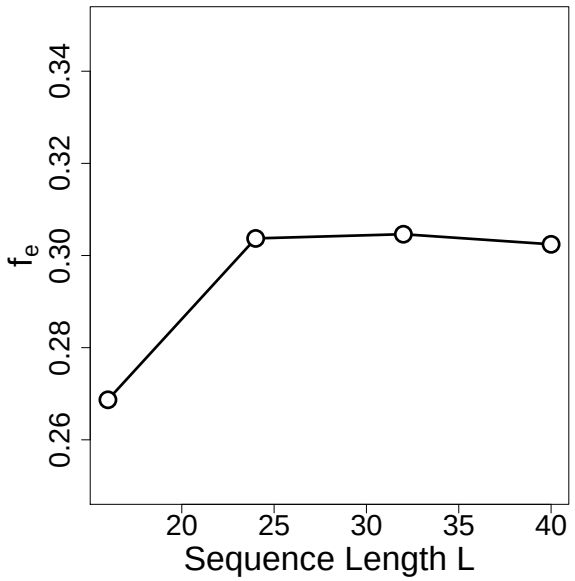
Supplementary Figure 10 – K_e and P_e for Simulated RNA Molecules of Length L

of either genotype's neighborhood. Therefore, we only need to consider the cases where both mutants produce phenotypes in the shared part of their neighborhoods, with probability $(1-f)^2$, or both produce phenotypes in the unshared portion of their neighborhoods, with probability f^2. In the former case, the probability that both mutants will have the same phenotype is $\frac{1}{(1-f)K_e}$, or $\frac{p_{clone}}{1-f}$. In the latter case, each mutant can be one of $P-(1-f)K$ possible phenotypes; if $K \ll P$ then the probability that they are the same phenotype is approximately $\frac{1}{P_e}$, or p_{random}. These considerations yield the following relationship among $p_{neighbor}$, p_{clone}, p_{random}, and f:

$$p_{neighbor} \approx (1-f)\, p_{clone} + f^2\, p_{random} \tag{6.1}$$

We use the equation above to define the effective value of f for the RNA landscape, after first measuring $p_{neighbor}$, p_{clone}, p_{random}. Supplementary Figure 11 shows values of f_e estimated in this way, for several sequence lengths L. These estimates complement the findings of (Sumedha et al. 2007), who examined the overlap in phenotypic distance of related, but not neighboring, RNA genotypes.

Together these estimates for the RNA folding landscape suggest that the space of possible phenotypes is very rich compared to the phenotypic neighborhood of a single genotype – i.e. $K \ll P$. Furthermore, the phenotypic neighborhoods of neutral neighbors differ significantly. These measurements in RNA confirm the central



Supplementary Figure 11 – f_e for simulated RNA molecules of length L.

premise of our model: a neutral mutation can have profound, epistatic consequences for the phenotypes of subsequent mutations.

6.2 Robustness and Evolvability in Evolving RNA Populations

To more directly test our abstract model using the RNA genotype-phenotype landscape, we performed evolutionary simulations with populations of RNA sequences. In these simulations, genotypes were 24-base-pair sequences and phenotypes were the minimum-free-energy structures of these sequences. For a set of simulations, we first chose M structures to designate as 'high-fitness;' these correspond to the single, high-fitness type in our abstract model. Adjusting M allows us to tune the ease of adapting to high fitness, much as we set the ratio $\frac{K}{P}$ in our general model. For each replicate, we chose genotypes at random until one was found that folds into a non-trivial structure which is not a high-fitness type. The chosen phenotype represents the fit wild-type; all other structures, except the high-fitness types, are then considered to be inviable. The chosen genotype is then used to found a population of N clones, which then evolve according to a Wright-Fisher, discrete-generation model until the first high-fitness phenotype arises. The 'adaptation time' is defined as the

number of generations which pass before a high-fitness type arises. Note that each replicate begins from a potentially unique genotype, but the high-fitness phenotype remain consistent throughout a set of replicates.

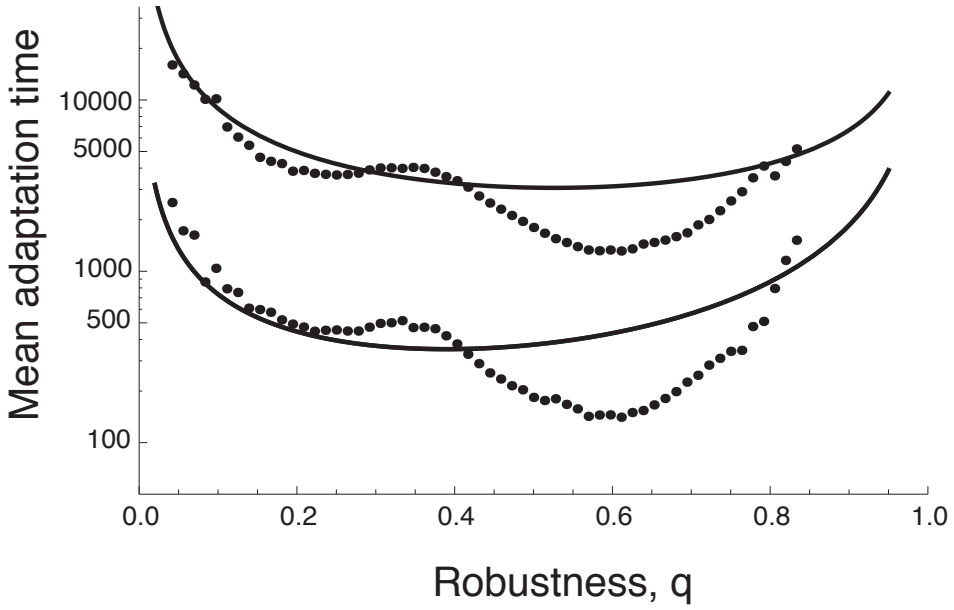
We measured the robustness, q, of each initial genotype as the frequency of point mutations that do not alter the phenotype, as we have done throughout our analysis. We used this initial q to predict adaptation times according to the analytical formulae derived above. For these simulations, the initial population was clonal, and so either all individuals were pre-adapted or none were pre-adapted. Thus, we obtained the expected adaptation time by suming expected adaptation times from each of these initial conditions, weighted by their probability:

$$\begin{aligned}\mathbb{P}\{\text{none pre-adapted}\}T^N(0) + \mathbb{P}\{\text{all pre-adapted}\}T^N(N) \\ \sim \mathbb{P}\{\text{none pre-adapted}\}T_1(0) + \mathbb{P}\{\text{all pre-adapted}\}T_2(1),\end{aligned}$$

for large N, where $T_1(0)$ is given by (1.10) and $T_2(1)$ is obtained from (1.5). To calculate predicted waiting times as a function of q, we measured three properties of these initial genotypes corresponding to the three parameters of our general model: the probability that a genotype is pre-adapted, the average mutation rate of pre-adapted genotypes to the high-fitness types, and the probability that a genotype which is not pre-adapted has a pre-adapted neighbor. These measurements correspond to $\frac{K}{P}$, $\frac{1}{K}$, and f in our abstract model. Although we estimated f above for RNA landscapes in general, we found that specific landscapes, determined by the particular set of M high-fitness types, have values of f which are more relevant to evolution of those landscape. Also, the mutation rate to high-fitness types declined monotonically with q in the RNA data, and so we used a linear regression of this relationship in our analytic predictions.

Supplementary Figure 12 shows the mean adaptation time as a function of robustness, q, for two sets of RNA simulations. In both cases, $N = 500$ and $\mu = 0.0002$; $M = 500$ in the lower curve, and $M = 200$ in the upper curve. Note that the empirical relationship between robustness and evolvability is somewhat more complex and differs from our analytical prediction; however, the primary trend matches the U-shaped prediction of our general, population-genetic model. Moroever, our general model accurately predicts the scale of the waiting times in the RNA simulations, over several orders of magnitude.

Our analysis of the RNA genotype-phenotype map suggests that robustness can facilitate adaptation under some circumstances. However, Ancel & Fontana (2000) found that evolving RNA populations became less able to adapt as their robustness increased. While several differences between their study and ours may contribute to this discrepancy, the most fundamental one is that Ancel and Fontana did not



Supplementary Figure 12 – The mean time before arrival of a beneficial mutation in RNA simulations. Each point is the mean of replicate simulations, and lines represent analytical predictions as described in the text. $N = 500$ and $\mu = 0.0002$; M, the number of high-fitness phenotypes, is 500 in the lower curve, and 200 in the upper.

manipulate robustness to test its effect on adaptation. In fact, populations typically became more robust as they evolved, such that robustness and distance from the optimum were confounded. By specifying a set of high-fitness types, rather than evolving population toward a distant optimum, our simulations reduce the effect of confounding variables. We also note that Cowperthwaite et al. (2008) found no relation between the size of the neutral network on which a population began, and its success at evolving to a distant phenotypic optimum; similarly, Wagner (2008) found no relationship between the robustness of a sequence, and a measure of its evolvability. The major difference between our results and those negative findings is that we designate a set of structures as high-fitness, as opposed to picking a single, optimal structure. As a result, the populations evolved much more rapidly in our simulations, compared with the slow, step-wise process of adaptation to a distant optimum. The relative speed of adaptation in our simulations suggests that the robustness of the initial genotype has a much greater influence on adaptation than in previous studies.

For all simulations we used version 1.6.1 of the Vienna package. Sequences were folded at 37°C with the option 'dangles' set to '2.' These settings produced the same landscapes as those described in (Cowperthwaite et al. 2008).

References

Abramowitz, M. & Stegun, I. A. (1965), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover Publications, New York.

Ancel, L. W. & Fontana, W. (2000), 'Plasticity, evolvability, and modularity in rna.', *J Exp Zool* **288**(3), 242–283.

Cowperthwaite, M. C., Economo, E. P., Harcombe, W. R., Miller, E. L. & Meyers, L. A. (2008), 'The ascent of the abundant: How mutational networks constrain evolution', *Plos Computational Biology* **4**(7).

Ethier, S. N. & Kurtz, T. G. (2005), *Markov Processes: Characterization and Convergence*, Wiley Interscience, Hoboken.

Ewens, W. J. (2004), *Mathematical population genetics*, Vol. v. 27, 2nd ed edn, Springer, New York.
URL: *http://www.loc.gov/catdir/enhancements/fy0818/2003065728-d.html*

Feller, W. (1954*a*), 'Diffusion processes in one dimension', *Trans. Amer. Math. Soc.* **77**(1), 1–31.

Feller, W. (1954*b*), 'The general diffusion operator and positivity preserving semigroups in one dimension', *Ann. of Math.* **60**(3), 417–436.

Forster, R., Adami, C. & Wilke, C. O. (2006), 'Selection for mutational robustness in finite populations', *Journal of Theoretical Biology* **243**(2), 181–190.

Joyce, P. & Tavaré, S. (1987), 'Cycles, permutations and the structure of the Yule process with immigration', *Stoch. Proc. Appl,* **25**, 309–314.

Karlin, S. & Taylor, H. M. (1981), *A Second Course in Stochastic Processes*, Academic Press, San Diego.

Slater, L. J. (1966), *Generalized hypergeometric functions*, University Press, Cambridge.

Sumedha, Martin, O. C. & Wagner, A. (2007), 'New structural variation in evolutionary searches of rna neutral networks', *Biosystems* **90**(2), 475–485.

van Nimwegen, E., Crutchfield, J. & Huynen, M. (1999), 'Neutral evolution of mutational robustness', *Proceedings of the National Academy of Sciences of the United States of America* **96**(17), 9716–9720.

Wagner, A. (2008), 'Robustness and evolvability: a paradox resolved.', *Proc Biol Sci* **275**(1630), 91–100.