

# Supporting Information

## Analysis of weak-mutation Markov chain

We proceed in somewhat more generality than in the main text. Suppose that evolution under mutation alone proceeds as a reversible, continuous-time Markov chain on a finite state space with rate matrix (infinitesimal generator)  $\mathbf{Q}_M$  and equilibrium distribution  $\boldsymbol{\pi}_M$ . If the scaled Malthusian fitness of genotype  $i$  is given by  $\mathbf{F}(i)$ , then evolution under weak mutation is a Markov chain with rate matrix  $\mathbf{Q}$  whose  $i, j$ -th entry is:

$$\mathbf{Q}(i, j) = \begin{cases} \frac{\mathbf{F}(j) - \mathbf{F}(i)}{1 - e^{-(\mathbf{F}(j) - \mathbf{F}(i))}} \mathbf{Q}_M(i, j) & \text{for } i \neq j \\ -\sum_{k \neq i} \mathbf{Q}(i, k) & \text{for } i = j. \end{cases} \quad (\text{S1})$$

It is easy to verify that the equilibrium distribution of this chain is given by the vector  $\boldsymbol{\pi}$ , where  $\boldsymbol{\pi}(i) \propto \boldsymbol{\pi}_M(i) e^{\mathbf{F}(i)}$ , and that this equilibrium satisfies detailed balance, so that the chain defined by  $\mathbf{Q}$  is also reversible. Note that the more limited definition of  $\mathbf{Q}_M$  in the main text based on some finite number of bi-allelic sites with non-zero forward and reverse mutation rates necessarily results in a reversible Markov chain, since it is simply the rate matrix for a collection of independent two-state chains with non-zero transition rates, and any two-state continuous-time chain with non-zero transition rates is reversible.

Because the Markov chain defined by  $\mathbf{Q}$  is reversible, the definition of detailed balance implies that the matrix  $\mathbf{D}_\pi^{1/2} \mathbf{Q} \mathbf{D}_\pi^{-1/2}$  is symmetric, where  $\mathbf{D}_x$  is the diagonal matrix whose diagonal entries are given by the vector  $\mathbf{x}$ . We can thus expand  $\mathbf{D}_\pi^{1/2} \mathbf{Q} \mathbf{D}_\pi^{-1/2}$  in terms of its eigenvalues and eigenvectors as

$$-\mathbf{D}_\pi^{1/2} \mathbf{Q} \mathbf{D}_\pi^{-1/2} = \sum_{k=1}^n \lambda_k \mathbf{u}_k \mathbf{u}_k^T, \quad (\text{S2})$$

where  $0 = \lambda_1 < \lambda_2 \leq \lambda_3 \leq \dots \leq \lambda_n$  are the eigenvalues of  $-\mathbf{D}_\pi^{1/2} \mathbf{Q} \mathbf{D}_\pi^{-1/2}$  and the eigenvectors  $\mathbf{u}_k$  form an orthonormal basis of  $\mathbb{R}^n$ . Multiplying the above equation by  $\mathbf{D}_\pi^{-1/2}$

from the left and  $\mathbf{D}_\pi^{1/2}$  from the right, then gives us:

$$-\mathbf{Q} = \sum_{k=1}^n \lambda_k \mathbf{r}_k \mathbf{l}_k^\top, \quad (\text{S3})$$

where  $\mathbf{l}_k = \mathbf{D}_\pi^{1/2} \mathbf{u}_k$  and  $\mathbf{r}_k = \mathbf{D}_\pi^{-1/2} \mathbf{u}_k$  are the left and right eigenvectors of  $-\mathbf{Q}$  associated with  $\lambda_k$ . Note that  $\mathbf{l}_k^\top \mathbf{D}_\pi^{-1} \mathbf{l}_m = \mathbf{r}_k^\top \mathbf{D}_\pi \mathbf{r}_m = 1$  for  $k = m$  and 0 otherwise, and  $\mathbf{l}_k(i) = \pi(i) \mathbf{r}_k(i)$ .

The transition probabilities for the Markov chain can then be written in terms of this expansion of  $\mathbf{Q}$ . In particular, let  $\mathbf{P}_t$  be the matrix whose  $i, j$ -th element is the probability that a population that begins at time 0 fixed for genotype  $i$  is fixed for genotype  $j$  at time  $t$ . Then we can write:

$$\mathbf{P}_t(i, j) = \sum_{k=1}^n e^{-\lambda_k t} \mathbf{r}_k(i) \mathbf{l}_k(j). \quad (\text{S4})$$

As a result, for any function on the state space of the Markov chain, the expected value of that function at time  $t$  for a population that begins fixed for genotype  $i$  at time 0 is given by

$$\sum_{k=1}^n e^{-\lambda_k t} \mathbf{r}_k(i) \mathbf{l}_k^\top \mathbf{g}. \quad (\text{S5})$$

where  $\mathbf{g}(i)$  is the value of the function at genotype  $i$ . Equation 4 follows by choosing  $\mathbf{g}(i) = \mathbf{F}(i)$  and noting that because the rows of  $\mathbf{Q}$  sum to zero, we must have  $\mathbf{r}_1 = \mathbf{1}$ , where  $\mathbf{1}$  is the vector of all 1s, for all  $i$  and thus  $\mathbf{l}_1 = \pi$ .

## Matching the fitness trajectory using symmetric mutation rates

Our main result can be extended to the case where we specify that the non-epistatic fitness landscape that produces the same fitness trajectory as an epistatic landscape must have equal forward and backward mutation rates at each site (i.e.  $\mu_l = \nu_l$  for all  $l$ ). This additional constraint of symmetric mutation rates limits the size of the  $c_k$  that can be accounted for by a

single non-epistatic site (in particular, the greatest value  $c_k$  that can be accounted for by a single site is .278, which occurs when  $S_l = -1.28$ ; negative values of  $c_k$  can be matched regardless of their magnitude). However, even with symmetric rates, one can still construct a non-epistatic landscape to exactly match the mean fitness trajectory of an arbitrary epistatic landscape by having multiple sites corresponding to a single term in Equation 4. In particular, one can generalize Equation 7 to

$$c_k = - \sum_{l \in \mathcal{L}_k} S_l \frac{\alpha_l}{\alpha_l + \beta_l} \quad (\text{S6})$$

where  $\mathcal{L}_k$  is the set of sites in the non-epistatic fitness landscape corresponding to the term  $c_k e^{-\lambda_k t}$  in Equation 4, and  $\alpha_l + \beta_l = \lambda_k$  for all  $l \in \mathcal{L}_k$ . This flexibility of using additional sites also means that one can alter the higher moments of the time-dependent fitness distribution of a non-epistatic fitness landscape while keeping the mean (i.e. mean fitness trajectory) unchanged.

## Bound on error when attempting to match fitness trajectory using a small number of sites

Next, we turn to deriving the bound on the error of the approximation of a mean fitness trajectory from an arbitrary epistatic fitness landscape using a fitness landscape with  $m$  sites. In particular, we will show that for any mean fitness trajectory  $f(t)$  produced by an arbitrary, finite fitness landscape whose mutational dynamics take the form of a reversible Markov chain, one can always construct an  $m$ -site non-epistatic fitness landscape and choice of starting genotype such that the resulting mean fitness trajectory  $f^*(t)$  satisfies

$$\sup_{t \geq 0} |f^*(t) - f(t)| \leq \frac{1}{m+1} \sqrt{\frac{\text{Var}_{\boldsymbol{\pi}} \mathbf{F}}{\boldsymbol{\pi}(i)}}}. \quad (\text{S7})$$

It is clear that this bound is far from being tight, as our main result shows that we can always exactly match the mean fitness trajectory if  $m \geq 2^L - 1$ . Indeed, in our experience it is often

possible to produce fits with a much smaller error than that given by the bound, particularly if we only consider a finite range of times.

The derivation has two parts. First we show that we can construct a landscape such that the error is at most  $\sum_{k=2}^n |c_k|/(m+1)$ , where the  $c_k$  are defined in Equation 4. The proof is based on a closely related argument from Kammler (1976, pg. 768), which the interested reader should also consult concerning the relation to completely monotone functions and the Laplace-Stieltjes transform. The second part of the proof then uses Hölder's inequality together with some linear algebra to bound  $\sum_{k=2}^n |c_k|$  in terms of the equilibrium frequency of the initial genotype and the variance in fitness at equilibrium for the original epistatic fitness landscape.

It is sufficient to construct approximations to mean fitness trajectories of the form

$$f(t) = \sum_{k=2}^n c_k e^{-\lambda_k t} \quad \text{with } \lambda_k > 0, \quad (\text{S8})$$

where we have assumed without loss of generality that  $c_1 = 0$  (we could match any value by appropriately choosing the initial fitness on the non-epistatic fitness landscape) and  $c_k \neq 0$  for  $k \geq 2$ . Now, by our main result, we can construct an  $m$ -site non-epistatic fitness landscape that produces any mean fitness trajectory of the form

$$f^*(t) = \sum_{i=1}^{m'} c_i^* e^{-\lambda_i^* t} \quad \text{with } \lambda_i^* > 0, \quad (\text{S9})$$

where we choose  $1 \leq m' \leq m$ . Furthermore, we can pick the  $\lambda_i^*$  and  $c_i^*$  such that, for any  $\lambda \geq 0$ , we have

$$\left| \left( \sum_{i:\lambda_i^* \leq \lambda} c_i^* \right) - \left( \sum_{k:\lambda_k \leq \lambda} c_k \right) \right| \leq \frac{1}{m+1} \sum_k |c_k|. \quad (\text{S10})$$

For instance, we can choose

$$\lambda_i^* = \sup \left\{ \lambda > \lambda_{i-1}^* : \sum_{k:\lambda_{i-1}^* < \lambda_k \leq \lambda} |c_k| \leq \frac{1}{m+1} \sum_k |c_k| \right\} \quad (\text{S11})$$

$$c_i^* = \sum_{k:\lambda_{i-1}^* < \lambda_k \leq \lambda_i^*} c_k, \quad (\text{S12})$$

where we interpret  $\lambda_0^*$  as 0 and define the  $\lambda_i^*$  iteratively for  $i = 1, 2, \dots$  until we either reach  $m$  or the sup no longer exists in which case we set  $m'$  equal to the last value of  $i$  for which the sup exists. To see why this solution works, note that for  $\lambda \in \{\lambda_1^*, \dots, \lambda_{m'}^*\}$

$$\left| \left( \sum_{i:\lambda_i^* \leq \lambda} c_i^* \right) - \left( \sum_{k:\lambda_k \leq \lambda} c_k \right) \right| = 0 \quad (\text{S13})$$

and that the sum  $\sum_{k:\lambda_k \leq \lambda} c_k$  viewed as a function of  $\lambda$  can change its value by at most  $\sum_k |c_k|/(m+1)$  in each of the intervals  $[0, \lambda_1^*), [\lambda_1^*, \lambda_2^*), \dots, [\lambda_{m'}^*, \infty)$ .

Having specified our approximating mean fitness trajectory  $f^*(t)$ , we can now bound its error relative to  $f(t)$ . Note that for any  $x, t > 0$ , we can write

$$e^{-xt} = t \int_x^\infty e^{-\lambda t} d\lambda = t \int_0^\infty \chi(\lambda - x) e^{-\lambda t} d\lambda, \text{ where } \chi(y) = 1 \text{ for } y \geq 0 \text{ and } 0 \text{ otherwise.}$$

Thus we have, for  $t > 0$ :

$$|f^*(t) - f(t)| = \left| \left( \sum_{i=1}^{m'} c_i^* e^{-\lambda_i^* t} \right) - \left( \sum_{k=1}^m c_k e^{-\lambda_k t} \right) \right| \quad (\text{S14})$$

$$= \left| \left( \sum_{i=1}^{m'} c_i^* t \int_0^\infty \chi(\lambda - \lambda_i^*) e^{-\lambda t} d\lambda \right) - \left( \sum_{k=1}^m c_k t \int_0^\infty \chi(\lambda - \lambda_k) e^{-\lambda t} d\lambda \right) \right| \quad (\text{S15})$$

$$= \left| t \int_0^\infty \left( \left( \sum_{i=1}^{m'} c_i^* \chi(\lambda - \lambda_i^*) \right) - \left( \sum_{k=1}^m c_k \chi(\lambda - \lambda_k) \right) \right) e^{-\lambda t} d\lambda \right| \quad (\text{S16})$$

$$= \left| t \int_0^\infty \left( \left( \sum_{i:\lambda_i^* \leq \lambda} c_i^* \right) - \left( \sum_{k:\lambda_k \leq \lambda} c_k \right) \right) e^{-\lambda t} d\lambda \right| \quad (\text{S17})$$

$$\leq t \int_0^\infty \left| \left( \sum_{i:\lambda_i^* \leq \lambda} c_i^* \right) - \left( \sum_{k:\lambda_k \leq \lambda} c_k \right) \right| e^{-\lambda t} d\lambda \quad (\text{S18})$$

$$\leq \left( \sup_\lambda \left| \left( \sum_{i:\lambda_i^* \leq \lambda} c_i^* \right) - \left( \sum_{k:\lambda_k \leq \lambda} c_k \right) \right| \right) t \int_0^\infty e^{-\lambda t} d\lambda \quad (\text{S19})$$

$$= \sup_{\lambda} \left| \left( \sum_{i: \lambda_i^* \leq \lambda} c_i^* \right) - \left( \sum_{k: \lambda_k \leq \lambda} c_k \right) \right| \quad (\text{S20})$$

$$\leq \frac{1}{m+1} \sum_k |c_k|. \quad (\text{S21})$$

This establishes the required inequality for  $t > 0$ ; the inequality must then also hold at  $t = 0$  by the continuity of  $f(t)$  and  $f^*(t)$ .

It remains to derive an upper bound on  $\sum_k |c_k| = \sum_{k=2}^n |\mathbf{r}_k(i) \mathbf{l}_k^T \mathbf{F}|$ . By Hölder's inequality we have

$$\sum_{k=2}^n |\mathbf{r}_k(i) \mathbf{l}_k^T \mathbf{F}| \leq \sqrt{\sum_{k=2}^n (\mathbf{r}_k(i))^2} \sqrt{\sum_{k=2}^n (\mathbf{l}_k^T \mathbf{F})^2}. \quad (\text{S22})$$

Now,  $\sum_{k=2}^n (\mathbf{r}_k(i))^2 \leq \sum_{k=1}^n (\mathbf{r}_k(i))^2$  and the latter sum is the squared Euclidean norm of the  $i$ -th row of the matrix  $\mathbf{D}_{\pi}^{-1/2} \mathbf{U}$ , where  $\mathbf{U}$  is the matrix with  $\mathbf{u}_k$  as its  $k$ -th column. Since  $\mathbf{U}$  is an orthogonal matrix, its rows are orthonormal and hence have a squared Euclidean norm equal to 1. Because the  $i$ -th row of  $\mathbf{U}$  is multiplied by  $1/\sqrt{\pi(i)}$  in the matrix product  $\mathbf{D}_{\pi}^{-1/2} \mathbf{U}$ , we have  $\sum_{k=1}^n (\mathbf{r}_k(i))^2 = 1/\pi(i)$ . Indeed, since  $\mathbf{r}_1(i) = 1$  for all  $i$ , we have

$$\sum_{k=2}^n (\mathbf{r}_k(i))^2 = \frac{1 - \pi(i)}{\pi(i)} \leq \frac{1}{\pi(i)}. \quad (\text{S23})$$

As for the other sum, since  $\mathbf{l}_1 = \boldsymbol{\pi}$ , we have

$$\sum_{k=2}^n (\mathbf{l}_k^T \mathbf{F})^2 = \left( \sum_{k=1}^n (\mathbf{l}_k^T \mathbf{F})^2 \right) - (\boldsymbol{\pi}^T \mathbf{F})^2 \quad (\text{S24})$$

$$= \left( \sum_{k=1}^n \left( (\mathbf{D}_{\pi}^{1/2} \mathbf{u}_k)^T \mathbf{F} \right)^2 \right) - (\boldsymbol{\pi}^T \mathbf{F})^2 \quad (\text{S25})$$

$$= \left( \sum_{k=1}^n (\mathbf{u}_k^T (\mathbf{D}_{\pi}^{1/2} \mathbf{F}))^2 \right) - (\boldsymbol{\pi}^T \mathbf{F})^2 \quad (\text{S26})$$

$$= \left( \sum_{k=1}^n \left( \sqrt{\pi(k)} \mathbf{F}(k) \right)^2 \right) - (\boldsymbol{\pi}^T \mathbf{F})^2 \quad (\text{S27})$$

$$= \boldsymbol{\pi}^T \mathbf{F}^2 - (\boldsymbol{\pi}^T \mathbf{F})^2 \quad (\text{S28})$$

$$= \text{Var}_{\boldsymbol{\pi}} \mathbf{F}, \quad (\text{S29})$$

where  $\mathbf{F}^2(i) = \mathbf{F}(i)^2$  and we have used the fact that  $\mathbf{U}$  is orthonormal and hence preserves the squared Euclidean norm of a vector. This completes the derivation of the bound.

## Dynamics at equilibrium

To study evolution at equilibrium, we again consider an ensemble of populations, but instead of assuming that all populations in the ensemble begin at some specified genotype, we let the initial genotype of each population be drawn from  $\boldsymbol{\pi}$ , the equilibrium distribution of the Markov chain defined by  $\mathbf{Q}$ . Using the definition of covariance, the covariance between the fitness of a population at time  $t' \geq 0$  whose genotype is drawn from  $\boldsymbol{\pi}$  at time 0 and its fitness at time  $t' + t$  is given by

$$a(t) = \sum_{i=1}^n \sum_{j=1}^n \boldsymbol{\pi}(i) \mathbf{P}_t(i, j) (\mathbf{F}(i) - \boldsymbol{\pi}^T \mathbf{F}) (\mathbf{F}(j) - \boldsymbol{\pi}^T \mathbf{F}). \quad (\text{S30})$$

Defining the centered fitness vector  $\mathbf{F}' = \mathbf{F} - (\boldsymbol{\pi}^T \mathbf{F}) \mathbf{1}$ , we can rewrite this in matrix notation as

$$a(t) = (\mathbf{F}')^T \mathbf{D}_{\boldsymbol{\pi}} \mathbf{P}_t \mathbf{F}'. \quad (\text{S31})$$

Using Equation S4, we can then expand  $\mathbf{P}_t$  in terms of its eigenvalues and eigenvectors and simplify to get

$$a(t) = \sum_{k=1}^n e^{-\lambda_k t} \left( (\mathbf{F}')^T \mathbf{D}_{\boldsymbol{\pi}} \mathbf{r}_k \right) (\mathbf{1}_k^T \mathbf{F}') \quad (\text{S32})$$

$$= \sum_{k=1}^n e^{-\lambda_k t} (\mathbf{1}_k^T \mathbf{F}')^2 \quad (\text{S33})$$

$$= \sum_{k=2}^n e^{-\lambda_k t} (\mathbf{1}_k^T \mathbf{F})^2 \quad (\text{S34})$$

where the last line follows because by construction  $\mathbf{l}_1^\top \mathbf{F}' = \boldsymbol{\pi}^\top \mathbf{F}' = 0$  and, for  $k \geq 2$ ,  $\mathbf{l}_k^\top \mathbf{1} = \mathbf{l}_k^\top \mathbf{r}_1 = 0$ , so that for  $k \geq 2$

$$\mathbf{l}_k^\top \mathbf{F}' = \mathbf{l}_k^\top \mathbf{F} - (\boldsymbol{\pi}^\top \mathbf{F}) (\mathbf{l}_k^\top \mathbf{1}) \tag{S35}$$

$$= \mathbf{l}_k^\top \mathbf{F}. \tag{S36}$$

Equation 15 in the main text then follows from Equation S34 by noting that  $(\mathbf{l}_k^\top \mathbf{F})^2$  is non-negative.