

Supplementary Online Material

In the following sections, we present our approach to calculating Lyapunov exponents.

We derive our central result

$$\Lambda = \frac{1}{\tau} \sum_{i=1}^n p_i \tau_i \lambda_1(A_i) - \frac{1}{\tau} \sum_{i,j=1}^n p_j b_{ij} T_{ij}^* \lambda_1(A_i) \quad (1)$$

directly from the growth equation

$$\dot{\mathbf{x}}(t) = A_{\mathcal{E}(t)} \mathbf{x}(t) \quad (2)$$

We then use perturbation theory to show how the lag times (expressed in terms of eigenvector projections) can be calculated to arbitrary precision in terms of the switching rates H_{ij} . We give a more general formulation than in the main text, allowing the number of environments (n) to be different from the number of phenotypes (m). We use notation introduced in Figure 1, main text.

1. Lyapunov Exponents for Structured Population Growth

In order for our general approximation to be valid, we require the top eigenvalue of the matrices A_i to be real, and for there to be a gap in the eigenvalue spectrum between the two eigenvalues of largest magnitude. Furthermore, we will restrict our discussion to cases when the top eigenvector is non-negative. In the case of responsive switching, these conditions hold, as seen in section 3. In the case of spontaneous stochastic switching, these conditions are guaranteed by the Perron-Frobenius theorem. We apply the theorem to the nonnegative matrix $A_i + \gamma I$, where I is the identity matrix and γ is a large positive constant. The eigenvectors of this matrix are identical to those of A_i . As

long as each phenotype can give rise to any other phenotype, via some sequence of switchings, the matrix $A_l + \gamma I$ is primitive, and the theorem may be applied. Primitivity holds trivially if all switching rates are strictly positive.

We subdivide time into consecutive intervals in which the environment variable $\mathcal{E}(t)$ does not change. The duration of the l -th such interval is denoted T_l , and the total time elapsed by the end of the L -th interval is denoted $t_L = \sum_{l=1}^L T_l$, where $t_0 = 0$. The state of the environment during interval l is denoted $\mathcal{E}(l)$.

We introduce the generalized eigenvectors \mathbf{v}_r^k which bring the matrix A_k to its Jordan-block form. That is, if we define the matrix $M_k = \begin{bmatrix} \mathbf{v}_1^k & \cdots & \mathbf{v}_m^k \end{bmatrix}$, then the matrix $M_k^{-1} A_k M_k$ is in Jordan-block diagonal form (I). The eigenvalue associated with the r -th eigenvector is denoted $\lambda_r(A_k)$, and these are decreasing with increasing r , and appear with multiplicity (the top eigenvalue is non-degenerate).

If durations T_l are sufficiently long (we will explain what this means shortly), the direction of the population vector at the end of the $(l-1)$ -th interval is very close to that of the top eigenvector of environment $\mathcal{E}(l-1)$. Thus $\mathbf{x}(t_{l-1}) \approx N(t_{l-1}) \mathbf{v}_1^{\mathcal{E}(l-1)}$, where we choose the top eigenvectors normalized so their entries sum to one: $\sum_{s=1}^m (\mathbf{v}_1^k)_s = 1$ for all k (this is always possible due to the positivity of the top eigenvectors). When the environment changes, we may then simply project this top eigenvector onto the new eigenbasis to describe the dynamics in the new environment. Projecting $\mathbf{x}(t_{l-1})$ onto the eigenbasis of environment $\mathcal{E}(l)$, the component along the new top eigenvector $\mathbf{v}_1^{\mathcal{E}(l)}$ is given by

$q_{\varepsilon(l)\varepsilon(l-1)}N(t_{l-1})$, where $q_{ij} \equiv \mathbf{e}_i \cdot M_i^{-1} M_j \mathbf{e}_i$ and $\mathbf{e}_i = (1, 0, \dots, 0)$. Note that q_{ij} is independent of the magnitudes of all eigenvectors other than the top eigenvectors, and thus our prescribed normalization of \mathbf{v}_1^k uniquely determines q_{ij} .

The time evolution of $N(t)$ for $t_{l-1} \leq t < t_l$ is then given by

$$N(t) = \left(q_{\varepsilon(l)\varepsilon(l-1)} e^{\lambda_1(A_{\varepsilon(l)})(t-t_{l-1})} + G(t-t_{l-1}) \right) N(t_{l-1}) \quad (3)$$

where $G(T)$ is a function that grows slower than $e^{\lambda_1(A_{\varepsilon(l)})T}$, and $G(0) = 1 - q_{\varepsilon(l)\varepsilon(l-1)}$. The exact form of $G(T)$ depends on the lower eigenvalues of $A_{\varepsilon(l)}$ and their multiplicities, and on the projection of $\mathbf{v}_1^{\varepsilon(l-1)}$ onto their corresponding eigenvectors. The Lyapunov exponent is then given by the following limit:

$$\begin{aligned} \Lambda &= \lim_{L \rightarrow \infty} \frac{\log N(t_L)}{t_L} = \lim_{L \rightarrow \infty} \frac{1}{t_L} \sum_{l=1}^L \log \left(q_{\varepsilon(l)\varepsilon(l-1)} e^{\lambda_1(A_{\varepsilon(l)})T_l} + G(T_l) \right) \\ &\approx \lim_{L \rightarrow \infty} \frac{1}{t_L} \sum_{l=1}^L \log \left(q_{\varepsilon(l)\varepsilon(l-1)} e^{\lambda_1(A_{\varepsilon(l)})T_l} \right) \\ &= \lim_{L \rightarrow \infty} \frac{1}{L\tau} \sum_{l=1}^L T_l \lambda_1(A_{\varepsilon(l)}) + \lim_{L \rightarrow \infty} \frac{1}{L\tau} \sum_{l=1}^L \log q_{\varepsilon(l)\varepsilon(l-1)} \end{aligned} \quad (4)$$

where $\tau = \lim_{L \rightarrow \infty} \frac{t_L}{L}$ (the average duration of environments).

The approximation in (4) is valid when $q_{\varepsilon(l)\varepsilon(l-1)} e^{\lambda_1(A_{\varepsilon(l)})T_l} \gg G(T_l)$. This can be achieved if T_l are sufficiently long. We have an explicit bound on $G(T)$ when the other eigenvalues are non-degenerate: $G(T) \leq m K_{\varepsilon(l)\varepsilon(l-1)} e^{\lambda_2(A_{\varepsilon(l)})T}$, where

$$K_{ij} = \max_{r \neq 1} \left| (\mathbf{e}_r \cdot M_i^{-1} \mathbf{v}_1^j) \sum_{s=1}^m (\mathbf{v}_r^i)_s \right| \text{ and } \mathbf{e}_r \text{ are the standard Euclidean basis vectors. Thus the}$$

following condition is sufficient:

$$T_l \gg \frac{1}{\lambda_1(A_{\varepsilon(l)}) - \lambda_2(A_{\varepsilon(l)})} \log \left(\frac{mK_{\varepsilon(l)\varepsilon(l-1)}}{q_{\varepsilon(l)\varepsilon(l-1)}} \right) \quad (5)$$

If we define $T_{\min} = \max_{i,j} \left\{ \frac{1}{\lambda_1(A_i) - \lambda_2(A_i)} \log \left(\frac{mK_{ij}}{q_{ij}} \right) \right\}$, then our approximation is valid for

environmental durations $T_l \gg T_{\min}$. The true Lyapunov exponent approaches the approximation exponentially fast in T_l , for $T_l > T_{\min}$, due to the exponential decay of the correction term neglected in (4), so the approximation is a very good one in this regime.

We can further simplify equation (4) if we assume that environmental changes follow the Markov chain b_{ij} . We denote the duration of the k -th occurrence of environment i using the random variable $T_k^{(i)}$ (this is just a regrouping of the random variables T_l). For fixed i , we assume that the variables $T_k^{(i)}$ are independent, identically-distributed variables, with mean τ_i . If L is the number of intervals elapsed, then for large L , the number of occurrences of environment i approaches $p_i L$, and of the environment pair j followed by i approaches $p_j b_{ij} L$, so the Lyapunov exponent is

$$\begin{aligned} \tau \Lambda &= \lim_{L \rightarrow \infty} \frac{1}{L} \sum_{i=1}^n \sum_{k=1}^{p_i L} T_k^{(i)} \lambda_1(A_i) + \frac{1}{L} \sum_{i,j=1}^n p_j b_{ij} L \log q_{ij} \\ &= \sum_{i=1}^n p_i \tau_i \lambda_1(A_i) + \sum_{i,j=1}^n p_j b_{ij} \log q_{ij} \end{aligned} \quad (6)$$

This is our central result allowing computation of Lyapunov exponents for structured population growth. It can be interpreted in terms of delay times as follows. The quantity q_{ij} is the projection of the population vector at the end of environment j onto the top eigenvector of the new environment i (the projection operation uses the new eigenbasis, via the matrix M_i^{-1}). If the population size is N , then $N q_{ij}$ is the size of the

subpopulation initially growing at the rate $\lambda_1(A_i)$ when the environment changes.

Defining the delay time T_{ij}^* as the amount of time it takes this subpopulation to reach a size of N , we find $T_{ij}^* = -(\log q_{ij}) / \lambda_1(A_i)$, and substituting this into the above equation, we obtain equation (1). Rewriting condition (5) using this definition, we find

$$T \gg \frac{1}{\lambda_1(A_i) - \lambda_2(A_i)} \left(\log(mK_{ij}) + T_{ij}^* \lambda_1(A_i) \right), \text{ where } T \text{ is the environmental duration.}$$

Computation of Lyapunov exponents is now reduced to computation of $\log q_{ij}$ and $\lambda_1(A_i)$, which is done for spontaneous and responsive switching in the next two sections. It is sometimes easier to directly compute the delay times T_{ij}^* , and this gives identical results, as shown in section 4.

2. Lyapunov Exponent for Stochastic Switching

Assuming that the stochastic switching rates are small compared to the growth rates, we can use perturbation theory to write the eigenvalues and eigenvectors of A_k to first order in $H_{ij}^{(k)}$, and use these expressions in equation (6) to compute the long-term growth rate.

The formulae for the eigenvectors, \mathbf{v}_r^k , and eigenvalues, $\lambda_r(A_k)$, of the matrix A_k are as follows (2): $\mathbf{v}_r^k = \mathbf{e}_r + \sum_{i \neq r} C_{ir}^{(k)} \mathbf{e}_i$, and $\lambda_r(A_k) = f_r^{(k)} - H_{rr}^{(k)}$, where $\Delta_{ij}^{(k)} \equiv f_j^{(k)} - f_i^{(k)}$,

$C_{ij}^{(k)} = H_{ij}^{(k)} / \Delta_{ij}^{(k)}$, for $i \neq j$, and $C_{ii}^{(k)} = 0$ for all i . We will use $C^{(k)}$ to refer to the matrix with entries $C_{ij}^{(k)}$.

Returning to the definition of q_{ij} , and using α_i to denote the fastest-growing phenotype in environment i , we can write $q_{ij} = \mathbf{e}_{\alpha_i} \cdot M_i^{-1} M_j \mathbf{e}_{\alpha_j}$. This expression gives the correct value for q_{ij} when the order of the eigenvectors appearing in the matrices M_i and M_j is arbitrary. We expand $M_i^{-1} M_j$ to first order in the switching rates:

$$M_i^{-1} M_j = (I + C^{(i)})^{-1} \cdot (I + C^{(j)}) \approx (I - C^{(i)}) \cdot (I + C^{(j)}) \approx I + C^{(j)} - C^{(i)}$$

where I is the identity matrix. It follows that $q_{ij} = \delta_{\alpha_i \alpha_j} + C_{\alpha_i \alpha_j}^{(j)} - C_{\alpha_i \alpha_j}^{(i)}$. If $\alpha_i = \alpha_j$, we have $\log(q_{ij}) = 0$ to first order, and if $\alpha_i \neq \alpha_j$, $\log(q_{ij}) \approx \log(C_{\alpha_i \alpha_j}^{(j)} - C_{\alpha_i \alpha_j}^{(i)})$. Using these expressions in equation (7) gives the Lyapunov exponent to first order:

$$\tau \Lambda_S = \sum_i p_i \tau_i (f_{\alpha_i}^{(i)} - H_{\alpha_i \alpha_i}^{(i)}) + \sum_{i,j: \alpha_i \neq \alpha_j} p_j b_{ij} \log \left(\frac{H_{\alpha_i \alpha_j}^{(j)}}{\Delta_{\alpha_i \alpha_j}^{(j)}} + \frac{H_{\alpha_i \alpha_j}^{(i)}}{\Delta_{\alpha_j \alpha_i}^{(i)}} \right)$$

In the case that $H_{ij}^{(k)} = H_{ij}$, and $\alpha_i = i$, we recover the expression given in the main text.

If the perturbation expansion were carried out to second order, terms linear in H would appear in the second sum above, with coefficients that are independent of τ_i . Since the linear terms in the first sum are proportional to τ_i , for large τ_i we are safe in ignoring any linear contribution coming from the second sum. Using higher order perturbation theory for the eigenvalues and eigenvectors, the Lyapunov exponent can be calculated as a series expansion in H_{ij} , if desired. Notice that the use of α_i above allows one to examine the general case in which the number of phenotypes and environments are not necessarily equal. We note that the case of n environments and m phenotypes, with $n > m$, can be mapped, by appropriate choice of parameters, to a problem of m

phenotypes in m environments, with the Lyapunov exponent taking the same general form as the expression given in the main text.

3. Lyapunov Exponent for Responsive Switching

In the case of responsive switching, the matrices A_k take the following form:

$$A_k = \begin{pmatrix} f_1^{(k)} - H_m & 0 & \cdots & \cdots & 0 \\ 0 & f_2^{(k)} - H_m & & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ H_m & & & f_{\alpha_k}^{(k)} & \cdots & H_m \\ 0 & \cdots & & f_{m-1}^{(k)} - H_m & 0 \\ 0 & \cdots & & 0 & f_m^{(k)} - H_m \end{pmatrix}$$

The matrix of eigenvectors of A_k and its inverse are as follows, where $r_j^{(k)} = -\frac{H_m}{H_m + \Delta_{j\alpha_k}^{(k)}}$:

$$M_k = \begin{pmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & & 0 & \cdots & 0 \\ \vdots & & \ddots & \vdots & & \vdots \\ r_1^{(k)} & \cdots & r_{\alpha_{k-1}}^{(k)} & 1 & r_{\alpha_{k+1}}^{(k)} & \cdots & r_m^{(k)} \\ & & & \ddots & & & \\ 0 & \cdots & 0 & 1 & 0 \\ 0 & \cdots & 0 & 0 & 1 \end{pmatrix} \quad M_k^{-1} = \begin{pmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & & 0 & \cdots & 0 \\ \vdots & & \ddots & \vdots & & \vdots \\ -r_1^{(k)} & \cdots & -r_{\alpha_{k-1}}^{(k)} & 1 & -r_{\alpha_{k+1}}^{(k)} & \cdots & -r_m^{(k)} \\ & & & \ddots & & & \\ 0 & \cdots & 0 & 1 & 0 \\ 0 & \cdots & 0 & 0 & 1 \end{pmatrix}$$

Since the leading eigenvector in environment k is \bar{e}_{α_k} , we find $q_{ij} = -r_{\alpha_j}^{(i)} = \frac{H_m}{H_m + \Delta_{\alpha_j\alpha_i}^{(i)}}$.

Using equation (6), the Lyapunov exponent is given by

$$\tau \Lambda_R = \sum_{i=1}^n p_i \tau_i f_{\alpha_i}^{(i)} - \sum_{i,j=1}^n p_j b_{ij} \log(1 + \Delta_{\alpha_j\alpha_i}^{(i)} / H_m)$$

Taking $\alpha_i = i$ and replacing $f_i^{(i)}$ by $f_i^{(i)} - c$, we obtain the expression given in the main text.

4. Method of Delay-Times for Computing Lyapunov Exponents

The results of the previous two sections can also be derived by calculating the delay times T_{ij}^* introduced in section 1, and using the relation $\log q_{ij} = -\lambda_1(A_i)T_{ij}^*$ in equation (6) to compute the Lyapunov exponent. As explained previously, the delay time T_{ij}^* is the amount of time it takes, upon a change of environment from j to i , for the subpopulation growing at the fastest rate in environment i to reach a population size equal to the total population size that was reached at the end of the environment j . These times can therefore be computed directly by solution of differential equations, as follows.

We will make repeated use of the following pair of equations

$$\begin{aligned}\dot{x}_j &= \gamma_j x_j \\ \dot{x}_i &= \gamma_i x_i + \delta x_j\end{aligned}$$

whose solution is $x_i(T) = \left(x_i(0) + \frac{\delta}{\gamma_i - \gamma_j} x_j(0)\right) e^{\gamma_i T} + \frac{\delta}{\gamma_j - \gamma_i} x_j(0) e^{\gamma_j T}$.

For responsive switching, assume the population has reached a size N at the end of environment j . The population is almost entirely composed of phenotype j . When the environment changes to i , phenotypes j switch to phenotype i at a rate H_m . T_{ij}^* is the time it takes for phenotype i to reach a size of N . In the above equations, this corresponds to

$\gamma_j = f_j^{(i)} - c - H_m$, $\gamma_i = f_i^{(i)} - c$, and $\delta = H_m$, and T_{ij}^* is found by solving $x_i(T_{ij}^*) = N$.

For large H_m , $\gamma_i \gg \gamma_j$, we can simply solve $\frac{H_m}{f_i^{(i)} - f_j^{(i)} + H_m} N e^{(f_i^{(i)} - c) T_{ij}^*} = N$, finding

$T_{ij}^* = (f_i^{(i)} - c)^{-1} \log(1 + \Delta_{ji}^R / H_m)$. Note that this expression holds for small H_m as well,

since small H_m implies T_{ij}^* is large, so we are again justified in ignoring term $e^{\gamma_j T}$. Since

the best phenotype in environment i does not switch to any other phenotype,

$$\lambda_1(A_i) = f_i^{(i)} - c.$$

For stochastic switching with small switching rates, the fastest phenotype in environment j has reached a population size $\approx N$, and there is a small amount of phenotype i due to switching at rate H_{ij} from phenotype j . This amount is found by taking $\gamma_j = f_j^{(j)} - H_{jj}$, $\gamma_i = f_i^{(j)} - H_{ii}$, and $\delta = H_{ij}$ in the above equations, and assuming that T is long enough and switching rates H are small, $x_i(T) = \frac{H_{ij}}{f_j^{(j)} - f_i^{(j)}} x_j(0) e^{f_j^{(j)} T}$. Since $x_j(0) e^{f_j^{(j)} T} \approx N$, we find the amount of phenotype i is equal to $H_{ij} N / (f_j^{(j)} - f_i^{(j)})$. When the environment switches to i , T_{ij}^* is approximately the time it takes phenotype i to reach size N . This is found by solving $x_i(T_{ij}^*) = N$ with $\gamma_j = f_j^{(i)} - H_{jj}$, $\gamma_i = f_i^{(i)} - H_{ii}$, $\delta = H_{ij}$, $x_j(0) \approx N$, and $x_i(0) \approx H_{ij} N / (f_j^{(j)} - f_i^{(j)})$, giving

$$T_{ij}^* = -\frac{1}{f_i^{(i)}} \log \left(\frac{H_{ij}}{f_j^{(j)} - f_i^{(j)}} + \frac{H_{ij}}{f_i^{(i)} - f_j^{(i)}} \right) = \frac{1}{f_i^{(i)}} \log \left(\Delta_{ij}^S / H_{ij} \right)$$

to lowest order in H_{ij} . Since H_{ij} is small, T_{ij}^* can be written as in the responsive case: $T_{ij}^* \approx \frac{1}{f_i^{(i)}} \log \left(1 + \Delta_{ij}^S / H_{ij} \right)$. The top eigenvalue to first order in switching rates is $\lambda_1(A_i) = f_i^{(i)} - H_{ii}$.

5. Lyapunov Exponents and Environmental Fluctuations

The Lyapunov exponent determines the long-term growth rate of a population characterized by growth/switching matrices A_k when presented with a changing environment given by the process $\mathcal{E}(t)$. Remarkably, equation (1) shows that, provided the environments remain constant for long enough periods, the long-term growth rate will

depend only on the mean durations of the environments, τ_i , and on the pairwise correlations between them, b_{ij} and will not depend on other characteristics of the environmental fluctuations.

To see this directly, we undertook the following numerical exercise. We considered two environments and two phenotypes, described by a pair of 2-by-2 matrices, A_1 and A_2 . In this case, b_{ij} is trivial, as environment 1 always follows environment 2, and vice versa. The only non-trivial environmental fluctuation is due to the process generating the durations of each environment. For simplicity, we kept the duration of environment 2 constant, that is, if $T_k^{(i)}$ is the duration of the k -th occurrence of environment i , we took $T_k^{(2)} = \tau_2$ for all k . The random variables $T_k^{(2)}$ thus have the delta function probability distribution, centered at the value τ_2 .

For the duration of environment 1 we examined three different distributions: uniform, exponential, and delta function. For each distribution, we generated many realizations of the environment, and calculated the Lyapunov exponent numerically. In Figure S1A, we plot the Lyapunov exponent as a function of τ_1 . The exponent depends strongly on the distribution that determines the environmental fluctuations. According to equations (1) and (5) we expect the exponent to become independent of the exact distribution if $T_k^{(1)}$ is always larger than some cutoff value. Taking this cutoff to be 10, and using the same three distributions, we find in Figure S1B that this is indeed the case: the Lyapunov exponents calculated using different distributions for $T_k^{(1)}$ are identical, and depend only on the mean duration, τ_1 .

This conclusion holds only if the environmental durations $T_k^{(i)}$ are all sufficiently large, more precisely, if $T_k^{(i)} \gg T_{\min}$ for all k and i . In this case, equation (1) also gives a way to calculate the exponent. The result of this exact calculation is given by the solid line in Figure S1B, showing that equation (1) is in excellent agreement with the numerical results.

6. Lyapunov Exponents for Finite Populations

So far we have allowed the population size to grow without bound. Suppose instead that a maximum population size, N , is imposed, for example by periodic resampling. The Lyapunov exponent, as defined by $\Lambda \equiv \lim_{t \rightarrow \infty} \frac{1}{t} \log N(t)$, would be zero, because the total population size asymptotically would not grow. If two different strains of organism were competing within a fixed population size, however, one would eventually go extinct. A more general definition of the long-term growth rate must therefore exist, which we now describe.

From the original population growth given in (2), we see that the total population size, $N(t)$, obeys the equation $\dot{N}(t) = \sum_i f_i(t) x_i(t) = \left(\sum_i f_i(t) n_i(t) \right) N(t)$, where $f_i(t)$ and $n_i(t) \equiv x_i(t) / N(t)$ are the growth rate and frequency of phenotype i at time t . Solution of this equation yields $N(t) = N(0) \exp \left(\int_0^t \sum_i f_i(t') n_i(t') dt' \right)$, suggesting the following definition of the long-term growth rate:

$$\Lambda \equiv \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t \sum_i f_i(t') n_i(t') dt' \quad (8)$$

This definition has the advantage of depending only on the frequencies of phenotypes, so it is meaningful even when population size is fixed. For unlimited growth, $\Lambda = \lim_{t \rightarrow \infty} \frac{1}{t} \log N(t)$, so it agrees with the previous definition. Calculation of Λ for finite-size populations using stochastic simulations was performed and was in excellent agreement with the calculated value given by equation (1), for population sizes $N \gg 1/q_{ij}$. For smaller population sizes, deviations from this value of Λ were observed, and could be accounted for by replacing T_{ij}^* by the appropriate expectation of the delay time (to be described elsewhere), provided $N > 1/q_{ij}$. When population size is so small that slower phenotypes are not sufficiently represented ($N < 1/q_{ij}$), the theory presented here may not hold.

7. Phenotypic Memory

Suppose the probability of a transition to environment i depends both on the current environment, j , and on the previous environment, k . We define composite indices $I = (i, j)$ and $J = (j, k)$, and write b_{IJ} to mean the probability of a transition to environment i , given that the pair of environments consisting of k followed by j has occurred. In a similar manner, we can consider the phenotypic history of an individual, *i.e.* the series of phenotypic transitions that occurred in its ancestral lineage going backwards in time. By phenotypic memory we mean the ability to remember a finite number of these transitions, including the current phenotype. We stress that such memory is long-term in the sense that the given individual remembers not only its immediate ancestor's phenotype (this

will usually be identical to its own phenotype, if switching rates are small), but also the phenotypic states of the last few phenotypic transitions in its ancestral lineage.

We use H_{IJ} to denote the rate of switching to phenotype i from phenotypes j that were previously k (in the ancestral lineage). The generalization to memory of m phenotypes, and m -point environmental correlations, is by composite indices of the form $I = (i_0, i_1, \dots, i_{m-1})$ and $J = (i_1, i_2, \dots, i_m)$. Here our convention is that the ability to remember only the current phenotype, in a fluctuating environment whose transition probabilities depend only on the current environment, corresponds to $m = 1$, *i.e.* the case considered in the main text.

With this notation, the expression for Λ_s given in the main text holds, for small switching rates, when all indices are replaced by composite indices. To see this, for example if $m = 2$, consider the derivation of equation (1) and suppose that an environmental transition $J = (i_1, i_2)$ to $I = (i_0, i_1)$ occurs. The leading phenotype at the end of environment J is of the type i_1 (more precisely, it is an i_1 that came from i_2 , or simply phenotype J). There is also a subpopulation of i_0 at the end of environment J , specifically i_0 that came from i_1 , or simply phenotype I . When the environment switches to i_0 , this phenotype I will be amplified until it dominates the population.

The optimal switching rates are again given by $H_{IJ}(\text{optimal}) = b_{IJ} / \tau_J$. As in the case of sensors, there is a maximal cost for which memory is beneficial. If we let $I_{env}(m)$ be the environmental entropy when m -point correlations are considered, and $\Lambda_s(m)$ be the corresponding Lyapunov exponent, then a basic theorem in information theory (3) states that $I_{env} = I_{env}(1) \geq I_{env}(2) \geq I_{env}(3) \dots$ etc. The difference in growth rate between organisms with m -point memory vs. 1-point memory is given by

$\tau(\Lambda_S(m) - \Lambda_S(1)) = I_{env}(1) - I_{env}(m)$, assuming that τ_i and $f_i^{(k)}$ are unchanged and depend only on the current environment. If the cost of such memory is c_m , memory is beneficial for $c_m < \frac{1}{\tau}(I_{env}(1) - I_{env}(m))$.

We stress that our analysis and results pertaining to phenotypic memory are justified only in the limit of very large populations, and small switching rates. For finite populations, a more delicate treatment is necessary, which we will not undertake here. To see why, consider a small population and a very long-lasting environment $J = (i_1, i_2)$. Eventually, the population will be dominated by phenotype i_1 , but because of the small population size and long environmental duration, we cannot neglect the fact that other phenotypes will eventually switch to phenotype i_1 . Thus, the population at the end of environment J may be dominated by a mixture of phenotypes (i_1, i_k) , for various values of k , rather than only the phenotype J .

8. Natural Selection and Phenotype Switching Mechanisms

Natural selection can maintain a phenotype switching mechanism, as follows. Suppose that in environment i a mutation arises (with frequency ω) abolishing switching. The new genotype grows at a rate $f_i^{(i)}$, while the switching genotype grows at a rate $f_i^{(i)} - H_{ii}$. The time to fixation is $\log(1/\omega)/H_{ii} = \tau_i \log(1/\omega)$, for optimal switching rates. Since ω is small ($\ll 1$), this time is longer than τ_i . If typical durations of the environment are close to τ_i , the mutation will not reach fixation. The same holds for non-optimal switching rates, provided that H_{ij} are small. The mechanism can be lost,

however, if a very long environmental duration occurs. The width of the distribution of environmental durations and the behavior of its tail thus play a role in determining whether a switching mechanism can be maintained. Certain strains of *Candida albicans*, for example, do not switch phenotypes, and may have lost the ability to do so.

References

1. S. Roman, *Advanced Linear Algebra*, Graduate Texts in Mathematics (Springer-Verlag, New York, 1992), vol. 135.
2. R. L. Liboff, *Introductory Quantum Mechanics* (Addison-Wesley, New York, ed. 2nd, 1992).
3. T. M. Cover, J. A. Thomas, *Elements of Information Theory* (John Wiley & Sons, Inc., New York, 1991).
4. A. Crisanti, G. Paladin, A. Vulpiani, *Products of Random Matrices* (Springer-Verlag, Berlin, 1993).

Figure Caption

Figure S1: Dependence of the Lyapunov exponent on environmental fluctuations.

Numerical computation used the matrices $A_1 = \begin{pmatrix} 2 & 0.05 \\ 10^{-6} & -0.05 \end{pmatrix}$ and $A_2 = \begin{pmatrix} -7 & 0.05 \\ 10^{-6} & -0.05 \end{pmatrix}$.

To calculate each point, we generated a realization of the changing environment in which each environment occurred L times, the duration of environment 2 was fixed at 20 hours, and the duration of the k -th occurrence of environment 1 was a random variable $T_k^{(1)}$. We

computed the matrix product $G = \prod_{k=1}^L \left(e^{20A_2} e^{T_k^{(1)}A_1} \right)$, and then estimated the Lyapunov

exponent using the formula $\tau \Lambda = \log(\text{Tr}(G)) / 2L$, taking $L = 100$ (see (4)). We averaged

this value over 100 separate runs. **A.** The Lyapunov exponent plotted for $T_k^{(1)}$ having an

exponential (large open squares), a uniform (medium open squares), or a delta (small

filled squares) distribution. The uniform distribution extended from 0 to $2\tau_1$, the rate of

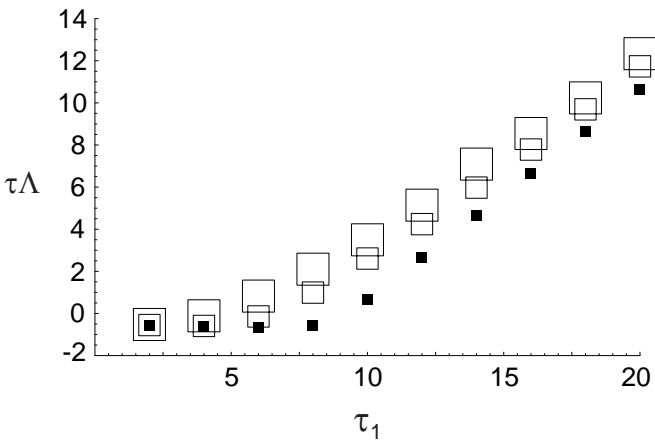
the exponential distribution was $1/\tau_1$, and the delta distribution was centered at τ_1 . **B.**

The Lyapunov exponent plotted for the three distribution from panel A, but each

distribution was shifted by 10 hours. This was done using $G = \prod_{k=1}^L \left(e^{20A_2} e^{T_k^{(1)}A_1} e^{10A_1} \right)$.

The solid line is the computation of $\tau \Lambda$ using equation (6).

Figure S1

A**B**