doi:10.1038/nature09905

MAINTENANCE OF THE FITTEST AND THE FLATTEST

R. E. BEARDMORE, I. GUDELJ, D. A. LIPSON AND L. D. HURST

Table of Contents

1. Notation

We use the following *Nature* convention: 'scalars', meaning real numbers, will be written using roman font as in the statement $u \in \mathbb{R}$, whereas 'vectors', meaning members of $\mathbb{R}^n$ for $n \geq 2$, will be written in bold, roman font as in $\boldsymbol{u} \in \mathbb{R}^2$. Matrices with be capitalised, so that M will represent a matrix that can act on a vector by multiplication, written as $M\boldsymbol{u}$, and I will denote the identity matrix that satisfies $I\boldsymbol{u} = \boldsymbol{u}$ for all $\boldsymbol{u}$. We shall write $0_{n\times n}$ highlighting the dimension, $n \times n$, of a matrix containing only zero entries. The notation $\mathrm{diag}(\boldsymbol{u})$ will be used to denote the square, diagonal matrix whose entries along the main diagonal are given by a vector $\boldsymbol{u}$.

If $F(\boldsymbol{u}, \boldsymbol{v})$ is a smooth function (meaning Fréchet differentiable) of two vector-valued quantities, $F : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^r$, we say that $d_{\boldsymbol{u}}F(\boldsymbol{u}, \boldsymbol{v})$ is the partial derivative of F with respect to the first variable. This derivative is a linear mapping from $\mathbb{R}^p$ to $\mathbb{R}^r$ that can be identified with an $r \times p$ matrix. We will write $d^2_{\boldsymbol{uu}}F(\boldsymbol{u}, \boldsymbol{v}), d^2_{\boldsymbol{uv}}F(\boldsymbol{u}, \boldsymbol{v})$ and $d^2_{\boldsymbol{vv}}F(\boldsymbol{u}, \boldsymbol{v})$ to represent second partial derivatives. Analogously, the notation $d_{(\boldsymbol{u},\boldsymbol{v})}G(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w})$ will denote the derivative with respect to the first two vector arguments of a function $G(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w})$.

The notation $\langle \boldsymbol{u}, \boldsymbol{v}\rangle$ represents the finite-dimensional, Euclidean inner product, $\langle \boldsymbol{u}, \boldsymbol{v}\rangle = \sum_{i=1}^n u_i v_i$, defined whenever $\boldsymbol{u} = (u_1, ..., u_n)$ and $\boldsymbol{v} = (v_1, ..., v_n)$ are two vectors of the same dimension; to reduce notational clutter where no confusion arises, column and row vectors will be used interchangeably. So, $\langle \boldsymbol{u}, \boldsymbol{u}\rangle = \sum_{i=1}^n u_i^2$ and so $\|\boldsymbol{u}\| = +\langle \boldsymbol{u}, \boldsymbol{u}\rangle^{1/2}$ is the Euclidean norm (or length) of $\boldsymbol{u}$. If $A = (a_{ij})_{i,j=1}^n$ is a matrix, we will write $\|A\| = +\left(\sum_{i,j=1}^n a_{ij}^2\right)^{1/2}$ for its Frobenius norm. When two vectors of the same dimension are juxtaposed and written $\boldsymbol{uv}$, we interpret this as a pointwise product: $\boldsymbol{uv} = (u_1v_1, u_2v_2, ..., u_nv_n)$. The uniform vector $(1, 1, ..., 1)$ will be written with bold font as $\mathbf{1}$ and so $\langle \mathbf{1}, \boldsymbol{u}\rangle = \sum_{i=1}^n u_i$.

If $\boldsymbol{u} \in \mathbb{R}^n$ is a vector and s is a scalar, by $\boldsymbol{u} - s$ we mean the vector $\boldsymbol{u} - s\mathbf{1} = (u_1 - s, u_2 - s, ..., u_n - s)$; this dimensionally incompatible notation for vector-scalar addition is consistent with the numerical package Matlab. By $\mathrm{span}\{\boldsymbol{u}\}$ we mean the one-dimensional linear space $\{\lambda \cdot \boldsymbol{u} : \lambda \in \mathbb{R}\}$ and then $\mathrm{span}\{\boldsymbol{u}\}^{\perp} = \{\boldsymbol{v} \in \mathbb{R}^n : \langle \boldsymbol{u}, \boldsymbol{v}\rangle = 0\}$. Following convention, the vector $\boldsymbol{u}$ is said to satisfy the inequality $\boldsymbol{u} > 0$ if both $u_i \geq 0$ for all $i = 1, ..., n$ and there is a j such that $u_j > 0$. We shall write $\boldsymbol{u} \geq 0$ if $u_i \geq 0$ for all i and then $\boldsymbol{u} \gg 0$ means that $u_i > 0$ for all such i.

A *diversity measure* is a functional $\mathcal{H} : \mathbb{R}^n \to \mathbb{R}$ that is scale-invariant, permutation-invariant, maximised at the uniform vector $\mathbf{1}$ and minimised at the competitive exclusion state $(1, 0, ..., 0)$, moreover the functional should be non-constant: $\mathcal{H}(1, 0, ..., 0) < \mathcal{H}(\mathbf{1})$. Hence, $\mathcal{H}(\lambda\boldsymbol{u}) = \mathcal{H}(\boldsymbol{u})$ for all scalars $\lambda > 0$ and all non-negative vectors $\boldsymbol{u} > 0$ and if P is a permutation matrix then $\mathcal{H}(P\boldsymbol{u}) = \mathcal{H}(\boldsymbol{u})$. Although not often written using this notation, the following version of Simpson's index is a diversity measure in this sense:

$$\mathcal{H}_S(\boldsymbol{u}) = -\left\| \frac{\boldsymbol{u}}{\langle \boldsymbol{u}, \mathbf{1}\rangle} - \frac{1}{n}\mathbf{1} \right\|^2.$$

Diversity measures are not unique, we could add any constant to $\mathcal{H}_S$ and it would still satisfy our requirements for being a diversity measure. Indeed, entropy may be used to measure diversity.

2. Introduction

The purpose of this supplement is to present three different theoretical models of microbial growth and evolution in a spatially unstructured environment, each of which can exhibit the phenomenon of *co-maintenance of the fittest and the flattest*. The structure shared by all the models is a *fitness landscape* that determines

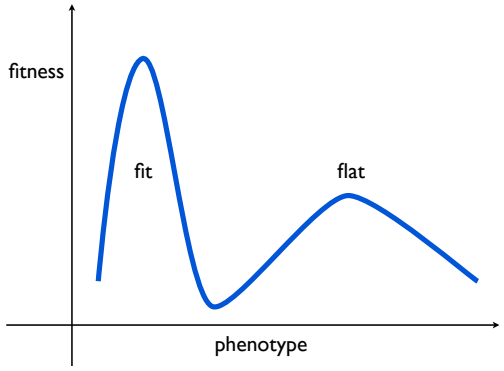
doi:10.1038/nature09905

which microbial cells have the highest *absolute fitness*, a synonym for their growth rates, and whereby the landscape possesses the following two features.

Firstly, the landscape has both a *fit* and a *flat* peak. While the term *fittest* simply refers to those microbial cells at the pinnacle of the fitness landscape, the *flattest* [46] refers to local fitness peaks with a lower absolute fitness than the fittest but which are more *mutationally robust* (see supplement fig. 1). Mutational robustness means that while a mutant offspring cell will suffer a reduction in fitness relative to its parent at the local peak, the smaller this reduction the more robust the peak is to mutation.

Secondly, the fitness landscape depends on the concentration of a single, limiting carbon source present in the environment. This carbon source is necessary for the cells to divide and, in doing so, they must consume some of the carbon and so deplete the quality of the environment. This, in turn, necessarily reduces the absolute fitnesses of all the cells.

supplement fig. 1 – Fit and flat peaks: assuming a correlation between genotype and phenotype, 'flat' describes a low fitness peak that is more robust to mutations than the fit peak.



We show that these two properties can be sufficient to maintain a population of cells at both fit and flat peaks together in equilibrium, provided mutation rates are neither too high nor too low. As the two peaks represent microbial cells that express different phenotypes within one population, this provides a framework to understand when multiple phenotypes may coexist in an unstructured envrionment. We will argue that the peaked structure found in the fitness landscape may be due to metabolic trade-offs whose existence has been postulated by others, between growth rate and growth yield for example. However, microbial cells can be the subject of other trade-offs that may conspire to produce such a coexistence effect, but a detailed discussion of this phenomenon is beyond the scope of this document and will form the basis of a series of future publications.

3. The Mutation-Selection Chemostat model

Our first mathematical model is the *mutation-selection chemostat* (MSC) equation, the following ordinary differential equation

$$\frac{d}{dt}\boldsymbol{f} = \epsilon(M-I)\boldsymbol{f} + \boldsymbol{G}(S)\boldsymbol{f} - \langle \boldsymbol{G}(S), \boldsymbol{f}\rangle \boldsymbol{f}, \tag{1a}$$

$$\frac{d}{dt}\Delta = \Delta(-d + \langle \boldsymbol{G}(S), \boldsymbol{f}\rangle), \tag{1b}$$

$$\frac{d}{dt}S = d(S_0 - S) - \Delta \langle \boldsymbol{U}(S), \boldsymbol{f}\rangle \tag{1c}$$

As regards the use of the term 'mutation' in this model, DNA-based mutations may apply, but *mutation* is better interpreted as a transition between different heritable states. For example, heritable epigenetic changes,

including transitions between alternative metabolic and regulatory configurations [31, 43] are possible. Thus, instead of ecotypes or genotypes, we will refer throughout the main manuscript to heritable 'types'.

The mathematical notation used in the MSC model are the following.

(A1). S, measured in micrograms per millilitre ($\mu g/ml$) unless otherwise stated, is the concentration of the limiting carbon source in the culture vessel of a chemostat at time t hours (h), Δ is the total density of cells per ml, S_0 ($\mu g/ml$) is the concentration of the carbon source held in a supply vessel, ϵ is the phenomenological mutation rate described in the main text of the manuscript and assumed to be the same for all types, d (per h) is the washout rate of the chemostat. (The volume of the culture vessel of the chemostat is not specified.)

(A2). There are n cell *types*, each type is labelled by a unique index i where $i = 1, ..., n$; the *mutational neighbours* of type i are types $i+1$ and $i-1$, analogous terminology is used at the boundaries 1 and n. (See supplement figs. 2(a) and 2(b) below for an illustration.)

(A3). The vector $\boldsymbol{f}(t) = (f_1(t), f_2(t), ..., f_n(t))$ contains the frequencies of each type, so that $\sum_{i=1}^{n} f_i(t) = 1$ holds for all times $t \geq 0$;

(A4). the i-th type has maximal uptake rate $V^{\max} \cdot x_i$, where $V^{\max}$ is a fixed constant measured in units of μg per cell per unit time and $x_i = i/n$ is a dimensionless value between 0 and 1; the latter will be called the *normalised maximal uptake rate* throughout.

(A5). The i-th type has a half-saturation constant for resource, $K(x_i)$, where $K(x)$ is a smooth function; $K(x)$ is measured in the same units as S, $\mu g/ml$.

(A6). The column vector $\boldsymbol{U}(S) = (U(x_1, S), U(x_2, S), ..., U(x_2, S))$ of uptake rates of all types is determined by the Monod function $U(x, S) = V^{\max} \frac{xS}{K(x)+S}$.

(A7). The yield of biomass per unit of resource of type i is c_i (in units of cells per μg) and we assume that c_i can be derived from the maximal uptake rate of type i. As a result, we shall write $c_i = c(x_i)$ where $c(\cdot)$ is a smooth function. The entries in the vector $\boldsymbol{G}(S) = (G(x_1, S), G(x_2, S), ..., G(x_2, S))$, where

$$G(x, S) = c(x) \cdot U(x, S)$$

define the per hour growth rates of each cell type at resource concentration S.

Throughout, cell densities computed from the mutation-selection chemostat equation (1) will be quoted in units of $\log_{10}$(cells) per ml unless otherwise stated (and 'log' will denote the base-10 logarithm). By the term *quasispecies* we mean a *cluster* of types whose densities form a unimodal distribution where a *cluster* is defined by each type being a mutational neighbour of at least one other type in that cluster.

When we speak of the *fitness landscape* in the remainder, we refer to the two-dimensional surface defined by the graph of the function $G(x, S)$ for $0 \leq x \leq 1$ and $S \geq 0$, note that this depends both on cell type through the normalised maximal uptake rate, x, but also on the resource concentration, S. There are natural constraints on this function, denoted G for 'growth'. For example, the availability of more resource should lead to a higher fitness of each cell type and therefore

$$G(x, S_1) > G(x, S_2) \qquad \text{whenever} \qquad S_1 > S_2$$

should hold. In the complete absence of the limiting carbon source, $S = 0$, it is assumed that no cell can grow and so G should satisfy

$$G(x, 0) = 0 \qquad \text{for all} \qquad x \in [0, 1].$$

These two properties will be assumed to hold in this article and, indeed, because of assumptions (A6) and (A7) above we will use the following specific form of fitness landscape

$$G(x,S) = c(x)\cdot V^{\max}\frac{xS}{K(x)+S}; \tag{2}$$

we shall also impose the mild restriction that G is a smooth function of all its arguments. Throughout K or $K(x)$ will be called a *half-saturation constant* and $1/K$ or $1/K(x)$ will be called the *affinity* for resource and $K(x)$ will never be zero. Similarly, $V^{\max}$ will always denote the maximum uptake rate of a carbon source into a cell.

The definition in (2) can be used to encode two trade-offs within the MSC model. (i) If $K(x)$ is an increasing function of x, then the half-saturation constant of each cell type increases with increasing maximal uptake rate, this is said to be a *rate-affinity trade-off* (see [27]). (ii) If $c(x)$ is a decreasing function of x, then cell yield decreases with increasing maximal uptake rate, this is said to be a *rate-yield trade-off* (see [32]). In this framework, the fittest cells are those with the highest possible uptake rates of resource but they do pay a cost for this in terms of having reduced yield, the cells with higher yields therefore have lower resource uptake rates.

If we define the vector of type densities, $\boldsymbol{b}(t) = \boldsymbol{f}(t)\cdot\Delta(t) = (b_1(t), b_2(t), ..., b_n(t))$, then the pair $(\boldsymbol{b}(t), S(t))$, with $\boldsymbol{b}(t)$ measured in units of cells per ml, satisfies the differential equation

$$\frac{d}{dt}\boldsymbol{b} = \epsilon(M-I)\boldsymbol{b} + (\boldsymbol{G}(S)-d)\boldsymbol{b}, \tag{3a}$$

$$\frac{d}{dt}S = d(S_0 - S) - \langle \boldsymbol{U}(S), \boldsymbol{b}\rangle \tag{3b}$$

and it is this form of the MSC equtation on which we concentrate throughout the remainder. Of particular interest here is the question of how the diversity of solutions of (1) and (3) depends on system parameters, in particular the mutation rate ϵ, with respect to some diversity measure $\mathcal{H}$.

The mathematical theory of the chemostat described, for example, in [39, 8] does not apply directly to (3) and although one can prove results on the global convergence to equilibrium for (3), one cannot easily exclude the possibility that it supports periodic solutions and chaotic attractors. Such solutions are already known to occur in simpler mutation-selection models than (1), for examples see [14]. As a result, estimates of long-term diversity supported by equation (3) will be obtained by studying the steady-state of this equation and by *a posteriori* numerical verification that the steady-states found in this way are locally and asymptotically stable. A straightforward eigenvalue computation is sufficient to establish this in practise and these are performed routinely in the computations that follow; while the algorithms developed in Matlab to rapidly solve (3) are able to report on any absence of local stability, no violations of such conditions were encountered[1]. We could go further and use techniques from rigorous computation to establish these statements rigorously, but this lies outside the scope of the present manuscript.

3.1. **The MSC model generates a dissipative dynamical system.** Under the restrictions imposed on equation (3), there is an efficiency constant $C > 0$ (independent of S, d, S_0 and ϵ) such that the inequality

$$\boldsymbol{G}(S) \le C\cdot\boldsymbol{U}(S)$$

[1]In addition, changes in local stability induce visible changes in the geometry of equilibrium solution branches, namely bifurcations, and it is shown below in Proposition 3 that no bifurcations are possible in non-trivial solution branches under certain restrictions, a theoretical result entirely consistent with our numerical computations. Eigenvalues located on the imaginary axis is a cause of oscillatory solutions and these were also tested for, but none were found.

holds in a pointwise sense, a natural condition stating that the growth rate of any type is no larger than some constant multiplied by uptake rate, in any environment. Define $\Sigma = S_0 - S - \frac{1}{C}\langle \mathbf{1}, \boldsymbol{b}\rangle$ and let us compute the time derivative of Σ:

$$\begin{aligned}\frac{d}{dt}\Sigma &= -\frac{d}{dt}S - \frac{1}{C}\left\langle \mathbf{1}, \frac{d}{dt}\boldsymbol{b}\right\rangle = -d(S_0 - S) + \langle \boldsymbol{U}(S), \boldsymbol{b}\rangle - \frac{1}{C}\langle \mathbf{1}, \epsilon(M-I)\boldsymbol{b} + (\boldsymbol{G}(S) - d)\boldsymbol{b}\rangle \\ &= -d\left(S_0 - S - \frac{1}{C}\langle \mathbf{1}, \boldsymbol{b}\rangle\right) + \left\langle \boldsymbol{U}(S) - \frac{1}{C}\boldsymbol{G}(S), \boldsymbol{b}\right\rangle \;=\; -d\Sigma \;+\; \text{positive} \\ &\geq -d\Sigma.\end{aligned}$$

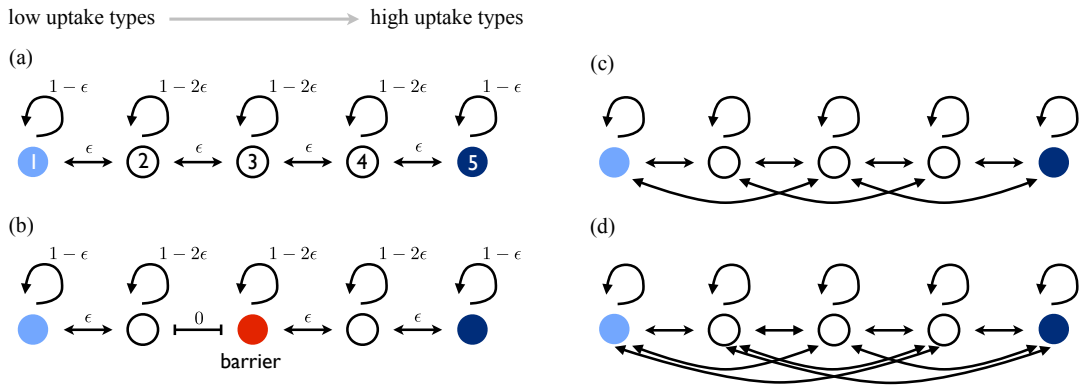
It follows from this inequality that for any positive number η, there is a time T such that $\Sigma(t) \geq -\eta$ for all $t > T$ and so we obtain the following dissipative bound on the long-term behaviour of solutions of (3):

$$S(t) + \frac{1}{C}\langle \mathbf{1}, \boldsymbol{b}(t)\rangle \leq S_0 + \eta \tag{4}$$

for all sufficiently large t. As a corollary of this argument, the inequality (4) holds for steady-state solutions of (3) with $\eta = 0$.

3.2. **The mutation matrix, M.** We will assume throughout that the matrix $M = (m_{ij})_{i,j=1}^{n}$ is a stochastic matrix (meaning $\mathbf{1}^T M = \mathbf{1}^T$) with zero elements on the diagonal and non-negative elements on the off-diagonal. The matrix M is constructed from the probabilities that a cell of type j exhibits a change of type, to i, due to a mutational event. Unless explicitly stated otherwise, we will assume that M is *irreducible*, meaning that for each pair (i, j) there is a power $p = p(i, j)$ such that the (i, j)-th entry of M^p is non-zero. This can be interpreted as a connectivity assumption on the set of types as it follows from this assumption that there is a finite sequence of mutations that provides a path between any two types.

supplement fig. 2 – Differences between the parent and offspring maximal uptake rate (and other cell phenotypes) arise when a mutation occurs, the parameter ϵ indicated on each transition represents mutation rates, the indicated values $1-\epsilon$ and $1-2\epsilon$ are derived from the probability that a change in maximal uptake rate does not occur. In (a) there is a small change in uptake rate with each mutation, (b) is the same as (a) except that a *mutational barrier* is introduced; in (a) and (b) the *mutational neighbours* of type i are types $i-1$ and $i+1$. If larger changes in cell uptake rate occur with each mutation, this is represented by diagrams in (c) and (d) and, note, the latter has a higher *mutational connectivity* than the former. In all cases, the cell types coloured blue represent 'type boundaries' at the extreme low and high limits of possible resource uptake rates.



Remark 1. *The so-called* connectivity of mutations *is determined from the pattern of non-zero entries in the matrix $M = (m_{ij})$, type i is said to be connected to type j by a single mutation if m_{ij} is non-zero. For example, the matrix denoted M above and associated with the mutational structures illustrated in supplement fig. 2(a) and (b), respectively, are the following tridiagonal matrices:*

$$M_a = \begin{bmatrix} 0 & \frac{1}{2} & 0 & 0 & 0 \\ 1 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 1 \\ 0 & 0 & 0 & \frac{1}{2} & 0 \end{bmatrix} \quad \textit{and} \quad M_b = \left[\begin{array}{cc|ccc} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & \frac{1}{2} & 0 \end{array}\right];$$

note that the column sums of both matrices are unity. The former matrix is irreducible and this *is the structure of the mutation matrix used throughout this paper unless specified otherwise, suitably adapted to the appropriate number of types, whereas the latter matrix, M_b, is not reducible by construction. To see this, note that the red block of zeros in M_b arises because no mutation can occur between types '2' and '3' in supplement fig. 2(b). For comparison, the mutational structures illustrated in supplement fig. 2(c) and (d) do not correspond to tridiagonal matrices.*

If the probability of mutation is the same for all cell types, ϵ, one can associate a Markov process $\mathcal{M}$ with the transition rates between different cell types by defining

$$\mathcal{M} = \overbrace{(1-\epsilon)I}^{no\ mutation} + \overbrace{\epsilon M}^{mutation} = I + \epsilon(M - I).$$

If ϵ is a small quantity, this expresses the idea that most offspring cells express the same type as their parent, as depicted in supplement figs. 2(a) and 2(b), but if a mutation does occur then the matrix M contains the probabilities of a transition occurring to each of the other types. Thus M must have zeros on its main diagonal, as do M_a and M_b in the above examples.

One consequence of the irreducibility and non-negativity of M is the existence of a unique invariant density $\boldsymbol{\nu} = (\nu_1, ..., \nu_n)^T$ associated with this mutation matrix, meaning that $\boldsymbol{\nu}$ satisfies

$$M\boldsymbol{\nu} = \boldsymbol{\nu} \qquad \text{and} \qquad \sum_{i=1}^{n} \nu_i = 1, \tag{5}$$

where $\nu_j > 0$ for all j.

In the case of a neutral fitness landscape whereby $\boldsymbol{G}(S) = g(S)\mathbf{1}$ for some scalar function $g(\cdot)$, but also in the limit whereby the mutation rate ϵ is 'infinite', after re-scaling time if necessary and noting that $\boldsymbol{f}$ has entries that sum to unity, equation (1a) reduces to the following linear diffusion equation

$$\frac{d}{dt}\boldsymbol{f} = (M - I)\boldsymbol{f}$$

whose solution operator $e^{(M-I)t}$ is mass-preserving. So, if $\boldsymbol{f}(0)$ has entries that sum to unity, it follows that $\boldsymbol{f}(t) = e^{(M-I)t}\boldsymbol{f}(0)$ also has entries that sum to unity for all $t \geq 0$ and so it converges to $\boldsymbol{\nu}$ as $t \to \infty$. As a result, it is trivially possible to chose mutation matrices, M, in such a way that (1) has (in an approximate sense) *any* unit vector $\boldsymbol{f}$ we desire as its equilibrium in a neutral landscape or when ϵ is large enough.

In order to avoid trivialities like this, whereby any equilibrium of (1) can be constructed, including very diverse states, but only by virtue of a structure imposed within M, in the remainder we shall restrict attention to those matrices M for which $\boldsymbol{\nu}$ is equal to the uniform distribution $\frac{1}{n}\mathbf{1}$. We shall also impose a non-neutrality

condition on our fitness landscape: mathematically speaking, for each pair of distinct indexes, i and j, by this we mean that $G_i(S) \neq G_j(S)$ for infinitely many S.

We re-iterate an assumed property of all the numerical computations and simulations conducted in this supplementary article: M will denote a process whereby a single mutational event forces a change of type of one unit in type space. As a result, M will be a tridiagonal matrix and it will have $\frac{1}{n}\mathbf{1}$ as its invariant density. We will utilise a working assumption that M is *reversible* a.k.a *symmetric*, meaning $M = M^T$ (equivalently $m_{ij} = m_{ji}$) to help simplify the theoretical presentation that follows. This assumption can be substantially weakened in a number of directions that we discuss later. Do note that when one assumes M to be an irreducible, non-negative, stochastic matrix, symmetric matrix, because $M^T\mathbf{1} = \mathbf{1}$ and $M = M^T$ it necessarily follows that $\boldsymbol{\nu}$ must be equal to the uniform distribution, $\boldsymbol{\nu} = \frac{1}{n}\mathbf{1}$.

3.3. **Comment: mutations are measured *per unit time*.** The structure of the MSC equations in (1) ensures that this model does not represent mutations originating from DNA replication errors alone as it predicts changes in cell type even when cell division does not occur, namely when setting $\boldsymbol{G}(S) \equiv 0$. The resulting linear differential equation $\frac{d}{dt}\boldsymbol{b} = \epsilon(M - I)\boldsymbol{b} - d\boldsymbol{b}$ is a diffusion equation that can generate novel types[2]. However, the phenotype of a cell, broadly defined, may indeed be dynamic in time and due, for example, to regulatory changes or to mutations induced by UV radiation. Stresses on the cell, the introduction of mutagens or the suppression, for example, of the SOS cycle may mean that biologically-realistic mutation rates are themselves dynamic quantities and subject to selection, a property not contained within the MSC model. Consistent with prior population genetics models (see for example [14]) the mutation rate ϵ in (1) is measured per unit time (in fact, per individual cell per unit time) and it is a fixed, unchanging quantity for all cell types.

As pointed out in [44, p. 1018] in the context of Ising models of sequence evolution, measuring mutation rates either per cell division or per generation requires a different model structure than if they were measured per unit time. As a result, we propose the following alternative equation to (3) that we do not analyse here: let $(\boldsymbol{b}(t), S(t))$ satisfy the differential equation

$$\frac{d}{dt}\boldsymbol{b} = (I + \tilde{\epsilon}(M - I))(\boldsymbol{G}(S)\boldsymbol{b}) - d\boldsymbol{b}, \tag{6a}$$

$$\frac{d}{dt}S = d(S_0 - S) - \langle \boldsymbol{U}(S), \boldsymbol{b}\rangle\,, \tag{6b}$$

where I is the identity matrix. If we were to set $\boldsymbol{G}(S) \equiv 0$ in equation (6a), we would obtain the exponential decay equation $\frac{d}{dt}\boldsymbol{b} = -d\boldsymbol{b}$ which contains no diffusion terms and so generates no novel cell types, furthermore setting the mutation rate $\tilde{\epsilon}$ equal to zero in (6) recovers the competitive chemostat equations, as it must.

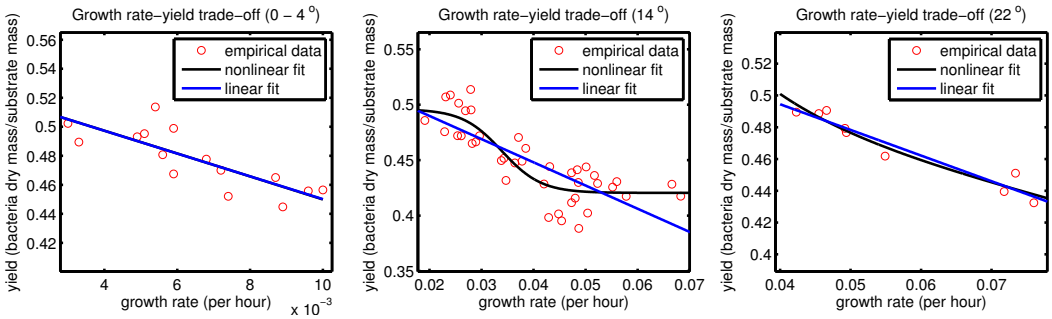
The matrix $I + \tilde{\epsilon}(M - I)$ in (6) can be written $(1 - \tilde{\epsilon})I + \tilde{\epsilon}M$ which expresses the idea that the dimensionless quantity $\tilde{\epsilon}$ is the probability of mutation and then M, whose columns sum to unity, identifies the change in cell type that subsequently occurs from that mutation. Like (3), equation (6) generates a dissipative dynamical system, an observation that may be exploited to produce conditions under which both classes of model have globally attractive, non-trivial steady-state solutions.

[2]If the *support* of a vector $\boldsymbol{b} = (b_1, ..., b_n)$ is the set of all indexes $j \in \{1, 2, ..., n\}$ such that $b_j \neq 0$, the differential equation $\frac{d}{dt}\boldsymbol{b} = F(t, \boldsymbol{b})$ is said to generate *no novel types* if the support of $\boldsymbol{b}(t)$ is equal to that of $\boldsymbol{b}(0)$ for all $t \geq 0$. If there is a t such that the support of $\boldsymbol{b}(t)$ and $\boldsymbol{b}(0)$ are different, then the equation is said to generate novel types.

4. Trade-offs: theory and evidence

4.1. **Rate-yield trade-off: definitions.** A substrate, such as glucose, is converted to ATP at a certain rate per unit time, this in turn is converted to biomass and hence to cellular growth. In [2], the authors observe that if microbes are limited by their energetic resource, the amount of biomass formed per unit of ATP is approximately constant and does not depend on the mode of ATP production. Therefore, as highlighted in [33], if the rate of ATP production increases, the rate of biomass formation and thus the growth rate of an organism also increases. Thus the 'rate' component of a *rate-yield trade-off* can be variously measured as moles of substrate per unit time, moles of ATP per unit time, biomass production per unit time or growth rate, computed from cell divisions, per unit time. Yield, by contrast, considers production per unit of substrate. This again can be moles of ATP per mole of substrate, biomass per mole of substrate or number of cell divisions per mole of substrate.

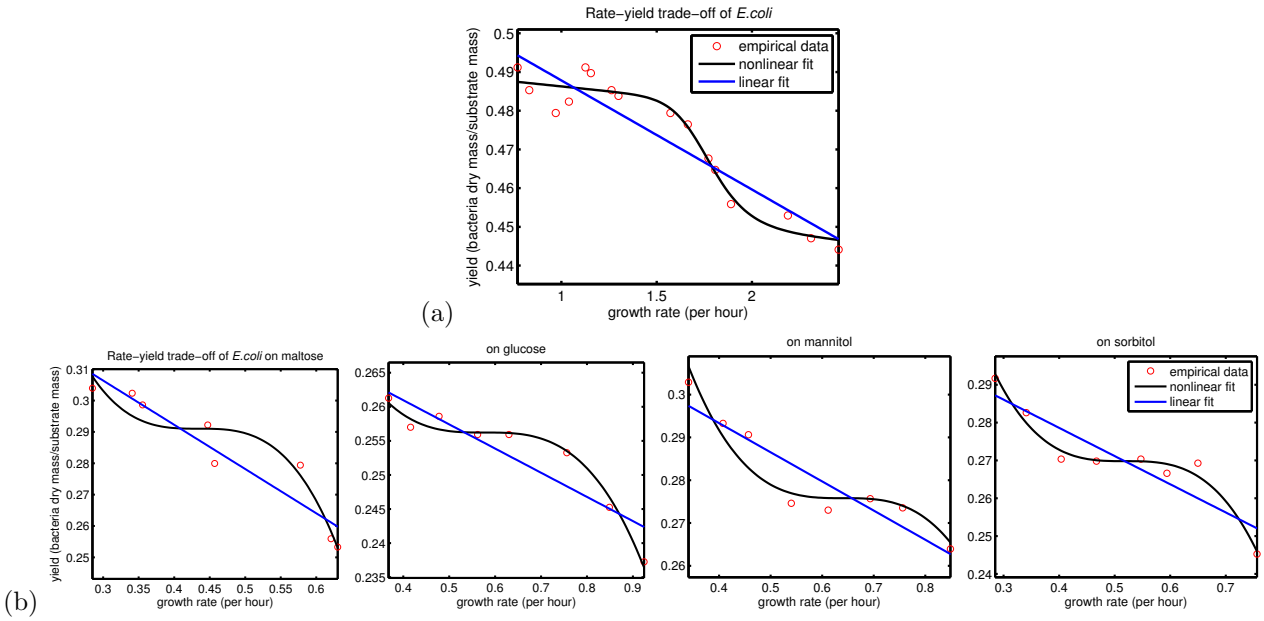
supplement fig. 3 – Linear and sigmoidal rate-yield trade-offs for soil microbes at three different temperatures: in all three cases we fit both a linear function and a nonlinear function whereby the yield equals $a - bR + c/(1 + \exp(d + eR))$, R denotes growth rate. In the middle plot the 95% confidence interval for parameter c is $(0.043, 0.108)$ which does not contain zero, thus indicating a better fit for the nonlinear form (data taken from [18]).



4.2. **Rate-yield trade-off: evidence.** Perhaps the best description of this trade-off and of its form comes from yeast, see [45], supplement figure 5(a) and [19, Figure 1] where the latter provides a summary of rate-yield trade-off data for various yeast species using data from [23, 35]. Evidence for the rate-yield trade-off is provided by Novak *et al.* who examined *E. coli* and showed that when comparing individuals within a population, such a trade-off was found [30]. Other studies indicate that maximal yield and maximal rate are incompatible. *Holophaga foetida*, for example, can double its maximum specific growth rate at the cost of a halved growth yield [17]. Likewise, *Acetobacter methanolicus* (see [28]) and *Saccharomyces kluyveri* (see [25]) can increase growth rates by shifting catabolic flow into metabolic branches with lower biomass yield. *Pseudomonas fluorescens* cultured in the chemostat has also been claimed to exhibit a rate-yield trade-off [12, supplementary text].

While the form of the trade-off is well-described in yeast, the literature also describes this trade-off within a single prokaryote species and within a community of microbes with sufficient detail to determine one or more related nonlinear forms. As regards microbial communities, Lipson *et al.* [18] examine rate-yield relationships for soil microbes under three temperature regimes and found a negative correlation between rate and yield in each. For two cases a linear fit is best, however, at 14 degrees a function with a sigmoidal structure containing both concave and convex regions provides a better fit to their data (see supplement fig. 3).

supplement fig. 4 – (a) An empirically-determined rate-yield trade-off in *E.coli* derived from data presented in [41]. A nonlinear Hill function is fitted to the data, illustrating a significant sigmoidal form with concave and convex regions. (b) Empirical rate-yield data from [26] of *E.coli* cultured on four different carbon sources also exhibits a significantly sigmoidal form. (If r denotes the rate data and y the yield data, both scaled linearly by their minimum and maximum values to lie between 0 and 1, the datafit in (a) uses the Hill function '$y = a - br - cr^e/(d^e + r^e)$' whereas (b) uses the monotonic cubic model '$y = a - br - c(r^2/2 + cr^3/(12b))$' and the linear fit is shown for comparison. Significantly, in all cases, the estimated 95% confidence interval for c does not contain zero.)

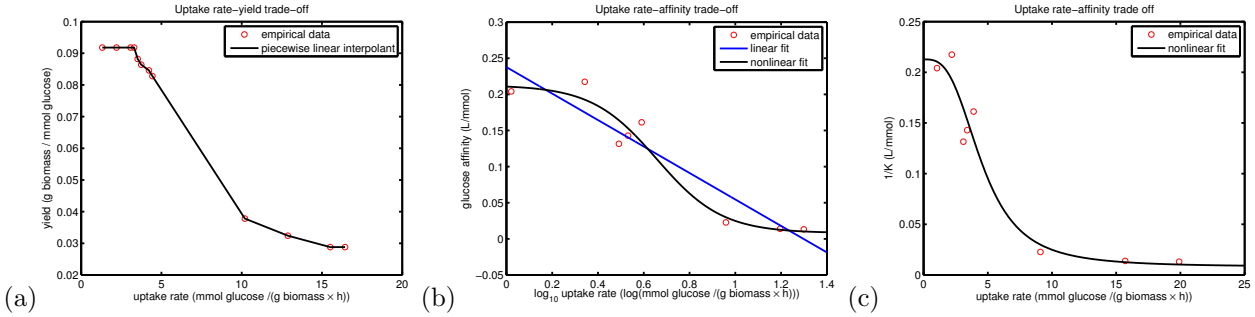


Evidence for a negative relationship between growth rate and yield within one bacterial species, *E. coli*, has existed in the classic microbiology literature for over half a century and by collating two sources we can determine a possible form for this trade-off. Monod reported growth rate and cell yield of *E. coli* cultures grown at a range of temperatures and a variety of substrates [26]. The individual relationships of rate and yield versus temperature from this work are presented within a few pages of each other in the classic *Bacterial Metabolism* by Stephenson [41], but to our knowledge these two experiments have not been integrated to elucidate a possible relationship between rate and yield. It is known that the optimum growth temperature of *E. coli* is close to 37^oC. Noting that growth yield declines from suboptimal to optimal temperatures, interpolating the data from the graphs of Figure 10, page 171 and Figure 16, page 178 in [41] to allow for a 1^o discrepancy in temperatures and merging the two data sets, this results in the negative and indeed sigmoidal relationship shown in supplement fig. 4(a).

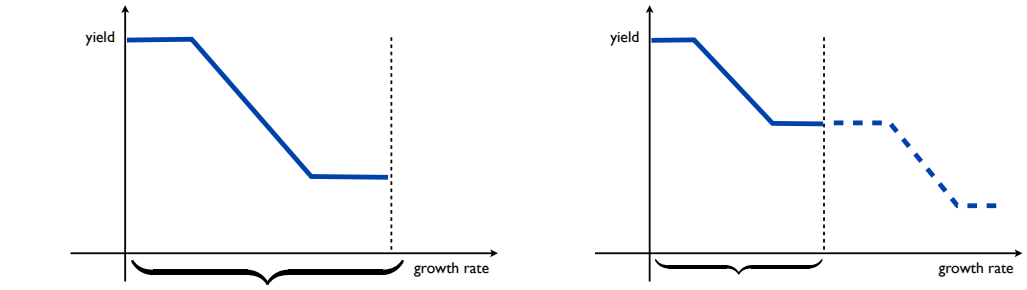
We can corroborate this with further evidence of a within-species rate-yield trade-off using data found in Table 28, p. 110 and Table 30, p. 111 of [26] where the yield and rate experiments were performed on *E. coli* at the same temperature and so no interpolation is necessary, moreover, the growth medium (*milieu S*) had either glucose, sorbitol, mannitol or maltose as carbon source. We note that the fits to this data shown in supplement fig. 4(b) all have a sigmoidal nonlinear form. We conclude from the evidence presented here that in all the trade-off forms examined, including yeast, *E. coli* and soil communities, the rate-yield trade-off in microbes may follow a sigmoidal structure and so possess at least one convex component.

doi:10.1038/nature09905

supplement fig. 5 – (a) An empirically-determined rate-yield trade-off in yeast indicates that different uptake rates can correspond to equal, or nearly equal yields (curve derived using data from [45]). (b) Empirical rate-affinity data in *S. cerevisiae* [9] is consistent with a sigmoidal trade-off (*uptake rate* was log-transformed and the data fitted using 'affinity $= a - bR + c/(1 + \exp(d+eR))$' where R denotes the base-10 logarithm of uptake rate, the 95% confidence interval for the parameter c is $(0.0895, 0.320)$ and the linear fit is shown for comparison). (c) This is plot (b) reproduced with a linear scale for the uptake rate and where $1/K$ on the y-axis denotes glucose affinity.



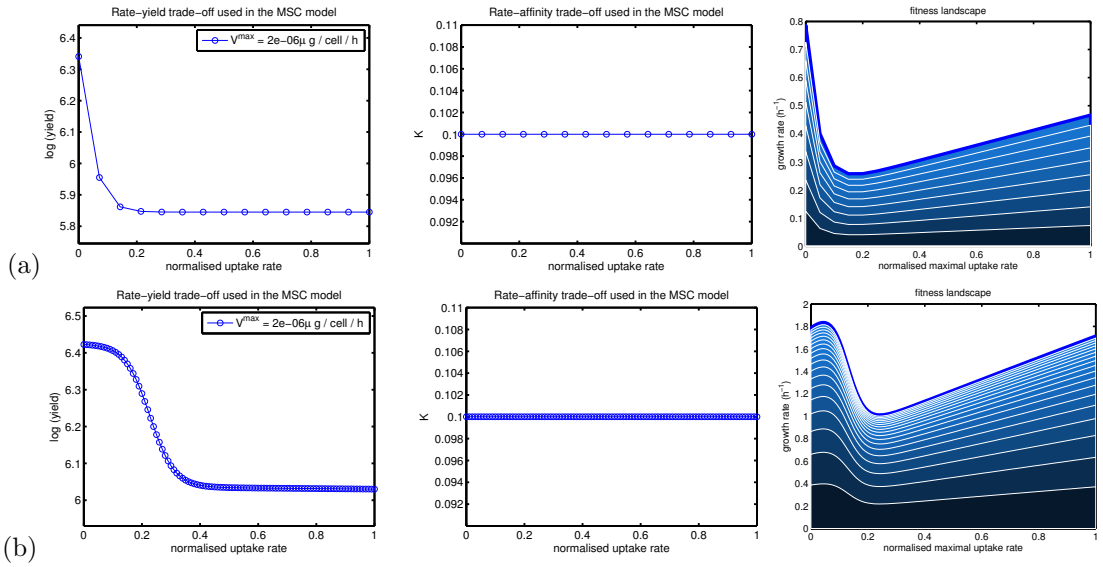
supplement fig. 6 – (a) A cartoon form of the empirical rate-yield trade-off from supplement fig. 5(a). (b) A possible trade-off obtained by extending (a) to a wider physiological regime.



Supplement fig. 5(a) displays an empirically-determined rate-yield trade-off obtained using data taken from [45] illustrating that microbial cells with different uptake rates may exhibit very similar yields, supplement fig. 15 below illustrates how this empirical curve is translated into a simplified form to be used within the mathematical model (3). As a result, all of the trade-offs we use to understand the diversity properties generated by solutions of equation (3) are shown in supplement figs. 3, 4, 5, 7, 8 and 15 where they exhibit linear, sigmoidal and staircase forms. The staircase trade-offs encode some redundancy in the sense that the same, or very similar, yield and affinity phenotypes may be expressed by cells with different uptake rates. Our rationale for our occasional use of staircase or multiply sigmoidal trade offs is illustrated in supplement fig. 6. We reason that trade-off data may take different forms when measured in different physiological regimes, as can be seen in supplement fig. 3 and so, when extending the trade-off data given for example in supplement fig. 5(a) and 5(b), we hypothesise that several regimes may exhibit sigmoidal behaviour. When this behaviour is concatenated to form a single trade-off, it is possible that a staircase-like shape will result.

4.3. **Rate-yield trade-off: biochemical basis.** For several of the above cases, the biochemical basis of the rate-yield trade-off involves alternative routings for metabolism that differ both in the ATP production and

supplement fig. 7 – The left-hand diagram in each plot is a rate-yield trade-off, the middle plot is a rate-affinity trade-off and the right-hand diagram is the resulting fitness landscape: (a) a convex rate-yield trade-off and (b) a sigmoidal rate-yield trade-off. (Darker regions in the rightmost diagrams indicate lower resource concentrations and the highest peaks (approaching the thick blue line) are only accessible when resources are abundant.)

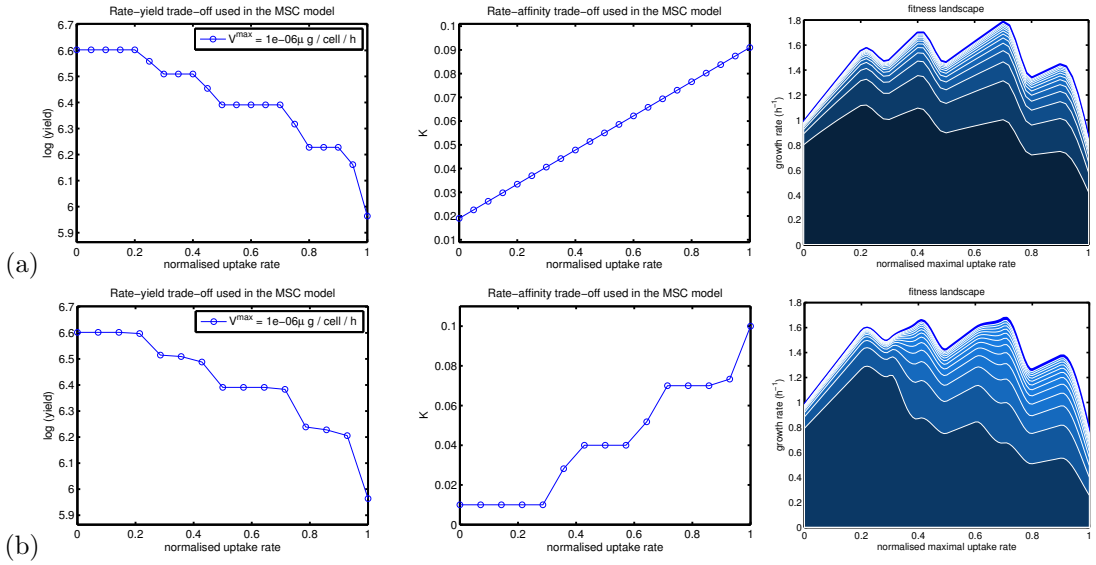


the conditions under which they are employed. For example, *E. coli* can employ a less efficient fermentative metabolism when glucose and oxygen are abundant. The Crabtree effect in yeast involving the diversion of pyruvate to ethanol, rather than using the ATP-producing citric acid cycle, provides one example of such an alternative routing. In this example, as with others, the biochemical pathways divide between a long, ATP-high yield pathway and a short, low-yield overflow pathway [29], the high yield pathway being utilised when resources are scarce. Similarly, the trade-offs in *Holophaga foetida*, *Acetobacter methanolicus* and *Saccharomyces kluyveri* all involve shifts from catabolic into overflow metabolic branches with lower biomass yield, when resources are abundant [28, 25].

In the absence of alternative routings, a rate-yield trade-off may yet be found. For example, with an abundant substrate, anabolism and catabolism can be unbalanced. This can lead to ATP spillage [36] and may be mediated by the usage of ATP in futile cycles, such as futile cycles of protons, again see [36]. More important, possibly, are direct costs of toxic schemes of sugar usage. Most notable in this context is methylglyoxal, the toxic product of intermediates of glycolysis through non-enzymic elimination of phosphate groups [36], this is highly cytotoxic being both inhibitory of cell division and affecting potassium concentrations. It is thought that if there were insufficient ADP to run normal glycolysis, for example when there is an excess of energy, triose phosphates can be employed in the non-ATP-generating methylglyoxal shunt [5]. A consequence of this is that when intermediates of glycolysis are abundant and cannot be processed sufficiently quickly, toxic products accumulate and possible ATP producing molecules not produced, thus reducing yield at high resource concentrations. Others suggest that at high ATP production rates, rates of protein production must also be high, which is likely to come at a direct cost owing to the reduced rate of production of other key proteins, thus high rate can come at a cost in terms of yield [48].

4.4. **Is the rate-yield trade-off likely to be common?** While the above evidence suggests both that the rate-yield trade-off can be found and what the underlying biochemical basis may be, it does not suggest

supplement fig. 8 – (a) A nonlinear rate-yield and a linear rate-affinity trade-off, (b) the same rate-yield trade-off as (a) and a different rate-affinity trade-off. The right-hand column illustrates that the fitness landscape resulting from each trade-off combination depends less on the form of the rate-affinity trade-off at higher resource concentrations, only at low concentrations does the rate-affinity relationship matter.



whether such trade-offs are likely to be general, beyond the fact that the underlying metabolism, such as overflow metabolism, is commonly observed. Two approaches have been taken to address this issue, both of which suggest trade-offs are either inevitable or likely. Perhaps the most convincing approach is to argue that tradeoffs are themselves evolutionarily optimal solutions. To see this, consider the problem of when low-yield overflow metabolism is, and is not used [24].

There are several types of model that address this issue [24]. For example, if toxic metabolic products build up under high resource conditions [20] it is simple to see why overflow metabolism, rapidly purging the system, might be favoured. While this is logically robust, it is unclear how general an explanation this might be. Molenaar *et al.* suggest what they propose to be a potentially more general explanation [24], one that supposes there to be costs of metabolism, including the costs of manufacture of importer channels and enzymes (see also [48]). Under low resource conditions, the costs of obtaining resources must be high as resources are scarce, ensuring that the optimal strategy requires high ATP payback for a necessarily high investment. Under resource abundance metabolites are cheap, it is thus not worth investing more to obtain more ATP, rather the optimal strategy is high rate low yield, employing the overflow metabolism.

To illustrate this they consider a choice between a metabolically efficient (high ATP yield) pathway and a catalytically efficient but metabolically inefficient pathway (low yield overflow). An optimal cell, they show, shifts from the catalytically efficient to the metabolically efficient pathway when going from high to low substrate concentrations. With decreasing substrate concentrations there comes a point at which channelling substrate through a catalytically efficient resource-wasting pathway does not pay off anymore in terms of growth rate, rather, high pay-off for the high investment is needed and hence routing through the slow, but ATP rich, pathway.

This model correctly predicts not just that the ATP rich pathway should be used less under resource abundance, but that it should be actively down-regulated to save costs [24]. Consistent with such models is evidence that different optimality criteria apply at high and low resource levels, evidence that has come from

fitting flux balance analysis models to empirical data at a range of resource levels [38]. As expected, given the above, unlimited growth on glucose is best described by the maximisation of ATP yield per flux unit, while under nutrient scarcity in continuous cultures, by contrast, the maximisation of the overall ATP or biomass yields achieves the highest predictive accuracy.

An alternative line of argument postulates that rate-yield trade-off is a thermodynamic necessity [32] and approaches including explicit kinetic descriptions of ATP production are argued to lead to similar results [40]. While data from *Holophaga foetida* is argued to be consistent with this model [17], the principle has been asserted as operating only when the cell operates near thermodynamic equilibrium [34]. Nevertheless, if y here denotes the yield of ATP per glucose, the function

$$\mathcal{J}_{ATP}(y) = y\left(\exp\left(\frac{y\Delta_1 - \Delta_2}{kT}\right) - 1\right)$$

is the flux obtained for a simple model of glycolysis [13, equation (43)] where k denotes Boltzmann's constant, T (here only) denotes temperature, Δ_1 and Δ_2 are certain free energy differences. As both Δ_1 and Δ_2 are negative [13], this provides a kinetic derivation of a rate-yield trade-off, at least for the case of glycolysis.

4.5. **Rate-affinity trade-off.** Organisms may be subject to multiple trade-offs simultaneously and so, in addition to the rate-yield trade-off described in some detail above, we will implement the so-called *rate-affinity* trade-off in our models. Evidence for a trade-off between uptake rate and affinity has been observed in *S. cerevisiae* [9] and *E. coli* [47] and data from the former has been used to produce supplement figs. 5(b and c). This trade-off, whereby higher affinity of a receptor for a substrate comes at a cost of lower rate of transportation of that substrate across the cell membrane is not peculiar to microbes and has already been postulated as one possible reason why no single best strategy for resource competition has evolved in plants [4, 11].

5. The bifurcation structure of steady-state solutions

Returning to the MSC equation into which we can encode the rate-yield and rate-affinity trade-offs discussed in the previous section, we now fix the values of d and S_0 and consider ϵ to be a free parameter in equation (3). Note that the steady-state problem of (3) can be written

$$0 = \epsilon(M - I)\boldsymbol{b} + (\boldsymbol{G}(S) - d)\boldsymbol{b}, \tag{7a}$$

$$0 = d(S_0 - S) - \langle \boldsymbol{U}(S), \boldsymbol{b}\rangle, \tag{7b}$$

and it will be helpful to cast equation (7a) as the S-dependent eigenproblem

$$d \cdot \boldsymbol{b} = (\epsilon(M - I) + \boldsymbol{G}(S))\,\boldsymbol{b}, \tag{8}$$

with d as the eigenvalue and $\boldsymbol{b}$ as the eigenvector. As M is assumed to be a symmetric, irreducible and stochastic matrix, by the Perron-Frobenius theorem unity is its largest eigenvalue. Hence $\max_{\boldsymbol{b}\neq\boldsymbol{0}} \frac{\langle M\boldsymbol{b},\boldsymbol{b}\rangle}{\|\boldsymbol{b}\|^2} = 1$ and so

$$\frac{\langle (M - I)\boldsymbol{b}, \boldsymbol{b}\rangle}{\|\boldsymbol{b}\|^2} = -1 + \frac{\langle M\boldsymbol{b}, \boldsymbol{b}\rangle}{\|\boldsymbol{b}\|^2} < 0$$

for every non-zero vector $\boldsymbol{b}$ that is not a scalar multiple of the invariant density of M.

The variational characterisation of eigenvalues and eigenvectors can now be used to show that the MSC equation obeys the following optimsation principle in steady-state:

$$d = \max_{\boldsymbol{b}\neq\boldsymbol{0}} \overbrace{\left(\frac{\epsilon\,\langle (M - I)\boldsymbol{b}, \boldsymbol{b}\rangle}{\|\boldsymbol{b}\|^2} + \frac{\langle \boldsymbol{b}, \boldsymbol{G}(S)\boldsymbol{b}\rangle}{\|\boldsymbol{b}\|^2}\right)}^{\text{negative} \quad + \quad \text{positive}} \quad \overset{\text{defines}}{=} \quad \rho^*(\epsilon, S). \tag{9}$$

The eigenvector providing the relative frequency of different bacterial types that achieves the maximum in (9) will be written throughout the remainder as $\boldsymbol{b}^*(\epsilon, S)$ and we shall assume for definiteness that it is normalised so that $\langle \mathbf{1}, \boldsymbol{b}^*(\epsilon, S)\rangle = 1$. Note, when d and ϵ are fixed, equation (9), that is

$$d = \rho^*(\epsilon, S) \tag{10}$$

is a single equation for the equilibrium resource concentration S. To help make the notation a little more transparent, notice that the two components of the steady-state $(\boldsymbol{b}, S)$ of (3) can be written in terms of one another as $\boldsymbol{b} = \lambda \cdot \boldsymbol{b}^*(\epsilon, S)$, where λ is a proportionality constant whose significance will become apparent in the following analysis.

The purpose of this analysis is to identify just two different bifurcation structures of (7) according to the value of the parameter

$$d_{crit}(S_0) \overset{\text{defines}}{=} \frac{1}{n}\sum_{i=1}^{n} G_i(S_0).$$

Note that this function of S_0, d_{crit}, is the mean absolute fitness of the population at the resource supply concentration and this parameter plays a fundamental role in understanding the two bifurcation structures alluded to above. The first of these structures can be stated as follows.

Proposition 1. *Suppose that $S_0 > 0$ is fixed and $d < d_{crit}(S_0)$, then for each $\epsilon > 0$ there is exactly one strictly positive steady-state $(\boldsymbol{b}(\epsilon), S(\epsilon))$ of (3). Morever, as $\epsilon \to \infty$ the distribution of types converges to the maximal diversity state (the uniform distribution) in the sense that*

$$\lim_{\epsilon\to\infty} \frac{\boldsymbol{b}(\epsilon)}{\langle \boldsymbol{b}(\epsilon), \mathbf{1}\rangle} = \frac{1}{n}\mathbf{1}.$$

As $\epsilon \to 0$ the distribution of types converges to the minimal diversity, competitive exclusion state in the sense that if $\lim_{\epsilon\to 0} \boldsymbol{b}(\epsilon) \neq 0$, then

$$\lim_{\epsilon\to\infty} \frac{\boldsymbol{b}(\epsilon)}{\langle \boldsymbol{b}(\epsilon), \mathbf{1}\rangle} = P \cdot (1, 0, ..., 0)^T$$

for some permutation matrix P.

Proof. First fix $S_0 > 0$. The first part follows from the fact that the eigenproblem (8) can be re-written

$$(d + \epsilon) \cdot \boldsymbol{b} = \epsilon M \boldsymbol{b} + \boldsymbol{G}(S)\boldsymbol{b}. \tag{11}$$

As M is irreducible and non-negative, for $\epsilon > 0$ so too is the matrix $\epsilon M + \mathrm{diag}(\boldsymbol{G}(S))$ and therefore, for each $S \geq 0$, the latter has an algebraically simple eigenvalue $e(\epsilon, S)$ corresponding to a positive eigenvector $\alpha\boldsymbol{b}$, where $\alpha > 0$ is any constant. The matrix $\epsilon M + \mathrm{diag}(\boldsymbol{G}(S))$ has entries that increase monotonically with respect to both ϵ and S and so its dominant eigenvalue also increases monotonically with respect to both variables. As a result, for each fixed d and ϵ, the resulting equation

$$d + \epsilon = e(\epsilon, S)$$

has at most one solution for the steady-state sugar concentration S and equation (7b) then uniquely determines the value of α. This argument ensures that when a solution of (7) exists with $\boldsymbol{b}$ non-negative, this solution is unique.

We now establish the existence of steady-state solutions by solving (7). To achieve this, first set $\delta = 1/\epsilon$ and then write $\boldsymbol{b} = \theta\mathbf{1} + \boldsymbol{v}$ where $\boldsymbol{v} \in \mathrm{span}\{\mathbf{1}\}^\perp$ in (7), so that $\langle \boldsymbol{v}, \mathbf{1}\rangle = 0$. As a result, we find that (7) defines an equation $F(\boldsymbol{v}, \theta, S, \delta) = 0$ where the mapping $F : \mathrm{span}\{\mathbf{1}\}^\perp \times \mathbb{R}^3 \to \mathbb{R}^{n+1}$ is differentiable. Let us write $\Pi : \mathbb{R}^n \to \mathrm{span}\{\mathbf{1}\}^\perp$ for the orthogonal projection that satisfies $\Pi[\boldsymbol{b}] = \boldsymbol{v}$ and $\Pi[\mathbf{1}] = 0$. (Note that $\delta = 0$ corresponds to '$\epsilon = \infty$' in this formulation.)

We now tackle the equation $F(\boldsymbol{v},\theta,S,\delta)=0$, which can be written in full as

$$0 = (M-I)\boldsymbol{v}+\delta\cdot\Pi[(\boldsymbol{G}(S)-d)(\theta\mathbf{1}+\boldsymbol{v})], \qquad (\in \mathrm{span}\{\mathbf{1}\}^{\perp}) \tag{12a}$$

$$0 = \langle \mathbf{1}, (\boldsymbol{G}(S)-d)(\theta\mathbf{1}+\boldsymbol{v})\rangle\,, \tag{12b}$$

$$0 = d(S_0-S)-\langle \boldsymbol{U}(S), \theta\mathbf{1}+\boldsymbol{v}\rangle\,, \tag{12c}$$

and we do this by first setting $\delta=0$. This yields a solution of $F=0$, namely $F(\boldsymbol{v}_*,\theta_*,S_*,0)=0$ where

$$\boldsymbol{v}_*=0\in\mathrm{span}\{\mathbf{1}\}^{\perp},\ \theta_*=\frac{\langle \boldsymbol{U}(S_*),\mathbf{1}\rangle}{d(S_0-S_*)}>0 \qquad \text{where} \qquad \frac{1}{n}\langle \boldsymbol{G}(S_*),\mathbf{1}\rangle = d. \tag{13}$$

The existence of an S_* satisfying the latter in (13) necessitates a solution $S_*\le S_0$ of the equation $d=d_{crit}(S_*)$, a solution that cannot exist if $d>d_{crit}(S_0)$. So, assuming $d<d_{crit}(S_0)$, the monotonicity of the function $d_{crit}(\cdot)$ with respect to its argument and the fact that $d_{crit}(0)=0$ yields the existence of an $S_*<S_0$ satisfying $d=d_{crit}(S_*)$, as required.

A short computation shows these conditions are sufficient for the derivative $d_{(\boldsymbol{v},\theta,S)}F(\boldsymbol{v}_*,\theta_*,S_*,0)$ to be an isomorphism from its domain of definition to its range and therefore the implicit function theorem can be applied to deduce the local existence of a unique, smooth curve of solutions $(\boldsymbol{v}(\delta),\theta(\delta),S(\delta))$ such that $F(\boldsymbol{v}(\delta),\theta(\delta),S(\delta),\delta)=0$ for all δ near zero. Now, the first component of the solution of (7), $\boldsymbol{b}$, can be written $\boldsymbol{b}(\epsilon)=\theta(1/\epsilon)\mathbf{1}+\boldsymbol{v}(1/\epsilon)$ where $\theta(0)=\theta_*$ and $\boldsymbol{v}(0)=0$, so that $\boldsymbol{b}(\epsilon)=\theta_*\mathbf{1}+O(1/\epsilon)$. This establishes the last part of the proposition because $\boldsymbol{b}(\epsilon)$ is proportional to the vector $\mathbf{1}$ in the limit $\epsilon\to\infty$.

There remains to establish the existence of a solution of (7) for all $\epsilon>0$, equivalently for all $\delta>0$. For this we can appeal to either modern global bifurcation theory [3] or to more classical Leray-Schauder continuation results described in [49], either way we can deduce the existence of a continuum $\mathcal{C}$ (a maximal, closed, connected set) of solutions of the equation $F(\boldsymbol{v},\theta,S,\delta)=0$, unbounded in the solution measure $\nu(\boldsymbol{v},\theta,S,\delta)=\|\boldsymbol{v}\|+|\theta|+|S|+|\delta|$. So, each element of $\mathcal{C}$ has the form $(\boldsymbol{v},\theta,S,\delta)$ and satisfies $F(\boldsymbol{v},\theta,S,\delta)=0$.

Note that a maximum principle can be established using (11): if $\boldsymbol{b}\ge 0$ then from the irreducibility of M we deduce that $\boldsymbol{b}\gg 0$, moreover $S=0$ is not possible at any solution in $\mathcal{C}$. From this it follows that if $(\boldsymbol{v},\theta,S,\delta)\in\mathcal{C}$ and the vector $\boldsymbol{b}=\boldsymbol{v}+\theta\mathbf{1}$ contains any negative entries, by the connectedness of $\mathcal{C}$ there must exist a non-negative $(\boldsymbol{v}',\theta',S',\delta')\in\mathcal{C}$ for which $\boldsymbol{b}'=\boldsymbol{v}'+\theta'\mathbf{1}$ contains a zero entry. However, we then deduce that $\boldsymbol{b}'=\boldsymbol{v}'+\theta'\mathbf{1}\ge 0$ has a zero entry, but this is a contradiction as $\boldsymbol{b}'$ must satisfy (11) with $S=S'\ge 0$. As a result, for any $(\boldsymbol{v},\theta,S,\delta)\in\mathcal{C}$ there results $\boldsymbol{v}+\theta\mathbf{1}\gg 0, S>0$. In other words, $\mathcal{C}$ defines an unbounded continuum of strictly positive solutions of (7).

From (4), $\|\boldsymbol{v}\|+|\theta|+|S|$ must be bounded within $\mathcal{C}$ and so this continuum must be unbounded with respect to the solution measure $\nu(\boldsymbol{v},\theta,S,\delta)=|\delta|=|1/\epsilon|$. As a result, the set $\mathcal{C}$ contains strictly positive solutions of (7) associated with arbitrarily small values of ϵ, that is, with arbitrarily large values of δ. As $\mathcal{C}$ is connected, with solutions of arbitrarily large and small, positive ϵ, the set $\mathcal{C}$ contains a solution of (7) for every value of $\epsilon>0$.

Finally, as $\epsilon\to 0$ along any solution sequence within $\mathcal{C}$, the component $\boldsymbol{b}(\epsilon)$ must converge to a solution, $\boldsymbol{b}$, of the equation $(\boldsymbol{G}(S)-d)\boldsymbol{b}=0$. However, because $\boldsymbol{G}(\cdot)$ is strictly monotone, the diagonal matrix $\boldsymbol{G}(S)-dI$ can only have one zero entry on the main diagonal and so $\boldsymbol{b}$ must have the minimal-diversity form $(0,0,...,0,*,0,...,0)$ where $*$ is a non-negative number. The result now follows. □

The second bifurcation structure is described in the following result.

Proposition 2. *Suppose that $S_0>0$ is fixed and $d>d_{crit}(S_0)$. There is a differentiable function $\epsilon_{crit}(d,S_0)$ such that if $\epsilon\in(0,\epsilon_{crit}(d,S_0))$, then there is exactly one strictly positive steady-state $(\boldsymbol{b}(\epsilon),S(\epsilon))$ of (3).*

However, if $\epsilon > \epsilon_{crit}(d, S_0)$ *then the trivial steady-state* $(\boldsymbol{b}, S) = (0, S_0)$ *is globally stable and* eventually exponentially stable *for the dynamics of (3). As a result, there exists a region in the* (d, ϵ) *plane whose boundary is given by the graph of a function* $d = J(\epsilon)$ *and the line* $d = d_{crit}(S_0)$ *in which the cell population cannot persist (see supplement fig. 9).*

Proof. For the same reason as explained in the proof of Proposition 1, when (3) has a non-zero steady-state solution, it is unique. Now, fix $S_0 > 0$ and write the steady-state equation (7) in the form $F(\boldsymbol{b}, S, \epsilon) = 0$ by suitably defining a smooth mapping F and note the existence of a trivial solution branch of this equation: $F(0, S_0, \epsilon) = 0$ for all $\epsilon \geq 0$.

Consider the linearisation of F about this trivial solution with respect to $(\boldsymbol{b}, S)$, evaluated at $(0, S_0)$:

$$d_{(\boldsymbol{b},S)}F(0, S_0) = \begin{bmatrix} \epsilon(M - I) + \mathrm{diag}(\boldsymbol{G}(S_0)) - dI & 0 \\ -\boldsymbol{U}(S_0)^T & -d \end{bmatrix}.$$

Applying the theorem on bifurcation from a simple eigenvalue [3], the equation $F(\boldsymbol{b}, S, \epsilon) = 0$ undergoes a biologically-relevant bifurcation from the trivial solution when there is a non-zero vector $\boldsymbol{k} \geq 0$ such that $[\epsilon(M - I) + \mathrm{diag}(\boldsymbol{G}(S_0)) - dI]\boldsymbol{k} = 0$, that is

$$\epsilon(M - I)\boldsymbol{k} + \boldsymbol{G}(S_0)\boldsymbol{k} = d\boldsymbol{k} \qquad \text{where} \qquad \langle \boldsymbol{k}, \mathbf{1} \rangle = 1. \tag{14}$$

Using the variational characterisation of the eigenvalues of the matrix in (14), we must find an ϵ such that

$$d = \max_{\boldsymbol{k} \neq 0} \left(\epsilon \frac{\langle \boldsymbol{k}, (M - I)\boldsymbol{k} \rangle}{\langle \boldsymbol{k}, \boldsymbol{k} \rangle} + \frac{\langle \boldsymbol{k}, \boldsymbol{G}(S_0)\boldsymbol{k} \rangle}{\langle \boldsymbol{k}, \boldsymbol{k} \rangle} \right) \overset{\text{defines}}{=} J(\epsilon).$$

Now, the function $J(\cdot)$ so-defined satisfies

$$J(0) = \max_{1 \leq i \leq n} G_i(S_0) \qquad \text{and} \qquad J(0) > J(\epsilon) \geq \frac{\langle \mathbf{1}, \boldsymbol{G}(S_0)\mathbf{1} \rangle}{\langle \mathbf{1}, \mathbf{1} \rangle} = d_{crit}(S_0)$$

for all $\epsilon > 0$ as $J(\cdot)$ is a monotone decreasing function.

It follows that either (i) there is a unique solution of $J(\epsilon) = d$, or else (ii) there is no solution at all of this equation because $d < d_{crit}(S_0)$. Case (ii) arises in the situation described in Proposition 1 in which case there is no bifurcation of solutions from the trivial solution branch and an unbounded set of non-trivial solutions can be globally defined (the set $\mathcal{C}$ defined in the proof of Proposition 1). We are therefore left with (i), in which case to each d we can associate a value of ϵ such that $J(\epsilon) = d$, let us call this solution ϵ_{crit}. We are now able to define the function $\epsilon_{crit} = \epsilon_{crit}(S_0, d)$ given in the statement of the proposition using the inverse function theorem and it satisfies $J(\epsilon_{crit}(S_0, d)) \equiv d$. In this case, the theorem on bifurcation from a simple eigenvalue [3] can be applied to demonstrate the existence of a locally unique solution branch $\mathcal{B}$ of the equation $F(\boldsymbol{b}, S, \epsilon) = 0$ that extends from the trivial solution at the bifurcation point $\epsilon = \epsilon_{crit}(S_0, d)$.

In order to describe the region in the (d, ϵ)-plane at which the bifurcation of $\mathcal{B}$ occurs from the trivial solution branch, namely the 1-dimensional set whereby $J(\epsilon) = d$, consider the following. By the Perron-Frobenius theorem, note that the largest real eigenvalue $d + \epsilon$ of the non-negative, irreducible matrix $\epsilon M + \mathrm{diag}(\boldsymbol{G}(S_0))$ corresponding to the eigenvector $\boldsymbol{k} \geq 0$ in equation (14) is algebraically simple when $J(\epsilon) = d$. Therefore, we can apply the implicit function theorem to deduce that because $J(\epsilon_{crit}(S_0, d)) = d$ identically, the function $\epsilon_{crit}(\cdot, \cdot)$ is smooth in its arguments.

Let $\delta = 1/\epsilon$ and define $\tau(\delta) = \delta \cdot J(\epsilon)$. Noting that $\boldsymbol{k}$ defined in (14) also depends smoothly on ϵ (see [3] for a derivation of the fact that algebraically simple eigenvalues and eigenvectors depend smoothly on system parameters) we will write $\boldsymbol{k}(\delta)$ to emphasise this smooth dependence. Now differentiate the expression

$$\epsilon(M - I)\boldsymbol{k} + \boldsymbol{G}(S_0)\boldsymbol{k} = J(\epsilon)\boldsymbol{k} \qquad \text{i.e.} \qquad (M - I)\boldsymbol{k} + \delta\boldsymbol{G}(S_0)\boldsymbol{k} = \tau(\delta)\boldsymbol{k}$$

with respect to δ. Using a prime ($'$) to denote the derivative with respect to δ, this gives

$$(M-I)\boldsymbol{k}' + \boldsymbol{G}(S_0)\boldsymbol{k} + \delta\boldsymbol{G}(S_0)\boldsymbol{k}' = \tau'(\delta)\boldsymbol{k} + \tau(\delta)\boldsymbol{k}'. \tag{15}$$

When taking the limit $\delta \to 0$, because $(M-I)\mathbf{1} = 0$ to remove any singularities we must set $\boldsymbol{k}(0) = \frac{1}{n}\mathbf{1}$ and $\tau(0) = 0$, now taking the inner product of (15) with the vector $\mathbf{1}$ yields $\tau'(0) = \langle \mathbf{1}, \boldsymbol{G}(S_0)\rangle / \langle \mathbf{1}, \mathbf{1}\rangle = d_{crit}(S_0)$, and so

$$\lim_{\epsilon\to\infty} J(\epsilon) = \lim_{\delta\to 0} \frac{\tau(\delta)}{\delta} = \lim_{\delta\to 0} \frac{\tau(\delta)-\tau(0)}{\delta - 0} = \tau'(0) = d_{crit}(S_0).$$

This aspect of the proof is illustrated in supplement fig. 9 where the region in the (d, ϵ) plane for which $J(\epsilon) = d$ is illustrated as a thick black line.

That the non-trivial, bifurcating solution branch $\mathcal{B}$ contains the unique, non-trivial solution of (7) only for $0 < \epsilon < \epsilon_{crit}(S_0, d)$ is proven as follows. Recall that $S_0 > 0$ and d are given with $d > d_{crit}(S_0)$. Now suppose that $\epsilon > \epsilon_{crit}(S_0, d)$ and let $\boldsymbol{\kappa} > 0$ be the vector that satisfies $\epsilon(M-I)\boldsymbol{\kappa} + (\boldsymbol{G}(S_0) - J(\epsilon))\boldsymbol{\kappa} = 0$ and note that $d = J(\epsilon_{crit}(S_0, d)) > J(\epsilon)$ because $\epsilon > \epsilon_{crit}(S_0, d)$ and J is a decreasing function. Let us compute $\frac{d}{dt}\langle \boldsymbol{b}(t), \boldsymbol{\kappa}\rangle$ along solutions of (3):

$$\begin{aligned}
\frac{d}{dt}\langle \boldsymbol{b}(t), \boldsymbol{\kappa}\rangle &= \left\langle \frac{d}{dt}\boldsymbol{b}(t), \boldsymbol{\kappa} \right\rangle = \langle \epsilon(M-I)\boldsymbol{b}(t) + (\boldsymbol{G}(S) - d)\boldsymbol{b}(t), \boldsymbol{\kappa}\rangle\,, \\
&= \langle \epsilon(M-I)\boldsymbol{b}(t) + (\boldsymbol{G}(S) - d)\boldsymbol{b}(t), \boldsymbol{\kappa}\rangle = \langle \epsilon(M-I)\boldsymbol{\kappa} + (\boldsymbol{G}(S) - d)\boldsymbol{\kappa}, \boldsymbol{b}(t)\rangle\,, \\
&= \langle \epsilon(M-I)\boldsymbol{\kappa} + (\boldsymbol{G}(S) - \boldsymbol{G}(S_0) + \boldsymbol{G}(S_0) - J(\epsilon) + J(\epsilon) - d)\boldsymbol{\kappa}, \boldsymbol{b}(t)\rangle\,, \\
&= \langle (\boldsymbol{G}(S) - \boldsymbol{G}(S_0) + J(\epsilon) - d)\boldsymbol{\kappa}, \boldsymbol{b}(t)\rangle\,, \\
&= \langle (\boldsymbol{G}(S) - \boldsymbol{G}(S_0))\boldsymbol{\kappa}, \boldsymbol{b}(t)\rangle + (J(\epsilon) - d)\langle \boldsymbol{\kappa}, \boldsymbol{b}(t)\rangle\,.
\end{aligned}$$

From (4), for all $\eta' > 0$ we can find a T such that for all $t > T$, $S(t) \le S_0 + \eta'$. By the monotonicity and continuity of $\boldsymbol{G}(\cdot)$, for all $\eta > 0$ we can find a T such that for all $t > T$, $\boldsymbol{G}(S(t)) \le \boldsymbol{G}(S_0) + \eta$ and we choose the positive quantity $\eta = \frac{1}{2}(d - J(\epsilon))$. It follows that

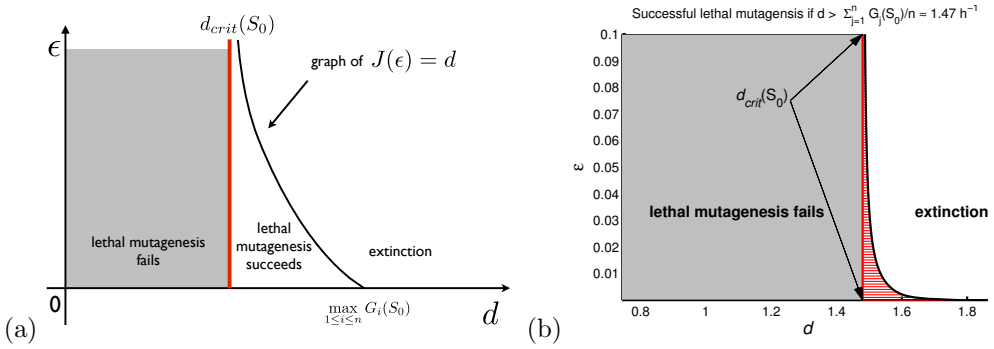
$$\frac{d}{dt}\langle \boldsymbol{b}(t), \boldsymbol{\kappa}\rangle = \langle (\boldsymbol{G}(S) - \boldsymbol{G}(S_0))\boldsymbol{\kappa}, \boldsymbol{b}(t)\rangle + (J(\epsilon) - d)\langle \boldsymbol{\kappa}, \boldsymbol{b}(t)\rangle \le \frac{1}{2}(J(\epsilon) - d)\langle \boldsymbol{\kappa}, \boldsymbol{b}(t)\rangle$$

and so there are constants $a, b > 0$ that depend on T such that for all large enough t, $\langle \boldsymbol{b}(t), \boldsymbol{\kappa}\rangle \le a \cdot e^{-b\cdot t}$ whence $\lim_{t\to\infty} \boldsymbol{b}(t) = 0$ eventually exponentially because $\boldsymbol{\kappa} \gg 0$. This establishes the last part of the proposition and also shows that non-negative solutions of (7) only exist if $\epsilon < \epsilon_{crit}(S_0, d)$ whenever $d > d_{crit}(S_0)$. Standard global bifurcation arguments can now be used to show that $\mathcal{B}$ can be extended to all values of ϵ for which $0 \le \epsilon < \epsilon_{crit}$. □

Remark 2. *All numerical computations performed beyond this point were obtained using a pseudo arc-length strategy* [16, 6] *to rapidly determine equilibrium states of the MSC equation with ϵ as the bifurcation parameter. This numerical technique is an extension of Newton's method that can cope with the presence of multiple solutions, bifurcations, phase transitions and rapid continuous transitions between different solutions.*

5.1. **Lethal Mutagenesis.** Proposition 2 identifies a relationship between S_0 and d whereby *lethal mutagenesis* may be a successful antibacterial strategy in (3). Namely, if $d > \frac{1}{n}\sum_{i=1}^{n} G_i(S_0) = d_{crit}(S_0)$, then by raising the mutation rate ϵ it is possible to push the population towards, and beyond a bifurcation point whereafter the fate of each cell is to be lost from the chemostat because the cells cannot divide sufficiently quickly. The red area in supplement fig. 9(b) illustrates a parameter region where lethal mutagenesis can be

supplement fig. 9 – (a) The key ingredients in the proof of Proposition 2: it highlights a region in the (d, ϵ) plane where the evolving population can be driven to extinction by increasing mutation rates. (b) An instance of (a) computed numerically using the MSC equation.



successfully applied and supplement fig. 10(b)) illustrates the reduction in bacterial density obtained from increasing the mutation rate.

Supplement fig. 10(b) also shows that as the mutation rate is increased, a bifurcation point is eventually encountered at $\epsilon = \epsilon_{\text{crit}}$ beyond which extinction is assured. This bifurcation point is readily shown to be a transcritical bifurcation in the sense of [1]. This establishes the result that along the equilibrium solution branch that supports non-zero cell densities illustrated in the left-hand plot of supplement fig. 10(b), there is a region within which bacterial density, $\langle \mathbf{1}, b(\epsilon) \rangle = \sum_{i=1}^{n} b_i(\epsilon)$, reduces approximately linearly near the bifurcation point as ϵ increases. (This is particularly clear in supplement fig. 10(b) at the label marked TB.)

Standard bifurcation theory gives the existence of an estimate of the form

$$\sum_{i=1}^{n} b_i(\epsilon) = \text{Constant} \cdot |\epsilon - \epsilon_{\text{crit}}| + O\left((\epsilon - \epsilon_{\text{crit}})^2\right) \tag{16}$$

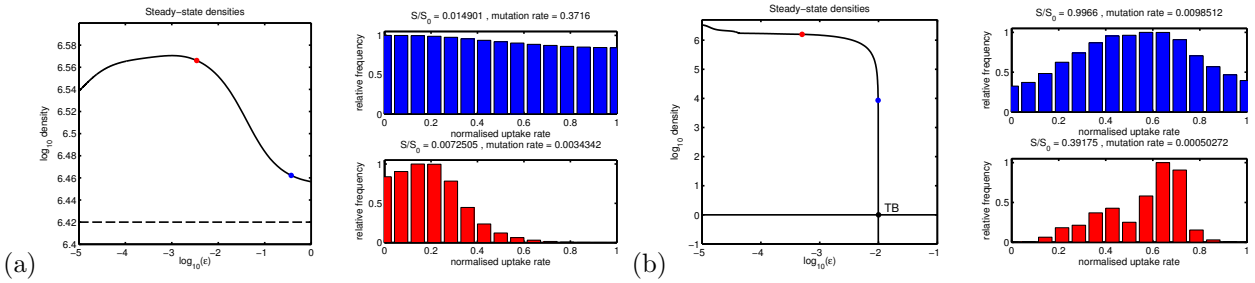
as $\epsilon \to \epsilon_{\text{crit}}$ from below but the region over which this approximation is valid, near to where $\log(\epsilon) \approx -2$, is very narrow in supplement fig. 10(b). As the mutation rate increases, the cell population is almost constant in density up until the point where a rapid and critical decline finally occurs. This rapid decline may well be system-specific, but the linear approximation (16) is a universal feature of equation (3) within (A1-A7) and the set of assumptions given in Proposition 2; such a linear reduction in density has been found in other theories of successful lethal mutagenesis [22].

Conversely, Proposition 1 states that lethal mutagenesis may fail as an antibacterial strategy if $d < d_{crit}(S_0)$. In this case, raising mutation rates may still reduce the density of cells in the chemostat but there is no bifurcation point at which extinction is assured and the population will be driven towards a stable, diverse state. The grey regions in supplement fig. 9 illustrate parameter regions where lethal mutagenesis may fail and supplement fig. 10(a) provides a numerical example of this: note how cell densities do not fall below approximately $10^{6.44}$ cells per ml in this computation.

6. Predictions of the MSC model: the coexistence hypothesis

We are now interested in mutation rates in the MSC equation that are not so high that the population is driven towards extinction, as described in Proposition 2, nor towards the high-diversity state described in Proposition 1. Moreover, Proposition 1 informs us that very low mutation rates are associated with a state of competitive exclusion where *survival of the fittest* applies. As we seek to understand how the MSC model can support the coexistence of different quasispecies of cell types, we must turn our attention to a cell population

supplement fig. 10 − Respectively, (a) and (b) are numerical illustrations of the results contained in Propositions 1 and 2 when rate-yield and rate-affinity trade-offs are invoked simultaneously. The relative frequencies of cell types in the population are shown as red and blue bar diagrams. The antibacterial strategy of *lethal mutagenesis* is successful in diagram (b) as equation (7) possesses a transcritical bifurcation at a critical mutation rate ϵ_{crit} (shown as the label TB where $\log_{10}(\epsilon_{\text{crit}}) \approx -2$) whereas (a) does not possess such a bifurcation and the strategy fails. Common with all transcritical bifurcations, when it occurs the solution branch emanates from the bifurcation point at TB as a locally linear path.



at 'intermediate' mutation rates where mathematical and analytic results are hardest to obtain. As a result, we now resort to numerical computations with the one goal in mind of finding conditions under which the 'fittest' and the 'flattest' can coexist in the system of equilibrium equations (7).

6.1. **Steady-state sugar consumption decreases with increasing mutation rate.** Equation (10) can be used to probe the relation between the mutation rate and consumption of the limiting carbon source in steady-state. Assuming $(\boldsymbol{b}(\epsilon), S(\epsilon))$ to be an equilibrium of (3) parameterised by mutation-rate, upon differentiating (10), that is $\rho^*(\epsilon, S(\epsilon)) \equiv d$, with respect to ϵ we find that

$$\frac{dS}{d\epsilon}(\epsilon) = -\frac{\langle \boldsymbol{b}(\epsilon), (M-I)\boldsymbol{b}(\epsilon)\rangle}{\langle \boldsymbol{b}(\epsilon), \boldsymbol{G}'(S(\epsilon))\boldsymbol{b}(\epsilon)\rangle} > 0.$$

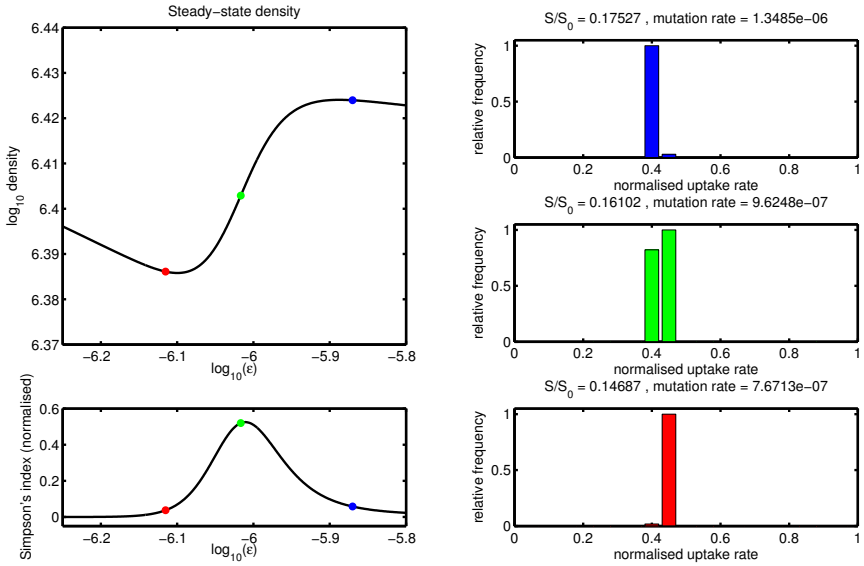
From this inequality we deduce that the equilibrium value of *consumed* resource in the chemostat reduces as mutation rate increases.

Although a decline in consumed resource is expected to be correlated with a reduction in population density, thus providing a rationale for the use of lethal mutagenesis, we saw in the previous section that this logic may fail. Whether or not the population density declines depends crucially on the nature of the specific trade-offs that form the fitness landscape. Inspect, for example, supplement fig. 11 wherein population densities can increase with an increasing mutation rate (the fitness landscape shown in supplement fig. 8(a) is derived using both rate-yield and rate-affinity trade-offs and this landscape was used for the computation that produced supplement fig. 11).

6.2. **Cell density scales linearly with S_0.** Equation (10) can be viewed as an equation for S that can be solved (via the implicit function theorem) for S as a function of (d, ϵ) and, as a result, it is clear that the equilibrium value of S is independent of S_0 in the MSC model. Put more simply, once d and ϵ are fixed, so too is S whatever value S_0 may happen to be.

This is a feature the MSC model (3) has in common with the chemostat models presented in [39] and it can be exploited to make a prediction: using (7), as S depends only on (d, ϵ), so too does the eigenvector $\boldsymbol{b}$ in (7a) up to the scalar multiple λ defined earlier in the text that we can now determine from (7b). Indeed, $\boldsymbol{b}$ must be proportional to the vector $\boldsymbol{b}^*(\epsilon, S(d, \epsilon))$ as defined in §5 and it follows that we can write the vector $\boldsymbol{b}$

supplement fig. 11 – Although an increase in mutation rates decreases resource consumption at equilibrium in the MSC model, trade-offs can conspire to create increases in density by locally flattening fitness peaks. Hence, as this example shows, population density may increase when mutation rates increase.



in (7a) as $\boldsymbol{b} = \lambda \cdot \boldsymbol{b}^*(\epsilon, S(d,\epsilon))$, including the constant λ. In order to determine the value of λ, using (7b) we find that

$$\lambda = \lambda(S_0, d, \epsilon) = \frac{d(S_0 - S(d,\epsilon))}{\langle \boldsymbol{U}(S(d,\epsilon)), \boldsymbol{b}^*(\epsilon, S(d,\epsilon))\rangle}.$$

Thus, provided there is sufficient resource to support the population, and $S_0 > S(d,\epsilon)$ is a critical condition needed for this, biomass scales approximately linearly with the resource supply concentration in the sense that

$$\langle \boldsymbol{b}, \mathbf{1}\rangle = \lambda(S_0, d, \epsilon) = \mathsf{Constant} \cdot S_0 + O(1)$$

as $S_0 \to \infty$; here $O(1)$ is standard notation for any bounded function of S_0.

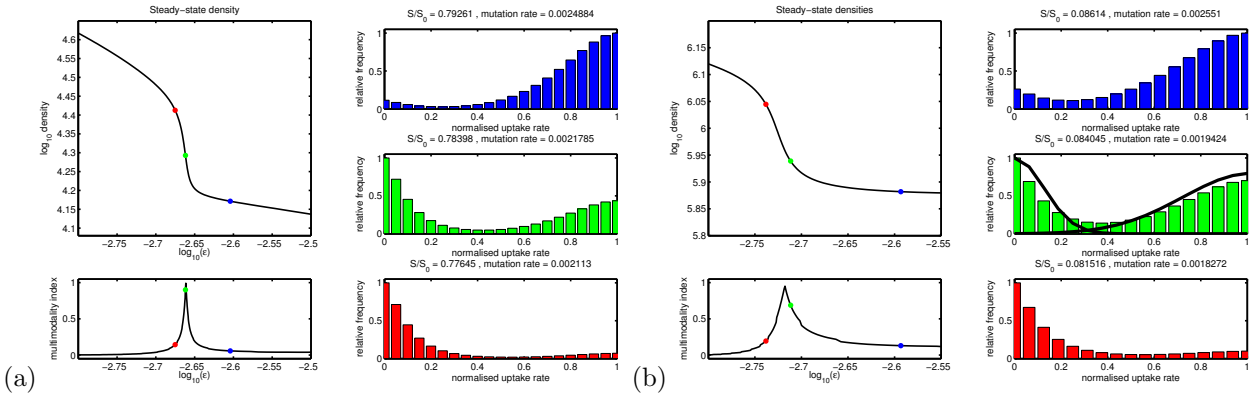
This analysis highlights a non-physical feature of chemostat models like (1), there is no explicit 'volume' parameter of the culture vessel and we simply measure cells per unit volume. However, volume effects are real and it cannot be true that biomass scales linearly with resource supply for arbitrarily large values of S_0, eventually space itself will be limiting. Spatial effects are observed in chemostat experiments and polymorphism can be maintained between producers and non-producers of adhesins that allows cells to stick to the vessel wall or to other cells (see, for example, [50]). As bacterial populations can maintain a diversity of types due to spatial structure, it is important that we do not introduce spatial effects into (1) but with this stipulation there must come the caveat that our models cannot describe experimental devices at high sugar concentrations. Nevertheless, the prediction that biomass scales linearly with S_0 stated above is borne out in practise (see, for example, [15, Fig. 1] for *E.coli* limited by glucose).

6.3. **Diversity is independent of S_0.** With the discussion of the previous section in mind, we now suppose that $\mathcal{H}$ is a diversity measure on n-types, such as the Simpson index $\mathcal{H}_S$. We imposed the condition that any diversity measure should be scale-invariant in §1 and so the diversity of the population in the chemostat at equilibrium given by

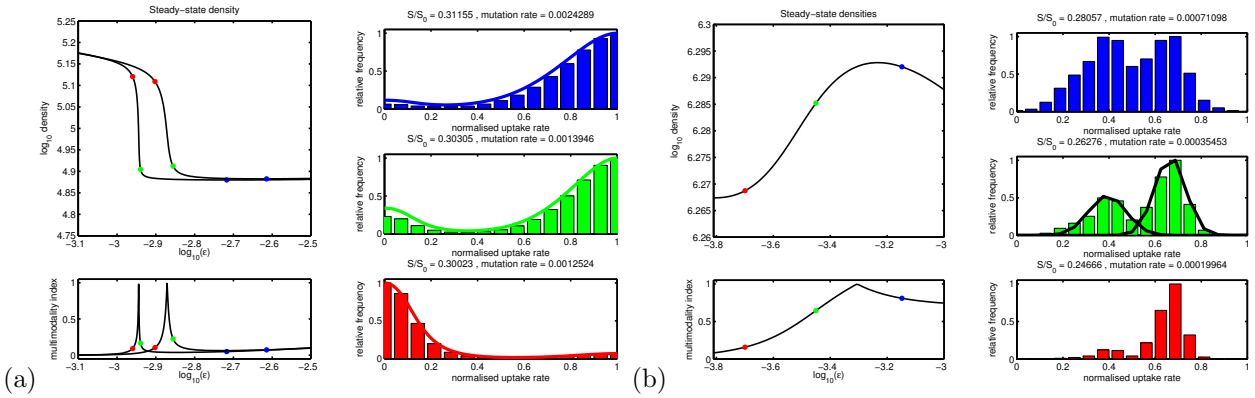
$$\mathcal{H}(\boldsymbol{b}(S_0, d, \epsilon)) = \mathcal{H}(\lambda(S_0, d, \epsilon) \cdot \boldsymbol{b}^*(\epsilon, S(d,\epsilon))) = \mathcal{H}(\boldsymbol{b}^*(\epsilon, S(d,\epsilon))),$$

a number that is clearly *independent* of the supply concentration S_0. Our observation here is consistent with a prediction from Thompson's Geographic Mosaic Theory [42] which states that diversity is constant across environmental gradients for systems having only G×E interactions and, indeed, the MSC equation is a model with this type of interaction.

supplement fig. 12 − Bar diagrams of the relative frequencies of cell types illustrating co-maintenance of the fit and the flat at intermediate mutation rates in the MSC model when there is (a) no rate-affinity trade-off but a convex rate-yield trade off (shown in supplement fig. 7(a)), (b) a linear rate-affinity trade-off with a convex rate-yield trade-off (resulting fitness landscape not shown). (b/right column/middle) A superposition of Gaussians has been plotted as black lines to highlight each quasispecies.



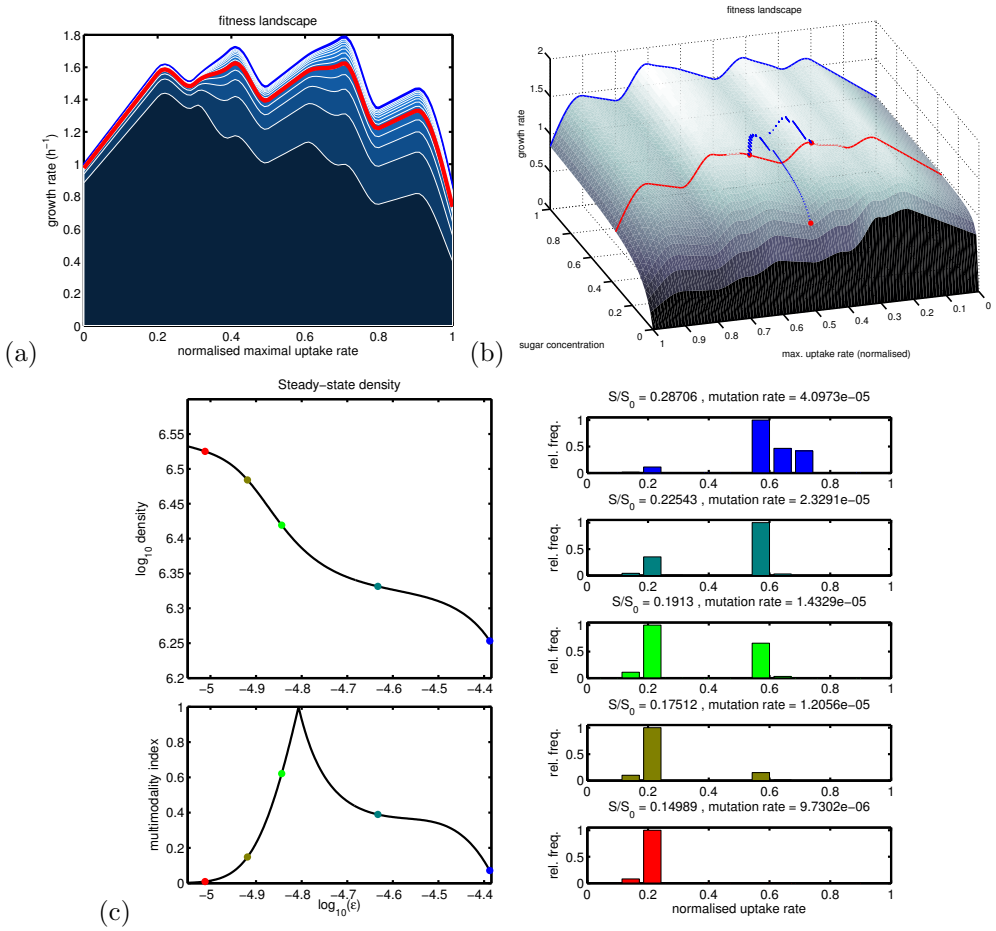
supplement fig. 13 − Co-maintenance of the fit and the flat at intermediate mutation rates in the MSC model. (a) No rate-affinity trade-off and a sigmoidal rate-yield trade off was used (see supplement fig. 7(b)) and two datasets are shown: when $n = 14$ cell relative type frequencies are shown as histograms, when $n = 100$ relative frequencies shown as lines in the right-hand column. (b) The trade-offs illustrated in supplement fig. 8(a) were used to produce this diagram showing it has a wider coexistence zone than in (a). (b/right column/middle) A superposition of Gaussians has been plotted as black lines to highlight each quasispecies.



6.4. **The MSC model can maintain both fit and flat quasispecies.** A fourth prediction of the MSC model under our assumptions is the *coexistence hypothesis*: if survival of the fittest prevails at low mutation rates and survival of the flattest is found at high mutation rates, under the structural impositions placed upon equation (3) there exists at least one interval of *intermediate* mutation rates for which both the fit and the flat coexist in steady-state.

If fit and flat quasispecies are dominant at low and high mutation rates respectively, the co-maintenance principle to which we allude is a consequence of the continuous dependence of the *unique* steady-state of (3) on the mutation rate. Namely, as we change that mutation rate, we continuously morph the density of types from a unimodal fit distribution into a unimodal flat distribution, at some interval of intermediate mutation rates we expect to observe a bimodal distribution of both fit and flat types. If the existence-uniqueness property of this steady-state were to fail, as illustrated in supplement fig. 16 below, then the coexistence hypothesis can fail.

supplement fig. 14 – (a) A resource-dependent fitness landscape encoding both rate-yield and rate-affinity trade-offs (for the trade-offs see supplement fig. 8(b)) where darker regions correspond to lower resource concentrations. (b) Branching of an initially monomorphic quasispecies (the red dot) is traced through this fitness landscape: the red line denotes the equilibrium fitness landscape upon which two quasispecies are stably maintained. (c) The passage from fittest to flattest with increasing mutation rate for the fitness landscape depicted in (a) passes through an interval of coexisting quasispecies. Here, a monotone decrease in density with increasing mutation rates *is* observed.



The co-maintenance of two quasispecies is illustrated in supplement figs. 12 and 13 whereby the mutation rate is increased and three different type distributions are illustrated, at low, intermediate and high mutation rates. Supplement fig. 12(a) was produced when the only trade-off implemented in the MSC model was the *rate-yield trade off*, as a result all the cell affinities for the carbon source are equal but the rate-yield trade-off follows a convex form. To produce supplement fig. 12(b), a linear rate-affinity trade-off was invoked in addition

to the convex rate-yield trade-off, this has the effect of widening the interval of mutation rates for which we obtain coexistence.

Supplement fig. 13(a) was obtained using a sigmoidal form for the rate-yield trade-off, a form found for the natural soil microbes illustrated in supplement figs. 3 and 4, but no rate-affinity trade-off was used here. Supplement fig. 13(b) was obtained using the theoretical landscape shown in supplement fig. 8(a) and supplement fig. 13(b,right column,middle plot) illustrates how two approximately normally distributed quasispecies can persist across a range of mutation rates.

Supplement figs. 12, 13 and 14 use a *multimodality index* (the ratio of biomass contained within each quasispecies) to illustrate a window of mutation rates for which co-maintenance is possible: the latter occurs in a narrow window in supplement fig. 12(a), but supplement fig. 12(b) then invokes an additional rate-affinity trade-off which serves to widen the window of mutation rates for which co-maintenance is possible. Supplement fig. 13(a) shows that a sigmoidal rate-yield trade-off alone is also sufficient to support multiple quasispecies but the region of mutation rates for which co-maintenance is largest in these examples can be seen found in supplement fig. 13(b). This illustrates that the width of the co-maintenance region depends on both the nature and on the number of the trade-offs used, moreover supplement fig. 13(a) shows that the number of types in the model may alter the size of the co-maintenance region.

The preceding numerical examples show that the theoretical principle of co-maintenance of quasispecies can be realised in the MSC model and in the search for a rationale to support wider regions of coexistence, consider the following example. Supplement fig. 14 illustrates that a yet wider region of mutation rates supporting coexistence can be found, but in this figure we have invoked both rate-yield and rate-affinity trade-offs of the form illustrated in supplement fig. 8(b). This lends credence to the argument that more complex, undulating fitness landscapes stem from more complex trade-offs and that they, in turn, promote coexistence.

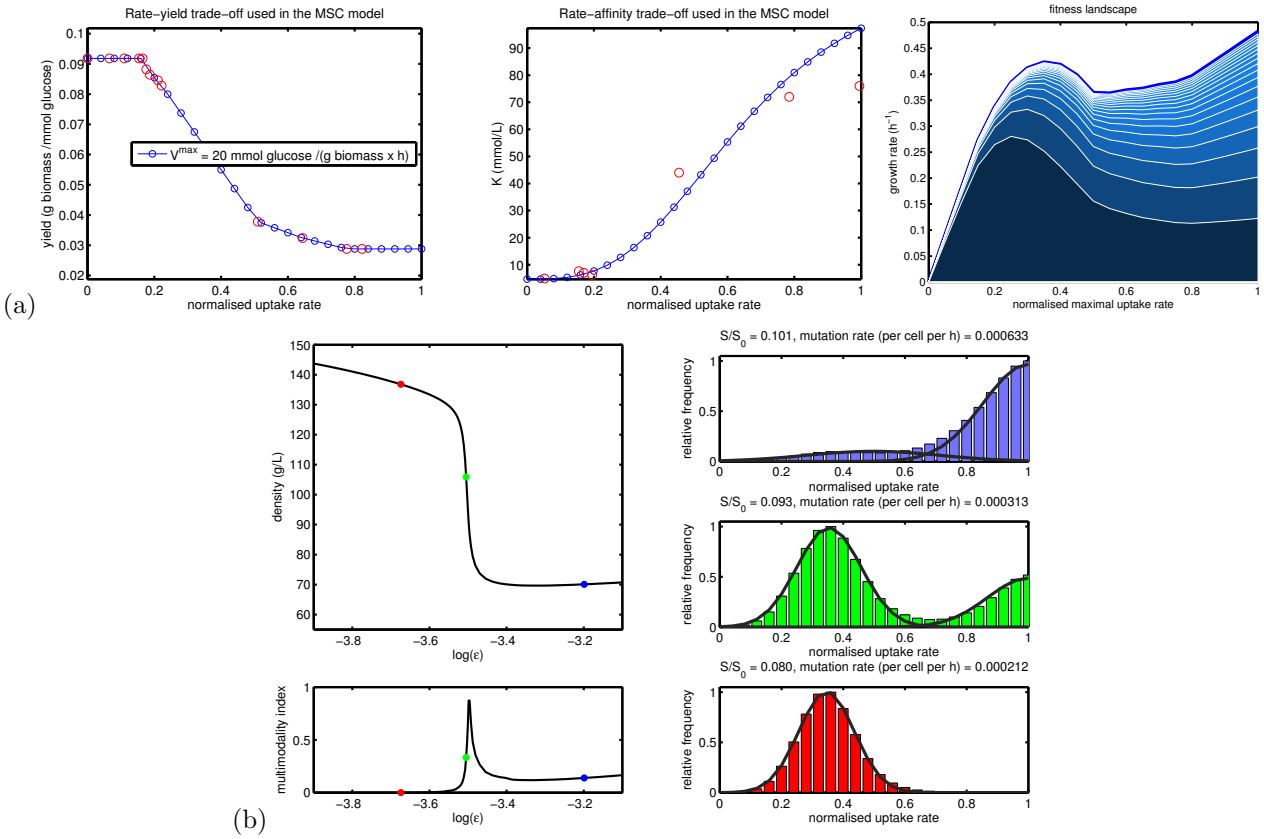
6.4.1. *Co-maintenance of* S. cerevisiae *types from existing trade-off data.* When known data for the rate-yield and rate-affinity trade-offs estimated from the available literature for *S. cerevisiae* under glucose limitation (see [45, 9] and supplement fig. 5) was deployed within the MSC model, we obtained the results shown in supplement fig. 15. To produce this figure we used a washout rate of the chemostat of $d = 0.375$ per hour and $S_0 = 2500$ mmol glucose/L, cells were measured in grams of biomass with cell density measured in grams of biomass per litre (g/L). The uptake rate of glucose by a cell was then measured in units of mmol glucose per gram of biomass per hour. We assumed for the purposes of this model that glucose uptake rates given in [45] in fact represent maximal uptakes rates, a working assumption not necessarily true in practice but necessary to permit us to formulate the data shown in supplement fig. 5 within the framework of the MSC model.

A comparison between the empirical data and how it has been processed for use within the MSC model is given in supplement fig. 15(a), with the resulting fitness landscape illustrated in the same figure (rightmost plot). Given these parameter values and prior empirical data, the MSC model predicts that co-maintenance regions of both fast and efficient metabolic strategies can be found in the chemostat provided the mutation rate lies within a region illustrated in supplement fig. 15(b). Although this is a narrow region, based on the discussion of the previous section we anticipate that the inclusion of more biological detail into the model would widen the region of mutation rates for which this co-maintenance property is found.

6.4.2. *Are observed mutation rates consistent with the co-maintenance principle?* Using our most realistically parameterised model, namely the one illustrated in §6.4.1 above, from supplement fig. 15(b) we argue that mutation rates at, or somewhat above an approximate threshold of $\sim 10^{-3.5} = 0.0003$ per genome per hour would be sufficient to see the maintenance of both quasi-species. Assuming such an approximate figure

doi:10.1038/nature09905

supplement fig. 15 – (a) The rate-yield and rate-affinity trade-off data estimated for *S. cerevisiae* shown in supplement fig. 5 (repeated here as red circles) were interpolated and fitted to a nonlinear model for deployment within the MSC equation. The result is a fitness landscape with two peaks at high sugar concentrations (shown in the rightmost plot). (b) The resulting steady-state solutions of the MSC equation (1) possess a region of co-maintenance of different quasispecies at the appropriate mutation rates, the model also exhibits maintenance of the fit and the flat at low and high mutation rates, respectively. A superposition of normal distributions has been fitted to the type frequency histogram to highlight each quasispecies.



to be representative of more complex situations, and indeed of prokaryotes, we now ask whether observed mutation rates are approximately of the correct order of magnitude to permit co-maintenance in practise. From supplement fig. 15(b) it is seen that co-maintenance may be possible even as the mutation rate passes beyond this threshold and so we ask whether it is likely that the experimental organism (*E.coli* in the case of [21]) with the enigmatic recovery of mutiple types in a chemostat has a mutation rate that is sufficiently high. Assuming *E. coli* exhibits about three generations per hour in a laboratory microcosm, the critical mutation rate for such an organism would be of the order of $0.0003/3 = 0.0001$ per genome per generation. Most microorganisms, *E. coli* included, have a per genome mutation rate that is about an order of magnitude higher than this at about 0.003 per genome per generation [7]. Only of a fraction of all the possible mutations are likely to be relevant to the trade-offs, but if more than $0.0001/0.003 = 3.33\%$ of the *E. coli* genome can mutate and affect the relevant metabolism and physiology, then observed mutation rates are consistent with our co-maintenance principle.

Any adjustment of these assumptions may affect this conclusion. For example, if we supposed that a bacterial strain had an order-of-magnitude higher mutation rate, for example owing to stress or hypermutation (as employed in [21]) with a genomic mutation rate of 0.03 per genome per generation, then just 0.3% of the

genome would need to be involved. Similarly, if we permitted for each observable DNA-based mutation there to be some rate of epimutation, again the fraction of sites involved in the trade-offs could be less than 3% for the principle to apply.

6.5. **Extension: asymmetric mutation matrices.** Our analysis of the equilibria of the MSC equation reveals that coexisting quasispecies can be maintained by negative-frequency dependence (NFD). Indeed, the property of NFD follows *a fortiori* if a locally stable equilibrium of equation (1) possesses coexisting quasispecies clouds, although the ecological reasoning that underpins the presence of NFD in this model will not be discussed in any detail until §8 in this supplement.

If, for the moment, we accept the assumption of a stable, coexistence steady-state and therefore also the presence of NFD in the MSC model, it follows that changing the mutational connectivity of types by altering the mutation matrix M need not destroy the coexistence state and its local stability can be preserved under such an alteration. As this will also maintain NFD, we claim that NFD may also be stable to the loss of mutational connections between clouds. However, it is also true by the construction of the MSC model that there must also be *large* changes to the connectivity encoded by M that will remove the quasispecies clouds entirely. For example, removing all mutational connections is equivalent to setting $\epsilon = 0$ and this will certainly be sufficient to see the loss of coexistence.

A particularly important issue here concerns the bifurcation structures presented in Propositions 1 and 2 and how they change when we alter the mutation matrix M so that mutations between types are permitted to occur asymmetrically between types. All the computations so far have utilised the mutation structure illustrated in supplement fig. 2 extended to the appropriate number of types, it is therefore important that we consider such asymmetries. The prior assumption of symmetry was invoked merely to shorten the proofs of Propositions 1 and 2, one possible extension to cover certain asymmetric cases now follows.

Proposition 3. *Suppose that (A1-A7) apply and that M is a symmetric, stochastic, irreducible $n \times n$ matrix and $\mathcal{A}$ is any $n \times n$ matrix. Consider the following asymmetric perturbation of (7):*

$$0 = \epsilon(M + \mathcal{A} - I)\boldsymbol{b} + (\boldsymbol{G}(S) - d)\boldsymbol{b}, \tag{17a}$$

$$0 = d(S_0 - S) - \langle \boldsymbol{U}(S), \boldsymbol{b} \rangle . \tag{17b}$$

For any strictly positive solution $(\boldsymbol{b}^*, S^*)$ of (17) defined when $\mathcal{A} = 0_{n\times n}$, there is an $\alpha > 0$ and a smoothly parameterised family of solutions of (17), $(\boldsymbol{b}(\mathcal{A}), S(\mathcal{A}))$, defined for all $\|\mathcal{A}\| < \alpha$ such that $(\boldsymbol{b}(0_{n\times n}), S(0_{n\times n})) = (\boldsymbol{b}^*, S^*)$.

Proof. Write the steady-state equation (17) using a smooth mapping in the form $F(\boldsymbol{b}, S; \mathcal{A}) = \boldsymbol{0}$ where we have highlighted the fact that the perturbation of the mutation matrix M, namely $\mathcal{A}$, is a parameter. Consider the linearisation of $F(\boldsymbol{b}, S; \mathcal{A})$ with respect to $(\boldsymbol{b}, S)$ about any strictly positive non-trivial solution that, because of Propositions 1 and 2, we know exists when $\mathcal{A} = 0_{n\times n}$ and that we now denote $(\boldsymbol{b}^*, S^*)$ for convenience:

$$d_{(\boldsymbol{b},S)}F(\boldsymbol{b}^*, S^*; 0_{n\times n}) = \begin{bmatrix} \epsilon(M - I) + \mathrm{diag}(\boldsymbol{G}(S^*)) - dI & \boldsymbol{G}'(S^*)\boldsymbol{b}^* \\ -\boldsymbol{U}(S^*)^T & -(d + \langle \boldsymbol{U}'(S^*), \boldsymbol{b}^* \rangle) \end{bmatrix}.$$

We now ask, can there be a vector $(\boldsymbol{k}, \sigma) \in \mathbb{R}^{n+1}$ such that $d_{(\boldsymbol{b},S)}F(\boldsymbol{b}^*, S^*; 0_{n\times n})[(\boldsymbol{k}, \sigma)] = (\boldsymbol{0}, 0)$? Writing this linear equation in terms of its two components we require

$$(\epsilon(M - I) + \mathrm{diag}(\boldsymbol{G}(S^*)) - dI)\boldsymbol{k} + \sigma\boldsymbol{G}'(S^*)\boldsymbol{b}^* = \boldsymbol{0} \tag{18a}$$

$$-\langle \boldsymbol{U}(S^*), \boldsymbol{k} \rangle - \sigma(d + \langle \boldsymbol{U}'(S^*), \boldsymbol{b}^* \rangle = 0, \tag{18b}$$

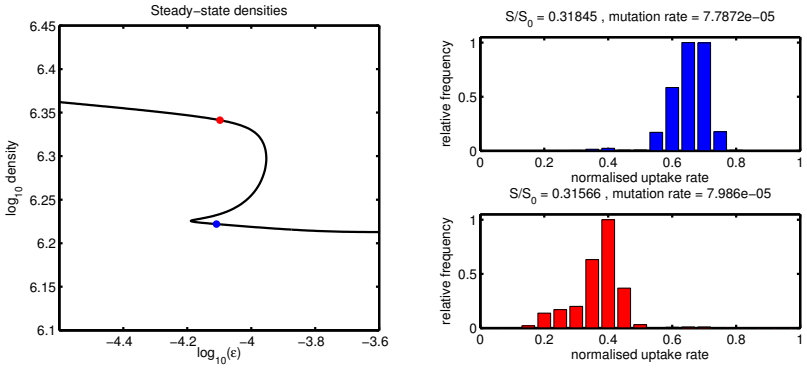
where $\boldsymbol{G}'(S)$ denotes the derivative of $\boldsymbol{G}(S)$ with respect to S, this is a strictly positive vector by earlier assumptions. Taking the inner product of (18a) with the strictly positive vector $\boldsymbol{b}^*$ and using the symmetry of M (we assumed $M = M^T$) we find $\sigma \cdot \langle \boldsymbol{b}^*, \boldsymbol{G}'(S^*)\boldsymbol{b}^* \rangle = 0$ from where $\sigma = 0$ must follow. From equation (18a) we then deduce that $\boldsymbol{k} = \theta \cdot \boldsymbol{b}^*$ for some real θ so that $\langle \boldsymbol{U}(S^*), \theta \cdot \boldsymbol{b}^* \rangle = 0$ must hold by (18b), it therefore follows that $\theta = 0$.

As $(\boldsymbol{k}, \sigma)$ must be the zero vector, the derivative $d_{(\boldsymbol{b},S)}F(\boldsymbol{b}^*, S^*; 0_{n\times n})$ is an isomorphism and so we can apply the implicit function theorem to deduce the existence of a locally-defined path of solutions of (17) that depends smoothly on $\mathcal{A}$:

$$F(\boldsymbol{b}(\mathcal{A}), S(\mathcal{A}); \mathcal{A}) \equiv \mathbf{0} \qquad \text{such that} \qquad \boldsymbol{b}(0_{n\times n}) = \boldsymbol{b}^*, S(0_{n\times n}) = S^*.$$

According to the implicit function theorem, the path $(\boldsymbol{b}(\mathcal{A}), S(\mathcal{A}))$ is smoothly-defined on some neighbourhood of the matrix $0_{n\times n}$, as claimed. □

supplement fig. 16 – Using a reducible mutation matrix in (1) can lead to hysteresis: increasing the mutation rate creates two fold bifurcations and three steady-states, of which two are asymptotically stable and one is unstable. Here, the fit types (red) give way to types on flatter peaks (blue) as the mutation rate increases.



Proposition 3 states that whenever we find a steady-state solution of (1), introducing asymmetric mutation rates between different cell types will not change the structure of the steady-state discontinuously. So, if we find the co-maintenance of fit and flat cell types in some system, that co-maintenance property will persist even if we alter the mutation matrices to be asymmetric. However, we re-iterate that there are certain symmetric perturbations of the mutation matrix M that can destroy coexistence and, as a result, there are also asymmetric perturbations that will have the same effect.

Remark 3. *Do note that the set of restrictions (A1-A7) we have imposed on the MSC model must be taken as sufficient conditions leading to the coexistence hypothesis at the appropriate mutation rates. They are not necessary conditions, nor do we claim that all models of the form (1) have such a coexistence region. Indeed, modifications of our underlying assumptions could result in the destruction of the coexistence region, in particular the existence of more than one steady-state can lead to hysteresis in the bifurcation structure (see supplement fig. 16) and under these conditions our coexistence hypothesis, as with other quasispecies models* [37]*, may well be false.*

7. An individual-based, stochastic model

In order to test the predictions obtained using the MSC model (3) and presented in the previous section, we now develop an individual-based, stochastic chemostat model into which we can encode rate-yield and rate-affinity trade-offs within single cells in competition for one limiting resource. Rather than develop a version of the MSC model as a stochastic ODE, we quite purposefully create a rule-based system that we know *cannot* be well-described in every detail by the MSC equations. For example, while the mutation rate was defined *per unit time* in the MSC model, mutations will be assumed to occur *per division* for the model that we describe below. Our rationale for this is the following. If the predictions of the previous section are to have any broad-scale relevance, we ought to be able to see those predictions borne out in quite different theoretical model systems.

The stochastic model, completely absent of any spatial structure, is encoded by the following rules that are evaluated each *computational generation*:

(SM1). There are at most n different possible cell types where each type is labelled by a number i between 1 and n. A cell of type i has that number of surface transporters for the resource molecules and each transporter can only contain one molecule at a time. As a result, the per unit time maximal uptake of resource by each cell, its $V^{\max}$, coincides with its type.

(SM3). Free resource molecules, finite in number (denoted S) are associated and bound randomly to each of the cells' multiple transporters such that the probability of binding is $S/(K_i + S)$ per transporter per unit time; here, the phenotype K_i is a half-saturation constant associated to each cell type. Bound transporters port molecules immediately across the cell membrane which are then held within the cell at no energy cost.

(SM3). The cell is assumed to convert internally held resource into 'protein' and when sufficient protein has accumulated fission is triggered, at this step a cell of type i is assumed to require at least e_i molecules of resource before cell division can occur.

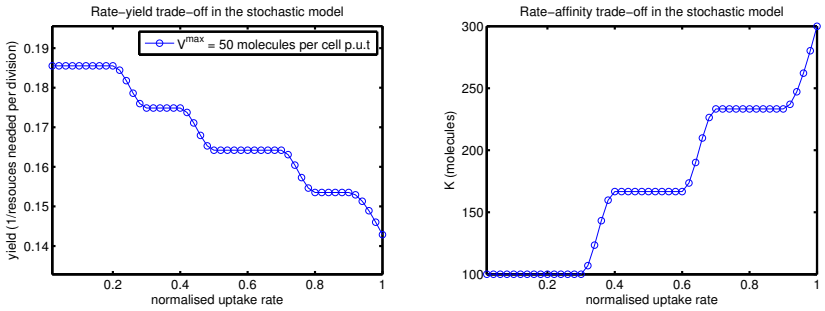
(SM4). *Mutation rate is measured per cell division*: when a cell divides, a copy is made with a fixed probability that the cell's offspring has type $i + 1$ or $i - 1$, given a parent cell of type i. Mutations that occur when a parent with types 1 and n divides can only yield offspring of types 2 and $n - 1$, respectively.

(SM5). When a cell divides, the internally held resources of the parent are apportioned symmetrically to create a parent and offspring pair with identical internal states.

(SM6). At each computational generation, d% of all biotic and abiotic matter is lost from the chemostat and resource molecules are supplied to the device at a constant rate.

As i is the maximal uptake rate of cell type i, a rate-yield trade-off can be incorporated into this stochastic model by having e_i increase as a function of i, noting that $1/e_i$ is the cell yield per unit resource. A rate-affinity trade-off is encoded in the stochastic model by having K_i increase as a function of i and the trade-offs used throughout this section, unless otherwise stated, are shown in supplement fig. 17. For the washout parameter we used the value $d = 2\%$, unless stated otherwise, so that the mean length of stay of any particle in the simulated chemostat device is fifty computational generations and the resource supply concentration was chosen to yield a carrying capacity of approximately three thousand cells in steady-state. These values were chosen for the convenience of rapidly simulating the model in Matlab and we chose a number n of fifty types throughout.

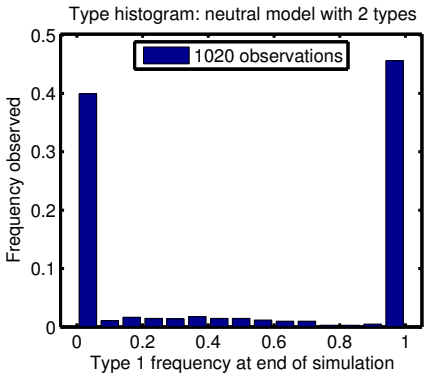
supplement fig. 17 − The rate-yield trade-off and the rate-affinity trade-offs used in the stochastic model: this is analogous to supplement fig. 8(b), adapted to the units of the stochastic model.



7.1. **Neutrality in the stochastic model.** The vectors (c_i) and (K_i) that associate cell types with uptake rate, cell yields and resource affinities S determine the nature of the fitness landscape, along with the resource concentration S. If (c_i) and (K_i) are constant across cell types then the cell fitnesses form a neutral landscape and so their frequencies change through drift alone, although the total density of the population can change in time.

The long-term dynamics in this neutral scenario are illustrated in supplement fig. 18 when there are just two types, so $n = 2$, where the washout rate is $d = 10\%$ and the mutation rate was set to zero for the production of this figure. For these parameter values, one type is eventually lost with probability one and the between-population invariant density of the stochastic model is then a weighted sum of Dirac masses supported on each of the two types. This claim is consistent with the histogram depicted in supplement fig. 18 formed from a finite number of observations.

supplement fig. 18 − The neutral stochastic model with $n = 2$ cell types: in each realisation of the model, the fixation of one of the two types occurs eventually with probability one. Averaging between populations yields an estimate of the invariant density, a weighted sum of two Dirac masses supported at zero and one: $\frac{1}{2}(\delta(\cdot) + \delta(\cdot - 1))$.



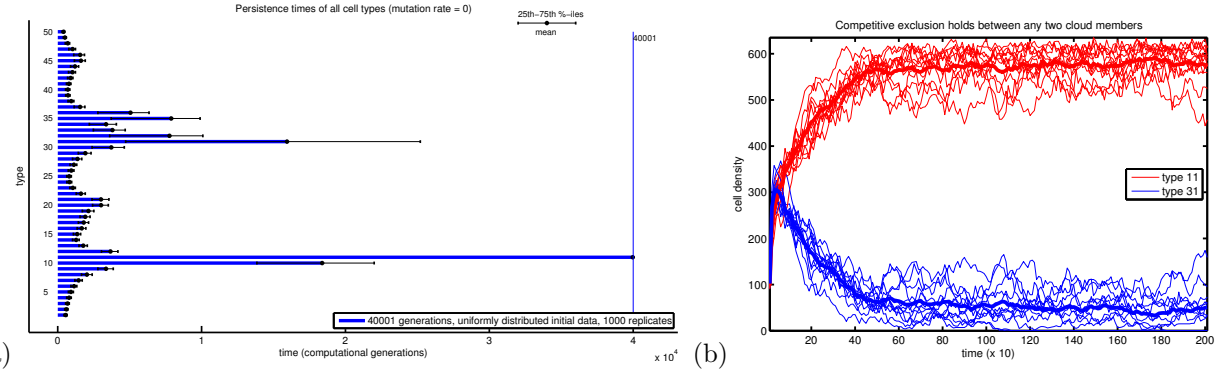
7.2. **Competitive exclusion in the stochastic model.** Consider the non-neutral fitness landscape pictured in supplement fig. 14(a), itself derived from the rate-yield and rate-affinity trade-offs shown in supplement fig. 7. When we previously employed this set of trade-offs in the MSC model we obtained the cell type densities depicted in supplement fig. 14(c). This figure illustrates that there is a wide range of mutation rates for which two quasispecies can coexist with approximate normalised uptake rates of 0.2 and 0.6 units, when taking into

account the 50 cell types of the stochastic model, the mode of the coexisting fit and flat quasispecies should therefore be located near to cell types $10 = 0.2 \times 50$ and $30 = 0.6 \times 50$.

We then ran the stochastic model for 40,001 computational generations with the mutation rate set to zero, as shown in supplement fig. 19(a), and this established type 11 as the fittest under the prevailing model parameters. Furthermore, supplement fig. 19(b) shows the outcome of a direct competition between type 11 and the only other type that also persisted to the end of the computation in 1,000 replicates, namely type 31. Again, type 11 prevailed in all replicates, just ten replicate trajectories are shown as an illustration in supplement fig. 19(b).

These computations verify that the stochastic model is consistent with survival of the fittest when mutation rate is set to zero and indicates that type 11 is the fittest cell type for the parameters used. Based on the analysis and computations of the previous section, we therefore anticipate that a cloud of quasispecies containing type 11 will coexist with different quasispecies at sufficiently high mutation rates.

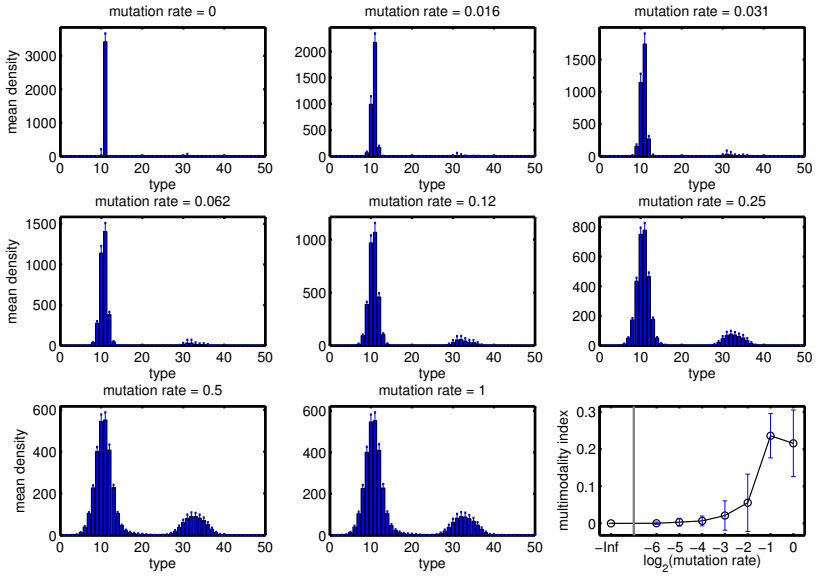
supplement fig. 19 – With selection acting through both rate-yield and rate-affinity trade-offs but zero mutation rate, we anticipate competitive exclusion. (a) This plot of persistence times for each cell type shows that one emerges as the fittest. (After 40,001 computational generations and of 50 possible types, type 11 fixed in 898/1,000 replicates and was the only type never to be lost from the population. It also had the highest density by the end of every simulation; type 31 was the only other type that persisted to the final computational generation of any simulation.) (b) Competition between just two types, 11 and 31, indicates 11 to be the fitter (shown are ten superimposed replicates and their means, each lasting 2,000 computational generations with $d = 10\%$).



7.3. **Co-maintenance of the fit and the flat in the stochastic model.** Having established consistency of the stochastic model with prior neutral and competitive systems under the appropriate model restrictions, in order to estimate equilibrium densities and type frequencies of this model we computed mean type distributions both within and between different populations at different mutation rates. The results, shown in supplement fig. 20 for a range of mutation rates, are based on means taken between two hundred different populations, all differing in the initial distribution of types and where the initial allocation of internal resources was randomly distributed among the 3,000 cells.

It is clear from supplement fig. 20 that a bimodal distribution of cell types is obtained at sufficiently high mutation rates whereas a near-monomorphic distribution in mutation-selection balance is obtained at the lowest mutation rates. We used a *multimodality index* defined as the ratio of total densities found in the two observed quasispecies to further probe the onset of coexistence. As can be seen in supplement fig. 20(bottom-right), before the onset of coexistence the multimodality index resides near zero, however fluctuations in this

supplement fig. 20 − A plot of the long-term densities of cell types measured between populations using the stochastic model for increasing mutation rates and indicating the presence of two quasispecies for sufficiently high mutation rates. The very bottom-right diagram shows the ratio of mean biomass under each cloud and indicates that a critical mutation rate must be exceeded for two clouds to coexist. (Histograms of between-population means of type densities were computed using two hundred randomly generated initial populations; standard errors are indicated.)

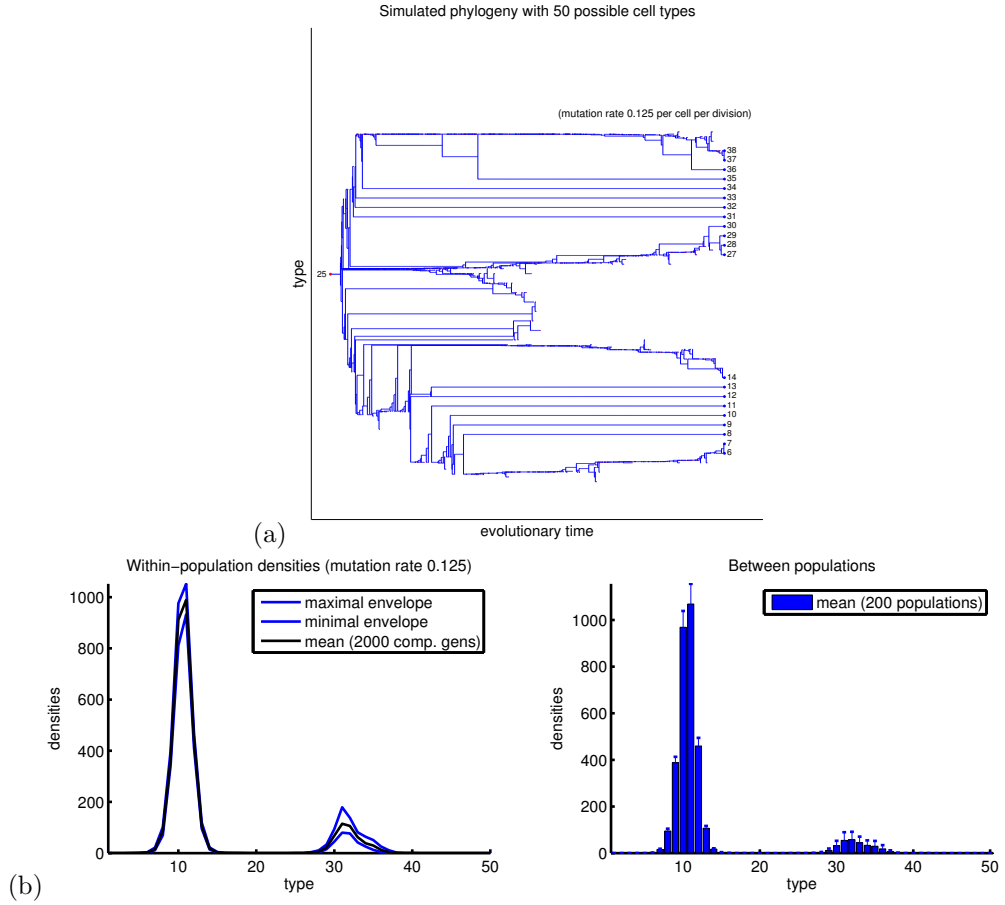


measure become more apparent as the flatter quasispecies persists for larger times as the transition is neared. Once a critical mutation rate is passed, variability of the multimodality index initially reduces but the index itself increases sharply and two quasispecies are co-maintained, one containing cell type 11 and the other type 31.

In order to establish whether the bimodality of between-population means shown in supplement fig. 20 was found within each population, samples were taken of temporal trajectories of the stochastic model for fixed initial conditions with a single seeding type (chosen to lie at the mid-point of type space {1,2,...,50}, namely 25). The individual states of each of the 3,000 seeding cells were again randomly chosen, where the latter value was chosen to be approximately equal to the carrying capacity of the simulation and the mutation probability per division was set to 1/8. A typical result is shown in supplement fig. 21 with similar results obtained for different mutation rates between 1/8 and 1 (data not shown). Diagram (a) of supplement fig. 21 shows a simulated phylogeny obtained after 20,001 computational generations in which the seeding monomorphic population branches into three lineages of which one is eventually purged.

The mean type densities associated with the final 2,000 generations of the single population depicted in supplement fig. 21(a) are shown in supplement fig. 21(b,left), this is a bimodal distribution and satisfies the Hartigan dip test for bimodality ($p < 10^{-14}$). To check that such bimodality was maintained for all large enough times, the *lower and upper envelopes* of the same trajectory are superimposed as blue lines on the diagram and a histogram of type densities taken between 200 populations for the same initial states is shown in supplement fig. 21(b) for comparison. The upper and lower envelopes are defined to be the maximal and minimal values taken by type densities throughout the simulated trajectory and so provide upper and lower bounds on that trajectory. So, if the data ensemble of length $N = 2,000$ type densities on $n = 50$ types, is

supplement fig. 21 – (a) An initially monomorphic population branches into three quasispecies of which one is eventually lost (simulation of 20,001 computational generations). (b) The long-term cell type densities found within the population in (a) (shown as a black line) are compared with long-term averages computed between populations for the same mutation rate (blue histogram). Also shown (blue lines) are the upper and lower envelopes of the final 1,000 computational generations from the phylogeny in (a) illustrating that the bimodality of efficient and inefficient types is maintained within one population for all sufficiently large times.



$\mathcal{E} = (\Delta_t^{(1)}, \Delta_t^{(2)}, ..., \Delta_t^{(n)})_{t=1}^N$ then the upper and lower envelopes are defined to be the vectors

$$\mathrm{upp}(\mathcal{E}) = \left(\max_{1\le t\le N} \Delta_t^{(1)}, \max_{1\le t\le N} \Delta_t^{(2)}, ..., \max_{1\le t\le N} \Delta_t^{(n)} \right)$$

and

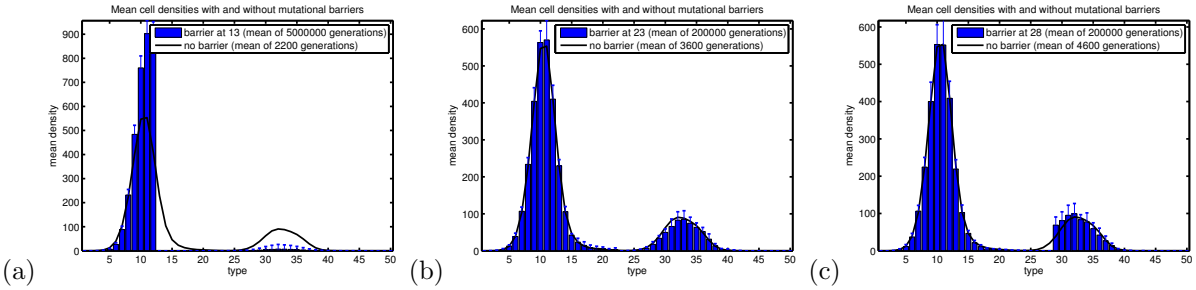
$$\mathrm{low}(\mathcal{E}) = \left(\min_{1\le t\le N} \Delta_t^{(1)}, \min_{1\le t\le N} \Delta_t^{(2)}, ..., \min_{1\le t\le N} \Delta_t^{(n)} \right).$$

Both of these are bimodal, as can be seen in supplement fig. 21(b,left), we therefore deduce that the type distribution contains two clouds or clusters for each of the 2,000 computational generations sampled.

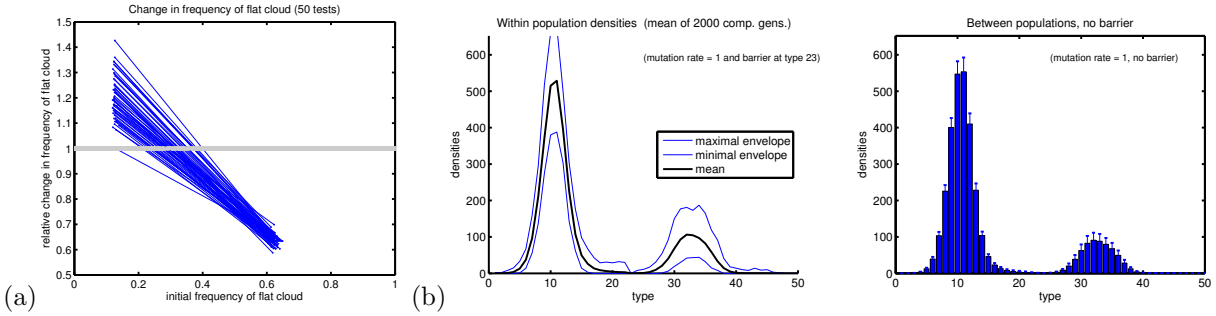
7.4. **Preventing valley-crossing lineages: mutational barriers.** As the phylogeny in supplement fig. 21(a) indicates, the observed bimodality of the long-term type distributions consisting of efficient and inefficient cell types in supplement fig. 21(b) does not appear to be due to lineages traversing fitness valleys from the fit to the flat and so replenishing the flatter quasispecies. In order to test this hypothesis more thoroughly, we performed two further tests.

First, after executing a single simulation of the stochastic model and running it to a near-equilibrium state (based on the mean type distributions shown in supplement fig. 20) we then imposed a *mutational barrier*

supplement fig. 22 – The presence of two quasispecies persists within populations even after the imposition of *mutational barriers* between quasispecies. Comparing the black line which is the between-population type distributions taken from supplement fig. 20 with the resulting within-population distributions obtained after a mutational barrier has been imposed in the stochastic model, diagram (b) shows little change, (a) shows the loss of the flattest cloud and (c) shows a small increase in the density of some of the flat types. (In all three diagrams, the mutation rate has been set equal to its highest possible value, namely 1 per cell per division.)



supplement fig. 23 – (a) Diversity of quasispecies is maintained at all times by negative frequency dependence (fifty tests, mutation rate equals 1) even in the presence of a mutational barrier between those quasispecies (at type 23). (b) This is illustrated by long-term time averages within a single population trajectory and in the lower and upper envelopes of that trajectory (after transients) all of which illustrate the persistence of both quasispecies. For comparison, the right-hand diagram shows between-population averages (taken from supplement fig. 20) with no mutational barrier.



between quasispecies and observed changes in the long-term distribution of type densities that resulted from this. A mutational barrier in a linear mutational topology is a pair of cell types between which the probability of mutation is zero, see supplement fig. 2(b) for an illustration, such a mutational structure is asymmetrical in the sense of §6.5. The consequences of imposing a barrier are shown in supplement fig. 22 where two clouds of types may be found even after the imposition of a barrier, although supplement fig. 22(c) shows that barriers can enhance the densities of one quasispecies and supplement fig. 22(a) shows that one of the quasispecies may be lost after a barrier is introduced.

Second, we tested the stochastic model for negative frequency dependence (NFD) at the level of quasispecies by imposing mutational barriers between clusters and then suppressing the densities of each of the two emergent quasispecies clusters in turn. The resulting change in frequencies of the types in time within each cluster, whether increasing (a relative change in frequency larger than one) or decreasing (a relative change in frequency less than one) is shown in supplement fig. 23(a) for fifty replicates of one initial condition supporting two quasispecies. Our point is that this figure clearly supports the coexistence hypothesis through negative

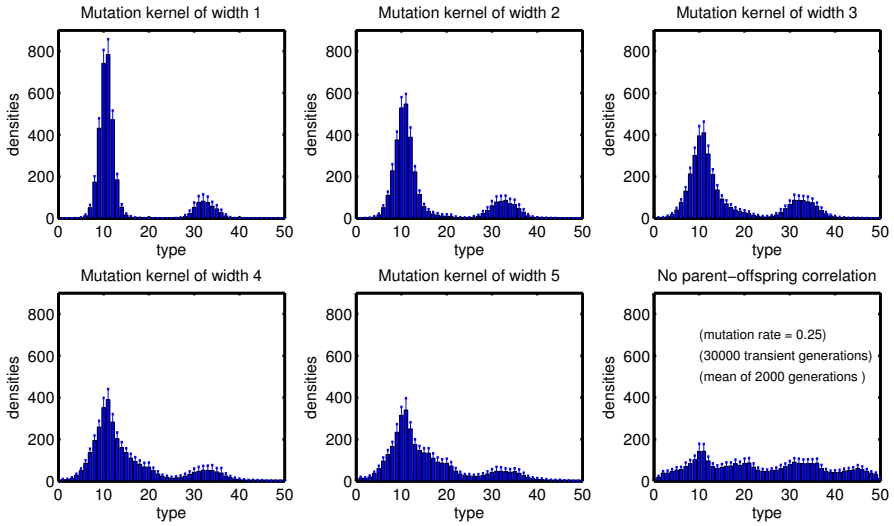
frequency-dependent selection. Moreover, the mean, lower and upper envelopes of the simulated distribution of types of just of one of these replicates is shown in supplement fig. 23(b,left) illustrating that bimodality of the distribution of types was maintained at all times.

7.5. **Robustness to changes in mutation kernel.** Let us momentarily modify the nature of mutations in the stochastic model in the following manner, changing (SM4) above to be the following rule:

(SM4′). When a cell divides, a copy is made but with a fixed probability that the cell's offspring has a different type to its parent. If w is a fixed parameter, the type of the offspring is a randomly chosen, uniformly distributed member of the set $\{i-w,...,i-1,i+1,...,i+w\}$ given a parent cell of type i. Mutations that occur when a parent with types 1 and n divides can only yield offspring of types $\{2,...,w\}$ and $\{n-w,...,n-1\}$, respectively.

When using assumption (SM4′) we say that the *mutation kernel* is the uniform distribution of width w.

supplement fig. 24 – The co-maintenance of two quasispecies is robust to changes in mutational structure, illustrated here as the width w of a uniformly-distributed mutation kernel is increased. In the extreme, the complete absence of correlation between parent and offspring types leads to the loss of both quasispecies (bottom-right figure). (Shown are histograms of cell type densities obtained after a transient period and averaged over a trajectory of 2,000 computational generations.)



Supplement fig. 24 shows the density of types obtained when simulating the stochastic model after making the choice of $w = 1,2,3,4$ and 5, thus increasing the width of the mutation kernel. The same figure also illustrates the density of types when choosing each mutated offspring type randomly from the set $\{1,2,...,n\}$, in this case we say that there is no *parent-offspring correlation* of types. The purpose of supplement fig. 24 is to show that a population structure consisting of co-maintained quasispecies is robust to changes in mutation structure but that a correlation between parent and offspring must be present in order that the quasispecies are stably maintained.

8. Co-maintenance: a toy model

To understand what might be the underlying cause of the maintenance of different quasispecies by negative-frequency dependent selection, consider the following reasoning. Returning to (1), note that the model is subject to both frequency and density-dependent selection. However, when the mutation rate ϵ is small a

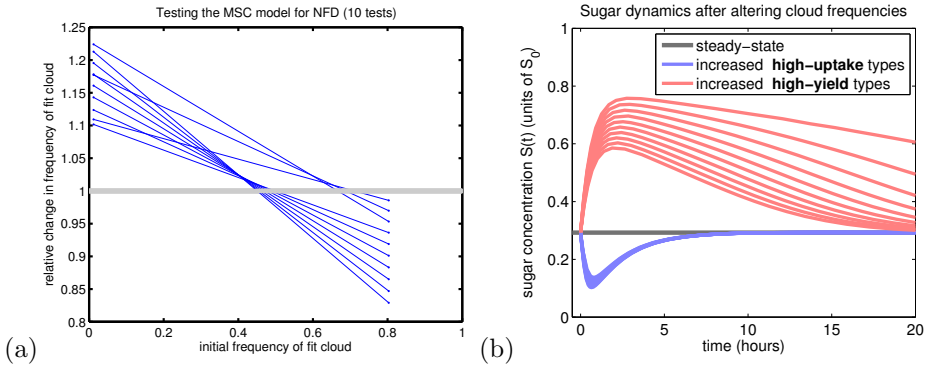
mathematical argument based on the implicit function theorem (and used in Proposition 1) tells us that one type, the fittest, dominates and all others are small in frequency. More precisely, up to a re-ordering of coordinates the steady-state vector $\boldsymbol{f} = (f_1, f_2, ..., f_n)$ of (1) is in a state of *mutation-selection balance* in the sense that there is an ϵ-independent constant $\kappa > 0$ such that

$$1 - \kappa\epsilon < f_1 \leq 1 \qquad \text{and} \qquad 0 \leq f_j < \kappa\epsilon \text{ for } j \geq 2,$$

for all small enough ϵ.

Once a critical mutation rate is surpassed, a second quasispecies can be supported even in the presence of a mutational barrier because equation (1a) tells us that one cloud can mediate the dynamics of the other through their common environment, S. To see this, imagine the thought experiment whereby we run the chemostat to steady-state conditions and isolate two quasispecies, one fit and one flat, and we then perform the following exchange of cells. (1) First *remove* one cell from the flat cloud and then *add* one cell to the fit cloud and observe the dynamics of the system, (2) in a second population remove one cell from the fit cloud and add one cell to the flat. How does the system return to its equilibrium state?

supplement fig. 25 – (a) The standard test of tracking the response in frequencies of cell types demonstrates that two quasispecies are supported by negative frequency-dependent selection in the MSC model (1). (b) Response of the resource concentration after increasing the frequency of fit types (left) and flat types (right): the former case leads to a momentary decrease in sugar concentration, the latter to a momentary increase. This perturbation of the environment sets the conditions for both quasispecies and the resource concentrations to return to equilibrium (shown as red lines). (The trade-offs used in (a) and (b) are illustrated in supplement fig. 8(b).)



The answer lies in the fact that these two operations momentarily affect the dynamics of the sugar in the environment in different ways, as depicted in supplement fig. 25(b). In the case where the frequency of fit types is increased, note that these types are fitter because of their higher uptake rates, hence the resource concentration reduces from its previous steady-state value. However, at such lower resource concentrations it is the 'flatter cells' with their higher efficiency that are advantaged by this change which then acts to increase and so re-set sugar levels back towards their equilibrium value. The case where the flat types are increased in frequency behaves entirely analogously, but by first increasing sugar levels (see supplement fig. 25(b)).

When this thought experiment is implemented in simulations of (1), the negative frequency-dependent dynamics that result are shown in supplement fig. 25(a) whereby the fit and flat types are both returned to *intermediate* frequencies when rare or common.

8.1. **A toy population genetical model.** Having established the cause of environmentally-mediated, negative frequency-dependent selection, we now continue our intuitive explanation regarding mechanisms enabling coexistence of the fittest and the flattest by developing a toy model. So, consider two quasispecies each labeled a and A formed from two distinct clusters of types. The fitness of quasispecies a is denoted by G_a while the fitness of quasispecies A is denoted by G_A. The relative frequency of the set of all types in quasispecies A is given by f_A and we assume that negative frequency dependence is exhibited by this model in the sense that:

$$\frac{G_A}{G_a} = r_0 - f_A(r_0 - r_1) \tag{19}$$

where $r_0 > r_1 > 1$ are fixed constants and, by assumption, $0 \leq f_A \leq 1$. These inequalities imply that $G_A > G_a$, meaning A is the *fitter* of the two quasispecies.

Next, we assume that there are three types of mutational event that occur between quasispecies that arise due to the mutation of different cell types within them:

(ME1). $a \to a$ or $A \to A$: these cases correspond to mutational events that keep offspring cell types within the same quasispecies as their parent type (this is an intra-quasispecies mutation);

(ME2). $a \to A$ or $A \to a$: these cases correspond to mutational events that move offspring cell types to different quasispecies from their parent types (an inter-quasispecies mutation);

(ME3). $a \to$ dead or $A \to$dead: here 'dead' represents a highly-deleterious mutation that produces a non-viable offspring cell type neither in quasispecies a nor A.

To discount mutation-selection equilibrium we assume that mutational events of type (ME2) do not occur, moreover the fitness effects of mutational events of type (ME1) are completely ignored in this simple model.

We now suppose that the loss of cell types from a and A due to a mutational event (ME3) reduces the fitness of a and A in the sense that

$$w_a = G_a - \mu_a \qquad \text{and} \qquad w_A = G_A - \mu_A, \tag{20}$$

where μ_a and $\mu_A > 0$; thus w_a and w_A are the fitnesses of each quasispecies in the presence of a constant per-type, per-cell-division mutation rate. We shall assume that $\mu_A > \mu_a$, expressing the idea that a is the flatter quasispecies and the more mutationally robust, it follows that mutations that occur to cell types within quasispecies a are less likely to be non-viable. We now assume for reasons of simplicity that there is an α such that

$$\alpha = \mu_a/\mu_A \tag{21}$$

with $0 < \alpha < 1$ from our construction.

Now, the principle of *survival of the fittest* is relevant when $w_A > w_a$ and a little algebra shows that this occurs when the following inequality applies: $\frac{G_a(r_1-1)}{1-\alpha} > \mu_A$. On the other hand, *survival of the flattest* is relevant when $w_A < w_a$ and, again, a little algebra shows that this occurs if $\frac{G_a(r_0-1)}{1-\alpha} < \mu_A$. Thus, our toy model so-constructed is consistent with the survival of the fittest and flattest at low and high mutation rates, respectively. Coexistence of the fittest and the flattest quasispecies is relevant when the equality $w_A = w_a$ holds, based on our construction we find that this occurs when mutation rates take on *intermediate* values in such a way that

$$\frac{G_a(r_1 - 1)}{1 - \alpha} < \mu_A < \frac{G_a(r_0 - 1)}{1 - \alpha};$$

see supplement fig. 26(a) for an illustration.

This co-maintenance principle requires two core properties, namely $G_A > G_a, \mu_A > \mu_a$ and that negative frequency-dependent selection acts between the two quasispecies. If we were to assume that mutations to

types in quasispecies a have the same probability of being non-viable as mutations to types in quasispecies A, in the sense that $\mu_A = \mu_a$, it would then follow that A is a fitter quasispecies than a. Similarly, suppose there were an absence of negative frequency dependence, $r_0 = r_1$, in this instance (19) would become

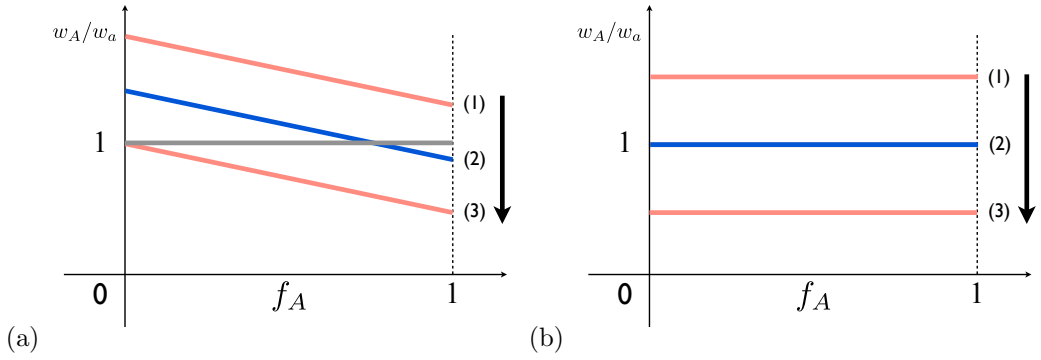
$$G_A = r_0 G_a. \tag{22}$$

Allowing for $\mu_a = \alpha\mu_A$ with $0 < \alpha < 1$ and substituting (22) into (20) we obtain

$$\frac{w_A}{w_a} = \frac{G_a r_0 - \mu_A}{G_a - \alpha\mu_A}. \tag{23}$$

From (23) we deduce that there could only be a single value of μ_A for which coexistence of the fittest and the flattest could occur, namely for $\mu_A = \frac{G_a(r_0-1)}{1-\alpha}$and thus, in the absence of these two core properties, it is only possible to derive the non-generic or non-robust form of coexistence depicted in supplement fig. 26(b).

supplement fig. 26 – The relative fitness of quasispecies 'A' as a function of the frequency of all the types contained within it, f_A. (a) The toy population genetical models predicts that the transition of the fittest quasispecies 'A' (see label (1)), to the flattest 'a' (see label (3)) passes through a regime of co-maintenance (see label 2) in the direction indicated by the arrow as the mutation rate is increased. (b) Only a non-robust form of co-maintenance (label (2)) is possible in the absence of negative frequency-dependence.



References

[1] A. Ambrosetti and G. Prodi, *A Primer of Nonlinear Analysis*, Cambridge University Press, 1993.

[2] T. Bauchop and S. R. Elsden, *The growth of micro-organisms in relation to their energy supply*, J. Gen. Microbiol., 23 (1960), pp. 457–469.

[3] B. Buffoni and J. Toland, *Analytic Theory of Global Bifurcation*, Princeton Press, 2003.

[4] F. S. ChapinIII, *The mineral nutrition of wild plants*, Annual Review of Ecology and Systematics, 11 (1980), pp. 233 –260.

[5] R. A. Cooper, *Metabolism of methylglyoxal in microorganisms.*, Annu Rev Microbiol, 38 (1984), pp. 49–68.

[6] E. J. Doedel, R. Paffenroth, A. Champneys, T. Fairgrieve, Y. Kuznetsov, B. Sandstede, and X. Wang, *Auto 2000: Continuation and bifurcation software for ordinary differential equations (with homcont)*. Technical Report, Caltech; see also `http://indy.cs.concordia.ca/auto/`, 2001.

[7] J. W. Drake, *A constant rate of spontaneous mutation in DNA-based microbes.*, Proc Natl Acad Sci, 88 (1991), pp. 7160–7164.

[8] D. E. Dykhuizen and D. L. Hartl, *Selection in chemostats.*, Microbiol Rev, 47 (1983), pp. 150–168.

[9] K. Elbing, C. Larsson, R. M. Bill, E. Albers, J. L. Snoep, E. Boles, S. Hohmann, and L. Gustafsson, *Role of hexose transport in control of glycolytic flux in* saccharomyces cerevisiae, Appl. Environ. Microbiol., 70 (2004), pp. 5323–5330.

[10] S. E. Forde, R. E. Beardmore, I. Gudelj, S. S. Arkin, J. N. Thompson, and L. D. Hurst, *Understanding the limits to the generalizability of experimental evolutionary models*, Nature, 455 (2008), pp. 220–223.

[11] J. B. Grace and D. Tilman, *Perspectives on Plant Competition*, Blackwell, Oxford., 1990.

[12] I. Gudelj, R. E. Beardmore, S. S. Arkin, and R. C. MacLean, *Constraints on microbial metabolism drive evolutionary diversification in homogeneous environments.*, J Evol Biol, 20 (2007), pp. 1882–1889.

[13] R. Heinrich, F. Montero, E. Klipp, T. G. Waddell, and E. Melendez-Hevia, *Theoretical approaches to the evolutionary optimization of glycolysis: thermodynamic and kinetic constraints.*, Eur J Biochem, 243 (1997), pp. 191–201.

[14] J. Hofbauer, *The selection mutation equation*, J. Mathematical Biology, (1985), pp. 41–53.

[15] J. Ihssen and T. Egli, *Specific growth rate and not cell density controls the general stress response in Escherichia coli*, Microbiology, 150 (2004), pp. 1637–1648.

[16] H. B. Keller, *Numerical Solution of Bifurcation and Nonlinear Eigenvalue Problems*, Academic Press, New York, ed. P. Rabinowitz, 1977.

[17] J.-U. Kreft, *Biofilms promote altruism.*, Microbiology, 150 (2004), pp. 2751–2760.

[18] D. Lipson, R. Monson, S. Schmidt, and M. Weintraub, *The trade-off between growth rate and yield in microbial communities and the consequences for under-snow soil respiration in a high elevation coniferous forest*, Biogeochemistry, 95 (2009), pp. 23–35. 10.1007/s10533-008-9252-1.

[19] R. C. MacLean, *The tragedy of the commons in microbial populations: insights from theoretical, comparative and experimental studies*, Heredity, 100 (2007), pp. 471–477.

[20] R. C. MacLean and I. Gudelj, *Resource competition and social conflict in experimental populations of yeast*, Nature, 411 (2006), pp. 498–501.

[21] R. Maharjan, S. Seeto, L. Notley-McRobb, and T. Ferenci, *Clonal adaptive radiation in a constant environment*, Science, 313 (2006), pp. 514–517.

[22] G. Martin and S. Gandon, *Lethal mutagenesis and evolutionary epidemiology*, Philos Trans R Soc Lond B, 365 (2010), pp. 1953–1963.

[23] A. Merico, P. Sulo, J. Piskur, and C. Compagno, *Fermentative lifestyle in yeasts belonging to the saccharomyces complex.*, FEBS J, 274 (2007), pp. 976–989.

[24] D. Molenaar, R. van Berlo, D. de Ridder, and B. Teusink, *Shifts in growth strategies reflect tradeoffs in cellular economics.*, Mol Syst Biol, 5 (2009), p. 323.

[25] K. Moller, B. Christensen, J. Forster, J. Piskur, J. Nielsen, and L. Olsson, *Aerobic glucose metabolism of saccharomyces kluyveri: growth, metabolite production, and quantification of metabolic fluxes.*, Biotechnol Bioeng, 77 (2002), pp. 186–193.

[26] J. Monod, *Recherches sur la croissance des cultures bactériennes*, Hermann et cie, Paris, first ed., 1942.

[27] M. Muir, L. Williams, and T. Ferenci, *Influence of transport energization on the growth yield of escherichia coli.*, J Bacteriol, 163 (1985), pp. 1237–1242.

[28] R. H. Müller and W. Babel, *Oxidative capacity determines the growth rate with acetobacter methanolicus*, Acta biotechnologica, 13 (1993), pp. 3–11.

[29] O. M. Neijssel and D. W. Tempest, *The role of energy-spilling reactions in the growth of klebsiella aerogenes nctc 418 in aerobic chemostat culture.*, Arch Microbiol, 110 (1976), pp. 305–311.

[30] M. Novak, T. Pfeiffer, R. E. Lenski, U. Sauer, and S. Bonhoeffer, *Experimental tests for an evolutionary trade-off between growth rate and yield in E. coli.*, Am Nat, 168 (2006), pp. 242–251.

[31] A. Novick and M. Weiner, *Enzyme induction as an all-or-none phenomenon.*, Proc Natl Acad Sci U S A, 43 (1957), pp. 553–566.

[32] T. Pfeiffer and S. Bonhoeffer, *Evolutionary consequences of tradeoffs between yield and rate of atp production*, Z. Phys. Chem., 216 (2002), pp. 51—63.

[33] ———, *Evolution of crossfeeding in microbial populations*, Am. Nat., 163 (2004), pp. E126–E135.

[34] T. Pfeiffer, S. Schuster, and S. Bonhoeffer, *Cooperation and competition in the evolution of ATP-producing pathways*, Science, 292 (2001), pp. 504–507.

[35] E. Postma, C. Verduyn, W. A. Scheffers, and J. P. Van Dijken, *Enzymic analysis of the crabtree effect in glucose-limited chemostat cultures of saccharomyces cerevisiae.*, Appl Environ Microbiol, 55 (1989), pp. 468–477.

[36] J. B. Russell and G. M. Cook, *Energetics of bacterial growth: balance of anabolic and catabolic reactions.*, Microbiol Rev, 59 (1995), pp. 48–62.

[37] J. Sardanye, S. F. Elena, and R. V. Sole, *Simple quasispecies models for the survival-of-the-flattest effect: The role of space*, J. Theoretical Biology, 250 (2008), pp. 560—568.

[38] R. Schuetz, L. Kuepfer, and U. Sauer, *Systematic evaluation of objective functions for predicting intracellular fluxes in escherichia coli.*, Mol Syst Biol, 3 (2007), p. 119.

[39] H. L. Smith and P. Waltman, *The Theory of the Chemostat*, Cambridge University Press, 2003.

[40] A. Stephani, J. C. Nuno, and R. Heinrich, *Optimal stoichiometric designs of atp-producing systems as determined by an evolutionary algorithm.*, J Theor Biol, 199 (1999), pp. 45–61.

[41] M. Stephenson, *Bacterial metabolism*, Monographs on biochemistry, Longmans Green and Co., London, 1949.

[42] J. Thompson, *The Coevolutionary Process*, University of Chicago Press, 1994.

[43] J.-W. Veening, W. K. Smits, and O. P. Kuipers, *Bistability, epigenetics, and bet-hedging in bacteria.*, Annu Rev Microbiol, 62 (2008), pp. 193–210.

[44] H. Wagner, E. Baake, and T. Gerisch, *Ising quantum chain and sequence evolution*, J. Stat. Phys, 92 (1999), pp. 1017–52.

[45] R. A. Weusthuis, J. T. Pronk, P. J. van den Broek, and J. P. van Dijken, *Chemostat cultivation as a tool for studies on sugar transport in yeasts.*, Microbiol Rev, 58 (1994), pp. 616–630.

[46] C. O. Wilke, J. L. Wang, C. Ofria, R. E. Lenski, and C. Adami, *Evolution of digital organisms at high mutation rates leads to survival of the flattest*, Nature, 412 (2001), pp. 331–333.

[47] K. W. Wirtz, *A genetic model for changes in microbial kinetic coefficients*, J. Biotechnol., 97 (2002), pp. 147–162.

[48] W. W. Wong, L. M. Tran, and J. C. Liao, *A hidden square-root boundary between growth rate and biomass yield.*, Biotechnol Bioeng, 102 (2009), pp. 73–80.

[49] E. Zeidler, *Nonlinear Functional Analysis and its Applications I: Fixed-Point Theorems*, Springer-Verlag, 1986.

[50] S. Zhong, S. P. Miller, D. E. Dykhuizen, and A. M. Dean, *Transcription, translation, and the evolution of specialists and generalists.*, Mol Biol Evol, 26 (2009), pp. 2661–2678.

Appendix A. Model parameters

The parameter values employed for the MSC model for all figures apart from supplement fig. 15 were adapted from ones used previously by the authors [10] in the study of *E. coli B* under glucose limitation. Cell concentration is given in cells/ml, values of $V^{\max}$ and resource uptake by each cell are given in μg per cell per unit time, the chemostat washout parameter d is measured per hour so growth and uptake rates are also measured per hour. Cell yield is given in cells per μg of resource and resource concentrations for S_0 are stated in $\mu g/ml$, S is then stated as number between 0 and 1 as a fraction of S_0. The precise parameter values used cannot be presented succinctly due to the trade-offs used, but throughout d was taken between 0.2 and $1.4h^{-1}$, S_0 was $1\mu g/ml$, K between 0.01 and $0.1\mu g/ml$, $V^{\max} \approx 10^{-6}\mu g$/cell/h with cell yield approximately 10^6cells/μg. As a result, growth rates of all cell types range between 0 and 2 per hour, with most types around one division per hour. This gives equilibrium cell densities in the culture vessel of the chemostat between 10^6 and 10^8 cells per ml, with lower densities generally found at higher mutation rates.

Different physical units were used for the implementation of the stochastic model. The large concentration of cells found in chemostat devices prevents the use of parameters in the same units as the MSC model because of constraints on computational resources. So, parameters were used to yield a carrying capacity of 3,000 cells, a number that could be simulated rapidly to generate enough replicates to ensure reliable statistics. Available resources are measured as a number of molecules in the chemostat device but volume is not specified, cells are given as a number, the unit of time is arbitrary and measured according to iteration number of the model (we call this a *computational generation*). Unless stated otherwise, washout is 2% per iteration yielding an expected time to loss from the device of 50 computational generations for every cell and resource molecule.

Cell growth is modelled by having each cell accumulate a minimal amount of *protein* in order to divide and this provides an efficiency measure because it is assumed that each resource molecule can be converted into a fixed and constant amount of protein. All cells have the same *mutation rate* which is the probability of a change in type per division. A mutation changes the type of a cell by a value of at most w to ensure a positive correlation between the type of a parent cell and the type of the offsprings derived from it. This assumption

was purposefully invoked to prevent mutations in a parent cell of one quasispecies or fitness peak that could yield offspring cells in a different quasispecies or fitness peak.

Supplement figure 15(b) is based on known *Saccharomyces cerevisiae* data and was produced using the rate-yield and rate-affinity data provided in supplement fig. 15(a) where a chemostat washout rate of $d = 0.375$ per hour was used, with a glucose supply rate of $S_0 = 2500$ mmol/L. Glucose uptake here is measured in mmol per gram of biomass per hour with a maximum possible value of $V^{\max} = 20$ mmol/g biomass/h, this is the maximum value for which data is available in supplement fig. 5(c). As can be seen in supplement fig. 15(a), this yields a prediction of between 70g and 140g of biomass per litre.