1 Model Description

Conceptually, the Flask model simulates the ecology and evolution of microbial communities. References to other work describing the Flask model and the typical dynamics of the flask ecosystems can be found in the main paper. Each community is suspended in a liquid medium held in a flask subjected to a continuous fixed-rate chemical flux, giving growing conditions similar to those found in a chemostat. Individual microbes grow and reproduce dependent on food supply and environmental conditions within the flask, and nutrient cycling loops and stable ecologies emerge from the indirect interaction of individuals via the flask environment. Mutation may occur during microbe reproduction, allowing the genesis of new microbial strains.

The composition of the liquid medium in each flask determines the environment of the microbes. Some of the chemicals present may be consumed as food by the microbial population and converted to biomass, while others are non-consumable and form part of the abiotic environment. In addition it is assumed that the liquid medium has properties such as temperature, pH, salinity, etc., and that these both affect, and can be affected by, microbial activity. We will refer to these non-consumable chemicals and physical properties of the flask environment collectively as 'abiotic factors' for ease of discussion; while chemical nutrients are also abiotic we feel that their role as the subjects of metabolism justifies this notational convenience. The effect of the microbes on abiotic factors is modelled here as a side-effect of metabolism, with a genetically specified effect caused by each microbe for every unit of biomass created.

The composition of the abiotic environment resulting from the interaction of the input and output fluxes with the collective actions of the microbial population forms a 'phenotypic' ecosystem trait that is used as the basis for selection. Offspring ecosystems are formed by innoculating sterile flasks with seed populations sampled from the most successful ecosystems in the previous iteration. The response to selection is measured as the change over time in the distance of the environmental state variables from some pre-specified target vector. The artificial ecosystem selection method is described fully in the main paper.

1.1 The flask environment

Each flask contains a neutral liquid matrix in which is suspended a microbial population. There is a flow of liquid medium through the flask which occurs continuously at a fixed rate.

The inflow brings with it influxes of nutrients at fixed concentrations and steady inputs to abiotic factors, while the outflow removes a fixed proportion of stored nutrients and abiotic factors. The liquid medium in each flask is assumed to be well-mixed, so that in the absence of perturbation the composition of the medium in each flask will reach a homogeneous steady state equilibrium. Each microbe both consumes and excretes chemical nutrients, and also has an effect on the levels of the abiotic factors as a side-effect of metabolism (explained in Section 1.3.4 below).

The state of the flask environment is given by a vector V:

$$V = (n_1, ..., n_N, a_1, ..., a_A) = (v_1, ..., v_{N+A})$$

where n_i is the concentration of nutrient i, a_i is the level of abiotic factor i, or equivalently, v_i is the level of the i^{th} environmental state variable. N is the number of chemical nutrients, and A is the number of abiotic factors. The change in V over time is given by Equation 3, which relates the rate of change of each v_i over time to the rates of influx and outflux of that variable and the net effect of microbial activity. I_i is the rate of influx of v_i, O_i is the rate of outflux, and E_i is the effect on v_i of current microbial activity. The form of Equation 3 is general to nutrients and abiotic factors, although E_i is calculated differently for nutrients and abiotic factors (see Sections 1.3.1 and 1.3.4).

$$\frac{dv_i}{dt} = I_i - O_i v_i + E_i \tag{3}$$

1.2 Microbes

Microbes are modelled as simple organisms that consume and excrete nutrients in fixed proportions and affect the levels of abiotic factors in their environment as a side-effect of metabolism. The precise ratios in which nutrients are consumed and excreted, and the nature of the by-product effect on abiotic factors, are genetically encoded for each individual, as are its preferred abiotic conditions (i.e., the state of the abiotic environment in which its growth rate will be maximised). Microbes grow by converting consumed nutrients to biomass and reproduce by splitting when their biomass reaches a fixed threshold. Biomass is reduced at a fixed rate to represent the inevitable thermodynamic inefficiency of metabolism and the cost of maintaining cellular machinery. Microbes die if their biomass drops below a fixed threshold, which can happen in sustained periods of nutrient limitation.

A microbe can be represented by a vector M:

$$M = (B, \lambda, \mu, \alpha, \beta)$$

where B is the current biomass of the microbe, $\lambda = (\lambda_1, ..., \lambda_N)$ represents the ratio in which nutrients are consumed, $\mu = (\mu_1, ..., \mu_N)$ the ratio in which excreta is returned to the environment as nutrients, $\alpha = (\alpha_1, ...\alpha_A)$ the microbe's effect on abiotic factors as a side-effect of metabolism, and $\beta = (\beta_1, ...\beta_A)$ the relative proportions of abiotic factors in the environment at which growth rate is maximised. Clearly $\sum_{i=1}^{N} \lambda_i = 1$ and $\sum_{i=1}^{N} \mu_i = 1$ hold since all materials consumed and excreted must be accounted for; there is no such constraint on α since the effect of the microbe on the abiotic environmental factors does not necessarily involve mass transfer and is thus treated generally. $\sum_{i=1}^{A} \beta_i = 1$ also holds, since environmental

preference is represented here as a vector specifying the optimal proportion of each abiotic factor relative to the other abiotic factors in the flask environment. Of the quantities in the microbe state vector M, only B is a variable during the lifetime of an individual, since λ, μ, α and β are genetically encoded and thus fixed.

1.2.1 Genotype

The genotype for a microbe is an array with $2N + 2A$ loci taking values in the range $[-1.00, 1.00]$. The genotype is subdivided into two sets of N loci for consumption and excretion and two sets of A loci for influence on abiotic factors and preferred environmental conditions. The microbe phenotype is formed by mapping and transforming the values in its genotype according to fixed rules.

$$\overbrace{\underbrace{\begin{array}{cccc} \{N\ consumption\ loci\} & \{N\ excretion\ loci\} & \{A\ effect\ loci\} & \{A\ preference\ loci\} \\ \downarrow & \downarrow & \downarrow & \downarrow \\ \lambda & \mu & \alpha & \beta \end{array}}_{\text{Phenotype}}}^{\text{Genotype}}$$

The consumption ratio λ (specifying the fixed microbe-specific proportions in which nutrients are consumed) is found by linearly mapping the N alleles for consumption to the range $[0.00, 1.00]$ and normalising to give the fraction of total consumption that is made up of each nutrient. The excretion ratio μ (specifying the fixed microbe-specific proportions in which excreta is returned to the environment as nutrients) is found similarly. For example, if $N = 3$ and the consumption loci of the genotype are $(-0.4, 0.7, 0.1)$, this would map linearly to $(0.3, 0.85, 0.55)$ and give a normalised consumption ratio of $\lambda = (0.18, 0.5, 0.32)$.

The vector α determining a microbe's effect on the abiotic factors in the environment is found by directly mapping the A alleles from the relevant part of the genotype to the values for the phenotypic trait without scaling or transformation, i.e., values remain in the range $[-1.00, 1.00]$. These values give the alteration caused in the level of each abiotic factor by the creation of a single unit of biomass during microbe metabolism (see Section 1.3.4 for more details). A microbe's preferred abiotic environment β is determined by linear mapping and normalisation of the relevant A alleles of the genotype, using the same scheme as that for finding the consumption and excretion ratios λ and μ. The microbe's preferred environment is thus expressed as a vector of the relative proportions of each abiotic factor; the microbe's growth rate will be maximised when the state of the environment matches this preference (see Section 1.3.3).

Note that 'genotypes' as defined here are highly abstracted analogues of their biological inspiration; the developmental stage used here is direct and deterministic, and there is no possibility of significant epistatic interactions or pleiotropy. However, the use of the term 'genotype' here is justified since the genotype is the mechanism of microbe heredity, the determinant of microbe phenotype, and the subject of mutations leading to phenotypic variation.

1.2.2 Reproduction and mutation

If the genetic specification of the microbe causes it to be successful in its environment (i.e., if its nutrient demands and preferred abiotic conditions are suited to the current state of the liquid medium held in the flask), the microbe will consume nutrients and grow by increasing its biomass. If a microbe's biomass reaches the reproduction threshold T_R, it reproduces asexually by splitting. The parent microbe donates half of its biomass to the offspring microbe, which also receives a copy of the parent's genotype. Mutation of the offspring genotype may occur during reproduction, implemented as a potential for copying error at each locus of the genotype which causes the allele value for that locus to be randomly reassigned from the range $[-1.00, 1.00]$. Mutations occur at each locus with low probability P_{mut}. No mechanism for genetic recombination is implemented.

1.2.3 Maintenance cost and death

Unsuccessful microbes will not consume enough nutrients to grow and may reduce in biomass due to a fixed rate of biomass decrement which is incorporated in the model as a proxy for the combined energy costs of maintaining cellular machinery and metabolic inefficiency. This 'cost of living' reduces biomass at a fixed rate γ per simulation timestep, with the decrement assumed to be lost from the flask environment as unrecoverable heat radiation. The inclusion of this cost ensures that nutrients cannot be infinitely recycled and thus preserves the thermodynamic integrity of the model.

If the biomass of a microbe falls below a threshold T_D the microbe is assumed to die from starvation. In addition to this, each living microbe may die 'from natural causes' with a low probability P_D at each timestep. This mechanism is intended to be a catch-all for death by predation, senescence, etc., and serves to thin out the microbial population in an unbiased way and thus promote continuing selection and competition between microbes. Note that the value of P_D is related to the washout rate of living microbes in chemostat models. When a microbe dies it is assumed that its remaining biomass is washed out and lost from the system.

1.3 Microbe metabolism

1.3.1 Nutrient consumption / excretion

At each timestep of the simulation, each living microbe j will attempt to consume a total of C_j^{max} units of nutrient, with the contributions to this total made up of each different nutrient type fixed in the relative proportions defined by the microbe's genetically specified consumption ratio. The size of C_j^{max} is limited by a global maximum level C^{max} and is calculated on an individual basis for each microbe j. This calculation takes into account the match between the current state of the abiotic environment and the genetically specified preferences of microbe j and will be covered in Section 1.3.3 below. The actual amount consumed C_j^{act} is less than or equal to C_j^{max} and depends on nutrient availability.

In order to ensure that the microbe population doesn't consume more nutrients than currently exist in the flask environment, individual demands may need to be scaled, and to ensure that no artefacts are introduced into the model by this scaling, it must not favour any particular individual. At each timestep the total nutrient demand for the entire microbe population is calculated and compared to the amounts of nutrients available. It is assumed

that all microbes are continuously and simultaneously feeding, so in the case that there is an insufficient amount of a nutrient available to meet the entire population demand, the demand of every individual microbe that requires that nutrient is scaled down equally, so that the total amount consumed by the population matches what is available.

Mathematically, we have constraints $C_j^{max} \leq C^{max}$ and $C_j^{act} \leq C_j^{max}$, and then:

$$C_j^{act} = C_j^{max} \prod_{i=1}^{N} w_{ij} \tag{4}$$

where C_j^{act} is the actual total quantity of nutrients consumed by microbe j after scaling for nutrient limitation has been applied and w_{ij} is the scaling factor for nutrient i for microbe j. Values for w_{ij} are calculated sequentially by noting that the population demand D_i for nutrient i (after all individual demands for nutrient $i-1$ have been scaled appropriately) and w_{ij} are related. Recalling that λ_{ij} is the proportion of consumption taken as nutrient i by microbe j and n_i is the total amount of nutrient i currently available in the flask environment, we can derive the full set of all w_{ij} and D_i for every nutrient i and living microbe j by solving iteratively for each value of D and w, starting with an assumed value of $w_{0j} = 1$ (valid since nutrient 0 does not exist).

$$\begin{aligned} w_{0j} &= 1 && \forall j \\ w_{ij} &= \begin{cases} min(1, \frac{n_i}{D_i}) & \lambda_{ij} > 0 \\ 1 & \lambda_{ij} = 0 \end{cases} \end{aligned}$$

$$D_i = \sum_{j}^{living} \left(\lambda_{ij} C_j^{max} \prod_{k=0}^{i-1} w_{kj} \right)$$

Then having established the value of C_j^{act} we can go on to derive:

$$C_j = C_j^{act}(\lambda_{1j}, ..., \lambda_{Nj}) \tag{5}$$

$$C_i^{pop} = \sum_{j}^{living} \lambda_{ij} C_j^{act} \tag{6}$$

where C_j is the actual consumption vector for microbe j and details how much of each nutrient is consumed by microbe j at a particular timestep. C_i^{pop} is the total amount of nutrient i consumed by all living microbes.

This scheme means that each microbe always consumes nutrients in the relative proportions specified by its genetically determined consumption ratio. If nutrient limitation means that the amount of a particular nutrient consumed by a microbe is scaled down, the amounts of the other nutrients its consumes are also scaled down by an equivalent factor to maintain the fixed relative proportions of consumption.

Consumed nutrients are converted to biomass with a standard efficiency of θ, with the waste being excreted as nutrients (i.e., $C_j^{act} = 10$ units of food consumed with efficiency of $\theta = 0.6$ makes 6 units of biomass and 4 units of excreta). Excreta is returned to the environment as

nutrients in the fixed proportions specified by the microbe's genetically determined excretion ratio. We can thus define the excretion vector X_j for every microbe j, and an expression for the total amount X_i^{pop} of nutrient i collectively excreted and returned to the environment by the population:

$$X_j = (1-\theta)C_j^{act}(\mu_{1j}, ..., \mu_{Nj}) \tag{7}$$

$$X_i^{pop} = \sum_{j}^{living} (1-\theta)C_j^{act}\mu_{ij} \tag{8}$$

1.3.2 Growth

At each timestep of the simulation a microbe j will consume C_j^{act} units of nutrient, which are converted to biomass with a fixed efficiency of θ. Taking into account the previously defined maintenance cost γ, we can now state the growth rate (rate of change of biomass) of a microbe j as:

$$\frac{dB_j}{dt} = \theta C_j^{act} - \gamma \tag{9}$$

Note that Equation 9 specifies the growth of an individual. Growth of a population occurs only as a result of individual growth and reproduction, and is not specified *a priori* as in more traditional population ecology models such as Lotka-Volterra systems.

1.3.3 Effect of abiotic factors on metabolic rate

The model is designed so that the state of the abiotic environment affects the growth rate of microbes, and this is implemented as a feedback from the environmental state variables for abiotic factors onto the consumption demands of all microbes. As mentioned above, each microbe will attempt to consume a maximum amount C_j^{max} of nutrients at each timestep, with this demand being met depending on nutrient availability. The attempted consumption amount C_j^{max} is calculated for each microbe j as a function of the match between the current state of the abiotic environment and the microbe's genetically specified preferences. This function has Gaussian form and falls away smoothly from its maximum value as the distance between the current environment and the optimum increases. Mathematically, we can capture this as below:

$$C_j^{max} = \psi_j C^{max} \tag{10}$$

$$\psi_j = e^{-\tau(\rho_j)^2} \tag{11}$$

$$\rho_j = \sqrt{\sum_{i=1}^{A}(\hat{a}_i - \beta_{ij})^2} \tag{12}$$

$$\hat{a}_i = \frac{a_i}{\sum_{i=1}^{A} a_i} \tag{13}$$

where C^{max} is a universal constant defining the maximum rate of consumption for any microbe, ψ_j is a microbe-specific measure of the microbe's satisfaction with the current abiotic environment given its preferences, τ is a universal constant parameter that sets the level of influence of the abiotic environment on growth rate (high τ means a stronger influence, $\tau = 0$ means no influence), ρ_j is a measure of the distance between the current abiotic environment and the microbe's preferred environment, $\hat{a}_i$ is the normalised level of abiotic factor a_i, and β_{ij} is microbe j's preferred normalised level for factor a_i.

1.3.4 Effect of microbial activity on environment

Microbes can affect the amount of nutrients in the flask environment as well as the levels of the abiotic factors. During metabolism microbes remove nutrients from the environment by consumption and add nutrients to the environment by excretion. Also, during metabolism, microbes affect the levels of abiotic environmental factors as a side-effect of their metabolic action.

The metabolic effect on abiotic factors is implemented in the model as a mechanism by which microbes can affect their abiotic environment. In the real world, all organisms have an effect on their abiotic surroundings, and while metabolism is one way in which organisms do this (particularly in the microbial world), other kinds of behaviour are also significant. However, to keep the model as simple as possible we have only implemented the interaction of microbes with their abiotic environment as a side-effect of metabolism. This method has the advantage of making the size of a microbe's effect on the environment proportional to its growth rate, so that fast-growing and more fecund microbial strains have a greater influence than dormant or slow-growing strains, as seems intuitively correct.

Equation 3 relates the rate of change of each environmental variable v_i to the combined effect E_i of the microbial population on that variable. The calculation of E_i is treated differently for nutrients and abiotic factors. First of all, let the vector E be the vector of effects of the population on all environmental variables, and note that this vector can be sub-divided into nutrient and abiotic components.

$$E = (E_n, E_a) = (E_{n1}, ..., E_{nN}, E_{a1}, ..., E_{aA})$$

We can use our previous definitions of nutrient consumption/excretion to work out the population effect on nutrient state variables. This gives us an expression for the effect E_{ni} of the population on nutrient i:

$$E_{ni} = -C_i^{pop} + X_i^{pop} \tag{14}$$

The effect of each microbe on each abiotic factor is determined by its genetically specified effect vector α and is applied for every unit of biomass created. The expressions for microbe growth combined with this vector can be used to work out the population effect E_{ai} on each abiotic state variable i:

$$E_{ai} = \sum_{j}^{living} \frac{dB_j}{dt} \alpha_{ij} \tag{15}$$

1.4 Simulation method

At the start of each simulated timestep, nutrient/abiotic influx is added to the environment. This is followed by simultaneous update of all microbes in the population for metabolism, death (by starvation or random selection), and reproduction, in that order. When the biota are updated, nutrient/abiotic outflux is removed from the environment and the system update is complete for that timestep.

The update equation for the flask environment (Equation 3) and the microbial growth equation (Equation 9) are continuous differential equations. At each timestep these differential equations are discretised by calculating their instantaneous value and adding it to the existing values of the quantities concerned. For environmental update we have $\Delta V = \frac{dV}{dt}$ and then $V_t = V_{t-1} + \Delta V$. For microbe growth we have $\Delta B = \frac{dB}{dt}$ and $B_t = B_{t-1} + \Delta B$. In effect this is equivalent to numerical integration using Euler's forward method with an integration timestep equal to one simulated timestep in the flask ecosystem.

2 Additional Results

The results given in this section support those given in the main body of the paper. Some interpretation is given in figure/table captions here, but these results are best understood by reference to the main paper.