

# Defense versus growth in a hostile world - Lessons from phage and bacteria

Rasmus Skytte Eriksen<sup>1</sup> & Sandeep Krishna<sup>2</sup>

<sup>1</sup> Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark

<sup>2</sup> Simons Centre for the Study of Living Machines, National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore, India.

May 25, 2020

## Abstract

Bacterial communities are often highly diverse with several closely related species (or strains) coexisting together. These bacteria compete for resources and the competitive exclusion principle predicts that all but the fastest-growing bacteria will go extinct. When exposed to phage, it is predicted that bacterial strains with restriction-modification (RM) systems can circumvent the competitive exclusion principle and reach diversity on the order of the phage burst size. We show that with a trade-off between bacterial growth rates and the strength of their RM systems, the diversity of such an ecosystem can further increase several fold beyond the burst size limit. Moreover, we find that the ratio of the growth rate of a bacterial strain to the imperfection of its RM system is an excellent predictor of (i) whether the strain will go extinct or not, and (ii) the biomass of the strain if it survives. In contrast, the growth rate alone is not a determinant of either of these properties. Our work provides a quantitative example of a model ecosystem where the fitness of a species is determined not by growth rate, but by a trade-off between growth and defence against predators.

**Keywords:** Restriction-modification systems. Bacteriophage. Biodiversity. Predator partitioning. Growth-defense trade-off.

## 1 Introduction

Bacteria and their phage predators are the most abundant organisms on Earth and are present in almost every known ecosystem. Due to their ubiquitous presence, bacteria strongly influence local environments, while phages in turn strongly influence the composition of the bacterial population as well as the selection pressures on individual bacterial strains[1, 2]. It is therefore important to understand the dynamics of phage-bacteria ecosystems. A key feature of such ecosystems is the constant co-evolution of bacterial defence systems against phage, and phage mechanisms to avoid or bypass these defences[3, 4, 5, 6]. Here, we focus on the defence provided by restriction-modification (RM) systems in bacteria. These consist of a pair of enzymes, which target the same specific DNA sequences (typically 4-6 bases long). The restriction enzyme cuts the DNA at these sites if they are unmethylated, whereas the corresponding methyltransferase modifies the DNA at these sequences thus protecting it from the restriction activity. If a phage enters a bacterium with such an RM system, and it is not methylated at the target sequences, it will be destroyed with a high probability. However, there is some chance that it will be methylated at all the sites by the methyltransferase before the restriction enzyme can act, thus escaping the RM system to proliferate unhindered. We will refer to the probability with which an RM system successfully destroys an unprotected infecting phage as its “efficiency” or “strength”.

RM systems are found in 90 % of all bacterial genomes[7], and also occupy a significant portion of bacterial genomes, on average even more than CRISPR defences [5]. This is curious because at

first glance RM systems appear to be a very weak defence against phage. This is because if even a single phage particle manages to evade the defence by becoming appropriately methylated before it is destroyed by the restriction enzyme, then it and all of its future offspring will be immune to this RM system. The probability that any given phage evades the defenses is in the range  $\sim 10^{-2} - 10^{-8}$  [8, 9]. Given that phage burst sizes are on the order of hundreds [10], even with highly efficient RM systems a population of bacteria will eventually let a phage through its defenses. However, in an ecosystem consisting of many different strains of bacteria with different RM systems, it turns out that this apparently transient defence can result in effective cooperation between bacterial strains and a long-term boost in diversity [11, 12, 13]. A previous mathematical model [13] has argued that the burst size of the phage sets an upper bound on the diversity of bacterial strains – the number of strains that can co-exist. In contrast, in the absence of phages, competitive exclusion [14, 15] would allow only a single strain to survive in the long run.

This previous model does not, however, consider the correlation between the efficiency of an RM system and its cost, in terms of reduction of the growth rate of the bacterium. Rather, this model assumes the cost and efficiency are uncorrelated, which allows the appearance in the model of unrealistic, near cost-less near-perfect RM systems. In this paper, we implement such cost-efficiency correlations in the mathematical model and find that surprisingly high diversities are possible - the number of bacterial strains that can co-exist can be several-fold larger than the burst size of the phage. We argue that such high diversity is possible because the phage-bacteria interaction, in this case, allows even bacteria with quite low growth rates, and therefore low steady-state biomass, to stably co-exist with bacteria with much higher growth-rates and biomass.

## 2 Methods and Materials

### 2.1 An ecosystem of phage and bacteria with RM systems

We consider an ecosystem consisting of one type of phage and  $N$  bacterial strains all of which are potential hosts for the phage but who contain different, unique, RM systems. Due to the presence of different RM systems, there exists an equal number,  $N$ , of epigenetic variants of the phage which are appropriately methylated to avoid the corresponding RM systems. In a well-mixed ecosystem of this sort, the equations describing the dynamics of the densities of each bacterial strain,  $b_i$ , and each epigenetic phage variant,  $p_i$ , are (see [13]):

$$\dot{b}_i = \gamma_i b_i (1 - B/C) - \eta_i b_i p_i - \eta_i \omega_i b_i (P - p_i) - \alpha b_i, \quad (1)$$

$$\dot{p}_i = \eta_i \beta_i b_i p_i + \eta_i \omega_i \beta_i b_i (P - p_i) - \eta_i p_i B - \delta_i p_i. \quad (2)$$

These equations model growth of bacteria, phage adsorption followed by either cell lysis and a burst of offspring phage or the destruction of the phage by the RM system (see figure 1), as well as the death of phage and bacteria. Here,  $\gamma_i$  is the growth rate of the  $i$ th bacterial strain,  $C$  is the total bacterial carrying capacity of the environment,  $\alpha$  is the rate at which bacteria are diluted from the ecosystem, and  $\omega_i$  is the imperfection of the  $i$ th RM system, i.e. the probability that a phage that is *not* correctly methylated will escape the RM defence.  $\omega$  is thus inversely related to the efficiency of the RM system. The rest of the parameters describe phage properties:  $\eta_i$  is the rate of adsorption to the  $i$ th bacterial strain,  $\beta_i$  is the burst size, i.e., the number of phage progeny released during lysis of the  $i$ th bacterial strain, and  $\delta_i$  is the rate at which the  $i$ th phage variant decays. Various choices can be made for these parameters – we will return to this point in a moment. The only difference from the equations of [13] is that we explicitly include the carrying capacity  $C$ , which they took to be unity.

### 2.2 Open system with periodic addition of new bacterial strains

With a fixed number of bacterial strains, the bacterial and phage densities typically reach a steady-state after order 10 bacterial generations [13]. However, over even longer timescales, one may expect that new strains of bacteria (with different, unique RM systems) arise by invasion, mutation, or acquisition of RM systems from the environment or other bacteria. To model this, we begin with an ecosystem consisting of  $M$  strains of bacteria and the corresponding  $M$  phage variants, and periodically, at times  $T, 2T, 3T, \dots$ , introduce a new bacterial strain with a unique RM system

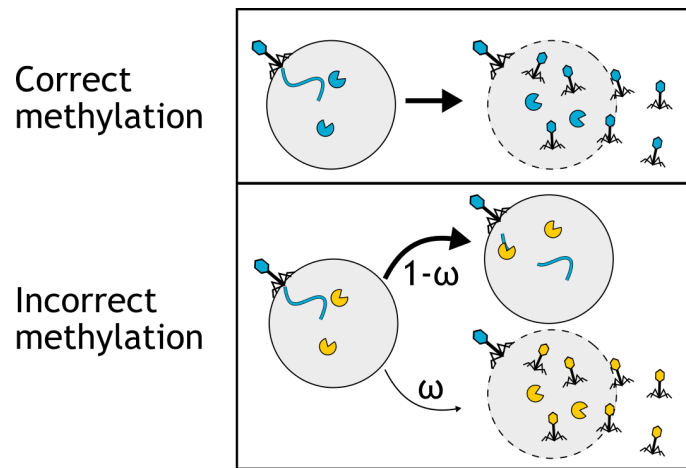


Figure 1: Illustration of phage infections. The fate of the phage infection depends on the methylation state of the phage genome. If all recognition sites on the phage genome have been methylated, the RM system is ineffective and the phage will ultimately lyse the bacterium. If the phage does not have the methylated recognition sites, the RM system will almost always cleave the phage genome and prevent the proliferation of the phage, but with probability  $\omega_i$  the phage be able to escape the RM systems and lyse the bacterium. Note that the progeny phage will now be methylated differently than the parent phage.

along with the corresponding phage variant, both at a very low density of  $1/C$ , which corresponds to a single individual.  $T$  is typically chosen large enough to allow the system to reach steady-state between additions of new strains. At the same time as new species are added, we remove all bacterial strains and phage variants whose density has fallen below the threshold density of  $1/C$ .

### 2.3 Choice of parameter values and correlation between cost and efficiency of RM systems

In [13], the phage parameters  $\eta$ ,  $\beta$ , and  $\delta$  were chosen to be the same for all phage variants, whereas  $\omega_i$  and  $\gamma_i$  were allowed to vary between strains. We do the same (see table S1 for a full list of parameter values), except that we choose the  $\omega_i$  and  $\gamma_i$  values differently. Ref. [13] sampled  $\omega_i$  from a log-uniform probability distribution between  $10^{-4}$  and 1, and the  $\gamma_i$  independently from a uniform probability distribution between  $\alpha$  and 1. We will refer to such sampling as the “uncorrelated RM system” case. We deviate slightly from Ref. [13] and sample the growth rates  $\gamma_i$  uniformly between 0 and 1. We also examine the “correlated RM system” case, where we draw the pair  $(\gamma_i, \omega_i)$  from a joint probability distribution  $P(\gamma, \omega)$ , chosen so that the two quantities are correlated, with an increased strength of the RM system resulting in an increased cost to the bacterium via a decreased growth rate. We compute the  $\gamma(\omega)$  values by taking a product of appropriately scaled (log) uniform random numbers. The correlations arise because the scalings applied to the random numbers are controlled by a common parameter (see supplement section S2 for details). Wherever we compare the correlated with the uncorrelated RM case, we make sure that the correlated distribution is normalized such that the (marginal) expectation values for the growth rates and RM strengths are the same as in the uncorrelated distribution (i.e.  $\langle \gamma \rangle = 0.5$ , and  $\langle \log_{10}(\omega) \rangle = -2$ ).

Figure 2A schematically exhibits the cost-efficiency correlations in our model. Here, the unique RM systems are shown in different colours, and the corresponding epigenetic phage variants are correspondingly coloured. The barred arrows connecting the phage to the primary hosts indicate that phage variants strongly suppress the bacterial strain with the corresponding RM systems but only weakly suppress the other strains. The correlations result in strains with weak RM systems (less pac-men) reproducing faster, and vice-versa.

For the correlated RM system, we investigate how the important the structure of the joint distribution  $P(\gamma, \omega)$  is in two ways: 1) We change the “steepness” of the distribution, i.e., how much the efficiency of an RM system reduces the growth rate of a bacterium, and 2) we change

the Pearson correlation coefficient of the distribution while keeping the marginal means constant. For these two purposes we use two different choices of  $P(\gamma, \omega)$  which are parametrized differently to make it easier to change the steepness or the correlation coefficient.

In the first case, we decrease the costs of the RM systems (see section S2 for the details). Essentially, we scale the  $\gamma$  values which allows us to control the steepness of the joint probability distribution in the  $\gamma - \omega$  plane (quantified as the slope  $a$  determined by a least-square fit to  $\gamma = a \log 10(\omega) + b$ ).

The above tuning of the steepness does result in a small change the correlation coefficient. Therefore, to understand how changing the correlation coefficient affects our results, we generate  $(\gamma, \omega)$  pairs by sampling from a truncated multivariate normal distribution. This new distribution has the property that the Pearson correlation coefficient,  $\rho$ , between  $\gamma$  and  $\omega$  is directly controllable. To compare with the previous distributions, we normalize the multivariate normal distribution such that the pairs generated have the same marginal expectation values for as the uncorrelated and correlated distributions above. See section S2 for reference.

## 2.4 Solution to the dynamical equations

In supplementary section S3, we solve eq. (2) for the steady-state and we construct a simplified method for determining the diversity of the system. The steady-state analysis produces: (i) equations for the total bacterial density,  $B$ , and the total phage density,  $P$ , in terms of the set of  $\omega_i$  and  $\gamma_i$  values (see eqs. (S21) and (S14)); (ii) an equation for each individual bacterial density,  $b_i$ , also in terms of the set of  $\omega_i$  and  $\gamma_i$  values. From the latter, (S22), we observe that in order for all the  $b_i$ s to be positive in steady-state, each pair  $(\gamma_i, \omega_i)$  must fulfill:

$$f(\gamma_i, \omega_i) = \gamma_i(1 - B/C) - \alpha - \eta\omega_i P > 0. \quad (3)$$

Thus, given the set of  $\omega_i$  and  $\gamma_i$  values, we can calculate which bacterial strains will go extinct once the system reaches steady-state.

### 2.4.1 Iterative approximation of the model dynamics

The steady-state solution from section S3 above allows us to simulate the open ecosystem using the following iterative procedure:

1. Set  $B = 0$ ,  $P = 0$  and  $t = 0$  ( $t$  is the current time).
2. Draw a  $(\gamma_i, \omega_i)$  pair from a specified joint distribution  $P(\gamma, \omega)$  (which may be correlated or not).
3. If the pair fulfills:  $f(\gamma_i, \omega_i) > 0$ , go to step 4 (computing densities), otherwise go to step 6 (advance time).
4. Re-compute  $B$ , all  $b_i$ s, and  $P$  based on the set of  $(\gamma, \omega)$ , using eqs. (S21), (S14) and (S22).
5. If any  $(\gamma_i, \omega_i)$  pair no longer fulfills  $f(\gamma_i, \omega_i) > 0$ , then remove the pair with the smallest  $b_i$ . Go back to step 4 (recomputing densities).
6. Add  $T$  to the current time  $t$ . Go back to step 2 (add a new RM system).

This iterative procedure is an approximation of the full dynamics, but it is a good one whenever the time  $T$  is large enough that the system reaches close to a steady-state before the addition of each new strain. Supplementary figure S2 and S3 show evidence that the diversity predicted by the iterative model matches the full dynamical model with  $T = 10^4$  very well.

## 2.5 Implementation and data availability

We implemented our algorithm in MATLAB (R2018b version 9.5.9.944444) and used the ODE solver ODE45 to solve the dynamical equations. The code and generated data files are available at the online repository located here[16]:

<https://github.com/RasmusSkytte/CorrelatedRestrictionModificationSystems/tree/v1.0>

## 3 Results

### 3.1 Correlated RM systems enable higher diversity

We first examine an open ecosystem model, where at periodic intervals of time a new bacterial system with a unique RM strain is introduced, along with the corresponding phage variant (see Methods and Materials section for details). The interval between each such injection is chosen to be long enough for the system to reach close to steady-state. When the RM system strengths,  $\omega_i$ , and the bacterial growth rates,  $\gamma_i$ , are uncorrelated, we find, as did Ref. [13], that the diversity is limited by the burst size (see figure 2B, light blue line). This limit was explained in Ref. [13] — the effective cooperation between bacteria results in the bacterial densities becoming roughly equal, and this, combined with the requirement that at least one phage on average survives from each burst to successfully infect a bacterium, means the number of strains cannot exceed the burst size of the phage. Indeed, we find that in this case, the bacterial density distribution exhibits a narrow peak (see figure 2C, hatched bars).

However, when we introduce correlations by making the bacterial growth rates inversely related to the strength of their RM systems, then we find that the diversity can break through this burst size limit and increase up to two-fold beyond it (figure 2B, dark blue line). Simultaneously, the equipartition of bacterial density is violated; the bacterial density distribution becomes much broader (figure 2C, shaded bars).

We conclude that having such a broad distribution of bacterial density (in particular the possibility for bacterial strains with quite low densities to co-exist with other strains with much higher densities) is crucial for obtaining such high diversity. The introduction of correlations also changes the final distribution of growth rates and RM strengths (see figure 2D-E). In the uncorrelated case (hatched bars), the growth rates,  $\gamma$ , are narrowly distributed near the maximal value of 1, while the strength of the RM systems are broadly distributed between the smallest possible value ( $10^{-4}$ ) and the highest value at which RM systems provide utility ( $1/\beta$ ). When correlations are introduced (solid bars), we see a broader distribution of growth rates, albeit still biased towards faster growth. Interestingly, we observe a narrowing of the distribution of RM strengths whereby highly effective RM systems are no longer observed and bacteria with weaker RM systems can survive.

Note that the presence of the phage is crucial to obtain this high diversity. In the absence of the phage, competitive exclusion would result in only the bacterial strain with the highest growth rate surviving.

### 3.2 Explanation for increased diversity with correlated RM systems

In the  $\gamma-\omega$  plane shown schematically in figure 2F, the added strains are picked from the enclosed region which depicts the set of possible values determined by the probability distribution  $P(\gamma, \omega)$ .

From the steady-state solution to the model (see section 2.4 in Methods and eq. (S22) in Supplementary section S3), we know that in order for a strain to be viable, it must fulfill the condition:  $f(\gamma_i, \omega_i) = \gamma_i(1 - B/C) - \alpha - \eta\omega_i P > 0$ . Thus, a strain will survive and co-exist with the existing ecosystem if its  $\gamma$  and  $\omega$  values place it in the green region, below and to the right of the red curve (which indicates  $f(\gamma, \omega) = 0$ ), whose location is determined by the current total bacterial and phage densities. Conversely, it will go extinct if it is above or to the left of the red curve. In turn, the successful addition of a strain causes the total bacterial density  $B$  to increase, which results in the red line moving to the right. Thus, successive successful additions of new strains will eventually reduce the area of the viable (green) region that will result in survival. Further, eq. (S22) shows that if a strain survives, the closer it is to the red line, the lower its steady state biomass.

In the absence of correlations, the distribution  $P(\gamma, \omega)$  uniformly covers a rectangle in the  $\gamma-\omega$  space, as shown in figure 2G. In such a case, the red line typically settles to the far right, resulting in the surviving species having a narrow distribution of large growth rates, close to 1.

In contrast, when we introduce a trade-off, the shape of the grey region changes because of the correlation between the RM strengths and the growth rates, as shown in figure 2H. Here, the red line is unable to move as far to the right as in the previous case, since the points are closer to the red line and thus contribute less biomass to the system. The red line typically settles in a region that allows a broader distribution of growth rates in the surviving strains. Hence, one obtains a broader distribution of bacterial densities and a higher diversity.

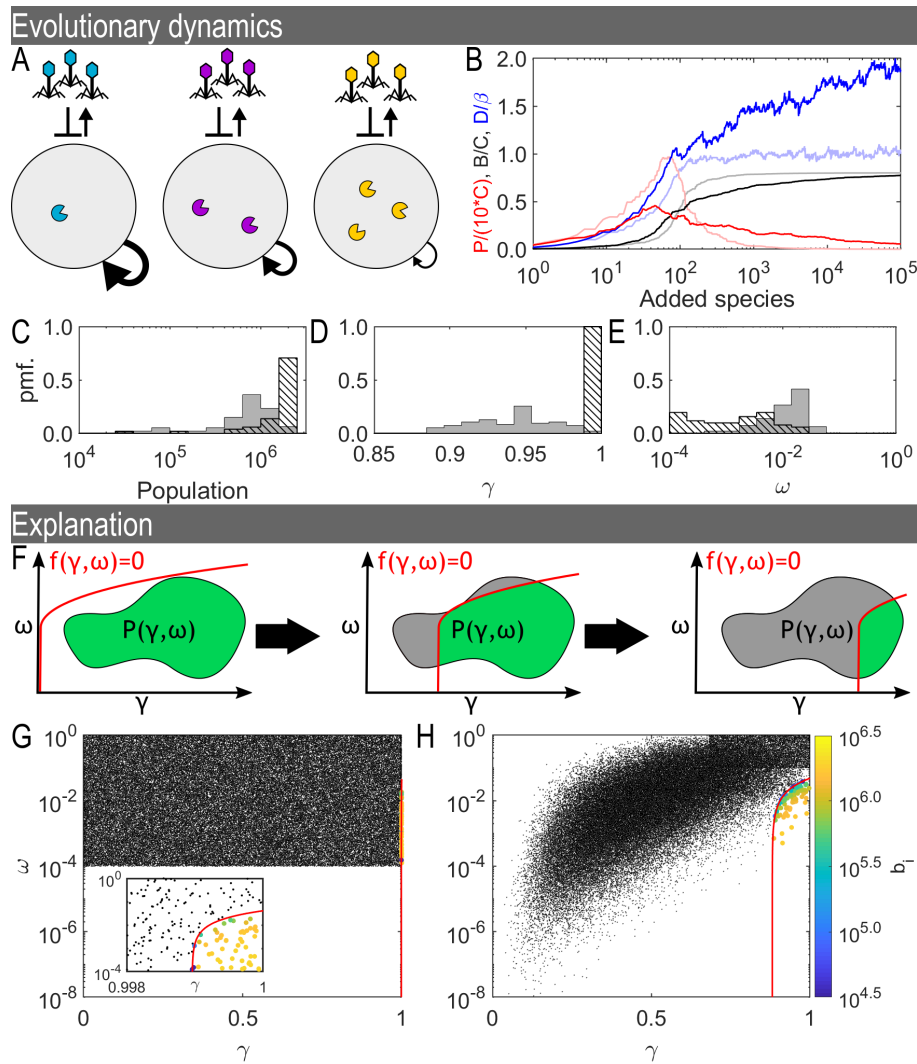


Figure 2: Correlations and diversity. (A) Schematic of the model. The system consists of populations of bacterial strains with different, unique RM systems, each represented by a colour. Each of the bacterial strains reproduces at a rate  $\gamma_i$ , which is inversely proportional to the strength of the RM system it has (correlated case). For each such strain, there is a corresponding phage variant (with matching colour) which has the pattern of methylation required to avoid that RM system. Phage with wrong patterns of methylation can only infect a bacterium with probability  $\omega_i$ . (B) Time courses from a simulation with correlated RM systems showing the total bacterial density,  $B$ , and the total phage density  $P$ , in units of the carrying capacity  $C$ . Also shown is the diversity  $D$  (the number of co-existing bacterial strains) in units of the phage burst size  $\beta$ . The dark (light) colours correspond to the case of (un)correlated RM systems. (C-E) Distributions at the end of the simulation. (un)correlated RM systems shown as solid (hatched) bars. (C) Distribution of strain population sizes. (D) Distribution of growth rates. (E) Distribution of RM imperfections. (F) Schematic of the simplified model in  $\gamma$ - $\omega$  space. The red line indicates the boundary between viable and non-viable strains (representing eq. (3) in the main text). The filled area (grey+green) indicates the region from which  $(\gamma, \omega)$  pairs are drawn. The green region depicts the portion of this area where species are viable. As species are added, the region corresponding to viable strains constricts. (G) Black dots indicate the available strains from panel (B, light curves, uncorrelated). Coloured points indicate the species surviving at the end of the simulation. Colour scale is proportional to  $\log_{10}$  of the population. Inset zooms in on the relevant region. (H) Same as (G) but for correlated RM systems.

By viewing the dynamics in  $\gamma - \omega$  space, we can now understand why correlated RM systems change the distribution of growth rates and RM strengths (see figure 2D-E). Without the correlations, strong RM systems are available for all growth rates and they drive the bacteria to a narrow distribution of growth rates. With correlations, the lack of low  $\omega$  - high  $\gamma$  species leads to a broader distribution of growth rates.

We can understand this increased diversity intuitively if we modify the argument used in Ref. [13]: In the steady-state, at least one phage from a burst must find a new host to propagate on or it would go extinct. This may be the host which the phage is epigenetically protected against, or it may be any of the  $(D - 1)$  other bacterial strains in the system if the phage can escape restriction during the infection. The probability for a phage of variant  $j$  to successfully propagate after adsorbing is then:

$$p_j = \frac{b_j}{B} + \sum_{i \neq j}^D \omega_i \frac{b_i}{B} \quad (4)$$

For at least one phage per burst to successfully reproduce requires  $p_j \geq 1/\beta$ . This must be true for all phages in the system, so we can construct the aggregate statement by summing over all  $D$  epigenetic variants:

$$\frac{D}{\beta} \leq \sum_{i=1}^D \frac{b_i}{B} + (D-1) \sum_{i=1}^D \omega_i \frac{b_i}{B} = 1 + (D-1) \sum_{i=1}^D \omega_i \frac{b_i}{B} \quad (5)$$

If we rearrange this statement we get:

$$\frac{D}{\beta} \leq \frac{1 - \sum_{i=1}^D \omega_i \frac{b_i}{B}}{1 - \beta \sum_{i=1}^D \omega_i \frac{b_i}{B}} \quad (6)$$

This expression highlights that if there is no leakage ( $\omega_i \sim 0$ ), diversity is indeed limited by  $\beta$ , and each bacterial strain reaches a population of  $\delta/(\eta(\beta-1))$ . As the leakage increases, the phage begin to reliably propagate by attacking the hosts that are not epigenetically protected against them. This, in turn, leads to stronger suppression of the strains with the weaker RM systems (large  $\omega$ 's) which broadens the biomass distribution and allows new bacterial strains to claim the unoccupied space in the ecosystem.

### 3.3 The effect of parameters on diversity

To explore how our robust our results about diversity are to varying the different parameter values, we introduce a simplification of the dynamical model above to save computational time. Instead of simulating the entire set of dynamical equations, we evolve the ecosystem in the following iterative way:

1. Add bacterial strain with growth rate,  $\gamma_i$ , and RM strength,  $\omega_i$ , chosen from a given probability distribution  $P(\gamma, \omega)$ .
2. Compute steady-state.
3. Remove non-viable strains.
4. Go to step 1.

More details are provided in the Methods and Materials section, but essentially this step-by-step process is an approximate way of determining which added strains will survive and which will go extinct at any given point in time. This procedure is an approximate representation of the full dynamics. However, it is a very good approximation when the interval between additions is large (see supplementary figure S2 and S3).

### 3.3.1 The influence of $P(\gamma, \omega)$

The “steepness” or slope of the cloud of points in 2H is an indication of the average cost to a bacterium of increasing the efficiency of its RM system. In reality, even very efficient RM systems reduce the growth rate of a bacterium by at most a few percent [9], which would correspond to a much steeper cloud of points. Therefore, in figure 3A, we decrease the cost of having an RM system as described in section 2.3 in Methods. Note that despite the large change in slope, the underlying correlations does not change substantially. The results are surprisingly robust to such changes and despite the now smaller variation in growth rates the system retains the increased diversity and the high bacterial biomass, while the phage population does drop substantially as the distribution steepens.

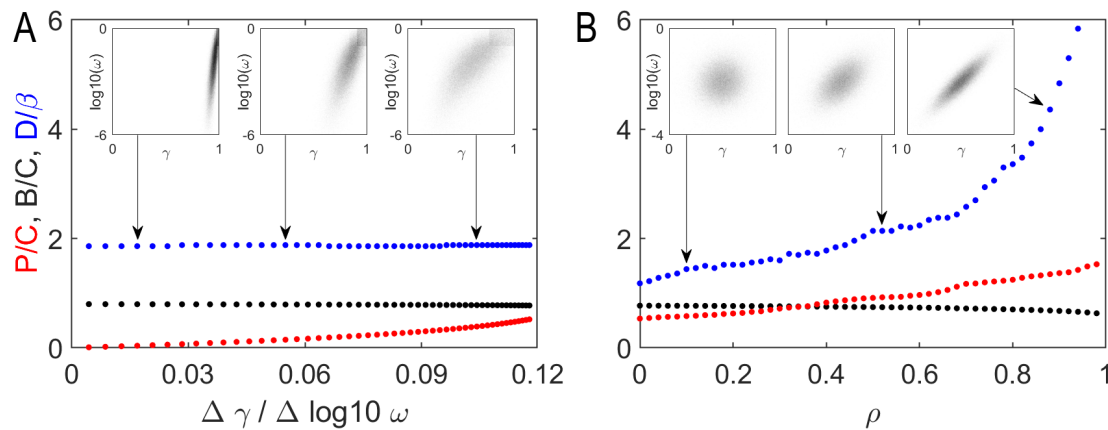


Figure 3: Distribution shape and diversity. We vary the  $\gamma - \omega$  distributions and use the iterative algorithm to predict the final populations. (A) By limiting the cost of individual RM systems, we change the overall slope of the  $(\gamma, \omega)$  distribution without changing the correlations. (B) Using a joint normal distribution for  $\gamma$  and  $\log_{10}(\omega)$  can change from an uncorrelated distribution to a fully correlated one by increasing  $\rho$  without changing the mean values. (A-B) Insets indicate the sampled values of  $(\gamma, \omega)$  at the indicated locations.

Next we test how increased correlations changes the outcome of the simulation. In figure 3B, we draw our  $(\gamma, \log_{10}(\omega))$  values from a multivariate normal distribution where we change the covariance between the random variables via the parameter  $\rho$  such that when  $\rho = 0$ , the distributions are uncorrelated, and when  $\rho = 1$  the distributions are fully correlated.

In this imagined case, the ecosystem changes drastically as the correlations are increased. When uncorrelated ( $\rho \sim 0$ ), we recover the diversity limit of  $D \sim \beta$  as from figure 2B, but as the correlations increase we observe a drastic increase in the diversity which reaches more than 5 times the burst size. Interestingly, this increase in diversity is accompanied by only a moderate increase in the phage population.

From these tests and the analytical solution we speculate that it is the limited access to high  $\gamma$  - low  $\omega$  pairs that are key to high diversity ecosystems. In fact, we find that even when in the uncorrelated RM system case, we can still obtain diversity higher than the burst size when we no longer allow for strong RM systems — in such a case the bacterial density distribution becomes broad (see supplementary figure S4).

### 3.4 The effect of phage and bacterial parameters on diversity

In figure 4, we use the iterative algorithm to investigate how the diversity varies as we vary the bacterial death rate  $\alpha$ , the phage burst size  $\beta$ , the phage decay rate  $\delta$ , and the phage adsorption rate  $\eta$ . Notice that the values we plot are final values obtained by the simulated evolution after adding  $10^5$  species. These values are not the steady-state values as adding more species would change the results. What the graph does show, is the relative difficulty for the ecosystems of bacteria to overcome the phage predation. We see that the diversity and population levels will vary as we change the parameters but in general it remains very stable across a wide range of



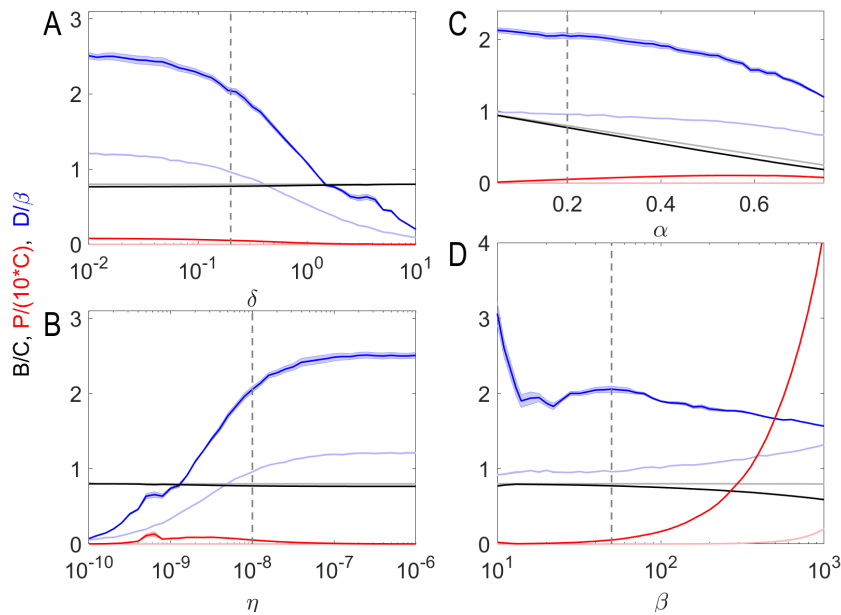


Figure 4: Diversity dependence on parameter values. We show the final populations and diversity after  $10^5$  rounds of the iterative algorithm under varying parameters: (A) the phage decay rate,  $\delta$ , (B) the phage adsorption rate,  $\eta$ , (C) the bacterial death rate,  $\alpha$  and (D) the phage burst size,  $\beta$ . Dark (light) colours indicate simulations using (un)correlated RM strengths and bacterial growth rates. Shaded areas indicate standard error of the mean over 5 replicates but this error is often too small to see on the figure. Dashed lines indicate the default parameter values.

parameters. In this stage, the phages are suppressed to low levels and the bacteria are close to their maximal population size while retaining a high diversity of  $\sim 2\beta$ . From panel (C), we see that as the bacterial death rate increases, a corresponding loss of bacterial biomass and diversity follows since the bacteria need higher growth rates to compensate for the increased death rate.

The last three panels (A-B, D) show varying phage parameters. In general, we can consider these changes as either an increase in virulence or decreases in virulence. For example, the virulence decreases as the phage death rate increases (A), the adsorption rate decreases (B), or when the burst size decreases (D). For the death rate and the adsorption rate, the results are similar: as the phage becomes less potent, the bacteria are suppressed less and begin to compete with each other causing a decrease in the diversity through increased competitive exclusion. This is not observed when the burst size decreases, but note that diversity is measured in terms of the burst size and therefore the diversity *does* drop but not more than expected ( $D \sim 2\beta$ ).

In contrast, we can consider what happens when the virulence increases by (A) decreasing death rates, (B) increasing adsorption rate, or (D) increasing burst size. Again the death rate change and the adsorption rate change yield similar results where the phage are less likely to decay before finding a bacterium to infect and thus they more strongly control the ecosystem and cement the high diversity state. The burst size of the phage does influence the stability of the ecosystem as the increase in burst size gives a smaller increase in diversity than expected. This is because the increases phage pressure requires stronger RM systems to overcome. Since the underlying probability distribution for  $P(\gamma, \omega)$  does not change, it takes longer for fit bacteria to arise — making it harder for the bacteria to reach the high diversity state.

### 3.5 Strength of the RM system as a co-determinant of survival

In our simulations, e.g., figure 2B, bacteria can coexist despite different bacterial strains having up to 10 % variation in growth rates (figure 2D). The question is now: why does this wide distribution not lead to the slower-growing bacteria being out-competed by competitive exclusion?

One key observation from (S22), is that the amount of biomass each point contributes to the whole increases not just with the bacterial growth rate, but also with increasing RM strength.

In figure 4, we plot the biomass of each bacterial strain as a function of 1) the effective growth rates  $\gamma'_i$ , 2) the strength of the RM systems  $1/\omega_i$ , and 3) the ratio  $\gamma'_i/\omega_i$ . In panel (A-C), we use data at the end of the simulated evolution (figure 2B) and in panels (D-F), we use a simulated batch culture with 100 bacterial strains.

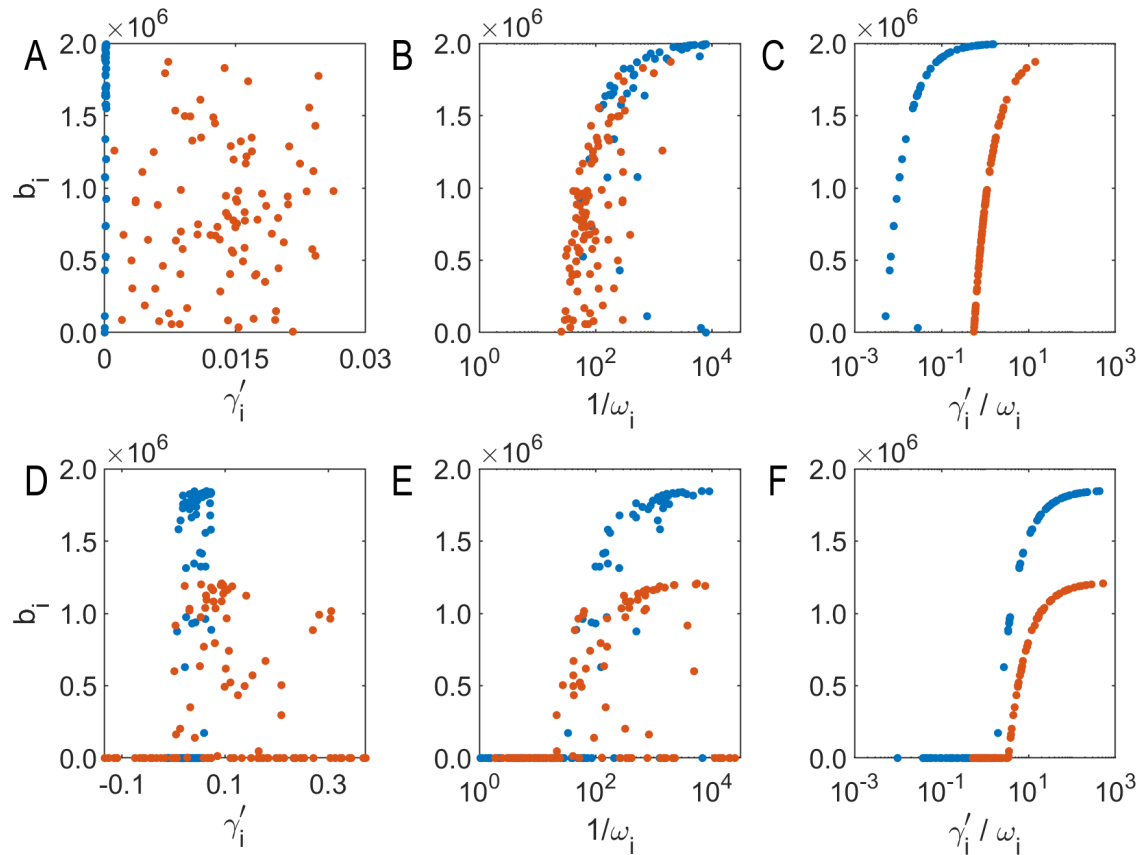


Figure 5: Predictors of survival. Using the data from the end of the simulated evolution, we show the population size as a function of (A) Effective growth rates  $\gamma'_i = \gamma_i(1 - B/C) - \alpha$ , (B) Strength of RM system  $1/\omega_i$ , and (C) ratio of effective growth rate  $\gamma'_i$  to RM imperfection  $\omega_i$ . (D-F) same as (A-C) but using data from a simulated batch culture consisting of 100 bacterial strains. In all cases red (blue) points correspond to (un)correlated RM systems.

In both cases, we see that neither the growth rate nor the RM strength can fully explain the achieved biomasses of the bacterial strains, but that the combination of the two can. The batch culture is particularly interesting since it clearly shows that the strains with the highest growth rates and those with the strongest RM systems go extinct.

As one would expect, we see in the simulations that the effective growth rate must be positive for the bacteria to survive, but beyond that, the growth rate has little influence on the achieved biomass. For the surviving bacteria, the strength of the RM systems becomes a good predictor for the biomass but the product of the effective growth rate and the RM strength is an almost perfect predictor of the bacterial biomass.

## 4 Discussion

We have demonstrated that when there is a trade-off between bacterial growth rate and the efficiency of its RM defence system, a remarkably large number bacterial strains can coexist on the same resource. The work of Ref. [9] does provide evidence for such a trade-off and our simulations (see figure 3A) indicate that even a weak trade-off of this kind is sufficient to produce coexistence of many strains.

We show that in these phage limited ecosystems, it is not the growth rate of the bacteria alone that determines whether they go extinct or not - this is determined instead by a combination of growth rate *and* the strength of the RM system (see figure 5). Further, for strains that do not go extinct, this combination -  $\gamma'/\omega$  - determines their relative biomass. To be more precise, there exists an “extinction curve” in the growth rate vs RM efficiency plane (the red curve in Fig. 2), which marks out a region where strains would go extinct, and the shape of this curve depends on  $\gamma'/\omega$  and not  $\gamma$  alone. In the region where they do not go extinct, the distance from the extinction curve (proportional to  $\gamma'/\omega$ ) determines the biomass as shown in Fig. 5. In our model ecosystem, the distance may thus be thought of as a measure of “fitness” of a bacterial strain, and  $\gamma'/\omega$  acts as a quantitative (albeit nonlinear) measure of fitness. There are several other systems where such trade-offs between growth and defense control the fitness: One more example from the world of bacteria is when bacteria are repeatedly exposed to periods of high nutrient availability followed by periods of starvation[17]; another example, in plants, is described in [18].

Our results mirror the lessons learned from studies of coexistence of species, in particular how the partitioning of predators can have a stabilizing effect on a competitive ecosystem[19]. Since the bacterial strains in our analysis are highly similar, resource partitioning is assumed to be negligible, and thus it is the strong partitioning of predators generated by the RM systems that causes substantial niche differentiation and allows for high diversity. We speculate that this effect is enhanced in our system because the predator changes type when preying on its secondary targets. This results in a negative feedback loop such that when a prey (bacterial host) grows in abundance, it is preyed on more by both its secondary predators and primary predators (phage of different epigenetic types) as expected. However, the predation turns the secondary predators into primary predators which then strongly inhibit the prey. Thus, in our system, *intraspecies* interactions result in a strong *interspecies* interaction - an interesting twist on the predator partitioning discussed in [19].

We further show that such high diversity ecosystems are quite robust to the parameters of the phage, but that this state starts breaking down when the phage is not virulent enough to control the bacterial biomass (see figure 4). In such cases, the ecosystems are not phage limited but rather resource-limited and the competitive exclusion drives the system towards a winner-takes-all ecosystem. This behaviour is similar to what is found in the discrete spatial model of Ref. [12], where the phages almost die out and cannot control the bacterial biomass. Here they find, as we do, that the diversity of the ecosystem is very limited.

The increased difficulty with reaching high diversity ecosystems observed at some parameter values is consistent with the “narrowing staircase of coexistence” discussed in Ref. [20]. In our system, the “narrowing” of the sustainable parameter space is explicitly visualized by the red line in the  $\gamma - \omega$  space as it moves down and to the right over time. The gradual evolution of the ecosystems drives increasing constraints on the parameters of new bacterial strains.

In their paper[21], the authors suggest that the wide distribution of growth rates found in nature can be explained by a receptor based trade-off between nutrient uptake (favoring receptors) and phage infections, as phage often use nutrient receptors as sites of recognition and infection. The bacterium can thus limit its phage exposure at the cost of a loss of growth rate by expressing fewer receptors. Here we demonstrate an alternate mechanism by which correlations in  $\gamma - \omega$  space can widen the distribution of growth rates. Since RM systems are very efficient, even very small correlation yields large reduction in phage exposure.

Bacteria typically evolve resistance to phages by loss of a receptor, and while such resistance gives perfect protection against a phage variant, the loss of receptor cause a substantial loss in growth rate[22]. In our framework, this can be understood as a zero- $\omega$  - small- $\gamma$  strain, and in accordance with what Ref. [11] found, such strains will be out-competed in a sufficiently diverse ecosystem. This highlights that, in comparison with reducing receptor expression levels, RM systems may be able to achieve similar efficiency of defence against a phage but at a much smaller cost.

Furthermore, when communities of several species coexist, several synergistic mechanisms such as cross-feeding may come into play that may counteract the growth-rate loss of investing in the RM system[23, 24, 25].

RM systems are believed to influence the speciation of bacteria due to their ability to control horizontal gene transfer and thus genetically isolate the bacteria[7]. In addition, it is believed that RM systems diversify the genetic pool by promoting DNA recombination[9]. From the perspective

of evolution, in the ecosystems that develop in our model, we may speculate that the high diversity allows for more genetic variation, and thus presumably a faster rate of evolution, while the wide distribution in growth rates gives more room for bacteria to evolve, e.g., by reciprocal sign epistasis[26].

We end by noting again that the life of bacteria is a harsh life, and that they exist in a constant stressful environment, be that stress stemming from predation, from starvation, or from toxins / antibiotics, or abiotic factors. Our work reinforces the idea that in order to understand what constitute “fit” bacteria in a realistic environment, one must consider the trade-off between growth and defence, or growth and strategies to alleviate other stressors.

## 5 Funding

R.S.E has received funding for this project from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme under grant agreement No 740704.

S.K. thanks the Simons Foundation for funding, as well as the Department of Atomic Energy, Government of India, under project no. 12-R&D-TFR-5.04-0800.

## References

- [1] J. R. Brum et al. “Seasonal time bombs: dominant temperate viruses affect Southern Ocean microbial dynamics”. In: *The ISME Journal* 10.2 (Feb. 2016), pp. 437–449. ISSN: 1751-7362. DOI: 10.1038/ismej.2015.125. URL: <http://dx.doi.org/10.1038/ismej.2015.125>. 20<http://www.nature.com/articles/ismej2015125>.
- [2] M. K. Mirzaei and C. F. Maurice. “Ménage à trois in the human gut: interactions between host, bacteria and phages”. In: *Nature Reviews Microbiology* 15.7 (July 2017), pp. 397–408. ISSN: 1740-1526. DOI: 10.1038/nrmicro.2017.30. URL: <http://dx.doi.org/10.1038/nrmicro.2017.30>20<http://www.nature.com/articles/nrmicro.2017.30>.
- [3] J. S. Weitz, H. Hartman, and S. A. Levin. “Coevolutionary arms races between bacteria and bacteriophage”. In: *Proceedings of the National Academy of Sciences* 102.27 (July 2005), pp. 9535–9540. ISSN: 0027-8424. DOI: 10.1073/pnas.0504062102. URL: <http://www.pnas.org/cgi/doi/10.1073/pnas.0504062102>.
- [4] T. W. Berngruber, S. Lion, and S. Gandon. “Evolution of suicide as a defence strategy against pathogens in a spatially structured environment”. In: *Ecology Letters* 16.4 (Apr. 2013). Ed. by P. Thrall, pp. 446–453. ISSN: 1461023X. DOI: 10.1111/ele.12064. URL: <http://doi.wiley.com/10.1111/ele.12064>.
- [5] K. S. Makarova, Y. I. Wolf, and E. V. Koonin. “Comparative genomics of defense systems in archaea and bacteria”. In: *Nucleic Acids Research* 41.8 (Apr. 2013), pp. 4360–4377. ISSN: 1362-4962. DOI: 10.1093/nar/gkt157. URL: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkt157>.
- [6] C. Xue and N. Goldenfeld. “Coevolution Maintains Diversity in the Stochastic “kill the Winner” Model”. In: *Physical Review Letters* 119.26 (2017), pp. 1–5. ISSN: 10797114. DOI: 10.1103/PhysRevLett.119.268101. arXiv: 1706.02666.
- [7] K. Vasu and V. Nagaraja. “Diverse Functions of Restriction-Modification Systems in Addition to Cellular Defense”. In: *Microbiology and Molecular Biology Reviews* 77.1 (2013), pp. 53–72. ISSN: 1092-2172. DOI: 10.1128/MMBR.00044-12. eprint: <https://mmbbr.asm.org/content/77/1/53.full.pdf>. URL: <https://mmbbr.asm.org/content/77/1/53>.
- [8] M.-È. Dupuis et al. “CRISPR-Cas and restriction–modification systems are compatible and increase phage resistance”. In: *Nature Communications* 4.1 (Oct. 2013), p. 2087. ISSN: 2041-1723. DOI: 10.1038/ncomms3087. URL: <http://www.nature.com/articles/ncomms3087>.
- [9] M. Pleška et al. “Bacterial autoimmunity due to a restriction-modification system”. In: *Current Biology* 26.3 (2016), pp. 404–409. ISSN: 09609822. DOI: 10.1016/j.cub.2015.12.041.

- [10] M. De Paepe and F. Taddei. “Viruses’ life history: Towards a mechanistic basis of a trade-off between survival and reproduction among phages”. In: *PLoS Biology* 4.7 (June 2006). Ed. by T. Kirkwood, pp. 1248–1256. ISSN: 15457885. DOI: 10.1371/journal.pbio.0040193. URL: <http://dx.plos.org/10.1371/journal.pbio.0040193>.
- [11] S. A. Frank. “POLYMORPHISM OF BACTERIAL RESTRICTION-MODIFICATION SYSTEMS: THE ADVANTAGE OF DIVERSITY”. In: *Evolution* 48.5 (Oct. 1994), pp. 1470–1477. ISSN: 00143820. DOI: 10.1111/j.1558-5646.1994.tb02189.x. URL: <http://doi.wiley.com/10.1111/j.1558-5646.1994.tb02189.x>.
- [12] L. Pagie. “Individual- and Population-based Diversity in Restriction-modification Systems”. In: *Bulletin of Mathematical Biology* 62.4 (July 2000), pp. 759–774. ISSN: 00928240. DOI: 10.1006/bulm.2000.0177. URL: <http://link.springer.com/10.1006/bulm.2000.0177>.
- [13] K. Sneppen et al. “Restriction modification systems as engines of diversity”. In: *Frontiers in Microbiology* 6 (2015), p. 528. ISSN: 1664-302X. DOI: 10.3389/fmicb.2015.00528. URL: <https://www.frontiersin.org/article/10.3389/fmicb.2015.00528>.
- [14] G. F. Gause. “The Struggle for Existence”. In: *Soil Science* 41.2 (Feb. 1936), p. 159. DOI: 10.1097/00010694-193602000-00018.
- [15] G. Hardin. “The Competitive Exclusion Principle”. In: *Science* 131.3409 (Apr. 1960), pp. 1292–1297. ISSN: 0036-8075. DOI: 10.1126/science.131.3409.1292. URL: <https://www.sciencemag.org/lookup/doi/10.1126/science.131.3409.1292>.
- [16] R. S. Eriksen and S. Krishna. *Data from: Defense versus growth in a hostile world - Lessons from phage and bacteria*. GitHub online repository. May 2020. DOI: 10.5281/zenodo.3842682. URL: <https://github.com/RasmusSkytte/CorrelatedRestrictionModificationSystems>.
- [17] Y. Himeoka and N. Mitarai. “Dynamics of bacterial populations under the feast-famine cycles”. In: *Physical Review Research* 2.1 (Mar. 2020), p. 013372. ISSN: 2643-1564. DOI: 10.1103/PhysRevResearch.2.013372. arXiv: 1910.05673. URL: <https://link.aps.org/doi/10.1103/PhysRevResearch.2.013372>.
- [18] S. W. Makhabu et al. “Resource availability and plant antiherbivore defense.” In: *Plant ecology*. 184.1 (2006), pp. 895–899. ISSN: 1385-0237. DOI: 10.1007/978-94-010-9117-6.
- [19] P. Chesson and J. J. Kuang. “The interaction between predation and competition”. In: *Nature* 456.7219 (2008), pp. 235–238. ISSN: 14764687. DOI: 10.1038/nature07248.
- [20] J. O. Haerter, N. Mitarai, and K. Sneppen. “Phage and bacteria support mutual diversity in a narrowing staircase of coexistence”. In: *The ISME Journal* 8.11 (Nov. 2014), pp. 2317–2326. ISSN: 1751-7362. DOI: 10.1038/ismej.2014.80. URL: <http://dx.doi.org/10.1038/ismej.2014.80>  
<http://www.nature.com/articles/ismej201480>.
- [21] T. F. Thingstad. “Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems”. In: *Limnology and Oceanography* 45.6 (Sept. 2000), pp. 1320–1328. ISSN: 00243590. DOI: 10.4319/lo.2000.45.6.1320. URL: <http://onlinelibrary.wiley.com/doi/10.4319/lo.2000.45.6.1320/abstract%7B%5C%7D5Cnpapers2://publication/uuid/7D985307-E9B4-4904-9A7F-A71877950469%20http://doi.wiley.com/10.4319/lo.2000.45.6.1320>.
- [22] M. Middelboe. “Bacterial growth rate and marine virus-host dynamics”. In: *Microbial Ecology* 40.2 (2000), pp. 114–124. ISSN: 00953628. DOI: 10.1007/s002480000050.
- [23] G. D’Souza et al. “Less is more: Selective advantages can explain the prevalent loss of biosynthetic genes in bacteria”. In: *Evolution* 68.9 (2014), pp. 2559–2570. ISSN: 15585646. DOI: 10.1111/evo.12468.
- [24] S. Pande et al. “Fitness and stability of obligate cross-feeding interactions that emerge upon gene loss in bacteria”. In: *The ISME Journal* 8.5 (May 2014), pp. 953–962. ISSN: 1751-7362. DOI: 10.1038/ismej.2013.211. URL: <http://www.nature.com/articles/ismej2013211>.
- [25] A. Goyal and S. Maslov. “Diversity, Stability, and Reproducibility in Stochastically Assembled Microbial Ecosystems”. In: *Phys. Rev. Lett.* 120 (15 Apr. 2018), p. 158102. DOI: 10.1103/PhysRevLett.120.158102. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.120.158102>.

- [26] F. J. Poelwijk et al. “Empirical fitness landscapes reveal accessible evolutionary paths”. In: *Nature* 445.7126 (Jan. 2007), pp. 383–386. ISSN: 0028-0836. DOI: 10.1038/nature05451. URL: <http://www.nature.com/articles/nature05451>.

# Supplementary Material for: Defense versus growth in a hostile world - Lessons from phage and bacteria

Rasmus Skytte Eriksen<sup>1</sup> & Sandeep Krishna<sup>2</sup>

<sup>1</sup> Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark

<sup>2</sup> Simons Centre for the Study of Living Machines, National Centre for Biological Sciences,  
Tata Institute of Fundamental Research, Bangalore, India.

May 25, 2020

## S1 Parameter values used in simulations

Throughout the paper, we use the parameter values listed in the table below unless otherwise specified.

Name	Value	Description
$C$	$10^8$	Carrying capacity of the ecosystem
$\eta$	$10^{-8}$	Adsorption rate of the phages
$\alpha$	0.2	Dilution rate of the ecosystem
$\beta$	50	Burst size of the phages
$\delta$	0.2	Decay rate of the phages
$T$	$10^4$	Interval between bacterial additions
$lb$	4	log 10 of lower bound for omega
$\mu$	4	Parameter controlling correlation
$t$	$2 - \left(\frac{1}{2}\right)^{\left(\frac{1}{\mu}-1\right)}$	Scaling factor for the growth rates
$M$	5	Number of starting species

Table S1: Default parameter values used in the simulations.

## S2 Generating correlated RM systems

When generating a correlated  $(\gamma, \omega)$  pair in our simulations we use the following algorithm:

1. Draw  $n$  from a Poisson distribution with mean  $\mu$ .
2. For each  $r = 1, 2, \dots, n$ , draw two independent and identically distributed numbers  $u_{1r}$  and  $u_{2r}$  from a uniform distribution between  $0 \leq u \leq 1$ .

3. Compute  $\gamma_r$  and  $\omega_r$  defined as:

$$\omega_r = 10^{-\frac{lb}{\mu} u_{1r}}$$

$$\gamma_r = 1 - t \cdot u_{2r}$$

4. Compute the  $(\gamma, \omega)$  pair by:

$$\gamma = \prod_{r=1}^n \gamma_r$$

$$\omega = \prod_{r=1}^n \omega_r$$

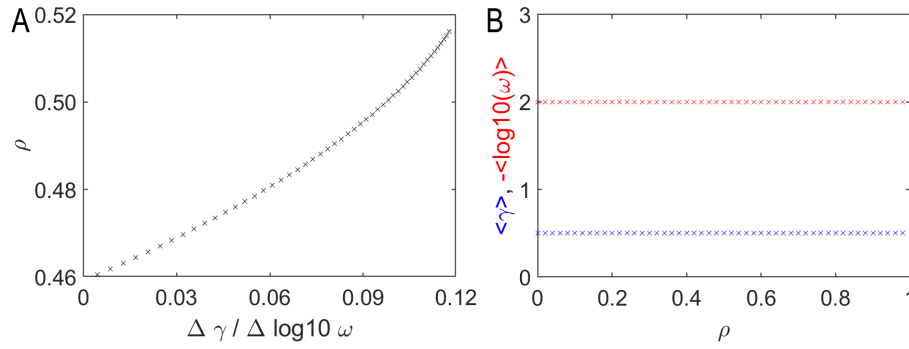


Figure S1: Change of correlation coefficient and marginal expectation values. In the main text figure 3, we investigate how the underlying distribution  $P(\gamma_i, \omega_i)$  changes the dynamics. Here we show (A) how the correlation coefficient changes as we change the marginal distributions in figure 3A, and (B) how the marginal distribution change as we change the correlation coefficient in figure 3B.

Here  $lb$  sets the lower bound on  $\omega$ , and  $\mu$  sets the mean of  $\gamma$  as well as the correlation coefficient (see table S1 for default values).

This somewhat indirect method to generate correlations was used for historical reasons - arising from a separate study of strains containing multiple independent RM systems. In this context, we emphasise that this should not be taken as a proposed mechanism for why such correlations arise in bacteria with RM systems; it is simply one parametrization of a  $P(\gamma, \omega)$  distribution that exhibits correlations and conveniently allows us to change the “steepness” of the cost of RM systems by varying the parameter  $t$ .

In the main text, we compare such correlated  $(\gamma, \omega)$  pairs with uncorrelated pairs and, to allow direct comparison, the expectation values should be identical for the uncorrelated pairs and the correlated pairs. We achieve this by setting  $t = 2 - \left(\frac{1}{2}\right)^{\left(\frac{1}{\mu}-1\right)}$  which is derived by setting  $\langle(1-tu)^\mu\rangle = \langle u \rangle$  when  $\langle u \rangle = \frac{1}{2}$

## S2.1 Changing the cost of RM systems

In the main text, we test the effect of changing the “steepness” of the distribution by scaling the growth rates. This is done by sweeping the parameterization variable  $t$ . This parameter changes the marginal expectation values while keeping the correlation coefficient  $\rho$  almost unchanging. Across the sweep of parameter  $t$ , the “slope” of the distribution changes – we quantify the slope as  $\Delta\gamma/\Delta\log_{10}(\omega)$  – and the correlations drop from  $\rho \approx 0.51$  at the shallowest distribution to  $\rho \approx 0.46$  at the steepest distribution (see figure S1 for details).

## S2.2 Changing the correlation coefficient

In our investigation, we also test the effect of changing the correlation coefficient without changing the marginal expectation values. Here we use a multivariate normal distribution to draw the values of  $\gamma$  and  $\log_{10}(\omega)$ .

The multivariate normal distribution  $P_{MVN}(\gamma, \log_{10}(\omega))$  is defined by a location vector  $\vec{\mu}$ , which sets the means of the marginal distributions:

$$\vec{\mu} = \begin{bmatrix} 0.5 \\ -2 \end{bmatrix} \quad (\text{S1})$$

By defining  $\vec{\mu}$  so, the drawn growth rates will have mean 0.5 and the drawn  $\omega$  values will have mean  $10^{-2}$ . The distribution is further defined by a covariance matrix  $\vec{\Sigma}$  which defines the spread of each of the marginal distribution and the correlations between the drawn values:

$$\vec{\Sigma} = \begin{bmatrix} \sigma_\gamma^2 & \rho\sigma_\gamma\sigma_{\log_{10}(\omega)} \\ \rho\sigma_\gamma\sigma_{\log_{10}(\omega)} & \sigma_{\log_{10}(\omega)}^2 \end{bmatrix} \quad (\text{S2})$$



Where  $\sigma_\gamma = 0.125$  and  $\sigma_{\log_{10}(\omega)} = 0.75$ , defines the marginal spreads, and  $\rho$  is the Pearson correlation coefficient. Note that we truncate the distribution by redrawing pairs where either  $\gamma < 0$  or  $\log_{10}(\omega) > 0$ .

We show in figure S1B that, as expected, the means of the marginal distributions do not change as we sweep the correlations coefficient from 0 to 1.

### S3 Analytical solution to the model

To begin, we remind the reader that the model equations we want to find the steady-state of are the following:

$$\dot{b}_i = \gamma_i b_i (1 - B/C) - \eta_i b_i p_i - \eta_i \omega_i b_i (P - p_i) - \alpha b_i \quad (\text{S3})$$

$$\dot{p}_i = \eta_i \beta_i b_i p_i + \eta_i \omega_i \beta_i b_i (P - p_i) - \eta_i p_i B - \delta_i p_i \quad (\text{S4})$$

In the steady state, the growth and decay terms balance:

$$b_i (\gamma_i (1 - B/C) - \alpha) = \eta_i b_i p_i + \eta_i \omega_i b_i (P - p_i) \quad (\text{S5})$$

$$\eta_i \beta_i b_i p_i + \eta_i \omega_i \beta_i b_i (P - p_i) = \eta_i p_i B + \delta_i p_i \quad (\text{S6})$$

We start by combining eq. (S3) and eq. (S4)

$$\beta_i b_i (\gamma_i (1 - B/C) - \alpha) = \eta_i p_i B + \delta_i p_i \quad (\text{S7})$$

We then rearrange eq. (S5) and assume  $p_i > 0$  to obtain the ratio:

$$\frac{b_i}{p_i} = \frac{\eta_i B + \delta_i}{\beta_i (\gamma_i (1 - B/C) - \alpha)} \quad (\text{S8})$$

Next, we isolate  $p_i$  in (S3) when  $b_i > 0$

$$\gamma_i (1 - B/C) - \alpha = \eta_i p_i + \eta_i \omega_i P - \eta_i \omega_i p_i \quad (\text{S9})$$

$$\gamma_i (1 - B/C) - \alpha = \eta_i (1 - \omega_i) p_i + \eta_i \omega_i P \quad (\text{S10})$$

$$p_i = \frac{\gamma_i (1 - B/C) - \alpha - \eta_i \omega_i P}{\eta_i (1 - \omega_i)} \quad (\text{S11})$$

$$p_i = \frac{\gamma_i (1 - B/C)}{\eta_i (1 - \omega_i)} - \frac{\alpha}{\eta_i (1 - \omega_i)} - P \frac{\omega_i}{(1 - \omega_i)} \quad (\text{S12})$$

The total phage mass is then found by perform a sum on eq. (S10)

$$P = (1 - B/C) \sum_j \frac{\gamma_j}{\eta_j (1 - \omega_j)} - \alpha \sum_j \frac{1}{\eta_j (1 - \omega_j)} - P \sum_j \frac{\omega_j}{(1 - \omega_j)} \quad (\text{S13})$$

$$P = (1 - B/C) \frac{\sum_j \frac{\gamma_j}{\eta_j (1 - \omega_j)}}{1 + \sum_j \frac{\omega_j}{(1 - \omega_j)}} - \alpha \frac{\sum_j \frac{1}{\eta_j (1 - \omega_j)}}{1 + \sum_j \frac{\omega_j}{(1 - \omega_j)}} \quad (\text{S14})$$

Next we isolate  $b_i$  in eq. (S4)

$$\eta_i \beta_i b_i p_i + \eta_i \omega_i \beta_i b_i (P - p_i) = \eta_i p_i B + \delta_i p_i \quad (\text{S15})$$

$$b_i p_i (1 - \omega_i) + b_i \omega_i P = p_i B \frac{1}{\beta_i} + p_i \frac{\delta_i}{\eta_i \beta_i} \quad (\text{S16})$$

We assumed  $p_i > 0$ , so we can divide through by  $p_i$ .

$$b_i (1 - \omega_i) + \frac{b_i}{p_i} \omega_i P = B \frac{1}{\beta_i} + \frac{\delta_i}{\eta_i \beta_i} \quad (\text{S17})$$

We then insert eq. (S6) into eq. (S15).

$$b_i (1 - \omega_i) + \frac{\eta_i B + \delta_i}{\beta_i (\gamma_i (1 - B/C) - \alpha)} \omega_i P = B \frac{1}{\beta_i} + \frac{\delta_i}{\eta_i \beta_i} \quad (\text{S18})$$

We can now isolate  $b_i$  in eq. (S16)

$$b_i = B \frac{1}{\beta_i(1-\omega_i)} + \frac{\delta_i}{\eta_i\beta_i(1-\omega_i)} - \frac{\eta_i B + \delta_i}{\beta_i(\gamma_i(1-B) - \alpha)} \frac{\omega_i}{1-\omega_i} P \quad (\text{S19})$$

The total number of bacteria is then found by summing over the  $b_i$ 's.

$$B = B \sum_j \frac{1}{\beta_j(1-\omega_j)} + \sum_j \frac{\delta_j}{\eta_j\beta_j(1-\omega_j)} - \sum_j \frac{\eta_j B + \delta_j}{\beta_j(\gamma_j(1-B/C) - \alpha)} \frac{\omega_j}{1-\omega_j} P \quad (\text{S20})$$

By inserting our expression for  $P$  (eq. (S12)) we obtain the self consistent equation for  $B$ .

$$\begin{aligned} B = & B \sum_j \frac{1}{\beta_j(1-\omega_j)} + \sum_j \frac{\delta_j}{\eta_j\beta_j(1-\omega_j)} \\ & - \sum_j \frac{\eta_j B + \delta_j}{\beta_j(\gamma_j(1-B/C) - \alpha)} \frac{\omega_j}{1-\omega_j} \left( (1-B/C) \frac{\sum_j \frac{\gamma_j}{\eta_j(1-\omega_j)}}{1 + \sum_j \frac{\omega_j}{(1-\omega_j)}} \right. \\ & \left. - \alpha \frac{\sum_j \frac{1}{\eta_j(1-\omega_j)}}{1 + \sum_j \frac{\omega_j}{(1-\omega_j)}} \right) \end{aligned} \quad (\text{S21})$$

Finally, we can rearrange eq. (S16) to simple expression for the  $b_i$ 's.

$$b_i = \frac{B + \delta_i/\eta_i}{\beta_i(1-\omega_i)} \left( 1 - \frac{\eta_i\omega_i P}{(\gamma_i(1-B/C) - \alpha)} \right) \quad (\text{S22})$$

## S4 Performance of the iterative algorithm

The above analytical solution allows us to use a simplified, iterative algorithm to predict the results from the full model.

### S4.1 Comparing with the simulated evolution

In figure S2, we run the iterative algorithm in conditions which mimic those use full dynamical model shown in figure 2. The overall behaviour is remarkably similar to that found in figure 2.

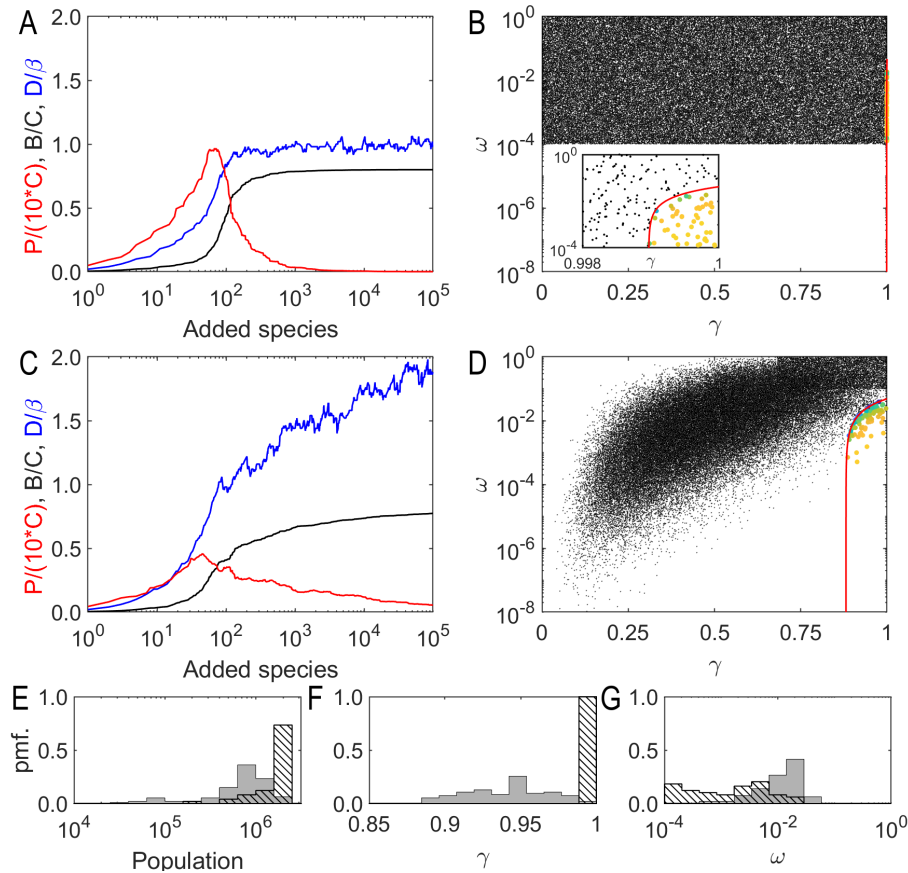


Figure S2: The iterative algorithm on an open system. We use our iterative algorithm to model the simulated evolution of figure 2B. We test the algorithm using an uncorrelated distribution for the RM systems (A-B) and correlated RM systems (C-D). Both cases are qualitatively similar to the results from the full dynamical model. (E-F) shows the distributions of (E)  $\gamma$  and (F)  $\omega$  at the end of the simulation with (un)correlated RM systems shown as (hatched) solid bars.

### S4.2 Finite species

We also more directly compare the predictions of the dynamical system with the predictions of our iterative map. To do so, we limit the number of possible  $(\gamma, \omega)$  pairs available in the system. In doing so, the dynamical system will revisit  $(\gamma, \omega)$  pairs until a steady-state is reached, rather than continuously generating new  $(\gamma, \omega)$  pairs which prevent the system from ever settling down fully. Since the dynamical system now visits a known set of  $(\gamma, \omega)$  pairs, we can directly compare the performance of our map with the dynamical system for different set sizes (see figure S3). Using the setup described above, we vary the number of  $(\gamma, \omega)$  pairs from 10 to 1000 and compare the prediction between the two methods in two conditions: (i) using the same  $(\gamma, \omega)$  values in both methods and (ii) using different realizations of  $(\gamma, \omega)$  values from the same distribution. In the former case, we see an almost perfect match between the two methods while we in the latter case

see good agreement for the diversity and total bacterial population but some variation in the total phage population.

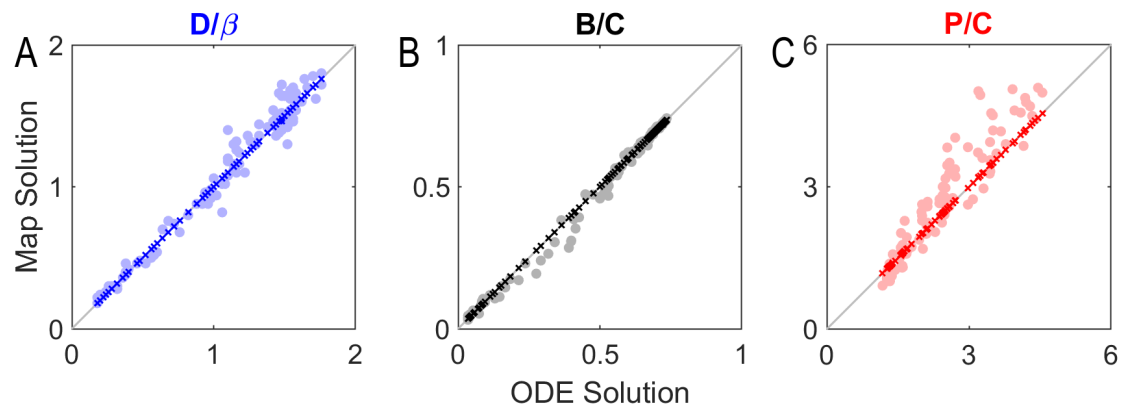


Figure S3: Comparison of the map and the dynamical system using same  $(\gamma, \omega)$  values (dark crosses) and identically distributed  $(\gamma, \omega)$  values (light circles). (A)  $D/\beta$ , (B)  $B/C$ , (C)  $P/C$ .

## S5 High diversity with the uncorrelated RM systems

From our steady-state solution in section S3, we now understand that a high diversity state can be induced by limiting the number of high  $\gamma$  - low  $\omega$  species. This means that we should be able to get a high diversity state from an uncorrelated distribution if the imperfections are drawn from the more limited interval:  $10^{-2}$  to 1. In figure S4, we test this scenario and find that in fact we do get diversity substantially higher than  $\beta$  under these conditions.

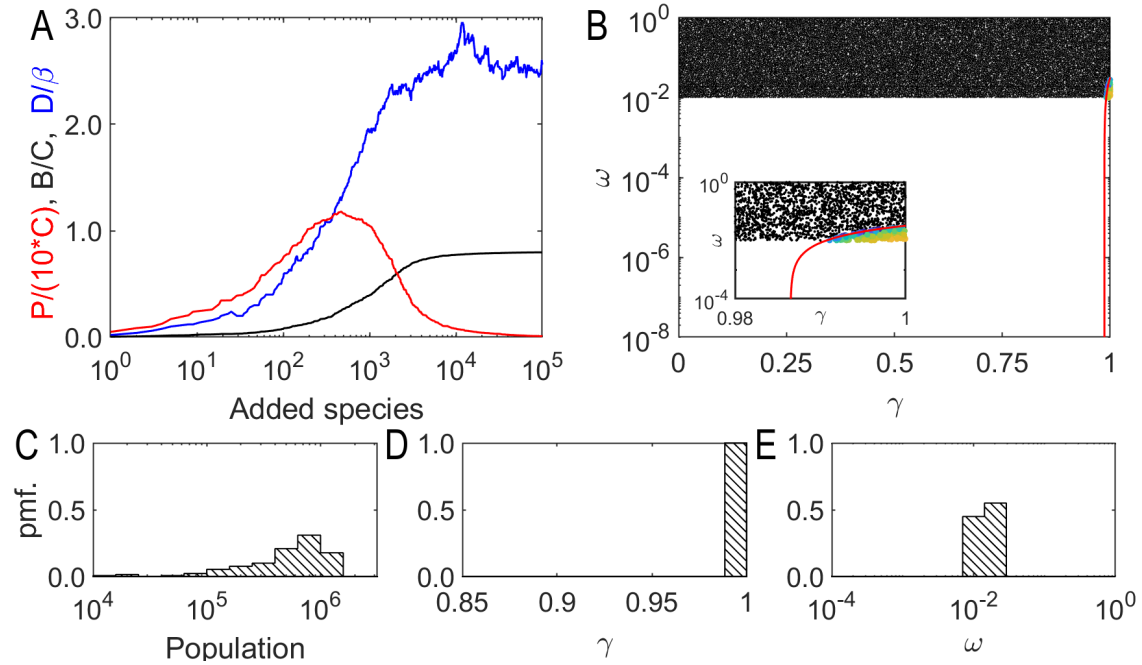


Figure S4: High diversity with uncorrelated RM systems. Using the iterative algorithm, we use an uncorrelated  $\gamma - \omega$  distribution where the RM imperfections  $\omega$  have values between  $10^{-2}$  and 1. (A) Simulated evolution. (B)  $\gamma - \omega$  distribution (C-E) Final distributions.