Constrained proteome allocation affects coexistence in models of competitive microbial communities

Leonardo Pacciani-Mori[1,2], Samir Suweis[1], Amos Maritan[1,†] & Andrea Giometto[2,3,4,†]

[1]*Department of Physics and Astronomy "Galileo Galilei", University of Padua, Via Francesco Marzolo 8, 35131 Padua (Italy)*

[2]*Department of Physics, Harvard University, 17 Oxford St, Cambridge 02138 MA*

[3]*Department of Molecular and Cellular Biology, Harvard University, 52 Oxford St, Cambridge 02138 MA*

[4]*School of Civil and Environmental Engineering, Cornell University, 220 Hollister Dr, Ithaca 14853 NY*

†*These authors contributed equally.*

Microbial communities are ubiquitous and play crucial roles in many natural processes. Despite their importance for the environment, industry and human health, there are still many aspects of microbial community dynamics that we do not understand quantitatively. Recent experiments have shown that the metabolism of species in a community is intertwined with its composition, suggesting that properties at the intracellular level such as the allocation of cellular proteomic resources must be taken into account when describing microbial communities with a population dynamics approach. In this work we reconsider one of the theoretical frameworks most commonly used to model population dynamics in competitive ecosystems, MacArthur's consumer-resource model, in light of experimental evidence showing how pro-

teome allocation affects microbial growth. This new framework allows us to describe community dynamics at an intermediate level of complexity between classical consumer-resource models and biochemical models of microbial metabolism, accounting for temporally-varying proteome allocation subject to constraints on growth and protein synthesis in the presence of multiple resources, while preserving analytical insight into the dynamics of the system. We first show experimentally that proteome allocation needs to be accounted for to properly understand the dynamics of even the simplest microbial community, i.e. two bacterial strains competing for one common resource. We then study the model analytically and numerically to determine the conditions that allow multiple species to coexist in systems with arbitrary numbers of species and resources.

Introduction

Microbes are among the most abundant life forms on Earth, in terms of biomass[1]. They are found in almost every habitat of our planet, and continue to surprise us with their ability to survive in places that were thought to be inhospitable and barren. For example, microbial communities have been found in the deep terrestrial subsurface[2,3], and it has been estimated that the first five kilometers beneath the Earth's surface could be habitable for them[4]. Because of their ubiquity, microbial communities play fundamental roles in countless natural processes of vital importance, from the digestion and overall health of their host organism[5] to the regulation of bio-geochemical cycles[6,7]. Despite their importance, however, we still know very little about the fundamental mechanisms that regulate microbial communities, partly because we are only able to grow in the lab a very

small fraction of all the microbes found in nature[8], and partly because microbial communities are complex, non-linear systems[9] whose dynamics is difficult to predict. For these reasons, scientists from many disciplines have long been fascinated by the challenging theoretical questions posed by the study of microbial communities' structure and dynamics, and serious efforts are being made to understand how competition[10–12] and metabolic interactions[13,14] allow such systems to maintain the very high levels of biodiversity found in nature.

Recent experimental studies have shown that the structure and composition of microbial communities are tightly linked to the metabolism of the species that comprise them[15,16] (e.g., communities with different taxonomic compositions can nevertheless exhibit the same metabolic functional structure[17,18]). We can therefore speculate that the ways with which microbes uptake and use different resources for growth and proliferation can affect the dynamics of an entire community. Resource uptake is constrained by the other functions that cells must perform to grow and proliferate, and the balance between such functions is governed by the allocation of the internal resources of the cell (e.g., the proteome, the set of proteins expressed by a cell) to different tasks. It is therefore important to understand how microbial community dynamics is influenced by the proteome allocation of its members, and new insights in this direction might help us make more powerful predictions of how microbial communities assemble and evolve[19,20]. However, accounting for the dynamics of metabolism and gene expression of each species in a microbial community explicitly (e.g., via community flux balance analysis[21]) can be very challenging, and the large dimensionality of the mathematical models that attempt to do so poses limits to our understanding of the dynamics of microbial communities and of the fundamental properties that affect species coexistence.

Scott *et al.*[22] showed that, despite the complexity of bacterial metabolism, there are simple relationships that link the fraction of the proteome allocated for nutrient uptake and protein synthesis to the growth rate of bacteria grown in isolation, and that reducing these fractions by forcing cells to express a useless protein reduces their growth rate. Such relationships are very powerful because they describe how bacterial growth is influenced by proteome allocation and gene expression without requiring an explicit representation of the underlying molecular mechanisms. These relationships, which were also based on earlier observations by Schaechter *et al.*[23] on how the ribosomal component of the proteome of a microbial species scales with the growth rate, have recently been applied in many different contexts[24] and were instrumental in improving our knowledge of microbial metabolism, both experimentally[25] and computationally[26]. However, as the experiments by Scott *et al.*[22] were performed with single-species populations in exponential phase, it is still an open question if their approach can also be used to describe the population dynamics of different interacting microbial species competing for multiple resources.

In this work we fill this gap by linking the results by Scott *et al.*[22] to one of the most widely adopted theoretical frameworks for modeling competitive ecosystems, MacArthur's consumer-resource model[27–29], and tailor it to describe the dynamics of microbial species (or strains) competing for one or more resources. MacArthur's model describes how the population abundances of N_S species competing for a common pool of N_R resources change over time, and has been used in several recent studies[10–12,30–32] to understand under which conditions multiple species can coexist while competing for few resources. We show that generalizing Scott *et al.*'s proteome-growth relationships and including them into a consumer-resource framework allows us to build

a community dynamics model where all parameters can in principle be measured experimentally and have a precise biological interpretation. This "consumer-proteome-resource" model describes community dynamics at an intermediate level of complexity between classical consumer-resource models and biochemical models of microbial metabolism[21]. By adopting such an intermediate level of complexity and realism, we can take into account the dynamics of gene expression and microbial metabolism, while preserving analytical insights on the microbial community dynamics and identifying the key intracellular properties affecting species coexistence.

In the next section we describe our "consumer-proteome-resource" model for a general number of species/strains and resources. We then apply this model to the simplest implementation of an experimental microbial community, i.e. two *Escherichia coli* strains competing for glucose as the only carbon source. The experiment's results highlight how the proposed theoretical framework accounting for proteome allocation allows us to understand species population dynamics in competitive communities. We finally study (both analytically and numerically) the "consumer-proteome-resource" model for communities composed of arbitrary numbers of species and resources to identify the conditions allowing the coexistence of multiple species in the community. A discussion section and some future perspectives conclude this work.

Results

The consumer-proteome-resource model The phenomenological framework proposed by Scott *et al.*[22] prescribes that the proteome of a single microbial species growing on a single resource

can be minimally divided into three sectors: one dedicated to nutrient uptake and metabolism (the "P-sector"), one dedicated to ribosomal proteins responsible for biomass production and growth (the "R-sector"), and a third one dedicated to housekeeping functions (the "Q-sector"), which was shown to be incompressible[22]. Naming φ^P, φ^R and φ^Q the proteome fractions corresponding to these sectors, we must have $\varphi^P + \varphi^R + \varphi^Q = 1$ (since all proteome fractions must sum to one), and Scott *et al.* have shown that φ^P and φ^R are linear functions of the species' growth rate g, i.e:

$$\varphi^P = \frac{\rho}{\overline{\kappa}^n(c)} g \ , \tag{1a}$$

$$\varphi^R = \frac{\rho}{\kappa^t} g + \varphi^0 \ . \tag{1b}$$

Here ρ is a conversion factor (equal to the ratio between the total mass of the ribosomal proteins and the total RNA mass of the cells) and $\overline{\kappa}^n(c) = \kappa^n \cdot r(c)$, where $r(c) = c/(K+c)$ is a Michaelis-Menten (or Monod) function, which encapsulates the dependence on the resource concentration c. K is the half-saturation constant of the resource and κ^n is the "nutritional capacity" of the (only) limiting resource. This parameter measures how much protein biomass is produced per unit ribosomal mass per unit time, and therefore depends on how much energy the resource contains and how efficiently the microbial species can metabolize it (see Supporting Online Material in Scott *et al.*[22] for a molecular interpretation of κ^n). The parameter κ^t is the "translational capacity" of the microbial species, measuring how much protein biomass is produced per unit ribosomal mass per unit time; it is therefore a measure of how fast the microbial species expresses its genome to synthesize proteins. Finally, φ^0 is the incompressible core of φ^R, representing the fact that ribosomal

proteins are present in the cells also when microbes are not growing. All these parameters involve the ribosomal mass of the microbial species because the measurements by Scott *et al.*[22] were done by assaying the RNA/protein ratio in exponentially growing *Escherichia coli*.

The way we generalize Scott *et al.*'s[22] framework to a system with multiple species and resources is shown in Figure 1a: we assume that the proteome fraction φ_σ^P allocated by species σ to nutrient uptake and metabolism can be further sub-divided into smaller fractions $\varphi_{\sigma i}^P$, each one dedicated to one of the available resources. In other words, we call $\varphi_{\sigma i} = \varphi_{\sigma i}^P$ the proteome fraction allocated by species σ to the uptake and metabolization of resource i. With this choice, in order to ensure that the sum of all the proteome fractions is equal to one we must have:

$$\varphi_\sigma^Q + \varphi_\sigma^R + \sum_{i=1}^{N_R} \varphi_{\sigma i}^P = 1 \ . \tag{2}$$

This constraint represents the finiteness of a microbial species' proteome, i.e. the fact that each species in a community has a limited "proteomic budget" that can be spent for all the necessary biological functions: for example, if more proteins need to be produced for metabolizing complex substrates (i.e., if the nutrient fraction φ_σ^P increases), then a smaller part of the proteome will be available for biomass production (i.e., the ribosomal fraction φ_σ^R decreases). In order to achieve optimal growth, microbial species must balance this trade-off[22].

In Figure 1b we show a schematic representation of the "classic" consumer-resource model. Within this framework, a community is a set of N_S species that can only uptake some (or all) of the N_R available resources. The species' growth is determined by the types and the amout of resources

they uptake, and is also regulated by a "maintenance cost" (representing the fact that species need to uptake a minimum amount of resources in order to survive). The resources, on the other hand, can be thought of as organic substrates that are supplied to the system with given (constant) rates s_i. The model describes explicitly the dynamics of both species and resources, and the ways in which the species uptake the available substrates are encoded in parameters that in the literature are called "metabolic strategies" or "resource preferences". Therefore, in the consumer-resource framework the interactions between species are indirect and mediated by the abundance of resources, and other types of direct inter-specific interactions (like the exchange of useful metabolic byproducts) are not included.

Figure 1c depicts schematically the assumptions underlying our consumer-proteome-resource model. Each species σ uptakes resource i with a rate $J_{\sigma i}$ that is proportional to the proteome fraction $\varphi_{\sigma i}$. Then, resource i is converted into a growth term $g_\sigma^{(i)}$ proportionally to the uptake rate $J_{\sigma i}$. For our purposes, we assume that all resources in the system are substitutable, so that they can be used interchangeably and we can write the total growth rate g_σ of a given species as the sum of all the terms $g_\sigma^{(i)}$. Eventually (see Materials and Methods for all the detailed computations) we obtain a mathematical model that has the traditional structure of a consumer-resource model, but with the added merit of describing population dynamics using parameters and variables that have a precise biological meaning at the intracellular scale of the system and that can in principle be measured experimentally[22]. For example, "metabolic strategies" in our framework correspond to the proteome fractions $\varphi_{\sigma i}$. Additionally, our model also includes the proteome finiteness constraint given by Eq (2). The final expression of this constraint in our equations is significantly

different from similar constraints that have been studied in the consumer-resource framework[31,33], and all the results that we will show are ultimately a repercussion of this finiteness of the microbial proteome, as encoded in our model.

The most important consequence of the proteome finiteness constraint is that it implies that the proteome fractions $\varphi_{\sigma i}$ *cannot* be fixed parameters, but must change over time and therefore must be dynamical variables (see Materials and Methods). It is thus necessary to introduce some form of dynamics on the proteome fractions that each species allocates for nutrient uptake and metabolization. Our approach is to require that all $\varphi_{\sigma i}$ evolve dynamically with a characteristic timescale in order to maximize the instantaneous growth rate[33] of species σ in what we call an "adaptive process", while we ensure that the proteome finiteness constraint is satisfied at all times. The mathematical details are discussed in the Materials and Methods section.

Application of the model to an experimental competitive microbial community. We have applied our consumer-proteome-resource model to the simplest case of a competitive community, i.e. two species competing for one common resource. We competed experimentally two strains of *E. coli* grown in a liquid minimal medium with glucose as the sole carbon source, transferring a fraction of the community to fresh medium daily and measuring the relative abundance of the two strains at each transfer (see Materials and Methods). The two strains had the same genetic background and expressed constitutively from their genome two different fluorescent proteins, which allowed us to measure their relative abundance via flow-cytometry. Additionally, we introduced in strain $\sigma = 1$ a plasmid containing a red fluorescent protein (RFP) whose expression could be con-

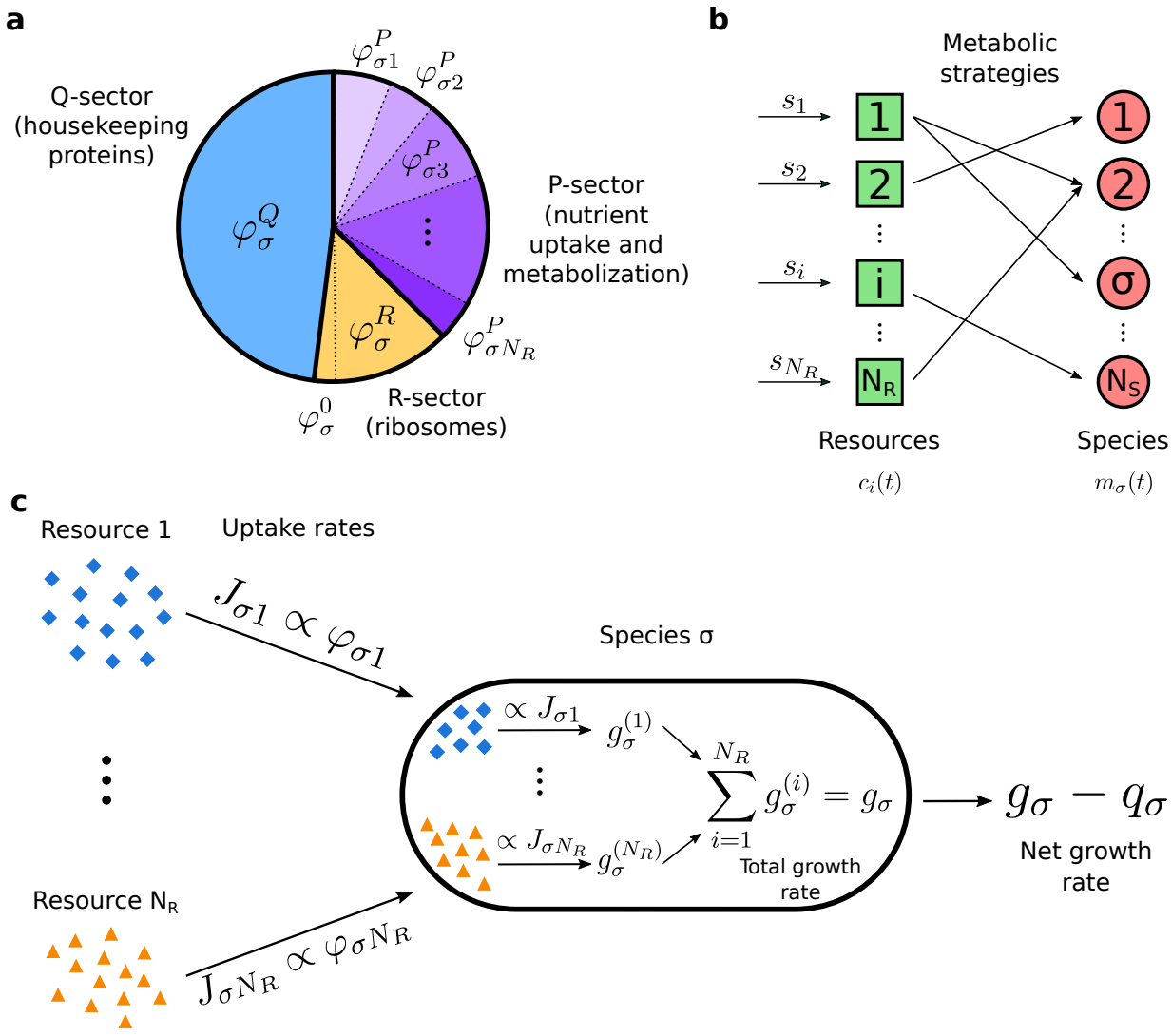

Figure 1: Assumptions made in building our model. **(a):** Generalization of the proteome subdivision introduced by Scott et al.[22] to the case of N_R resources: we assume that the sector allocated by species σ for nutrient uptake and metabolization is subdivided into smaller fractions $\varphi_{\sigma i} = \varphi^P_{\sigma i}$, each one dedicated to a specific resource. **(b):** Schematic representation of a consumer-resource model with N_R resources and N_S species. In this framework, the concentrations c_i of the resources and the biomass densities m_σ of the species are described by systems of coupled differential equations. Resources are supplied with (constant) rates s_i, and are uptaken by the species (arrows represent resource flows). The ways in which each species uptakes resources are encoded in the "metabolic strategies". **(c):** Assumptions used to write the equations of our consumer-resource model with proteome allocation. Each species σ uptakes resource i with a rate $J_{\sigma i}$ proportional to the proteome fraction $\varphi_{\sigma i}$. Then, each resource contributes a growth term $g^{(i)}_\sigma$ (proportional to the resource uptake rate) to the total growth rate of the cells. The net growth rate of species σ is the difference between this total growth rate and a maintenance cost q_σ.

trolled by adding Isopropyl β-D-1-thiogalactopyranoside (IPTG, a molecular mimic of allolactose that cannot be metabolized by *E. coli*) to the medium. Thus, by varying the concentration of IPTG in the medium we could vary the proteome allocation of strain 1 by forcing it to produce a useless protein. The experiment is sketched in Figure 2.

Applying our model to such a simple community, and introducing assumptions consistent with our experimental conditions (e.g., the fact that the strains are grown in medium-rich conditions, and that they share the same genetic background), our model predicts that the selective advantage $\mathcal{S}$ of strain 1 over strain 2 (a measure for the difference in fitness between the two strains) is:

$$\mathcal{S} = \frac{d}{dt} \ln \frac{f}{1-f} \propto \Phi_1 - \Phi_2 \ , \tag{3}$$

where f is the relative abundance (or frequency) of strain 1, and Φ_σ is the total proteome fraction allocated by species σ for metabolism and growth (see the Materials and Methods section for all the details). In other words, the selective advantage of strain 1 over strain 2 is proportional to the difference between these two proteome fractions.

According to Eq (3), the ratio between the relative abundances of the two strains decreases or grows exponentially with time, depending on the sign of $\Phi_1 - \Phi_2$, which then sets the outcome of competition: for example, if $\Phi_2 > \Phi_1$ (i.e., strain 2 allocates a larger fraction of its proteome to metabolism and biomass production than strain 1) then $\mathcal{S} < 0$ and strain 2 outcompetes strain 1. Coexistence between the two strains is possible uniquely when $\Phi_1 = \Phi_2$ and thus $\mathcal{S} = 0$. The system therefore exhibits two regimes where only one of the two strains survives (competitive

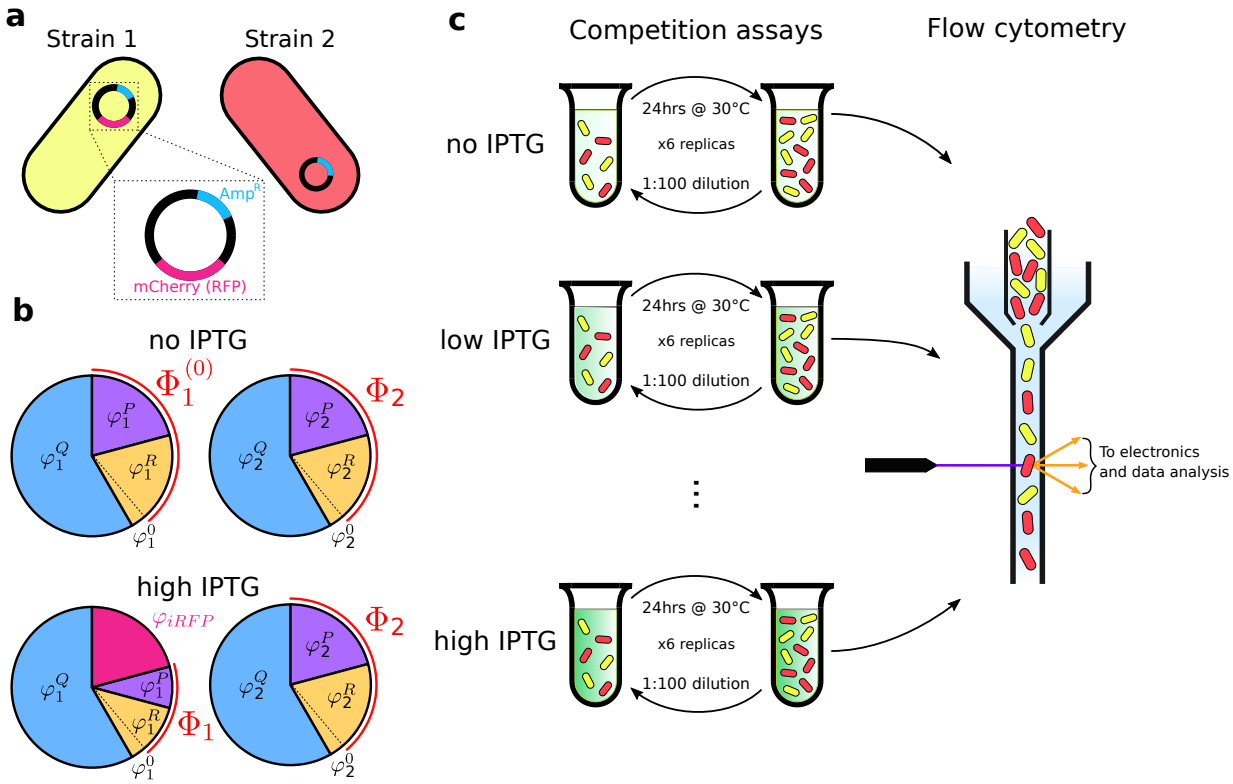

Figure 2: Schematic representation of the experiment. **(a):** Two *E. coli* strains were used: strain 1 constitutively expresses a yellow fluorescent protein (mVenus) and carries a plasmid with the ampicillin resistance cassette (cyan Amp^R in the plasmid magnification) and a red fluorescent protein (RFP), mCherry (magenta), under the control of the *trc* promoter, an hybrid of the *trp* and *lac* promoters. Strain 2 constitutively expresses a red fluorescent protein (mKate2Hyb) and carries a plasmid with the ampicillin resistance cassette. **(b):** Proteome allocation of the two strains at different concentration of IPTG in the medium. When strain 1 grows in the presence of IPTG, a fraction φ_{iRFP} of the strain's proteome is allocated for the expression of the RFP mCherry, thus reducing the fraction Φ_1 allocated for metabolism and growth. The proteome allocation of strain 2, instead, is not affected by the presence of IPTG. **(c):** The two strains were co-cultured in minimal medium at different IPTG concentrations, they were diluted daily into fresh medium and their relative abundance was measured at every transfer via flow-cytometry.

exclusion), separated by the coexistence point $\Phi_1 = \Phi_2$. Eq (3) thus connects a well known concept of population genetics, the selective advantage in exponentially growing populations, with the differential proteome allocation $\Phi_1 - \Phi_2$ between microbial strains. According to this model, a 1% difference in proteome allocation for growth between two strains (i.e., $\Phi_1 - \Phi_2 = 1\%$) corresponds to a selection coefficient of $\mathcal{S} \approx 1.1 \cdot 10^{-2}$ 1/h (see Materials and Methods for details on this estimation). Considering now that we can force strain 1 to produce a useless RFP, we can investigate how changing its proteome allocation affects the outcome of competition. Indicating with φ_{iRFP} the fraction of proteome allocated by strain 1 to the synthesis of the IPTG-inducible RFP, the proteome fraction allocated for nutrient uptake and growth is $\Phi_1 = \Phi_1^{(0)} - \varphi_{iRFP}$ (with $\Phi_1 = \Phi_1^{(0)}$ in the absence of IPTG), and we find that the selective advantage $\mathcal{S}$ is predicted to decay linearly with φ_{iRFP} as $\mathcal{S} = \alpha - \beta \cdot \varphi_{iRFP}$ with α and β positive constants (see the Materials and Methods section for all details and the explicit expression of $\mathcal{S}$ in this case).

As shown in Figure 3a (magenta data points), the experimental selective advantage $\mathcal{S}$ decreases linearly with the production rate of the IPTG-inducible RFP of strain 1 over a broad range or RFP production rates (the mean cell's fluorescence measured after 8 h at 105 μM IPTG is 22 times higher than at 0 μM IPTG, Figure 3d), which are proportional to φ_{iRFP}. In the absence of IPTG and at low concentrations of it, strain 1 outcompetes strain 2 ($\mathcal{S} > 0$). At an IPTG concentration of approximately 30 μM, the two strains coexist by maintaining a stable relative fraction. At IPTG concentrations larger than 30 μM, strain 1 is outcompeted by strain 2 (i.e., $\mathcal{S} < 0$). By manipulating the proteome allocation of strain 1 we are thus able to explore the expected regimes of the system. Figure 3a also shows that strain 1 has a fitness advantage over strain 2 in the ab-

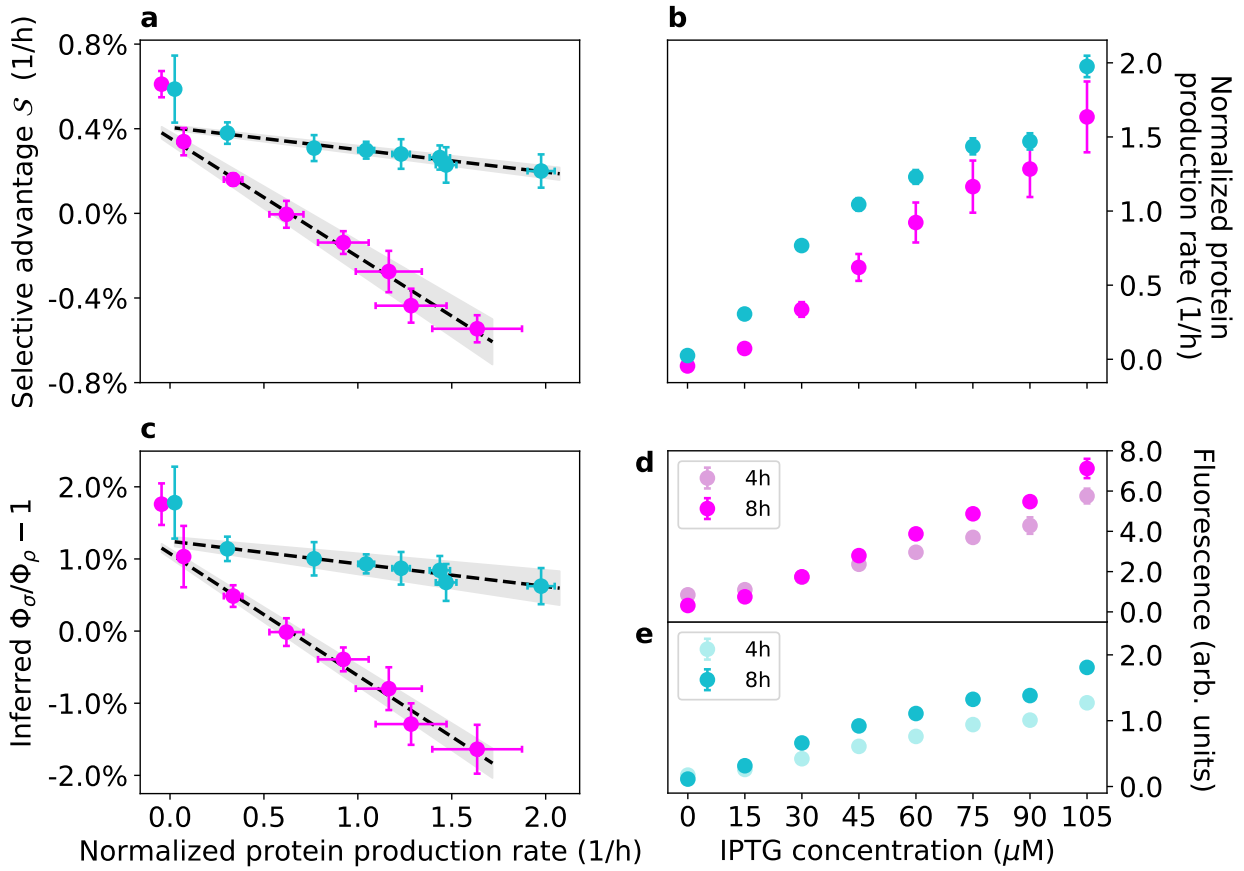

Figure 3: Experimental results. Magenta points represent data from experiments with strains 1 and 2. Cyan points represent data from experiments with strains 3 and 4, where strain 3 expresses constitutively mKate2Hyb and the IPTG-inducible Venus yellow fluorescent protein (YFP) and strain 4 expresses mVenus constitutively (see Materials and Methods and Figure S.1). Error bars represent two standard deviations. Note that the normalized protein production rates are not directly comparable across magenta and cyan points (see Materials and Methods). **a:** The experimental selection coefficient $\mathcal{S}$ (y axis) decreases linearly with the normalized production rate of the inducible protein, measured as the temporal variation of the mean cell fluorescence signal at different concentrations of the inducer IPTG, accounting for dilution of such protein via cell division (see Materials and Methods). The grey band represent the 68% confidence interval of the linear fit. The time series of $\ln[f/(1-f)]$ for the two experiments are reported in Figures S.2 and S.3. **b:** Induced protein production rates as functions of IPTG concentration (see Materials and Methods). **c:** Inferred values of the ratios Φ_1/Φ_2 and Φ_3/Φ_4 (minus one) as functions of the induced protein production rate (normalized). Also shown are the linear fits of the data with their 68% confidence interval. **d:** Mean (induced) red fluorescence of strain 1 at 4h and 8h after inoculation in the conditions used for our experiment (see Materials and Methods). **e:** Mean (induced) yellow fluorescence of strain 3 at 4h and 8h after inoculation in the conditions used for our experiment (see Materials and Methods).

sence of IPTG, since $\mathcal{S} > 0$ at low protein production rates. In our theoretical framework, such an advantage implies that $\Phi_1^{(0)} - \Phi_2 > 0$. It is also possible to use our data to estimate the ratio Φ_1/Φ_2 at different protein production rates, shown in Figure 3c (see Materials and Methods for details). This ratio is approximately $\Phi_1^{(0)}/\Phi_2 \approx 1.02$ for low protein production rates and then decays linearly up to $\Phi_1/\Phi_2 \approx 0.98$.

Figure 3 also shows the results of a second experiment performed with two different strains (cyan data points). These strains have different fluorescent protein combinations with respect to strains 1 and 2 (see Materials and Methods and Figure S.1): strain 3 expresses constitutively a red fluorescent protein (mKate2Hyb) and carries a plasmid with an IPTG-inducible yellow fluorescent protein (Venus YFP), while strain 4 expresses constitutively the yellow fluorescent protein mVenus (see Materials and Methods). Also in these independent sets of experiments, the selective advantage decreases linearly as the protein production rate is increased over a broad range (the mean cell's fluorescence measured after 8 h at 105 μM IPTG is 16 times higher than at 0 μM IPTG, Figure 3e). In the first set of experiments (magenta points in Figure 3), and to a lesser degree in the second set of experiments (cyan points), the data points at the lowest production rate (i.e., at 0 μM IPTG) appear to deviate from the linear trend, and so the fits in Figure 3a-c were calculated by excluding those data points (including those points in the fit doesn't affect the results, see Figure S.20). The flow-cytometry data suggest that the average fluorescent intensity of strain 1 from the induced RFP decreased over the course of the experiment at 0 μM IPTG, which may partly explain the deviation of the first magenta point in Figure 3a from the linear trend via a reduction in protein production rate throughout the experiment at 0 μM IPTG. Another factor that may cause

deviations from a linear trend is an increased gene-expression heterogeneity between cells in the absence of IPTG, a well-known property of the lac operon whose constituent parts we have used in our genetic constructs[34], which might confer heterogeneous growth rates to different cells in the population. Please note that the normalized protein production rates of the two sets of experiments (magenta and cyan data points) are not directly comparable (see Materials and Methods).

Coexistence of multiple species An analytical and numerical analysis of our consumer-proteome-resource model in the general case of multiple species and multiple resources can provide some insights into the conditions needed for species coexistence. In particular, looking for stationary solutions of our model where all species have non-null biomass densities yields two necessary conditions (see the Materials and Methods section for all detailed expressions and computations). The first one is that the maintenance cost q_σ of species σ must be proportional to the total proteome fraction allocated for metabolism and growth, i.e. $q_\sigma \propto \Phi_\sigma$, with a species-dependent proportionality constant. This requirement is biologically reasonable, since allocating a larger fraction of the proteome to such functions requires additional energy to synthesize the necessary proteins.

The second condition resembles similar ones that have been shown to hold for consumer-resource models with metabolic trade-offs[31], and can be interpreted graphically as follows (see Materials and Methods for all the mathematical details). A system with N_R resources can be represented on an $(N_R - 1)$–dimensional simplex, where each vertex corresponds to one of the available resources; considering for example the case $N_R = 3$, the system can be represented on a triangle (i.e., a bi-dimensional simplex) as shown in Figure 4. On this simplex we can draw

the vectors $\vec{\hat{s}}$ and $\vec{\hat{\varphi}}^*_\sigma$, whose components are appropriately rescaled versions of (respectively) the resource supply rates s_i and the stationary proteome fractions $\varphi^*_{\sigma i}$ (i.e., the values of the proteome fractions once the species' populations have reached a stationary state). The second condition for species coexistence is therefore that $\vec{\hat{s}}$ must belong to the *convex hull* of the vectors $\varphi^*_{\sigma i}$, as shown in Figure 4. Notice that this condition involves the *stationary* proteome fractions $\varphi^*_{\sigma i}$. Since the model is highly non-linear, it is impossible to predict *a priori* these stationary values in general.

However, by exploring different regions of the parameter space it is possible to make predictions. In fact, the parameters that are relevant in this sense are the following: the ratios $\gamma_{\sigma i} = \kappa^n_i/\kappa^t_\sigma$ between the nutritional and translational capacities and the characteristic timescales τ_σ of the adaptive process that maximizes the growth rate g_σ in the dynamics of $\varphi_{\sigma i}$ (see the Materials and Methods section for details). These timescales τ_σ measure how fast the dynamics of the proteome fractions $\varphi_{\sigma i}$ vary: the smaller τ_σ, the faster species σ can switch between different resources. Biologically speaking, this parameter can be thought of as a measure of how "fast" the regulatory mechanisms of a microbial species can respond to changes in the availability of resources.

The first regime that we have explored is $\tau_\sigma \gg 1$ and $\gamma_{\sigma i} \sim 0$, or in other words: the species' adaptive process is very slow (i.e., they respond very slowly to changes in resource abundance) and the nutritional capacity is much smaller than the translational capacity (which happens, for example, when species are grown in very low quality nutrients). In this case the model predicts that the stationary values $\hat{\varphi}^*_{\sigma i}$ of the proteome fractions allocated by the species to nutrient uptake and metabolization change negligibly, and therefore the rescaled nutrient supply rate vector $\vec{\hat{s}}$ must

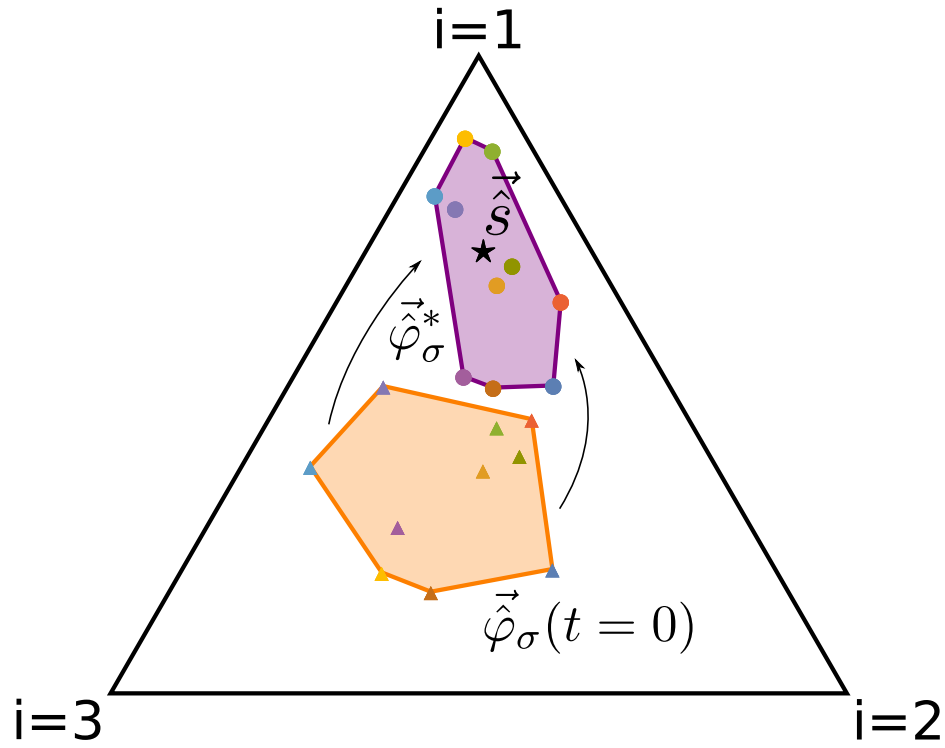

Figure 4: Graphical representation of the second condition necessary for coexistence. Here we consider a system with $N_S = 10$ species and $N_R = 3$ resources (for ease of representation). In this case, the system can be represented on a bi-dimensional simplex (i.e., a triangle) where each vertex corresponds to one of the available resources. On this simplex, we can draw the rescaled nutrient supply rate vector $\vec{\hat{s}}$ (black star) and the rescaled initial proteome fractions $\vec{\hat{\varphi}}_\sigma(t=0)$ allocated by the species to the uptake and metabolism of the resources (colored triangles); their convex hull is drawn in orange. We have also drawn the stationary values $\hat{\vec{\varphi}}^*_\sigma$ of the proteome fractions (colored circles), and their convex hull is drawn in purple. In this representation, if $\vec{\hat{s}}$ lies on one on the sides of the simplex, it means that only two of the available resources are being externally supplied to the system, and analogously if one of the $\hat{\vec{\varphi}}^*_\sigma$ lies on one of the sides of the simplex, it means that its corresponding species is uptaking and metabolizing only two of the available resources. In general, the positions of $\vec{\hat{s}}$ and $\hat{\vec{\varphi}}^*_\sigma$ depend on the relative ratios with which the resources are supplied or uptaken by the species. Using this representation, our model prescribes that coexistence will be possible if $\vec{\hat{s}}$ lies inside the convex hull of the stationary fractions $\hat{\vec{\varphi}}^*_\sigma$ as shown in this case.

lie in the convex hull of the rescaled *initial* proteome fractions $\hat{\varphi}_{\sigma i}$, as shown in Figure 5.

In the second regime we still have $\tau_\sigma \gg 1$, but now $\gamma_{\sigma i} \gtrsim 1$ (i.e., the ratios between the nutritional and translational capacities have slightly larger values). In this case the dynamics of $\hat{\varphi}_{\sigma i}$ allows the proteome fractions to move inside the simplex. Therefore, the system can reach stationary states where all species coexist even when $\vec{\hat{s}}$ is not necessarily close to the convex hull of the initial $\vec{\hat{\varphi}}_{\sigma i}$. On the other hand, we observed that if $\vec{\hat{s}}$ is too far away from the initial convex hull of $\hat{\varphi}_{\sigma i}$ there might be extinctions. However, if $\vec{\hat{s}}$ lies at an intermediate distance between these two cases, the system can reach diverse stationary states only if the resource supply rates s_i are sufficiently large. For example, multiplying each resource supply rate by a factor $x > 1$, i.e. $s_i \to x s_i$ (this rescaling leaves $\hat{s}_i$ unchanged, see Materials and Methods), we observe a transition between two different states of the system for increasing values of x: when $x \sim 1$, only one or a few strains survive, whereas for larger x the stationary biomass densities m^*_σ of the other species increase until all of them coexist. Figure 6 shows an example of such transition. This phenomenon occurs only when $\vec{\hat{s}}$ lies in specific areas of the simplex, whose shape and position can be determined numerically, but depend on the particular values of the model parameters used.

In this same regime, if $\gamma_{\sigma i}$ assume increasingly large values (which happens for example, if the species are grown in nutrients with increasingly higher qualities) coexistence will be possible even if $\vec{\hat{s}}$ lies at increasingly large distances from the convex hull of the initial $\hat{\varphi}_{\sigma i}$.

Finally, the last regime that we explored is $\tau \lesssim 1$, i.e. the adaptive process that maximizes the species' growth rates is fast. In this case, the smaller the timescales τ_σ are, the faster the

proteome fractions $\varphi_{\sigma i}$ will reach their stationary values, and coexistence will always be possible. In other words, if the timescales τ_σ have sufficiently small values, all species will be able to coexist independently of the initial values of the proteome fractions $\varphi_{\sigma i}$ and of the resource supply rates s_i. However, as the τ_σs grow, fewer and fewer species will be able to coexist. This can be shown by multiplying τ_σ by a factor $y > 0$: Figure 7 shows how the species' stationary biomasses change as y increases, and we can see that as species adaptation becomes slower (i.e., for larger y), fewer and fewer species survive in the community.

Discussion

Motivated by our experiment that shows how varying proteome allocation can have strong effects on the dynamics of a very simple microbial community, we have formulated a consumer-resource model that generalizes and incorporates the phenomenological laws discovered by Scott *et al.*[22]. In this way we have related microbial growth to proteome allocation in competitive communities and have investigated the conditions that lead to species coexistence in the presence of multiple resources.

Our model describes the population dynamics of a purely competitive microbial community, i.e. an ensemble of species that compete directly for the same pool of resources. Our main contribution is introducing a physiological, experimentally-validated constraint on the amount of resources that cells can devote to growth and metabolism in consumer-resource models with temporally-varying nutrient uptake rates, differing from previous works[31,33] that considered phenomenologi-

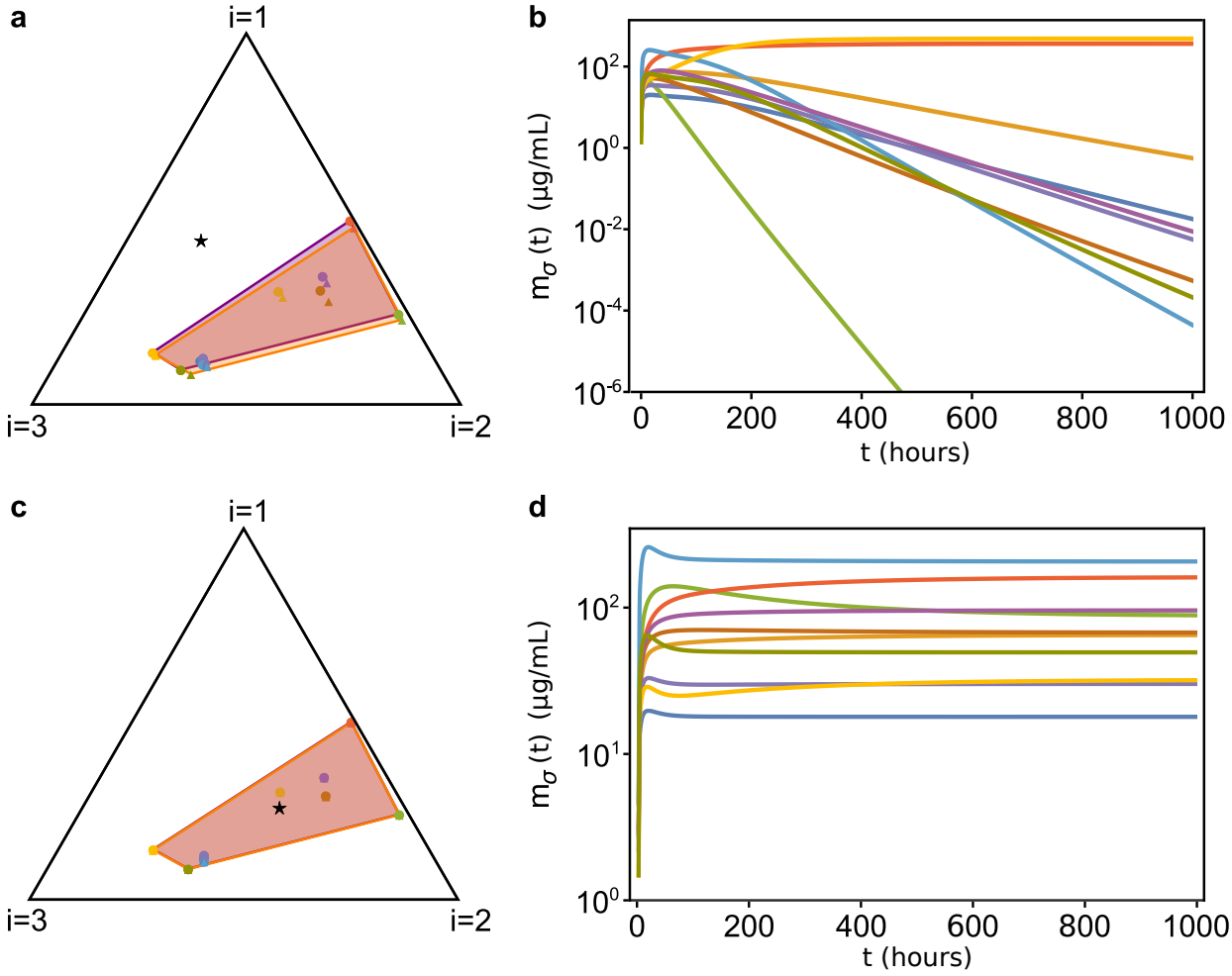

Figure 5: Temporal evolution of our consumer-proteome-resource model when $\tau_\sigma \gg 1$ and $\gamma_{\sigma i} \sim 0.1$. **(a):** Initial conditions for the $\vec{\varphi}_\sigma$ of a system with 10 species and 3 resources. We are using the graphical representation[31] shown in Figure 4: the black triangle is the simplex to which the $\hat{\varphi}_{\sigma i}$ (colored dots) and the $\hat{s}_i$ (black star) belong. The initial $\hat{\varphi}_{\sigma i}$ are represented as colored triangles, and their convex hull is colored in orange, while $\hat{\varphi}^*_{\sigma i}$ are represented as circles of the same colors, and their convex hull is in purple. With good approximation, $\hat{\varphi}^*_{\sigma i} \sim \hat{\varphi}_{\sigma i}(t = 0)$. **(b):** Time evolution of the species' biomasses m_σ relative to the case shown in (a). Since $\vec{\hat{s}}$ lies outside of the convex hull of the $\vec{\hat{\varphi}}_\sigma$, most species go extinct. **(c):** Same as in (a), but with $\vec{\hat{s}}$ belonging to the convex hull of the $\vec{\hat{\varphi}}_\sigma$. **(d):** Population dynamics of the system corresponding to the case shown in (c). In this case all species coexist. The parameters and the initial conditions were drawn from random distributions (see SI for more information). All parameters other than $\vec{\hat{s}}$ are identical in the four panels).

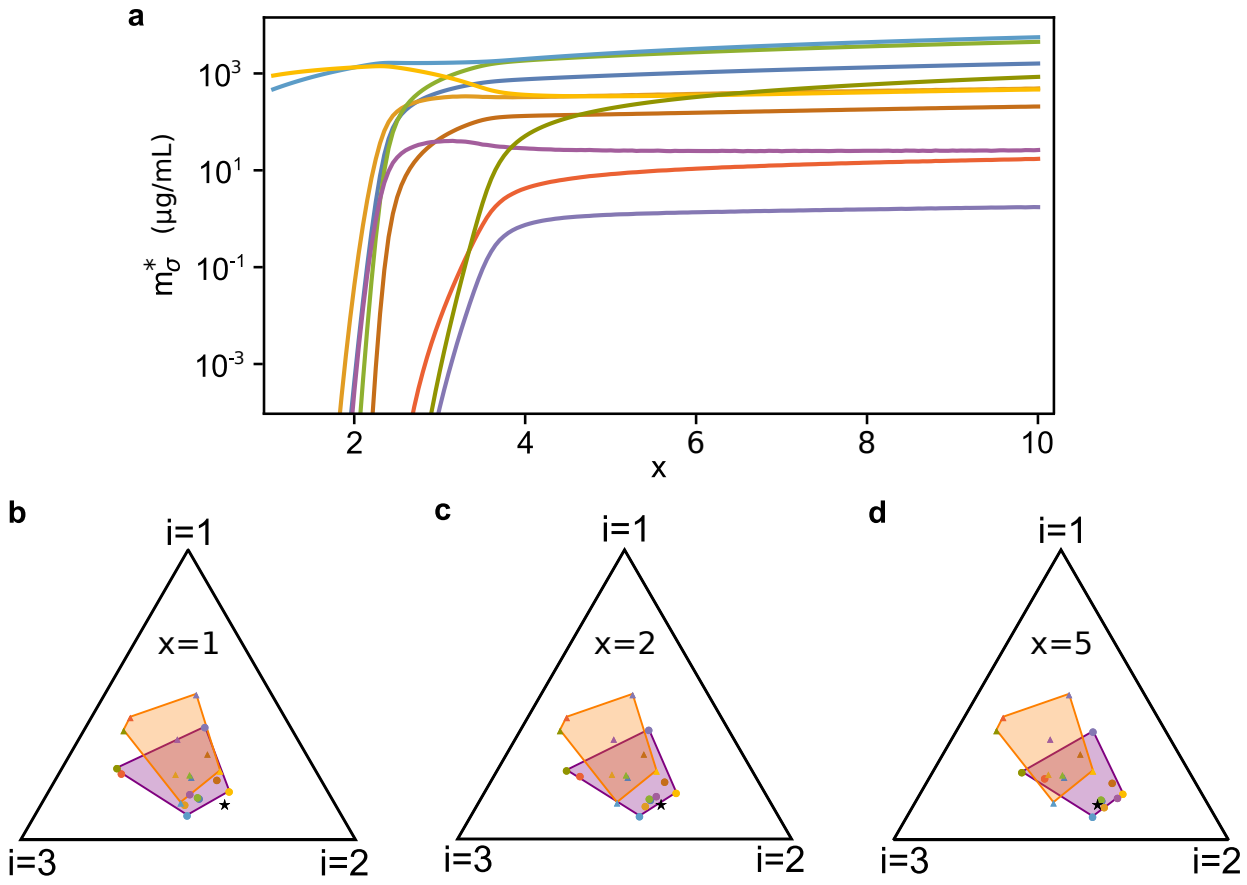

Figure 6: Species coexistence as a function of the rescaled resource supply rate $x\vec{s}$ (with $x > 1$). As for Figure 5, the $\vec{\varphi}_\sigma$ evolve according to our model with $\tau_\sigma \gg 1$, $\gamma_{\sigma i} \gtrsim 1$, $N_S = 10$ and $N_R = 3$. Here, $\vec{s}$ was drawn randomly outside the convex hull of the initial $\hat{\varphi}_{\sigma i}$ (same $\vec{s}$ for all panels) and we varied $x > 1$. **(a):** Stationary values of the species' biomasses for different values of x. When $x \simeq 1$ the system is in an oligodominant phase where only one or a few species survive, but as x grows larger the system shifts to a diverse phase in which all species coexist. Notice that the relative ratios of the stationary abundances m^*_σ are not constant as x grows. **(b-d):** Initial (orange) and stationary (purple) convex hull of the rescaled proteome fractions $\hat{\varphi}_{\sigma i}$ for different values of x. For small x, the resource supply (black star) is not large enough to allow the $\hat{\varphi}_{\sigma i}$ to move so that coexistence condition is satisfied. Increasing x (d), this becomes possible and thus all species coexist. The parameters and the initial conditions were drawn from random distributions (see SI for more information). All parameters other than $\vec{s}$ and the initial conditions $m_\sigma(0)$ and $c_i(0)$ are identical in the four panels).

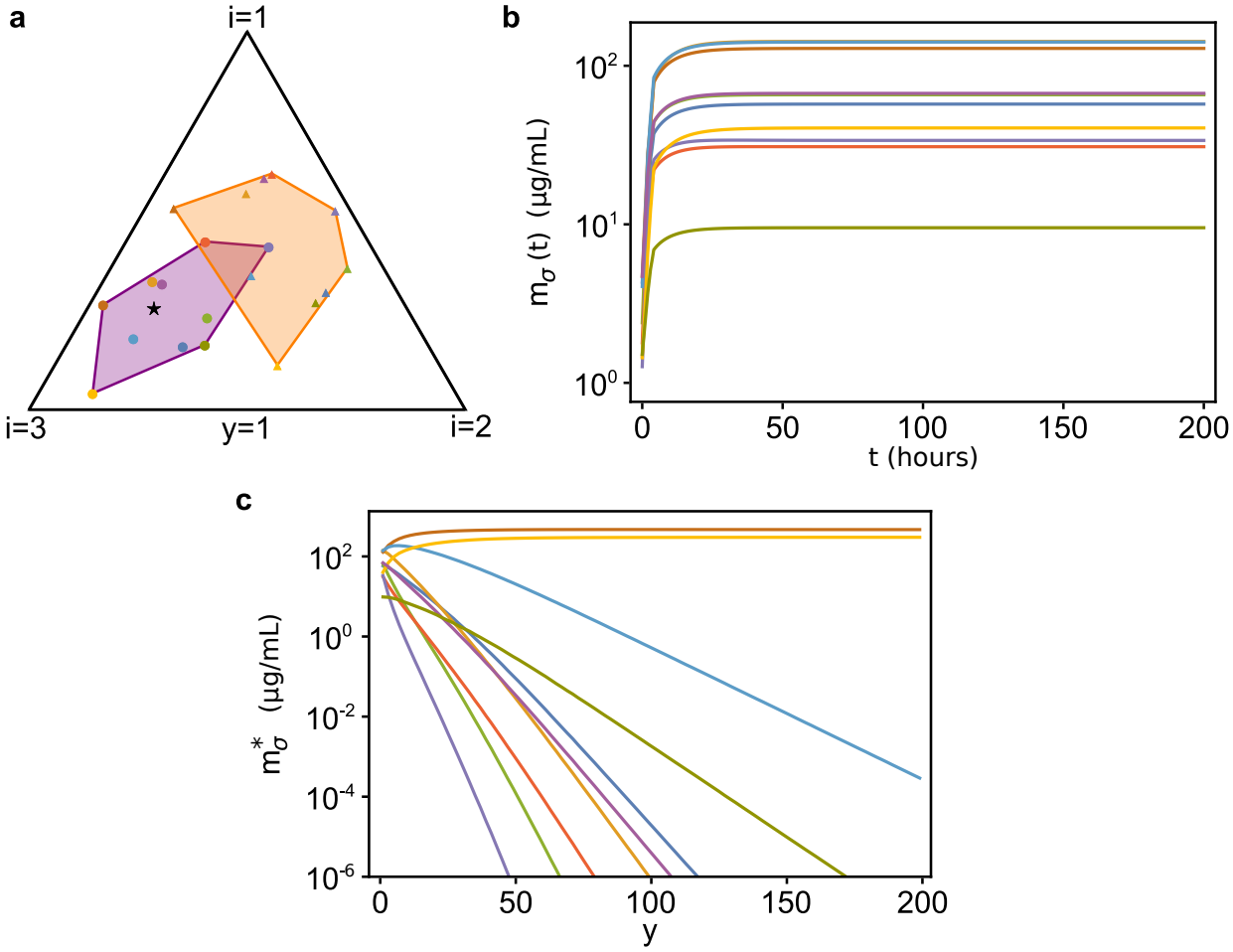

Figure 7: Temporal evolution of our consumer-proteome-resource model with $\tau_\sigma \lesssim 1$, $N_S = 10$ and $N_R = 3$. **(a):** Initial (orange) and stationary (purple) convex hull of the rescaled proteome fractions $\hat{\varphi}_{\sigma i}$. **(b):** Population dynamics of the system represented in (a). **(c):** Stationary values of the species' biomasses as function of y, where in this case we have used $y\tau_\sigma$ with $y > 1$ as the timescales over which the $\vec{\varphi}_\sigma$ evolve (and all other parameters have been kept equal to the cases shown in (a) and (b)). As we can see, as y increases the system shifts from a diverse stationary state to one where only few species survive. The parameters and the initial conditions were drawn from random distributions (see SI for more information). All parameters other than y are identical in the three panels).

cal constraints not based on direct experimental measurements. Introducing the right constraint in such models is particularly important, because the exact conditions that allow species coexistence depend on the specific form of the constraint (see Materials and Methods).

We have shown that the model can make meaningful predictions when applied to a system of two bacterial strains competing for one common resource. Furthermore, when larger communities with multiple species and resources are considered, the model predicts that high levels of biodiversity can be achieved only if certain conditions apply. In particular, we find that coexistence is possible when two conditions apply. The first one is that the maintenance cost must be proportional to the total proteome fraction allocated by the species to metabolism and growth, i.e. $q_\sigma \propto \Phi_\sigma$. The second one can be interpreted graphically as described in the Results section, and summarized as follows: *i)* if the timescales τ_σ over which the species shift between different resources are large (i.e., $\tau_\sigma \gg 1$) and if the quality of the resources is low, coexistence will be possible only if the resource supply rates have particular values (i.e., the rescaled nutrient supply rate vector $\vec{\tilde{s}}$ belongs to the convex hull of $\vec{\tilde{\varphi}}_\sigma$); *ii)* if again $\tau_\sigma \gg 1$, but the resources are of higher quality, coexistence becomes possible (and in some cases the magnitude of the resource supply rates must be large enough); *iii)* if the resources' quality is higher, coexistence is favored; and *iv)* coexistence is favored for smaller values of the timescales τ_σ.

The dynamics of microbial communities has traditionally been studied at the ecological level by using models of population dynamics describing how the population abundances of different species in the community change over time as the result of competition for resources. While this

approach is undoubtedly useful and effective, it often cannot describe the system at a level of detail necessary to make predictions from measurable quantities. In fact, it is becoming increasingly clear that the structure and dynamics of microbial communities are affected by the metabolic activity of the species that comprise them[15–18]. As shown here, mathematical models of community dynamics that take explicitly into account how different species allocate their proteome to regulate nutrient uptake can provide new insights into the link between the ecological properties of microbial communities, i.e. population dynamics and species coexistence, and their intracellular ones, i.e. metabolism and gene expression[20].

Direct competition for resources is only one of the many known interactions that can take place between microbial species: exchange of metabolic byproducts[14], production of toxins[13] and environmental conditioning[35] are only a few of the ways in which we know microbes interact within a community. Each of these processes provide both growth benefits and proteomic costs to microbial species, and can in principle be included in our framework by appropriately taking into account how they affect proteome allocation and species' fitness. With our framework it would therefore be possible to make quantitative predictions involving such phenomena, and testing them against experimental data.

Materials and Methods

The consumer-proteome-resource equations The general structure of a consumer-resource model is:

$$\dot{m}_\sigma = m_\sigma(g_\sigma - q_\sigma) \quad \sigma = 1, \dots, N_S \,, \tag{4a}$$

$$\dot{c}_i = s_i - \sum_{\sigma=1}^{N_S} J_{\sigma i} m_\sigma \quad i = 1, \dots, N_R \,, \tag{4b}$$

where m_σ is the biomass density of species σ and g_σ is its growth rate. The parameter q_σ is a maintenance cost, due to the fact that each species requires a minimum amount of energy per unit time to survive without growing. Finally, c_i is the density of resource i, s_i is the (constant) resource supply rate, and $J_{\sigma i}$ is the rate at which species σ uptakes resource i per unit biomass. To write these equations explicitly, we introduce the following assumptions: *i)* the uptake rate $J_{\sigma i}$ is proportional to the proteome fraction $\varphi_{\sigma i} = \varphi^P_{\sigma i}$ allocated by species σ for the uptake and metabolization of resource i and *ii)* each resource contributes to the growth of species σ through a term $g^{(i)}_\sigma$ proportional to the uptake rate $J_{\sigma i}$, so that the total growth rate g_σ of species σ can be written as the sum of all the terms $g^{(i)}_\sigma$. Specifically, we rewrite Eq (1a) as:

$$\varphi^P_{\sigma i} = \frac{\rho_\sigma}{\overline{\kappa}^n_i(c_i)} g^{(i)}_\sigma \,, \tag{5}$$

where ρ is considered to be species-dependent, $\overline{\kappa}_i^n(c_i) = \kappa_i^n \cdot r_i(c_i)$ (with $r_i(c_i) = c_i/(K_i + c_i)$), and $g_\sigma^{(i)}$ is the contribution to the growth rate of species σ due to the uptake of resource i, i.e.:

$$g_\sigma = \sum_{i=1}^{N_R} g_\sigma^{(i)} \ , \tag{6}$$

and we generalize Eqs (1a) and (1b) to:

$$g_\sigma = \sum_{i=1}^{N_R} \frac{\overline{\kappa}_i^n(c_i)}{\rho_\sigma} \varphi_{\sigma i}^P \ , \tag{7a}$$

$$\varphi_\sigma^R = \frac{\rho_\sigma}{\kappa_\sigma^t} g_\sigma + \varphi_\sigma^0 \ . \tag{7b}$$

Eq (6) implies that the N_R resources are substitutable (e.g., different carbon sources), otherwise their contribution to the growth rate may satisfy a different equation (e.g., their contributions may be multiplicative rather than additive). We can use Eq (7a) to write Eq (7b) in terms of the fractions $\varphi_{\sigma i}^P$. By doing so we get that the normalization condition given by Eq (2) reads:

$$\sum_{i=1}^{N_R} \varphi_{\sigma i} \left[1 + \frac{\overline{\kappa}_i^n(c_i)}{\kappa_\sigma^t} \right] = 1 - \varphi_\sigma^Q - \varphi_\sigma^0 := \Phi_\sigma \ , \tag{8}$$

where we have written $\varphi_{\sigma i}$ instead of $\varphi_{\sigma i}^P$ for simplicity and Φ_σ is the total proteome fraction that species σ allocates to metabolism and growth.

We generalize the results of Scott *et al.* to the case of multiple resources by assuming that

the uptake rate $J_{\sigma i}$ of resource i per unit biomass is proportional to $\varphi_{\sigma i}$, i.e.:

$$J_{\sigma i} = \xi_i r_i(c_i)\varphi_{\sigma i} \ , \tag{9}$$

where the proportionality constant ξ_i can be interpreted biologically as the maximum catalytic rate of the enzyme used to metabolize resource i (see SI). By comparing Eqs (9) and (5) we can see that the contribution to the growth rate of species σ due to the uptake of resource i is proportional to its uptake rate, i.e. $g_\sigma^{(i)} = \chi_{\sigma i} J_{\sigma i}$ with

$$\chi_{\sigma i}\xi_i = \frac{\kappa_i^n}{\rho_\sigma} \ . \tag{10}$$

With the considerations above, we obtain the final equations of our model:

$$\dot{m}_\sigma = m_\sigma \left[\sum_{i=1}^{N_R} \eta_{\sigma i} r_i(c_i)\varphi_{\sigma i} - q_\sigma \right] \ , \tag{11a}$$

$$\dot{c}_i = s_i - \xi_i r_i(c_i) \sum_{\sigma=1}^{N_S} m_\sigma \varphi_{\sigma i} \ , \tag{11b}$$

$$\sum_{i=1}^{N_R} \varphi_{\sigma i} \left[1 + \gamma_{\sigma i} r_i(c_i)\right] = \Phi_\sigma \ , \tag{11c}$$

where we have written explicitly $\overline{\kappa}_i^n(c_i) = \kappa_i^n r_i(c_i)$ with $r_i(c_i) = c_i/(K_i + c_i)$, and we have defined $\eta_{\sigma i} := \kappa_i^n/\rho_\sigma$ and $\gamma_{\sigma i} := \kappa_i^n/\kappa_\sigma^t$ to simplify the notation. Regardless of the particular form of $\overline{\kappa}^n(c)$ chosen, for our purposes we only need to assume that $\overline{\kappa}^n(c)$ is a monotonically increasing function of c, and that $\lim_{c\to 0} \overline{\kappa}^n(c)/c = \kappa^n/K$ and $\lim_{c\to\infty} \overline{\kappa}^n(c) = \kappa^n$. The parameter ξ_i can

be interpreted as the maximum catalytic rate of the enzyme used to metabolize resource i, and Φ_σ is the total proteome fraction allocated by species σ for nutrient uptake and growth (see Materials and Methods and SI).

The constraint in Eq (11c) is the explicit expression of Eq (2) in our framework, and can be interpreted geometrically: considering species σ, the N_R-dimensional vector $\vec{\varphi}_\sigma = (\varphi_{\sigma 1}, \ldots, \varphi_{\sigma N_R})$ belongs to a hyperplane whose normal vector $\hat{n}_\sigma$ has components $1 + \gamma_{\sigma i} r_i(c_i)$. This means that as the system evolves, the components of $\hat{n}_\sigma$ vary with time and therefore the hyperplane to which $\vec{\varphi}_\sigma$ belongs moves in the N_R-dimensional space. This is also the reason why the proteome fractions $\varphi_{\sigma i}$ must be dynamical variables: the coefficients $1+\gamma_{\sigma i} r_i(c_i)$ in Eq (11c) are not fixed, but change with time depending on the system's dynamics through $r_i(c_i)$. This implies that for the constraint to be satisfied at all times, the proteome fractions $\varphi_{\sigma i}$ *cannot* be fixed but must be, in turn, dynamical variables: an increase (decrease) of $1 + \gamma_{\sigma i} r_i(c_i)$ must be balanced by a decrease (increase) of some of the $\varphi_{\sigma i}$s. This constraint reflects the well known fact that microbes can vary their enzyme synthesis with time and switch between nutrients according to environmental conditions[33,36–38].

Dynamics of the proteome fractions $\varphi_{\sigma i}$ We call $\vec{c} = (c_1, \ldots, c_{N_R})$ the vector of resource concentrations and define

$$F_\sigma(\vec{\varphi}_\sigma, \vec{c}) := \sum_{i=1}^{N_R} \varphi_{\sigma i} \left[1 + \gamma_{\sigma i} r_i(c_i)\right] - \Phi_\sigma \tag{12}$$

so that the constraint given by Eq (11c) can be written more simply as $F_\sigma(\vec{\varphi}_\sigma, \vec{c}) = 0$. Since this constraint must hold at every instant, any equation for $\dot{\vec{\varphi}}_\sigma$ must satisfy

$$\dot{F}_\sigma(\vec{\varphi}_\sigma, \vec{c}) \equiv \dot{\vec{\varphi}}_\sigma \cdot \vec{\nabla}_\varphi F_\sigma + \dot{\vec{c}} \cdot \vec{\nabla}_c F_\sigma = 0 \ , \tag{13}$$

where $\vec{\nabla}_\varphi$ and $\vec{\nabla}_c$ are, respectively, the gradients taken with respect to the components of $\vec{\varphi}_\sigma$ and $\vec{c}$. The "minimal" equation for $\varphi_{\sigma i}$, i.e. the simplest one (in the sense that it does not introduce extra terms orthogonal to $\vec{\nabla}_\varphi F_\sigma$, which would lead to a proliferation of new parameters) that satisfies Eq (13), is therefore:

$$\dot{\vec{\varphi}}_\sigma = -\frac{\vec{\nabla}_\varphi F_\sigma}{(\vec{\nabla}_\varphi F_\sigma)^2} \dot{\vec{c}} \cdot \vec{\nabla}_c F_\sigma \ , \tag{14}$$

where, however, we are not taking into account the fact that with such an equation some of the $\varphi_{\sigma i}$ might become negative with time (see SI for detailed computations on how this can be taken into account).

Microbes are able to switch between nutrients when cultured in mediums containing more than one resource[36]. For this reason, we can implement an adaptive approach[33] and ask that $\vec{\varphi}_\sigma$ evolves in time so that the growth rate g_σ of species σ is maximized respecting the constraint $F_\sigma(\vec{\varphi}_\sigma, \vec{c}) = 0$, i.e. Eq (11c) is satisfied. In this case the evolution equation for $\vec{\varphi}_\sigma$ becomes:

$$\dot{\vec{\varphi}}_\sigma = \frac{1}{\tau_\sigma} \vec{\nabla}_\varphi g_\sigma - \frac{\vec{\nabla}_\varphi F_\sigma}{(\vec{\nabla}_\varphi F_\sigma)^2} \left(\frac{1}{\tau_\sigma} \vec{\nabla}_\varphi g_\sigma \cdot \vec{\nabla}_\varphi F_\sigma + \dot{\vec{c}} \cdot \vec{\nabla}_c F_\sigma \right) \ , \tag{15}$$

where we have introduced τ_σ, the characteristic timescale over which $\vec{\varphi}_\sigma$ changes[33] (detailed com-

putations are shown in the SI). We can recover Eq (14) from Eq (15) by sending τ_σ to infinity. Geometrically, Eq (14) represents the case in which $\vec{\varphi}_\sigma$ is dragged along by the hyperplane to which it belongs, as the hyperplane moves because of Eq (11c). On the other hand, according to Eq (15) (with small enough values of τ_σ) the $\vec{\varphi}_\sigma$ are free to move on the hyperplane to find the maximum instantaneous growth rate compatible with the constraint given by Eq (11c).

In this work we have used a generalization of Eq (15) that ensures $\varphi_{\sigma i}(t) \geq 0 \ \forall t$, and varied the values of τ_σ when needed (see SI for details).

Conditions for coexistence Evaluating Eqs (11a)–(11c) at stationarity we obtain:

$$\sum_{i=1}^{N_R} \eta_{\sigma i} r_i^* \varphi_{\sigma i}^* = q_\sigma \ , \tag{16a}$$

$$s_i = \xi_i r_i^* \sum_{\sigma=1}^{N_S} m_\sigma^* \varphi_{\sigma i}^* \ , \tag{16b}$$

$$\sum_{i=1}^{N_R} \varphi_{\sigma i}^* (1 + \gamma_{\sigma i} r_i^*) = \Phi_\sigma \ , \tag{16c}$$

where we are denoting with the symbol "*" the quantities computed at stationarity, and we have assumed $m_\sigma^* \neq 0$. It is easily seen by substitution that a possible solution for r_i^* in Eqs (16a) and (16c) is

$$r_i^* = \left[\kappa_i^n \left(\frac{\Phi_\sigma}{\rho_\sigma q_\sigma} - \frac{1}{\kappa_\sigma^t} \right) \right]^{-1} \ . \tag{17}$$

This solution is acceptable only if its right hand side is independent of σ, i.e. if

$$\frac{\Phi_\sigma}{\rho_\sigma q_\sigma} - \frac{1}{\kappa_\sigma^t} = \Theta \ , \tag{18}$$

with Θ some given constant independent of σ. Using Eqs (17) and (18) in Eq (16c) or (16a) we get

$$\sum_{i=1}^{N_R} \varphi_{\sigma i}^* = \frac{\Phi_\sigma}{1 + \frac{1}{\Theta \kappa_\sigma^t}} \ . \tag{19}$$

From Eq (17) we have:

$$r_i^* = \frac{1}{\kappa_i^n \Theta} \qquad \Rightarrow \qquad c_i^* = \frac{K_i}{\kappa_i^n \Theta - 1} \ , \tag{20}$$

and since we need $r_i^* < 1$ (or equivalently $c_i^* > 0$), we need $\Theta > \max_i 1/\kappa_i^n$. Notice that Eq (18) can be rewritten as

$$q_\sigma = \frac{\Phi_\sigma}{\rho_\sigma (\Theta + 1/\kappa_\sigma^t)} \ , \tag{21}$$

which is the explicit expression of the relationship between q_σ and Φ_σ. Notice that Eq (19) is a consequence of the system's constraint in Eq (16c), which is Eq (11c) computed at stationarity. Therefore, the expression of the maintenance cost given in Eq (21) is a consequence of the constraint introduced in our consumer-proteome-resource model.

If we now define:

$$\hat{s}_i = \frac{s_i \kappa_i^n / \xi_i}{\sum_{j=1}^{N_R} s_j \kappa_j^n / \xi_j} \tag{22a}$$

$$\hat{\varphi}_{\sigma i}^* = \frac{\varphi_{\sigma i}^*}{\sum_{j=1}^{N_R} \varphi_{\sigma j}^*} \ , \tag{22b}$$

$$z_\sigma = \frac{m^*_\sigma \rho_\sigma q_\sigma}{\sum_\lambda m^*_\lambda \rho_\lambda q_\lambda} \ , \tag{22c}$$

(so that z_σ are positive coefficients that sum to one) Eq (16b) can be indeed rewritten as

$$\hat{s}_i = \sum_{\sigma=1}^{N_S} z_\sigma \hat{\varphi}^*_{\sigma i} \tag{23}$$

(see SI for the detailed computations).

Since $\sum_i \hat{s}_i = \sum_i \hat{\varphi}^*_{\sigma i} = 1$, the vectors $\vec{\hat{s}}$ and $\vec{\hat{\varphi}}^*_\sigma$ belong to an $(N_R - 1)$–dimensional simplex. Furthermore, since z_σ are positive coefficients that sum to one, Eq (23) means that $\vec{\hat{s}}$ belongs to the convex hull of the vectors $\vec{\hat{\varphi}}^*_\sigma$. Since Eq (23) derives from requiring that all species have non-null stationary biomasses, we can see how this is the second condition necessary for coexistence. Notice also that Eq (23) depends on the (rescaled) value of $\varphi_{\sigma i}$ *at stationarity*, meaning that the proteome fractions $\varphi_{\sigma i}$ vary over time as the system evolves to satisfy Eq (23), i.e. to include $\vec{\hat{s}}$ in the convex hull of the vectors $\vec{\hat{\varphi}}^*_\sigma$.

If we now suppose that $\tau_\sigma \gg 1$, so that we can use Eq (14) for the dynamics of $\varphi_{\sigma i}$, observing that the i-th component of the gradients $\vec{\nabla}_\varphi F_\sigma$ and $\vec{\nabla}_c F_\sigma$ are

$$(\vec{\nabla}_\varphi F_\sigma)_i \equiv \frac{\partial F_\sigma}{\partial \varphi_{\sigma i}} = 1 + \gamma_{\sigma i} r_i(c_i) \tag{24a}$$

and

$$(\vec{\nabla}_c F_\sigma)_i \equiv \frac{\partial F_\sigma}{\partial c_i} \propto \gamma_{\sigma i} \ , \tag{24b}$$

we find that if $\gamma_{\sigma i} \sim 0$ then $\dot{\vec{\varphi}}_\sigma \sim 0$ and therefore, if stationarity is reached sufficiently fast, $\varphi^*_{\sigma i} \sim \varphi_{\sigma i}(t=0)$. In other words, if the $\gamma_{\sigma i}$ are small, the proteome fractions $\varphi_{\sigma i}$ at stationarity will be close to their initial values. Therefore in this case, with good approximation, Eq (23) gives the condition for all species to coexist, i.e. $\vec{\hat{s}}$ must be inside the convex hull of $\hat{\varphi}_{\sigma i} = \varphi_{\sigma i}(0)/\sum_j \varphi_{\sigma j}(0)$. If $\gamma_{\sigma i} \sim 0$ as discussed in the Results section, coexistence will be possible if the components of $\dot{\vec{\varphi}}_\sigma$ are not too small for a sufficiently long period of time so as to allow them to reach values satisfying Eq (23) and thus for the species to coexist. This can be obtained by using large supply rates s_i so that $r_i(c_i) \sim 1$ for a sufficiently long time, as discussed in the Results. Finally, if the ratios $\gamma_{\sigma i}$ have larger values the proteome fractions $\varphi_{\sigma i}$ will be able to move more quickly.

Strains used in the experiment The *Escherichia coli* strains used in our experiment have the same genetic background MG1655. The strains used in the experiments were constructed starting from the ancestor strain 0Y (expressing constitutively the yellow fluorescent protein mVenus from the genome, with genotype attTN7::pRNA1_mVenus) or the ancestor strain 0R (expressing constitutively the red fluorescent protein mKate2Hyb from the genome, with genotype attTN7::pRpsL_mKate2Hyb).

Strain 1 was obtained by transforming strain 0Y with the plasmid pR (see Table S.1), which contains the ampicillin resistance cassette, the red fluorescent protein mCherry under the control of the *trc* promoter, a hybrid of the *trp* and *lac* promoters, and the *lac* repressor, *lacI*. The expression of mCherry could thus be induced by adding IPTG, which binds to the repressor encoded by *lacI* allowing the expression of genes promoted by the *trc* promoter (here, mCherry). Because IPTG

cannot be metabolized by *E. coli*, its concentration remains constant during our experiment and is unaltered by bacterial growth.

Strain 2 was obtained by transforming strain 0R with the plasmid pAMP (see Table S.1), which was obtained by removing the inducible red fluorescent protein mCherry from plasmid pR using traditional cloning.

Strain 3 was obtained by transforming strain 0R with plasmid pY (see Table S.1), which is identical to plasmid pR, except for the fluorescent protein induced by the *trc* promoter, which is Venus YFP instead of mCherry.

Strain 4 was obtained transforming strain 0Y with plasmid pAMP.

Because all strains had the ampicillin resistance cassette in the plasmids used to transform them, we performed the experiments by adding ampicillin to the medium to prevent contamination and plasmid loss.

Experimental protocol The competition assays were performed as follows:

1. The strains were cultured overnight from a stock culture in M63 medium with 1% w/v glucose, and ampicillin. Then, the strains were mixed to perform competition assays aiming for 50:50 relative frequencies.

2. The mixtures were inoculated in a 96-well plate containing M63 medium with 1% w/v glu-

cose and ampicillin at eight different IPTG concentrations: 0μM, 15μM, 30μM, 45μM, 60μM, 75μM, 90μM, 105μM (six technical replicates per concentration).

3. The well plate was covered with a porous rayon film that allowed gas exchange and was cultured for 24 hours at 30°C on a microplate shaker set at 1050rpm.

4. After 24 hours, the plate was reinoculated in a new 96-well plate with fresh medium (with the appropriate concentrations of IPTG in each well) with a dilution factor of 100. The new plate was cultured for another cycle at 30°C for 24 hours with constant shaking at 1050rpm, while the old one was diluted with a dilution factor of 2000 to be analyzed at the flow cytometer.

IPTG calibration and computation of the normalized protein production rate We measured how the fluorescence intensity of individual cells, a proxy for the total amount of fluorescent protein produced, varied as a function of the IPTG concentration. To do so, we inoculated strains 1 and 3 in a 96-well plate containing M63 minimal medium with ampicillin, 1% w/v glucose and the same IPTG concentrations used in our experimental protocol (six technical replicates per concentration, per strain). The plate was incubated at 30°C for 8 hours with constant shaking at 1050rpm. At times $t = 4$h and $t = 8$h after inoculation we measured at the flow-cytometer the mean fluorescence intensity of cells due to the induced fluorescent proteins at the various concentrations of IPTG (Figures 3d and 3e). From these data, we estimated the normalized fluorescent protein production rate as follows.

We call $k(C_I)$ the rate at which the fluorescence of the inducible protein increases when cells are exposed to a concentration C_I of IPTG, and we call d_{FP} the fluorescent protein degradation

rate. The fluorescent intensity $\mathcal{I}$ of a cell (due to the production of the IPTG-inducible fluorescent protein) in between two successive cell divisions thus satisfies $d\mathcal{I}/dt = k(C_I) - d_{FP}\mathcal{I}$. At a cell division event, the fluorescent intensity of a cell is reduced by a factor 2. Indicating with $\mathcal{I}_0$ the cell's fluorescent intensity at the first measurement time ($t = 4$h), it can be shown (see SI) that according to this model the cell's fluorescent intensity changes with time as:

$$\mathcal{I}(t) = 2\mathcal{I}_0 e^{-gt\left(1+\frac{d_{FP}}{g}\right)} + \frac{k(C_I)}{d_{FP}}\left[1 - e^{-gt\left(1+\frac{d_{FP}}{g}\right)}\right]\left(1 - \frac{1}{2^{1+\frac{d_{FP}}{g}} - 1}\right) , \tag{25}$$

where g is the cell's growth rate. Fluorescent proteins have small degradation rates compared to the cellular growth rate, so assuming $d_{FP} \ll g$ we can approximate Eq 25 as:

$$\mathcal{I}(t) = 2\mathcal{I}_0 e^{-gt} + 2\ln(2)\frac{k(C_I)}{g}\left(1 - e^{-gt}\right) . \tag{26}$$

We used Eq 26 and the data Figures 3d and 3e to compute the quantity $k(C_I)$. Because the absolute value of $k(C_I)$ depends on the arbitrary units returned by the flow cytometer (the intensity $\mathcal{I}$ is measured as a cell's pulse area at the flow cytometer), we normalized the values of $k(C_I)$ dividing them by the mean fluorescent intensity $\langle\mathcal{I}\rangle$ of cells measured in the absence of IPTG at the first measurement in the calibration experiment (see Figures 3d and 3e). Such a normalization affects only the absolute value of such rates, and not their relative magnitude. This also means that the normalized production rates shown for the two experiments in Figure 3 cannot be compared directly. The normalized $k(C_I)/\langle\mathcal{I}\rangle$ are the protein production rates (with dimensions 1/time) reported in Figure 3.

Estimation of the selection coefficient $\mathcal{S}$ Applying Eqs (11a)–(11c) to the case of two populations and one resource, we obtain:

$$\dot{m}_\sigma = m_\sigma \left(\eta_\sigma r(c)\varphi_\sigma - q_\sigma\right) \qquad \sigma = 1,\ 2\ , \tag{27a}$$

$$\dot{c} = s - \xi r(c)(m_1\varphi_1 + m_2\varphi_2)\ , \tag{27b}$$

$$\varphi_\sigma(1 + \gamma_\sigma r(c)) = \Phi_\sigma \qquad \sigma = 1,\ 2\ , \tag{27c}$$

where now Eq (27c) gives the explicit expression of the (only) proteome fraction φ_σ as a function of the resource concentration. Because the ancestors of our two strains (i.e., strains 1 and 2 without the plasmids conferring ampicillin resistant and the inducible RFP, see Materials and Methods) have equal growth rates (see Figures S.8 and S.9), we set $\eta_1 = \eta_2 = \eta$, $q_1 = q_2 = q$ and $\gamma_1 = \gamma_2 = \gamma$ in Eqs (27a)–(27c). Note that, instead, $\Phi_1 \neq \Phi_2$ because the proteome allocation of strain 1 could be varied experimentally and because the plasmids introduced in the ancestor strains have different maintenance costs. Given that cells in the experiment are grown in nutrient-rich conditions, we assume that the maintenance cost is negligible, i.e. $q \simeq 0$. Furthermore, because most of the dynamics (i.e., the relative change in abundance of the two strains) occurs in the early phases of growth when glucose is abundant, we assume that $r(c) \approx 1$ at all times so that we can neglect Eq (27b) and we are left with:

$$\dot{m}_\sigma = m_\sigma \frac{\eta}{1+\gamma}\Phi_\sigma \qquad \sigma = 1,\ 2\ . \tag{28}$$

Notice again that this expression, and in particular the fact that the growth rate of species σ is proportional to Φ_σ, is a consequence of the constraint in Eq (27c).

We therefore have that the expression of the selective advantage $\mathcal{S}$ is:

$$\mathcal{S} = \frac{d}{dt} \ln \frac{f}{1-f} = \frac{\dot{m}_1}{m_1} - \frac{\dot{m}_2}{m_2} = \frac{\eta}{1+\gamma} \left(\Phi_1 - \Phi_2\right), \tag{29}$$

where $f = m_1/(m_1+m_2)$ and $1-f = m_2/(m_1+m_2)$ are the relative abundances (or "frequencies") of strain 1 and strain 2, respectively. In the SI we show that if we lift the assumption that $r(c) = 1$ at all times, we obtain:

$$\mathcal{S} = \frac{\eta \cdot r(c)}{1+\gamma \cdot r(c)} \left(\Phi_1 - \Phi_2\right) \ . \tag{30}$$

Evaluation of the ratios Φ_1/Φ_2 and Φ_3/Φ_4 Consider the competition assay with strains 1 and 2 (the results are the same also for the competition assay between strains 3 and 4, after all subscripts are appropriately changed). For a given IPTG concentration C_I, from Eq (28) the growth rate of strain 1 is $g_1[k(C_I)] = \Phi_1[k(C_I)] \cdot r(c)\eta/(1+\gamma r(c))$ (where we have inserted explicitly the dependence on $r(c)$, and $k(C_I)$ is the protein production rate induced by C_I). Dividing $\mathcal{S}$ in Eq (30) by g_1, for any value of $r(c)$ we obtain:

$$\frac{\mathcal{S}[k(C_I)]}{g_1[k(C_I)]} = \frac{\Phi_1[k(C_I)] - \Phi_2[k(C_I)]}{\Phi_1[k(C_I)]} \ , \tag{31}$$

which is easily rearranged into:

$$\frac{\Phi_1}{\Phi_2}[k(C_I)] = \frac{1}{1 - \mathcal{S}[k(C_I)]/g_1[k(C_I)]} \tag{32}$$

(which are the values plotted in Figure 3c). Notice that, this result does *not* depend on the assumption that $r(c) = 1$ at all times, i.e. Eq (32) is valid for any value of $r(c)$.

Selective advantage and proteome fraction allocated to the useless RFP Because we did not measure RNA/protein ratios in our experiments, we can only estimate the values of κ^t and κ^n by taking them from the literature for *E. coli* strains grown at 30°C in conditions similar to our experiments. Rosset *et al.*[39,40] measured the RNA/protein ratio of several *E. coli* strains grown at 30°C in M63 medium. Using their data and the relationship[22] $r = r_0 + g/\kappa^t$, where r is the RNA/protein ratio and r_0 a constant, we can estimate the translational capacity as $\kappa^t = 3.0 \pm 0.5$ μg protein/μg RNA $\cdot$ 1/h (mean$\pm$SD). An estimate for the nutritional capacity κ^n, instead, can be obtained via the equation[22] $g = g_{\max}\kappa^n/(\kappa^n + \kappa^t)$, where $g_{\max}$ is the maximum growth rate obtainable by our strain at a given temperature (for us, 30°C), when nutrients are abundant. Van Derlinden and Van Impe[41] report a maximum growth rate $g_{\max} \approx 1.2$ 1/h for *E. coli* MG1655 grown at 30°C in rich medium with glucose (no error estimate was reported). Solving for κ^n and using the growth rate value g measured for strain 1 in the absence of IPTG, we find $\kappa^n = 1.2 \pm 0.2$ μg protein/μg RNA $\cdot$ 1/h. These values allow us to estimate $\gamma = \kappa^n/\kappa^t = 0.4 \pm 0.1$ and $\eta = \kappa^n/\rho = 1.57 \pm 0.07$ 1/h using the value for $\rho = 0.76$ μg protein/μg RNA $\cdot$ 1/h reported in Scott *et al.*[22]. With these estimations, from the expression of the selective advantage in Eq (30) we have

that a 1% difference in proteome allocation for metabolism and growth between the two strains (i.e., $\Phi_1 - \Phi_2 = 1\%$) leads to $\mathcal{S} \approx 1.1 \cdot 10^{-2}$. Finally, with these calculations we can estimate the maximum percentage of proteome $\max \varphi_{iRFP}$ and $\max \varphi_{iYFP}$ allocated to the production of, respectively, the inducible red and yellow proteins in our two experiments. In particular, for the first experiment we have $\max \varphi_{iRFP} = \Phi_1^{(0)} - \Phi_2 - (1+\gamma)/\eta \cdot \mathcal{S}_{105} = (1+\gamma)/\eta \cdot (\mathcal{S}_0 - \mathcal{S}_{105}) \approx 1.1\%$ (where $\mathcal{S}_0$ and $\mathcal{S}_{105}$ are, respectively, the mean selection coefficients in the 0 μM and 105 μM IPTG treatments). For the experiment involving strains 3 and 4, using the same procedure we find $\max \varphi_{iYFP} \approx 0.4\%$. Of course, given that we had to rely on measurements taken from the literature, these should be regarded as only rough estimates.

Acknowledgements We thank David R Nelson and Andrew W Murray for hosting L. P.-M. during the experiment and the initial development of the model and for insightful comments and suggestions. We thank Daniel Eaton for providing the ancestor bacterial strains used in the experiment. A. M. and L. P.-M. acknowledge the Cariparo Foundation for funding. S. S. acknowledges the University of Padua for STARS ReACT grant. A. G. was supported by research fellowships from the Swiss National Science Foundation, Projects P2ELP2_168498, P400PB_180823 and P400PB_180823 / 2.

Competing Interests The authors declare that they have no competing financial interests.

Correspondence Correspondence and requests for materials should be addressed to L. P.-M. (email: leonardopaccianimori@gmail.com).

1. Bar-On, Y. M., Phillips, R. & Milo, R. The biomass distribution on earth. *Proceedings of the National Academy of Sciences* **115**, 6506–6511 (2018).

2. Colman, D. R., Poudel, S., Stamps, B. W., Boyd, E. S. & Spear, J. R. The deep, hot biosphere: Twenty-five years of retrospection. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 6895–6903 (2017).

3. Puente-Sánchez, F. *et al.* Viable cyanobacteria in the deep continental subsurface. *Proceedings of the National Academy of Sciences* **115**, 10702–10707 (2018).

4. Gold, T. The deep, hot biosphere. *Proceedings of the National Academy of Sciences* **89**, 6045–6049 (1992).

5. Sekirov, I., Russell, S. L., Antunes, L. C. M. & Finlay, B. B. Gut microbiota in health and disease. *Physiological Reviews* **90**, 859–904 (2010).

6. Singh, B. K., Bardgett, R. D., Smith, P. & Reay, D. S. Microorganisms and climate change: terrestrial feedbacks and mitigation options. *Nature Reviews Microbiology* **8**, 779–790 (2010).

7. Cavicchioli, R. *et al.* Scientists' warning to humanity: microorganisms and climate change. *Nature Reviews Microbiology* **17**, 569–586 (2019).

8. Stewart, E. J. Growing unculturable bacteria. *Journal of Bacteriology* **194**, 4151–4160 (2012).

9. Gonze, D., Coyte, K. Z., Lahti, L. & Faust, K. Microbial communities as dynamical systems. *Current Opinion in Microbiology* **44**, 41 – 49 (2018).

612 10. Tikhonov, M. & Monasson, R. Collective Phase in Resource Competition in a Highly Diverse
613 Ecosystem. *Physical Review Letters* **118**, 1–5 (2017).

614 11. Butler, S. & O'Dwyer, J. P. Stability criteria for complex microbial communities. *Nature*
615 *Communications* **9**, 2970 (2018).

616 12. Landmann, S. & Engel, A. Systems of random linear equations and the phase transition in
617 MacArthur's resource-competition model. *Epl* **124** (2018).

618 13. Niehaus, L. *et al.* Microbial coexistence through chemical-mediated interactions. *Nature*
619 *Communications* **10** (2019).

620 14. Marsland, R., III *et al.* Available energy fluxes drive a transition in the diversity, stability,
621 and functional structure of microbial communities. *PLOS Computational Biology* **15**, 1–18
622 (2019).

623 15. Rivett, D. W. & Bell, T. Abundance determines the functional role of bacterial phylotypes in
624 complex communities. *Nature Microbiology* **3**, 767–772 (2018).

625 16. Enke, T. N. *et al.* Modular Assembly of Polysaccharide-Degrading Marine Microbial Com-
626 munities. *Current Biology* 1–8 (2019).

627 17. Zelezniak, A. *et al.* Metabolic dependencies drive species co-occurrence in diverse microbial
628 communities. *Proceedings of the National Academy of Sciences* **112**, 201522642 (2015).

629 18. Louca, S. *et al.* High taxonomic variability despite stable functional structure across microbial
630 communities. *Nature Ecology & Evolution* **1**, 0015 (2016).

631 19. Basan, M. Resource allocation and metabolism: the search for governing principles. *Current*
632 *Opinion in Microbiology* **45**, 77–83 (2018).

633 20. Bajic, D. & Sanchez, A. The ecology and evolution of microbial metabolic strategies. *Current*
634 *Opinion in Biotechnology* **62**, 123–128 (2020).

635 21. Budinich, M., Bourdon, J., Larhlimi, A. & Eveillard, D. A multi-objective constraint-based
636 approach for modeling genome-scale microbial ecosystems. *PLOS ONE* **12**, 1–22 (2017).

637 22. Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z. & Hwa, T. Interdependence of cell
638 growth and gene expression. *Science* **330** (2010).

639 23. Schaechter, M., Maaløe, O. & Kjeldgaard, N. O. Dependency on medium and temperature
640 of cell size and chemical composition during balanced growth of salmonella typhimurium.
641 *Microbiology* **19**, 592–606 (1958).

642 24. Scott, M. & Hwa, T. Bacterial growth laws and their applications. *Current Opinion in Biotech-*
643 *nology* **22**, 559–565 (2011).

644 25. Basan, M. *et al.* Overflow metabolism in Escherichia coli results from efficient proteome
645 allocation. *Nature* **528**, 99–104 (2015).

646 26. Mori, M., Hwa, T., Martin, O. C., De Martino, A. & Marinari, E. Constrained allocation flux
647 balance analysis. *PLOS Computational Biology* **12**, 1–24 (2016).

648 27. MacArthur, R. Species Packing, and What Competition Minimizes. *Proceedings of the Na-*
649 *tional Academy of Sciences* **64**, 1369–1371 (1969).

650 28. MacArthur, R. Species packing and competitive equilibrium for many species. *Theoretical*
651 *Population Biology* **1**, 1–11 (1970).

652 29. Chesson, P. MacArthur's consumer-resource model. *Theoretical Population Biology* **37**, 26–
653 38 (1990).

654 30. Tikhonov, M. Community-level cohesion without cooperation. *eLife* **5** (2016).

655 31. Posfai, A., Taillefumier, T. & Wingreen, N. S. Metabolic Trade-Offs Promote Diversity in a
656 Model Ecosystem. *Physical Review Letters* **118**, 28103 (2017).

657 32. Advani, M., Bunin, G. & Mehta, P. Statistical physics of community ecology: a cavity solution
658 to MacArthur's consumer resource model. *Journal of Statistical Mechanics: Theory and*
659 *Experiment* **2018**, 033406 (2018).

660 33. Pacciani-Mori, L., Giometto, A., Suweis, S. & Maritan, A. Dynamic metabolic adaptation can
661 promote species coexistence in competitive communities. *PLOS Computational Biology* **16**,
662 1–18 (2020).

663 34. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a
664 single cell. *Science* **297**, 1183–1186 (2002).

665 35. Ratzke, C. & Gore, J. Modifying and reacting to the environmental pH can drive bacterial
666 interactions. *PLoS Biology* **16**, e2004248 (2018).

667 36. Monod, J. The growth of bacterial cultures. *Annual Review of Microbiology* **3**, 371–394
668 (1949).

669 37. Stülke, J. & Hillen, W. Carbon catabolite repression in bacteria. *Current opinion in microbi-*
670 *ology* **2**, 195–201 (1999).

671 38. Görke, B. & Stülke, J. Carbon catabolite repression in bacteria: Many ways to make the most
672 out of nutrients. *Nature Reviews Microbiology* **6**, 613–624 (2008).

673 39. Rosset, R., Monier, R. & Julien, J. Rna composition of escherichia coli as a function of growth
674 rate. *Biochemical and Biophysical Research Communications* **15**, 329 – 333 (1964).

675 40. Rosset, R., Julien, J. & Monier, R. Ribonucleic acid composition of bacteria as a function of
676 growth rate. *Journal of Molecular Biology* **18**, 308 – 320 (1966).

677 41. Van Derlinden, E. & Van Impe, J. F. Modeling growth rates as a function of temperature:
678 Model performance evaluation with focus on the suboptimal temperature range. *International*
679 *Journal of Food Microbiology* **158**, 73 – 78 (2012).