ISME

ARTICLE

Constrained proteome allocation affects coexistence in models of competitive microbial communities

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

Abstract

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 structure and composition of microbial communities are intertwined with the metabolism of the species that inhabit them, 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 proteome 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 with a simple experiment 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. Then, we study our consumer-proteome-resource 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 biogeochemical 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 authors contributed equally: Amos Maritan, Andrea Giometto

Supplementary information The online version of this article (https://doi.org/10.1038/s41396-020-00863-0) contains supplementary material, which is available to authorized users.

✉ Leonardo Pacciani-Mori
leonardopaccianimori@gmail.com

✉ Andrea Giometto
giometto@cornell.edu

1 Dipartimento di Fisica e Astronomia "Galileo Galilei", Università degli Studi di Padova, Via Francesco Marzolo 8, 35131 Padova, Italy

2 Department of Physics, Harvard University, 17 Oxford St, Cambridge, MA 02138, USA

3 School of Civil and Environmental Engineering, Cornell University, 220 Hollister Dr, Ithaca, NY 14853, USA

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 inhabit 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 use it to describe the dynamics of microbial species 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. These studies, however, did not account for the fact that proteome allocation constraints limit the rates at which microbes can uptake different resources, which, as shown here, affects the conditions that lead to the coexistence of multiple microbial species in competitive communities. 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" (CPR) 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.

There have been attempts in the past at deriving models to describe the dynamics and/or structure of microbial communities by incorporating some insight into the metabolism of their species and the molecular aspects of their growth. One of the earliest and most notable efforts in this direction was performed by Droop [33], who developed a model that describes microalgal growth by taking into account intracellular quotas of the (single) supplied resource. In more recent times, the problem has been addressed by applying Flux Balance Analysis to genome-scale models in order to reveal how metabolic fluxes can influence community dynamics [34, 35]. This approach, however, leads to models that are extremely complicated and strongly dependent on the identity of the species in the community, since they require detailed knowledge of metabolic networks with hundreds of different reactions for every species, as well as the metabolic interactions among the members of the community. More recently, it has been shown that introducing some information on the metabolism of microbial species in models of community dynamics (without all the details that a Flux Balance Analysis model requires) can provide us with useful insights on the

properties of the community [36, 37]. Our work sits conceptually in this latter context, but unlike what has already been done in this direction does not make assumptions on the metabolism of the species and relies on quantities (like the proteome fractions) that can be measured directly.

In the next section, we describe the CPR model for a general number of species/strains and resources. First, we review the proteome allocation framework of Scott et al. [22] and discuss how we generalize it to multiple resources. Second, we review the fundamental structure of consumer-resource models. Third, we construct our consumer-resource model which incorporates proteome allocation. Then, we consider the simplest implementation of an experimental microbial community, i.e. two *Escherichia coli* strains competing for glucose as the only carbon source, to illustrate that it is necessary to account for proteome allocation in consumer-resource models to describe the dynamics (and the conditions for coexistence) of even the simplest microbial community. The experiment described here constitutes a proof of the concept that one needs to account for proteome allocation dynamics when adopting consumer-resource theory to describe competitive microbial communities. Finally, we study (both analytically and numerically) the CPR 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

Microbial proteome allocation

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 the Monod function which encapsulates the dependence on the resource concentration c. Most of our results do not actually depend on the exact functional form of $r(c)$, as long as $r(c)$ is a monotonically increasing function that saturates for large values of c (see Materials and Methods). K is the half-saturation constant of the resource and κ^n is the "nutritional capacity" [22] 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 Supplementary Information and [22] for a molecular interpretation of κ^n). The parameter κ^t is the "translational capacity" [22] 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*.

Scott et al.'s results apply to microbes growing on a single resource. We generalize their framework to a system with multiple species and resources as shown in Fig. 1a: indicating with $\varphi^P_{\sigma i}$ the proteome fraction allocated by species σ to the uptake and metabolization of resource i, the total proteome fraction allocated by species σ to nutrient uptake and metabolism is given by $\varphi^P_\sigma = \sum_{i=1}^{N_R} \varphi^P_{\sigma i}$. To ensure that the sum of all the proteome fractions is equal to one we must have:

$$\varphi^Q_\sigma + \varphi^R_\sigma + \sum_{i=1}^{N_R} \varphi^P_{\sigma i} = 1. \tag{2}$$

This constraint represents the finiteness of a 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^P_σ increases), then a smaller part of the proteome will be available for biomass production (i.e., the ribosomal fraction φ^R_σ decreases). In order to achieve optimal growth, microbial species must balance this trade-off [22].
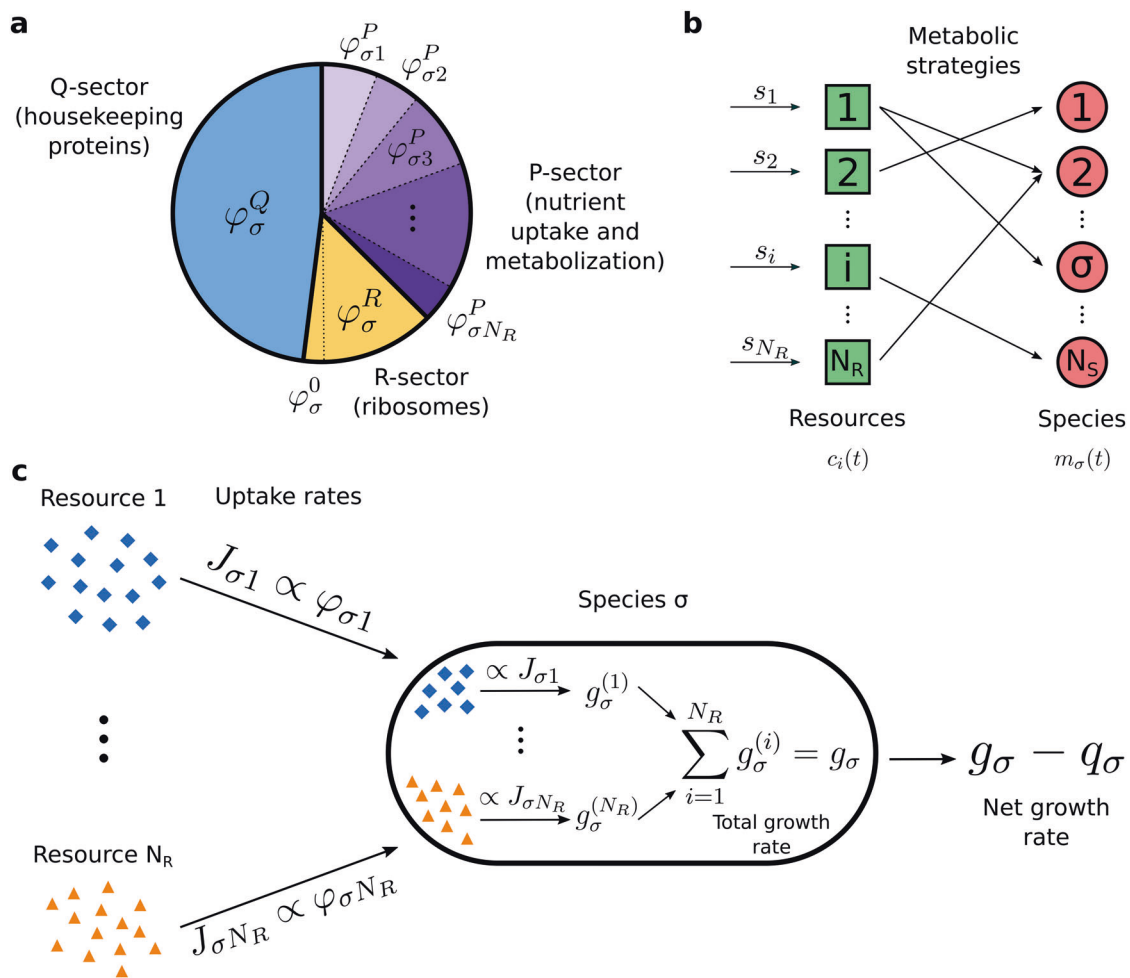
Fig. 1 Assumptions of the CPR model. a Generalization of Scott et al.'s [22] proteome subdivision to the case of N_R resources: the proteomic sector allocated by species σ for nutrient uptake and metabolization is subdivided into smaller fractions $\varphi_{\sigma i} = \varphi^P_{\sigma i}$, each dedicated 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". In our framework we are not considering the exchange of metabolic byproducts between species (i.e., cross-feeding). **c** Assumptions used to write the equations of the CPR model. 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. The net growth rate of species σ is the difference between the sum of these contributions and the maintenance cost q_σ.

Consumer-resource models

In Fig. 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. Species' growth rates are determined by the types and the amount of resources they uptake, and are 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 substrates that are supplied to the system with given (constant) rates s_i, and they are uptaken by species in the community. Overall, the model describes explicitly the dynamics of both species and resources through equations with the following structure:

$$\dot{m}_\sigma = m_\sigma(g_\sigma - q_\sigma) \qquad \sigma = 1, \dots, N_S, \qquad (3a)$$

$$\dot{c}_i = s_i - \sum_{\sigma=1}^{N_S} J_{\sigma i} m_\sigma \qquad i = 1, \dots, N_R, \qquad (3b)$$

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. The ways in which species uptake the available substrates are encoded in $J_{\sigma i}$ with parameters that in the literature are called "metabolic strategies" or "resource preferences". In particular, consumer-resource models are generally setup so that $J_{\sigma i} \propto \alpha_{\sigma i}$, with $\vec{\alpha}_\sigma = (\alpha_{\sigma 1}, \dots, \alpha_{\sigma N_R})$ the metabolic strategy (or resource preference) of species σ. Therefore, in the consumer-resource framework the interactions between species are indirect and mediated by the abundance of resources and the species' resource preferences. Other types of direct inter-specific interactions (like cross-feeding through the exchange of metabolic byproducts), though undoubtedly important in natural microbial ecosystems, are not addressed in this work.

The consumer-proteome-resource model

Here, we incorporate proteome allocation constraints into consumer-resource models and show that proteome fractions allocated to the uptake of different resources must vary with time as resource concentrations vary. Figure 1c depicts schematically the assumptions underlying the CPR model. Each species σ uptakes resource i with a rate $J_{\sigma i}$ that is proportional to the proteome fraction $\varphi_{\sigma i}^P$. Then, resource i accounts for 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 $g_\sigma^{(i)}$. This assumption is consistent with previous works [38, 39] that considered the proteome allocation introduced by Scott et al. [22] in systems with two substitutable resources. Eventually, we obtain the following mathematical model (see Materials and Methods for the detailed derivation):

$$\dot{m}_\sigma = m_\sigma \left[\sum_{i=1}^{N_R} \frac{\kappa_i^n}{\rho_\sigma} r_i(c_i) \varphi_{\sigma i} - q_\sigma \right], \tag{4a}$$

$$\dot{c}_i = s_i - \xi_i r_i(c_i) \sum_{\sigma=1}^{N_S} m_\sigma \varphi_{\sigma i}, \tag{4b}$$

$$\sum_{i=1}^{N_R} \varphi_{\sigma i} \left[1 + \frac{\kappa_i^n}{\kappa_\sigma^t} r_i(c_i) \right] = \Phi_\sigma, \tag{4c}$$

where we have written $\varphi_{\sigma i} = \varphi_{\sigma i}^P$ for simplicity. 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 metabolism and biomass synthesis, which is fixed as shown by Scott et al. [22]. These equations have the traditional

structure of a consumer-resource model given by Eqs. (3a) and (3b), 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 a species growing on a single resource, the parameters that are most easily measured experimentally are the per-biomass resource uptake rate $\xi r(c) \varphi_\sigma$ and the yield (expressed as biomass per grams of resource), which in our framework is given by $Y = \kappa^n / \rho \xi$ (see Supplementary Information).

Notice that the metabolic strategies in our framework correspond to the proteome fractions $\varphi_{\sigma i}$. If we interpreted the $\varphi_{\sigma i}$ as fixed parameters, the CPR model would be placed within the field of classic substitutable consumer-resource theory. However, we show below that the proteome fractions $\varphi_{\sigma i}$ are actually dynamical variables that vary according to the concentration of resources, and thus the CPR model constitutes a generalization of classic consumer-resource theory with substitutable resources, based on experimental evidence of microbial proteome allocation and growth. In the CPR model the proteome fractions are subject to the constraint encoded by Eq. (4c), which derives from the proteome finiteness given by Eq. (2). The expression of this constraint is significantly different from other ones that have been studied in the consumer-resource framework [40]. Posfai et al. [31], for example, considered a classic consumer-resource model with fixed metabolic strategies, and a metabolic constraint that in our notation would read $\sum_{i=1}^{N_R} \varphi_{\sigma i} = \Phi$, where the sum does not depend on the resource concentrations through $r_i(c_i)$, and it is assumed that $\Phi_\sigma = \Phi$ for all σ (i.e., the value of Φ_σ is exactly the same for all species). Such a model, however, cannot reproduce the fact that microbial species vary their metabolic strategies with time according to the concentration of resources, and the constraint $\sum_{i=1}^{N_R} \varphi_{\sigma i} = \Phi$ does not account for the fact that, as a species invests more resources in nutrient uptake and metabolization (the $\varphi_{\sigma i}$) to achieve a higher growth rate, such an investment must be balanced by an increased investment in ribosomal proteins (the φ_σ^R), both of which are constrained by the finiteness of the proteome.

The proteome finiteness constraint, as encoded by Eq. (4c), yields one important consequence that has important repercussions on the properties the CPR model. In particular, it implies that the proteome fractions $\varphi_{\sigma i}$ *cannot* be fixed parameters, but must change as the resources' concentrations c_i change, and therefore they must be dynamical variables. This can be easily seen by considering a system with only one resource, for which Eq. (4c) reads

$$\varphi_\sigma \left[1 + \frac{\kappa^n}{\kappa_\sigma^t} r(c) \right] = \Phi_\sigma, \tag{5}$$

and thus the φ_σ *must* change as functions of the resource concentration:

$$\varphi_\sigma = \frac{\Phi_\sigma}{1 + \frac{\kappa^n}{\kappa_\sigma^l} r(c)}. \tag{6}$$

In particular, φ_σ must *decrease* as the resource concentration c *increases* (recall that $r(c)$ is a monotonically increasing function). This occurs because if, for example, the available resource becomes scarce, cells will need to produce more catabolic proteins to meet their energy requirements. In the presence of multiple resources, the proteome finiteness constraint of Eq. (4c) implies that if the concentration of one resource c_j decreases, then either $\varphi_{\sigma j}$ or some of the $\varphi_{\sigma i}$ with $i \neq j$ *must* increase to satisfy the constraint, since Φ_σ is constant. Thus, it is *necessary* to introduce some form of dynamics on the proteome fractions that each species allocates for nutrient uptake and metabolization. This observation should not come as a surprise, given that microbes are known to adapt their proteome allocation and metabolic strategies according to which resources are available. Our approach is to require that all $\varphi_{\sigma i}$ evolve dynamically with a characteristic timescale to maximize the instantaneous growth rate of species σ in an adaptive process, while ensuring that the proteome finiteness constraint is satisfied at all times. The model equations and the mathematical details are discussed in the Materials and Methods.

Experimental example of the influence of proteome allocation on population dynamics

Traditional consumer-resource models do not account explicitly for proteome allocation to different tasks and assume that metabolic strategies are fixed with time. Here, we show experimentally that it is necessary to take into account proteome allocation within consumer-resource models to reproduce the dynamics of even the simplest 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. We introduced in strain $\sigma = 1$ a plasmid containing a Red Fluorescent Protein (RFP) whose expression could be controlled by adding to the medium Isopropyl β-D-1-thio-galactopyranoside (IPTG, a molecular mimic of allolactose that cannot be metabolized by *E. coli*). 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. We performed competition experiments at different concentrations of IPTG, measured the fluorescent protein production rates at these concentrations, and computed the selective advantage of strain 1 over strain 2, a measure for the difference in reproductive fitness between the two strains defined as:

$$S := \frac{d}{dt} \ln \frac{f}{1-f}, \tag{7}$$

where f is the relative abundance (or frequency) of strain 1, i.e. $f = m_1/(m_1 + m_2)$. The experiment is sketched in Fig. 2.

Figure 3a (magenta data points) shows that the selective advantage S decreased 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, Fig. 3d), which are proportional to φ_{iRFP}. In the absence of IPTG and at low concentrations of it, strain 1 outcompeted strain 2 ($S > 0$). At an IPTG concentration of ~30 µM, the two strains coexisted by maintaining a stable relative fraction for the duration of the experiment. At IPTG concentrations larger than 30 µM, strain 1 was outcompeted by strain 2 (i.e., $S < 0$). This experiment illustrates that, in the presence of the same concentration of a single resource, manipulating the proteome allocation of one of the two strains results in different outcomes for their competition dynamics. Consumer-resource theory, which neglects proteome allocation dynamics, would not be able to predict competition dynamics in these settings.

Figure 3 also shows the results of a second experiment performed with two different strains (cyan data points). These strains had different fluorescent protein combinations with respect to strains 1 and 2 (see Materials and Methods and Fig. S.1): strain 3 expressed constitutively a red fluorescent protein (mKate2Hyb) and carried a plasmid with an IPTG-inducible yellow fluorescent protein (Venus YFP), while strain 4 expressed constitutively the yellow fluorescent protein mVenus (see Materials and Methods). Also in these independent sets of experiments, the selective advantage decreased linearly as the protein production rate was increased over a broad range (the mean cell's fluorescence measured after 8 h at 105 µM IPTG was 16 times higher than at 0 µM IPTG, Fig. 3e). In this case, strain 3 always outcompeted strain 4, even at high concentrations of IPTG. This may be explained by the fact that the two proteins expressed by strains 1 and 3 have a different fitness cost (see Supplementary Information for more details).

It is natural to ask whether the CPR model can reproduce the results of our experiment. Applying the CPR framework to such a simple community, using assumptions consistent with our experimental settings (e.g., the fact that the strains
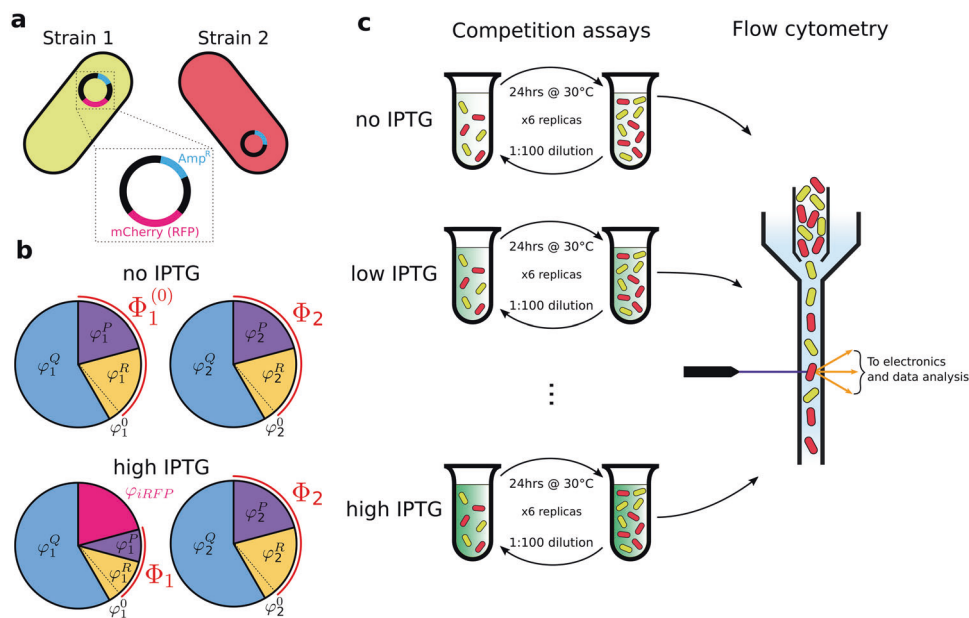
Fig. 2 Schematic representation of the experiment. a Two E. coli strains were used: strain 1 constitutively expresses a yellow fluorescent protein (mVenus) and carries a plasmid with the ampicillin resistance cassette (cyan Amp[R] in the plasmid magnification) and a red fluorescent protein (RFP), mCherry (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.

are grown in medium-rich conditions, and that they share the same genetic background), leads to the prediction that the selective advantage S of strain 1 over strain 2 is given by (see Materials and Methods):

$$S = \frac{d}{dt} \ln \frac{f}{1-f} \propto \Phi_1 - \Phi_2. \quad (8)$$

The same result could be obtained by assuming that the findings of Scott et al. [22] on how the exponential growth rate in isolation depends on proteome allocation can be applied to our experiment, in which cells were grown in co-culture dilution experiments and were not always in exponential phase. According to Eq. (8), 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 $S < 0$ and strain 2 outcompetes strain 1. Coexistence between the two strains is possible uniquely when $\Phi_1 = \Phi_2$ and thus $S = 0$. The system, therefore, exhibits two regimes where only one of the two strains survives (competitive exclusion), separated by the coexistence point $\Phi_1 = \Phi_2$. Equation (8) 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.

In our experiment, we forced strain 1 to produce a useless RFP at different rates depending on the IPTG concentration. Indicating with φ_{iRFP} the fraction of proteome allocated by strain 1 to the synthesis of the IPTG-inducible RFP (proportional to the fluorescent protein production rate), the proteome fraction allocated for nutrient uptake and growth is given by $\Phi_1 = \Phi_1^{(0)} - \varphi_{iRFP}$ (with $\Phi_1 = \Phi_1^{(0)}$ in the absence of IPTG). Thus, the selective advantage S is predicted to decay linearly with φ_{iRFP} as $S = \alpha - \beta \cdot \varphi_{iRFP}$ with α and β positive constants (see the Materials and Methods section for all details and the explicit expression of S in this case). This prediction is thus consistent with the experimental observation of a linear decrease of S with the fluorescent protein production rate.

Coexistence of multiple species in the consumer-proteome-resource model

We now analyze the CPR model in the general case of multiple species and multiple resources both analytically and numerically, to provide some insights into the conditions required for the coexistence of all species in the community. Specifically, we look for stationary solutions
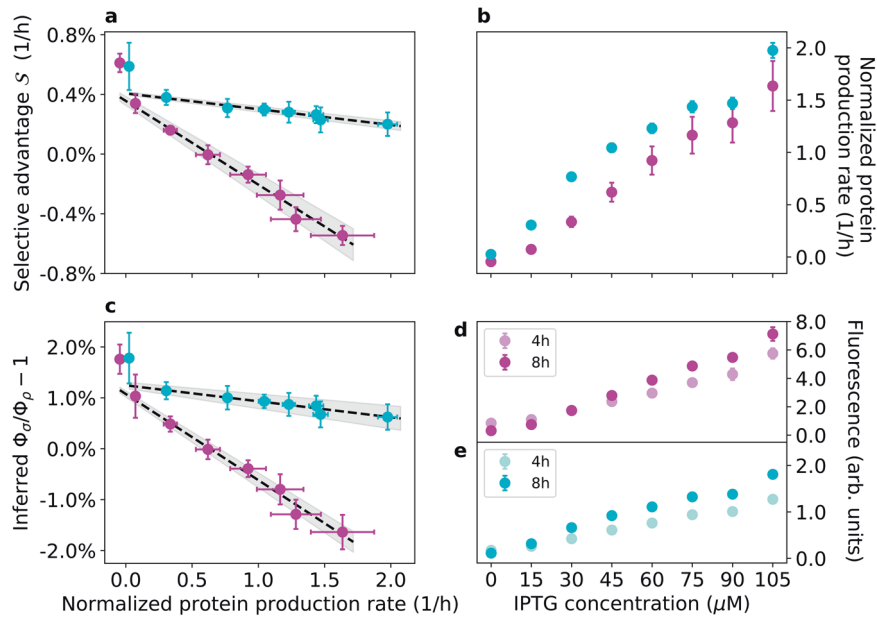
Fig. 3 Experimental results. Magenta points represent data from experiments with strains 1 and 2. Cyan points represent data from experiments with strains 3 and 4, where strain 3 expresses constitutively mKate2Hyb and the IPTG-inducible Venus yellow fluorescent protein (YFP) and strain 4 expresses mVenus constitutively (see Materials and Methods and Fig. 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 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 gray band represents the 68% confidence interval of the linear fit. The time series of $\ln[f/(1-f)]$ for the two experiments are reported in Figs. S.19 and S.20. **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 4 h and 8 h after inoculation in the conditions used for our experiment (see Materials and Methods). **e** Mean (induced) yellow fluorescence of strain 3 at 4 h and 8 h after inoculation in the conditions used for our experiment (see Materials and Methods).

where *all* species have non-null biomass densities. Doing so yields two necessary conditions for the coexistence of all species (see the Materials and Methods for all detailed expressions and computations). The first condition, which holds when there are more species than resources in the system ($N_S > N_R$), 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 condition is also required for all species to coexist if there are fewer species than resources ($N_S \leq N_R$) and all proteome fractions at stationarity $\varphi^*_{\sigma i}$ are larger than zero. If, instead, there are fewer species than resources ($N_S \leq N_R$) and some proteome fractions at stationarity are equal to zero, it is possible to find particular solutions for which all species coexist, without requiring $q_\sigma \propto \Phi_\sigma$. This happens, for example, when $N_S \leq N_R$ and the vectors $\vec{\varphi}^*_\sigma = \left(\varphi^*_{\sigma 1}, \dots, \varphi^*_{\sigma N_R}\right)$ are non-overlapping (i.e., $\vec{\varphi}^*_\sigma \cdot \vec{\varphi}^*_\rho = 0$ for $\sigma \neq \rho$), which means that each species uses resources

that are not used by other species. Further details can be found in the Materials and Methods.

The second condition, which holds in all the scenarios discussed in the previous paragraph, can be interpreted as follows using a graphical representation introduced by Posfai et al. [31] (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 Fig. 4. On this simplex one can draw the vectors $\vec{s}$ and $\vec{\hat{\varphi}}^*_\sigma$, whose components are appropriately rescaled versions of (respectively) the resource supply rates s_i and the stationary proteome fractions $\varphi^*_{\sigma i}$ (see Materials and Methods). The second condition for species coexistence prescribes, therefore, that $\vec{s}$ must belong to the *convex hull* of the vectors $\vec{\hat{\varphi}}^*_\sigma$, as shown in Fig. 4.

Notice that, differently from similar results of earlier investigations of consumer-resource models [31], this condition involves the *stationary* proteome fractions $\varphi^*_{\sigma i}$, and thus the community has the opportunity to coexist even if
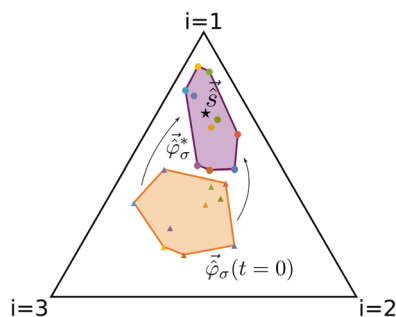
Fig. 4 Graphical representation of the second condition necessary for coexistence. Here we consider a system with $N_S = 10$ species and $N_R = 3$ resources (for ease of representation). In this case, the system can be represented on a bi-dimensional simplex (i.e., a triangle) where each vertex corresponds 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 $\vec{\hat{\varphi}}_\sigma^*$ 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 $\vec{\hat{\varphi}}_\sigma^*$ 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 $\vec{\hat{\varphi}}_\sigma^*$ depend on the relative ratios with which the resources are supplied or uptaken by the species.

the rescaled resource supply rate vector is not within the convex hull of the proteome fractions at the start of the temporal evolution.

Because the CPR model is highly non-linear, it is impossible to predict a priori the values of the stationary fractions $\varphi_{\sigma i}^*$ once all the other parameters are set. However, it is possible to understand how the various parameters affect the dynamics of the system by exploring different regions of the parameter space. The dynamics of the system, in fact, will depend on *how* the proteome fractions $\varphi_{\sigma i}$ evolve, and therefore the dynamics of the system will inevitably be influenced by some of the model parameters. In this sense the relevant parameters are the ratios $\gamma_{\sigma i} = \kappa_i^n / \kappa_\sigma^t$ 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 for details). The timescales τ_σ measure how fast the dynamics of the proteome fractions $\varphi_{\sigma i}$ vary: the smaller τ_σ is, 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 explored is $\tau_\sigma \gg 1$ and $\gamma_{\sigma i} \sim 0$. In this regime, the adaptive process that regulates the dynamics of the proteome fractions $\varphi_{\sigma i}$ is very slow (i.e., species 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 rescaled proteome fractions allocated by the species to nutrient uptake and metabolization change negligibly, and therefore all species survive only if the rescaled nutrient supply rate vector $\vec{\hat{s}}$ lies in the convex hull of the rescaled *initial* proteome fractions $\vec{\hat{\varphi}}_\sigma$, as shown in Fig. 5.

The second regime we explored is $\tau_\sigma \gg 1$ and $\gamma_{\sigma i} \gtrsim 1$. 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 if $\vec{\hat{s}}$ is not necessarily close to the convex hull of the initial $\vec{\hat{\varphi}}_\sigma$. On the other hand, we observed that if $\vec{\hat{s}}$ is too far away from the convex hull of the initial $\vec{\hat{\varphi}}_\sigma$ there might still 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. $s_i \rightarrow x s_i$ (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 a few species survive, whereas for larger values of 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 quality) coexistence will be possible even if $\vec{\hat{s}}$ lies at increasingly large distances from the convex hull of the initial $\vec{\hat{\varphi}}_\sigma$.

Finally, the last regime we explored is $\tau_\sigma \lesssim 1$, i.e. the adaptive process maximizing 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 independently of the initial values of the proteome fractions $\varphi_{\sigma i}$ and of the resource supply rates s_i. However, as the τ_σ grow, fewer and fewer species will be able to coexist. This can be seen by multiplying τ_σ by a factor $y > 1$: Fig. 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.

The results of this section can be summed up as follows. If metabolic adaptation is slow, i.e. if the characteristic relaxation times τ_σ of the proteome fractions $\vec{\varphi}_\sigma$ are large (or in other words, if the species shift slowly between different resources), coexistence will be favored if the system
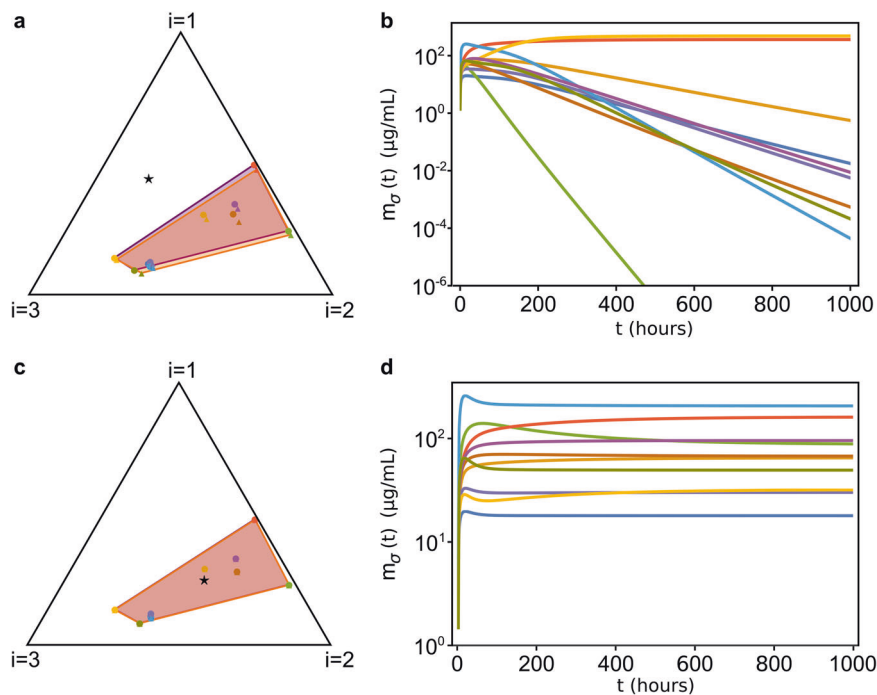
Fig. 5 Temporal evolution of the CPR model when $\tau_\sigma \gg 1$ and $\gamma_{\sigma i} \sim 0.1$. a Initial conditions for the $\vec{\varphi}_\sigma$ of a system with 10 species and 3 resources, depicted using the same graphical representation [31] of Fig. 4: the black triangle is the simplex to 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 $\vec{\hat{\varphi}}_\sigma$. **d** Biomass 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 Supplementary Information). All parameters other than $\vec{\hat{s}}$ are identical in the four panels.

contains high-quality nutrients (i.e., the $\gamma_{\sigma i}$ have larger values). If the system contains low-quality nutrients, coexistence will be possible only if the resources are supplied in particular ratios that depend on the species' proteome allocation. In particular, coexistence will be possible if the rescaled nutrient supply rate vector $\vec{\hat{s}}$ lies inside the convex hull of the rescaled proteome fractions $\vec{\hat{\varphi}}_\sigma$. On the other hand, fast metabolic adaptation (i.e., small values of τ_σ) always favor coexistence.

Discussion

Motivated by our experiment that shows how varying proteome allocation can have strong effects on the dynamics of even 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 bridged microbial growth with proteome allocation constraints in competitive communities, and we have investigated the conditions that lead to species coexistence in the presence of multiple resources.

This CPR 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. The main contribution of this work is introducing a physiological, experimentally-validated constraint on the amount of resources that cells can devote to growth and metabolism in consumer-resource models (i.e., Eq. (15c)). The introduction of this constraint makes it *necessary* to introduce some dynamics on the proteome fractions allocated for nutrient uptake and metabolization, and we have done so using an adaptive approach that assumes that microbial species are evolutionary well adapted to their environment. This work differs (both in scope and approach) from previous ones that involve adaptation on some species' internal variables [41], and in particular differs from previous works involving the consumer-resource framework [31, 40] that considered phenomenological constraints that were not based on direct experimental measurements, nor on an interpretation of such constraints as arising from the finiteness of the proteome. 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
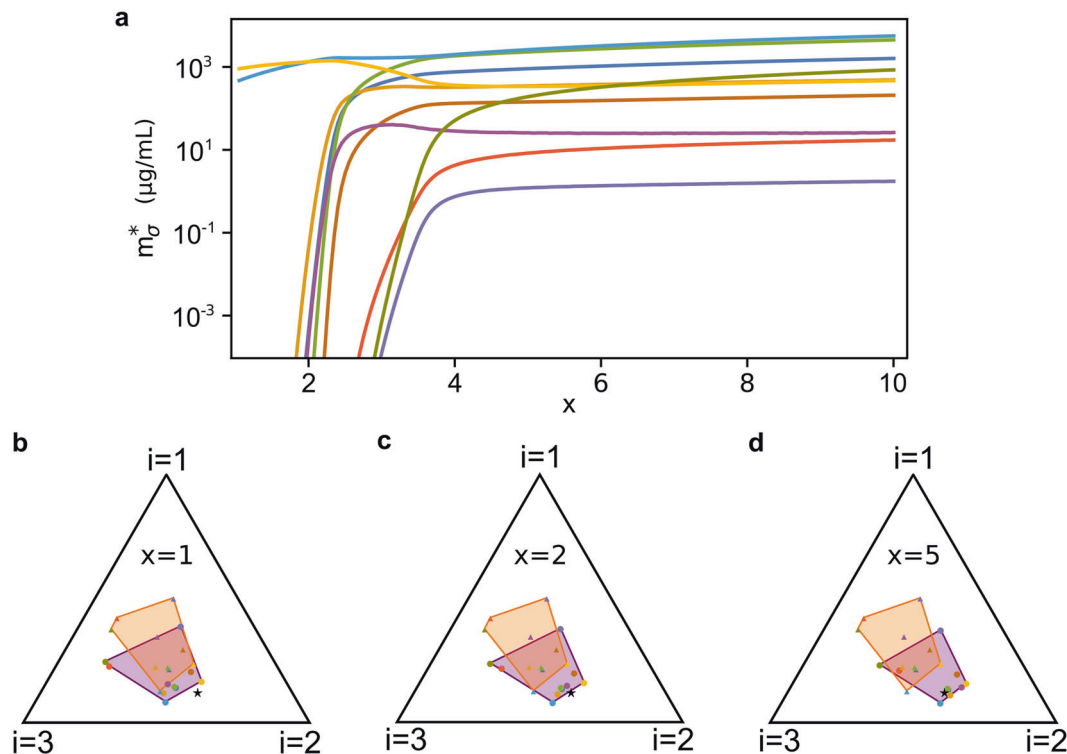
Fig. 6 Species coexistence as a function of the rescaled resource supply rate $x\vec{s}$ (with $x > 1$). As for Fig. 5, the $\vec{\varphi}_\sigma$ evolve according to the CPR model with $\tau_\sigma \gg 1$, $\gamma_{\sigma i} \gtrsim 1$, $N_S = 10$ and $N_R = 3$. Here, $\vec{\hat{s}}$ was drawn randomly outside the convex hull of the initial $\vec{\hat{\varphi}}_\sigma$ (same $\vec{\hat{s}}$ 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 in which 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 ratio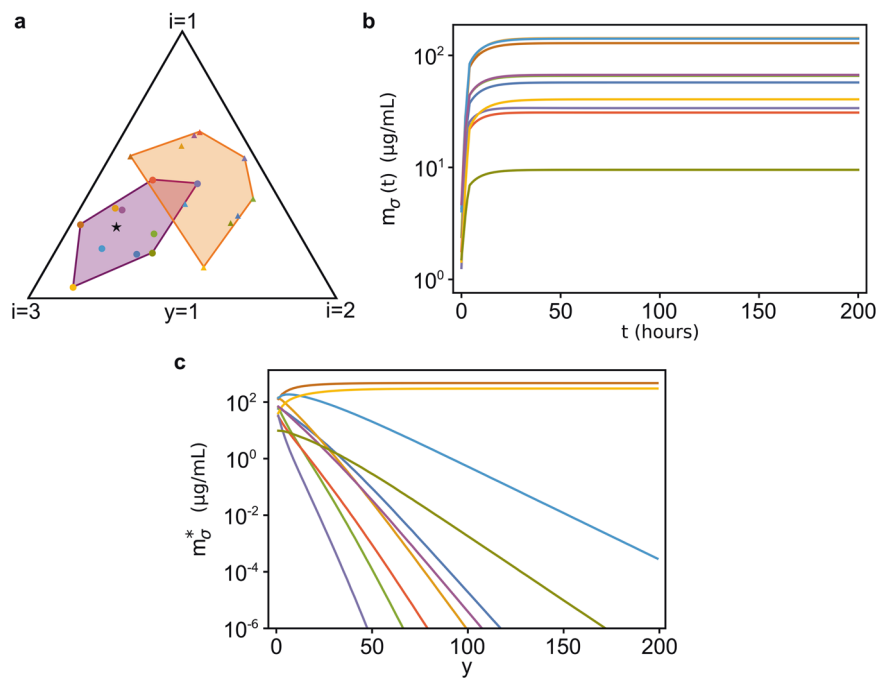s of the stationary abundances m_σ^* are not constant as x grows. **b–d** Initial (orange) and stationary (purple) convex hull of the rescaled proteome fractions $\hat{\varphi}_{\sigma i}$ for different values of x. For small x, the resource supply (black star) is not large enough to allow the $\hat{\varphi}_{\sigma i}$ to move so that the coexistence condition is satisfied. Increasing x (d), this becomes possible and thus all species are able to coexist. The parameters and the initial conditions were drawn from pre-assigned random distributions (see Supplementary Information). All parameters other than $\vec{\hat{s}}$ and the initial conditions $m_\sigma(0)$ and $c_i(0)$ are identical in the four panels.

and Methods). A further discussion on the differences between the CPR model and previous ones can be found in the Supplementary Information.

We have then shown that the CPR model predicts that high levels of biodiversity can be achieved only if two conditions apply. The first condition 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 condition 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 $\vec{\hat{s}}$ belongs to the convex hull of $\vec{\hat{\varphi}}_\sigma$); (ii) if again $\tau_\sigma \gg 1$, but the resources are of higher quality, coexistence is possible (in some cases the magnitude of the resource supply rates must be large enough), and if the resources' quality is higher, coexistence is favored; and (iii) coexistence is favored for smaller values of the timescales τ_σ. From the

biological point of view, these points can be interpreted as follows: (i) if the species switch slowly between different resources and the quality of the resources is low, coexistence will be possible only if the resources are supplied with particular ratios (which depend on the proteome allocation of all the species); (ii) if again the species switch slowly between different resources, coexistence will be favored if the resources have higher quality; (iii) fast metabolic adaptation (i.e., the species can switch quickly between different resources) favors coexistence. Our approach, therefore, makes it possible to quantify precisely in what ways the internal cellular dynamics make coexistence possible in a broad range of environmental contexts.

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

Fig. 7 Coexistence in the CPR model is affected by the values of the adaptation timescales τ_σ. a Initial (orange) and stationary (purple) convex hull of the rescaled proteome fractions $\hat{\varphi}_{\sigma i}$, for a system with $N_S = 10$ and $N_R = 3$ when $\tau_\sigma \lesssim 1$. **b** Temporal biomass dynamics of the system represented in **a**. **c** Increasing the values of τ_σ leads to extinctions. Shown are the stationary values of species' biomasses calculated by multiplying the values τ_σ of panels **a**, **b** by a factor $y > 1$, while keeping the other parameters unchanged. As y increases, the system shifts from a diverse stationary state for $y = 1$ to states in which only few species survive. The parameters and the initial conditions were drawn from pre-assigned random distributions (see Supplementary Information for more information). All parameters other than y are identical in the three panels.

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 [42] 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 derivation of the CPR models equations starts from Eqs. (3a) and (3b). To write these equations explicitly, we introduce the following assumptions: (i) the uptake rate $J_{\sigma i}$ is proportional to the proteome fraction $\varphi_{\sigma i} = \varphi_{\sigma i}^P$ 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 $g_\sigma^{(i)}$ 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 $g_\sigma^{(i)}$. Specifically, we rewrite Eq. (1a) as:

$$\varphi_{\sigma i}^P = \frac{\rho_\sigma}{\overline{\kappa}_i^n(c_i)} g_\sigma^{(i)}, \tag{9}$$

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.:

$$g_\sigma = \sum_{i=1}^{N_R} g_\sigma^{(i)}, \tag{10}$$

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

$$g_\sigma = \sum_{i=1}^{N_R} \frac{\overline{\kappa}_i^n(c_i)}{\rho_\sigma} \varphi_{\sigma i}^P, \tag{11a}$$

$$\varphi_\sigma^R = \frac{\rho_\sigma}{\kappa_\sigma^t} g_\sigma + \varphi_\sigma^0. \tag{11b}$$

Equation (10) 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. (11a) to write Eq. (11b) in terms of the fractions $\varphi_{\sigma i}$. By doing so we get that the normalization condition given by Eq. (2) reads:

$$\sum_{i=1}^{N_R} \varphi_{\sigma i}\left[1 + \frac{\overline{\kappa}_i^n(c_i)}{\kappa_\sigma^t}\right] = 1 - \varphi_\sigma^Q - \varphi_\sigma^0 := \Phi_\sigma, \tag{12}$$

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 biomass synthesis.

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.:

$$J_{\sigma i} = \xi_i r_i(c_i) \varphi_{\sigma i}, \tag{13}$$

where the proportionality constant ξ_i can be interpreted biologically as the maximum catalytic rate of the enzyme used to metabolize resource i (see Supplementary Information). By comparing Eqs. (13) and (9) 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

$$\chi_{\sigma i} \xi_i = \frac{\kappa_i^n}{\rho_\sigma}. \tag{14}$$

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

$$\dot{m}_\sigma = m_\sigma\left[\sum_{i=1}^{N_R} \eta_{\sigma i} r_i(c_i) \varphi_{\sigma i} - q_\sigma\right], \tag{15a}$$

$$\dot{c}_i = s_i - \xi_i r_i(c_i) \sum_{\sigma=1}^{N_S} m_\sigma \varphi_{\sigma i}, \tag{15b}$$

$$\sum_{i=1}^{N_R} \varphi_{\sigma i}[1 + \gamma_{\sigma i} r_i(c_i)] = \Phi_\sigma, \tag{15c}$$

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 $r(c)$ chosen, for our purposes we only need to assume that $r(c)$ is a monotonically increasing function of c, and that $\lim_{c\to 0} r(c)/c = 1/K$ and $\lim_{c\to\infty} r(c) = 1$.

The constraint in Eq. (15c) is the explicit expression of Eq. (2) in our framework, and can be interpreted geometrically: considering species σ, the N_R-dimensional vector $\vec{\varphi}_\sigma = (\varphi_{\sigma 1}, \dots, \varphi_{\sigma N_R})$ 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. (15c) 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 $\varphi_{\sigma i}$. This constraint reflects the well known fact that microbes can vary their enzyme synthesis with time and switch between nutrients according to environmental conditions [40, 43–45].

Dynamics of the proteome fractions $\varphi_{\sigma i}$

We call $\vec{c} = (c_1, \dots, c_{N_R})$ the vector of resource concentrations and define

$$F_\sigma(\vec{\varphi}_\sigma, \vec{c}) := \sum_{i=1}^{N_R} \varphi_{\sigma i}[1 + \gamma_{\sigma i} r_i(c_i)] - \Phi_\sigma \tag{16}$$

so that the constraint given by Eq. (15c) 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 $\vec{\varphi}_\sigma$ must satisfy

$$\dot{F}_\sigma(\vec{\varphi}_\sigma, \vec{c}) \equiv \dot{\vec{\varphi}}_\sigma \cdot \vec{\nabla}_\varphi F_\sigma + \dot{\vec{c}} \cdot \vec{\nabla}_c F_\sigma = 0, \tag{17}$$

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. (17) is therefore:

$$\dot{\vec{\varphi}}_\sigma = -\frac{\vec{\nabla}_\varphi F_\sigma}{(\vec{\nabla}_\varphi F_\sigma)^2} \vec{c} \cdot \vec{\nabla}_c F_\sigma, \tag{18}$$

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 Supplementary Information 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 [43]. For this reason, we can implement an adaptive approach [40] 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. Equation (15c) is satisfied. In this case the evolution equation for $\vec{\varphi}_\sigma$ becomes:

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

(19)

where we have introduced τ_σ, the characteristic timescale over which $\vec{\varphi}_\sigma$ changes [40] (detailed computations are shown). We can recover Eq. (18) from Eq. (19) by sending τ_σ to infinity. Geometrically, Eq. (18) 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. (15c). On the other hand, according to Eq. (19) (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. (15c).

In this work we have used a generalization of Eq. (19) that ensures $\varphi_{\sigma i}(t) \geq 0 \forall t$, and varied the values of τ_σ when needed (see Supplementary Information for details).

The introduction of this dynamics on the proteome fractions $\varphi_{\sigma i}$ in consumer-resource models allows our model to reproduce phenomena that classic consumer-resource theory cannot describe, like diauxic shifts (see Fig. S.2).

Conditions for coexistence

Evaluating Eqs. (15a)–(15c) at stationarity we obtain:

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

(20a)

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

(20b)

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

(20c)

where we are denoting with the symbol "*" the quantities computed at stationarity, and we have assumed $m_\sigma \neq 0$. If we now assume $\varphi_{\sigma i}^* \neq 0$ for all i and all species, it is easily seen by substitution that a possible solution for r_i^* in Eqs. (20a) and (20c) is

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

(21)

Under our assumption (i.e., $\varphi_{\sigma i}^* \neq 0$ for all i, for all species), and if $N_S > N_R$ (i.e., the number of species is larger than the number of resources) this solution is acceptable only if its right-hand side is independent of σ, i.e. if

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

(22)

with Θ some given constant independent of σ. Using Eqs. (21) and (22) in Eqs. (20c) or (20a) we get

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

(23)

From Eq. (21) we have:

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

(24)

and since we need $r_i^* < 1$ (or equivalently $c_i^* > 0$), we need $\Theta > \max_i 1/\kappa_i^n$. Therefore, Eq. (22) can be rewritten as

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

(25)

which is the explicit expression of the relationship between q_σ and Φ_σ. Equation (23) is a consequence of the system's constraint in Eq. (20c), which is Eq. (15c) computed at stationarity. Therefore, the expression of the maintenance cost given in Eq. (25) is a consequence of the constraint introduced in the CPR model.

Notice, again, that this holds under the assumption that $\varphi_{\sigma i}^* \neq 0$ for all i and σ, and $N_S > N_R$. If we remove these assumptions, then it is possible to find solutions with $N_S \leq N_R$ where Eq. (22) does not hold. For example, if the species' stationary proteome fractions $\vec{\varphi}_\sigma^*$ are non-overlapping (i.e., $\vec{\varphi}_\sigma^* \cdot \vec{\varphi}_\rho^* = 0$ when $\sigma \neq \rho$), then r_i^* as given in Eq. (21) can be a valid solution without requiring Eq. (22). Consider as an example the particular case $N_S = N_R = 3$ and $\varphi_{\sigma i}^* \propto \delta_{\sigma i}$ (where δ is Kronecker's delta), i.e. a system with three species where each one uptakes only one resource, and no two species uptake the same resource. It is easy to imagine that the three species should be able to coexist, since their niches (defined in this context as the set of resources used for sustenance) do not overlap. This is indeed the case, given that a solution for r_i^* in Eqs. (20a) and (20c) is given by:

$$r_i^* = \left[\kappa_i^n \left(\frac{\Phi_i}{\rho_i q_i} - \frac{1}{\kappa_i^t} \right) \right]^{-1},$$

(26)

where we have identified each species index σ with the only resource i it consumes, and we don't need to require

Eq. (22) to hold for this solution to be feasible. This will be of course true even for systems where the species and/or resource labels are permuted (e.g., species 1 uptakes resource 2, species 2 uptakes resource 3 and species 3 uptakes resource 1, instead of species 1 uptaking resource 1, species 2 uptaking resource 2, and species 3 uptaking resource 3). This will be true even when $N_R > N_S$, as long as the vectors $\vec{\varphi}_\sigma^*$ are still non-overlapping and the inverse of r_i is written as the product of κ_i^n and $\Phi_\sigma/(\rho_\sigma q_\sigma) - 1/\kappa_\sigma^t$ where σ is the (only) species uptaking that resource. If one of the resources, e.g. resource j, is not uptaken by any species one has $\dot{c}_j = s_j$, i.e. c_j will grow linearly indefinitely. On the other hand, if $N_S > N_R$ then Eq. (22) is necessary in order to have feasible solutions, even if we remove the assumption that $\varphi_{\sigma i}^* = 0$ for all species and resources.

Going back to Eqs. (20a)–(20c), if we now define:

$$\hat{s}_i = \frac{s_i \kappa_i^n / \xi_i}{\sum_{j=1}^{N_R} s_j \kappa_j^n / \xi_j} \tag{27a}$$

$$\hat{\varphi}_{\sigma i}^* = \frac{\varphi_{\sigma i}^*}{\sum_{j=1}^{N_R} \varphi_{\sigma j}^*}, \tag{27b}$$

$$z_\sigma = \frac{m_\sigma^* \rho_\sigma q_\sigma}{\sum_\lambda m_\lambda^* \rho_\lambda q_\lambda}, \tag{27c}$$

(so that z_σ are positive coefficients that sum to one), and Eq. (20b) can be rewritten as

$$\hat{s}_i = \sum_{\sigma=1}^{N_S} z_\sigma \hat{\varphi}_{\sigma i} \tag{28}$$

(see Supplementary Information 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. (28) means that $\vec{\hat{s}}$ belongs to the convex hull of the vectors $\vec{\hat{\varphi}}_\sigma^*$. Since Eq. (28) derives from requiring that all species have non-null stationary biomasses, we can see how this is the other condition necessary for coexistence.

At first glance, the result in Eq. (28) looks similar to what has been observed in consumer-resource model with metabolic trade-offs by Posfai et al. [31]. However, our result has an important difference with respect to that model: Eq. (28) depends in fact on the (rescaled) value of $\varphi_{\sigma i}$ at stationarity. In the CPR model, therefore, the proteome fractions $\varphi_{\sigma i}$ vary over time to satisfy Eq. (28), i.e. to include $\vec{\hat{s}}$ in the convex hull of the vectors $\vec{\hat{\varphi}}_\sigma^*$, unlike in Posfai et al. [31] where metabolic strategies (which in our framework correspond to the $\varphi_{\sigma i}$) are fixed and thus coexistence is only possible if $\vec{\hat{s}}$ is within the convex hull of the $\varphi_{\sigma i}$ from the very start.

If we now suppose that $\tau_\sigma \gg 1$, so that we can use Eq. (18) 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

$$\left(\vec{\nabla}_\varphi F_\sigma\right)_i \equiv \frac{\partial F_\sigma}{\partial \varphi_{\sigma i}} = 1 + \gamma_{\sigma i} r_i(c_i) \tag{29a}$$

and

$$\left(\vec{\nabla}_c F_\sigma\right)_i \equiv \frac{\partial F_\sigma}{\partial c_i} \propto \gamma_{\sigma i}, \tag{29b}$$

we find that if $\gamma_{\sigma i} \sim 0$ then $\dot{\vec{\varphi}}_\sigma \sim 0$ and therefore $\varphi_{\sigma i}^* \sim \varphi_{\sigma i}(t = 0)$. 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. (28) 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 $\gamma_{\sigma i} \gtrsim 1$ as discussed in the Results section, on the other hand, 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. (28) 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.

Figures S.3–S.14 show the results of fitness assays performed with all the strains used in our experiments and the ancestor strains.

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 glucose and ampicillin at eight different IPTG concentrations: 0, 15, 30, 45, 60, 75, 90, 105 μM (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 h at 30 °C on a microplate shaker set at 1050 rpm.
4. After 24 h, 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 h with constant shaking at 1050 rpm, 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 h with constant shaking at 1050 rpm. 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 (Fig. 3d, e). 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 I of a cell (due to the production of the IPTG-inducible fluorescent protein) in between two successive cell divisions thus satisfies $dI/dt = k(C_I) - d_{FP}I$. At a cell division event, the fluorescent intensity of a cell is reduced by a factor 2. Indicating with I_0 the cell's fluorescent intensity at the first measurement time ($t = 4$ h), it can be shown (see Supplementary Information) that according to this model the cell's fluorescent intensity changes with time as:

$$I(t) = 2I_0 e^{-gt\left(1+\frac{d_{FP}}{g}\right)} + \frac{k(C_I)}{d_{FP}}\left[1 - e^{-gt\left(1+\frac{d_{FP}}{g}\right)}\right] \\ \left(1 - \frac{1}{2^{1+\frac{d_{FP}}{g}} - 1}\right), \tag{30}$$

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. (30) as:

$$I(t) = 2I_0 e^{-gt} + 2\ln(2)\frac{k(C_I)}{g}(1 - e^{-gt}). \tag{31}$$

We used Eq. (31) and the data in Fig. 3d, e 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 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 $\langle I \rangle$ of cells measured in the absence of IPTG at the first measurement in the calibration experiment (see Fig. 3d, e). 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 Fig. 3 cannot be compared directly.

The normalized $k(C_I)/\langle I \rangle$ are the protein production rates of strains 1 and 3 (with dimensions 1/time) reported in Fig. 3.

The growth curves and the growth rates of strains 1 and 3 for the different IPTG concentrations used in our experiments are shown in Figs. S.15–S.18.

Estimation of the selection coefficient S

To first approximation, we can use the results of Scott et al. [22] on the dependence of the exponential growth rate of *E. coli* strains grown in isolation in rich medium (which in our notation corresponds to $r(c) = 1$) to estimate the outcome of our competition experiment, and in particular to estimate the

dependence of the selection coefficient on the Φ_σ. From Eq. (11b) the growth rate of species σ is given by:

$$g_\sigma = \frac{\kappa_\sigma^t}{\rho_\sigma}\left(\varphi_\sigma^R - \varphi_\sigma^0\right). \tag{32}$$

Using also Eq. (11a) with $N_R = 1$ and the definition of Φ_σ from Eq. (12), we can rewrite this as:

$$g_\sigma = \frac{\kappa_\sigma^t}{\rho_\sigma}\left(\Phi_\sigma - \frac{\rho_\sigma}{\kappa^n}g_\sigma\right), \tag{33}$$

which is easily rearranged into:

$$g_\sigma = \frac{\kappa_\sigma^t \kappa^n}{\rho_\sigma\left(\kappa_\sigma^t + \kappa^n\right)}\Phi_\sigma \tag{34}$$

(see also Eq. (S23) in [22, Online Supporting Material]). Therefore, if we assume $\kappa_1^t = \kappa_2^t$ and $\rho_1 = \rho_2$ (which can happen, for example, if the two populations are different strains of the same microbial species with similar genetic backgrounds) and $m_\sigma(t) = m_\sigma(0)\exp(g_\sigma t)$ (which is a good approximation for populations growing in batch cultures with nutrient-rich medium), the selection coefficient is given by:

$$S = \frac{d}{dt}\ln\frac{f}{1-f} = \frac{\dot{m}_1}{m_1} - \frac{\dot{m}_2}{m_2} = \frac{\kappa^t \kappa^n}{\rho(\kappa^t + \kappa^n)}(\Phi_1 - \Phi_2), \tag{35}$$

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. The time series of the values of $\ln(f/1 - f)$ for the two experiments are shown in Figs. S.19 and S.20.

If we now apply the CPR model, i.e. Eqs. (15a)–(15c), to the case of two populations and one resource, we obtain:

$$\dot{m}_\sigma = m_\sigma(\eta_\sigma r(c)\varphi_\sigma - q_\sigma) \qquad \sigma = 1, 2, \tag{36a}$$

$$\dot{c} = s - \xi r(c)(m_1\varphi_1 + m_2\varphi_2), \tag{36b}$$

$$\varphi_\sigma(1 + \gamma_\sigma r(c)) = \Phi_\sigma \qquad \sigma = 1, 2, \tag{36c}$$

where $\eta_\sigma = \kappa^n/\rho_\sigma$, $\gamma_\sigma = \kappa^n/\kappa_\sigma^t$, and now Eq. (36c) 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 0Y and 0R) have the same genetic background (see, for example, Figs. S.3 and S.4), we set $\eta_1 = \eta_2 = \eta$, $q_1 = q_2 = q$ and $\gamma_1 = \gamma_2 = \gamma$ in Eqs. (36a)–(36c). Notice 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. Furthermore, note that assuming $\eta_1 = \eta_2$ is equivalent to assuming $\rho_1 = \rho_2$, and on the other hand $\gamma_1 = \gamma_2$ is equivalent to $\kappa_1^t = \kappa_2^t$. 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. (36b) and we are left with:

$$\dot{m}_\sigma = m_\sigma \frac{\eta}{1+\gamma}\Phi_\sigma \sigma = 1, 2. \tag{37}$$

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. (36c). We, therefore, have that the expression of the selective advantage S in this case is:

$$S = \frac{d}{dt}\ln\frac{f}{1-f} = \frac{\dot{m}_1}{m_1} - \frac{\dot{m}_2}{m_2} = \frac{\eta}{1+\gamma}(\Phi_1 - \Phi_2). \tag{38}$$

From the definitions of $\eta = \kappa^n/\rho$ and $\gamma = \kappa^n/\kappa^t$ it is immediate to see that the coefficient in Eq. (38) is the same as the one in Eq. (35).

With our framework, however, we can show that this result continues to be true even when we remove the assumption that $r(c) = 1$ at all times. In our experiment, for example, it was not true that glucose was always abundant throughout the experiment, since the density of the cells saturated well before the following re-inoculation in fresh medium was made (i.e., 24 h). In fact, the typical growth rate of the strains, estimated from growth curves measured in the same experimental conditions used for the competition assays, is 0.3 1/h. The competition assays started from a cellular density of ~$8 \cdot 10^6$ cells/mL, thus if growth was exponential the density after 24 h would have been be ~$1.4 \cdot 10^{10}$, which is much higher than the typical density (~10^9 cells/mL) that E. coli cells reach at saturation. With a growth rate of 0.3 1/h, the time needed to reach a cellular density that is hundredfold the initial one (and therefore the time needed to reach saturation after a re-inoculation) is ~15.4 h.

A model better suited to describe the population dynamics of the two strains in our experiment would be as follows. The temporal dynamics of biomass and glucose concentration between two consecutive dilutions satisfies:

$$\dot{m}_\sigma = m_\sigma\eta_\sigma r(c)\varphi_\sigma \qquad \sigma = 1, 2 \tag{39a}$$

$$\dot{c} = -r(c)(m_1\eta_1\varphi_1 + m_2\eta_2\varphi_2), \tag{39b}$$

where $c(t)$ is the concentration of glucose at time t and $r(c) = c/(c + K)$ is Monod's function. This model is somewhat similar to a classic consumer-resource model, with the

difference that there is no mortality term in Eq. (39a): between two consecutive dilutions, the biomass m_σ of strain σ will grow as long as there is glucose available, and because $r(0) = 0$ the strains will stop growing (i.e. they will enter the stationary phase) once glucose runs out.

We now make the following approximation: we assume that, after every reinoculation, glucose is initially abundant (i.e. $r \sim 1$) and that the transition of $r(c)$ from 1 to 0 as c decreases is abrupt, which happens if K is sufficiently small. In other words, we assume that K is sufficiently small so that $r(c) \approx 1$ until a given time T (the instant at which glucose is completely depleted), when $r(c)$ abruptly goes to zero (i.e., $r(c(t)) \sim H(T - t)$ with H the Heaviside's step function). This means that after a reinoculation m_σ will grow exponentially for a time interval of length T, after which it will stop until the next dilution. If we still set $\eta_1 = \eta_2 = \eta$ and $\gamma_1 = \gamma_2 = \gamma$ and call D the dilution factor between reinoculations, we have that the biomass $m_\sigma^{(N)}$ of strain σ at the N-th dilution is ($m_\sigma^{(0)}$ being the biomass at the initial inoculation):

$$m_\sigma^{(1)} = m_\sigma^{(0)} \exp\left(\frac{\eta}{1+\gamma}\Phi_\sigma T\right), \tag{40a}$$

$$m_\sigma^{(2)} = \frac{m_\sigma^{(1)}}{D}\exp\left(\frac{\eta}{1+\gamma}\Phi_\sigma T\right) = \frac{m_\sigma^{(0)}}{D}\exp\left(2\frac{\eta}{1+\gamma}\Phi_\sigma T\right), \tag{40b}$$

$$m_\sigma^{(3)} = \frac{m_\sigma^{(2)}}{D}\exp\left(\frac{\eta}{1+\gamma}\Phi_\sigma T\right) = \frac{m_\sigma^{(0)}}{D}\exp\left(3\frac{\eta}{1+\gamma}\Phi_\sigma T\right), \tag{40c}$$

$$\vdots$$

$$m_\sigma^{(N)} = \frac{m_\sigma^{(0)}}{D^{N-1}}\exp\left(N\frac{\eta}{1+\gamma}\Phi_\sigma T\right). \tag{40d}$$

Therefore, if we call $f^{(N)}$ the relative abundance of strain 1 at the N-th dilution, we have:

$$\ln\frac{f^{(N)}}{(1-f)^{(N)}} = \ln\frac{m_1^{(0)}}{m_2^{(0)}} + NT\frac{\eta}{1+\gamma}(\Phi_1 - \Phi_2), \tag{41}$$

which gives the same expression for the selection coefficients after deriving with respect to the time NT.

Comments on the experimental selection coefficient S

Figure 3a shows that strain 1 has a fitness advantage over strain 2 in the absence of IPTG, since $S > 0$ at low protein production rates, even though the only significant difference between the two strains is that strain 1 carries an extra copy of *lacI* and the inducible fluorescent protein mCherry (see

"Strains used in the experiment"); in our theoretical framework, such an advantage implies that $\Phi_1^{(0)} - \Phi_2 > 0$. This may be explained by the observation that expressing *lacI* is beneficial for *E. coli* strains growing on glucose because it represses expression of the *lac* operon. Stoebel et al. [46] have in fact found that cells with the genomic copy of *lacI* show some residual *lacA* activity when grown in glucose, and estimated the cost of expressing *lacA* as 1.85% per generation [46], which may be alleviated in the presence of an extra copy of *lacI*. See also Supplementary Information for a more detailed discussion. Using our data, it is possible to estimate the ratio Φ_1/Φ_2 at different protein production rates (Fig. 3c). This ratio is approximately $\Phi_1^{(0)}/\Phi_2 \approx 1.02$ for low protein production rates and then decays linearly up to $\Phi_1/\Phi_1 \approx 0.98$.

In the first set of experiments (magenta points in Fig. 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 $0\,\mu$M IPTG) appear to deviate from the linear trend, and so the fits in Fig. 3a–c were calculated by excluding those data points (including them in the fit doesn't affect the results, see Fig. S.21). 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 $0\,\mu$M IPTG, which may partly explain the deviation of the first magenta point in Fig. 3a from the linear trend via a reduction in protein production rate throughout the experiment at $0\,\mu$M 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 [47], which might confer heterogeneous growth rates to different cells in the population. Note that the normalized protein production rates of the two sets of experiments (magenta and cyan data points) are not directly comparable.

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. (37) the growth rate of strain 1 is $g_1[k(C_I)] = \Phi_1[k(C_I)] \cdot \eta r(c)/(1 + \gamma r(c))$ (where we have inserted explicitly the dependence on $r(c)$, and $k(C_I)$ is the protein production rate induced by C_I). On the other hand, the expression of the selective advantage for general values of $r(c)$ is:

$$S = \frac{\eta \cdot r(c)}{1 + \gamma \cdot r(c)}(\Phi_1 - \Phi_2) \tag{42}$$

(in fact, if we only remove the assumption that $r(c) \approx 1$, from Eq. (36a) we have $\dot{m}_\sigma/m_\sigma = \eta_\sigma r(c)\varphi_\sigma$ with $\sigma = 1, 2$ and the definition of S leads to this equation). Dividing S in

Eq. (42) by g_1, for any value of $r(c)$ we obtain:

$$\frac{S[k(C_I)]}{g_1[k(C_I)]} = \frac{\Phi_1[k(C_I)] - \Phi_2[k(C_I)]}{\Phi_1[k(C_I)]}, \tag{43}$$

which is easily rearranged into:

$$\frac{\Phi_1}{\Phi_2}[k(C_I)] = \frac{1}{1 - S[k(C_I)]/g_1[k(C_I)]} \tag{44}$$

(which are the values plotted in Fig. 3c). Notice again that this result does *not* depend on the assumption that $r(c) = 1$ at all times, i.e. Eq. (44) is valid for any value of $r(c)$.

Proteome fraction allocated to the inducible 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. [48, 49] 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 [50] 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. (42) 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 $S \approx 1.1 \cdot 10^{-2}$. Finally, with these calculations we can estimate the maximum percentage of proteome $\max\varphi_{iRFP}$ and $\max\varphi_{iYFP}$ allocated at full expression to the production of, respectively, the inducible red and yellow proteins in our two experiments. In particular, for the first experiment we have $\max\varphi_{iRFP} = \Phi_1^{(0)} - \Phi_2 - (1+\gamma)/\eta \cdot S_{105} = (1+\gamma)/\eta \cdot (S_0 - S_{105}) \approx 1.1\%$ (where S_0 and 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.

Data availability

The raw flow cytometry data studied in this work and the software used to analyze it are all available at the following GitHub repository: https://github.com/LeonardoPaccia niMori/CPR-model-experiment-data-analysis.

Compliance with ethical standards

References

1. Bar-On YM, Phillips R, Milo R. The biomass distribution on earth. Proc Natl Acad Sci USA. 2018;115:6506–11.

2. Colman DR, Poudel S, Stamps BW, Boyd ES, Spear JR. The deep, hot biosphere: Twenty-five years of retrospection. Proc Natl Acad Sci USA. 2017;114:6895–903.

3. Puente-Sánchez F, Arce-Rodríguez A, Oggerin M, García-Villadangos M, Moreno-Paz M, Blanco Y, et al. Viable cyanobacteria in the deep continental subsurface. Proc Natl Acad Sci USA. 2018;115::10702–7.

4. Gold T. The deep, hot biosphere. Proc Natl Acad Sci USA. 1992;89:6045–9.

5. Sekirov I, Russell SL, Antunes LCM, Finlay BB. Gut microbiota in health and disease. Physiol Rev. 2010;90:859–904.

6. Singh BK, Bardgett RD, Smith P, Reay DS. Microorganisms and climate change: terrestrial feedbacks and mitigation options. Nat Rev Microbiol. 2010;8:779–90.

7. Cavicchioli R, Ripple WJ, Timmis KN, Azam F, Bakken LR, Baylis M, et al. Scientists' warning to humanity: microorganisms and climate change. Nat Rev Microbiol. 2019;17:569–86.

8. Stewart EJ. Growing unculturable bacteria. J Bacteriol. 2012;194: 4151–60.

9. Gonze D, Coyte KZ, Lahti L, Faust K. Microbial communities as dynamical systems. Current Opin Microbiol. 2018;44:41–9.

10. Tikhonov M, Monasson R. Collective phase in resource competition in a highly diverse ecosystem. Phys Rev Lett. 2017;118:1–5.

11. Butler S, O'Dwyer JP. Stability criteria for complex microbial communities. Nat Commun. 2018;9:2970.

12. Landmann S, Engel A. Systems of random linear equations and the phase transition in MacArthur's resource-competition model. EPL 2018;124;18004.

13. Niehaus L, Boland I, Liu M, Chen K, Fu D, Henckel C, et al. Microbial coexistence through chemical-mediated interactions. Nat Commun. 2019;10:2052.

14. Marsland R III, Cui W, Goldford J, Sanchez A, Korolev K, Mehta P. Available energy fluxes drive a transition in the diversity, stability, and functional structure of microbial communities. PLOS Comput Biol. 2019;15:1–18.

15. Rivett DW, Bell T. Abundance determines the functional role of bacterial phylotypes in complex communities. Nat Microbiol. 2018;3:767–72.

16. Enke TN, Datta MS, Schwartzman J, Cermak N, Schmitz D, Barrere J, et al. Modular assembly of polysaccharide-degrading marine microbial communities. Curr Biol. 2019;29:1–8.

17. Zelezniak A, Andrejev S, Ponomarova O, Mende DR, Bork P, Patil KR. Metabolic dependencies drive species co-occurrence in diverse microbial communities. Proc Natl Acad Sci USA. 2015; 112:201522642.

18. Louca S, Jacques SMS, Pires APF, Leal JS, Srivastava DS, Parfrey LW, et al. High taxonomic variability despite stable functional structure across microbial communities. Nat Ecol Evol. 2016;1:0015.

19. Basan M. Resource allocation and metabolism: the search for governing principles. Curr Opin Microbiol. 2018;45:77–83.

20. Bajic D, Sanchez A. The ecology and evolution of microbial metabolic strategies. Curr Opin Biotechnol. 2020;62:123–8.

21. Budinich M, Bourdon J, Larhlimi A, Eveillard D. A multi-objective constraint-based approach for modeling genome-scale microbial ecosystems. PLOS One. 2017;12:1–22.

22. Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T. Interdependence of cell growth and gene expression. Science. 2010;330:1099–1102.

23. Schaechter M, Maaløe O, Kjeldgaard NO. Dependency on medium and temperature of cell size and chemical composition during balanced growth of salmonella typhimurium. Microbiology. 1958; 19:592–606.

24. Scott M, Hwa T. Bacterial growth laws and their applications. Curr Opin Biotechnol. 2011;22:559–65.

25. Basan M, Hui S, Okano H, Zhang Z, Shen Y, Williamson JR, et al. Overflow metabolism in Escherichia coli results from efficient proteome allocation. Nature. 2015;528:99–104.

26. Mori M, Hwa T, Martin OC, De Martino A, Marinari E. Constrained allocation flux balance analysis. PLOS Comput Biol. 2016;12:1–24.

27. MacArthur R. Species packing, and what competition minimizes. Proc Natl Acad Sci USA. 1969;64:1369–71.

28. MacArthur R. Species packing and competitive equilibrium for many species. Theor Popul Biol. 1970;1:1–11.

29. Chesson P. MacArthur's consumer-resource model. Theor Popul Biol. 1990;37:26–38.

30. Tikhonov M. Community-level cohesion without cooperation. eLife. 2016;5:e15747.

31. Posfai A, Taillefumier T, Wingreen NS. Metabolic trade-offs promote diversity in a model ecosystem. Phys Rev Lett. 2017; 118:28103.

32. Advani M, Bunin G, Mehta P. Statistical physics of community ecology: a cavity solution to MacArthur's consumer resource model. J Stat Mech. 2018;2018:033406.

33. Droop MR. The nutrient status of algal cells in continuous culture. J Marine Biol Assoc UK. 1974;54:825–55.

34. Khandelwal RA, Olivier BG, Röling WFM, Teusink B, Bruggeman FJ. Community flux balance analysis for microbial consortia at balanced growth. PLOS One. 2013;8:1–10.

35. Embree M, Liu JK, Al-Bassam MM, Zengler K. Networks of energetic and metabolic interactions define dynamics in microbial communities. Proc Natl Acad Sci USA. 2015;112:15450–5.

36. Liao C, Wang T, Maslov S, Xavier JB. Modeling microbial cross-feeding at intermediate scale portrays community dynamics and species coexistence. PLOS Comput Biol. 2020;16:1–23.

37. Muscarella ME, O'Dwyer JP. Species dynamics and interactions via metabolically informed consumer-resource models. Theor Ecol 2020;13:503–18.

38. Hermsen R, Okano H, You C, Werner N, Hwa T. A growth-rate composition formula for the growth of E. coli on co-utilized carbon substrates. Mol Syst Biol. 2015;11:801.

39. Erickson DW, Schink SJ, Patsalo V, Williamson JR, Gerland U, Hwa T. A global resource allocation strategy governs growth transition kinetics of Escherichia coli. Nature. 2017;551:119–23.

40. Pacciani-Mori L, Giometto A, Suweis S, Maritan A. Dynamic metabolic adaptation can promote species coexistence in competitive communities. PLOS Comput Biol. 2020;16:1–18.

41. Taillefumier T, Posfai A, Meir Y, Wingreen NS. Microbial consortia at steady supply. eLife. 2017;6:e22644.

42. Ratzke C, Gore J. Modifying and reacting to the environmental pH can drive bacterial interactions. PLoS Biol. 2018;16: e2004248.

43. Monod J. The growth of bacterial cultures. Ann Rev Microbiol. 1949;3:371–94.

44. Stülke J, Hillen W. Carbon catabolite repression in bacteria. Curr Opin Microbiol. 1999;2:195–201.

45. Görke B, Stülke J. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. Nat Rev Microbiol. 2008;6:613–24.

46. Stoebel DM, Dean AM, Dykhuizen DE. The cost of expression of escherichia coli lac operon proteins is in the process, not in the products. Genetics. 2008;178:1653–60.

47. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. Science. 2002;297:1183–6.

48. Rosset R, Monier R, Julien J. RNA composition of escherichia coli as a function of growth rate. Biochem Biophys Res Commun. 1964;15:329–33.

49. Rosset R, Julien J, Monier R. Ribonucleic acid composition of bacteria as a function of growth rate. J Mol Biol. 1966;18: 308–20.

50. Van Derlinden E, Van, Impe JF. Modeling growth rates as a function of temperature: Model performance evaluation with focus on the suboptimal temperature range. Int J Food Microbiol. 2012; 158:73–8.