

Master thesis title

Pablo Lechón Alonso

Imperial College London

Word count: 10

Abstract

Community assembly, the process whereby species come together and interact to form functioning and coherent aggregations, is an old-age unsolved problem in ecology. In the microbial world, it is common that whole communities come into contact with each other and reassemble into a new community. This process has been termed community coalescence. The mechanisms driving community coalescence are poorly understood, partly because theoretical work rarely considers communities with mutualistic interactions, which are pervasive in microbial consortia. In this work, I propose a measure of community cohesion that, taking into account both competitive and mutualistic interactions, predicts the outcome of microbial community coalescence. The proposed metric explicitly quantifies the so-called *cohesivness* exhibited by microbial communities, and captures previous results of coalescence experiments while pinning them down to more realistic assumptions about interactions in the community. My results demonstrate how considerations about the type and strength of the interactions in the community can shed light upon macroscopic properties and collective behaviour exhibited by coalescing communities. I suggest that accurately mapping the metabolic pathways of bacteria offers a way to experimentally measure their interactions in microbial consortia and outlines a path to understand and even engineer microbial communities with beneficial functioning.

Contents

1	Introduction notes	2
2	Introduction	3
3	Methods and Results	4
3.1	Consumer resource model with cross feeding interactions	4
3.2	Community Assembly	6
3.3	A metric of community cohesion	9
3.4	Community-level cohesion predicts the outcome of community co- alescence experiments.	12
4	Discussion	14
5	Discussion notes	15
6	Things to do in the future	16
7	Appendix	19
7.1	Reversible enzyme kinetics	19
7.2	Table of parameter values and meaning	20
7.3	Relationship between Θ , F , C , and Ψ	20
7.4	Calculation of individual fitness	20

1 Introduction notes

- State my aims/hypotheses/questions by the end of the introduction.

- Why is it interesting? Why don't we know the answer?

Following, what I want to talk about

- Will a coalesced community be more perssistent than a naive one upon a event if it has a history of coalescence?

- How cohesive a community is when we analyze it in terms of cohord and dominant species?

- say at some point how many possible networks there are.

- measure community productivity and check that the correlation breaks down because there is co-selection.

- microbial ensamblages

13 2 Introduction

14 Microbial communities are widespread throughout our planet, from the deep
15 ocean to the human gut, and they play a critical role in natural processes
16 ranging from animal development and host health (Huttenhower et al. 2012)
17 to biogeochemical cycles (Falkowski et al. 2008). These communities are very
18 complex, often harbouring hundreds of species (Gilbert et al. 2014), making
19 them hard to characterize. Recently, DNA sequencing has facilitated a high-
20 resolution mapping of these consortia, opening a niche for ambitious theorists
21 and experimentalists to collaboratively disentangle the complexity of these sys-
22 tems (Marsland et al. 2019, Goldford et al. 2018, Goyal & Maslov 2018, Fried-
23 man et al. 2017, Costello et al. 2012). One of the problems yet to be solved is
24 community assembly – the process by which species come together and inter-
25 act to establish a community. Contrary to what is found in the macroscopic
26 world, in microbial ecology, it is common that whole communities move to a
27 region where they encounter another community. The process by which two or
28 more communities that were previously separated join and reassemble into a new
29 community has been termed "community coalescence" (Rillig et al. 2015). This
30 type of event repeatedly happens in nature due to abiotic (wind, tides or river
31 flow), biotic (animal courtship, parent-offspring interactions or leaves falling),
32 and anthropogenic (industrial anaerobic digestion, agriculture, between-human
33 contact) factors (Castledine et al. 2020). Despite the frequency and importance
34 of microbial community coalescence, the mechanisms responsible for the com-
35 munity structure and function resulting from coalescence events remain poorly
36 understood.

37 Early mathematical models of community-community invasion revealed that
38 when two communities previously separated by a barrier merge due to its re-
39 moval, asymmetrical dominance of one community over the other one is likely to
40 occur (Gilpin 1994, Toquenaga 1997). As an explanation for this observation, it
41 was argued that, because communities have been assembled through a history of
42 competitive exclusion, they are likely to compete with each other as coordinated
43 entities, rather than as a random collection of species. New theoretical work uses
44 consumer-resource models to show that coalescing microbial communities exhibit
45 an emergent cohesiveness (Tikhonov 2016, Tikhonov & Monasson 2017). These
46 communities have been termed Metabolically Cohesive [microbial] Consortium
47 (MeCoCos) by Pascual-García et al. (2020) and suggested to be pervasive in
48 microbial ecology.

49 Recent results from coalescence experiments of methanogenic communities sug-
50 gest that during a coalescence event between two communities, multiple taxa
51 from the same community act as cohesive units and are selected together (eco-
52 logical co-selection) (Sierocinski et al. 2017). Further experimental evidence
53 of co-selection in community coalescence has been reported in Lu et al. (2018),
54 where it was shown that successful collective invasions are accompanied by strong
55 community-level interactions. The microbial communities used in these exper-
56 iments are characterized by complex cross-feeding interactions (Hansen et al.
57 2007, Lawrence et al. 2012, Embree et al. 2015). Furthermore, the type of trophic
58 interactions present in a community has been suggested as a factor that might
59 affect the outcome of community coalescence (Castledine et al. 2020). Yet, the-
60 oretical models used in community coalescence studies so far have considered

61 competition between species as the only force driving community assembly.
 62 In this work, I explore the role of other types of interactions, which appear
 63 to be ubiquitous in microbial communities. First, I use a consumer-resource
 64 model that includes both facilitation of metabolites via by-product secretion,
 65 and competition for substrates, to simulate many instances of community as-
 66 sembly. Second, I propose a metric of community cohesion that accounts for
 67 both competitive and mutualistic interactions and I measure the cohesion level
 68 in the simulated communities. Third, I apply the proposed metric to predict the
 69 outcome of microbial community coalescence events.

70 3 Methods and Results

71 3.1 Consumer resource model with cross feeding interactions

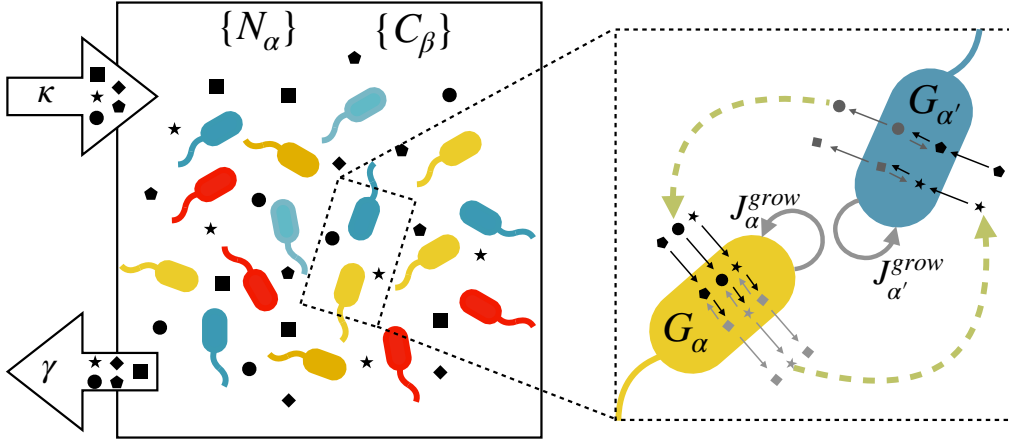


Figure 1: **Schematic of the model.** (left) Consider a chemostat where m metabolites are steadily supplied at rate κ and diluted at rate γ . Different bacterial strains coexist in the chemostat, and they consume the metabolites in the environment, $\{C_\beta\}$ through their reaction networks G_α (right), to obtain the necessary power J_α^{grow} to increase their abundance $\{N_\alpha\}$. The green arrows in the magnified portion emphasize that species α (yellow) facilitates metabolites to species α' (blue) and viceversa. The double arrows shown within the cells denote the fact that the reactions I consider are fully reversible.

72 In order to simulate communities with cross-feeding interactions, I use a
 73 consumer-resource model inspired in the the work of (Marsland et al. 2019).
 74 Consider an environment where a single limiting element \mathcal{R} is present in m forms
 75 with different concentrations C_β , where $\beta \in \{1 \dots m\}$. For example, this could
 76 be a carbon-limited environment where m sugars are supplied. Let now N_α
 77 denote the abundance of each bacterial strain α , present in the environment,
 78 where $\alpha \in \{1 \dots s\}$. Each species is uniquely characterized by the metabolic
 79 strategy it uses to harvest resources. This strategy is encoded in its reaction
 80 network G_α , a collection of chemical reactions between the metabolites in the
 81 environment that produce energy that is used by bacteria for survival and repli-
 82 cation (see Figure 1). A maintenance cost required for survival is imposed to

each species. This cost is determined by their reaction network: a higher maintenance cost is incurred by both being able to produce more types of enzymes, and those enzymes involving high energy yielding reactions.

If we now allow the dynamics of this system to unfold, the concentration of each metabolite C_β determines the dynamics of the abundances N_α of each species, which harvest resources through their different metabolic strategies. During this process the products of the chemical reactions of some species may act as substrates for other ones. This introduces cross-feeding interactions in the system (see Figure 1). The changes in species abundance therefore translate into changes in the total supply and demand of resources. In turn, resource concentrations C_β are depleted until equilibrium is reached. A more rigorous description of the model, along with its mathematical form will now be presented.

Consider the population dynamics of s consumers (eg. bacterial strains) that feed on m resources. In this model, a species is defined by the metabolic strategy it uses to harvest energy from the environment. Let $G_\alpha(\mathcal{M}, \mathcal{N})$ be the metabolic network of species α , where \mathcal{M} is a set of nodes $\mathcal{M} = \{x : x \text{ is an integer from the interval } [1, m] \text{ labeling the metabolite}\}$ and \mathcal{N} a set of uni-directed edges $\mathcal{N} = \{(x, y) : x \in \mathcal{M}, y \in \mathcal{M} \text{ and } x < y \text{ (} x \text{ and } y \text{ are the product and the substrate of a chemical reaction, respectively)}\}$. The growth power of species α , J_α^{grow} will be given by the product of the amount of generated energy per reaction event η_i and rate q_i of each reaction, summed across all reactions in \mathcal{N} .

$$J_\alpha^{grow} = \sum_{i=1}^{|\mathcal{N}|} q_i \eta_i \quad (1)$$

where $|\cdot|$ denotes cardinality of a set. Refer to subsection 7.1 for specifications on q and η .

Every species has a maintenance cost χ_α that represents the required energy to sustain life, which is assumed to take the form

$$\chi_\alpha = \chi_0 \sum_{\mathcal{N}} (y - x) \quad (2)$$

where χ_0 is the average cost per reaction, x and y are the substrate and the product of the reaction, respectively, and the summatory term adds up the metabolite gap of all reactions. Therefore, the maintenance cost of one species increases if one or both of the following quantities increases: (1) the amount of enzymes a species is capable of decomposing, and (2) the energy yielded by the reactions in which these enzymes are involved. The cost function (Eq. 2) ensures that neither generalists, nor specialists, are systematically favored during the community assembly.

Under this parametrization, the time evolution of the population of species α can be written as

$$\frac{dN_\alpha}{dt} = g_\alpha N_\alpha [J_\alpha^{grow} - \chi_\alpha] \quad (3)$$

where g_α is a proportionality constant relating energy to abundance of strain α . The dynamics of the resources depend on the incoming and outgoing resource fluxes due to the biochemical reactions taking place inside bacteria, as well as the resource external dynamics. The incoming resource flux of metabolite β

generated by strain α is its rate of consumption due to all the biochemical reactions possessed by α in which β is a substrate. The outgoing flux is that due to reactions in which β is a product.

$$\begin{aligned} v_{\alpha\beta}^{in} &= \sum_{\mathcal{S}} q & \text{with } \mathcal{S} &\equiv N \cap \{(x = \beta, y)\} \\ v_{\alpha\beta}^{out} &= \sum_{\mathcal{P}} q, & \text{with } \mathcal{P} &\equiv N \cap \{(x, y = \beta)\} \end{aligned} \quad (4)$$

The external resource dynamics are modelled as a supply rate minus a dilution rate that depends on the resource concentration to ensure convergent dynamics.

$$h_{\beta} = \kappa - \gamma C_{\beta} \quad (5)$$

Therefore, the variation with time of the concentration of metabolite β has the form

$$\frac{dC_{\beta}}{dt} = h_{\beta} + \sum_{\alpha=1}^s (v_{\alpha\beta}^{in} - v_{\alpha\beta}^{out}) N_{\alpha} \quad (6)$$

Thus, the model is a system of $s + m$ coupled differential equations completely specified by Eqs. 3 & 6.

In the following section, I characterize the communities that assemble after integrating a collection of these systems, where each system has 10 bacterial strains ($s = 10$) in an environment with 15 different resources ($m = 15$). For each system, species are assigned metabolic strategies drawn at random.

3.2 Community Assembly

Armed with this model I now simulate $n_s = 2 \cdot 10^3$ community assembly events of $s = 10$ species that interact in an environment with $m = 15$ substrates that are steadily supplied.

The values of the parameters of the model (subsection 7.2, table 1) remain constant throughout all simulations. The reason for this is that, my aim is not to parametrize the model to reveal large-scale patterns found in experiments (although that would be a fruitful endeavour because of the rich parameter space of this model). Rather, I use it to simulate a set of microbial communities with cross-feeding interactions that will be later used in the community coalescence experiments.

In order to do so, I first create $s \cdot n_s$ random reaction networks, $G_{\alpha}(\mathcal{M}, \mathcal{N})$ (one for each strain) using the following procedure. Consider, the $m \times m$ adjacency matrix A_{ij}^{α} , whose elements; the edges (i, j) of G_{α} , represent chemical reactions. Since the reaction network is hierarchical ($i < j$, subsection 7.1), the adjacency matrix is an upper triangular matrix with zeros in the main diagonal, and the reactions possessed by strain α can be expressed as $(i, i+k)$, where k represents the k^{th} diagonal of A ($k \in \{1, \dots, m-1\}$ with $k = 0$ being the main diagonal), and i is the row number of one of its elements ($i = 1 \dots m$). The reaction network G_{α} is constructed by sampling n_r reactions from different diagonals, with decreasing probability as the order of the diagonal increases. Thus, I choose n_r pairs of integers (i, k) according to the algorithm summarized below.

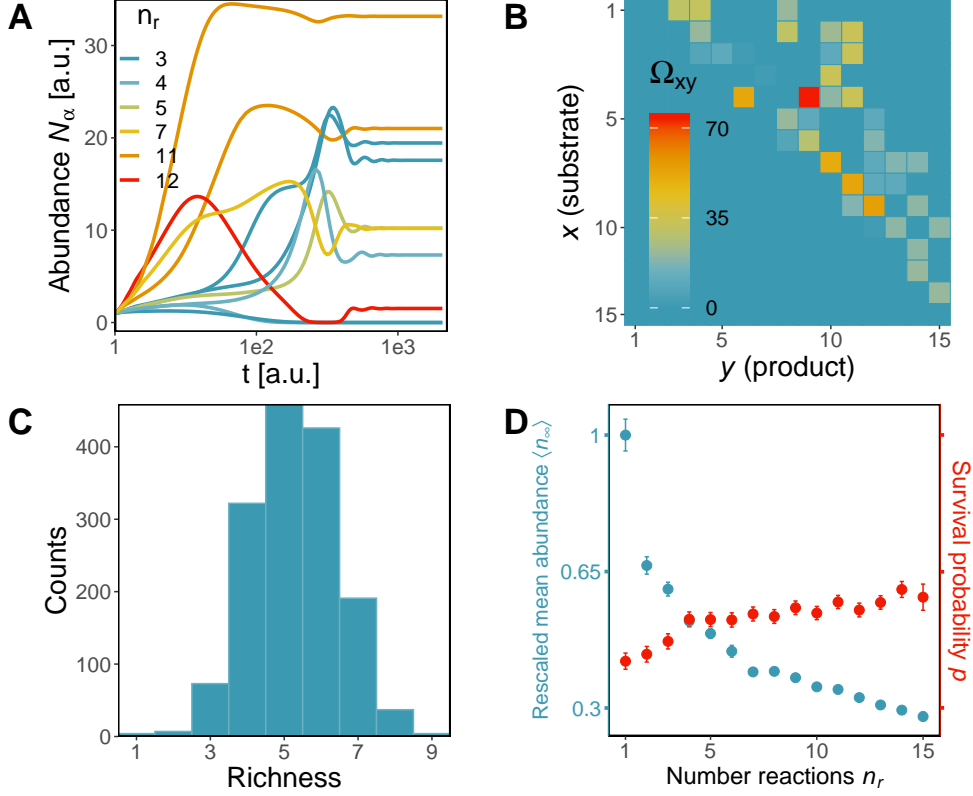


Figure 2: **Results from community assembly simulations.** Plots (A) and (B) exemplify one community assembly event and (C) and (D) convey results across simulations. (A) Time variation of species' abundance for one instance of community assembly with $m = 15$, $s = 10$, and a set of s randomly generated reaction networks. Time (x-axis) and population (y-axis) are measured in arbitrary units. Each time series is coloured according to n_r , the number of reactions possessed by the reaction network of each strain. (B) Community reaction network, obtained by summing the reaction network adjacency matrices of all species weighted by their respective carrying capacity: $\Omega = \sum_{k=1}^s N_\infty^k A_k$. The community reaction network is unique for each community, and it constitutes a blueprint of how that community depletes resources in the environment. Note that, according to the imposed constraints, reactions where $x > y$ are absent, and those where $y \gg x$, are rare. (C) Histogram of richness of the n_s simulations. (D). In blue, mean value of carrying capacity rescaled to 1 against the number of reactions n_r . In red, survival probability against number of reactions n_r . Species with less reactions (specialists) tend to be present at higher abundances than those with higher n_r (generalists), but they have a lower surviving probability.

- 167 1. Choose n_r by sampling it from a uniform distribution $U(1, m)$
- 168 2. Choose k by sampling one value from a truncated normal distribution
169 $N(1, \sqrt{m-1})$ with limits $[1, m-1]$, and rounding it to the closest integer.
- 170 3. Sample i from a uniform distribution of integers $U(0, m-k)$.
- 171 4. The reaction $(i, i+k)$ is stored, and the process is repeated until n_r reac-
172 tions have been sampled.

173 Several things are important to note about this algorithm. Firstly, sampling k
174 from a truncated normal distribution ensures that high metabolite gaps (very
175 energetic reactions) are not likely to happen. This introduces a bias against the
176 presence of organisms with few and very energetic reactions, which are rare in
177 microbial communities. Second, the truncation limits in step 2 have been chosen
178 to respect the imposed constraint that reactions can only be of the form $i < j$.
179 Third, the upper limit of the uniform distribution from which i is sampled is
180 bounded by k , the diagonal we are sampling from.

181 When the sampling of reaction networks is completed, Eqs. 6 and 3 are in-
182 tegrated using a Runge Kutta method (Dormand & Prince 1980) with initial
183 conditions $N_\alpha(t=0) = 2$ and $C_\beta(t=0) = 0$.

184
185 Relevant results stemming from the simulations of community assembly events
186 are plotted in figure 2. The first two figures convey information about the dy-
187 namics and the resource consumption map of one particular community. In
188 figure (2A), all abundances start increasing because all resources are present
189 and steadily supplied. As the dynamics evolve, the community engineers its own
190 environment by consumption of metabolites and secretion of by-products, caus-
191 ing the creation of ecological niches where some species reach equilibrium and
192 others go extinct. Figure (2B), shows that all metabolites are being consumed
193 (all rows have at least one non-zero element), which is to say, that all vacant
194 niches are being occupied.

195 On average, the community assembly simulations generated communities with
196 richness spanning from 1 to 9 species (figure 2C), where specialists tend to be
197 more abundant than generalists (figure 2D, blue points). This can be attributed
198 to several specialists being able to deplete all resources through their combined
199 action more efficiently than one generalist (Pascual-García et al. 2020), thus,
200 dominating the community at equilibrium. Note that in the simulated commu-
201 nities, while specialists tend to be present at higher abundances than generalists,
202 their survival probability is lower (figure 2D, red points). Upon competition,
203 generalists have more alternatives to obtain energy, while specialists do not, and
204 thus are more vulnerable to extinctions.

205
206 Following the assemblage of many synthetic communities, I propose a metric
207 of community cohesion that takes into account all the interactions, mutualistic
208 and competitive, between the species in the community. The cohesion of the
209 communities is then tracked during and after the process of community assembly.
210 This measure is later used to predict the outcome of community coalescence
211 events.

212 3.3 A metric of community cohesion

213 The cohesion of a community is ultimately determined by the nature of the inter-
 214 actions between its members. Since I am considering two types of interactions,
 215 namely, competition, and mutualism, the simplest way to render them into a
 216 mathematical expression is to subtract them.

$$217 \quad \text{Cohesion} = \text{Facilitation} - \text{Competition} \quad (7)$$

218 Measuring levels of facilitation and competition within a microbial community
 219 is experiementally challenging. However, the metabolic strategies of each species
 220 are well determined in this theoretical framework. Therefore, I use the reaction
 221 network of each bacterial strain to compute their competition and facilitation
 222 indices with the rest of the species in the community.

223 Let s_1 and s_2 be two sequences of integers labeling metabolites. I am interested
 224 in measuring their *overlapping degree* $\xi(s_1, s_2)$, ie, the proportion of metabolites
 225 of s_1 that intersect with s_2 summed with the proportion of metabolites of s_2
 226 that intersect with s_1 , normalized to 1.

$$227 \quad \xi(s_1, s_2) = \frac{1}{2} \sum_{k \in s_1 \cap s_2} \left(\frac{D_{s_1}(k)}{|s_1|} + \frac{D_{s_2}(k)}{|s_2|} \right) \quad (8)$$

228 Here, k takes the values in the set that result form intersecting s_1 and s_2 . $D_s(k)$
 229 is the number of elements from the sequence s that are equal to k . Vertical bars
 230 $| \cdot |$ express cardinality of a squence. The purpose of all denominators in equation
 231 8 is to normalize ξ to 1.

232 One way to capture the facilitation of a community is by calculating its facil-
 233 itation matrix F , which is composed of the facilitation indices of all possible
 234 ordered pairs i, j of species in the community. Precisely, the facilitation index
 235 f_{ij} of species i towards species j , is given by the overlapping degree of the se-
 236 quence of products y_i of species i , with the sequence of substrates x_j of species
 237 j . Equivalently, the competition matrix C gathers the competition level of the
 238 community. The competition index between species i and j , c_{ij} is given by the
 239 overlapping degree of the sequence of substrates x_i of species i , and the sequence
 240 of substrates x_j of species j . Thus,

$$241 \quad F_{ij} = \begin{cases} \xi(y_i, x_j) & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \quad C_{ij} = \begin{cases} \xi(x_i, x_j) & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \quad (9)$$

242 Note that facilitation is directional but competition is not. This implies that
 243 $F_{ij} \neq F_{ji}$ and F is not symmetric, but $C_{ij} = C_{ji}$ and C is symmetric.
 244 Following the idea sketched in equation 7, a cohesion matrix Ψ can be defined
 245 using equations in 9, as an upper triangular $s \times s$ matrix whose elements are
 246 given by

$$247 \quad \Psi_{ij} = \begin{cases} \frac{1}{2} (F_{ij} + F_{ji}) - C_{ij} & \text{if } i < j \\ 0 & \text{if } i \geq j \end{cases} \quad (10)$$

248 I choose to define Ψ as an upper triangular matrix because cohesion is not a
 249 directional measure; two species are not more cohesive if measured from i to j ,
 250 than from j to i . Instead, cohesion is a pairwise estimate independent of the

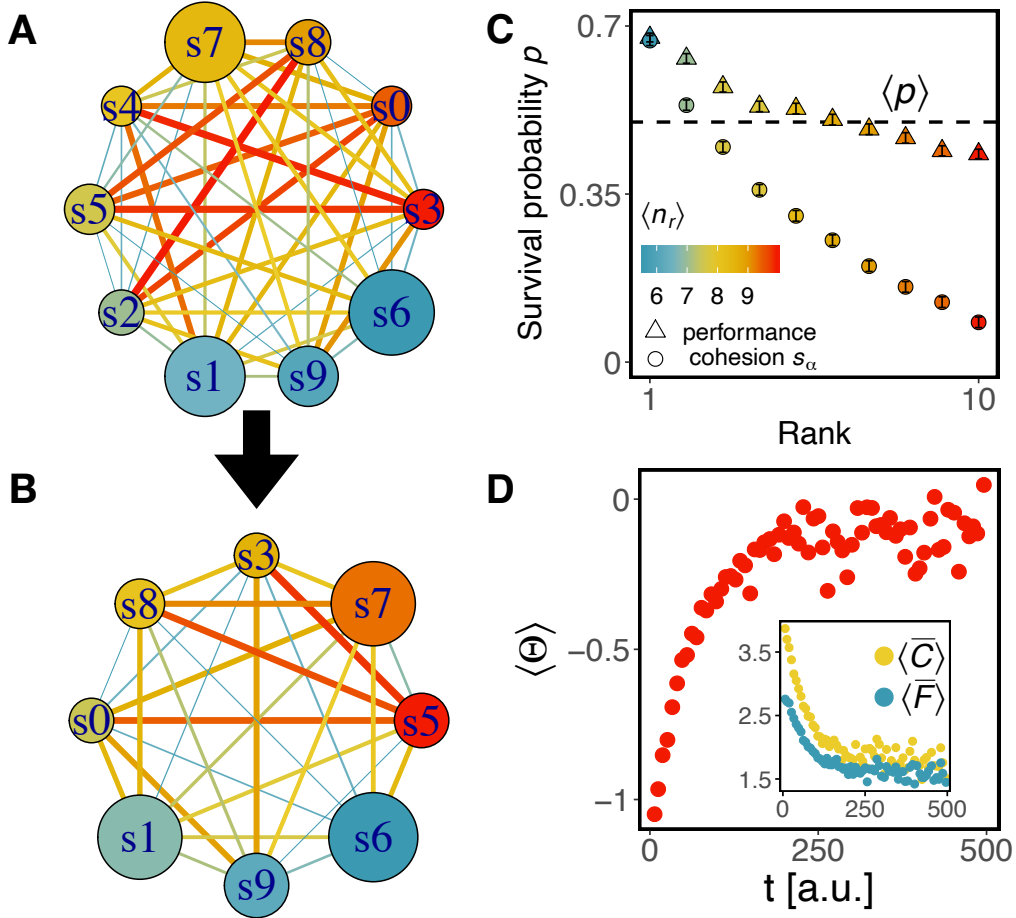


Figure 3: Cohesion metric across community assembly events. Cohesion network before (A) and after (B) community assembly. Thicker and red edges represent higher cohesion coefficient Ψ_{ij} between its nodes, species i and j . The color of the nodes changes from red to blue in anti-clockwise direction indicating decreasing total cohesion level s_α . The size of the node represents the number of reactions possessed by that species. (C) Median survival probability after community assembly as a function of cohesion rank (circles) and individual performance rank (triangles) of species in the random community (before assembly) across all simulations. Circles are weighted by the cohesion rank measured in the assembled community, e.g., weight is maximum when the first ranked species in the random community remains first ranked after assembly. The dashed line is the average number of extinctions across simulations. Species with higher cohesion and individual performance are less likely to go extinct during community assembly. However, cohesion predicts survival probability better than individual performance. The color of the points reflects the mean number of reactions of species in each rank. There is a slight correlation between the number of reactions and the survival probability. (D) Community-level cohesion averaged across all community assembly events, $\langle \Theta \rangle$, as a function of time. Every time a species goes extinct during community assembly, the community cohesion is recalculated with the remaining species. On average, community assembly follows trajectories where community cohesion increases. (D, inset) Community competition and facilitation levels averaged across all simulations. Both decrease during community assembly, but competition decreases faster. Decrease of facilitation is explained by the positive correlation between $\langle C \rangle$ and $\langle F \rangle$ (see Figure 4, A).

direction of measure, so its matrix representation should be either symmetric or triangular. The triangular definition of Ψ allows me to interpret it as the adjacency matrix of a directed weighted network of cohesion between species. An example of this network, corresponding to the instance of community assembly shown in figures **2A** and **2B**, is illustrated before and after assembly in figures **3A** and **3B**, respectively. Nodes represent species and edges are weighted by their corresponding element in the cohesion matrix (represented in the figures by line thickness and color). Each species has a different node size and color. The size of the species encodes the number of reactions in its reaction network. The color encodes the node strength s_α , which is the sum of the edge weights connected to node α . This represents the total cohesion level of species α with the rest of species in the community, and it is calculated as

$$s_\alpha = \sum_{j \neq \alpha} \Psi_{\alpha j} \quad (11)$$

The cohesion network is a powerful tool to fully track the changes in interactions between members in the community during its assembly. The two networks illustrated in figures **3A** and **3B** show that the top 4 species (s3, s0, s8 and s7) with higher total cohesion levels s_α remain extant after the community assembles. Additionally, the bottom 3 species (s1, s9 and s6) with lower s_α have more reactions in their reaction networks than the average. To test the generality of these observations I calculate the median survival probability after community assembly for each s_α rank position in the random community, across all instances of community assembly. Figure **3C** shows a clear correlation between these two measures, indicating that species with lower s_α go extinct more easily than more cohesive species. The survival probability is plotted also as a function of individual performance rank in figure **3C** (see subsection 7.4 for details on the calculation of the proxy of individual performance). The individual performance of a species is calculated here by measuring its equilibrium abundance in isolation. Interestingly, individual performance does not predict survival probability as well as total cohesion level, suggesting that in these experiments, the individual fitness of a species becomes decoupled from its probability of success. This observation reflects the well-known fact that the success of a species is context-dependent, and observing a species in isolation does not measure its performance in the relevant environment (McGill et al. 2006, McIntire & Fajardo 2014). The stabilization of the system is normally accompanied by a cascade of extinctions. The effect of these extinctions in the cohesion at a community level can be conveyed by averaging the cohesion network

$$\Theta = \frac{1}{T_s} \sum_{i < j} \Psi_{ij} \quad (12)$$

Where $T_s = \binom{s+1}{2}$ is the number of elements being summed: those in the upper diagonal of Ψ (the triangular number of order s). To investigate how the community-level cohesion changes over the course of community assembly, I measure Θ after the occurrence of every extinction across all simulations. This variation can be seen in figure **3D**, where the binned averaged value of all measures of Θ is represented during the first 500 units of time for all simulations. A systematic increase of community-level cohesion during the first

half of measured time, and following stabilization near $\Theta = 0$ is observed. This increase is due to a faster decrease in community competition and facilitation levels, $\langle \bar{C} \rangle$, and $\langle \bar{F} \rangle$, averaged across simulations. (**3D, inset**).

This section presented a measure of cohesion, Ψ , that proved to be a useful tool to capture relevant insights of the dynamics of community assembly when mutualistic interactions are present. In the next section, I perform community coalescence experiments to test the predictive power of my metric, finding that communities with higher community cohesion are more successful upon coalescence.

3.4 Community-level cohesion predicts the outcome of community coalescence experiments.

Consider a coalescence event, whereby two communities previously separated come into sudden contact. In general, monodominance of one community after the mix reaches stable state is not guaranteed. Instead, both communities will contribute species to the final equilibrium. Can we predict which community will do so more successfully?

To answer this question I firstly, use all the simulated communities to populate a facilitation-competition (F-C) diagram where the axes are $\langle F \rangle$ and $\langle C \rangle$; community-level facilitation and competition respectively. They are calculated by averaging the non-diagonal elements of facilitation and competition matrices. Communities are scattered across the plot, bringing out two regimes: a mutualistic regime where $\langle F \rangle > \langle C \rangle$, and a competitive regime where $\langle F \rangle < \langle C \rangle$. The former case has a community-level cohesion Θ satisfying, $\Theta > 0$ whereas the latter has $\Theta < 0$ for the former (see subsection 7.3 for a rigorous description of the relationships between these variables). I then select communities with 4 species from the extremes of the two regimes, $\Theta \gg 0$ and $\Theta \ll 0$. This puts at my disposal two groups of communities with higher and lower levels of cohesion (blue and red strips in figure **4A**). I now perform coalescence experiments where a resident community \mathcal{C}_R from one group is mixed with an invading one \mathcal{C}_I from the other group. One would expect that communities from the mutualistic regime are more successful on average than those from the competitive regime. To confirm this, results from an 'elimination assay' competing pairs of communities from each group is presented in figure **4C**. Near five hundred pairs of communities are mixed, and correspond to the columns in figure **4C**. For each pair, species from both communities are equilibrated together. The rows in figure **4C** correspond to these species, and are ordered increasingly according to their total cohesion level s_α . For each species that goes extinct during the coalescent event, its provenance is identified (i.e. does it come from the blue or the red community?), and the corresponding tile in figure **4C** is coloured accordingly. The dominant colour is red, confirming that communities in the competitive regime experiment more extinctions, and thus, are worse at contributing with their members to the final equilibrium.

Finally, I select the N communities with 5 species (figure **2B**), and perform all $\binom{N}{2}$ possible community coalescence events in which a resident community \mathcal{C}_R is mixed with an invading one \mathcal{C}_I . At each event, I calculate the similarity of between post-coalescence and resident communities as the normalized scalar

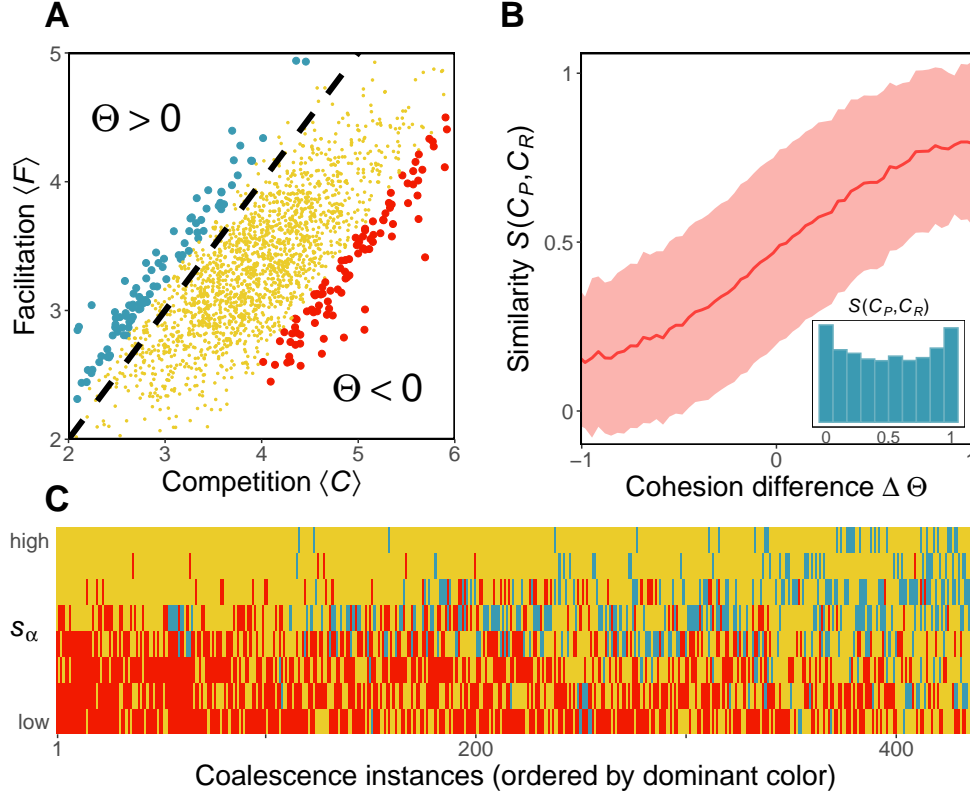


Figure 4: **Results from community coalescence experiments.** (A) Each simulated community is plotted in a competition-facilitation diagram. Communities above the dashed line $\langle C \rangle = \langle F \rangle$ have $\Theta > 0$, and thus they are in the facilitation-dominated regime. Communities below the dashed line have $\Theta < 0$, and therefore they belong to the competition-dominated regime. The extremes of each regime are selected (blue and red dots), and coalescence experiments where one community from the blue group mixes with one community from the red group are performed, only for communities of richness 4. (C) Altruistic communities ($\Theta > 0$) outperform competitive communities ($\Theta < 0$) in the latter experiments. In this elimination assay, each column represents one coalescence instance, and each element in a column is a species. Extinctions are coloured to match the group in plot A to which the extinct species belonged. There is a higher proportion of extinct species from the red group (more red tiles than blue tiles). (B) The outcome of community coalescence is predicted by community-level cohesion. The similarity between the post-coalescence community and the resident community, $S(C_P, C_R)$ is plotted as a function of the community cohesion difference $\Delta \Theta$ between them, for all possible coalescence events between 2 communities of richness 5. Shown is binned mean (100 bins) over communities with similar $\Delta \Theta$ (solid line) $\pm \sigma$ (shaded) (B, inset) Histogram of similarity showing that monodominance of one community after coalescence ($S = 0$, $S = 1$) is more frequent than a perfect mixing ($S = 0.5$)

product of their species abundance vector at stable state.

$$S(\mathcal{C}_R, \mathcal{C}_P) = \frac{\vec{N}_\infty^R \cdot \vec{N}_\infty^P}{\sqrt{|\vec{N}_\infty^R|} \sqrt{|\vec{N}_\infty^P|}} \quad (13)$$

Additionally, I calculate the community-level cohesion difference $\Delta\Theta$ between the two coalescing communities. A clear non-linear correlation emerges when I plot similarity versus cohesion difference (figure **4B**). The larger the difference between community cohesion, the more similar is the post-coalescent community to its more cohesive parent. The non-linearity of this curve is a manifestation of the asymmetrical dominance reported in the first works of community coalescence (Gilpin 1994). This is more evident when looking at the histogram of similarities (**4B, inset**) for the coalescence experiments performed, where monodominance of one community ($S = 1, S = 0$) is more frequent than a perfect mixing ($S = 0.5$)

4 Discussion

A frequent way in which microbial communities come to be is through the mix of two or more communities, an event that has been termed community coalescence (Rillig et al. 2015). Numerous theoretical and experimental studies suggest that coalescing communities behave as 'coherent wholes', and compete against each other like coordinated armies (Gilpin 1994, Toquenaga 1997, Livingston et al. 2013, Tikhonov 2016, Tikhonov & Monasson 2017, Sierocinski et al. 2017, Lu et al. 2018). To date, competition in coalescence is the most studied interaction, but more work is needed to understand how other types of interactions lead to different coalescence outcomes (Castledine et al. 2020). In this work, I investigated the behaviour of pairs of coalescent communities that harboured organism interdependence through metabolic complementarity (i.e., cross-feeding). How does including mutualistic interactions affect the outcome of community coalescence?

To answer this question, I proposed to quantify the coherence of these consortia through a metric of community cohesion, Ψ , which computes the level of positive feedback between every pair of species in the community. I found that Ψ increased on average during community formation (see Figure **2D**). Tracking cohesion in coalescence events revealed a non-linear relationship between cohesiveness and post-coalescence success, that is, more cohesive communities were coherently favoured during species sorting, and therefore dominated at equilibrium. This result constitutes strong evidence supporting that a community undergoing a coalescence event behaves as a 'coherent whole'. It is expected that members of cohesive communities have been ecologically co-selected; those individuals from a key taxon whose presence provides an advantage for individuals from other taxa are positively selected (Sierocinski et al. 2017). This contrasts with the alternative hypothesis suggesting that those communities harbouring species with higher individual performance are the ones that would dominate in the formation of communities. Only weak support was found for this hypothesis (see Figure **2C**). These two hypotheses are two extremes of a continuum. Although it certainly seems an exaggeration to view the community as a 'super-organism', it is also a perilous simplification to consider it as a mere collection

386 of individuals, ignoring the fact that coevolutionary processes can play an im-
387 portant role in it (Rillig & Mansour 2017).

388 A recent idea that smoothly interpolates between these to extremes has been
389 proposed in (Pascual-García et al. 2020). Metabolically Cohesive microbial
390 Consortium (MeCoCos) are groups of microbes that exhibit a positive feedback
391 loop, whereby they engineer their environment by both creating and using re-
392 sources. They constitute an intermediate level of organization between the com-
393 munity and the individual. These groups have been hypothesized to be resistant
394 against invasions because no other species would be able to harvest resources
395 rapidly enough to compete with the established members. My finding that more
396 metabolically cohesive consortia are more successful in community coalescence
397 experiments confirms their hypothesis. A possible line for future research would
398 be to use the metric of cohesion to identify MeCoCos in the synthetic communi-
399 ties, and track their behaviour to try to answer the question of whether or not
400 microbial community assembly can be understood as a succession of MeCoCos
401 coalescence events .

402 Another prediction of Pascual-García et al. (2020) is that MeCoCos efficiently
403 deplete resources to the lowest concentration. This result was obtained in the
404 absence of mutualistic interactions by Tikhonov (2016), who showed that when
405 two communities compete, the one that is more efficient at simultaneously de-
406 pleting all substrates will dominate. In his model, the microscopic dynamics con-
407 veniently took the form of optimizing a community level function (MacArthur
408 1969). In this more general model, collective dynamics were not reducible to
409 solving an optimization problem. Yet, the results here are consistent with his
410 work (see Figure 4). This confirms that the emergent cohesiveness reported here
411 and by many others, is a general consequence of ecological interactions, resource
412 partitioning, and the community shaping its own environment (niche construc-
413 tion).

414 Measuring cohesion in the synthetic communities used in this work was possible
415 because the theoretical framework of the model provided readily usable reaction
416 networks. The aim of these is to resemble bacterial metabolic pathways, and
417 thus lay out the foundations for the formulation of consumer-resource models
418 that characterize more realistic microbial communities. A promising direction of
419 research would be to focus on parametrizing the model based on available high-
420 resolution metabolic networks in the literature and then attempting to use it
421 to predict the outcome of community coalescence events. This could be applied
422 then to drive the community in question, through successive coalescence events,
423 towards states where certain functions are optimized (e.g. methane production,
424 as in Rillig et al. (2016), crops disease resistance as suggested in Calderón et al.
425 (2017) or the abundance of a healthy donor community in the gut microbiome,
426 reviewed in Wilson et al. (2019) and Wang et al. (2019)).

427 5 Discussion notes

- 428 • This thesis addresses the question of what are truly the mechanisms ex-
429 plaining what experiments show? An alternative measure of cohesiveness
430 that stems from more realistic modelling of microbial ecosystems is able
431 to reproduce these results and thus is closer to uncover what are the real
432 mechanisms behind community cohesion.

- 433 • community coalescence is a way to explicitly show and test the cohesiveness
434 of microbial communities while asking questions about how these commu-
435 nities came to be.
- 436 • discuss why I chose m as the upper limit for the number of reactions that
437 a strain can posses.
- 438 • Discuss why the traditional fitness (how fast resources are consumed) doesn't
439 correlate with what I call fitness: community cohesion. Show that in the
440 case of pure competition, it does (Tikhonov 2016), but in the case of purely
441 facilitation, it doesn't
- 442 • My measure of cohesion is an approximate one. Does facilitation help the
443 same degree that competition bothers?
- 444 • talk about environment engineering, and reference
- 445 • Maybe facilitation is actually not that important, But competition, and
446 functional groups, the ones that drive cohesion..
- 447 • There is no allusion to individual species fitness, because here its more
448 important the cohesion between them.
- 449 • Talk about innovation rather than improvement when it comes to facilita-
450 tion.
- 451 • One species can change the whole community because it affects all of it!
452 (all the elements in the matrix, or a good portion of them.)

453 6 Things to do in the future

- 454 • Ask Emma about papers of hierarchy of metabolites
- 455 • Find a paper that says that organisms with few and very energetic reactions
456 are rare.
- 457 • Should I include a page at the end specifying what things did I do, and
458 what things didn't I do, and that way I don't have to do it during the
459 paper?
- 460 • make a nice looking table of the parameters of the model.
- 461 • plot mean abundance as a function of number of reactions for both assembled
462 communities
- 463 • List of plots I want to make: 1. The s plot with richness == 5. 2. The
464 histogram of similarity with richness == 4. 3. The elimination assay. 4.
465 The cloud. 5. The evolution of cooperation. 6. The community reaction
466 network.
- 467 • Revise the cohesion of my thesis as a whole: are sections well separated?
468 Do they link well with each other? For example, at the end of the model
469 presentation section, I can introduce the next one by saying that I will
470 investigate the dynamics of community assembly, and then just start with

471 that right away. Additionally, at the end of the cohesion section, I can
 472 specify what is the type of coalescence events I am going to study next,
 473 namely, those in which the environment remains constant.

- 474 • When I talk about stable equilibrium, say that A stable (non- invade-
 475 able) equilibrium is characterized by an extra condition that all the absent
 476 species, if re-introduced, would be driven back to extinction
- 477 • Turn to my dictionary of cool words, and use them.
- 478 • Link my work with experimentalists.
- 479 • Find a reference for the claim: The cohesion of a community is ultimately
 480 determined by the nature of the interactions of its members.
- 481 • The cost of the model This cost model corresponds to the assumption of
 482 approximate neutrality.
- 483 • the proportionality constant is not important because I am only concerned
 484 with the equilibrium state where $(dN_\alpha/dt) = 0$
- 485 • Talk about competition minimization and link it to Pascual et al
- 486 • change color and labels of nodes in figure 3
- 487 • add a concluding paragraph to each section.
- 488 • Get rid of triangles, keep the weighted version. The point I am trying to
 489 make is not even that important.
- 490 • Consider transforming figure **3C** into a barplot, it seems more sensible.
- 491 • Talk about the positive correlation between $\langle F \rangle$ and $\langle C \rangle$
- 492 • In the presentation of the model, talk about the number of possible net-
 493 works that it has (effectively random), and also that different choice of
 494 initial biomass of each species would only alter the transient dynamics,
 495 but not the outcome of assembly: the equilibrium state where $\frac{dN_\alpha}{dt} = 0$
- 496 • Talk about simplifications of the model: deterministic dynamics and well-
 497 mixed environment (acknowledge their importance).
- 498 • Really important, plot survival rate against cohesion rank to uncover the
 499 fitness decoupling!
- 500 • Tend to be present at higher abundances, instead of tend to be more abun-
 501 dant.
- 502 • When talking about the fitness, mention that the success of a species is
 503 context dependent, and that organisms modify their own environment.
- 504 • Talk about avoiding priority effects and species sorting in the methods,
 505 when I say that the resources when mixing two communities are reset back
 506 to 2 equiabundant concentrations.

- 507 • Talk about diversifying selection when saying that specialists are more
508 abundant.
- 509 • Talk about species sorting when describing the model in biological terms:
510 species sorting results in competitive exclusion of species whose niches
511 overlap or are absent.
- 512 • Mutualisms explanation: many microbial interactions result in a net ben-
513 efit to the interacting partners.
- 514 • synonym for mutualism metabolic complementarity.
- 515 • Broadly interacting taxa, taxa with a high level of cohesion, are positively
516 co-selected in community coalescence.
- 517 • Consider changing altruistic communities (give without expecting some-
518 thing in return) with mutualistic communities (win-win) situation.
- 519 • cite Inferring metabolic mechanisms of interaction within a defined gut
520 microbiota?
- 521 • talk about how functionally redundant groups *those with similar metabolic
522 capabilities) tend to go extinct because they lead to competitive interac-
523 tions. On the other hand, members from different functional groups may
524 engage in mutualism or comensalism relationships. Members of related
525 through metabolic complementarity tend to occur.
- 526 • Include in the intro that including mutualisms has been suggested as a
527 next step.
- 528 • The same number of species to avoid selection effects.
- 529 • Change labels in nodes so that they are not s, which may be confused with
530 strength of the node.

531 7 Appendix

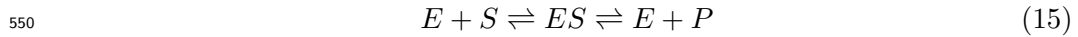
532 7.1 Reversible enzyme kinetics

533 Outside the the bacterial cell, the energy resides in the form of chemical potential
 534 μ held by the metabolites, and biochemical reactions inside the cell produce
 535 energy due to a difference in the chemical potentials of substrate and product. I
 536 assigned chemical potentials to each metabolite according to

$$537 \quad \mu_\beta = E \left(1 - \sqrt{\frac{\beta - 1}{m - 1}} \right) \quad (14)$$

538 where $\beta = 1, \dots, m$ and E is the energy of the most energetic metabolite. I have
 539 chosen this chemical potential function because I hope to find papers where they
 540 explain that there is a hierarchy on the metabolite energetic spectrum. This
 541 means that the energy produced by a reaction of the type $(\beta, \beta + 1)$ decreases as
 542 you go down the hierarchy. Reactions involving metabolites situated higher in
 543 the hierarchy are more energetic than reactions that involve those lower in the
 544 hierarchy.

545 The rate at which a given chemical reaction transforms substrate into product
 546 is modelled using reversible Michaelis-Menten enzyme kinetics. Thus, the model
 547 considers chemical reactions where a substrate S binds to an enzyme E to form
 548 an enzyme-substrate complex ES , which in turn produces a product P and
 549 recovers enzyme E .



551 The choice of fully reversible enzyme kinetics, instead of the traditional assump-
 552 tion of irreversibility in the second reaction, aims to capture more accurately the
 553 nature of biochemical reactions taking place in microbial communities. In these
 554 reactions the Gibbs energy change ΔG is not always big, which implies that the
 555 reaction of product formation can reach equilibrium at a similar time scale as
 556 the formation of the complex (Keener & Sneyd 2008). In this case, the tradi-
 557 tional irreversible Michaelis-Menten scheme breaks down, and more elaborated
 558 frameworks, like the fully reversible one that this model offers, need to be used.
 559 To comply with 2^{nd} law of thermodynamics, the network G_α is completely hier-
 560 archical, ie. the edges are unidirectional ($x < y$), going from the more energetic,
 561 to the less energetic metabolite. Thus, for the reaction scheme in 15 and the
 562 imposed thermodynamic constraint only reactions where $\Delta G^0 = \mu_P - \mu_S < 0$
 563 can take place.

564 With all the above considerations, the expression for the rate of reaction i poss-
 565 esed by strain α is given below. A formal derivation of equation 16 can be found
 566 in Hoh & Cord-Ruwisch (2000)

$$567 \quad q_{\alpha i} = \frac{q_m^{\alpha i} S_\alpha (1 - \theta_\alpha)}{K_S^{\alpha i} + S_\alpha (1 + k_R^{\alpha i} \theta_\alpha)} \quad (16)$$

568 Here, θ_α measures how far is the reaction from equilibrium (0 being the furthest,
 569 and 1 being equilibrium).

$$570 \quad \theta = \frac{[P]}{[S]K_{eq}} \quad (17)$$

571 where $[]$ denote concentration and K_{eq} is the equilibrium constant

$$572 \quad K_{eq} = \exp \left(\frac{-\Delta G^0 - \eta \Delta G_{ATP}}{RT} \right) \quad (18)$$

573 The energy produced by the reaction is then stored in the form of ATP molecules.
 574 In the model, η represents the moles of ATP molecules produced per mole of
 575 reaction. For a given reaction (x, y) eta I calculate eta as

$$576 \quad \eta = \frac{y - x}{m} \quad (19)$$

577 which represents the normalized metabolite gap between substrate and product
 578 of the reaction. Therefore, the higher the gap, the more energy will be stored.
 579

580 7.2 Table of parameter values and meaning

Parameter	Meaning	Value
m	Number of metabolites	100
s	Number of strains	10
ΔG_{ATP}	ATP Gibbs energy	$7.5 \cdot 10^4$
μ_0	Most energetic metabolite	$3 \cdot 10^4$
nATP	$\max \left(\frac{\Delta G_{S \rightarrow P}^0}{\Delta G_{ATP}} \right)$	4
η	Moles of ATP energy per reaction	0.5
q_m	Maximum reaction rate	1
K_S	Saturation constant	0.1
k_r	Reversibility constant	10
g	Growth factor	1
m	Maintenance factor	$0.2 \cdot J_{grow}$
κ	Externally supplied resource	1
γ	Dilution rate	0.5
N_0	Populations initial conditions	(1, 1, ..., 1)
C_0	Concentrations initial condition	(0, 0, ..., 0)

Table 1: Parameter meanings and their values

581 7.3 Relationship between Θ , F , C , and Ψ

582 7.4 Calculation of individual fitness

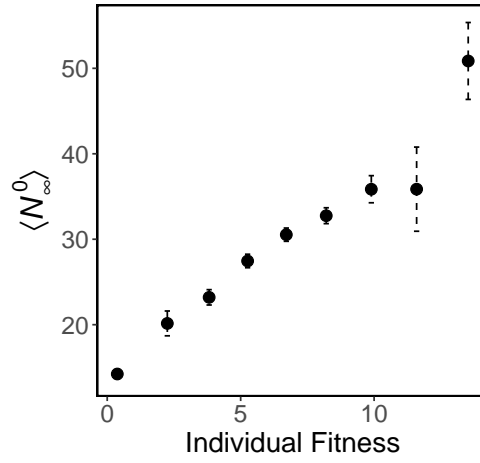


Figure 5: Abundance of 500 random species at isolated equilibrium as a function of proxy of individual fitness. Shown is binned mean (10 bins) over species with similar individual fitness. Errorbars are 1 standard error.

References

- Calderón, K., Spor, A., Breuil, M. C., Bru, D., Bizouard, F., Violle, C., Barnard, R. L. & Philippot, L. (2017), ‘Effectiveness of ecological rescue for altered soil microbial communities and functions’, *ISME Journal* .
- Castledine, M., Sierocinski, P., Padfield, D. & Buckling, A. (2020), ‘Community coalescence: An eco-evolutionary perspective’.
- Costello, E. K., Stagaman, K., Dethlefsen, L., Bohannan, B. J. & Relman, D. A. (2012), ‘The application of ecological theory toward an understanding of the human microbiome’.
- Dormand, J. R. & Prince, P. J. (1980), ‘A family of embedded Runge-Kutta formulae’, *Journal of Computational and Applied Mathematics* .
- Embree, M., Liu, J. K., Al-Bassam, M. M. & Zengler, K. (2015), ‘Networks of energetic and metabolic interactions define dynamics in microbial communities’, *Proceedings of the National Academy of Sciences of the United States of America* .
- Falkowski, P. G., Fenchel, T. & Delong, E. F. (2008), ‘The microbial engines that drive earth’s biogeochemical cycles’.
- Friedman, J., Higgins, L. M. & Gore, J. (2017), ‘Community structure follows simple assembly rules in microbial microcosms’, *Nature Ecology and Evolution* .
- Gilbert, J. A., Jansson, J. K. & Knight, R. (2014), ‘The Earth Microbiome project: Successes and aspirations’.
- Gilpin, M. (1994), ‘Community-level competition: Asymmetrical dominance’, *Proceedings of the National Academy of Sciences of the United States of America* .
- Goldford, J. E., Lu, N., Bajić, D., Estrela, S., Tikhonov, M., Sanchez-Gorostiaga,

- A., Segrè, D., Mehta, P. & Sanchez, A. (2018), ‘Emergent simplicity in microbial community assembly’, *Science* .
- Goyal, A. & Maslov, S. (2018), ‘Diversity, Stability, and Reproducibility in Stochastically Assembled Microbial Ecosystems’, *Physical Review Letters* .
- Hansen, S. K., Rainey, P. B., Haagenzen, J. A. & Molin, S. (2007), ‘Evolution of species interactions in a biofilm community’, *Nature* .
- Hoh, C. Y. & Cord-Ruwisch, R. (2000), ‘A practical kinetic model that considers endproduct inhibition in anaerobic digestion processes by including the equilibrium constant’, *Biotechnology and Bioengineering* .
- Huttenhower, C., Gevers, D., Knight, R. & Al., E. (2012), ‘Structure, function and diversity of the healthy human microbiome’, *Nature* .
- Keener, J. P. & Sneyd, J. (2008), ‘Mathematical Physiology’, *Book* .
- Lawrence, D., Fiegna, F., Behrends, V., Bundy, J. G., Phillimore, A. B., Bell, T. & Barraclough, T. G. (2012), ‘Species interactions alter evolutionary responses to a novel environment’, *PLoS Biology* .
- Livingston, G., Jiang, Y., Fox, J. W. & Leibold, M. A. (2013), ‘The dynamics of community assembly under sudden mixing in experimental microcosms’, *Ecology* .
- Lu, N., Sanchez-gorostiaga, A., Tikhonov, M. & Sanchez, A. (2018), ‘Cohesiveness in microbial community coalescence’, *bioRxiv* .
- MacArthur, R. M. (1969), ‘SPECIES PACKING, AND WHAT COMPETITION MINIMIZES’, *Proceedings of the National Academy of Sciences* .
- Marsland, R., Cui, W., Goldford, J., Sanchez, A., Korolev, K. & Mehta, P. (2019), ‘Available energy fluxes drive a transition in the diversity, stability, and functional structure of microbial communities’, *PLoS Computational Biology* .

- McGill, B. J., Enquist, B. J., Weiher, E. & Westoby, M. (2006), ‘Rebuilding community ecology from functional traits’, *Trends in Ecology and Evolution* .
- McIntire, E. J. & Fajardo, A. (2014), ‘Facilitation as a ubiquitous driver of biodiversity’, *New Phytologist* .
- Pascual-García, A., Bonhoeffer, S. & Bell, T. (2020), ‘Metabolically cohesive microbial consortia and ecosystem functioning’.
- Rillig, M. C., Antonovics, J., Caruso, T., Lehmann, A., Powell, J. R., Veresoglou, S. D. & Verbruggen, E. (2015), ‘Interchange of entire communities: Microbial community coalescence’.
- Rillig, M. C. & Mansour, I. (2017), ‘Microbial Ecology: Community Coalescence Stirs Things Up’.
- Rillig, M. C., Tsang, A. & Roy, J. (2016), ‘Microbial community coalescence for microbiome engineering’.
- Sierocinski, P., Milferstedt, K., Bayer, F., Großkopf, T., Alston, M., Bastkowski, S., Swarbreck, D., Hobbs, P. J., Soyer, O. S., Hamelin, J. & Buckling, A. (2017), ‘A Single Community Dominates Structure and Function of a Mixture of Multiple Methanogenic Communities’, *Current Biology* .
- Tikhonov, M. (2016), ‘Community-level cohesion without cooperation’, *eLife* .
- Tikhonov, M. & Monasson, R. (2017), ‘Collective Phase in Resource Competition in a Highly Diverse Ecosystem’, *Physical Review Letters* .
- Toquenaga, Y. (1997), ‘Historicity of a simple competition model’, *Journal of Theoretical Biology* .
- Wang, J. W., Kuo, C. H., Kuo, F. C., Wang, Y. K., Hsu, W. H., Yu, F. J., Hu, H. M., Hsu, P. I., Wang, J. Y. & Wu, D. C. (2019), ‘Fecal microbiota transplantation: Review and update’.
- Wilson, B. C., Vatanen, T., Cutfield, W. S. & O’Sullivan, J. M. (2019), ‘The super-donor phenomenon in fecal microbiota transplantation’.