Imperial College London

Coalescence of cohesive microbial communities

Pablo Lechón

Supervisor: Dr. Samraat Pawar

August 2020

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Imperial College London

Formatted in the journal style of *Nature Ecology and Evolution*

Submitted for the MSc in Computational Methods in Ecology and Evolution

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 that govern these events are poorly understood, partly because theoretical work in community coalescence rarely considers communities with mutualistic interactions, which are pervasive in microbial consortia. In this work, I use a new consumer-resource model to simulate communities harbouring competitive and mutualistic interactions, and propose a measure of community cohesion that predicts the outcome of microbial community coalescence. The proposed metric explicitly quantifies the so-called *cohesivness* exhibited by microbial communities. It reproduces an important previous result, i.e., that more cohesive communities are more successful in community coalescence events while pinning it down to more realistic assumptions about the interactions in the community. The consistency of my results with previous works demonstrates that the collective coherence exhibited by coalescing communities is a general consequence of ecological interactions, resource partitioning, and the community shaping its environment. The proposed cohesion measure can be used to guide coalescence experiments in which different communities are successively combined until a desired beneficial functioning is reached.

Contents

1 Introduction

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

Early mathematical models of community-community invasion revealed that when two communities previously separated by a barrier merge due to its removal, asymmetrical dominance of one community over the other one is likely to occur (Gilpin 1994, Toquenaga 1997). As an explanation for this observation, it was argued that, because communities have been assembled through a history of competitive exclusion, they are likely to compete with each other as coordinated entities, rather than as a random collection of species. This result is also shown in new theoretical work, where consumer-resource models are used to show that coalescing microbial communities exhibit an emergent cohesiveness (Tikhonov 2016, Tikhonov & Monasson 2017). These findings suggest that communities arising from the struggle for existence of its members display a certain level of coherence. These communities have been termed Metabolically Cohesive [microbial] Consortium (MeCoCos) by Pascual-García et al. (2020) and suggested to be pervasive in microbial guilds.

Recent results from coalescence experiments of methanogenic communities suggest that during a coalescence event between two communities, multiple taxa from the same community act as cohesive units and are selected together (ecological co-selection) (Sierocinski et al. 2017). Further experimental evidence of co-selection in community coalescence has been reported in Lu et al. (2018), where it was shown that the invasion success of a given taxon is determined by its community members. The microbial communities used in these experiments are characterized by complex cross-feeding interactions (Hansen et al. 2007, Lawrence et al. 2012, Embree et al. 2015), where the metabolic by-products of one species are substrates for others. Furthermore, the type of interactions

present in a community has been suggested as a factor that might affect the outcome of community coalescence (Castledine et al. 2020). Yet, theoretical models used in community coalescence studies so far have considered competition between species as the only force driving community assembly.
In this work, I explore the combined role of other types of interactions, namely, competition and mutualism, which appear to be ubiquitous in microbial communities. First, I use a new consumer-resource model that includes both facilitation of metabolites via by-product secretion, and competition for substrates, to simulate many instances of community assembly. Second, I propose a metric of community cohesion that accounts for both competitive and mutualistic interactions, and I measure the cohesion level in the simulated communities. Third, I apply the proposed metric to predict the outcome of microbial community coalescence events.

2 Methods

The structure of this section is threefold. First, I lay out the model used to simulate communities. Second, I detail the procedure I follow to assembly many synthetic microbial communities. Third, I present the metric of cohesion and use it on the simulated communities.

2.1 Consumer-resource model with cross feeding interactions

In order to simulate communities with cross-feeding interactions, I use a consumer-resource model based on the the work of Marsland et al. (2019).
Consider an environment with m resources present in different concentrations C_β ,where $\beta \in \{1 \dots m\}$. Let now N_α denote the abundance of each bacterial strain α, present in the environment, where $\alpha \in \{1 \dots s\}$. Each species is uniquely characterized by the metabolic strategy it uses to harvest resources. This strategy is encoded in its reaction network G_α, a collection of chemical reactions between the metabolites in the environment that produce energy that is used by bacteria for survival and replication (see Figure 1). 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. 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 (in the network theory sense), where $\mathcal{M}$ is a set of nodes $\mathcal{M} = \{x : x$ is an integer from the interval $[1, m]$ labeling the metabolite$\}$ and $\mathcal{N}$ a set of uni-directed edges $\mathcal{N} = \{(x, y) : x \in \mathcal{M}$x , $y \in \mathcal{M}$ and $x < y$ (x and y 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

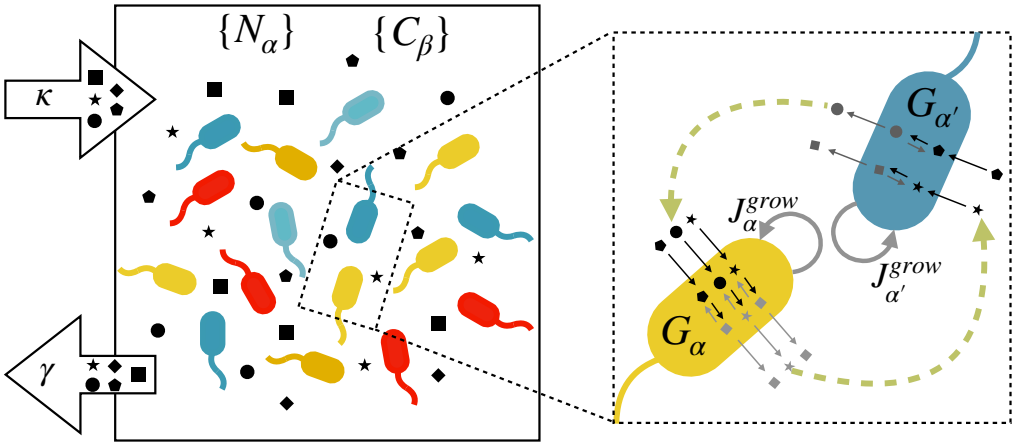

Figure 1: **Schematic of the model**. (left) Consider a chemostat where m metabolites are steadily supplied at rate κ and diluted at rate γ. Different bacterial strains coexist in the chemostat, and they consume the metabolites in the environment, C_β through their reaction networks G_α (right), to obtain the necessary power J_α^{grow} to increase their abundance N_α. 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 I am considering reversible reactions, so the reaction rate depends on both substrate and product concentrations.

across all reactions in $\mathcal{N}$.

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

where $|\ \ |$ 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

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

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 able to produce, and (2) the energy yielded by the reactions in which these enzymes are involved. Intuitively, the more enzymes a species is capable to express, the higher is the probability that at least two of them require starkly different conditions for optimal function. Therefore, expressing those enzymes simultaneously would imply maintaining two separate compartments, and incur in extra cost (see Supplementary material in Tikhonov & Monasson (2017)). The effect of the cost function (Eq. 2) is to ensure 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

$$\frac{dN_\alpha}{dt} = g_\alpha N_\alpha \left[J_\alpha^{grow} - \chi_\alpha\right] \tag{3}$$

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 possesed by α in which β is a substrate. The outgoing flux is that due to reactions in which β is a product.

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

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

$$h_\beta = \kappa - \gamma C_\beta \tag{5}$$

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

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

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

2.2 Community Assembly

I use the previous model to assemble $n_s = 2 \cdot 10^3$ communities by integrating the set of equations in 3 and 6. Each simulation starts with 10 bacterial strains ($s = 10$) in an environment with 15 different resources ($m = 15$) and a random realization of the metabolic strategy, G_α, for each species.
The values of the parameters of the model remain constant throughout all simulations, and have been chosen motived by biological processes (total free energy of photosintesis or glucose respiration) and to avoid pathological situations (species always being saturated or maintenance being higher than the maximum amount of energy that can be harvested) (subsection 7.2, table 1). 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 endavour 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 (see figure

2B), and the reactions possesed by strain α can be expressed as $(i, i+k)$, where k represents the kth 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.

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

Several things are important to note about this algorithm. Firstly, sampling k from a truncated normal distribution ensures that high metabolite gaps (very energetic reactions) are not likely to happen. This introduces a bias against the precence of organisms with few and very energetic reactions, which are rare in microbial communities. Second, the truncation limits in step 2 have been chosen to respect the imposed constraint that reactions can only be of the form $i < j$. Third, the upper limit of the uniform distribution from which i is sampled is bounded by k, the diagonal we are sampling from.
When the sampling of reaction networks is completed, Eqs. 6 and 3 are integrated using a Runge Kutta method (Dormand & Prince 1980) with initial conditions $N_\alpha(t=0) = 2$ and $C_\beta(t=0) = 0$.

2.3 A metric of community cohesion

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

$$Cohesion = Facilitation - Competition \tag{7}$$

Measuring levels of facilitation and competition within a microbial community is experimentally challenging. However, the metabolic strategies of each species are well determined in this theoretical framework. Therefore, I use the reaction network of each bacterial strain to compute their competition and facilitation indices with the rest of the species in the community.
Let s_1 and s_2 be two sequences of integers labeling metabolites. I am interested in measuring their *overlapping degree* $\xi(s_1, s_2)$, ie, the proportion of metabolites of s_1 that intersect with s_2 summed with the proportion of metabolites of s_2 that intersect with s_1, normalized to 1.

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

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

One way to capture the facilitation of a community is by calculating its facilitation matrix F, which is composed of the facilitation indices of all posible ordered pairs i, j of species in the community. Precisely, the facilitation index f_{ij} of species i towards species j, is given by the overlapping degree of the sequence of products y_i of species i, with the sequence of substrates x_j of species j. Equivalently, the competition matrix C gathers the competition level of the community. The competition index between species i and j, c_{ij} is given by the overlapping degree of the sequence of substrates x_i of species i, and the sequence of substrates x_j of species j. Thus,

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

Note that facilitation is directional but competition is not. This implies that $F_{ij} \neq F_{ji}$ and F is not symmetric, but $C_{ij} = C_{ji}$ and C is symmetric.

Following the idea sketched in equation 7, a cohesion matrix Ψ can be defined using equations in 9, as an upper triangular $s \times s$ matrix whose elements are given by

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

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

3 Results

Relevant results steming from the simulations of community assembly events are plotted in figure **2**. The first two figures convey information about the dynamics and the resource consumption map of one particular community. In figure (**2A**), all abundances start increasing because all resources are present and steadily supplied. As the dynamics evolve, the community engineers its own environment by consumption of metabolites and secretion of by-products, causing the creation of ecological niches. During this process, species sorting results in competititve exclusion of species whose niches overlap. Alternatively, species whose niches are separate may engage in mutualistic relationships through metabolic complementarity (see figure **1**), resulting in a net benefit to the interacting partners (Pascual-García et al. 2020). Figure (**2B**), shows that all metabolites are being consumed (all rows have at least one non-zero element), which is to say,

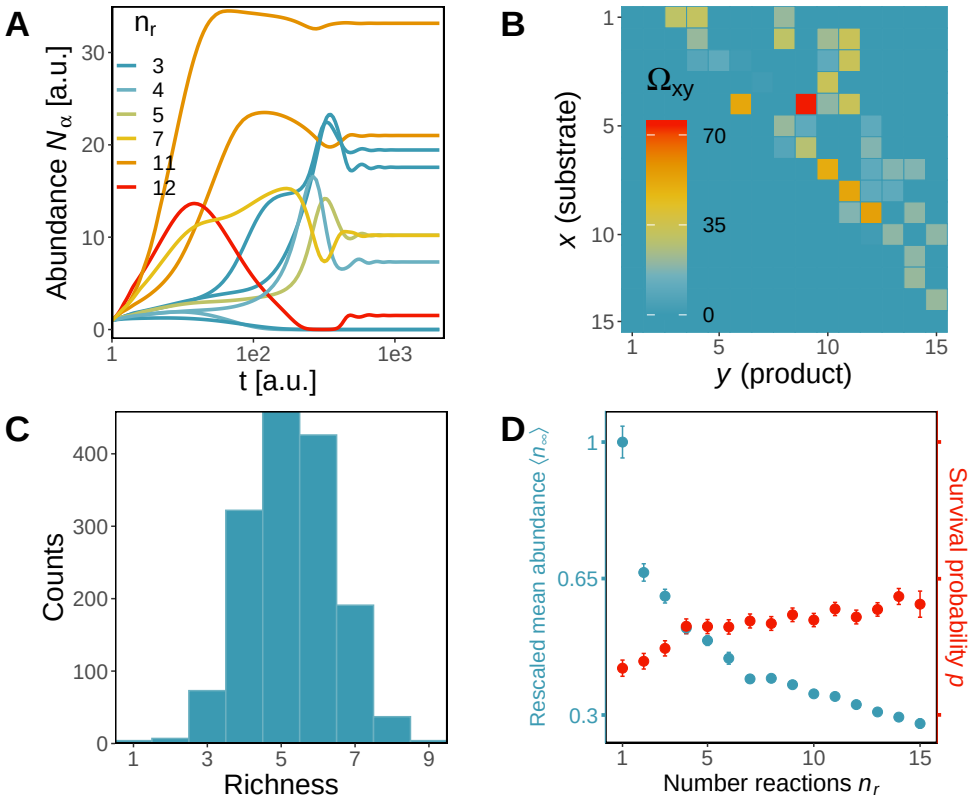

Figure 2: **Results from community assembly simulations**. Plots (**A**) and (**B**) exemplify one community assembly event and (**C**) and (**D**) convey results accross simulations. (**A**) Time variation of species' abundance for one instance of community assembly with $m = 15$, $s = 10$, and a set 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 possesed 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 abscent, 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.

that all vacant niches are being occupied.

The community assembly simulations generated communities with richness spanning fom 1 to 9 species (figure **2C**), where specialists are, on average, more abundant than generalists (figure **2D**, blue points). This can be attributed to several specialists being able to deplete all resources through their combined action more efficiently than one generalist (Pascual-García et al. 2020), thus, dominating the community at equilibrium. Note that in the simulated communities, while specialists tend to be present at higher abundances than generalists, their survival probability is lower (figure **2D**, red points). Generalists have more alternatives to obtain energy than specialists. Consequently, they are more resistant to extinctions.

During community assembly, I track the cohesion network adjacency matrix, Ψ. An example of this network is shown in figure **3A** (before community assembly) and figure **3B** (after community assembly). In these figures, the 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

$$s_\alpha = \sum_{j \neq \alpha} \Psi_{\alpha j} \tag{11}$$

The two networks ilustrated 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 also plotted 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 (Tikhonov 2016, McGill et al. 2006, Mcintire & Fajardo 2014).

The stabilization of the communities during the assembly 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 each time an extinction takes place.

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

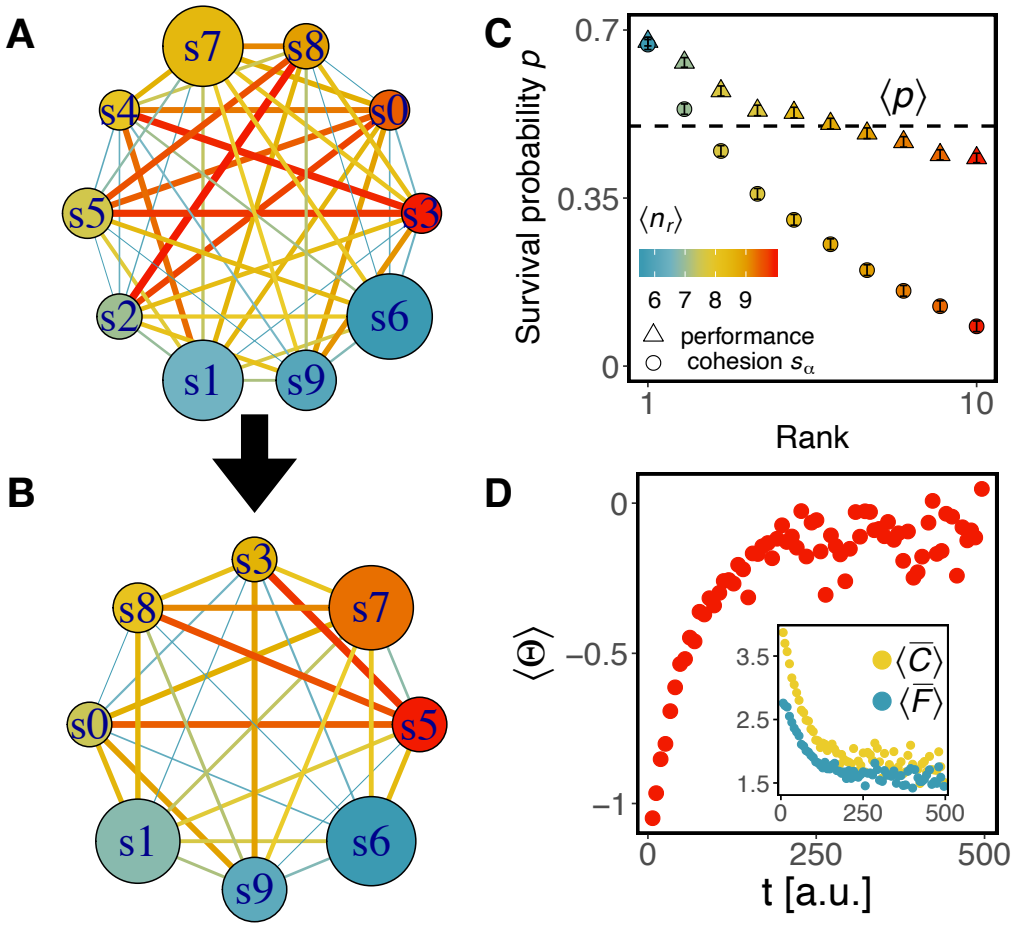

Figure 3: **Cohesion metric across community assembly events.** Cohesion network before (**A**) and after (**B**) community assembly. Thicker and red edges represent higher cohesion coefficient Ψ_{ij} between its nodes, species i and j. The color of the nodes changes from red to blue in anti-clockwise direction indicating decreasing total cohesion level s_α. The size of the node represents the number of reactions possesed 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 asesmbly) accross 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).

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**).

Having assembled many communities, and measured their cohesion levels, I am in good shape to perform community coalescence experiments of pairs of communities.

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 succesfully?

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, $\Theta >> 0$ and $\Theta << 0$. 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

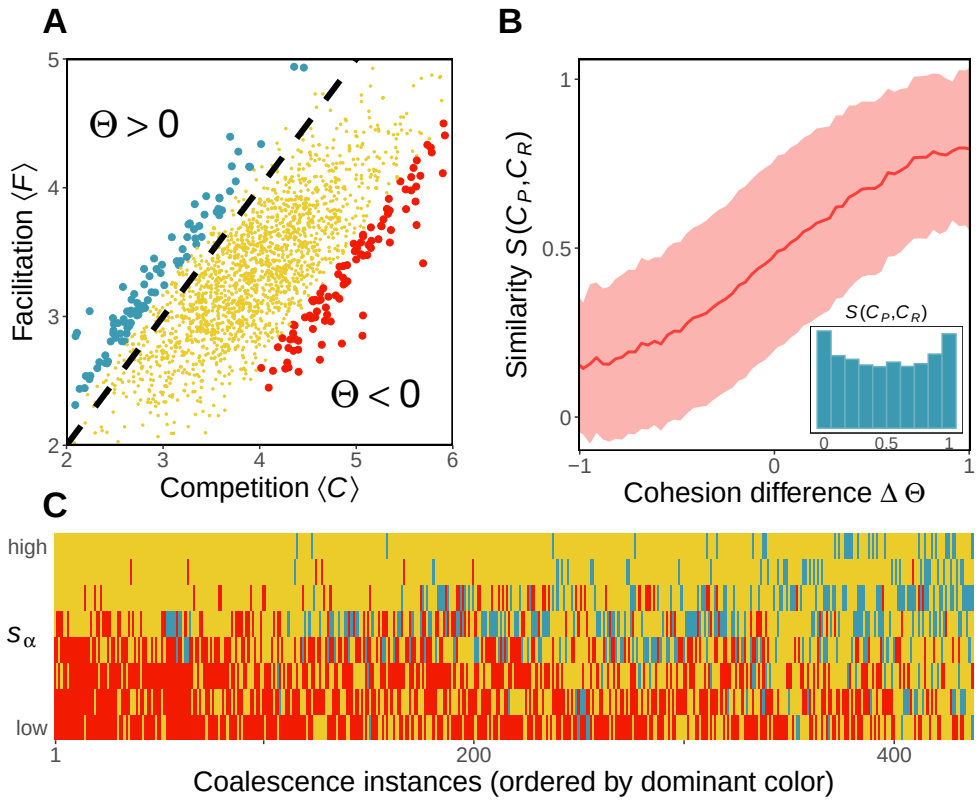

Figure 4: **Results from community coalescence experiments**. (**A**) Each simulated community is plotted in a competition-facilitation diagram. Communities above the dashed line $\langle C \rangle = \langle F \rangle$ have $\Theta > 0$, and thus they are in the facilitation-dominated regime. Communities 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, $S(\mathcal{C}_P, \mathcal{C}_R)$ 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.

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

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 *superorganism*, it is also a perilous simplification to consider it as a mere collection of

individuals, ignoring the fact that coevolutionary processes can play an important role in it (Rillig & Mansour 2017).

A recent idea that smoothly interpolates between these to extremes has been proposed in (Pascual-García et al. 2020). Metabolically Cohesive microbial Consortium (MeCoCos) are groups of microbes that exhibit a positive feedback loop, whereby they engineer their environment by both creating and using resources. They constitute an intermediate level of organization between the community and the individual. These groups have been hypothesized to be resistant against invasions because no other species would be able to harvest resources rapidly enough to compete with the established members. My finding that more metabolically cohesive consortia are more successful in community coalescence experiments confirms their hypothesis. A possible line for future research would be to use the metric of cohesion to identify MeCoCos in the synthetic communities, and track their behaviour to try to answer the following question: can we understand microbial community assembly as a succession of MeCoCos coalescence events?

Another prediction of Pascual-García et al. (2020) is that MeCoCos efficiently deplete resources to the lowest concentration. This result was obtained in the absence of mutualistic interactions by Tikhonov (2016), who showed that when two communities compete, the one that is more efficient at simultaneously depleting all substrates will dominate. Intuitively, the more efficient community (higher community-level fitness) succeeds because its able to engineer an environment suitable for its members. In his model, the dynamics conveniently took the form of optimizing a community level function (MacArthur 1969). In this more general model, collective dynamics were not reducible to solving an optimization problem, yet, the results here are consistent with Dr. Tikhonov's work. Specifically, the cohesive communities of my work behave in the same way as the more efficient (or more fit) communities in his work when it comes to community coalescence (see figure **4**). This confirms that more cohesive communities are also more efficient at depleting resources simultaneously.

The results obtained in theoretical works of community-community competition up to date have been consistent accross different models, and types of interactions. Therefore, the coherence exhibited by microbial communities seems to be a general consequence of ecological interactions, resource partitioning, and the community shaping its own environment (niche construction).

Several simplifications were made in this study. First, the parameters of the model were kept constant throughout the simulations. The reason behind this is that the aim of this study was to show that the strength and type of interactions present in microbial communities influence the outcome of community coalescence. For this purpose, experimental parametrization of the model was not strictly necessary, and was left to be developed in future work. Second, the model used here left out stocastic effects and spacial structure. Although these two considerations are tremendously important in most cases, their removal in this work helped to smoothly build on the work of Tikhonov (2016), and to make the dynamics computationally tractable. Third, the measure of community-level cohesion, Θ, was calculated as the rough average of all pairwaise cohesion levels. A more refined measure of community cohesion, for example, counting the number of closed feedback loops, weighted by their strenght, might help reduce the noise of the correlation (see shade in figure **4B**).

4.1 Empirical relevance

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

5 Discussion notes

- This thesis addresses the question of what are truly the mechanisms explaining what experiments show? An alternative measure of cohesiveness that stems from more realistic modelling of microbial ecosystems is able to reproduce these results and thus is closer to uncover what are the real mechanisms behind community cohesion.
- community coalescence is a way to explicitly show and test the cohesiveness of microbial communities while asking questions about how these communities came to be.
- discuss why I chose m as the upper limit for the number of reactions that a strain can posses.
- Discuss why the traditional fitnes (how fast resources are consumed) doesn't correlate with what I call fitness: community cohesion. Show that in the case of pure competition, it does (Tikhonov 2016), but in the case of purely facilitation, it doesn't
- My measure of cohesion is an aproximate one. Does facilitation help the same degree that competition bothers?
- talk about environment ingeneering, and reference
- Maybe facilitation is actually not that important, But competition, and functional groups, the ones that drive cohesion..
- There is no alusion to individual species fitness, because here its more important the cohesion between them.
- Talk about innovation rather than improvement when it comes to facilitation.
- One species can change the whole community because it affects all of it! (all the elements in the matrix, or a good portion of them.)

6 Things to do in the future

- Ask Emma about papers of hierarchy of metabolites
- Find a paper that says that organisms with few and very energetic reactions are rare.
- Should I include a page at the end specifying what things did I do, and what things didn't I do, and that way I don't have to do it during the paper?
- make a nice looking table of the paramters of the model.
- Revise the cohesion of my thesis as a whole: are sections well separated? Do they link well with each other? For example, at the end of the model presentation section, I can introduce the next one by saying that I will investigate the dynamics of community assembly, and then just start with that right away. Additionaly, at the end of the cohesion section, I can specify what is the type of coalescence events I am going to study next, namely, those in which the environmet remains constant.
- Turn to my dictionary of cool words, and use them.
- Find a reference for the claim: The cohesion of a community is ultimately determined by the nature of the interactions of its members.
- The cost of the model This cost model corresponds to the assumption of approximate neutrality.
- change color and lables of nodes in figure 3
- add a concluding paragraph to each section.
- Consider transforming figure **3C** into a barplot, it seems more sensible.
- In the presentation of the model, talka bout the number of possible networks that it has (efectively random), and also that different choice of initial biomass of each species would only alter the transient dynamics, but not the outcome of assembly,: the equilibrium state where $\frac{dN_\alpha}{dt} = 0$
- Talk about simplifications of the model: deterministic dynamics and well-mixed environment (acknowledge their importance).
- When talking about the fitness, mention that the success of a species is context dependent, and that organisms modify their own environment.
- Talk about avoiding priority effects and species sorting in the methods, when I say that the resources when mixing two communities are reset back to 2 equiabundant concentrations.
- Talk about diversifying selection when saying that specialists are more abundant.
- Talk about species sorting when describing the model in biological terms:
- synonim for mutualism metabolic complementarity.

- Broadly interacting taxa, taxa with a high level of cohesion, are positively co-selected in community coalescence.
- cite Inferring metabolic mechanisms of interaction within a defined gut microbiota?
- talk about how functionally redundant groups *those with similar metabolic capabilities) tend to go extinct because they lead to competitive interactions. On the other hand, members from different functional groups may engage in mutualism or comensalism relationships. Members of related through metabolic complementarity tend to occur.
- Include in the intro that including mutualisms has been suggested as a next step.
- The same number of species to avoid selection effects.
- Change labels in nodes so that they are not s, which may be confused with strength of the node.
- figure, or Figure. equation, Eqs, eq??
- Should I include any discussion on bottom-up vs top-down assembly?
- Talk about the upper bound and justify why my communities aren't richer than 9 speces. Read Cui et al. (2020)
- Make the discusssion experimental relevance more extensive by explaining the feneomenon in which it can help, cite and explain Calderón et al. (2017)
- Figure 2 B is not well explained either in the caption, or in the main text.

7 Appendix

7.1 Reversible enzyme kinetics

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

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

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

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

$$E + S \rightleftharpoons ES \rightleftharpoons E + P \tag{15}$$

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

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

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

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

$$\theta = \frac{[P]}{[S] K_{eq}} \tag{17}$$

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

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

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

$$\eta = \frac{y - x}{m} \tag{19}$$

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

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	$\max\left(\frac{\Delta G^0_{S\to P}}{\Delta G_{ATP}}\right)$	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

7.3 Relationship between Θ, F, C, and Ψ

7.4 Calculation of individual fitness

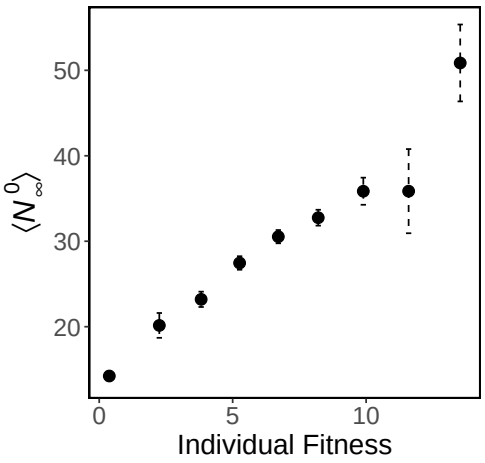

Figure 5: Abundance of 500 random species at isolated equilibrium as a function of proxy of individual fitness. Shown is binned mean (10 bins) over species with similar individual fitness. Errorbars are 1 standard error.

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'.

Cui, W., Marsland, R. & Mehta, P. (2020), 'Effect of Resource Dynamics on Species Packing in Diverse Ecosystems', *Physical Review Letters* .

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., Haagensen, 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'.