Master thesis title

Pablo Lechón Alonso

Imperial College London

Word count: 10

Abstract

Contents

1 Introduction notes

- State my aims/hypotheses/questions by the end of the introduction.
- Make sure I explain everything adequately. Provide more background in the introduction and methods that
- What is the problem I am tackling? you end up looking forward to read more
- the general way the problem has been approached.
- Clearly define my aims of the rearch project.
- Why is it interesting? Why don't we know the answer?
- Build from the most general and fundamental hypotheses to the most refined or teneous ones.
- How will I go about testing my hypothesis.

Following, what I want to talk about

- Microbial Community Coalescence in communities with mutualisms ie, cross-feeding.
- Thermodynamic constraints
- What drives a community to be succesful in a coalescence event?
- 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?
- Experimental studies report cohesivness in comunity coalescence events.

- talk about why this model is good with methanogenic communities, strong cross feeding
- say at some point how many possible networks there are.
- talk about coselection
- measure community productivity and check that the correlation breaks down because there is co-selection.
- stress importance of dominant community interaction types
- microbial ensamblages
- Theoretical efforts for community coalescence (Tikhonov 2016, Livingston et al. 2013, Toquenaga 1997, Gilpin 1994)
- Experimental efforts to understand community coalescence (Lu et al. 2018, Sierocinski et al. 2017)
- Community coalescence is just another expression of the fight competition vs facilitation...

2 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 harboring 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 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 usual 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 happens repeatedly 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. New theoretical work uses consumer-resource models to show that coalescing microbial communities exhibit an emergent cohesiveness (Tikhonov 2016, Tikhonov & Monasson 2017). These communities have been termed Metabolically Cohesive [microbial] Consortium (MeCoCos) by Pascual-García et al. (2020), and suggested to be pervasive in microbial ecology. 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 successful collective invasions are accompanied by strong community-level interactions. 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). Furthermore, the type of trophic 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 coalescencee studies so far have considered competition between species as the only force driving community assembly.

In this work, I explore the role of other types of interactions, which appear to be ubiquitous in microbial communities. Specifically, I propose a metric of community cohesion that accounts for both competitive and mutualistic interactions. I then use a consumer-resource model that includes both facilitation of metabolites via by-product secretion, and competition for substrates, to simulate many instances of community assembly . Finally, I measure the cohesion level on the simulated communities and use it to predict the outcome of microbial community coalescence events.

3 Methods and Results

3.1 Consumer resource model with cross feeding interactions

Simpler description propositions:

- The availability h_i of each resource i determines the dynamics of n_μ. The changes in species abundance translate into changes in the total demand for resources, denoted T_i . This total demand, in turn, depletes resource availability h_i .
- The cost of the model This cost model corresponds to the assumption of approximate neutrality.
- the proportionality constant is not important because I am only concerned with the equilibrium state where $(dN_\alpha/dt) = 0$

In order to simulate communities with cross-feeding interactions, I use a consumer-resource model inspired in the the work of (Marsland et al. 2019) that PhD student Jacob Cook has developed. I modify parts of this model to make it more suitable for my purposes.

I consider the population dynamics of s consumers (eg. bacterial strains) that feed on m resources. In this model, a species is defined by its metabolic stratetgy 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 $\mathcal{M} = \{x : x$ is an integer from the interval $[1, m]$ labeling the metabolite$\}$ and $\mathcal{N}$ a set of uni-directed edges $\mathcal{N}$ $= \{(x, y) : x \in \mathcal{M}$, $y \in \mathcal{M}$ and $x < y$ (x and y are product and substrate of a chemical reaction)$\}$. The growth power of species α, J_α^{grow} will be given by the product of the amount of generated energy η_i and rate q_i of each reaction, summed across all reactions in $\mathcal{N}$.

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

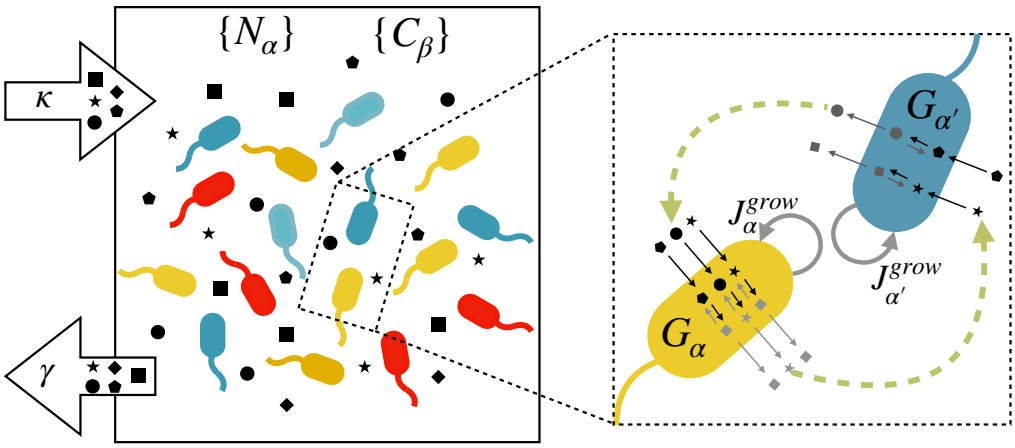

Figure 1: **Schematic of the model**. (left) All m metabolites are steadily supplied into a chemostat with different bacterial strains at rate κ and diluted at rate γ. (right) Bacteria use the metabolites in the environment, $\{C_\beta\}$ through their reaction networks G_α, to obtain the necessary power J_α^{grow} to increase their abundance $\{N_\alpha\}$ by replicating. The green arrows in the magnified portion emphasize that species α (yellow) facilitates metabolites to species α' (blue) and viceversa. The duble arrows in the reactions happening inside the cells stand for the reversible enzyme kinetics considered by this model.

where $|\ \ |$ denotes cardinality of a set. Refer to subsection 6.1 for specifications on q and η .

Every species has a mantenance cost χ_α that represents the required energy to sustain life, and I model as

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

where χ_0 is the average cost per reaction, and the summatory term adds up the metabolite gap of all reactions. Therefore, the more reactions a speciess has or/and the more energetically they are, the higher is the maintenance cost. 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 writtne 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 extrenal 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 consists of $s + m$ coupled differential equations completely specified by Eqs. 3 & 6.

3.2 Community Assembly

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

The values of the of the parameters of the model (subsection 6.2, 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 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, $G_\alpha(\mathcal{M}, \mathcal{N})$ using the following procedure. Consider, the $m \times m$ adjacency matrix A^α_{ij}, whose elements represent the edges (i, j) of G_α. Since the reaction network is hierarchical ($i < j$, subsection 6.1), the adjacency matrix is an upper triangular matrix with zeros in the main diagonal, and the reactions possesed by strain α can be expressed as $(i, i + k)$, where k represents the kth diagonal of A ($k = 1, \dots m - 1$ 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 pairs (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 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.

Some notes about this algorithm are, first, 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 super-organisms with few and very energetic reactions, which are rare in microbial communities.

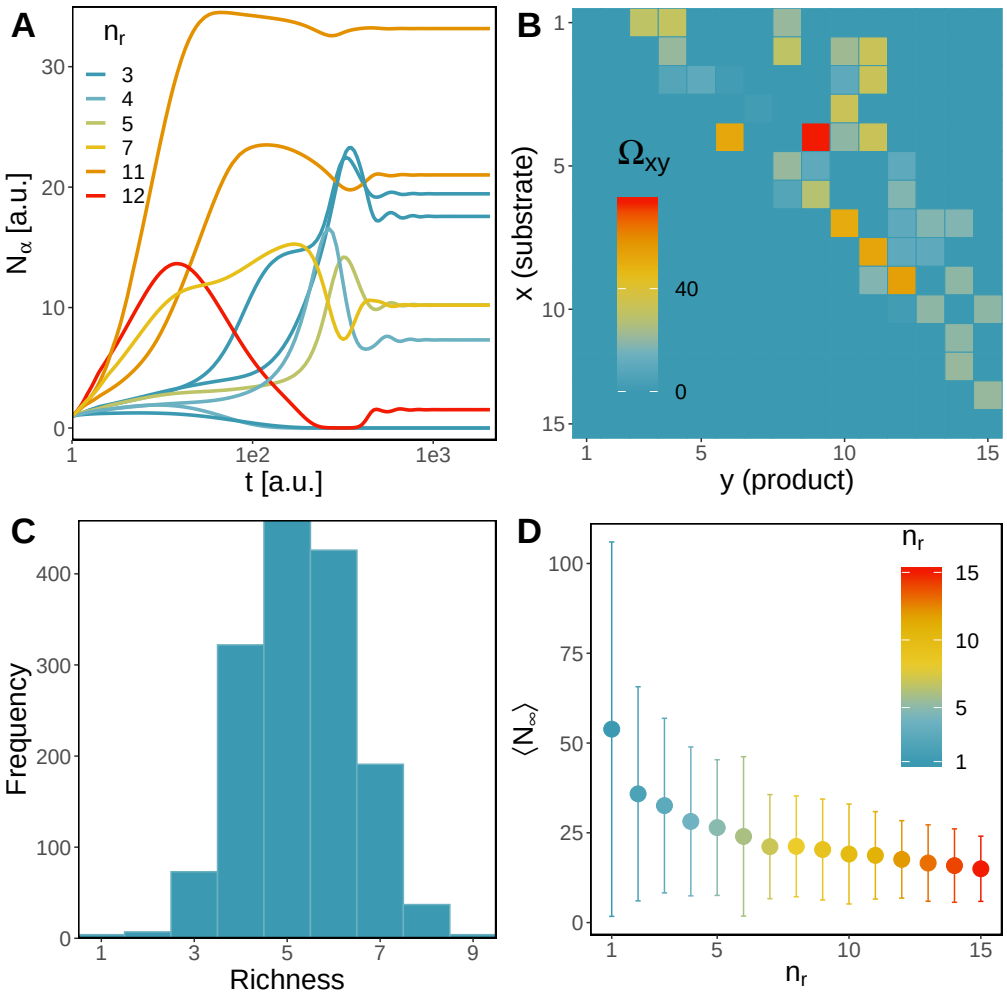

Figure 2: **Results from community assembly simulations**. Plots (**A**) and (**B**) exemplify one community assembly event and (**C**) and (**D**) convey results accross simulations. **A**) Time variation of species' abundance for one instance of community assembly with $m = 15$, $s = 10$, and a set 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 adjacency matrices of all species weighted by their respective carrying capacity: $\Omega = \sum_{k=1}^{s} N_\infty^k A_k$. (**C**) Histogram of richness of the n_s simulations. (**D**). Mean value of carrying capacity $\pm\sigma$ (error bars) against the number of reactions n_r: species with less number of reactions (specialists) are more abundant than those with higher n_r (generalists); several specialists deplete all resources through their combined action more efficiently than one generalist.

Second, the truncation limits in step 2 have been chosen to respect the hierarchical character of the network: $k \neq 0$ to avoid reactions of the form (i, i). Third, the upper limit of the uniform distribution from wich i is sampled is bounded by k, the diagonal we are sampling from.

When the sampling of reaction networks is completed, equations 6 and 3 are integrated using a Runge Kutta method (Dormand & Prince 1980) with initial conditions $N_\alpha(0) = 2$ and $C_\beta(0) = 0$. Relevant results steming from the simulations of community assembly events are plotted in figure 2

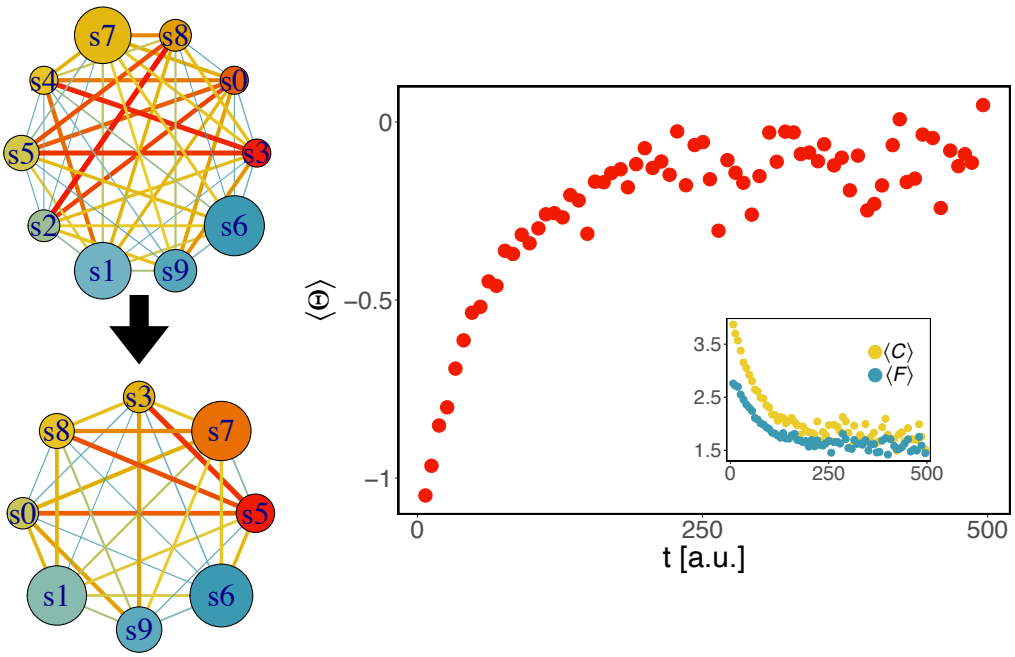

Figure 3: **Community cohesion time variation during community assembly. (left)** Cohesion coefficients for each pair of interacting species before and after community assembly. Thicker and red edges represesnt a higher cohesion coefficient. The nodes (species) are also arranged according to cohesion level, in anti-clockwise direction. The size of the node represents the number of reactions possesed by each strain, $s\alpha$. (**right**) Averaged community cohesion across all community assembly events as a function of time. Every time a species goes extinct during community assembly, the community cohesion is recalculated with the remaining species. The process of community assembly increases community cohesion. (**inset**) Averaged competition and facilitation indexes. Both decrease during community assembly, but competition decreases more. The decrease of facilitation wouldn't be expected, and may be atributed to the extinction of redundant functional groups.

3.3 A metric of community cohesion

Following the ensamble of many synthetic communities, I postulate a metric of community cohesion that is later used to predict the outcome of community coalescence.

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.

$$\xi(s_1, s_2) = \frac{1}{2} \sum_{k \in s_1 \cap s_2} \left(\frac{D_{s_1}(k)}{|s_1|} + \frac{D_{s_2}(k)}{|s_2|} \right) \tag{7}$$

Here, k takes the values in the set that results form intersecting s_1 and s_2. $D_s(k)$ is the number of elements from the sequence s that are equal to k. The purpose of all denominators in equation 7 is to normalize ξ to 1. Note that fractions inside the sumation term are equal only if the number of elements in both sequences s_1 and s_2 are the same.

The cohesion of a community – the degree to which species in a community colaborate with each other, can be thought of as the difference *cohesion = facilitation – competititon*. 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 of species i, y_i, with the sequence of substrates of species j, x_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 of species i, x_i and the sequence of substrates of species j, x_j. Thus,

$$f_{ij} = \begin{cases} \xi(y_i, x_j) & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \qquad c_{ij} = \begin{cases} \xi(x_i, x_j) & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \tag{8}$$

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.

Community-level cohesion can now be formally defined as the mean value of the matrix $F - C$

$$\Theta = \langle F - C \rangle \tag{9}$$

3.4 Θ predicts the outcome of community coalescence exeriments.

Following the community assembly and consequent meassure of cohesion in all communities, I perform community coalescence experiments to test the predictive power of my metric.

First, I select the N communities with 5 species, and perform all $\binom{N}{2}$ community coalescence events in which a resident community $\mathcal{C}_R$ is mixed with an invading one $\mathcal{C}_I$. At each event, calculate the similarity of between post-coalesecence and resident communities as the normalized scalar product of their species abundance vector at stable state.

$$S(\mathcal{C}_R, \mathcal{C}_P) = \frac{\vec{N_\infty^R} \cdot \vec{N_\infty^P}}{\sqrt{|\vec{N_\infty^R}|}\sqrt{|\vec{N_\infty^P}|}} \tag{10}$$

4 Discussion

- Although it certainly seems an exaggeration to view the communiity as such a tightly regimented entity, it is also perilous to ignore the fadct that coevolutionary processes can play an important role in communities.

- Cooperation reduces community stability, tho increases community fitness... (Coyte et al. 2015)

- This thesis addresses the question of what are truly the mechanisms explaining what experiments show? An alternative measure of cohesivenes that stems from a more realistic modeling of microbial ecosystems is able to reproduce this 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

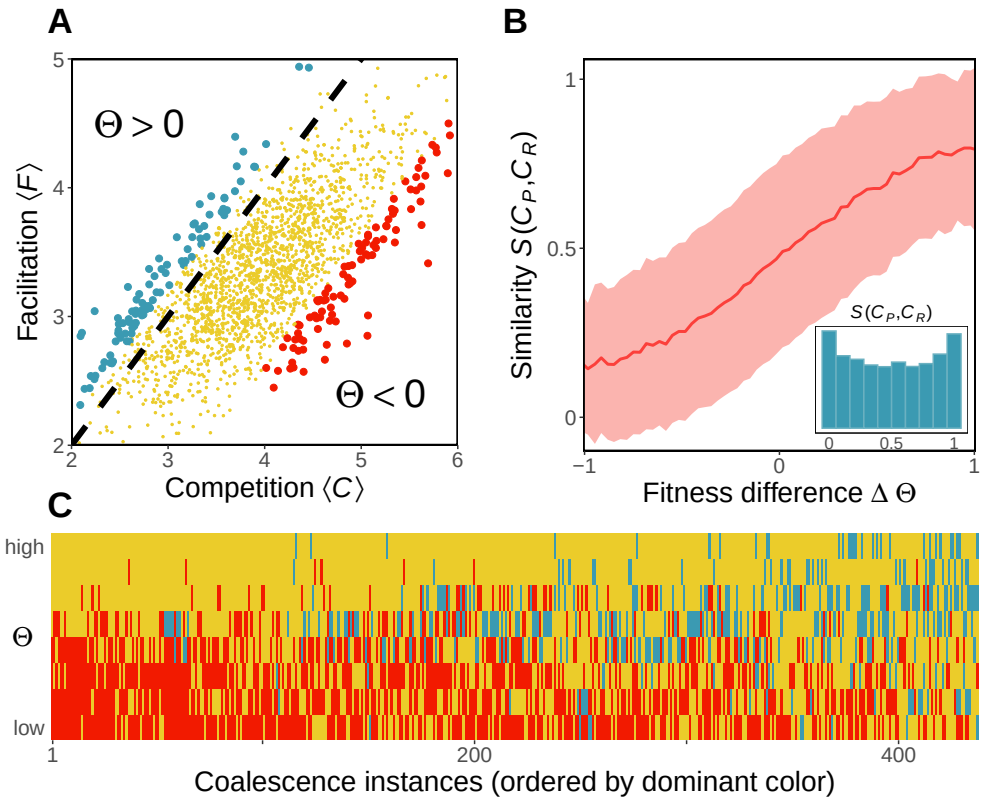

Figure 4: **Results from community coalescence experiments**. (**A**) Each simulated community is plotted in a competition-facilitation diagram. Communities above the dashed line $\langle C \rangle = \langle F \rangle$ have $\Theta > 0$; they are in the facilitation-dominated regime. Communities below the dashed line have $\Theta < 0$; they belong to the competition-dominated regime. The extremes of each regimes 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 ($S = 0.5$)

of microbial communities while asking questions about how these communities came to be.

- use Pascual-García et al. (2020) for the discussion
- discuss why I chose m as the upper limit for the number of reactions that a strain can posses.
- Discuss possible refinements of the measure of cohesion: instead of averaging, I could count the number of closed loops.
- 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.

5 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.
- plot mean abundance as a function of number of reactions for both assebled communities
- List of plots I want to make: 1. The s plot with richness == 5. 2. The histogram of similarity with richness == 4. 3. The elimination asssay. 4. The cloud. 5. The evolution of cooperation. 6. The community reaction network.
- Revise the cohesion of my thesis as a whole: are sections well separated? Do they link well with each other?
- When I talk about stable equilibrium, say that A stable (non- invadeable) equilibrium is characterized by an extra condition that all the absent species, if re-introduced, would be driven back to extinction
- Turn to my dictionary of cool words, and use them.
- In figure 3, why ¡F¿ - ¡C¿ $\neq$ ¡Θ¿? Or is it?

6 Appendix

6.1 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

$$\mu_\beta = E\left(1 - \sqrt{\frac{\beta - 1}{m - 1}}\right) \tag{11}$$

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 modeled 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.

$$E + S \rightleftharpoons ES \rightleftharpoons E + P \tag{12}$$

The choice of fully reverisble enzyme kinetics, instead of the traditional assumption of ireversibility 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 high, 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 tradi-

tional 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 12 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 13 can be found in Hoh & Cord-Ruwisch (2000)

$$q_{\alpha i} = \frac{q_m^{\alpha i} S_\alpha (1 - \theta_\alpha)}{K_S^{\alpha i} + S_\alpha (1 + k_R^{\alpha i} \theta_\alpha)} \tag{13}$$

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

$$\theta = \frac{[P]}{[S] K_{eq}} \tag{14}$$

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

$$K_{eq} = \exp\left(\frac{-\Delta G^0 - \eta \Delta G_{ATP}}{RT}\right) \tag{15}$$

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

$$\eta = \frac{y - x}{m} \tag{16}$$

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

6.2 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

References

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'.

Coyte, K. Z., Schluter, J. & Foster, K. R. (2015), 'The ecology of the microbiome: Networks, competition, and stability', *Science* .

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* .

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* .

Pascual-García, A., Bonhoeffer, S. & Bell, T. (2020), ‘, which can be found in subsection 3.3 (page 11).’.

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’.

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* .