

Supplementary Material for:
The role of competition versus cooperation in microbial
community coalescence

Pablo Lechón, Tom Clegg, Jacob Cook, Thomas P. Smith, Samraat Pawar

May 15, 2021

Contents

S1 Relationship with other microbial consumer-resource models	2
S2 Cost function	3
S3 Competition and Facilitation Metrics	3
S4 Modulating cohesion levels	6
S5 Adding consumer guild structure	8
S6 Further details of the community assembly simulations	10
S7 Further details of the community coalescence simulations	12
S8 Matrix representation of the model	12

S1 Relationship with other microbial consumer-resource models

Here we describe how our model is related to two other recent ones that have been used to study microbial community assembly and coalescence. The mapping between the notation used in Tikhonov (2016) (T), Marsland et al. (2019) (M) and those used here is provided in the following table:

Notation for...	M	Here	T
Species index	i	α	$\vec{\sigma}$
Species abundance	N_i	n_α	n_σ
Resource a species can harvest	\vec{c}_i	\vec{c}_α	σ_i
Resource supply	κ_α	κ_j	R_i
Dilution rate	τ_α	∞	NA
Minimal resource requirement	m_i	z_α	$\chi_{\vec{\sigma}}$
Resource weight	w_α	$\vec{1}$	$\vec{1}$
Resource \rightarrow biomass conversion factor	g_i	g_α	$(\tau_0 \chi_{\vec{\sigma}})^{-1}$
Leakage factor	l_α	l	0
Metabolic matrix	$D_{\alpha\beta}$	$(D_{kj})^T$	NA

With Marsland et al. (2019)'s Model

Our model differs from the version used in Marsland et al. (2019) in the following respects: (i) all resources contain the same amount of energy (taken to be 1 for simplicity), (ii) a type I functional response, (iii) binary consumer preferences, (iv) a shared core metabolism encoded in D , (v) a common leakage fractions for all species and resources, and (vi) a complex environment where all resources are externally supplied in equal amounts. We address the implications of these assumptions in the Discussion section.

1. All resources contain the same amount of energy ($\omega_j = 1$).
2. We only consider a type-1 functional response ($c(R_j) = R_j$).
3. Consumer preferences are binary, instead continuously distributed between 0 and 1.
4. Dilution rate is very high, such that $\tau^{-1} \approx 0$

With Tikhonov (2016)'s Model

Tikhonov (2016) did not explicitly model resource dynamics, but we can establish the relationship between that model and the one used here, as follows. Our consumer-resource model

is

$$\begin{aligned}\frac{dn_\alpha}{dt} &= g_\alpha n_\alpha \left((1-l) \sum_j c_{\alpha j} R_j - z_\alpha \right) \\ \frac{dR_j}{dt} &= \kappa_j - \sum_\alpha n_\alpha c_{\alpha j} R_j + l \sum_{\alpha k} n_\alpha D_{kj} c_{\alpha k} R_k.\end{aligned}\tag{S1}$$

In the limit when there is no leakage ($l = 0$), we can make the assumption that molecular (resource) dynamics are faster than population dynamics, and therefore that resource concentration R_j at any moment quickly equilibrates to reflect the instantaneous demand, i.e., $dR/dt \approx 0$. This allows us to separate the time scales between resource and population dynamics, recovering the model used by Tikhonov (2016):

$$\begin{aligned}\frac{dn_\alpha}{dt} &= g_\alpha N_\alpha \underbrace{\left(\sum_j c_{\alpha j} \frac{\kappa_j}{\sum_\alpha N_\alpha c_{\alpha j}} - z_\alpha \right)}_{\text{Resource surplus } \Delta_\alpha} \\ R_j &= \frac{\kappa_j}{\sum_\alpha N_\alpha c_{\alpha j}}\end{aligned}$$

S2 Cost function

The cost function used in our model (Eq S1) ensures that neither specialists or generalists are systematically favoured during community assembly, and corresponds to the assumption of approximate neutrality (Tikhonov 2016, Tikhonov & Monasson 2017, 2018). Species α will increment its abundance when its surplus term in Eq S1 is greater than 0;

$$\frac{1}{g_\alpha n_\alpha} \frac{dn_\alpha}{dt} = (1-l) \sum_j c_{\alpha j} R_j - z_\alpha > 0\tag{S2}$$

That is, species able to maintain positive growth for the lowest concentration of resources will be favoured. This can be achieved by either reducing the cost z_α or increasing the number of resource preferences in Eq S2. However, if one imposes the constraint $z_\alpha = \sum_j c_{\alpha j}$, the condition to maintain positive growth becomes

$$R_j - \chi_0 > 0$$

Which is independent of the number of preferences of the consumer; that is, species are neutral under this assumption. To break this degeneracy, and place our model in the regime of approximate neutrality, we introduce a small random fluctuation term, ϵ in Eq 2 of the main text.

S3 Competition and Facilitation Metrics

In the model system (Eq S1), competition for resources exists because the metabolic preferences vectors of all species in the community will generally overlap. We define intrinsic

competition between a pair of species (α, β) as the level of overlap between their metabolic preferences, independent of their environment's state. We calculate this metabolic overlap by counting the number of common preferences. Since these are binary, taking the scalar product of the two preference vectors yields the number of common elements between them,

$$(C_i)_{\alpha\beta} = \sum_k c_{\alpha k} c_{\beta k}.$$

This intrinsic competition becomes effective competition when we consider the environmental states experienced by the consumers, which change due to biotic and abiotic factors.

The resources available in the environment can either be externally supplied through κ or leaked by consumers through l and D . Depending on their origin, effective competition for resources is calculated differently. Consider the resource abundance landscape of a system as its dynamics unfold. Initially ($t = 0$) species are inoculated in an environment where all resources are externally supplied. As the dynamics play out, a fraction l of the externally supplied resource abundance vector is mapped into another vector of abundances through leakage according to the core metabolism D . When the system reaches equilibrium, both resource abundance profiles (the externally supplied and the one leaked by consumers) coexist, so effective competition (henceforth, just ‘competition’) needs to be calculated differently for each resource profile. We calculate the competition between a given pair of species on abiotically vs biotically generated resources as follows.

The competition for each abiotically-generated resource is weighted by the normalized resource supply

$$\tilde{\kappa} = \vec{\kappa}/\max(\vec{\kappa}),$$

with the interaction strength given by the non-leaked fraction $1 - l$ (fraction of any resource that is effectively consumed abiotically). Thus, the overall competition C_a between the species pair (α, β) for abiotically-generated resources must be

$$(C_a)_{\alpha\beta} = (1 - l) \sum_k \tilde{\kappa}_k c_{\alpha k} c_{\beta k}. \quad (\text{S3})$$

Now consider the competition between species α and β for biotically-generated resources (which have been leaked to the environment), C_b . There will only be competition for the k^{th} biotically-generated resource when both the following two conditions hold: 1) α or β consume resource j and leak a fraction of it in the form of resource k ; 2) α and β both consume k . Mathematically, this combination of conditions implies:

$$(C_b)_{\alpha\beta}^{k \rightarrow j} \propto D_{jk} (c_{\alpha j} + c_{\beta j}) c_{\alpha k} c_{\beta k}. \quad (\text{S4})$$

In the case of abiotic competition, it was straightforward to see that the interaction strength of abiotic competition is the consumption fraction $1 - l$. Intuitively, one would think that l is the interaction strength for competition on biotically-generated resources. Formally, we can think of the harvest-leaking process as a infinite discrete succession of consumption and leakage cycles. When a species harvests resource j from the environment, it intakes a fraction $1 - l$ and leaks back a fraction l in the form of other resources (first two arrows in Fig S1). The leaked resources become part of the available resource pool, and are competed

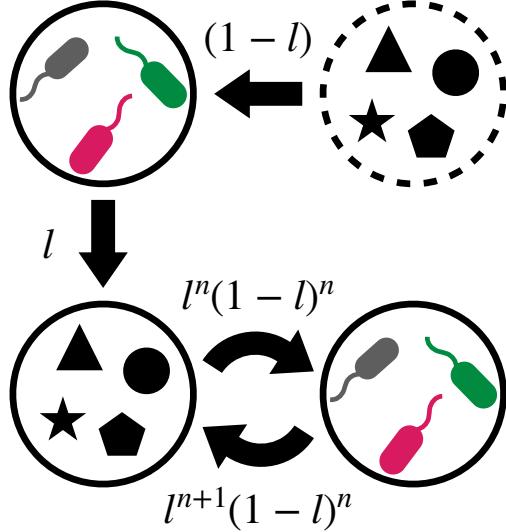


Figure S1: Mechanism of resource recycling. The species in this model consume the resources that come through the supply rate κ (abiotic uptake, top arrow) and leak a fraction l back to the environment (left arrow). The leaked resources can be harvested again by the species in the community (biotic uptake, bottom arrows). This can be modeled as an infinite series of consecutive cycles of uptake and leakage where the amount of resources consumed by each strain decreases by a factor of l , after each cycle.

for in a second cycle of consumption, with intake fraction $l(1 - l)$. This process extends up to n cycles, (with $n \rightarrow \infty$), as illustrated in Fig S1, so that at cycle n , the fraction of harvested resource is $l^n(1 - l)$. The strength of the biotic competition and facilitation links, B , is calculated by summing the fraction of ingested resource over all cycles as

$$B = \sum_{n=1}^{\infty} l^n(1 - l) = (1 - l) \left(\sum_{n=0}^{\infty} l^n - 1 \right) = (1 - l) \left(\frac{1}{1 - l} - 1 \right) = l.$$

As expected, the strength factor for biotic links is l , the fraction of resources originally leaked. Note that this result depends on the convergence of the geometric series, which is only true when $l < 1$. In the trivial case $l = 1$, then $B = 0$, and the system would have neither competitive nor facilitative interactions.

Then, the total competition between the species pair (α, β) for biotically-generated resources is the sum of the terms that arise from expression S4 when spanning the full resource set, weighted by l :

$$(C_b)_{\alpha\beta} = l \sum_{jk} \tilde{\kappa}_j D_{jk} (c_{\alpha j} + c_{\beta j}) c_{\alpha k} c_{\beta k}. \quad (\text{S5})$$

Note that the factor $\tilde{\kappa}_j$ is necessary in both measures of competition (equations S3 and S5) because all resources are externally-supplied when they first enter the system, and so we need to account for possible differences in the external resource supply rate.

The two types of competition presented here, are merely a formal classification, since consumers cannot tell apart biotically and abiotically-generated resources. Therefore, the

total competition between species pair (α, β) is the sum of the two terms

$$(C_T)_{\alpha\beta} = (C_a)_{\alpha\beta} + (C_b)_{\alpha\beta}. \quad (\text{S6})$$

Next, we consider facilitation. Facilitation links form when a species leaks resources (metabolic by-products) that are used by another species. Therefore, the facilitation of species α towards species β when the former consumes resource j and releases k will be non-zero when 1) species α produces k as a metabolic by-product when consuming j , and 2) species β consumes k . Analogous to the case of competition for biotically-generated resources (Eq S5), the strength of these interactions is also given by l and the factor $\tilde{\kappa}_j$. Combining these conditions, the total facilitation of species $\alpha \rightarrow \beta$ will be

$$F_{\alpha\beta} = l \sum_{jk} \tilde{\kappa}_j c_{\alpha j} D_{jk} c_{\beta k}. \quad (\text{S7})$$

Note that facilitation is directional, which implies that the community facilitation matrix $F_{\alpha\beta}$ is not symmetric. On the other hand, the community competition matrix $(C_T)_{\alpha\beta}$ is symmetric, since competition is not directional.

Averaging the matrices of pairwise metrics S6 and S7 yields a proxy of the community-level competition and facilitation. The matrix diagonals are left out of this averaging because we are only interested in inter-specific competition and facilitation. Additionally, we only consider the upper diagonal elements of the symmetric competition matrix. Then, the community-level competition and facilitation measures are given by

$$\mathcal{C} = \langle C_{T_{\alpha\beta}} \rangle_{\alpha \neq \beta} \quad \& \quad \mathcal{F} = \langle F_{\alpha\beta} \rangle_{\alpha \neq \beta},$$

respectively.

S4 Modulating cohesion levels

We are interested in generating communities spanning a broad range of cohesion levels, i.e., different competition and facilitation levels. These can be modified by sampling $c_{\alpha j}$ and D_{jk} respectively with certain constraints.

Modulating competition level

First, let us consider the problem of modulating the competition level of the community by sampling the metabolic preferences for each strain, encoded in $c_{\alpha j}$. Each species α has a binary vector \vec{c}_α of length m that specifies if resource j is used ($c_{\alpha j} = 1$) or not ($c_{\alpha j} = 0$). The metabolic preference matrix C is constructed by sampling its rows \vec{c}_α sequentially. The process of sampling \vec{c}_α is two-fold. We first sample m_α , the number of resources that species α uses, from an exponential distribution. This choice is supported by experimental evidence (Sung et al. 2017). Second, to determine which resources are used by species α , we sample m_α resources with probability vector \vec{p}_α . Note that in this sampling scheme, ‘iteration number’ and ‘species’ are equivalent, and denoted by index α .

For species $\alpha = 1$ all resources have the same probability of being sampled $1/m$, where m is

the total number of resources. This assumption is consequent with the absence of a resource hierarchy, since they all carry the same energy. After each iteration, the sampling probability of each resource changes according to what has been sampled previously. Let $d_{\alpha j}$, denote the cumulative demand of resource j when the metabolic preferences of species α are sampled. That is, the number of consumers of resource j at iteration α .

$$d_{\alpha j} = \sum_{i=1}^{\alpha} c_{ij}.$$

Based on $d_{\alpha j}$, I then compute the probability that species α is assigned resource j as one of its preferences

$$p_{\alpha j} = (1 - k_c) \frac{1}{m} + k_c \frac{d_{\alpha-1j}}{\sum_j d_{\alpha-1j}}, \quad (\text{S8})$$

where the denominator represents the total number of preferences sampled up until iteration $\alpha - 1$, and acts as a normalization constant, and together with the numerator represents the normalized cumulative demand (Eq 7 main text),

$$\tilde{d}_{\alpha-1j} = \frac{d_{\alpha-1j}}{\sum_j d_{\alpha-1j}} \quad (\text{S9})$$

The strength with which $p_{\alpha j}$ deviates from a uniform distribution is given by the parameter $k_c \in [0, 1]$, the competitiveness factor. When $k_c \rightarrow 1$, competition is maximized. Thus, highly demanded resources are more likely to be sampled in the next iteration. On the contrary, when $k_c = 0$ competition is random, since each resource is equally likely to be chosen.

The metabolic preferences sampling procedure is implemented in the following algorithm

Algorithm 1: Sampling of metabolic preferences

```

for  $\alpha \in \{1, \dots, s\}$  do
    Sample  $m_\alpha$  from an exponential distribution
    Sample vector  $\vec{v}$  of  $m_\alpha$  integers  $\in \{1, \dots, m\}$  with probability vector  $\vec{p}(\alpha)$ 
    Switch on sampled preferences  $\vec{c}_\alpha[\vec{v}] = 1$ 
    Update  $\vec{d}_\alpha$ 
    Update  $\vec{p}_\alpha$  using the new  $\vec{d}_\alpha$ 
end

```

Modulating facilitation level

Second, consider the problem of modulating the facilitation level of the community by sampling the metabolic cross-feeding topology encoded in the community metabolic matrix D . Each element of this matrix, D_{jk} , specifies the fraction of leaked energy from resource j that is released in the form of resource k . Note that by definition $\sum_j D_{jk} = 1$.

To broaden the range of facilitation levels that a community can achieve, we use a concentration parameter vector \vec{q}_k , whose elements vary proportionally to the demand of each resource d_j and the cooperation factor k_f as

$$q_{jk} = \frac{1}{u}(1 + k_f d_j). \quad (\text{S10})$$

Thus, k_f sets the amount of structure of D . When $k_f \rightarrow 0$ the metabolic matrix has no structure. All elements of the concentration parameter are the same (flat Dirichlet distribution), and therefore all resources are released at equiprobable fractions. As $k_f \rightarrow 1$ the structure of D becomes fully determined by the resource demands of the community, so that more coveted resources are released at higher fractions. The factor u in Eq S10 controls the sparsity of the metabolic network, ranging from a fully connected network when $u \rightarrow 0$ to a sparse one-to-one network when $u \rightarrow 1$.

Note that although the two methods for sampling c and D share similarities, they are conceptually different. First, the sampling of c is fully random, in the sense that a vector of probabilities is constructed first, and then preferences are randomly sampled from that distribution. On the other hand, the sampling of D has a random term, x_{jk} , and a non-random term that depends only on the vector of demands of the community \vec{d}_s . Another difference is that the competitiveness factor is imposed on the preference sampling probability vector, while the facilitation factor is imposed directly on the values of D . Finally, the sampling of D depends on the sampling of c , reflecting that high or low facilitation levels are achieved by tuning the secretion structure of the community to be symmetric to the profile of demands, or independent from it, respectively.

S5 Adding consumer guild structure

In order to further modulate competition levels of communities on biotically-generated resources, we add structure to the preference and metabolic matrices in the form of resource classes. As a result, we now have two levels of structure; inter-species, intra-guild facilitation due to preferential feeding (previous section), and inter-guild facilitation due to the existence of preferred resource classes. This allows us to decouple the terms C_b and F so that their respective effects can be correctly quantified, while keeping the levels of facilitation and competition adjustable through the constants k_c and k_f .

The algorithm that we use to construct to sample the metabolic preferences matrix C is analogous to that used in previous section. The only difference is that the probability $p_{\alpha j}$ that species α is assigned resource j as one of its preferences has a different form, which we now derive. Our aim is to write a normalized piece-wise function out of Eq S8, where each piece is weighted more or less depending on resource j being preferred by species α or not, respectively. Therefore, we can assume the form of this function to be

$$p_{\alpha j}^A = \begin{cases} M \left((1 - k_c) \frac{1}{m} + k_c \frac{d_{\alpha-1j}}{\sum_j d_{\alpha-1j}} \right) (1 + K_c) & \text{if } j \in A \\ \frac{N}{m - n_c} (1 - K_c) & \text{otherwise,} \end{cases} \quad (\text{S11})$$

where M and N are normalization constants that ensure $\sum_j p_{\alpha j} = 1$. Note that K_c modulates the amount of structure in the matrix, ranging from no structure when $K_c = 0$ to maximum guild structure when $K_c = 1$. In order to obtain expressions for the normalization constants M and N we impose the following constrains on each piece of Eq S11

$$p_\alpha^1 = \sum_{C(j) \in T} p_{\alpha j} = \frac{n_c}{m} (1 - K_c) + K_c \quad (\text{S12})$$

and

$$p_\alpha^0 = \sum_{C(j) \notin T} p_{\alpha j} = \left(1 - \frac{n_c}{m}\right) (1 - K_c). \quad (\text{S13})$$

These two constraints guarantee that two necessary conditions are satisfied; (1) the total probability sums to one, since $p_\alpha^1 + p_\alpha^0 = 1$, and (2) probability of sampling resources in (outside) the preferred class approaches 1 (0) when K_c increases (decreases). We then solve for M and N by expanding equations S12 and S13 using the definition in expression S11 for $p_{\alpha j}$

$$M \left((1 - k_c) \frac{n_c}{m} + k_c \frac{T_c}{T} \right) (1 + K_c) = \frac{n_c}{m} (1 - K_c) + K_c$$

$$M = \frac{K_c + \frac{n_c}{m} (1 - K_c)}{(K_c + 1) \left(\frac{n_c}{m} (1 - k_c) + \frac{T_c}{T} k_c \right)}$$

and

$$N = 1 - \frac{n_c}{m},$$

where $T = \sum_j d_{\alpha-1j}$, and we have used the following expressions

$$\sum_{C(j) \in T} 1 = n_c \quad \sum_{C(j) \notin T} 1 = m - n_c \quad \sum_{C(j) \in T} d_{\alpha-1j} = T_c.$$

Thus, the closed form of the sampling probability under the general scenario (Eq 7 main text) is

$$p_{\alpha j} = \begin{cases} \left(K_c + \frac{n_c}{m} (1 - K_c) \right) \frac{\left((1 - k_c) \frac{1}{m} + k_c \frac{d_{\alpha-1j}}{T} \right)}{\left((1 - k_c) \frac{n_c}{m} + k_c \frac{T_c}{T} \right)} & \text{if } j \in A \\ \frac{1}{m} (1 - K_c) & \text{otherwise.} \end{cases} \quad (\text{S14})$$

To also encode this guild structure in D , we sample each row from a Dirichlet distribution with concentration parameters q_{jk}

$$D_{jk} = \text{Dir}(q_{j1}, q_{j2}, \dots, q_{jm})_k,$$

where the concentration parameter depends on the cumulative demand as specified in the previous section (see Eq S10). Additionally, the value of q_{jk} decreases with the inter-guild facilitation factor K_f if uptaken and leaked resources belong to the same (resource) class, and increases with K_f in the opposite case. With these conditions, the expression for the concentration parameter is

$$q_{jk} = \begin{cases} \frac{(1 + k_f d_j)}{u M_{C(j)}} (1 - K_f) & \text{if } A(j) = A(k) \\ \frac{(1 + k_f d_j)}{u(M - M_{C(j)})} (1 + K_f) & \text{otherwise.} \end{cases} \quad (\text{S15})$$

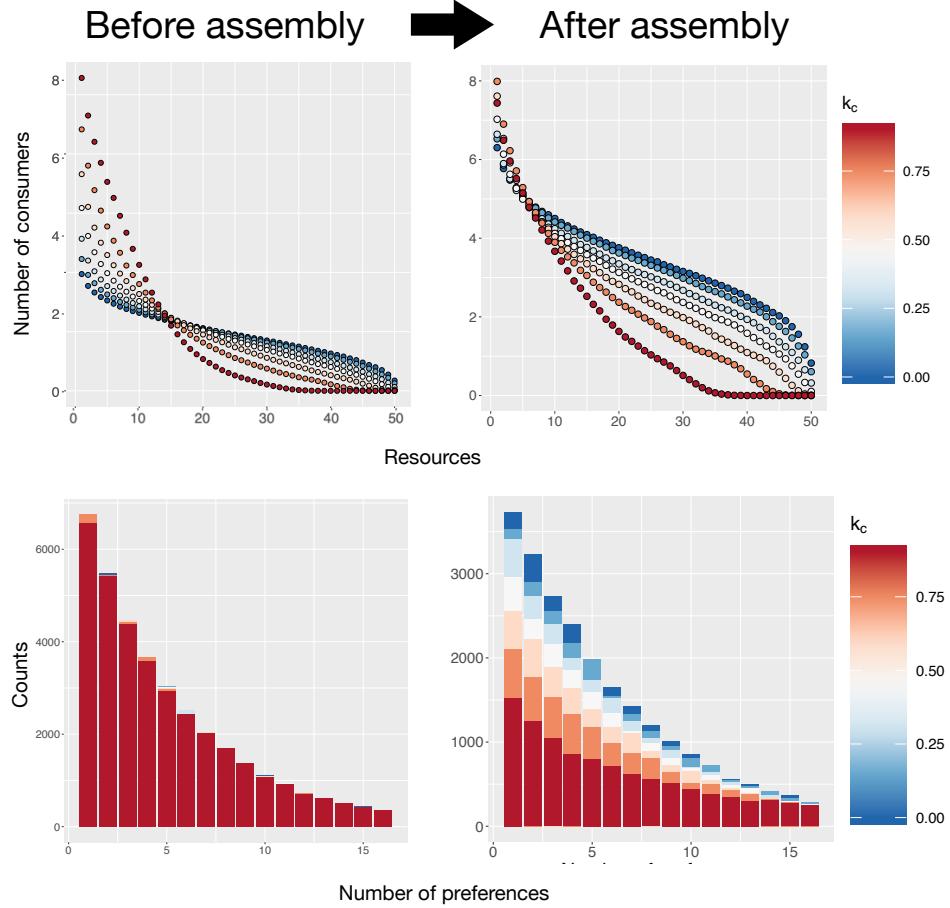


Figure S2: **Top row:** number of consumers per resource (arbitrarily labelled before and after community assembly, for increasing value of the competitiveness factor. Note that as k_c increases, the difference between the most demanded resource and the second-most demanded resource increases, for both the before, and after assembly cases. **Bottom row:** Histogram of number of consumers (y-axis) with n_c (x-axis) number of preferences. We find that the trend does not change before and after assembly, since in both plots it the probability to posses n_c preferences decreases as n_c increases to have a higher amount of resource preferences.

Here, u is the sparsity of the metabolic network, ranging from a fully connected network when $u \rightarrow 0$ to a sparse one-to-one network when $u \rightarrow 1$. M_A is the number of consumers in class A , to which resource j belongs. Note that in expressions S14 and S15 if we make $K_c = K_f = 0$, all the resources belong to the same class, and we recover equations S8 and S10 from the previous version.

S6 Further details of the community assembly simulations

We perform two sets of community assembly simulations. First, under preferential feeding, and second with resource guild structure added. Within each set, we perform 100 random

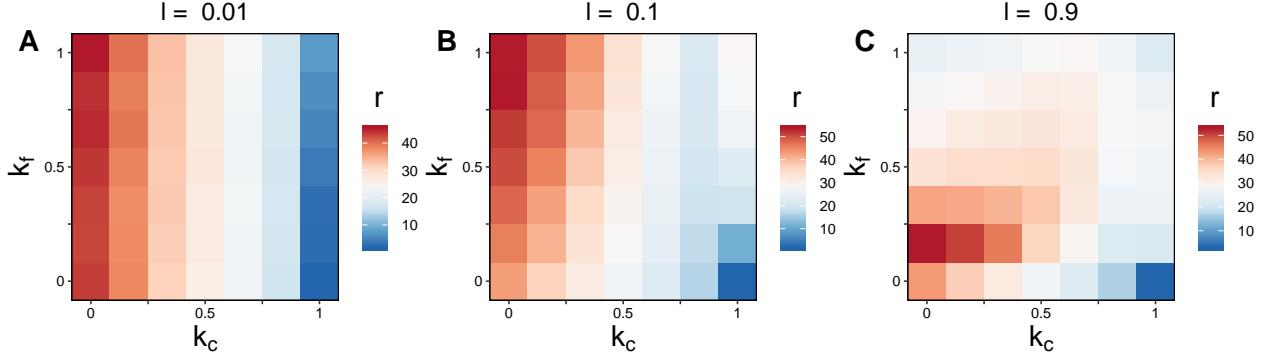


Figure S3: Heat-maps for each value of l of the average richness r as a function of k_c and k_f . **A:** When leakage is negligible, average species richness decreases as k_c increases, but does not change with k_f . That is, leakage is so low that the effect of the level of facilitation on the final community richness is negligible. **B:** At intermediate leakage level, the average richness is maximized where competition is lowest and facilitation highest, and minimized under the exact opposite set of conditions. **C:** At very high level of leakage, highest richness is reached in communities with low competition and facilitation levels.

assembly processes at every point of the parameter grid, which encompasses all possible combinations of the following parameter values $s = m = 60$, $l = k_c = k_f = [0.1, \dots, 0.9]$. $K_c = K_f = 0$, in the first set, and $\in [0.1, \dots, 0.9]$ in the second one. We then use the assembled communities to perform *in silico* community invasion assays.

We also need to ensure that these non-random structures remain in place after the communities have been assembled. To this end, we plot the number of consumers per resource, before and after assembly, as a function of k_c , and also the histograms of the number of metabolic preferences before and after assembly for each value of k_c . As can be seen in Fig S2, the constraints on resource preferences, although slightly modified, remains in place. The difference in the number of consumers per resource is much higher for high k_c than it is for lower values of k_c , corresponding with what we imposed before assembly. Likewise, the exponential distribution that we imposed on the number of preferences per consumer is also maintained after assembly, with slight differences in the final overall species richness values.

Interestingly, after assembling the parent communities, we find that each value of leakage represents a different regime, resulting in different effects on species richness. The result in Fig S3C may seem particularly counter-intuitive, as one would expect that high facilitation promotes species diversity. However, the species in this communities are not efficient at consuming resources, since they leak the majority of what they harvest. Thus, in order to most efficiently deplete the most efficiently available resources, they need to perform more cycles of consumption than a species with a lower leakage factor. The optimal interaction topology that ensures efficient resource depletion in several cycles of consumption and leakage is one that minimizes facilitation and competition. To sum up; predictably, the three figures show that biodiversity is favoured by niche separation, since less competitive communities are more diverse. Surprisingly, the benefit of cooperative interactions changes for each case. For individuals in very low leakage communities, cooperative interactions are irrelevant. These become beneficial as leakage levels become moderate. However, they are detrimental, and

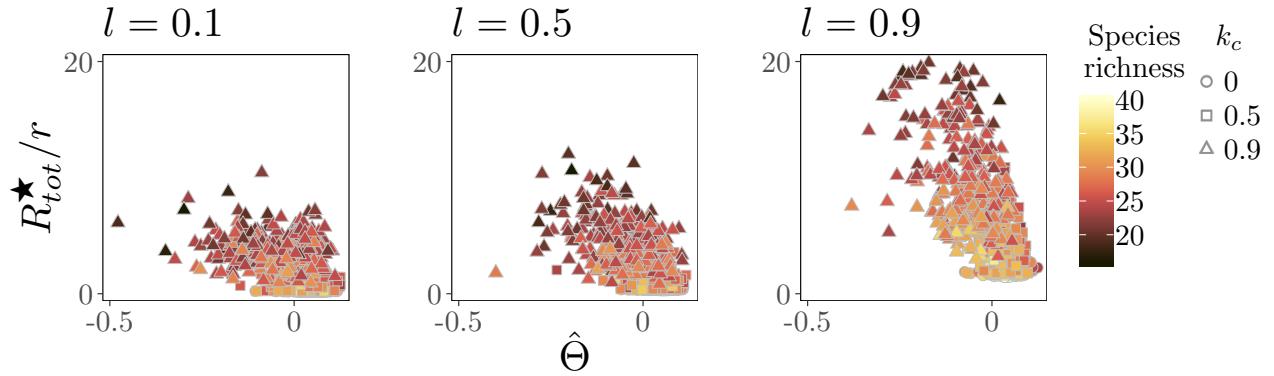


Figure S4: Resource depletion the under consumer guilds structured scenario

therefore minimized in situations where leakage is high.

Finally, we show the same plot as in Fig 3D, in for the scenario with guild structure. In this case, a negative correlation between resource depletion and cohesion is also seen, and it is higher the lower the leakage value is.

S7 Further details of the community coalescence simulations

In Fig S5 we can see how competition dominates even at high leakage values (top row). This is because facilitation has a weaker effect that only emerges if competition is very low, like in the case of the structured scenario with high leakage (bottom row). This is not the case of the top row, where, even tho $F > C$, competition is not negligible, so the effects of facilitation are not apparent.

Plots of each part of the cohesion measure for both the unstructured and structured scenarios and all leakage values can be found at the end of the supplement. In those figures it is clear that competition and facilitation are both negatively correlated with coalescence success, and that the best community upon a coalescence event is that which minimizes all interactions.

S8 Matrix representation of the model

Here we describe the matrix form of the model system (Eq S1), useful for efficient (vectorized) simulations. Eqs S1 can be expressed in matrix form as:

$$\begin{aligned}\frac{d\mathbf{n}}{dt} &= \mathbf{g} \circ \mathbf{n} (\text{diag}(\mathbf{R})C(\mathbf{1} - \mathbf{l}) - \mathbf{z}) \\ \frac{d\mathbf{R}}{dt} &= \boldsymbol{\kappa} - \text{diag}(\mathbf{R})C^T\mathbf{n} + D^T \text{diag}(\mathbf{l} \circ \mathbf{R})C^T\mathbf{n}.\end{aligned}$$

Here, $\mathbf{1}$ is a column vector of ones of appropriate dimension, and \circ denotes an element-wise operation. The interaction metrics presented above can also be vectorized. In the following

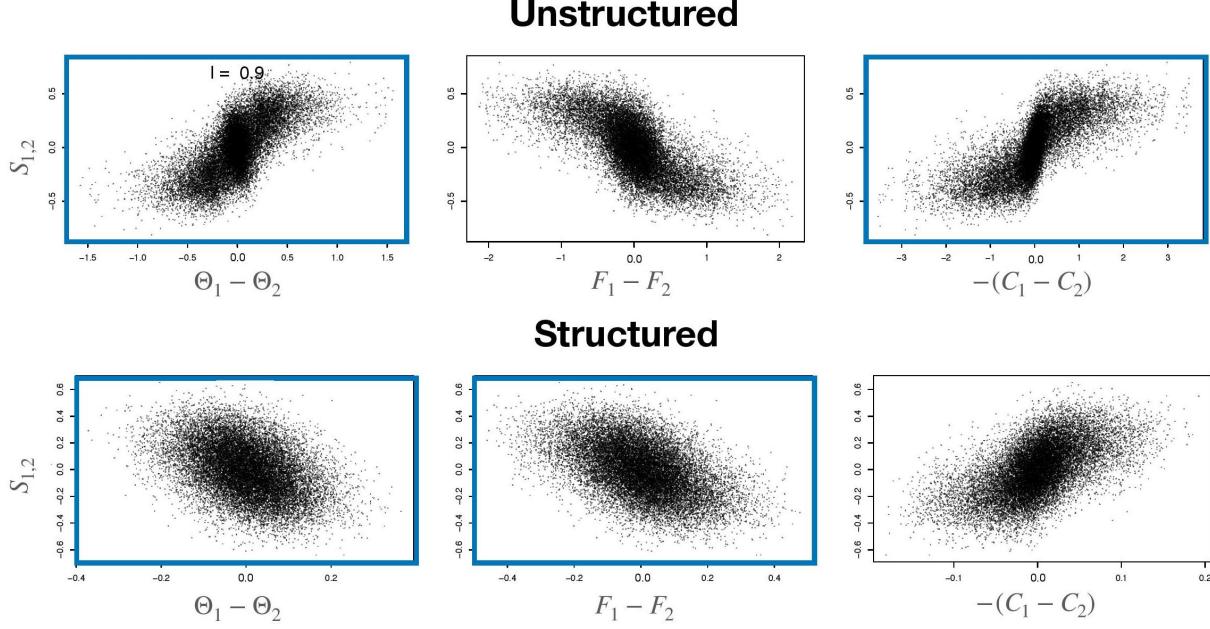


Figure S5: Plots of each component of the cohesion measure for both the unstructured and structured scenarios. Here, leakage is always 0.9, and x-axis is different across columns (cohesion difference, facilitation difference, and competition difference). Note that the x-axis of each column follows the cohesion equation i.e., $\Theta_1 - \Theta_2 = F_1 - F_2 - (C_1 - C_2)$. Whenever competition is present, it drives the dynamics of the system (top row, unstructured scenario). The effects of facilitation are only apparent when competition is negligible (bottom bottom row, structured scenario).

we use that the factor $\tilde{\kappa}_j = 1$ because in our simulations, all the resources are being supplied in the same amount. For instance, in the equations of competition, since metabolic preferences are binary, taking the scalar product of the two preference vectors yields the number of common elements between them as

$$C_a = (1 - l)CC^T.$$

For Eq S5, only indices j and k can be vectorized, taking the form

$$(C_b)_{\alpha\beta} = l(\vec{c}_\alpha + \vec{c}_\beta)^T D (\vec{c}_\alpha \circ \vec{c}_\beta).$$

The vectorization of Eq S7 is expressed as

$$F = lCDC^T.$$

Matrix representation at coalescence event

Each coalescence event is simulated by mixing the species from each community after the assembly process in isolation has been completed. This can be mathematically expressed in matrix form through a system of $s_1 + s_2 + m$ differential equations, where s_1 and s_2 are

the number of species that remain in the first and second communities after their assembly, respectively:

$$\frac{d\mathbf{n}_e}{dt} = \mathbf{g} \circ \mathbf{n}_e (\text{diag}(\mathbf{R}_e)C_e(1 - \mathbf{l}_e) - \mathbf{m}_e) \quad (\text{S16})$$

$$\frac{d\mathbf{R}}{dt} = \boldsymbol{\kappa} - [\mathbb{1} \ \mathbb{1}] \left(\text{diag}(\mathbf{R}_e)C_e^T \mathbf{n}_e + D_e^T \text{diag}(\mathbf{l}_e \circ \mathbf{R}_e)C_e^T \mathbf{n}_e \right) \quad (\text{S17})$$

Here $[\mathbb{1} \ \mathbb{1}]$ represents the horizontal concatenation of two $m \times m$ identity matrices. The sub-index e stands for *extended*. To form any of the extended vectors in equations S16 and S17, one simply vertically concatenates the vectors from community 1 and community 2. Constructing an extended matrix in Eq S16 and S17 is done by joining the two matrices alongside their diagonal. For example, constructing the vector of species abundances \mathbf{n}_e and the metabolic matrix D_e would be done as

$$\mathbf{n}_e = \begin{bmatrix} \mathbf{n}^{(1)} \\ \mathbf{n}^{(2)} \end{bmatrix} \quad D_e = \begin{bmatrix} D^{(1)} & 0 \\ 0 & D^{(2)} \end{bmatrix},$$

where the superscripts indicate membership of community (1 or 2).

References

- 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* **15**(2), e1006793. doi: 10.1371/journal.pcbi.1006793.
- Sung, J., Kim, S., Cabatbat, J. J. T., Jang, S., Jin, Y. S., Jung, G. Y., Chia, N. & Kim, P. J. (2017), ‘Global metabolic interaction network of the human gut microbiota for context-specific community-scale analysis’, *Nature Communications* **8**, 1–12.
URL: <http://dx.doi.org/10.1038/ncomms15393>
- Tikhonov, M. (2016), ‘Community-level cohesion without cooperation’, *eLife* **5**, e15747. doi: 10.7554/eLife.15747.
- Tikhonov, M. & Monasson, R. (2017), ‘Collective Phase in Resource Competition in a Highly Diverse Ecosystem’, *Physical Review Letters* **118**, doi: 10.1103/PhysRevLett.118.048103.
- Tikhonov, M. & Monasson, R. (2018), ‘Innovation Rather than Improvement: A Solvable High-Dimensional Model Highlights the Limitations of Scalar Fitness’, *Journal of Statistical Physics* **172**(1), 74–104.
URL: <https://doi.org/10.1007/s10955-018-1956-6>

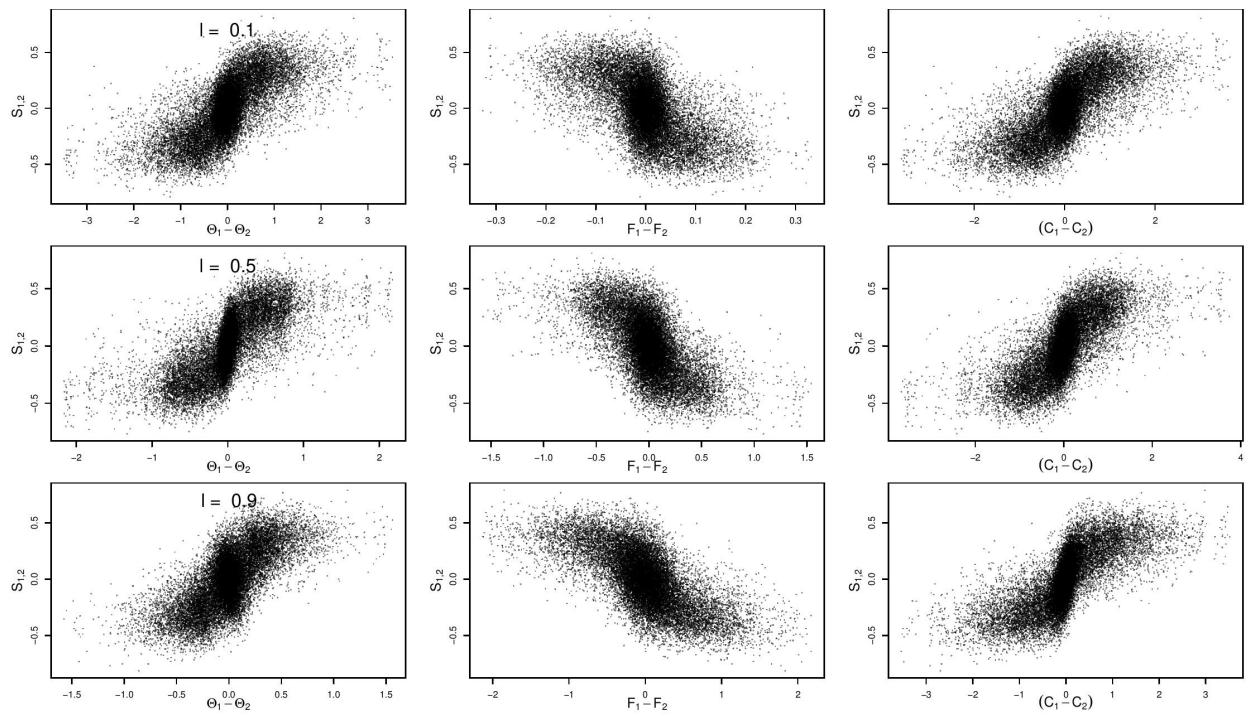


Figure S6: Unstructured scenario.

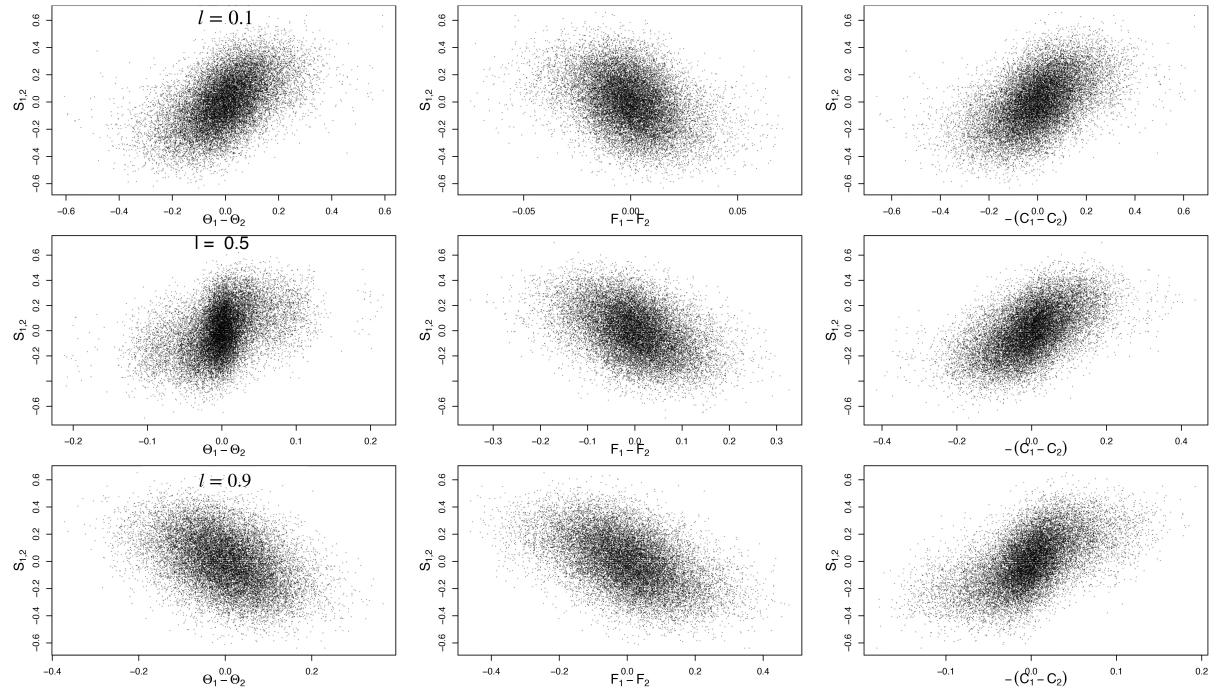


Figure S7: Structured scenario.