

The role of competition versus cooperation in microbial community coalescence

--Manuscript Draft--

Manuscript Number:	PCOMPBIOL-D-21-00903R1
Full Title:	The role of competition versus cooperation in microbial community coalescence
Short Title:	The role of competition versus cooperation in microbial community coalescence
Article Type:	Research Article
Keywords:	microbial ecology; community coalescence; species interactions
Corresponding Author:	Pablo Lechón-Alonso University of Chicago Department of Ecology and Evolution Chicago, UNITED STATES
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University of Chicago Department of Ecology and Evolution
Corresponding Author's Secondary Institution:	
First Author:	Pablo Lechón-Alonso
First Author Secondary Information:	
Order of Authors:	Pablo Lechón-Alonso Tom Clegg Jacob Cook Thomas P. Smith Samraat Pawar
Order of Authors Secondary Information:	
Abstract:	New microbial communities often arise through the mixing of two or more separately assembled parent communities, a phenomenon that has been termed ``community coalescence''. Understanding how the interaction structures of complex parent communities determine the outcomes of coalescence events is an important challenge. While recent work has begun to elucidate the role of competition in coalescence, that of cooperation, a key interaction type commonly seen in microbial communities, is still largely unknown. Here, using a general consumer-resource model, we study the combined effects of competitive and cooperative interactions on the outcomes of coalescence events. In order to do so, we simulate coalescence events between pairs of communities with different degrees of competition for shared carbon resources and cooperation through cross-feeding on leaked metabolic by-products (facilitation). We also study how structural and functional properties of post-coalescence communities evolve when they are subjected to repeated coalescence events. We find that in coalescence events, the less competitive and more cooperative parent communities contribute a higher proportion of species to the new community, because this endows superior ability to deplete resources and resist invasions. Consequently, when a community is subjected to repeated coalescence events, it gradually evolves towards being less competitive and more cooperative, as well as more species rich, robust and efficient in resource use. Encounters between microbial communities are becoming increasingly frequent as a result of anthropogenic environmental change, and there is great interest in how the coalescence of microbial communities affects environmental and human health. Our study provides new insights into the mechanisms behind microbial community coalescence, and a framework to predict outcomes based on the interaction structures of parent communities.
Suggested Reviewers:	Robert Marsland MIT: Massachusetts Institute of Technology marsland@alum.mit.edu

	<p>His work on modeling microbial communities partly underpins the work done here</p> <p>Sylvie Estrela Yale University sylvie.estrela@yale.edu</p> <p>Alberto Pascual-Garcia ETH Zurich: Eidgenossische Technische Hochschule Zurich alberto.pascual@env.ethz.ch</p> <p>Daniel Amor MIT: Massachusetts Institute of Technology damor@mit.edu</p>
Opposed Reviewers:	
Additional Information:	
Question	Response
Financial Disclosure	The author(s) received no specific funding for this work.
Enter a financial disclosure statement that describes the sources of funding for the work included in this submission. Review the submission guidelines for detailed requirements. View published research articles from PLOS Computational Biology for specific examples.	
This statement is required for submission and will appear in the published article if the submission is accepted. Please make sure it is accurate.	
<p>Unfunded studies</p> <p>Enter: <i>The author(s) received no specific funding for this work.</i></p>	
<p>Funded studies</p> <p>Enter a statement with the following details:</p> <ul style="list-style-type: none"> • Initials of the authors who received each award • Grant numbers awarded to each author • The full name of each funder • URL of each funder website • Did the sponsors or funders play any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript? • NO - Include this sentence at the end of your statement: <i>The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.</i> • YES - Specify the role(s) played. 	

* typeset	
Competing Interests	The authors have declared that no competing interests exist.
Use the instructions below to enter a competing interest statement for this submission. On behalf of all authors, disclose any competing interests that could be perceived to bias this work—acknowledging all financial support and any other relevant financial or non-financial competing interests.	
This statement will appear in the published article if the submission is accepted. Please make sure it is accurate. View published research articles from PLOS Computational Biology for specific examples.	
NO authors have competing interests	
Enter: <i>The authors have declared that no competing interests exist.</i>	
Authors with competing interests	
Enter competing interest details beginning with this statement:	
<i>I have read the journal's policy and the authors of this manuscript have the following competing interests: [insert competing interests here]</i>	
* typeset	
This statement is required for submission and will appear in the published article if the submission is accepted. Please make sure it is accurate and that any funding sources listed in your Funding Information later in the submission form are also declared in your Financial Disclosure statement.	
Data and Code Availability	All the code used in our simulations, as well as that for reproducing the figures can be found in the GitHub repository:

From the time of publication, Authors are required to make fully available and without restriction all data and computational code underlying their findings. Please see our [PLOS Data Policy page](#) for detailed policy information, and our [Code Sharing](#) page for specific information on code sharing.

https://github.com/pabolich/coalescence_paper_analysis

A **Data Availability Statement**, detailing where the data (and code, if applicable) can be accessed, is required at first submission. Insert your Data Availability Statement in the box below. The statement you provide **will be published in the article**, if accepted.

Please see the [Data Reporting](#) section of our submission guidelines for instructions on what you need to include in your Data Availability Statement.

If the data and code are all contained in your submission files, please state: *All relevant data are within the manuscript and its Supporting Information files.*

PLOS allows rare exemptions to address legal and ethical concerns. If you have legal or ethical restrictions, please detail these in your Data Availability Statement below for the Journal team to consider.

Responses to the Reviewers

Summary of Revisions

We thank all three reviewers for their detailed review and constructive comments, which have prompted us to make revisions that have greatly strengthened our study. Overall, five main issues were raised by the reviewers, which we have addressed as follows:

- (i) *The parameter sensitivity of our results:* To address this, we have relaxed certain key assumptions and re-run the simulations, and added more detailed justifications for other parameter choices (especially of the binary uptake rates / consumer preferences)
- (ii) *The need to better quantify properties of parent communities:* To address this, we have performed extensive new simulations and added a new section and figure that shows the key properties of parent communities (providing a better basis for relating them to coalescence outcomes)
- (iii) *The effect of repeated coalescence events on community-level properties:* To address this, we added a new serial coalescence simulation procedure and correspondingly, a new results figure
- (iv) *The validity of our community-level competition and cooperation measures and the need for more analytical insight into how they link to coalescence outcomes:* To address this, we have revised the competition and cooperation measures, which prompted us to add another type of coalescence procedure—recursive coalescence—to better address the role of cooperation in coalescence
- (v) *The need for further analysis of the feasibility and stability of the parent and coalesced communities:* To address this, we have made key revisions to the model structure (that have rendered it more general), added a new analytical section to the SM, and extensive new supporting simulations

Our detailed, point-wise responses follow. Reviewers' original comments are in **black**, and our responses in **blue**. The revised manuscript with key changes highlighted, as well as an unmarked version, are attached.

Reviewer #1

1. Community cohesion is defined as the difference between facilitation and competition within a community. However, these metrics are computed on the initial pools, i.e. before the community assembly and stabilization steps. Because both competition and facilitation interactions can change during community assembly (e.g. species become extinct, modify their density etc) I am left wondering how the reported results would change if cohesion was measured in the stabilized communities.

All properties were measured for initial (before assembly), parents (after assembly), and coalesced communities. We apologize for not being sufficiently clear previously, and have now revised the main text (Figure 1 text legend, and details of “Step 2: Assembly of parent communities”, lines 155–157) to clarify this issue.

2. Similarly, the authors consider equal growth rates for all resources and species (e.g. binary C matrices). I am similarly left wondering how would relaxing this assumption and considering different growth rates for different species and resources impact the results.

The model used in this paper is an extension of the one originally developed in Marsland et al. (2019). In the original work the authors considered the impact of two separate scenarios where deviations from equal growth rates would be observed. Firstly, the case

where the C matrices were either gamma or Gaussian distributed rather than being binary. Secondly, when the ‘value’ (w_i) of metabolites was allowed to vary rather than being set to one for all metabolites. More biomass growth occurs per amount uptaken for metabolites with a higher ‘value’. In our model the value is implicitly set to one for all metabolites. The authors of the original paper found (see their Figs S1–6) that neither scenario significantly changed the ecosystem properties at steady state. As such we do not expect that relaxing this assumption would change the dynamics we observe significantly at the ecosystem level, though it would clearly have impacts at the species level. Our metrics (Eqs 5&7) are set up in such a way that they could account for non-binary C matrices, and would merely be weighted by metabolite ‘values’ to account for variations in those. Therefore, we expect that our results relating coalescence success to the strength of competition and facilitation to be broadly unchanged if we consider scenarios more realistic than the equal growth scenario. A comment referring the interested reader to this previous work has been added to the Discussion section (lines 350–354).

Reviewer #2

Major comment

3. The results strongly focus on the success of parental communities at surviving the coalescence process. However, I was missing a comparison between the properties of the parental communities and the coalesced community. The only results shown in that regard are to be found in supplement’s Fig S2. Are coalesced communities more cohesive than both parental strains, or something in between? Do facilitation/competition levels increase or decrease during coalescence? I think that there is an important discussion missing that could shed light on community-level properties that might, or might not, be optimized during coalescence. If we iterate coalescence for a few rounds (the survival species of two coalesced communities are exposed to a new coalescence event), what should we expect? What is the speed at which iterative coalescence leads to a sort of convergence in community structure? I think that adding a figure addressing (at least some of) these questions in the main text could improve the significance of the work.

We thank the reviewer for this comment in particular as it has prompted us to generate new results that we feel strengthens the paper substantially. Specifically, we added a new type of coalescence simulation (Serial) which provide deeper mechanistic insights, including into the the community-level evolution of structural and functional properties. Please see lines 177–183 of the Methods, lines 247–261 of the Results, and lines 319–343 of the Discussion sections. In short, we find that repeated coalescence decreases competition, increases cooperation, and enhances invasion resistance of communities.

Minor comments

4. The second paragraph in the introduction seems to rush directly into the importance of interactions in coalescence. Around line 14, I think that it could be worth to introduce other potential drivers of coalescence. Just as possible examples to address: resource availability, species growth rates, dispersal rates (communities in the process of coalescence might still receive some degree of immigration from the environment), seasonality (or temperature, day-night light cycles, …).

Following the reviewer’s suggestion, we have now widened mention of the scope of factors that can play an important role in coalescence (lines 14–21).

5. Eq. 2: wouldn’t it be better to add subindexes to the noise term epsilon? Otherwise it looks like a universal constant in the equation. In any case, the supplementary text S2

could be a bit more precise about the parameters of the gaussian distribution from which epsilon is sampled, since the text in S2 only says that is a ‘small fluctuation term’. How much small? How sensitive the system is to this noise?

We have added a subindex to epsilon, since it is species-dependent. We have added details on the distribution from which ϵ is sampled in the new Supplementary material Section S1. Additionally, our results are conditional on the fact that epsilon is small (and would remain qualitatively unchanged if in fact these fluctuations were set to zero), which we now clarify in Supplementary material Section S1.1 .

6. The meaning of kappa in Eq 5 should be briefly introduced in the main text, even if it’s explained in more detail in the supplement.

We have changed the measures of competition and facilitation to address the comments of reviewers 2 and 3, which eliminates the parameter κ (see our responses to comments #22 and 23). In brief, competition and facilitation are not dependent on the external environment, but rather only on the preferences of the bacteria.

7. Before Eq. S4: why is it that consuming resource j is relevant to assess competition on resource k? It seems to me that a third species could be consuming j and leak k, and then the two focal species compete just for consuming k.

This comment no longer applies, since the measures have been changed (see our response to previous comment and comments #22 and 23).

8. I also have a hard time at understanding why competition for leaked resources and supplied resources ‘need to be calculated differently’ (page 4 in the supplement), meaning separately. I understand that it’s convenient to compute the two terms separately. Do the authors mean ‘can be calculated separately’ instead of ‘needs to be...’?

We have now substantially simplified the measures of competition and facilitation (see responses to previous two comments and comments #22 and 23), such that we no longer separately calculate competition for leaked resources and externally supplied resources.

9. Why is it that $(1-l)$ affects the metric for abiotic competition? In principle, microbes compete to uptake a full unity of a given resource from the environment. Once this happens, microbial metabolism converts $(1-l)$ into population density, and a fraction ‘l’ is leaked in the form of other resources. But I don’t think that competition itself should be weighted by the unleaked fraction in this case.

The reviewer raises a very good point. We have eliminated the factor $(1 - l)$ in the competition calculation, and substantially simplified the measure (see responses to previous three comments and comments #22 and 23).

10. The amount of parameters needed to describe the model makes it a bit hard for the reader to remember the meaning of each of them in the first reading. To make the paper more rapidly understandable at first glimpse, the authors could consider adding legends or equations for each term appearing on the figures. For example a legend in Fig 1 could include ‘cohesion = facilitation – competition’, ‘D = metabolic matrix’. Same for the meaning of ‘kc’, ‘Kc’ and ‘r’ in Fig 2 and 3.

In order not to clutter the figures too much, we have avoided this. Instead, we have now added a table of parameters with definitions to the main text’s Methods section.

11. In figure 3D, R^*/r is named ‘Resource depletion level’. Through this name, I would understand that a higher value on this quantity translates to a lower R^* . That is, communities with higher resource depletion levels should more efficiently deplete the resources. Unless

I am missing something, this is not how it is interpreted and, to me, it is a little bit counterintuitive in the present form.

Agreed - we have changed ‘resource depletion level’ to ‘Resource abundance at equilibrium’ in our revised figures.

12. I think that fig 4A is not referenced in the main text.

We thank the reviewer for their careful reading. We have now made sure that all figures in the main text are referred to.

Reviewer #3

Main Points

13. In LV systems, the objects encoding competition, mutualism, etc., are explicitly defined in the ODEs. This is an important advantage with respect to other models and possibly one of the reasons of its popularity, since changes in these objects have an immediate interpretation. Still, there is a long history of controversy discussing what is the relative role of ecological interactions in, e.g., the stability of ecosystems, and much of this discussion arises from the different modelling choices. Nowadays, the large body of theory around this model allow us to link very explicitly objects such as the effective competition, which can encode all types of interactions present in the system, with properties of the system such as its global stability [1]. In the model the authors are dealing with, however, such connections are not so clear, since competition is indirect through consumption of resources.

In the model the authors are dealing with, however, such connections are not so clear, since competition is indirect through consumption of resources.

The Reviewer has raised a very interesting and important point. Yes, an analytical understanding of this (new) class of (microbial) consumer resource models lags behind that of LV type models. As the reviewer says, derivation of the effective interaction matrix is a key step towards an in-depth analysis of the link between the structure and dynamics (especially, feasibility and stability) of these systems. Recent work has indeed made an advance in that direction (Marsland et al., 2020). We think that an in-depth analysis of our model based on derivation of the effective interaction matrix is outside the scope of our paper. However, prompted by further helpful comments by the Reviewer, we have now performed an analysis of the model that addresses the link between interaction structure, feasibility, and coalescence outcomes, and provides deeper insights into our numerical results (next response). We also now check local stability of the assembled parent and coalesced communities (also explained in our response to the next comment).

14. The authors define \mathcal{C} and \mathcal{F} and they refer to these quantities as “competition” and “facilitation”, respectively. We should be sure that the definitions of \mathcal{C} and \mathcal{F} are correct, since the ecological interpretation arises from them. In addition, we should understand how these definitions are related to feasibility and dynamical stability, to rule out potential “side effects” responsible for the observed behaviour. For instance, the authors state that “minimizing competition ensures coalescence success”, and we should note that the minimization of competition does not only come from minimizing resource consumption overlap, but from the indirect (positive) effect of facilitation. This has been shown for mutualistic systems, for which mutualism increases structural stability by minimizing the effective competition Bastolla et al. (2009). However, in the second section of Results the title states “cooperation undermines coalescence success” which contradicts previous reasoning (since now facilitation has a negative effect), and they just refer to some empirical references without further mechanistic investigation. And there are many possible reasons

explaining this results (I suggest some below) and a very economic one, which is that the definitions selected for competition and facilitation are inaccurate. Of course, it would also be possible that everything is right and that there are two different regimes in which facilitation acts differently on the stability of the system. We found such behaviour in Pascual-García and Bastolla (2017), but we found that mutualism could be detrimental exactly on the opposite regime the authors describe – for high competition.

I think it is important that the authors provide an analysis on how the different choices of parameters influence feasibility and dynamical stability, and which is the connection with competition and facilitation, or at least that they provide evidence that they are sufficiently controlling them. For example, an important point regarding the parent communities is that the authors do not mention if they are verifying that the assembled communities are dynamically stable. It seems to me that they are implying that by, integrating the system, feasibility implies dynamical stability. I am not aware of an analytical result for this system as it is for LV, so I would like to ask them to provide analytical or numerical verification. This is important, in particular for those communities with high \mathcal{F} values, which may be trivially disrupted in the encountering simply because they are not dynamically stable despite being feasible. Feasibility itself is another potential confounding factor. For example, in the discussion of equation S2 the authors state that positivity is ensured if $R_j - \chi_0 > 0$ but I don't think this is true, since the term $1 - l$ does not multiply z_α , and hence we should expect that the feasible space shrinks for higher values of l . I would expect that, by having more facilitation, the system becomes less feasible (because there are more constraints to be fulfilled). Hence, if the feasible space is very small due to our parameter choices, when we combine both communities the combination will more likely be unfeasible if there is more facilitation. Is this due to facilitation? One may argue that, given the conceptual relatedness of l and χ_0 , a fair comparison among systems with different l values should also consider different χ_0 values, in a way in which the feasible space is comparable. Otherwise, we may attribute to the interactions an effect which is rather related with our modelling choices (here, a constant χ_0)

The Reviewer has raised several important points here, which we address in turn:

- *Definitions of the competition and facilitation measures:* We have now revised the definitions of \mathcal{C} and \mathcal{F} - please see comment #22 below (and main text lines 101–117). This also means that we now have a new and simpler measure ($\mathcal{C} - \mathcal{F}$) which we call “net competition” that well-predicts coalescence outcomes.
- *The link between the interaction structure measures and coalescence:* The Reviewer's comment prompted us look more closely at the mathematical model and perform additional analyses. As the reviewer correctly noticed, we found that whether leakage is included in the consumer species' cost function matters to coalescence outcomes. Specifically, if the cost is independent of leakage level, communities with higher leakage are unable to deplete resources to a concentration as low as communities with lower leakage, ensuring that less leaky community would dominate in pairwise coalescence events. Therefore, we further generalized our model and analyses to consider the effect of varying the cost function (detailed in new SM section S1.2). Doing so shows that if leakage is included in the cost function (and which is also biologically realistic) it eliminates the inherent coalescence bias towards less leaky communities. This substantially clarifies the role of competition vs cooperation in community coalescence (see lines 265–273 in Discussion), and also help reconcile our results with the previous research cited by the Reviewer that cooperation enhances coexistence, but conditional on direct competition levels (Bastolla et al. (2009); Pascual-García and Bastolla (2017)). We now also mention this in the Discussion (lines 305–307 & 286–289).

- *The issue of feasibility:* In a model community under the revised mathematical model (see response to previous comment), feasibility is now independent of leakage because of the new cost function (see main text lines 85–94, SM section S1.1, and previous answer). This allows us to fairly coalescence communities with different leakage and correctly determine the role of cooperative interactions in the outcome of community coalescence.
- *The issue of stability:* We have now checked that all parent as well as coalesced communities are locally stable (described in new SM section S1.3, and mentioned in lines 199–201 of the revised main text).

Specific Points

Abstract

15. Please give context explaining why “Encounters between microbial communities are becoming increasingly frequent across the globe”

We have now added more rationale for this sentence (2nd from last sentence in the abstract).

Introduction

16. Line 18: Since dominance is asymmetric by definition, I think more context should be given for those not familiar with the publication.

We have removed the word “asymmetrical”, since the reviewer is right in that “dominance” is asymmetrical by definition. See new explanation in lines 22–24.

17. Line 28. I and collaborators suggested this possibility in Ref 35 for communities termed “metabolically cohesive consortia”, please consider referencing it here.

This reference was a significant influence when we were conceptualising the study performed in this manuscript. We have included an additional reference to it as suggested (lines 29–31).

18. Line 33. “efficient resource”... “consumption” may be missing?

We definitely missed a word here. We have now added “use”, as this is the terminology used in the reference (line 36).

19. Line 33: “the invasion of a given taxon is determined by its community members” how? please provide more context.

We have now clarified this (lines 36–39)

Methods

20. I would like to ask the authors to provide a more elaborated description of step 2 and 3 of their methodology, and to include some results in the Main Text. It is very unbalanced the description of step 1 Vs step 2 and 3 and there are some unclear steps (such as how dynamical stability was verified). I think that the assembly of parent communities is interesting per se, and much needed to understand downstream analysis. I can imagine that the authors may consider this analysis distracting, but given that there is no space limitation and that these results are also novel, I encourage the authors to dedicate a specific section in Results. The first thing I think is unclear is if the quantities they present as describing the “parent communities”, refer to the “starting communities” (before the

assemblage) or to the assembled communities. More specifically, I would like to ask the authors to provide the summary statistics for Θ , \mathcal{C} , and \mathcal{F} , species richness (and perhaps true diversity, i.e. the exponential of the shannon diversity), and metrics describing the abundances of the communities at steady state for starting, assembled and final communities (e.g. mean and CV (coefficient of variation, a measure of relative variance in abundance around the mean)). Some of these metrics may be as predictive as Θ or they may have a relevant influence. Since cohesion depends on some of the metrics I am asking the authors to compute, and they have a more direct interpretation, we would like to discard they have a more relevant influence than cohesion. Some of the caveats I mentioned above regarding dynamical stability and feasibility of parent communities could also be addressed in this section.

We agree, and have now added more explanation of the assembly of the parent communities and their structural and dynamical properties (see response to comment #14 above with regards to stability analyses, lines 199–201; and Methods, Step 2, lines 139–157). We have also added new results section describing key properties of the parent communities before and after their assembly (lines 203–228).

21. Equation 1. As I anticipated, I haven't found the specific values the authors used for some parameters, in particular κ , g , and χ_0 . Please consider providing a table with all parameters (including k_c , K_c , etc), their values, meaning, and any other useful information.

We have added a table (Table 1) of parameters with definitions to the main text.

- (a) Although the model was first proposed elsewhere, I would acknowledge if the authors justify their choice of the functional form for the increase in resources due to metabolic by-products. More specifically, why the term contains the resources abundances. For instance, if we consider metabolic by-products from central metabolism, the secretion would be proportional to species densities only, as suggested in Butler and O'Dwyer (2018).

We have now added justification in the main text in lines 74–83 for our choice of leakage dependency which rests on two points. Firstly we assume dependence of by-product formation in resource uptake as it reflects a much more general feature of microbial communities in nature. The leakage of by-products has been observed to be important in microbial communities Goldford et al. (2018) and in fact even the authors in the paper referred to by the reviewer note that their assumption of consumer mass dependence is only a “reasonable approximation for some processes”. Secondly we assume dependence as it allows a much wider range of cross-feeding interactions between individual consumer pairs. Under the consumer-mass dependence assumption positive interactions are global such that each individual consumer benefits the same from the other consumers in the system. As the main aim of this paper is to establish the effects of interactions and their specific structure on community coalescence such an assumption is too restrictive and would not allow full range of effects of cross-feeding to be explored.

- (b) I do not see why the authors find justified to remove the dilution term for resources, please justify.

We had originally dropped the dilution term to make the model more comparable to that of Tikhonov (2016). However, we agree with the reviewer that the choice is somewhat arbitrary even if it does emulate certain natural (e.g., in an isolated soil pocket) or experimental (e.g., single vial experiments without serial dilution) scenarios where resources are not diluted. Therefore we have now also generated results for chemostat type systems (with resource dilution), and show that the assembly and

coalescence results don't qualitatively change due to adding this parameter. The only difference is that the number of simulations that converge when there is a dilution term increases because resources that are not consumed by any species will not diverge (as they did in the absence of dilution), but reach a carrying capacity determined by the ratio between κ and τ . We now show only the results for this altered model in the paper.

Competition and facilitation metrics

22. Why it is needed a split between abiotic and biotic resources. It looks like the authors are trying a transformation of the system into an effective system, but I do not see the need and I think it is complicating things quite considerably. Firstly, If there is a split between abiotic and biotic resources it means that some of the κ parameters are zero, a possibility that should be stated from the beginning. Then a new parameter should be considered to quantify the fraction of resources that are externally supplied and understand its consequences for feasibility, dynamical stability, etc. I think that a microbe just "sees" a resource with some abundance, no matter its origin. Therefore, if species α consumes resources at rate $(1 - l)c_{\alpha k}$ and species β at rate $(1 - l)c_{\beta k}$, I would expect defining competition by the product:

$$C_{\alpha\beta} = (1 - l)^2 \sum_k c_{\alpha k} c_{\beta k},$$

and similarly for facilitation

$$F_{\alpha\beta} = l(1 - l) \sum_{jk} c_{\alpha j} D_{jk} c_{\beta k}.$$

This would be somewhat the "structural— competition and facilitation at each time-step of the simulation, and the simulation will tell us the effects, so I do not see the need, for instance, for considering the infinite loop of cycles (moreover, if its considered for $l(1 - l)$ it should also be considered for the object $c_{\alpha j} D_{jk} c_{\beta k}$.

The reviewer raises very good points. The separation between abiotic and biotic resources is unnecessary, and only serves to make things more complicated. Therefore have now simplified our competition and facilitation metrics, making them more parsimonious and intuitive (lines 101–117). Additionally, some of the changes made to the metrics also follow the recommendations of a related point raised by Reviewer 2 (see above comments # 7–9 and our response).

23. I do not understand the inclusion of $\tilde{\kappa}$ (possibly for the same reason I do not understand the need for the split of resources). But even if this distinction is needed, according to its definition, competition will be stronger for resources with high supply, and one could argue that species competing for only one available resource experience higher competition if the resource is scarce.

The reviewer is right; including $\tilde{\kappa}$ was an attempt to transform the system into an effective system. However, it is more rigorous to state simple measures of competition and facilitation, and let the simulations tell us the effect. Consequently, $\tilde{\kappa}$ has been removed from our metrics.

24. The authors average out the values of the matrix but they exclude the diagonal values. However, I do not see why should be excluded since it is very likely an important component of the stability of the system (think about diagonal dominance in LV systems), even more if we aim to compare the relative role of competition and facilitation.

We agree that this is an important issue to clarify. Therefore, we have changed our definitions of \mathcal{C} and \mathcal{F} to take into account intraspecific competition and facilitation. We now include these terms when calculating community-level competition and facilitation, finding that this they do not qualitatively change the results (see Fig 3C). We also agree that understanding the contribution of these intraspecific interaction values to the dynamics of these systems is important, but with reference to our comment # 13 above, that would require us to calculate the effective interaction matrix, which we think is an avenue for future work.

25. It should be clear whether intraspecific syntropy is considered (i.e. that a species consumes what it secretes). Although I can imagine that there are certain situations in which this has been observed when conditions change (similar to diauxic growth), in this model the metabolic strategies are fixed independently of the environmental conditions so I do not see it justified to consider this possibility, at least in general.

Consumption of leaked resources is a function present in the original consumer resource model upon which ours is based (Marsland et al., 2019). Leakage of essential metabolic resources back into the environment is a pervasive phenomenon in microbes (Morris, 2015; Kallus et al., 2017). Whilst it may seem counter-intuitive for microbes to secrete metabolites that they also consume, it has been suggested that such leakage can provide short-term growth benefits to bacteria, as a flux control or growth-dilution mechanism (Yamagishi et al., 2020a,b). We had previously mentioned this briefly in the discussion, but have now expanded further on this point, both in the Discussion (lines 295–303), and in the Methods where leakage in the model is described (lines 83–84).

26. By averaging across all pairs of species, the definition represents somewhat a mean-field competition value. But I think this may not be justified for the guild structure, which is far from mean field. Since, between guilds, competition is zero, I think that the quantity should be computed for each guild independently and then averaged out across guilds. This I think would solve the (to my taste artificially low values) arising from many pairs with no competition. Moreover, since competition is harsh within guilds, I would expect that the outcome of the encountering would exclude the species feeling stronger competition within each guild (a prediction we present in Ref 35, see Fig 4a which is the author's guild structure and Fig 4b and 5 for possible outcomes). This could be verified computing how the species extinctions are distributed across guilds.

We have now calculated community level competition as the average over the block diagonals, and community level facilitation as the average over the off-block diagonals. In SM Section S4.1, figure S6, we show that the results are qualitatively robust to this modification.

Simulations

27. Equation 7. Since α is used for both species and iterations it is confusing if it is not clarified (as the authors present in the SM).

We have changed the explanation of the simulation procedures, which can be found now in the Supplementary material S2.1.

28. My understanding of the procedure is that the author adapted the preferential attachment model of Barabási and Albert to their problem, perhaps it is worth mentioning it in the beginning since it is a widely known model.

We were not thinking of the preferential attachment model when we developed this procedure. However, we realize now the clear similarity and therefore have added a statement

noting the similarity and a reference to Barabási and Albert's original paper (see supplementary text section S2.1).

29. The authors state that “the resources that are highly demanded are also secreted in large fractions”. But if, for example, glucose is highly demanded, we will not expect glucose to be secreted in large fractions, but acetate. Might be they meant “the resources that are secreted in large fractions are highly demanded”? The question here is to be clear on whether a species is consuming what it secretes or not.

Our model is specified such that in highly facilitative communities ($k_f \rightarrow 1$), resources that are highly demanded by the community are also secreted in large fractions. We have now clarified in the text that this pertains to facilitative communities only (Supplementary text S2.1). The metabolic matrix D specifies the by-product leakage as a proportion of resource j which is transformed into resource k . It is therefore possible for species in our model to consume what they secrete, if their consumer preference matrix encodes for the by-product of another resource that they utilise. Ultimately the relative simplicity of our model, which does not contain a true stoichiometric matrix, prevents us from analyzing our results in terms of real metabolites.

Results and Discussion

30. In Figure 3D I would perhaps find more interesting and easier to understand plotting in the y axis the species richness (and colour the dots by resource depletion level), since it would be more directly connected with the stability (x-axis) diversity (y-axis) problem (with the colour being the mechanism)

We agree; this has been changed in the revised manuscript (see main text Figure 3).

31. Figure S2 top. How can the number of consumers of a resource increase after assembly? Typo in the caption “plots it the”

Corrected. Additionally, figure S2 has been removed, since we have included a section (lines 203–228) on community assembly in the main text that addresses the similar issues that Fig S2 was addressing.

32. SM Page 11. Type “this communities”

Corrected.

33. SM8. Please review this section, I think the first equation is not correct, if I am not wrong it should be $(\text{diag}(g \circ n)\text{CR})$. Also, l is a constant so no need for vector notation. Consider presenting the element-wise product \circ

The Reviewer is correct, and we changed the first line of the equation (2nd set of equations in SM section S1) to correct the error.

References

- Bastolla, U., Fortuna, M. A., Pascual-García, A., Ferrera, A., Luque, B. and Bascompte, J. (2009), ‘The architecture of mutualistic networks minimizes competition and increases biodiversity’, *Nature* **458**(7241), 1018–1020.
- Butler, S. and O’Dwyer, J. P. (2018), ‘Stability criteria for complex microbial communities’, *Nature Communications* **9**(1).
- URL:** <http://dx.doi.org/10.1038/s41467-018-05308-z>

- Goldford, J. E., Lu, N., Bajić, D., Estrela, S., Tikhonov, M., Sanchez-Gorostiaga, A., Segrè, D., Mehta, P. and Sanchez, A. (2018), ‘Emergent simplicity in microbial community assembly’, *Science* **361**(6401), 469–474.
- Kallus, Y., Miller, J. H. and Libby, E. (2017), ‘Paradoxes in leaky microbial trade’, *Nature Communications* **8**(1).
- Marsland, R., Cui, W., Goldford, J., Sanchez, A., Korolev, K. and 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.
- Marsland, R., Cui, W. and Mehta, P. (2020), ‘The minimum environmental perturbation principle: A new perspective on niche theory’, *American Naturalist* **196**(3), 291–305.
- Morris, J. J. (2015), ‘Black Queen evolution: The role of leakiness in structuring microbial communities’, *Trends in Genetics* **31**(8), 475–482.
URL: <http://dx.doi.org/10.1016/j.tig.2015.05.004>
- Pascual-García, A. and Bastolla, U. (2017), ‘Mutualism supports biodiversity when the direct competition is weak’, *Nature Communications* **8**.
- Tikhonov, M. (2016), ‘Community-level cohesion without cooperation’, *eLife* **5**, e15747. doi: 10.7554/eLife.15747.
- Yamagishi, J. F., Saito, N. and Kaneko, K. (2020a), ‘Advantage of Leakage of Essential Metabolites for Cells’, *Physical Review Letters* **124**, 048101.
URL: <https://doi.org/10.1103/PhysRevLett.124.048101>
- Yamagishi, J. F., Saito, N. and Kaneko, K. (2020b), ‘Microbial Potlatch: Cell-level adaptation of leakiness of metabolites leads to resilient symbiosis among diverse cells’, *bioRxiv* pp. 1–17.

The role of competition versus cooperation in microbial community coalescence

Pablo Lechón-Alonso^{1,2*}, Tom Clegg², Jacob Cook^{2,3}, Thomas P. Smith², Samraat Pawar²

1 Department of Ecology & Evolution, University of Chicago, Chicago, IL, USA.

2 Department of Life Sciences, Imperial College London, Silwood Park, Ascot, UK

3 Centre for Integrative Systems Biology and Bioinformatics, Imperial College London, UK

*plechon@uchicago.edu

Abstract

New microbial communities often arise through the mixing of two or more separately assembled parent communities, a phenomenon that has been termed “community coalescence”. Understanding how the interaction structures of complex parent communities determine the outcomes of coalescence events is an important challenge. While recent work has begun to elucidate the role of competition in coalescence, that of cooperation, a key interaction type commonly seen in microbial communities, is still largely unknown. Here, using a general consumer-resource model, we study the combined effects of competitive and cooperative interactions on the outcomes of coalescence events. In order to do so, we simulate coalescence events between pairs of communities with different degrees of competition for shared carbon resources and cooperation through cross-feeding on leaked metabolic by-products (facilitation). We also study how structural and functional properties of post-coalescence communities evolve when they are subjected to repeated coalescence events. We find that in coalescence events, the less competitive and more cooperative parent communities contribute a higher proportion of species to the new community, because this endows superior ability to deplete resources and resist invasions. Consequently, when a community is subjected to repeated coalescence events, it gradually evolves towards being less competitive and more cooperative, as well as more species rich, robust and efficient in resource use. Encounters between microbial communities are becoming increasingly frequent as a result of anthropogenic environmental change, and there is great interest in how the coalescence of microbial communities affects environmental and human health. Our study provides new insights into the mechanisms behind microbial community coalescence, and a framework to predict outcomes based on the interaction structures of parent communities.

Author summary

In nature, new microbial communities often arise from the fusion of whole, previously separate communities (community coalescence). Despite the crucial role that interactions among microbes play in the dynamics of complex communities, our ability to predict how these affect the outcomes of coalescence events remains limited. Here, using a general mathematical model, we study how the structure of species interactions confers an advantage upon a microbial community when it encounters another, and how

communities evolve after undergoing repeated coalescence events. We find that competitive interactions between species preclude their survival upon a coalescence event, while cooperative interactions are advantageous for post-coalescence survival. Furthermore, after a community is exposed to many coalescence events, the remaining species become less competitive and more cooperative. Ultimately, this drives the community evolution, yielding post-coalescence communities that are more species-rich, productive, and resistant to invasions. There are many potential environmental and health implications of microbial community coalescence, which will benefit from the theoretical insights that we offer here about the fundamental mechanisms underlying this phenomenon.

Introduction

Microbial communities are widespread throughout our planet [1], from the human gut to the deep ocean, and play a critical role in natural processes ranging from animal development and host health [2, 3] to biogeochemical cycles [4]. These communities are very complex, typically harbouring hundreds of species [5], making them hard to characterize. Recently, DNA sequencing has allowed high-resolution mapping of these communities, opening a niche for theoreticians and experimentalists to collaboratively decipher their complexity and assembly [6–11, 13].

Entire microbial communities are often displaced over space and come into contact with each other due to physical (e.g., dispersal by wind or water) and biological (e.g., animal-animal or animal-plant interactions, and leaves falling to the ground) factors [14–17]. The process by which two or more communities that were previously separated join and reassemble into a new community has been termed community coalescence [18]. Although microbial community coalescence is likely to be common, the effects of both intrinsic and extrinsic factors on the outcomes of such events remains poorly understood [19]. Among extrinsic factors, resource availability, immigration rate of new species, and environmental conditions (especially, pH, temperature, and humidity) are likely to be crucial [20–22]. Among intrinsic factors, the role of functional and taxonomic composition and the inter-species interaction structures of parent communities are expected to be particularly important [20, 21]. We focus on the role of species interactions on community coalescence in this study.

Early mathematical models suggested that in encounters between animal and plant communities, species in one community are more likely to drive those in the other extinct (community dominance) [23, 24]. This was explained as being the result of the fact that communities are a non-random collection of species assembled through a shared history of competitive exclusion, and therefore act as coordinated entities. Recent theoretical work [25] has more rigorously established this for microbial community coalescence events, showing that the dominant community will be the one more capable of depleting all resources simultaneously. Overall, these findings suggest that communities arising from competitive species sorting exhibit sufficient “cohesion” to prevent invasions by members of other communities [26, 27].

However, empirical support for the role of competition alone in coalescence outcomes is circumstantial, and the role of cooperation, which is commonly observed in microbial communities, is yet to be addressed theoretically. For example, during coalescence in methanogenic communities, “cohesive” units of taxa from the community with the most efficient resource use are co-selected [22]; and in aerobic bacterial communities, the invasion success of a given taxon is determined by its community members as a result of collective consumer-resource interactions and metabolic feedbacks between microbial growth and the environment [28]. Nonetheless, neither of these studies addressed the role of competition and cooperation in shaping coalescence success. Yet, these microbial

communities exhibit cooperation through a typically dense cross-feeding network, where leaked metabolic by-products of one species are shared as public goods across the entire community [29–31]. Indeed, several studies have suggested that a combination of competitive and cooperative interactions may determine the outcome of coalescence in microbial communities [21, 32, 33].

Here, we focus on the relative importance of competition and cooperation in community coalescence. We use a general consumer resource model that includes cross-feeding to assemble complex microbial communities having different degrees of competition and cooperation. We focus on determining the relative importance of the two types of interactions on outcomes of coalescence events, as well as the subsequent evolution of the structural and functional properties of coalesced communities.

Methods

Mathematical model

Our mathematical model for the microbial community dynamics is based on the work of Marsland et al. [6] (see Supporting text section 1, and Fig 1):

$$\begin{aligned} \frac{dN_\alpha}{dt} &= g_\alpha N_\alpha \left((1 - l_\alpha) \sum_j c_{\alpha j} R_j - z_\alpha \right), \\ \frac{dR_j}{dt} &= \kappa_j + \tau^{-1} R_j - \sum_\alpha N_\alpha c_{\alpha j} R_j + \sum_{\alpha k} N_\alpha l_\alpha D_{\alpha j k} c_{\alpha k} R_k. \end{aligned} \quad (1)$$

Here, N_α ($\alpha = 1, \dots, s$) and R_j ($j = 1, \dots, m$) are the biomass abundance of the α^{th} microbial (e.g., bacterial) species and the concentration of the j^{th} resource (e.g., carbon substrate). The growth of species α is determined by the resources it harvests minus the cost of maintenance (two terms in the brackets). Resource uptake depends on the resource concentration in the environment R_j , and on the uptake rate of species α , here assumed to be binary (j ($c_{\alpha j} = 1$ or $c_{\alpha j} = 0$)). The leakage term l_α determines the proportion of this uptake that species α releases back into the environment as metabolic by-products, with the remainder ($1 - l_\alpha$) being allocated to growth. The uptake that remains after subtracting a maintenance cost (z_α) is transformed into biomass with a proportionality constant of g_α , the value of which does not affect the results presented here.

The change in the concentration of resources in the environment (second line in Eq 1) is determined by four terms. The first and second terms represent the external supply and dilution of resource j , which give the rates at which the j^{th} resource enters and leaves the system. The third term is the uptake of the j^{th} resource from the environment, summed across all s consumers, and the fourth term represents resources entering the environmental pool via leakage of metabolic by-products. By-product leakage of species α is determined by the metabolic matrix D_α (or the “stoichiometric” matrix; [6]), with $D_{\alpha j k}$ representing the leaked proportion of resource j that is transformed into resource k by species α . Energy conservation dictates that D_α is a row stochastic matrix, meaning that its rows sum to 1. Note that in this model, rates of metabolic by-product formation are dependent on resource uptake (i.e., the amount of resource leaked into the environment depends on the amount being consumed). We use this specific structure as opposed to dependence of leakage directly on consumer biomass (e.g., [34], which would mean that the relative leakage to each resource type remains constant within a community), because it (i) more accurately reflects biological reality (microbes typically produce specific byproducts when feeding on specific

resources [7]) and (ii) allows greater variation in cross feeding between consumer pairs (as the contribution to leakage from each consumer is unique) allowing us to better explore the effects of interactions. Also, we allow species to leak metabolites that they also consume (we address this assumption further in the Discussion).

We define the consumer's maintenance cost to be:

$$z_\alpha = \chi_0(1 + \epsilon_\alpha)(1 - l_\alpha) \sum_j c_{\alpha j}, \quad (2)$$

where χ_0 is the average cost of being able to consume a given resource, the summation represents the total number of resources that species α is able to process, and ϵ_α is a small random fluctuation that introduces variation in the cost for species that have identical preferences. Eq 2 ensures that neither generalists nor specialists are systematically favoured during community assembly (by imposing a greater cost on species that consume a wider range of resources), and that all species are able to deplete resources to similar concentrations independently of their leakage level (see Supporting text section 1 for rationale, Fig S1 for results under different cost functions; and Discussion).

The above model entails the following assumptions: (i) all resources contain the same amount of energy (taken to be 1 for simplicity), (ii) a linear, non-saturating consumer functional response, (iii) binary consumer preferences (uptake rates), and (iv) an environment where all resources are externally supplied in equal amounts. We address the implications of these assumptions in the Discussion.

Parameter	Description	Value	Units
g_α	Biomass synthesised per unit energy harvested by species α	1	$B_a E_a^{-1}$
$c_{\alpha j}$	Uptake rate of metabolite j by species α	1/0	$s^{-1} B_a^{-1}$
D_{kj}	Fraction of metabolite k that's leaked in the form of j	[0, 1]	None
κ_j	Supply rate of metabolite j	2	mol s^{-1}
τ	Time scale of resource dilution	0.25	s
l_α	Fraction of energy intake leaked as metabolic byproducts	[0, 1)	None
χ_0	Average cost per metabolic pathway	0.1	E_a
ϵ_α	Random fluctuation in species α 's cost	$N(0, 0.1)$	None
s	Number of species in pre-assembly communities	60	-
m	Number of metabolites	60	-
k_c	Competition factor	[0,1)	-
k_f	Facilitation factor	[0,1]	-
K_c	Inter-guild competition factor	(0.1, 0.9)	-
K_f	Inter-guild facilitation factor	(0.1, 0.9)	-

Table 1. Table of parameters used in our model. The units E_a and B_a represent arbitrary energy and biomass units, respectively.

Competition and facilitation metrics

In the system of equations 1, competition for resources exists because all pairs of consumer species generally share some resource preferences (their metabolic preferences vectors are not orthogonal). We quantify the pairwise competition between a species pair (α, β) by counting the resource preferences they share through the scalar product of their preference vectors, $\mathbf{c}_\alpha \cdot \mathbf{c}_\beta$. Therefore, community-level competition (denoted as \mathcal{C}) can be calculated by taking the average of the competition matrix, which encodes the competition strengths between all species pairs, that is

$$\mathcal{C} = \langle CC^T \rangle, \quad (3)$$

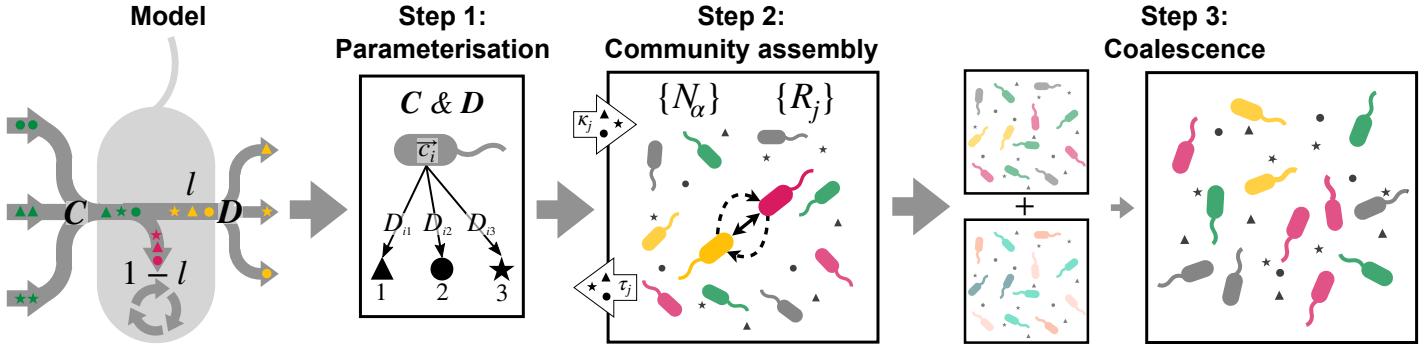


Fig 1. Overview of the coalescence modelling methodology. **Step 1.** The matrix of resource preferences (C) and the metabolic matrix (D) are sampled for each community. Black polygons are different resource types. **Step 2.** Dynamics of the system are allowed to play out (Eqs 1) until a locally stable equilibrium point is reached. Species composition and abundance, along with community-level competition \mathcal{C} (solid bidirectional arrows, Eq 3), and facilitation \mathcal{F} (dashed unidirectional arrows, Eq 4) are measured in assembled communities. **Step 3.** A pair of the assembled parent communities are mixed, and the resulting community integrated to steady state. For the random and recursive coalescence procedures, the contribution of each parent community to the final mix is analyzed ($S_{1,2}$, Eq 5) as a function of their interaction structures ($\mathcal{C}_{1,2}$ and $\mathcal{F}_{1,2}$) before they coalesced. In the case of the serial coalescence procedure, the properties of the resident community \mathcal{R} are tracked after each coalescence exposure.

where C is the $s \times m$ matrix of metabolic preferences of all the species in the community. 108

On the other hand, facilitation occurs when a species leaks metabolic by-products that are used by another species. We measure pairwise facilitation from species $\alpha \rightarrow \beta$ by calculating the fraction of secreted resources from species α that are consumed by species β per unit of resource abundance, $l_\alpha c_\alpha^T D_\alpha c_\beta$. Similar to competition, we compute community-level cooperation (denoted as \mathcal{F}), by taking the average of the facilitation matrix, which encodes the competition strengths between all species pairs, that is 110
111
112
113
114
115

$$\mathcal{F} = \langle \sum_\alpha \mathcal{D}(l) C D_\alpha C^T \rangle, \quad (4)$$

where where $\mathcal{D}(l)$ is a diagonal matrix with the leakage vector of each species in the community in its diagonal. 116
117

Henceforth, we refer to the quantity $\mathcal{C} - \mathcal{F}$ as “net competition”, which we later show is related to the “cohesion” defined in previous work [25]. 118
119

Simulations

In Fig 1 we present an overview of our simulations, which we now describe. For the parameter values used, see Table 1. 120
121
122

Step 1: Parameterization

We first set the parameters of the initial communities (before assembly) such that they span interactions across the spectrum of net competition ($\mathcal{C} - \mathcal{F}$). For each parent community, we modulate the structure of the C and D matrices (consisting of the resource preferences $c_{\alpha j}$'s and secretion proportions D_{jk} 's, respectively) by developing constrained random sampling procedures that guarantee specific levels of competition and facilitation at the community's steady state (see Supporting text section 2). In addition, we also add structure to C and D to emulate the existence of distinct resource classes and consumer guilds (see Supporting text section 4). With these procedures, net 123
124
125
126
127
128
129
130
131

competition in an assembled parent community can be regulated through four parameters: k_c (competition factor), k_f (facilitation factor), K_c (inter-guild competition factor), and K_f (inter-guild facilitation factor) (see Supporting text section 2). Note that we parameterize the initial communities by assuming (i) a shared core metabolism encoded in D , and (ii) a common leakage fraction l for all species (the implications of which we address in the Discussion), but we relax these assumptions in our coalescence simulations (Methods; Step 3).

Step 2: Assembly of parent communities

After parameterization, we numerically integrate Eqs 1 until steady state (a putative equilibrium) is reached. We perform 100 such assembly simulations with random sets of consumers for each combination of competition and facilitation factors (i.e., $k_c = k_f \in [0, 0.5, 0.9]$), repeating this for three values of leakage ($l \in [0.1, 0.5, 0.9]$). We compare species composition, abundances, and interaction structure of communities before and after assembly. In order to compare species composition, we calculate the difference between the proportion of species with a certain number of metabolic preferences, n_r (n-preference consumers, a measure of generalism), before and after assembly as

$$\Delta n_r = \frac{1}{p(n_r)} \left(\frac{T_{n_r}}{r} - \frac{T_{n_r}^0}{r^0} \right).$$

Here, $p(n_r)$ is the probability that a species has n_r metabolic preferences (Supporting text section 2), the 0 denotes before assembly, r is species richness, and T_{n_r} is the number of species with n_r preferences. Thus, when $\Delta n_r > 0$ the proportion of species with n_r metabolic preferences increases after assembly and vice versa. In order to analyze species abundances, we track the abundance fraction of consumers in each group of n-preference species, calculated as total abundance of the species in the group, divided by the total community biomass. Finally, we address the interaction structure after assembly by quantifying the levels of competition (\mathcal{C}) and facilitation (\mathcal{F}) in the assembled communities (Fig 2A).

Step 3: Coalescence

To simulate coalescence between a pair of assembled parent communities, we set all resources to their initial concentrations, and numerically integrate the new combined system to steady state. In order to disentangle the effects of competition versus cooperation and study the effect of repeated coalescence events, we simulate three coalescence scenarios: random, recursive, and serial, as follows (further details in Supporting text section 3).

Random coalescence. To address the effects of competition alone in the outcome of coalescence events, here we coalesce pairs of randomly sampled parent communities having the same leakage value l ($2 \cdot 10^4$ pairs for each leakage level, Fig 3C). That is, we fix the leakage level to ensure that the communities have, on average, similar cooperation levels, but leave k_c free to vary such that they span a broad range of competition levels.

Recursive coalescence. In order to study the effects of cooperation in particular on community coalescence, we repeatedly coalesce a given pair of communities A and B , slightly increasing the leakage of the latter in each iteration (Fig 4A). This allows us to modify the strength of cooperative interactions in the community, because facilitation is proportional to l (Eq 4), while keeping competition levels constant, because competition is independent of l (Eq 3) and the remaining parameters are kept fixed.

Serial coalescence. In the natural world, a community may be exposed to more than one coalescence event. Consequently, here we simulate a scenario where a local (“resident”) community \mathcal{R} harbouring species with leakage $l_{\mathcal{R}}$, and metabolism $D_{\mathcal{R}}$ is successively invaded by many other randomly sampled communities (“invaders”), \mathcal{I} with species of leakage $l_{\mathcal{I}}$ and metabolism $D_{\mathcal{I}}$ (Fig 5A). This allows us understand how the functional and structural properties of a microbial community evolve over time under successive encounters with other communities.

At the end of each random and recursive coalescence simulation, we quantify the dominance of either parent community in the post-coalescence community by measuring similarity of the latter to each of the two parents (indexed by 1 and 2) as:

$$S_{1,2} = \mathbf{p}_f \cdot \left(\frac{\mathbf{p}_2}{r_2} - \frac{\mathbf{p}_1}{r_1} \right), \quad (5)$$

where \mathbf{p}_f , \mathbf{p}_1 , and \mathbf{p}_2 are $(s_1 + s_2)$ -dimensional vectors of species presence-absence in the post-coalescent, and parent communities 1, and 2, respectively, with r_1 and r_2 the species richness values of the parent communities 1 and 2, respectively (calculated as $r_i = \sum p_i$). If $S_{1,2} = -1$, the coalesced community is identical to parent community 1, and if $S_{1,2} = 1$, it is identical to parent community 2. This measure is independent of the species richness. Thus we can mix communities with different species richness while avoiding a bias in similarity towards the richer one. We then analyze how this dominance measure depends on the interaction structure of the parent communities ($\mathcal{C}_{1,2}$ and $\mathcal{F}_{1,2}$; Eqs 3 and 4). After each coalescence event in the serial coalescence procedure, we measure competition and facilitation levels of the resident community, along with the average species maintenance cost, average resource abundance at equilibrium, species richness, and number of successful invasions, during the entire sequence of serial coalescence events. For all assembled parent as well as coalesced communities we confirmed that the steady state was a locally asymptotically stable equilibrium point (Supporting text section 1).

Results

Assembly of parent communities

In Fig 2 we show the key features of assembled communities. Figure 2A shows that as expected from Eqs 3 and 4, the levels of community-wide competition and facilitation are positively correlated, mediated by the structure of the C and D matrices. Figure 2B shows that the difference between the proportion of n-preference consumers before and after assembly (Δn_r , Methods; Step 1), increases for all simulated values of k_c , indicating that more generalist species are less prone to extinction during assembly. For the lowest value of k_c , Δn_r is in fact a monotonically increasing function of the number of preferences. This is expected because a species able to harvest energy from multiple resource pools is less likely to go extinct during community dynamics. As k_c increases, Δn_r reaches a minimum (Fig 2B), indicating that in more competitive environments pure specialists become more prevalent than moderate generalists. This is due to the fact that in a highly competitive environment the resource demands are concentrated on a subset of resources, while others are barely consumed (Supporting text section 2). In these communities, consumers that specialize exclusively on empty niches thrive. Figure 2C shows that specialist consumers are systematically present in higher abundance than generalists for all values of k_c . This is because several specialists are able to deplete all resources through their combined action more efficiently than one generalist [26], and as a result, although generalists are more persistent than specialists upon assembly (Fig 2B), they achieve lower abundances at equilibrium (Fig 2C). In

Figure 2B (inset), Δn_r is weighted by the abundance fraction (Methods; Step 2) of each group of n-preference consumers. This reveals an optimal group of consumers with a number of metabolic preferences that maximizes both survival probability and abundance at equilibrium. This optimal value increases for more competitive environments (as k_c increases). Finally, Figure 2D shows that more competitive communities tend to be less species rich, as expected from general competition theory.

223
224
225
226
227
228

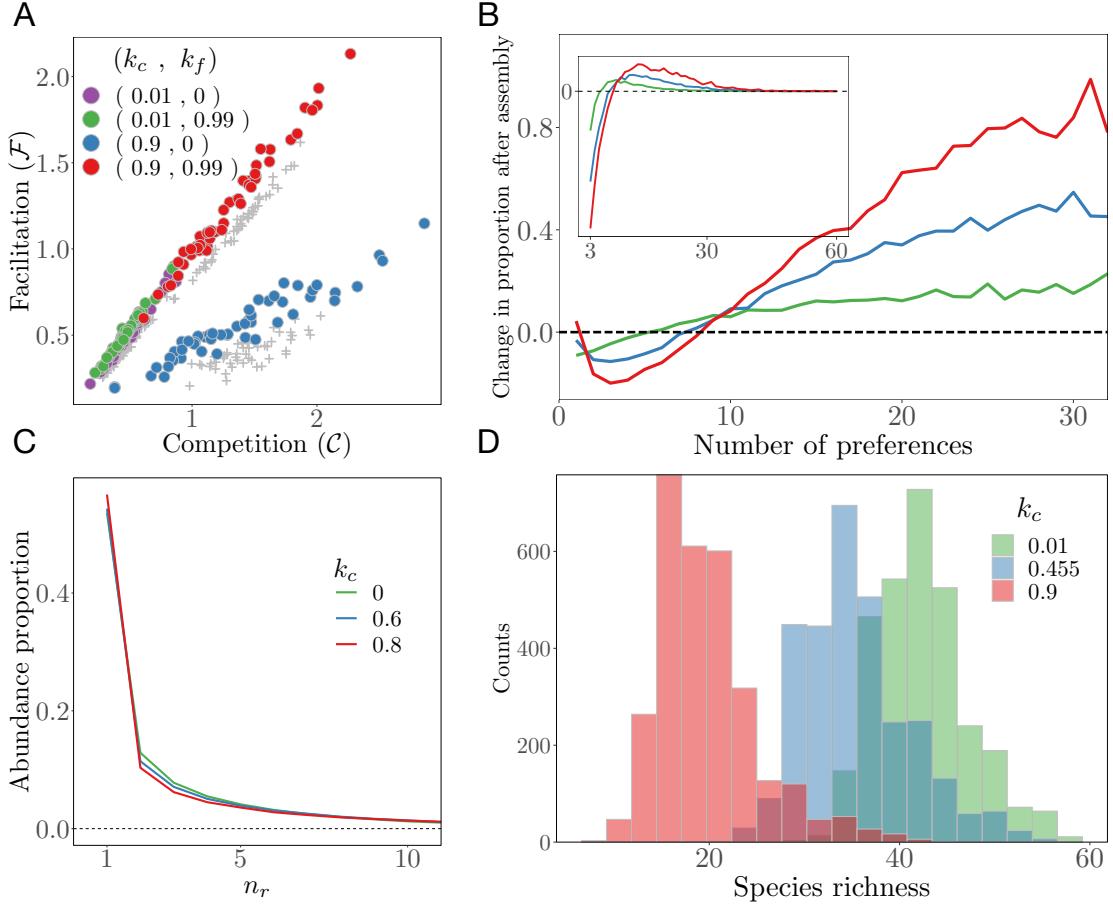


Fig 2. Features of assembled parent communities. **A:** Facilitation versus competition level in starting (grey dots) and assembled (coloured dots) communities for leakage $l = 0.9$ and different combinations of competition (k_c) and facilitation (k_f) factors. Communities assembled for each pair of $[k_c, k_f]$ values have the same colour. The assembled communities are always significantly more cooperative at the end of the assembly than at the start. **B:** Change in proportion of species in each n-preference consumer group (Δn_r , Methods; Step 2) before and after assembly, for different values of k_c (legend in panel C) indicating that more generalist species are less prone to extinction during assembly. Values for $n_r > 30$ had too much uncertainty due to low sampling and therefore have been removed for clarity. **Inset:** Δn_r is weighted by the abundance fraction at equilibrium of each n-preference consumers group. **C:** Abundance fraction of the n-preference species groups for different values of k_c . **D:** Distributions of species richness values of parent communities assembled under different k_c values. Increasing competitiveness tends to decrease species richness.

229
230
231

Reducing competition increases coalescence success

Figure 3 shows that communities with lower net competition values tend to perform better in coalescence as seen by the positive relationship between parent community

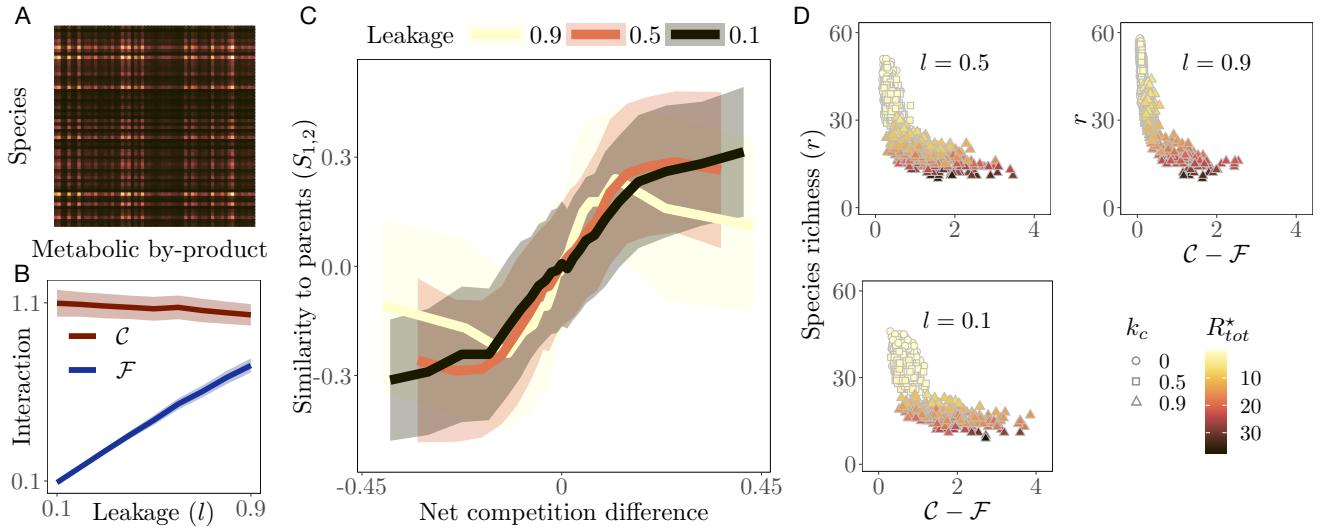


Fig 3. Community coalescence between pairs of randomly picked communities with same leakage.

A: Example of the secretion matrix with elements $(CD)_{\alpha k}$ representing the total leakage of resource k by species α . **B:** Community-level competition \mathcal{C} (dark red) and facilitation \mathcal{F} (blue) averaged across simulations for each leakage value. Since competition does not depend on the leakage, it remains consistently high throughout. Facilitation, on the other hand, increases linearly with leakage. **C:** Parent community dominance ($S_{1,2}$) as function of net competition difference $(\mathcal{C}_1 - \mathcal{F}_1) - (\mathcal{C}_2 - \mathcal{F}_2)$ (solid lines ± 1 standard deviation (shaded)), binned (20 bins) over communities with similar x axis values, for three community-wide leakage levels. The post-coalescence community is more similar to its less (net) competitive parent. **D:** Species richness (r) as a function of net competition in parent communities, coloured by total resource concentration at steady state (R_{tot}^*). The observed negative correlation for all values of leakage shows that communities with lower net competition tend to be more species-rich and also better at depleting resources (brighter coloured values, corresponding to lower levels of R_{tot}^* are scattered towards the top left of the plots).

dominance ($S_{1,2}$) and the quantity $(\mathcal{C}_1 - \mathcal{F}_1) - (\mathcal{C}_2 - \mathcal{F}_2)$ (Fig 3C). That is, communities that emerge following coalescence tend to have greater similarity with the less net competitive parent. This trend holds at higher values of leakage, where cooperation levels are significant (Fig 3B), but with a clear critical point (the yellow line reverses in direction at a value of effective competition difference). This pattern is driven by the fact that less competitive parent communities deplete resources more efficiently and achieve a higher species richness (Fig 3D; Supporting text section 1). All these results also qualitatively hold for microbial communities that have consumer guild structure (Supporting text section 4).

Cooperation further enhances coalescence success

Figure 4 shows that when a community (B) whose leakage fraction increases successively during recursive coalescence events with another (A) with a fixed leakage level, the former becomes increasingly dominant. The result is consistent for a range of leakage values of community A . This shows that increasing cooperation levels enhance coalescence success.

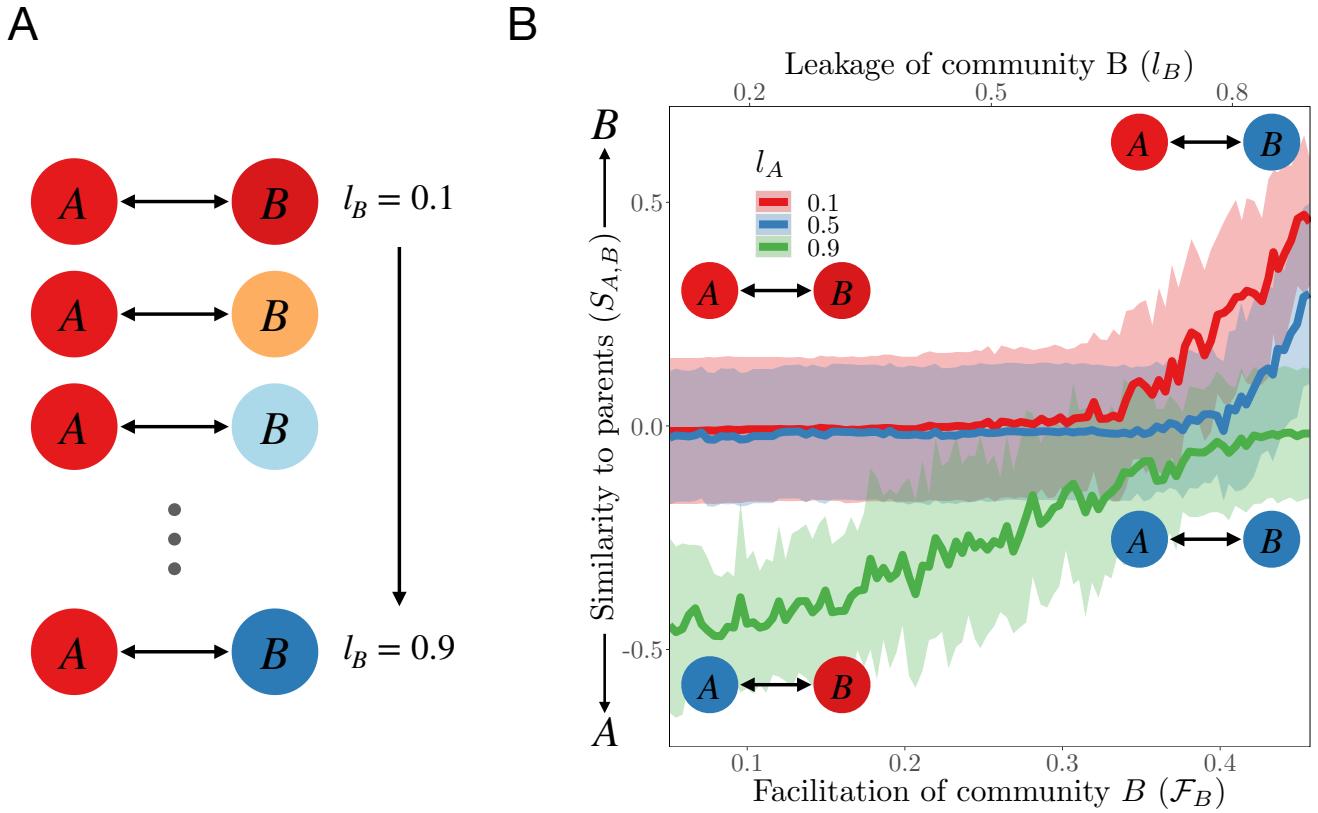


Fig 4. Recursive coalescence between microbial communities. **A:** Sketch of the simulation set up. The same pair of communities (A, B) is recursively coalesced, with the leakage of B gradually increasing after each coalescence event for three levels of leakage of A , and 25 replicates per l_A value. **B:** Parent community dominance after coalescence between communities A and B , as a function of facilitation level of community B , \mathcal{F}_B (bottom x axis), and leakage of community B , l_B (top x axis). Each curve corresponds to a different value of l_A . Shaded regions are $\pm\sigma$. Dominance of parent community B after coalescence increases with l_B , implying that higher cooperation levels enhance coalescence success.

Community evolution under repeated coalescence events

Figure 5 shows that on average, competition level significantly reduces and facilitation level increases during repeated coalescence events. Along with this, the average maintenance cost of species present in the resident community decreases with the number of coalescence exposures, and so does average resource abundance at equilibrium (Fig 5C and D), indicating that resource depletion ability improves in the process. In addition, the sub-population of resource specialists (that consume only one resource) increases with the number of coalescence events, while the rest of the species groups decrease in abundance (Fig 5E). Finally, the number of successful invasions into the resident community decreases function of number of coalescence events, while its species richness increases (Fig 5F). Taken together, these results show that communities composed of non-competing specialists that cooperate among themselves (Fig 5B and E) and reduce their respective metabolic costs (Fig 5C), improve their overall resource depletion ability (Fig 5D). This, in turn, makes them more resistant to multi-species invasions and therefore more successful in pairwise coalescence events (Fig 5F).

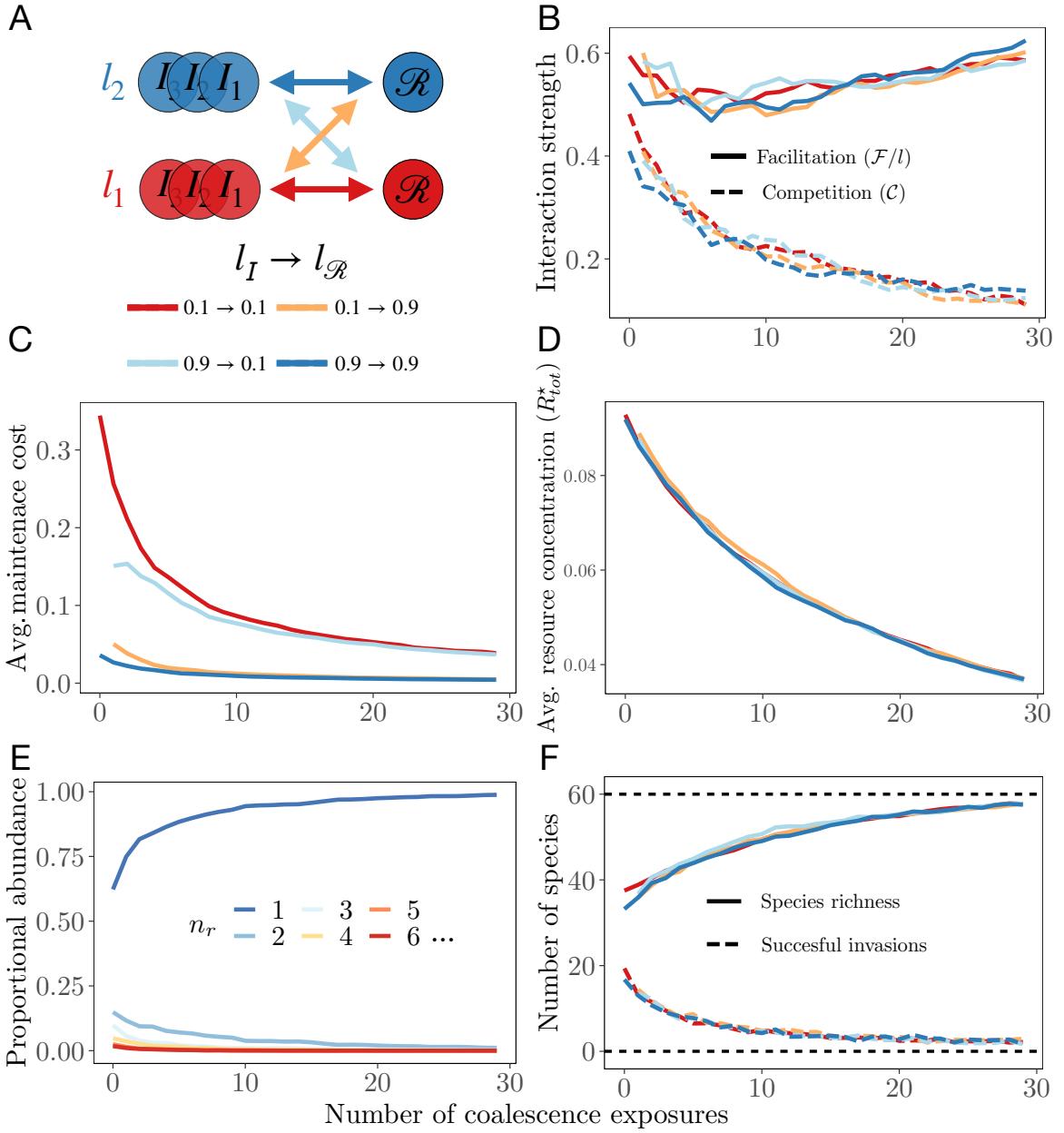


Fig 5. Serial coalescence of microbial communities. **A:** Sketch of the simulation set up. Resident communities (\mathcal{R} , upper circles) with leakage $l_{\mathcal{R}}$ are successively coalesced with randomly sampled invader communities (\mathcal{I} , lower circles) with leakage $l_{\mathcal{I}}$ for all possible combinations of leakage values (arrows) $l_{\mathcal{I}} = l_{\mathcal{R}} = [0.1, 0.9]$. For each serial coalescence sequence, we examine as a function of number of coalescence events, the following community properties of \mathcal{R} : **(B)** community-level competition (\mathcal{C} , dashed lines) and facilitation (\mathcal{F} , solid lines); **(C)** average species maintenance cost; **(D)** average resource concentration at equilibrium; **(E)** abundance fraction of each n_r -preference species group; and **(F)** number of successful invasions and with species richness. All the measures are averaged across 20 replicates. The standard deviation is decreasing along the x-axis and is never more than 40% of the mean for all the curves (not shown to reduce clutter). Abundance fraction of species with $n_r > 5$ was negligible and is not plotted for clarity.

Discussion

262

Our findings offer new mechanistic insights in the dynamics and outcomes of microbial community coalescence by explicitly considering the balance between competition and cooperation; two key interactions of real microbial communities [26, 37]. Specifically, we find that communities harbouring less competing and more cooperative species (that is, having lesser net competition) dominate after coalescence because they are better at depleting resources and resisting invasions. Therefore, when a community undergoes a series of coalescence events, its competitiveness decreases and cooperativeness increases, along with its species richness, resource use efficiency, and resistance to invasions. These results provide a theoretical foundation for hypotheses suggested recently [18, 21], and mechanistic insights into empirical studies that have demonstrated the importance of cross-feeding interactions on community coalescence [22].

Our result based on coalescence between pairs of random communities at very low leakage (black line in Fig 3C) essentially extends the results of [25] to communities with both competitive and cooperative interactions. Tikhonov showed that coalescence success is predicted by minimizing community-level competition through the optimisation of resource niche partitioning, which also guarantees maximization of resource depletion efficacy. Here we show that, similarly, the successful community is the one that achieves lower *net* competition ($\mathcal{C} - \mathcal{F}$), which also predicts community-level resource depletion efficacy as well as species richness (Fig 3D). Thus, simultaneously reducing competition and increasing cooperation together drives the outcome of community coalescence. Therefore, the quantity $-(\mathcal{C} - \mathcal{F}) = \mathcal{F} - \mathcal{C}$ is also a measure of the “cohesiveness” of a microbial community. However, we also find that at extreme value of leakage ($l = 0.9$), there is a critical level of net competition difference beyond which coalescence success decreases again (yellow line in Fig 3C). This suggests that in the regime of high cooperation and competition, (high leakage, and tail ends of the curve) facilitative links in fact become detrimental. A similar result has been reported in [65]. This critical value is not seen when the cost function does not include leakage (Fig. S2). Interestingly, we also find that this phenomenon is very weak when biologically-realistic guild structure is present (Fig. S6). These effects of extreme leakage (and facilitation) on coalescence success cannot be predicted by our model analyses (Supporting text section 1), and merit further investigation in future research, provided such high leakage levels are biologically feasible.

In our model systems, species compete not only for resources leaked by other species, but also for resources leaked by themselves, i.e., species may leak metabolic by-products that are also encoded in their consumer preferences vector. Leakage of metabolic resources is a pervasive phenomenon in the microbial world [38, 39], and has been shown to exist also in resources necessary for growth, even in situations when those essential metabolites are scarce [40, 41]. Although it may seem counter-intuitive for microbes to secrete metabolites essential for their own growth, such leakage can be advantageous, especially in bacteria, as “flux control” or growth-dilution mechanisms which provide short-term growth benefits in crowded environments [42, 43].

Our recursive coalescence simulations (Fig 4A) allowed us to establish that coalescence success is enhanced by cooperative interactions. This result is consistent with past theoretical work showing that mutualistic interactions are expected to increase structural stability by decreasing effective competition [44]. It is also consistent with recent theoretical results on single species invasions in microbial communities [45]. Nonetheless, this finding hinges on our choice of the cost function (Eq (2); Supporting text section 1). This cost function, which was motivated by biological considerations, imposed an efficiency cost to species with lower leakage, ensuring that all consumers, independently of their leakage fraction, depleted resources to the same concentration on average (see Supporting text section 1). This allowed us to perform coalescence events

between communities harbouring species with different leakage without introducing a bias towards the more efficient species. This choice corresponds biologically to the interpretation of leakage as an efficiency factor in the conversion from energy to biomass (Eq S2 in Supplementary information). As a consequence, higher leakage species reach, in general, lower abundances at equilibrium [46].

Our findings about the evolution of community-level properties in response to repeated community-community encounters (Fig 5) suggest that it might be possible to identify functional groups of microbes or microbial traits that are a “smoking gun” of past coalescence events experienced by a given community [18]. Additionally, our finding that members of communities with a history of coalescence are likely to become increasingly resistant to further community invasions suggests a novel and potentially economical way to assemble robust microbial communities. We also found that repeated coalescence events contributed to increase species richness, offering another mechanism that may help explain differences in microbial diversity across locations and environments [18].

Our finding that resident communities exposed to repeated community invasions were mainly composed of cooperative specialists (Fig 5B and E) is due to our assumption that all resources were supplied, and at a fixed rate. This allowed specialists to survive because their only source of energy was always provided. This property may not be as commonly seen in real communities, where fluctuations in resource supply are common. Ignoring environmental fluctuations allowed us to focus on coalescence outcomes in terms of the species interaction structure alone. While this assumption may be sensible in some cases [47], it is an oversimplification in others [33]. Therefore, studying the complex interplay between biotic interactions and environmental factors, e.g., by allowing substrate diversification from a single supplied resource [6, 7], or perturbing the supply vector periodically to simulate some form of seasonality, is a promising direction for future research. In such cases, we expect a more balanced mix of generalists and specialists, such that only the competitive interactions necessary to diversify the available carbon sources will persist upon coalescence events, but above that threshold, the results presented here (Figs 3, 4, and 5) would be recovered.

Assuming a core leakage and metabolism common to the whole community, made the assembly dynamics computationally tractable, while ensuring that the system was not far away from the conditions of real communities [6, 27]. The assumption of common leakage was relaxed in the recursive coalescence procedure. In addition, both the community-wide fixed leakage and core metabolism assumptions were relaxed in the serial coalescence simulations, thus using the model in its fully general form as seen in Eqs 1. Finally, to retain an analytically tractable theoretical setting for an otherwise complex system, we assumed binary consumer preferences, linear consumer functional responses, and resources of equal energetic value throughout. While this is a promising avenue for future work, we expect our results to be qualitatively robust to relaxation of these constraints, based on recent work on microbial community assembly dynamics using the same general model [6, 27].

Encounters between microbial communities are becoming increasingly frequent [52], and mixing of whole microbial communities is gaining popularity for bio-engineering [53], soil restoration [54], faecal microbiota transplantation [55, 56], and the use of probiotics [57]. We present a framework which relates the structure of species interactions in microbial communities to the outcome of community coalescence events. Although more work is required to bridge the gap between theory and empirical observations, this study constitutes a key step in that direction.

Supporting information	363
Supporting text section 1 – Further details of the mathematical model.	364
Supporting text section 2 – Modulating net competition levels.	365
Supporting text section 3 – Further details of the community coalescence simulations.	366
Supporting text section 4 – Adding consumer guild structure.	368
Fig S1 Consequence of relaxing the assumption of leakage dependence.	369
Fig S2 Results of random coalescence procedure without imposing the efficiency cost (leakage dependence).	370
Fig S3 Eigenvalues of J when evaluated at equilibrium for two example communities.	372
Fig S4 Examples of differently-structured preference (<i>C</i>) and metabolic (<i>D</i>) matrices.	373
Fig S5 Auto-correlation of vector of species abundances in community B for consecutive re-assemblies of this community, along the studied leakage range.	375
Fig S6 Community coalescence with consumer guilds present.	377
Author Contributions	378
Conceptualization: Pablo Lechón, Tom Clegg, Jacob Cook, Thomas P. Smith, Samraat Pawar	379
Methodology: Pablo Lechón	381
Software and Visualizations: Pablo Lechón	382
Writing – Original Draft Preparation: Pablo Lechón	383
Writing – Review & Editing: Pablo Lechón, Tom Clegg, Jacob Cook, Thomas P. Smith, Samraat Pawar.	384
Funding Acquisition: Samraat Pawar	386

References

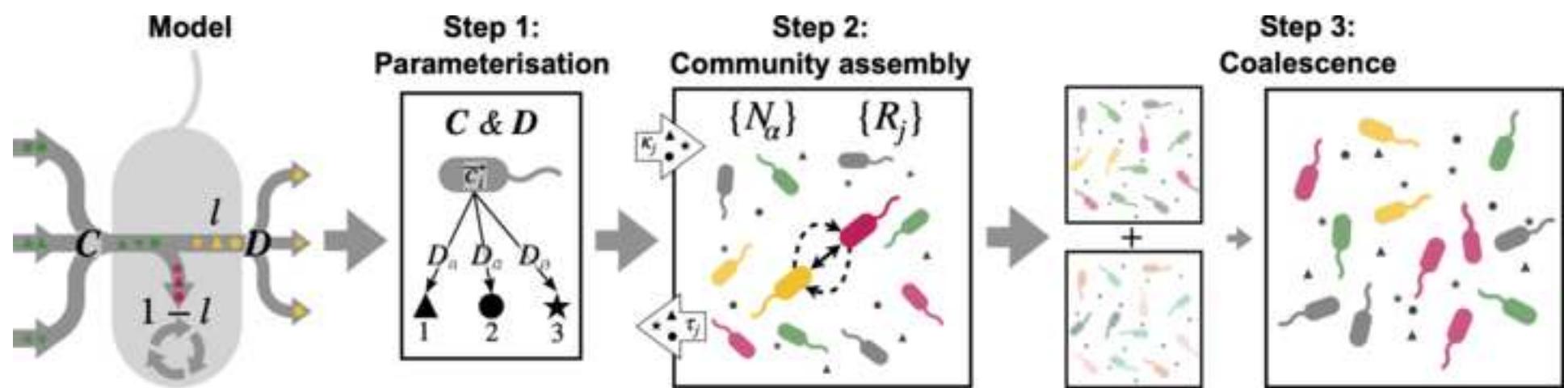
1. Fierer N, Jackson RB. The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America.* 2006;103(3):626–631. doi:10.1073/pnas.0507535103.
2. Huttenhower C, Gevers D, Knight R, Al E. Structure, function and diversity of the healthy human microbiome. *Nature.* 2012;486(7402):207–214. doi:10.1038/nature11234.
3. McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Lošo T, Douglas AE, et al. Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences of the United States of America.* 2013;110(9):3229–3236. doi:10.1073/pnas.1218525110.
4. Falkowski PG, Fenchel T, Delong EF. The microbial engines that drive earth's biogeochemical cycles. *Science.* 2008;320(5879):1034–1039. doi:10.1126/science.1153213.
5. Gilbert JA, Jansson JK, Knight R. The Earth Microbiome project: Successes and aspirations. *BMC Biology.* 2014;12(69):12915–014. doi:10.1186/s12915-014-0069-1.
6. Marsland R, Cui W, Goldford J, Sanchez A, Korolev K, Mehta P. Available energy fluxes drive a transition in the diversity, stability, and functional structure of microbial communities. *PLoS Computational Biology.* 2019;15(2):e1006793. doi: 10.1371/journal.pcbi.1006793. doi:10.1371/journal.pcbi.1006793.
7. Goldford JE, Lu N, Bajić D, Estrela S, Tikhonov M, Sanchez-Gorostiaga A, et al. Emergent simplicity in microbial community assembly. *Science.* 2018;361(6401):469–474. doi:10.1126/science.aat1168.
8. Goyal A, Maslov S. Diversity, Stability, and Reproducibility in Stochastically Assembled Microbial Ecosystems. *Physical Review Letters.* 2018;120(15):158102. doi:10.1103/PhysRevLett.120.158102.
9. Friedman J, Higgins LM, Gore J. Community structure follows simple assembly rules in microbial microcosms. *Nature Ecology and Evolution.* 2017;1(5):41559–017. doi:10.1038/s41559-017-0109.
10. Costello EK, Stagaman K, Dethlefsen L, Bohannan BJM, Relman DA. The application of ecological theory toward an understanding of the human microbiome. *Science.* 2012;336(6086):1255–1262. doi:10.1126/science.1224203.
11. Vila JCC, Jones ML, Patel M, Bell T, Rosindell J. Uncovering the rules of microbial community invasions. *Nature Ecology and Evolution.* 2019;3(8):1162–1171. doi:10.1038/s41559-019-0952-9.
12. Estrela S, Vila JCC, Lu N, Bajic D, Rebolleda-Gomez M, Chang CY, et al. Metabolic rules of microbial community assembly. *bioRxiv.* 2020. doi:10.1101/2020.03.09.984278.
13. Coyte KZ, Rao C, Rakoff-nahoum S, Foster KR. Ecological rules for the assembly of microbiome communities. *PLOS Biol.* 2021;19(2):e3001116. doi:10.1371/journal.pbio.3001116.

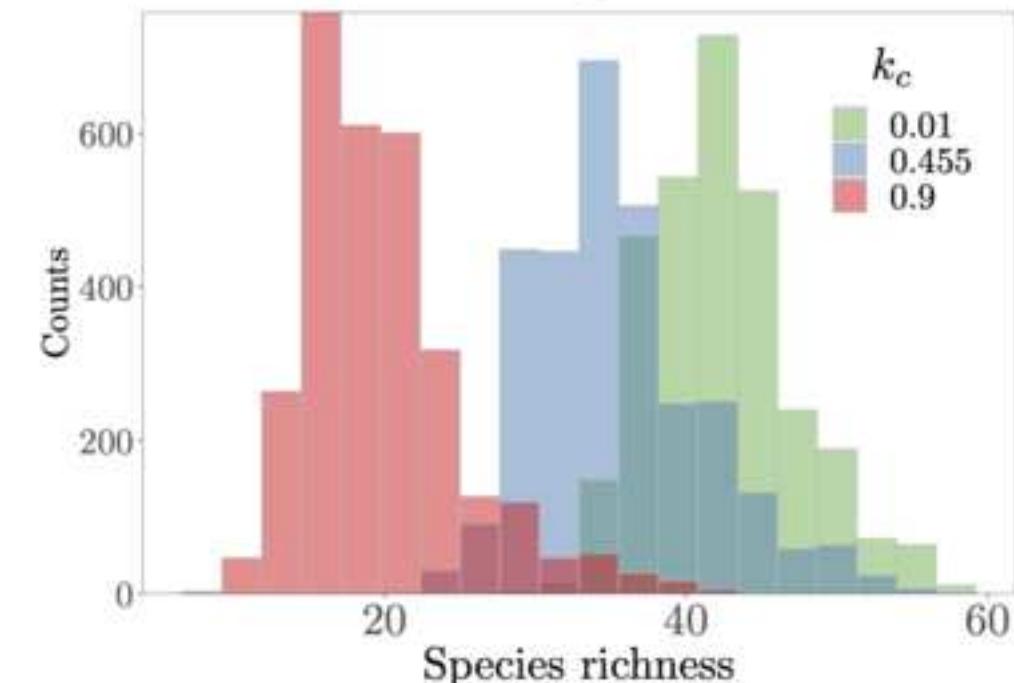
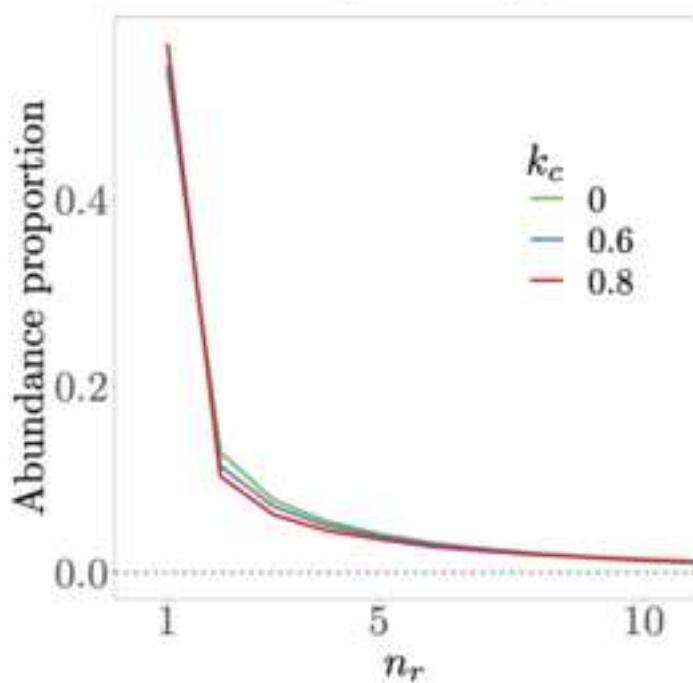
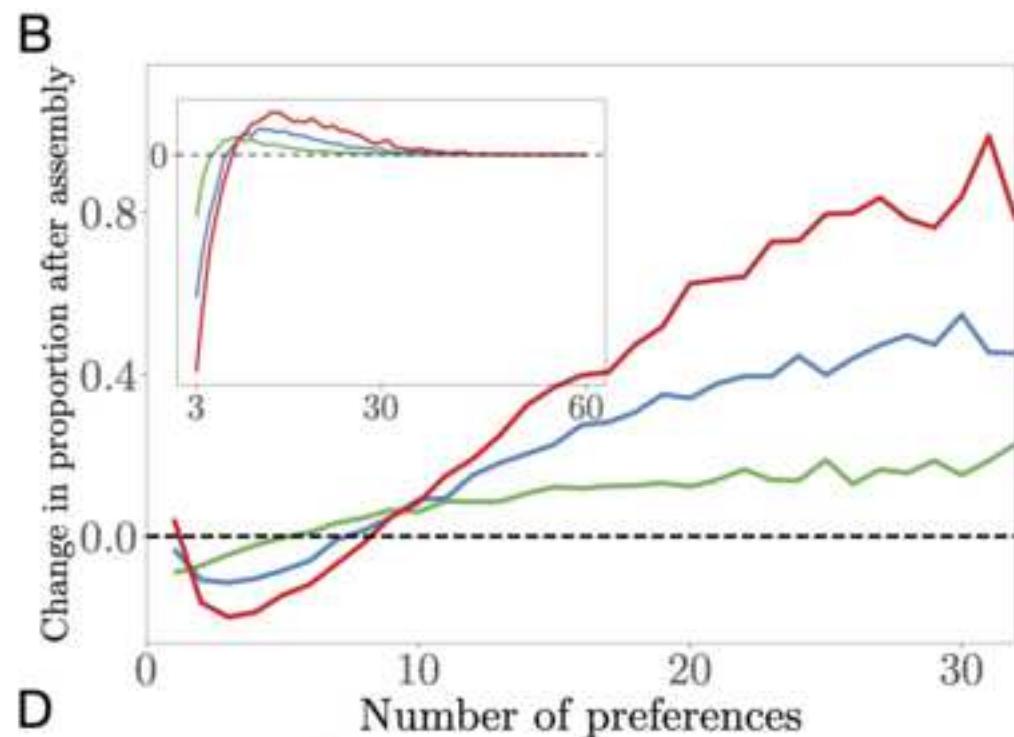
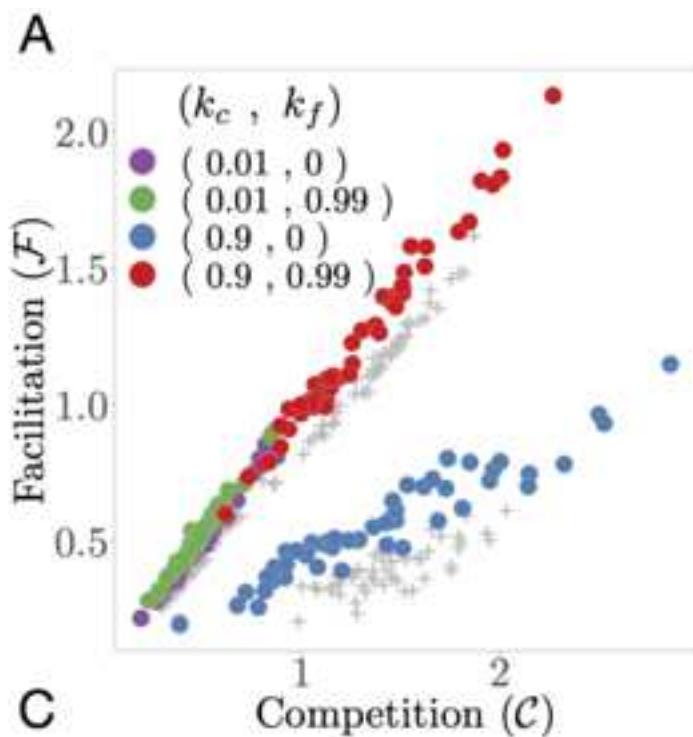
14. Luo X, Xiang X, Yang Y, Huang G, Fu K, Che R, et al. Seasonal effects of river flow on microbial community coalescence and diversity in a riverine network. *FEMS Microbiology Ecology*. 2020;96(8). doi:10.1093/femsec/fiaa132.
15. Vass M, Székely AJ, Lindström ES, Osman OA, Langenheder S. Warming mediates the resistance of aquatic bacteria to invasion during community coalescence. *Molecular Ecology*. 2021;doi:10.1111/mec.15800.
16. Kort R, Caspers M, van de Graaf A, van Egmond W, Keijser B, Roeselers G. Shaping the oral microbiota through intimate kissing. *Microbiome*. 2014;2(1):1–8. doi:10.1186/2049-2618-2-41.
17. Evans SE, Bell-Dereske LP, Dougherty KM, Kittredge HA. Dispersal alters soil microbial community response to drought. *Environmental Microbiology*. 2020;22(3):905–916. doi:10.1111/1462-2920.14707.
18. Rillig MC, Antonovics J, Caruso T, Lehmann A, Powell JR, Veresoglou SD, et al. Interchange of entire communities: Microbial community coalescence. *Trends in Ecology and Evolution*. 2015;30(8):470–476. doi:10.1016/j.tree.2015.06.004.
19. Rillig M, Tsang A, Roy J. Microbial community coalescence for microbiome engineering. *Front. Microbiol.* 2016;7:1967 doi: 10.3389/fmicb.2016.01967.
20. Rillig MC, Lehmann A, Aguilar-Trigueros CA. Soil microbes and community coalescence. *Pedobiologia*. 2016;59(1-2):37–40. doi:10.1016/j.pedobi.2016.01.001.
21. Castledine M, Sierociński P, Padfield D, Buckling A. Community coalescence: An eco-evolutionary perspective. *Philosophical transactions of the Royal Society of London Series B, Biological sciences*. 2020;375(1798):20190252. doi:10.1098/rstb.2019.0252.
22. Sierociński P, Milferstedt K, Bayer F, Großkopf T, Alston M, Bastkowski S, et al. A single community dominates structure and function of a mixture of multiple methanogenic communities. *Current Biology*. 2017;27(21):3390–3395. doi:10.1016/j.cub.2017.09.056.
23. Gilpin M. Community-level competition: Asymmetrical dominance. *Proceedings of the National Academy of Sciences of the United States of America*. 1994;91(8):3252–3254. doi:10.1073/pnas.91.8.3252.
24. Toquenaga Y. Historicity of a simple competition model. *Journal of Theoretical Biology*. 1997;187(2):175–181. doi:10.1006/jtbi.1997.0428.
25. Tikhonov M. Community-level cohesion without cooperation. *eLife*. 2016;5:e15747. doi: 10.7554/eLife.15747. doi:10.7554/eLife.15747.
26. Pascual-García A, Bonhoeffer S, Bell T. Metabolically cohesive microbial consortia and ecosystem functioning. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2020;375(1798):0190245. doi:10.1098/rstb.2019.0245.
27. Marsland, R., Cui, W. and Mehta, P. A minimal model for microbial biodiversity can reproduce experimentally observed ecological patterns.. *Sci Rep.* 2020;10, 3308. doi:doi.org/10.1038/s41598-020-60130-2.
28. Lu N, Sanchez-gorostiaga A, Tikhonov M, Sanchez A. Cohesiveness in microbial community coalescence. *bioRxiv*. 2018; p. 282723 doi:10.1101/282723.

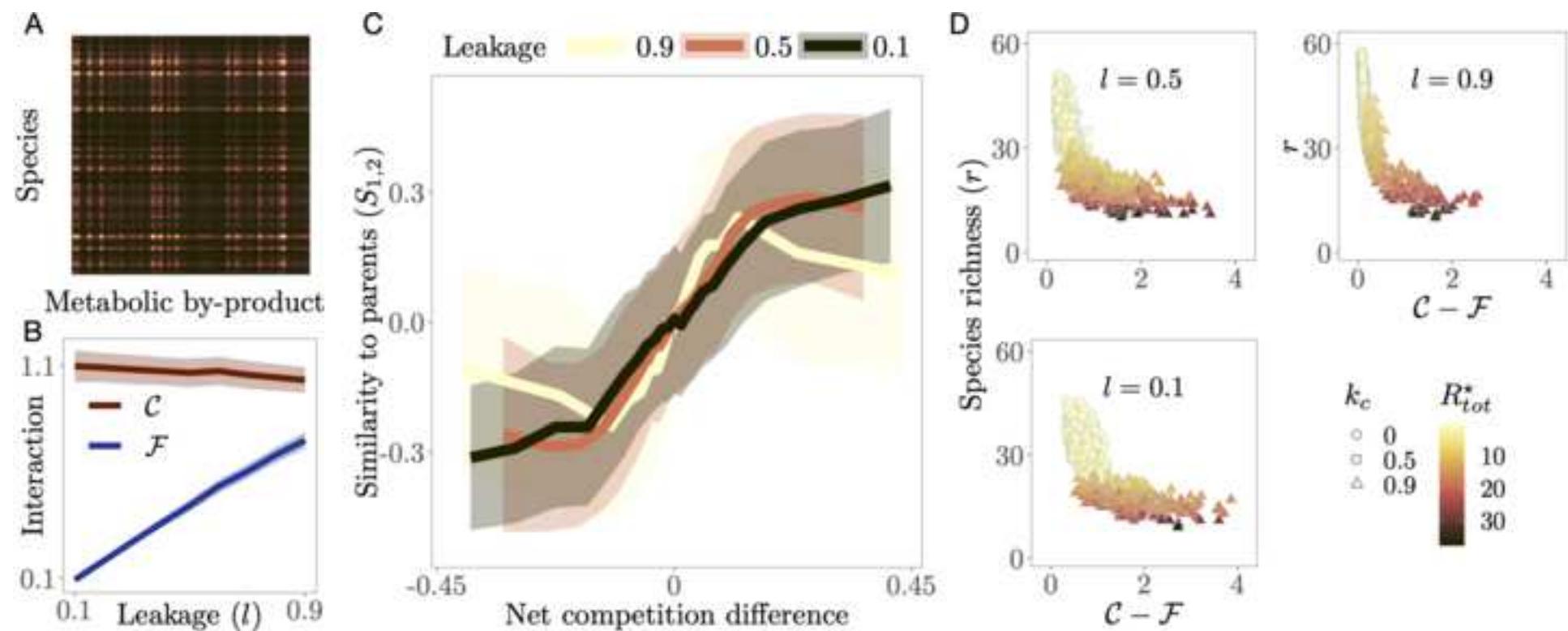
29. Hansen SK, Rainey PB, Haagensen JAJ, Molin S. Evolution of species interactions in a biofilm community. *Nature*. 2007;445(7127):533–536. doi:10.1038/nature05514.
30. Lawrence D, Fiegna F, Behrends V, Bundy JG, Phillimore AB, Bell T, et al. Species interactions alter evolutionary responses to a novel environment. *PLoS Biology*. 2012;10(5):e1001330. doi: 10.1371/journal.pbio.1001330. doi:10.1371/journal.pbio.1001330.
31. Embree M, Liu JK, Al-Bassam MM, Zengler K. Networks of energetic and metabolic interactions define dynamics in microbial communities. *Proceedings of the National Academy of Sciences of the United States of America*. 2015;112(50):15450–15455. doi:10.1073/pnas.1506034112.
32. Rivett DW, Jones ML, Ramoneda J, Mombrikotb SB, Ransome E, Bell T. Elevated success of multispecies bacterial invasions impacts community composition during ecological succession. *Ecology Letters*. 2018;21(4):516–524 doi:10.1111/ele.12916.
33. Albright MBN, Sevanto S, Gallegos-Graves LV, Dunbar J. Biotic interactions are more important than propagule pressure in microbial community invasions. *mBio*. 2020;11(5):1–16. doi:10.1128/mBio.02089-20.
34. Butler S, O'Dwyer JP. Stability criteria for complex microbial communities. *Nature Communications*. 2018;9(1): 2970. doi: 10.1038/s41467-018-05308-z.
35. Barabási A, Albert R. Emergence of scaling in random networks. *Science*. 1999;286(5439):509–512. doi:10.1126/science.286.5439.509.
36. Livingston G, Jiang Y, Fox JW, Leibold MA. The dynamics of community assembly under sudden mixing in experimental microcosms. *Ecology*. 2013;94(12):2898–2906. doi:10.1890/12-1993.1.
37. Machado D, Maistrenko OM, Andrejev S, Kim Y, Bork P, Patil KR, et al. Polarization of microbial communities between competitive and cooperative metabolism. *Nature Ecology and Evolution*. 2021;5(2):195–203. doi:10.1038/s41559-020-01353-4.
38. Morris JJ. Black Queen evolution: The role of leakiness in structuring microbial communities. *Trends in Genetics*. 2015;31(8):475–482. doi:10.1016/j.tig.2015.05.004.
39. Kallus Y, Miller JH, Libby E. Paradoxes in leaky microbial trade. *Nat. Comms.* 2017;8(1): 1361. doi: 10.1038/s41467-017-01628-8.
40. Silva LP, Northen TR. Exometabolomics and MSI: Deconstructing how cells interact to transform their small molecule environment. *Current Opinion in Biotechnology*. 2015;34:209–216. doi:10.1016/j.copbio.2015.03.015.
41. Paczia N, Nilgen A, Lehmann T, Gätgens J, Wiechert W, Noack S. Extensive exometabolome analysis reveals extended overflow metabolism in various microorganisms. *Microbial Cell Factories*. 2012;11:1–14. doi:10.1186/1475-2859-11-122.
42. Yamagishi JF, Saito N, Kaneko K. Advantage of Leakage of Essential Metabolites for Cells. *Physical Review Letters*. 2020;124:048101. doi:10.1103/PhysRevLett.124.048101.

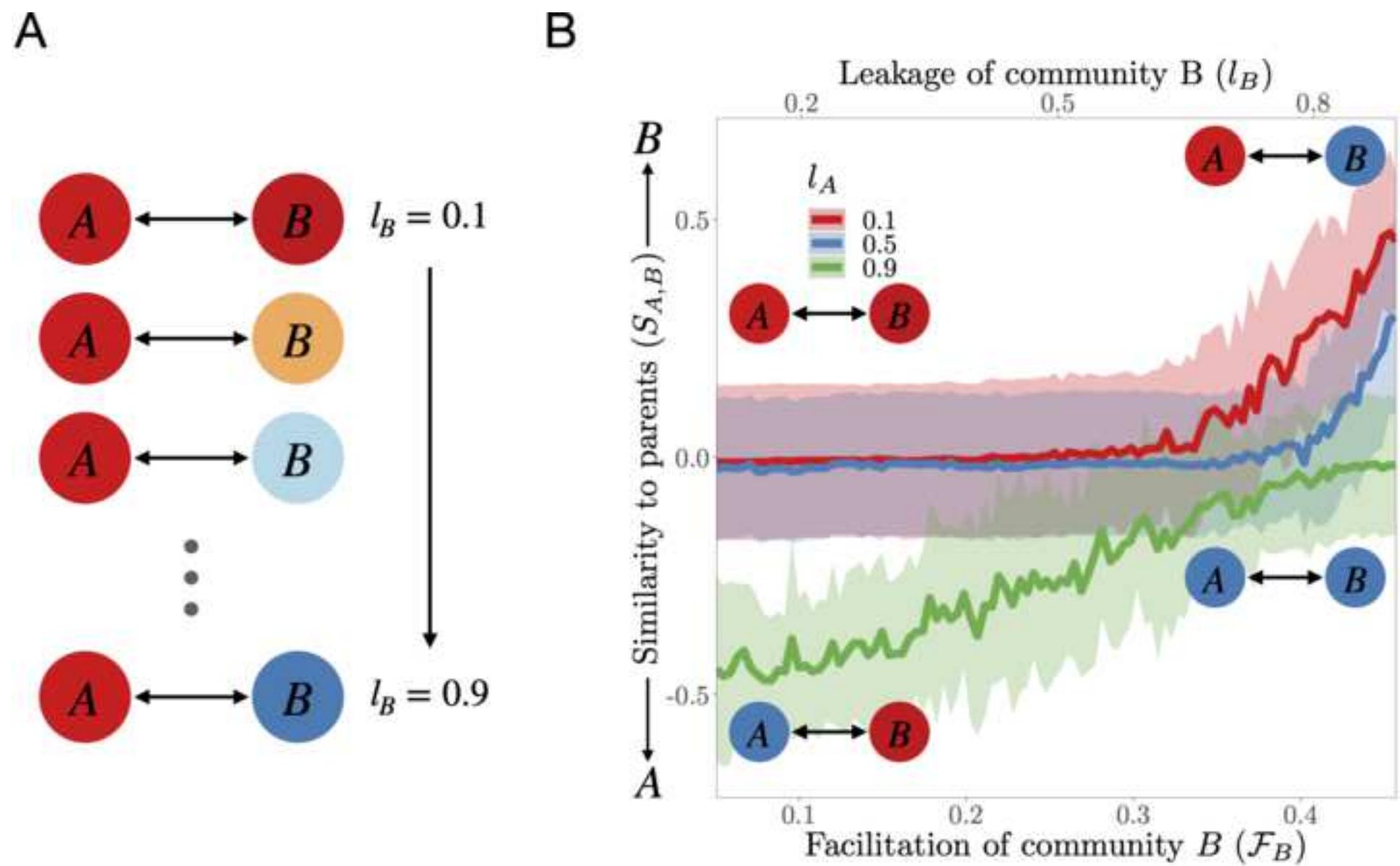
43. Yamagishi JF, Saito N, Kaneko K. Adaptation of metabolite leakiness leads to symbiotic chemical exchange and to a resilient microbial ecosystem. *PLOS Comp. Bio.* 2021;17(6): e1009143. doi: 10.1371/journal.pcbi.1009143.
44. Bastolla U, Fortuna MA, Pascual-García A, Ferrera A, Luque B, Bascompte J. The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature*. 2009;458(7241):1018–1020. doi: 10.1038/nature07950.
45. Kurkjian HM, Javad Akbari M, Momeni B. The impact of interactions on invasion and colonization resistance in microbial communities. *PLoS Computational Biology*. 2021;17(1):1–18. doi:10.1371/journal.pcbi.1008643.
46. Wieser W. Cost of growth in cells and organisms: General rules and comparative aspects. *Biol. Rev. Camb. Philos. Soc.* 1994;69(1):1–33. doi: 10.1111/j.1469-185X.1994.tb01484.x.
47. Acosta F, Zamor RM, Najar FZ, Roe BA, Hambright KD. Dynamics of an experimental microbial invasion. *Proceedings of the National Academy of Sciences of the United States of America*. 2015;112(37):11594–11599. doi:10.1073/pnas.1505204112.
48. Girvan MS, Campbell CD, Killham K, Prosser JI, Glover LA. Bacterial diversity promotes community stability and functional resilience after perturbation. *Environmental Microbiology*. 2005;7(3):301–313. doi:10.1111/j.1462-2920.2005.00695.x.
49. Eisenhauer N, Scheu S, Jousset A. Bacterial diversity stabilizes community productivity. *PLoS ONE*. 2012;7(3):1–5. doi:10.1371/journal.pone.0034517.
50. Maron PA, Sarr A, Kaisermann A, Lévéque J, Olivier M, Guigue J, et al. High Microbial Diversity Promotes Soil Ecosystem Functioning. *2018;84:02738–17*. doi:10.1128/AEM.02738-17.
51. Ratzke C, Barrere J, Gore J. Strength of species interactions determines biodiversity and stability in microbial communities. *Nature Ecology and Evolution*. 2020;4(3):376–383. doi:10.1038/s41559-020-1099-4.
52. Seebens H, Blackburn TM, Dyer EE, Genovesi P, Hulme PE, Jeschke JM, et al. No saturation in the accumulation of alien species worldwide. *Nature Communications*. 2017;8:1–9. doi:10.1038/ncomms14435.
53. Rillig MC, Tsang A, Roy J. Microbial community coalescence for microbiome engineering. *Frontiers in Microbiology*. 2016;7:1967. doi:10.3389/fmicb.2016.01967.
54. Calderón K, Spor A, Breuil MC, Bru D, Bizouard F, Viole C, et al. Effectiveness of ecological rescue for altered soil microbial communities and functions. *ISME Journal*. 2017;11:272–283. doi:10.1038/ismej.2016.86.
55. Wang JW, Kuo CH, Kuo FC, Wang YK, Hsu WH, Yu FJ, et al. Fecal microbiota transplantation: Review and update. *Journal of the Formosan Medical Association*. 2019;118(1):S23–S31. doi:10.1016/j.jfma.2018.08.011.
56. Wilson BC, Vatanen T, Cutfield WS, O’Sullivan JM. The super-donor phenomenon in fecal microbiota transplantation. *Frontiers in Cellular and Infection Microbiology*. 2019;9(2): doi:10.3389/fcimb.2019.00002.

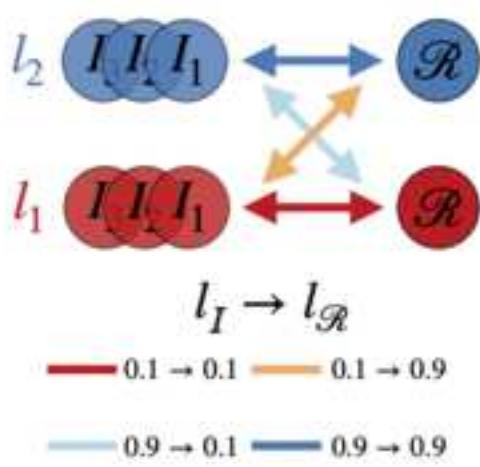
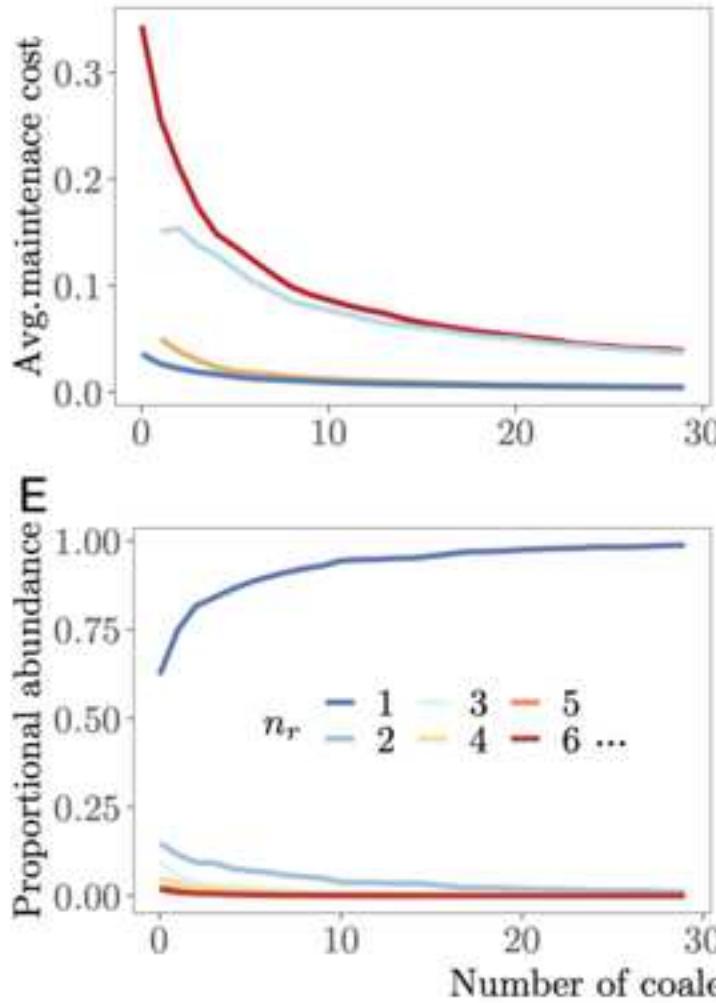
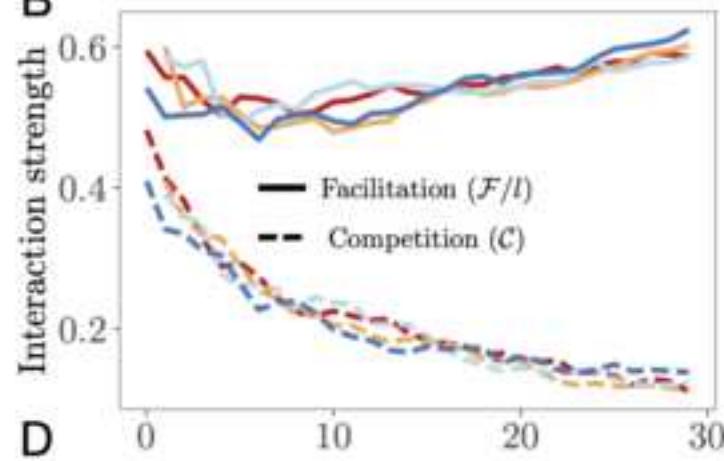
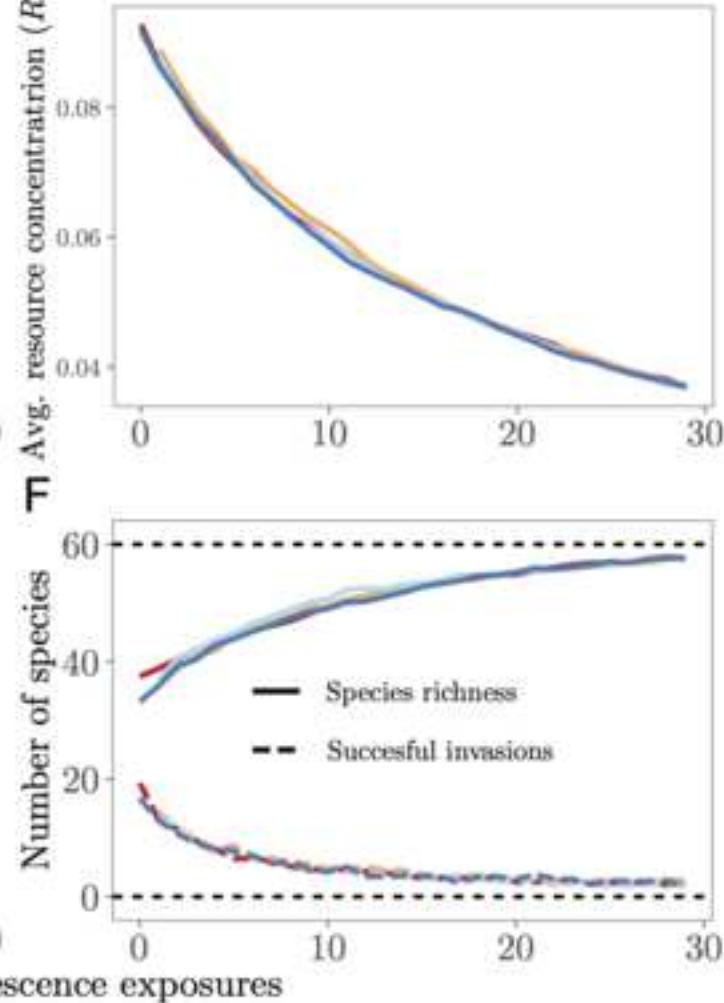
57. Lindemann SR, Bernstein HC, Song HS, Fredrickson JK, Fields MW, Shou W, et al. Engineering microbial consortia for controllable outputs. *ISME Journal*. 2016;10(9): 2077–2084. doi:10.1038/ismej.2016.26.
58. Louca S, Polz MF, Florent M, Albright MBN, Huber JHA, O'Connor MI, et al. Function and functional redundancy in microbial systems. *Nature Ecology and Evolution* 2018;2(6):936–943. doi:10.1038/s41559-018-0519-1.
59. Enke TN, Datta, MS, Schwartzman J, Cermak N, Schmitz D, Barrere J, Pascual-García A, Cordero OX. Modular Assembly of Polysaccharide-Degrading Marine Microbial Communities *Current Biology* 2019;29(9):1528–1535.e6 doi:10.1016/j.cub.2019.03.047.
60. Sung J, Kim S, Cabatbat JJT, Jang S, Jin Y, Jung GY, et al. Global metabolic interaction network of the human gut microbiota for context-specific community-scale analysis. *Nat. Commun.* 2017;8:15393. doi:10.1038/ncomms15393.
61. Tikhonov M, Monasson R. Collective phase in resource competition in a highly diverse ecosystem. *Physical Review Letters*. 2017;118:048103. doi:10.1103/PhysRevLett.118.048103.
62. Tikhonov M, Monasson R. Innovation rather than improvement: A solvable high-dimensional model highlights the limitations of scalar fitness. *J. Stat. Phys.* 2018;172:74–104. doi:10.1007/s10955-018-1956-6.
63. DeLong JP, Okie JG, Moses ME, Sibly RM, Brown JH. Shifts in metabolic scaling, production, and efficacy across major evolutionary transitions of life. *Proceedings of the National Academy of Sciences*. 2010;107(29): 12941–5. doi: 10.1073/pnas.1007783107.
64. Kempes CP, van Bodegom PM, Wolpert D, Libby E, Amend J, Hoehler T. Drivers of Bacterial Maintenance and Minimal Energy Requirements. *Front. Microbiol.* 2017;8:31. doi: 10.3389/fmicb.2017.00031.
65. Pascual-García A, Bastolla U. Mutualism supports biodiversity when the direct competition is weak. *Nature Communications*. 2017 8:14326. doi: 10.1038/ncomms14326.









A**C****B****D**



Click here to access/download
Supporting Information
SM.pdf

The role of competition versus cooperation in microbial community coalescence

Pablo Lechón^{1,2*}, Tom Clegg², Jacob Cook^{2,3}, Thomas P. Smith², Samraat Pawar²

1 Department of Ecology & Evolution, University of Chicago, Chicago, IL, USA.

2 Department of Life Sciences, Imperial College London, Silwood Park, Ascot, UK

3 Centre for Integrative Systems Biology and Bioinformatics, Imperial College London, UK

*plechon@uchicago.edu

Abstract

New microbial communities often arise through the mixing of two or more separately assembled parent communities, a phenomenon that has been termed “community coalescence”. Understanding how the interaction structures of complex parent communities determine the outcomes of coalescence events is an important challenge. While recent work has begun to elucidate the role of competition in coalescence, that of cooperation, a key interaction type commonly seen in microbial communities, is still largely unknown. Here, using a general consumer-resource model, we study the combined effects of competitive and cooperative interactions on the outcomes of coalescence events. In order to do so, we simulate coalescence events between pairs of communities with different degrees of competition for shared carbon resources and cooperation through cross-feeding on leaked metabolic by-products (facilitation). We also study how structural and functional properties of post-coalescence communities evolve when they are subjected to repeated coalescence events. We find that in coalescence events, the less competitive and more cooperative parent communities contribute a higher proportion of species to the new community, because this endows superior ability to deplete resources and resist invasions. Consequently, when a community is subjected to repeated coalescence events, it gradually evolves towards being less competitive and more cooperative, as well as more species rich, robust and efficient in resource use. Encounters between microbial communities are becoming increasingly frequent as a result of anthropogenic environmental change, and there is great interest in how the coalescence of microbial communities affects environmental and human health. Our study provides new insights into the mechanisms behind microbial community coalescence, and a framework to predict outcomes based on the interaction structures of parent communities.

Author summary

In nature, new microbial communities often arise from the fusion of whole, previously separate communities (community coalescence). Despite the crucial role that interactions among microbes play in the dynamics of complex communities, our ability to predict how these affect the outcomes of coalescence events remains limited. Here, using a general mathematical model, we study how the structure of species interactions confers an advantage upon a microbial community when it encounters another, and how

communities evolve after undergoing repeated coalescence events . We find that competitive interactions between species preclude their survival upon a coalescence event, while cooperative interactions are advantageous for post-coalescence survival. Furthermore, after a community is exposed to many coalescence events, the remaining species become less competitive and more cooperative. Ultimately, this drives the community evolution, yielding post-coalescence communities that are more species-rich, productive, and resistant to invasions. There are many potential environmental and health implications of microbial community coalescence, which will benefit from the theoretical insights that we offer here about the fundamental mechanisms underlying this phenomenon.

Introduction

Microbial communities are widespread throughout our planet [1], from the the human gut to the deep ocean, and play a critical role in natural processes ranging from animal development and host health [2,3] to biogeochemical cycles [4]. These communities are very complex, typically harbouring hundreds of species [5], making them hard to characterize. Recently, DNA sequencing has allowed high-resolution mapping of these communities, opening a niche for theoreticians and experimentalists to collaboratively decipher their complexity and assembly [6–11,13].

Entire microbial communities are often displaced over space and come into contact with each other due to physical (e.g., dispersal by wind or water) and biological (e.g., animal-animal or animal-plant interactions, and leaves falling to the ground) factors [14–17]. The process by which two or more communities that were previously separated join and reassemble into a new community has been termed community coalescence [18]. Although microbial community coalescence is likely to be common, the effects of both intrinsic and extrinsic factors on the outcomes of such events remains poorly understood [19]. Among extrinsic factors, resource availability, immigration rate of new species, and environmental conditions (especially, pH, temperature, and humidity) are likely to be crucial [20–22]. Among intrinsic factors, the role of functional and taxonomic composition and the inter-species interaction structures of parent communities are expected to be particularly important [20,21]. We focus on the role of species interactions on community coalescence in this study.

Early mathematical models suggested that in encounters between animal and plants communities, species in one community are more likely to drive those in the other extinct (community dominance) [23,24]. This was explained as being the result of the fact that communities are a non-random collection of species assembled through a shared history of competitive exclusion, and therefore act as coordinated entities. Recent theoretical work [25] has more rigorously established this for microbial community coalescence events, showing that the dominant community will be the one more capable of depleting all resources simultaneously. Overall, these findings suggest that communities arising from competitive species sorting exhibit sufficient “cohesion” to prevent invasions by members of other communities [26,27].

However, empirical support for the role of competition alone in coalescence outcomes is circumstantial, and the role of cooperation, which is commonly observed in microbial communities, is yet to be addressed theoretically. For example, during coalescence in methanogenic communities, “cohesive” units of taxa from the community with the most efficient resource use are co-selected [22]; and in aerobic bacterial communities, the invasion success of a given taxon is determined by its community members as a result of collective consumer-resource interactions and metabolic feedbacks between microbial growth and the environment [28]. Nonetheless, neither of these studies addressed the role of competition and cooperation in shaping coalescence success. Yet, these microbial

communities exhibit cooperation through a typically dense cross-feeding network, where leaked metabolic by-products of one species are shared as public goods across the entire community [29–31]. Indeed, several studies have suggested that a combination of competitive and cooperative interactions may determine the outcome of coalescence in microbial communities [21, 32, 33].

Here, we focus on the relative importance of competition and cooperation in community coalescence. We use a general consumer resource model that includes cross-feeding to assemble complex microbial communities having different degrees of competition and cooperation. We focus on determining the relative importance of the two types of interactions on outcomes of coalescence events, as well as the subsequent evolution of the structural and functional properties of coalesced communities.

Methods

Mathematical model

Our mathematical model for the microbial community dynamics is based on the work of Marsland et al. [6] (see Supporting text section 1, and Fig 1):

$$\begin{aligned} \frac{dN_\alpha}{dt} &= g_\alpha N_\alpha \left((1 - l_\alpha) \sum_j c_{\alpha j} R_j - z_\alpha \right), \\ \frac{dR_j}{dt} &= \kappa_j + \tau^{-1} R_j - \sum_\alpha N_\alpha c_{\alpha j} R_j + \sum_{\alpha k} N_\alpha l_\alpha D_{\alpha j k} c_{\alpha k} R_k. \end{aligned} \quad (1)$$

Here, N_α ($\alpha = 1, \dots, s$) and R_j ($j = 1, \dots, m$) are the biomass abundance of the α^{th} microbial (e.g., bacterial) species and the concentration of the j^{th} resource (e.g., carbon substrate). The growth of species α is determined by the resources it harvests minus the cost of maintenance (two terms in the brackets). Resource uptake depends on the resource concentration in the environment R_j , and on the uptake rate of species α , here assumed to be binary (j ($c_{\alpha j} = 1$ or $c_{\alpha j} = 0$)). The leakage term l_α determines the proportion of this uptake that species α releases back into the environment as metabolic by-products, with the remainder ($1 - l_\alpha$) being allocated to growth. The uptake that remains after subtracting a maintenance cost (z_α) is transformed into biomass with a proportionality constant of g_α , the value of which does not affect the results presented here.

The change in the concentration of resources in the environment (second line in Eq 1) is determined by four terms. The first and second terms represent the external supply and dilution of resource j , which give the rates at which the j^{th} resource enters and leaves the system. The third term is the uptake of the j^{th} resource from the environment, summed across all s consumers, and the fourth term represents resources entering the environmental pool via leakage of metabolic by-products. By-product leakage of species α is determined by the metabolic matrix D_α (or the “stoichiometric” matrix; [6]), with $D_{\alpha j k}$ representing the leaked proportion of resource j that is transformed into resource k by species α . Energy conservation dictates that D_α is a row stochastic matrix, meaning that its rows sum to 1. Note that in this model, rates of metabolic by-product formation are dependent on resource uptake (i.e., the amount of resource leaked into the environment depends on the amount being consumed). We use this specific structure as opposed to dependence of leakage directly on consumer biomass (e.g., [34], which would mean that the relative leakage to each resource type remains constant within a community), because it (i) more accurately reflects biological reality (microbes typically produce specific byproducts when feeding on specific

resources [7]) and (ii) allows greater variation in cross feeding between consumer pairs (as the contribution to leakage from each consumer is unique) allowing us to better explore the effects of interactions. Also, we allow species to leak metabolites that they also consume (we address this assumption further in the Discussion).

We define the consumer's maintenance cost to be:

$$z_\alpha = \chi_0(1 + \epsilon_\alpha)(1 - l_\alpha) \sum_j c_{\alpha j}, \quad (2)$$

where χ_0 is the average cost of being able to consume a given resource, the summation represents the total number of resources that species α is able to process, and ϵ_α is a small random fluctuation that introduces variation in the cost for species that have identical preferences. Eq 2 ensures that neither generalists nor specialists are systematically favoured during community assembly (by imposing a greater cost on species that consume a wider range of resources), and that all species are able to deplete resources to similar concentrations independently of their leakage level (see Supporting text section 1 for rationale, Fig S1 for results under different cost functions; and Discussion).

The above model entails the following assumptions: (i) all resources contain the same amount of energy (taken to be 1 for simplicity), (ii) a linear, non-saturating consumer functional response, (iii) binary consumer preferences (uptake rates), and (iv) an environment where all resources are externally supplied in equal amounts. We address the implications of these assumptions in the Discussion.

Parameter	Description	Value	Units
g_α	Biomass synthesised per unit energy harvested by species α	1	$B_a E_a^{-1}$
$c_{\alpha j}$	Uptake rate of metabolite j by species α	1/0	$s^{-1} B_a^{-1}$
D_{kj}	Fraction of metabolite k that's leaked in the form of j	[0, 1]	None
κ_j	Supply rate of metabolite j	2	mol s^{-1}
τ	Time scale of resource dilution	0.25	s
l_α	Fraction of energy intake leaked as metabolic byproducts	[0, 1)	None
χ_0	Average cost per metabolic pathway	0.1	E_a
ϵ_α	Random fluctuation in species α 's cost	$N(0, 0.1)$	None
s	Number of species in pre-assembly communities	60	-
m	Number of metabolites	60	-
k_c	Competition factor	[0,1)	-
k_f	Facilitation factor	[0,1]	-
K_c	Inter-guild competition factor	(0.1, 0.9)	-
K_f	Inter-guild facilitation factor	(0.1, 0.9)	-

Table 1. Table of parameters used in our model. The units E_a and B_a represent arbitrary energy and biomass units, respectively.

Competition and facilitation metrics

In the system of equations 1, competition for resources exists because all pairs of consumer species generally share some resource preferences (their metabolic preferences vectors are not orthogonal). We quantify the pairwise competition between a species pair (α, β) by counting the resource preferences they share through the scalar product of their preference vectors, $\mathbf{c}_\alpha \cdot \mathbf{c}_\beta$. Therefore, community-level competition (denoted as \mathcal{C}) can be calculated by taking the average of the competition matrix, which encodes the competition strengths between all species pairs, that is

$$\mathcal{C} = \langle CC^T \rangle, \quad (3)$$

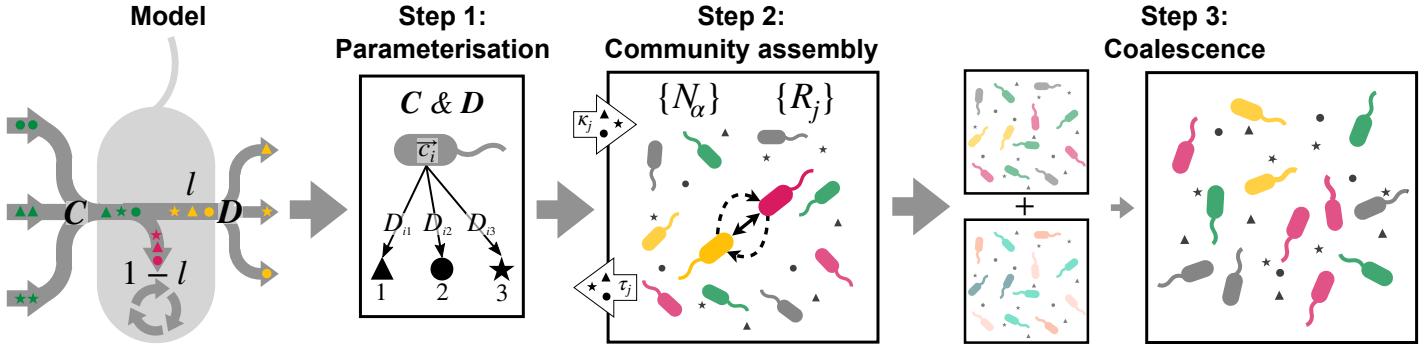


Fig 1. Overview of the coalescence modelling methodology. **Step 1.** The matrix of resource preferences (C) and the metabolic matrix (D) are sampled for each community. Black polygons are different resource types. **Step 2.** Dynamics of the system are allowed to play out (Eqs 1) until a locally stable equilibrium point is reached. Species composition and abundance, along with community-level competition \mathcal{C} (solid bidirectional arrows, Eq 3), and facilitation \mathcal{F} (dashed unidirectional arrows, Eq 4) are measured in assembled communities. **Step 3.** A pair of the assembled parent communities are mixed, and the resulting community integrated to steady state. For the random and recursive coalescence procedures, the contribution of each parent community to the final mix is analyzed ($S_{1,2}$, Eq 5) as a function of their interaction structures ($\mathcal{C}_{1,2}$ and $\mathcal{F}_{1,2}$) before they coalesced. In the case of the serial coalescence procedure, the properties of the resident community \mathcal{R} are tracked after each coalescence exposure.

where C is the $s \times m$ matrix of metabolic preferences of all the species in the community. 108

On the other hand, facilitation occurs when a species leaks metabolic by-products that are used by another species. We measure pairwise facilitation from species $\alpha \rightarrow \beta$ by calculating the fraction of secreted resources from species α that are consumed by species β per unit of resource abundance, $l_\alpha c_\alpha^T D_\alpha c_\beta$. Similar to competition, we compute community-level cooperation (denoted as \mathcal{F}), by taking the average of the facilitation matrix, which encodes the competition strengths between all species pairs, that is 109
110
111
112
113
114
115

$$\mathcal{F} = \langle \sum_{\alpha} \mathcal{D}(l) C D_{\alpha} C^T \rangle, \quad (4)$$

where where $\mathcal{D}(l)$ is a diagonal matrix with the leakage vector of each species in the community in its diagonal. 116
117

Henceforth, we refer to the quantity $\mathcal{C} - \mathcal{F}$ as “net competition”, which we later show is related to the “cohesion” defined in previous work [25]. 118
119

Simulations

In Fig 1 we present an overview of our simulations, which we now describe. For the parameter values used, see Table 1. 120
121
122

Step 1: Parameterization

We first set the parameters of the initial communities (before assembly) such that they span interactions across the spectrum of net competition ($\mathcal{C} - \mathcal{F}$). For each parent community, we modulate the structure of the C and D matrices (consisting of the resource preferences $c_{\alpha j}$'s and secretion proportions D_{jk} 's, respectively) by developing constrained random sampling procedures that guarantee specific levels of competition and facilitation at the community's steady state (see Supporting text section 2). In addition, we also add structure to C and D to emulate the existence of distinct resource classes and consumer guilds (see Supporting text section 4). With these procedures, net 123
124
125
126
127
128
129
130
131

competition in an assembled parent community can be regulated through four parameters: k_c (competition factor), k_f (facilitation factor), K_c (inter-guild competition factor), and K_f (inter-guild facilitation factor) (see Supporting text section 2). Note that we parameterize the initial communities by assuming (i) a shared core metabolism encoded in D , and (ii) a common leakage fraction l for all species (the implications of which we address in the Discussion), but we relax these assumptions in our coalescence simulations (Methods; Step 3).

Step 2: Assembly of parent communities

After parameterization, we numerically integrate Eqs 1 until steady state (a putative equilibrium) is reached. We perform 100 such assembly simulations with random sets of consumers for each combination of competition and facilitation factors (i.e., $k_c = k_f \in [0, 0.5, 0.9]$), repeating this for three values of leakage ($l \in [0.1, 0.5, 0.9]$). We compare species composition, abundances, and interaction structure of communities before and after assembly. In order to compare species composition, we calculate the difference between the proportion of species with a certain number of metabolic preferences, n_r (n-preference consumers, a measure of generalism), before and after assembly as

$$\Delta n_r = \frac{1}{p(n_r)} \left(\frac{T_{n_r}}{r} - \frac{T_{n_r}^0}{r^0} \right).$$

Here, $p(n_r)$ is the probability that a species has n_r metabolic preferences (Supporting text section 2), the 0 denotes before assembly, r is species richness, and T_{n_r} is the number of species with n_r preferences. Thus, when $\Delta n_r > 0$ the proportion of species with n_r metabolic preferences increases after assembly and vice versa. In order to analyze species abundances, we track the abundance fraction of consumers in each group of n-preference species, calculated as total abundance of the species in the group, divided by the total community biomass. Finally, we address the interaction structure after assembly by quantifying the levels of competition (\mathcal{C}) and facilitation (\mathcal{F}) in the assembled communities (Fig 2A).

Step 3: Coalescence

To simulate coalescence between a pair of assembled parent communities, we set all resources to their initial concentrations, and numerically integrate the new combined system to steady state. In order to disentangle the effects of competition versus cooperation and study the effect of repeated coalescence events, we simulate three coalescence scenarios: random, recursive, and serial, as follows (further details in Supporting text section 3).

Random coalescence. To address the effects of competition alone in the outcome of coalescence events, here we coalesce pairs of randomly sampled parent communities having the same leakage value l ($2 \cdot 10^4$ pairs for each leakage level, Fig 3C). That is, we fix the leakage level to ensure that the communities have, on average, similar cooperation levels, but leave k_c free to vary such that they span a broad range of competition levels.

Recursive coalescence. In order to study the effects of cooperation in particular on community coalescence, we repeatedly coalesce a given pair of communities A and B , slightly increasing the leakage of the latter in each iteration (Fig 4A). This allows us to modify the strength of cooperative interactions in the community, because facilitation is proportional to l (Eq 4), while keeping competition levels constant, because competition is independent of l (Eq 3) and the remaining parameters are kept fixed.

Serial coalescence. In the natural world, a community may be exposed to more than one coalescence event. Consequently, here we simulate a scenario where a local (“resident”) community \mathcal{R} harbouring species with leakage $l_{\mathcal{R}}$, and metabolism $D_{\mathcal{R}}$ is successively invaded by many other randomly sampled communities (“invaders”), \mathcal{I} with species of leakage $l_{\mathcal{I}}$ and metabolism $D_{\mathcal{I}}$ (Fig 5A). This allows us understand how the functional and structural properties of a microbial community evolve over time under successive encounters with other communities.

At the end of each random and recursive coalescence simulation, we quantify the dominance of either parent community in the post-coalescence community by measuring similarity of the latter to each of the two parents (indexed by 1 and 2) as:

$$S_{1,2} = \mathbf{p}_f \cdot \left(\frac{\mathbf{p}_2}{r_2} - \frac{\mathbf{p}_1}{r_1} \right), \quad (5)$$

where \mathbf{p}_f , \mathbf{p}_1 , and \mathbf{p}_2 are $(s_1 + s_2)$ -dimensional vectors of species presence-absence in the post-coalescent, and parent communities 1, and 2, respectively, with r_1 and r_2 the species richness values of the parent communities 1 and 2, respectively (calculated as $r_i = \sum p_i$). If $S_{1,2} = -1$, the coalesced community is identical to parent community 1, and if $S_{1,2} = 1$, it is identical to parent community 2. This measure is independent of the species richness. Thus we can mix communities with different species richness while avoiding a bias in similarity towards the richer one. We then analyze how this dominance measure depends on the interaction structure of the parent communities ($\mathcal{C}_{1,2}$ and $\mathcal{F}_{1,2}$; Eqs 3 and 4). After each coalescence event in the serial coalescence procedure, we measure competition and facilitation levels of the resident community, along with the average species maintenance cost, average resource abundance at equilibrium, species richness, and number of successful invasions, during the entire sequence of serial coalescence events. For all assembled parent as well as coalesced communities we confirmed that the steady state was a locally asymptotically stable equilibrium point (Supporting text section 1).

Results

Assembly of parent communities

In Fig 2 we show the key features of assembled communities. Figure 2A shows that as expected from Eqs 3 and 4, the levels of community-wide competition and facilitation are positively correlated, mediated by the structure of the C and D matrices. Figure 2B shows that the difference between the proportion of n-preference consumers before and after assembly (Δn_r , Methods; Step 1), increases for all simulated values of k_c , indicating that more generalist species are less prone to extinction during assembly. For the lowest value of k_c , Δn_r is in fact a monotonically increasing function of the number of preferences. This is expected because a species able to harvest energy from multiple resource pools is less likely to go extinct during community dynamics. As k_c increases, Δn_r reaches a minimum (Fig 2B), indicating that in more competitive environments pure specialists become more prevalent than moderate generalists. This is due to the fact that in a highly competitive environment the resource demands are concentrated on a subset of resources, while others are barely consumed (Supporting text section 2). In these communities, consumers that specialize exclusively on empty niches thrive. Figure 2C shows that specialist consumers are systematically present in higher abundance than generalists for all values of k_c . This is because several specialists are able to deplete all resources through their combined action more efficiently than one generalist [26], and as a result, although generalists are more persistent than specialists

upon assembly (Fig 2B), they achieve lower abundances at equilibrium (Fig 2C). In Figure 2B (inset), Δn_r is weighted by the abundance fraction (Methods; Step 2) of each group of n-preference consumers. This reveals an optimal group of consumers with a number of metabolic preferences that maximizes both survival probability and abundance at equilibrium. This optimal value increases for more competitive environments (as k_c increases). Finally, Figure 2D shows that more competitive communities tend to be less species rich, as expected from general competition theory.

223
224
225
226
227
228
229

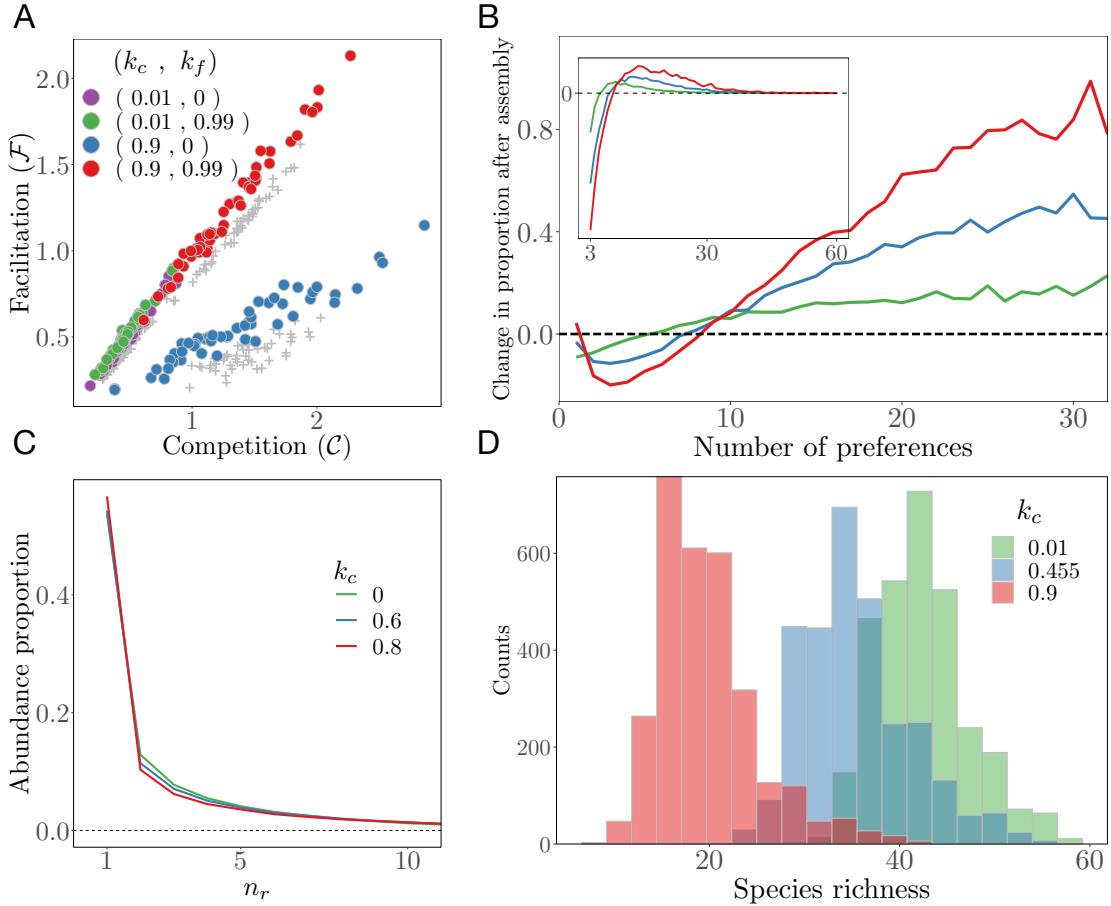


Fig 2. Features of assembled parent communities. **A:** Facilitation versus competition level in starting (grey dots) and assembled (coloured dots) communities for leakage $l = 0.9$ and different combinations of competition (k_c) and facilitation (k_f) factors. Communities assembled for each pair of $[k_c, k_f]$ values have the same colour. The assembled communities are always significantly more cooperative at the end of the assembly than at the start. **B:** Change in proportion of species in each n-preference consumer group (Δn_r , Methods; Step 2) before and after assembly, for different values of k_c (legend in panel C) indicating that more generalist species are less prone to extinction during assembly. Values for $n_r > 30$ had too much uncertainty due to low sampling and therefore have been removed for clarity. **Inset:** Δn_r is weighted by the abundance fraction at equilibrium of each n-preference consumers group. **C:** Abundance fraction of the n-preference species groups for different values of k_c . **D:** Distributions of species richness values of parent communities assembled under different k_c values. Increasing competitiveness tends to decrease species richness.

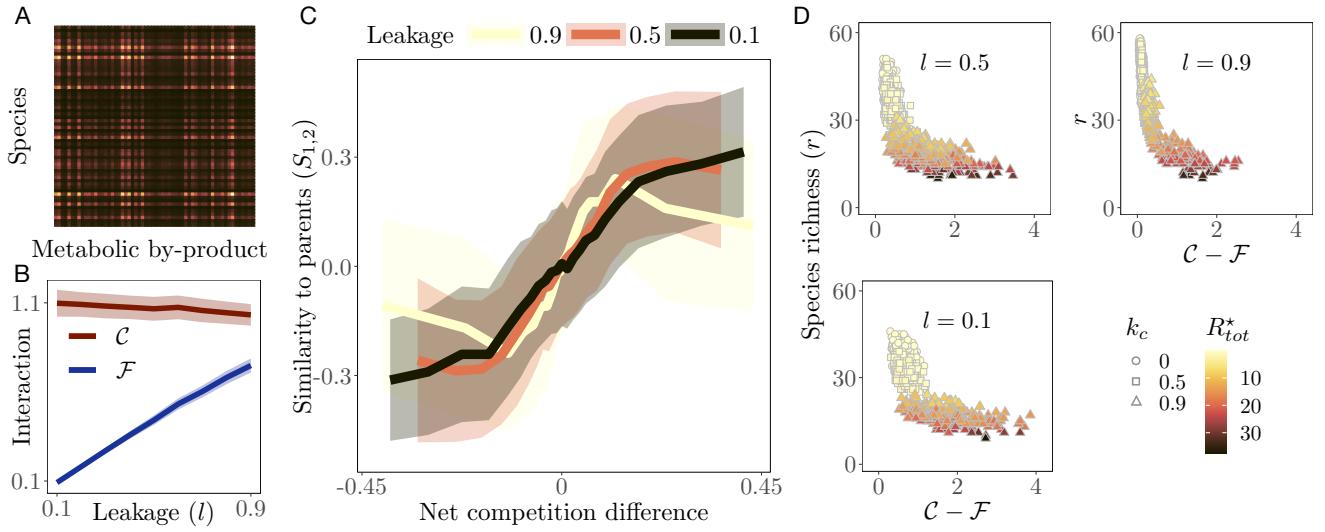


Fig 3. Community coalescence between pairs of randomly picked communities with same leakage.

A: Example of the secretion matrix with elements $(CD)_{\alpha k}$ representing the total leakage of resource k by species α . **B:** Community-level competition \mathcal{C} (dark red) and facilitation \mathcal{F} (blue) averaged across simulations for each leakage value. Since competition does not depend on the leakage, it remains consistently high throughout. Facilitation, on the other hand, increases linearly with leakage. **C:** Parent community dominance ($S_{1,2}$) as function of net competition difference $(\mathcal{C}_1 - \mathcal{F}_1) - (\mathcal{C}_2 - \mathcal{F}_2)$ (solid lines ± 1 standard deviation (shaded)), binned (20 bins) over communities with similar x axis values, for three community-wide leakage levels. The post-coalescence community is more similar to its less (net) competitive parent. **D:** Species richness (r) as a function of net competition in parent communities, coloured by total resource concentration at steady state (R_{tot}^*). The observed negative correlation for all values of leakage shows that communities with lower net competition tend to be more species-rich and also better at depleting resources (brighter coloured values, corresponding to lower levels of R_{tot}^* are scattered towards the top left of the plots).

Reducing competition increases coalescence success

Figure 3 shows that communities with lower net competition values tend to perform better in coalescence as seen by the positive relationship between parent community dominance ($S_{1,2}$) and the quantity $(\mathcal{C}_1 - \mathcal{F}_1) - (\mathcal{C}_2 - \mathcal{F}_2)$ (Fig 3C). That is, communities that emerge following coalescence tend to have greater similarity with the less net competitive parent. This trend holds at higher values of leakage, where cooperation levels are significant (Fig 3B), but with a clear critical point (the yellow line reverses in direction at a value of effective competition difference). This pattern is driven by the fact that less competitive parent communities deplete resources more efficiently and achieve a higher species richness (Fig 3D; Supporting text section 1). All these results also qualitatively hold for microbial communities that have consumer guild structure (Supporting text section 4).

Cooperation further enhances coalescence success

Figure 4 shows that when a community (B) whose leakage fraction increases successively during recursive coalescence events with another (A) with a fixed leakage level, the former becomes increasingly dominant. The result is consistent for a range of leakage values of community A . This shows that increasing cooperation levels enhance coalescence success.

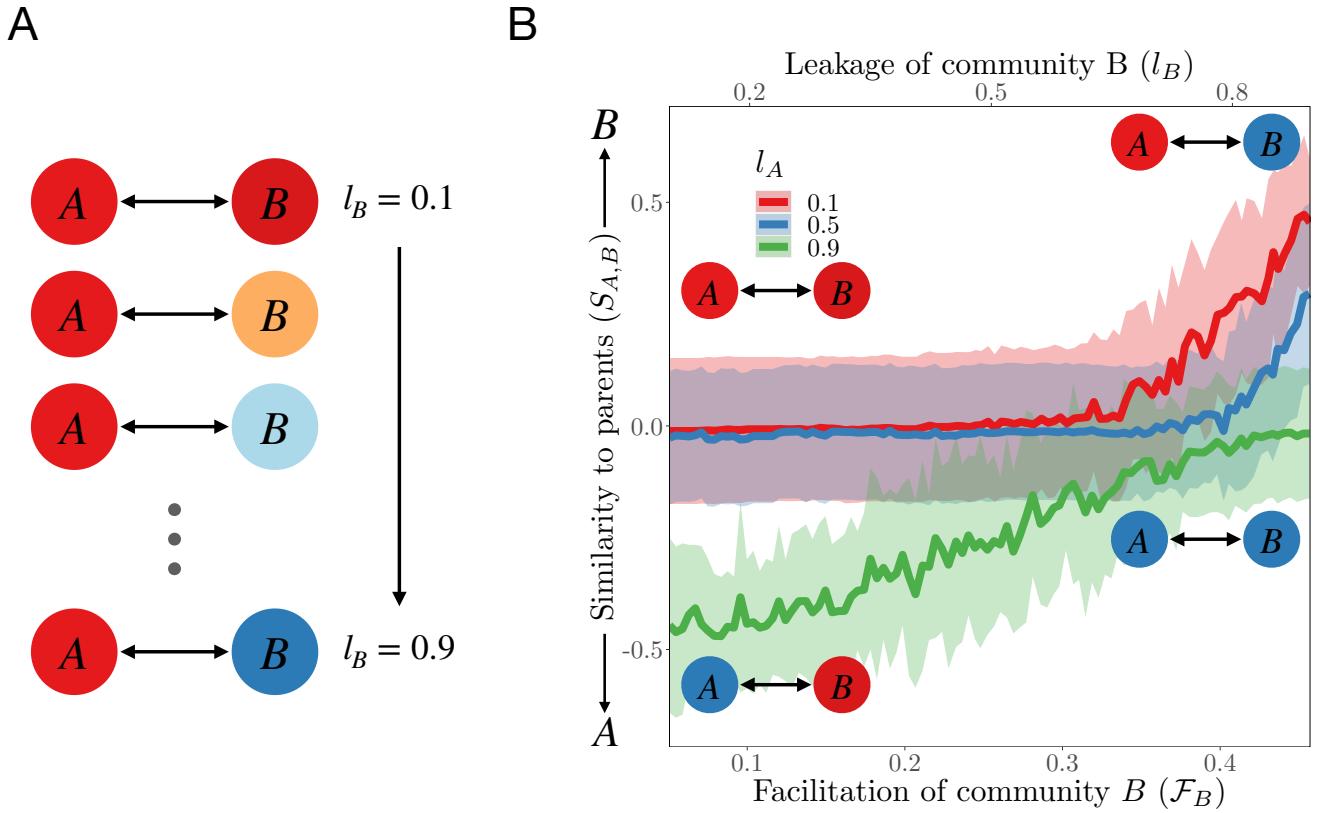


Fig 4. Recursive coalescence between microbial communities. **A:** Sketch of the simulation set up. The same pair of communities (A, B) is recursively coalesced, with the leakage of B gradually increasing after each coalescence event for three levels of leakage of A , and 25 replicates per l_A value. **B:** Parent community dominance after coalescence between communities A and B , as a function of facilitation level of community B , \mathcal{F}_B (bottom x axis), and leakage of community B , l_B (top x axis). Each curve corresponds to a different value of l_A . Shaded regions are $\pm\sigma$. Dominance of parent community B after coalescence increases with l_B , implying that higher cooperation levels enhance coalescence success.

Community evolution under repeated coalescence events

Figure 5 shows that on average, competition level significantly reduces and facilitation level increases during repeated coalescence events. Along with this, the average maintenance cost of species present in the resident community decreases with the number of coalescence exposures, and so does average resource abundance at equilibrium (Fig 5C and D), indicating that resource depletion ability improves in the process. In addition, the sub-population of resource specialists (that consume only one resource) increases with the number of coalescence events, while the rest of the species groups decrease in abundance (Fig 5E). Finally, the number of successful invasions into the resident community decreases function of number of coalescence events, while its species richness increases (Fig 5F). Taken together, these results show that communities composed of non-competing specialists that cooperate among themselves (Fig 5B and E) and reduce their respective metabolic costs (Fig 5C), improve their overall resource depletion ability (Fig 5D). This, in turn, makes them more resistant to multi-species invasions and therefore more successful in pairwise coalescence events (Fig 5F).

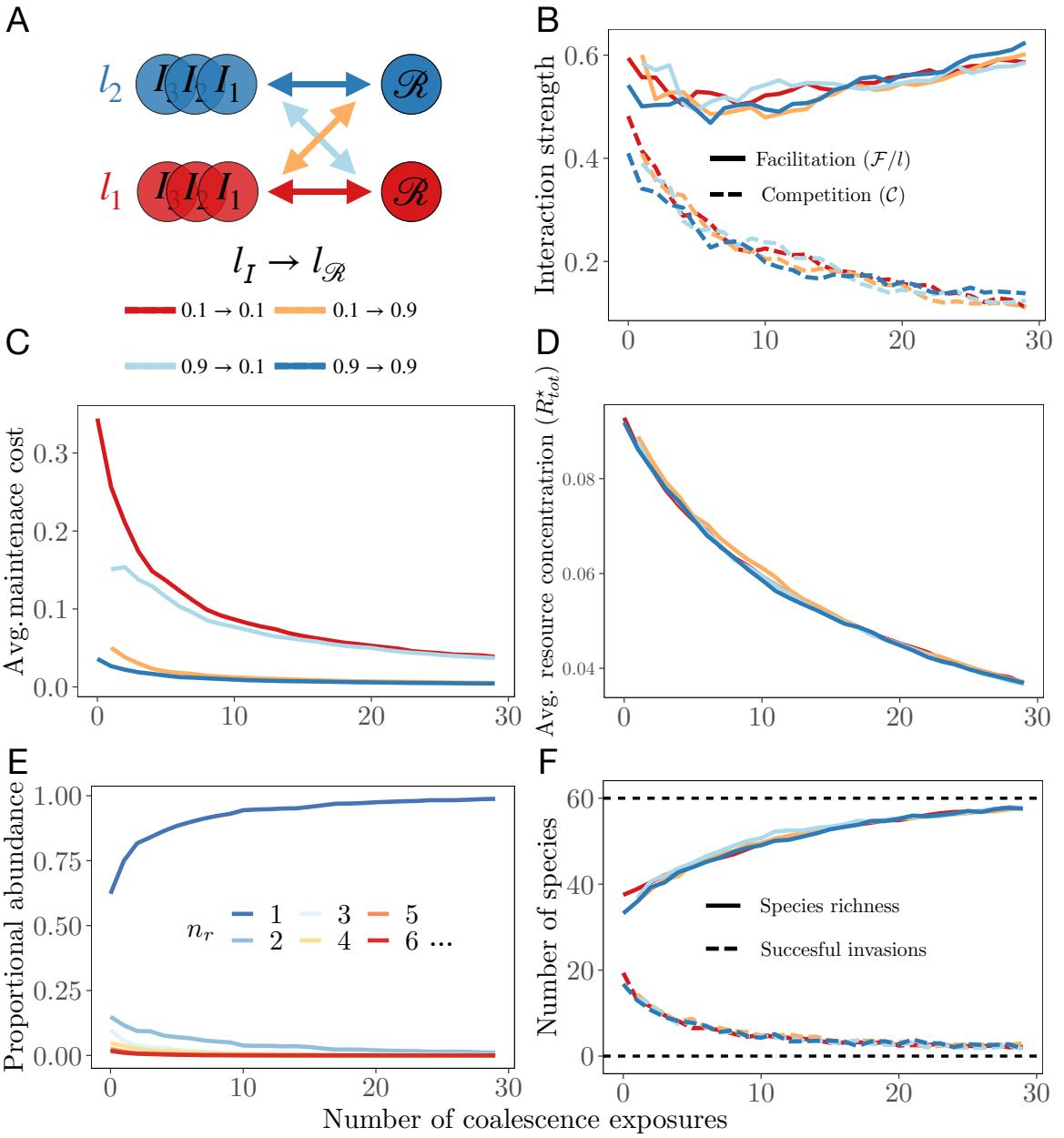


Fig 5. Serial coalescence of microbial communities. **A:** Sketch of the simulation set up. Resident communities (\mathcal{R} , upper circles) with leakage $l_{\mathcal{R}}$ are successively coalesced with randomly sampled invader communities (I , lower circles) with leakage l_I for all possible combinations of leakage values (arrows) $l_I = l_{\mathcal{R}} = [0.1, 0.9]$. For each serial coalescence sequence, we examine as a function of number of coalescence events, the following community properties of \mathcal{R} : **(B)** community-level competition (\mathcal{C} , dashed lines) and facilitation (\mathcal{F} , solid lines); **(C)** average species maintenance cost; **(D)** average resource concentration at equilibrium; **(E)** abundance fraction of each n -preference species group; and **(F)** number of successful invasions and with species richness. All the measures are averaged across 20 replicates. The standard deviation is decreasing along the x-axis and is never more than 40% of the mean for all the curves (not shown to reduce clutter). Abundance fraction of species with $n_r > 5$ was negligible and is not plotted for clarity.

Discussion

263

Our findings offer new mechanistic insights in the dynamics and outcomes of microbial community coalescence by explicitly considering the balance between competition and cooperation; two key interactions of real microbial communities [26, 37]. Specifically, we find that communities harbouring less competing and more cooperative species (that is, having lesser net competition) dominate after coalescence because they are better at depleting resources and resisting invasions. Therefore, when a community undergoes a series of coalescence events, its competitiveness decreases and cooperativeness increases, along with its species richness, resource use efficiency, and resistance to invasions. These results provide a theoretical foundation for hypotheses suggested recently [18, 21], and mechanistic insights into empirical studies that have demonstrated the importance of cross-feeding interactions on community coalescence [22].

Our result based on coalescence between pairs of random communities at very low leakage (black line in Fig 3C) essentially extends the results of [25] to communities with both competitive and cooperative interactions. Tikhonov showed that coalescence success is predicted by minimizing community-level competition through the optimisation of resource niche partitioning, which also guarantees maximization of resource depletion efficacy. Here we show that, similarly, the successful community is the one that achieves lower *net* competition ($\mathcal{C} - \mathcal{F}$), which also predicts community-level resource depletion efficacy as well as species richness (Fig 3D). Thus, simultaneously reducing competition and increasing cooperation together drives the outcome of community coalescence. Therefore, the quantity $-(\mathcal{C} - \mathcal{F}) = \mathcal{F} - \mathcal{C}$ is also a measure of the “cohesiveness” of a microbial community. However, we also find that at extreme value of leakage ($l = 0.9$), there is a critical level of net competition difference beyond which coalescence success decreases again (yellow line in Fig 3C). This suggests that in the regime of high cooperation and competition, (high leakage, and tail ends of the curve) facilitative links in fact become detrimental. A similar result has been reported in [65]. This critical value is not seen when the cost function does not include leakage (Fig. S2). Interestingly, we also find that this phenomenon is very weak when biologically-realistic guild structure is present (Fig. S6). These effects of extreme leakage (and facilitation) on coalescence success cannot be predicted by our model analyses (Supporting text section 1), and merit further investigation in future research, provided such high leakage levels are biologically feasible.

In our model systems, species compete not only for resources leaked by other species, but also for resources leaked by themselves, i.e., species may leak metabolic by-products that are also encoded in their consumer preferences vector. Leakage of metabolic resources is a pervasive phenomenon in the microbial world [38, 39], and has been shown to exist also in resources necessary for growth, even in situations when those essential metabolites are scarce [40, 41]. Although it may seem counter-intuitive for microbes to secrete metabolites essential for their own growth, such leakage can be advantageous, especially in bacteria, as “flux control” or growth-dilution mechanisms which provide short-term growth benefits in crowded environments [42, 43].

Our recursive coalescence simulations (Fig 4A) allowed us to establish that coalescence success is enhanced by cooperative interactions. This result is consistent with past theoretical work showing that mutualistic interactions are expected to increase structural stability by decreasing effective competition [44]. It is also consistent with recent theoretical results on single species invasions in microbial communities [45]. Nonetheless, this finding hinges on our choice of the cost function (Eq (2); Supporting text section 1). This cost function, which was motivated by biological considerations, imposed an efficiency cost to species with lower leakage, ensuring that all consumers, independently of their leakage fraction, depleted resources to the same concentration on average (see Supporting text section 1). This allowed us to perform coalescence events

between communities harbouring species with different leakage without introducing a bias towards the more efficient species. This choice corresponds biologically to the interpretation of leakage as an efficiency factor in the conversion from energy to biomass (Eq S2 in Supplementary information). As a consequence, higher leakage species reach, in general, lower abundances at equilibrium [46].

Our findings about the evolution of community-level properties in response to repeated community-community encounters (Fig 5) suggest that it might be possible to identify functional groups of microbes or microbial traits that are a “smoking gun” of past coalescence events experienced by a given community [18]. Additionally, our finding that members of communities with a history of coalescence are likely to become increasingly resistant to further community invasions suggests a novel and potentially economical way to assemble robust microbial communities. We also found that repeated coalescence events contributed to increase species richness, offering another mechanism that may help explain differences in microbial diversity across locations and environments [18].

Our finding that resident communities exposed to repeated community invasions were mainly composed of cooperative specialists (Fig 5B and E) is due to our assumption that all resources were supplied, and at a fixed rate. This allowed specialists to survive because their only source of energy was always provided. This property may not be as commonly seen in real communities, where fluctuations in resource supply are common. Ignoring environmental fluctuations allowed us to focus on coalescence outcomes in terms of the species interaction structure alone. While this assumption may be sensible in some cases [47], it is an oversimplification in others [33]. Therefore, studying the complex interplay between biotic interactions and environmental factors, e.g., by allowing substrate diversification from a single supplied resource [6, 7], or perturbing the supply vector periodically to simulate some form of seasonality, is a promising direction for future research. In such cases, we expect a more balanced mix of generalists and specialists, such that only the competitive interactions necessary to diversify the available carbon sources will persist upon coalescence events, but above that threshold, the results presented here (Figs 3, 4, and 5) would be recovered.

Assuming a core leakage and metabolism common to the whole community, made the assembly dynamics computationally tractable, while ensuring that the system was not far away from the conditions of real communities [6, 27]. The assumption of common leakage was relaxed in the recursive coalescence procedure. In addition, both the community-wide fixed leakage and core metabolism assumptions were relaxed in the serial coalescence simulations, thus using the model in its fully general form as seen in Eqs 1. Finally, to retain an analytically tractable theoretical setting for an otherwise complex system, we assumed binary consumer preferences, linear consumer functional responses, and resources of equal energetic value throughout. While this is a promising avenue for future work, we expect our results to be qualitatively robust to relaxation of these constraints, based on recent work on microbial community assembly dynamics using the same general model [6, 27].

Encounters between microbial communities are becoming increasingly frequent [52], and mixing of whole microbial communities is gaining popularity for bio-engineering [53], soil restoration [54], faecal microbiota transplantation [55, 56], and the use of probiotics [57]. We present a framework which relates the structure of species interactions in microbial communities to the outcome of community coalescence events. Although more work is required to bridge the gap between theory and empirical observations, this study constitutes a key step in that direction.

Supporting information	364
Supporting text section 1 – Further details of the mathematical model.	365
Supporting text section 2 – Modulating net competition levels.	366
Supporting text section 3 – Further details of the community coalescence simulations.	367
Supporting text section 4 – Adding consumer guild structure.	369
Fig S1 Consequence of relaxing the assumption of leakage dependence.	370
Fig S2 Results of random coalescence procedure without imposing the efficiency cost (leakage dependence).	371
Fig S3 Eigenvalues of J when evaluated at equilibrium for two example communities.	373
Fig S4 Examples of differently-structured preference (<i>C</i>) and metabolic (<i>D</i>) matrices.	374
Fig S5 Auto-correlation of vector of species abundances in community B for consecutive re-assemblies of this community, along the studied leakage range.	376
Fig S6 Community coalescence with consumer guilds present.	378
Author Contributions	379
Conceptualization: Pablo Lechón, Tom Clegg, Jacob Cook, Thomas P. Smith, Samraat Pawar	380
Methodology: Pablo Lechón	382
Software and Visualizations: Pablo Lechón	383
Writing – Original Draft Preparation: Pablo Lechón	384
Writing – Review & Editing: Pablo Lechón, Tom Clegg, Jacob Cook, Thomas P. Smith, Samraat Pawar.	385
Funding Acquisition: Samraat Pawar	386

References

1. Fierer N, Jackson RB. The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America.* 2006;103(3):626–631. doi:10.1073/pnas.0507535103.
2. Huttenhower C, Gevers D, Knight R, Al E. Structure, function and diversity of the healthy human microbiome. *Nature.* 2012;486(7402):207–214. doi:10.1038/nature11234.
3. McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Lošo T, Douglas AE, et al. Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences of the United States of America.* 2013;110(9):3229–3236. doi:10.1073/pnas.1218525110.
4. Falkowski PG, Fenchel T, Delong EF. The microbial engines that drive earth's biogeochemical cycles. *Science.* 2008;320(5879):1034–1039. doi:10.1126/science.1153213.
5. Gilbert JA, Jansson JK, Knight R. The Earth Microbiome project: Successes and aspirations. *BMC Biology.* 2014;12(69):12915–014. doi:10.1186/s12915-014-0069-1.
6. Marsland R, Cui W, Goldford J, Sanchez A, Korolev K, Mehta P. Available energy fluxes drive a transition in the diversity, stability, and functional structure of microbial communities. *PLoS Computational Biology.* 2019;15(2):e1006793. doi: 10.1371/journal.pcbi.1006793. doi:10.1371/journal.pcbi.1006793.
7. Goldford JE, Lu N, Bajić D, Estrela S, Tikhonov M, Sanchez-Gorostiaga A, et al. Emergent simplicity in microbial community assembly. *Science.* 2018;361(6401):469–474. doi:10.1126/science.aat1168.
8. Goyal A, Maslov S. Diversity, Stability, and Reproducibility in Stochastically Assembled Microbial Ecosystems. *Physical Review Letters.* 2018;120(15):158102. doi:10.1103/PhysRevLett.120.158102.
9. Friedman J, Higgins LM, Gore J. Community structure follows simple assembly rules in microbial microcosms. *Nature Ecology and Evolution.* 2017;1(5):41559–017. doi:10.1038/s41559-017-0109.
10. Costello EK, Stagaman K, Dethlefsen L, Bohannan BJM, Relman DA. The application of ecological theory toward an understanding of the human microbiome. *Science.* 2012;336(6086):1255–1262. doi:10.1126/science.1224203.
11. Vila JCC, Jones ML, Patel M, Bell T, Rosindell J. Uncovering the rules of microbial community invasions. *Nature Ecology and Evolution.* 2019;3(8):1162–1171. doi:10.1038/s41559-019-0952-9.
12. Estrela S, Vila JCC, Lu N, Bajic D, Rebolleda-Gomez M, Chang CY, et al. Metabolic rules of microbial community assembly. *bioRxiv.* 2020. doi:10.1101/2020.03.09.984278.
13. Coyte KZ, Rao C, Rakoff-nahoum S, Foster KR. Ecological rules for the assembly of microbiome communities. *PLOS Biol.* 2021;19(2):e3001116. doi:10.1371/journal.pbio.3001116.

14. Luo X, Xiang X, Yang Y, Huang G, Fu K, Che R, et al. Seasonal effects of river flow on microbial community coalescence and diversity in a riverine network. *FEMS Microbiology Ecology*. 2020;96(8). doi:10.1093/femsec/fiaa132.
15. Vass M, Székely AJ, Lindström ES, Osman OA, Langenheder S. Warming mediates the resistance of aquatic bacteria to invasion during community coalescence. *Molecular Ecology*. 2021;doi:10.1111/mec.15800.
16. Kort R, Caspers M, van de Graaf A, van Egmond W, Keijser B, Roeselers G. Shaping the oral microbiota through intimate kissing. *Microbiome*. 2014;2(1):1–8. doi:10.1186/2049-2618-2-41.
17. Evans SE, Bell-Dereske LP, Dougherty KM, Kittredge HA. Dispersal alters soil microbial community response to drought. *Environmental Microbiology*. 2020;22(3):905–916. doi:10.1111/1462-2920.14707.
18. Rillig MC, Antonovics J, Caruso T, Lehmann A, Powell JR, Veresoglou SD, et al. Interchange of entire communities: Microbial community coalescence. *Trends in Ecology and Evolution*. 2015;30(8):470–476. doi:10.1016/j.tree.2015.06.004.
19. Rillig M, Tsang A, Roy J. Microbial community coalescence for microbiome engineering. *Front. Microbiol.* 2016;7:1967 doi: 10.3389/fmicb.2016.01967.
20. Rillig MC, Lehmann A, Aguilar-Trigueros CA. Soil microbes and community coalescence. *Pedobiologia*. 2016;59(1-2):37–40. doi:10.1016/j.pedobi.2016.01.001.
21. Castledine M, Sierociński P, Padfield D, Buckling A. Community coalescence: An eco-evolutionary perspective. *Philosophical transactions of the Royal Society of London Series B, Biological sciences*. 2020;375(1798):20190252. doi:10.1098/rstb.2019.0252.
22. Sierociński P, Milferstedt K, Bayer F, Großkopf T, Alston M, Bastkowski S, et al. A single community dominates structure and function of a mixture of multiple methanogenic communities. *Current Biology*. 2017;27(21):3390–3395. doi:10.1016/j.cub.2017.09.056.
23. Gilpin M. Community-level competition: Asymmetrical dominance. *Proceedings of the National Academy of Sciences of the United States of America*. 1994;91(8):3252–3254. doi:10.1073/pnas.91.8.3252.
24. Toquenaga Y. Historicity of a simple competition model. *Journal of Theoretical Biology*. 1997;187(2):175–181. doi:10.1006/jtbi.1997.0428.
25. Tikhonov M. Community-level cohesion without cooperation. *eLife*. 2016;5:e15747. doi: 10.7554/eLife.15747. doi:10.7554/eLife.15747.
26. Pascual-García A, Bonhoeffer S, Bell T. Metabolically cohesive microbial consortia and ecosystem functioning. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2020;375(1798):0190245. doi:10.1098/rstb.2019.0245.
27. Marsland, R., Cui, W. and Mehta, P. A minimal model for microbial biodiversity can reproduce experimentally observed ecological patterns.. *Sci Rep.* 2020;10, 3308. doi:doi.org/10.1038/s41598-020-60130-2.
28. Lu N, Sanchez-gorostiaga A, Tikhonov M, Sanchez A. Cohesiveness in microbial community coalescence. *bioRxiv*. 2018; p. 282723 doi:10.1101/282723.

29. Hansen SK, Rainey PB, Haagensen JAJ, Molin S. Evolution of species interactions in a biofilm community. *Nature*. 2007;445(7127):533–536. doi:10.1038/nature05514.
30. Lawrence D, Fiegna F, Behrends V, Bundy JG, Phillimore AB, Bell T, et al. Species interactions alter evolutionary responses to a novel environment. *PLoS Biology*. 2012;10(5):e1001330. doi: 10.1371/journal.pbio.1001330. doi:10.1371/journal.pbio.1001330.
31. Embree M, Liu JK, Al-Bassam MM, Zengler K. Networks of energetic and metabolic interactions define dynamics in microbial communities. *Proceedings of the National Academy of Sciences of the United States of America*. 2015;112(50):15450–15455. doi:10.1073/pnas.1506034112.
32. Rivett DW, Jones ML, Ramoneda J, Mombrikotb SB, Ransome E, Bell T. Elevated success of multispecies bacterial invasions impacts community composition during ecological succession. *Ecology Letters*. 2018;21(4):516–524 doi:10.1111/ele.12916.
33. Albright MBN, Sevanto S, Gallegos-Graves LV, Dunbar J. Biotic interactions are more important than propagule pressure in microbial community invasions. *mBio*. 2020;11(5):1–16. doi:10.1128/mBio.02089-20.
34. Butler S, O'Dwyer JP. Stability criteria for complex microbial communities. *Nature Communications*. 2018;9(1): 2970. doi: 10.1038/s41467-018-05308-z.
35. Barabási A, Albert R. Emergence of scaling in random networks. *Science*. 1999;286(5439):509–512. doi:10.1126/science.286.5439.509.
36. Livingston G, Jiang Y, Fox JW, Leibold MA. The dynamics of community assembly under sudden mixing in experimental microcosms. *Ecology*. 2013;94(12):2898–2906. doi:10.1890/12-1993.1.
37. Machado D, Maistrenko OM, Andrejev S, Kim Y, Bork P, Patil KR, et al. Polarization of microbial communities between competitive and cooperative metabolism. *Nature Ecology and Evolution*. 2021;5(2):195–203. doi:10.1038/s41559-020-01353-4.
38. Morris JJ. Black Queen evolution: The role of leakiness in structuring microbial communities. *Trends in Genetics*. 2015;31(8):475–482. doi:10.1016/j.tig.2015.05.004.
39. Kallus Y, Miller JH, Libby E. Paradoxes in leaky microbial trade. *Nat. Comms.* 2017;8(1): 1361. doi: 10.1038/s41467-017-01628-8.
40. Silva LP, Northen TR. Exometabolomics and MSI: Deconstructing how cells interact to transform their small molecule environment. *Current Opinion in Biotechnology*. 2015;34:209–216. doi:10.1016/j.copbio.2015.03.015.
41. Paczia N, Nilgen A, Lehmann T, Gätgens J, Wiechert W, Noack S. Extensive exometabolome analysis reveals extended overflow metabolism in various microorganisms. *Microbial Cell Factories*. 2012;11:1–14. doi:10.1186/1475-2859-11-122.
42. Yamagishi JF, Saito N, Kaneko K. Advantage of Leakage of Essential Metabolites for Cells. *Physical Review Letters*. 2020;124:048101. doi:10.1103/PhysRevLett.124.048101.

43. Yamagishi JF, Saito N, Kaneko K. Adaptation of metabolite leakiness leads to symbiotic chemical exchange and to a resilient microbial ecosystem. *PLOS Comp. Bio.* 2021;17(6): e1009143. doi: 10.1371/journal.pcbi.1009143.
44. Bastolla U, Fortuna MA, Pascual-García A, Ferrera A, Luque B, Bascompte J. The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature*. 2009;458(7241):1018–1020. doi: 10.1038/nature07950.
45. Kurkjian HM, Javad Akbari M, Momeni B. The impact of interactions on invasion and colonization resistance in microbial communities. *PLoS Computational Biology*. 2021;17(1):1–18. doi:10.1371/journal.pcbi.1008643.
46. Wieser W. Cost of growth in cells and organisms: General rules and comparative aspects. *Biol. Rev. Camb. Philos. Soc.* 1994;69(1):1–33. doi: 10.1111/j.1469-185X.1994.tb01484.x.
47. Acosta F, Zamor RM, Najar FZ, Roe BA, Hambright KD. Dynamics of an experimental microbial invasion. *Proceedings of the National Academy of Sciences of the United States of America*. 2015;112(37):11594–11599. doi:10.1073/pnas.1505204112.
48. Girvan MS, Campbell CD, Killham K, Prosser JI, Glover LA. Bacterial diversity promotes community stability and functional resilience after perturbation. *Environmental Microbiology*. 2005;7(3):301–313. doi:10.1111/j.1462-2920.2005.00695.x.
49. Eisenhauer N, Scheu S, Jousset A. Bacterial diversity stabilizes community productivity. *PLoS ONE*. 2012;7(3):1–5. doi:10.1371/journal.pone.0034517.
50. Maron PA, Sarr A, Kaisermann A, Lévéque J, Olivier M, Guigue J, et al. High Microbial Diversity Promotes Soil Ecosystem Functioning. *2018;84:02738–17*. doi:10.1128/AEM.02738-17.
51. Ratzke C, Barrere J, Gore J. Strength of species interactions determines biodiversity and stability in microbial communities. *Nature Ecology and Evolution*. 2020;4(3):376–383. doi:10.1038/s41559-020-1099-4.
52. Seebens H, Blackburn TM, Dyer EE, Genovesi P, Hulme PE, Jeschke JM, et al. No saturation in the accumulation of alien species worldwide. *Nature Communications*. 2017;8:1–9. doi:10.1038/ncomms14435.
53. Rillig MC, Tsang A, Roy J. Microbial community coalescence for microbiome engineering. *Frontiers in Microbiology*. 2016;7:1967. doi:10.3389/fmicb.2016.01967.
54. Calderón K, Spor A, Breuil MC, Bru D, Bizouard F, Viole C, et al. Effectiveness of ecological rescue for altered soil microbial communities and functions. *ISME Journal*. 2017;11:272–283. doi:10.1038/ismej.2016.86.
55. Wang JW, Kuo CH, Kuo FC, Wang YK, Hsu WH, Yu FJ, et al. Fecal microbiota transplantation: Review and update. *Journal of the Formosan Medical Association*. 2019;118(1):S23–S31. doi:10.1016/j.jfma.2018.08.011.
56. Wilson BC, Vatanen T, Cutfield WS, O’Sullivan JM. The super-donor phenomenon in fecal microbiota transplantation. *Frontiers in Cellular and Infection Microbiology*. 2019;9(2): doi:10.3389/fcimb.2019.00002.

57. Lindemann SR, Bernstein HC, Song HS, Fredrickson JK, Fields MW, Shou W, et al. Engineering microbial consortia for controllable outputs. *ISME Journal*. 2016;10(9): 2077–2084. doi:10.1038/ismej.2016.26.
58. Louca S, Polz MF, Florent M, Albright MBN, Huber JHA, O'Connor MI, et al. Function and functional redundancy in microbial systems. *Nature Ecology and Evolution* 2018;2(6):936–943. doi:10.1038/s41559-018-0519-1.
59. Enke TN, Datta, MS, Schwartzman J, Cermak N, Schmitz D, Barrere J, Pascual-García A, Cordero OX. Modular Assembly of Polysaccharide-Degrading Marine Microbial Communities *Current Biology* 2019;29(9):1528–1535.e6 doi:10.1016/j.cub.2019.03.047.
60. Sung J, Kim S, Cabatbat JJT, Jang S, Jin Y, Jung GY, et al. Global metabolic interaction network of the human gut microbiota for context-specific community-scale analysis. *Nat. Commun.* 2017;8:15393. doi:10.1038/ncomms15393.
61. Tikhonov M, Monasson R. Collective phase in resource competition in a highly diverse ecosystem. *Physical Review Letters*. 2017;118:048103. doi:10.1103/PhysRevLett.118.048103.
62. Tikhonov M, Monasson R. Innovation rather than improvement: A solvable high-dimensional model highlights the limitations of scalar fitness. *J. Stat. Phys.* 2018;172:74–104. doi:10.1007/s10955-018-1956-6.
63. DeLong JP, Okie JG, Moses ME, Sibly RM, Brown JH. Shifts in metabolic scaling, production, and efficacy across major evolutionary transitions of life. *Proceedings of the National Academy of Sciences*. 2010;107(29): 12941–5. doi: 10.1073/pnas.1007783107.
64. Kempes CP, van Bodegom PM, Wolpert D, Libby E, Amend J, Hoehler T. Drivers of Bacterial Maintenance and Minimal Energy Requirements. *Front. Microbiol.* 2017;8:31. doi: 10.3389/fmicb.2017.00031.
65. Pascual-García A, Bastolla U. Mutualism supports biodiversity when the direct competition is weak. *Nature Communications*. 2017 8:14326. doi: 10.1038/ncomms14326.

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

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

September 1, 2021

Contents

S1 Further details of the mathematical model	2
S1.1 Cost function	2
S1.2 Effect of the cost function on community feasibility and dominance	3
S1.3 Stability	6
S1.4 Relationship with other microbial consumer-resource models	7
S2 Modulating net competition levels	8
S2.1 Modulating competition level	8
S2.2 Modulating facilitation level	10
S3 Further details of the community coalescence simulations	11
S3.1 Coalescence simulation procedures	11
S4 Adding consumer guild structure	13
S4.1 Coalescence between communities with guild structure	15

S1 Further details of the mathematical model

The model is (same as main text Eq 1):

$$\begin{aligned}\frac{dN_\alpha}{dt} &= g_\alpha N_\alpha \left((1 - l_\alpha) \sum_j c_{\alpha j} R_j - z_\alpha \right), \\ \frac{dR_j}{dt} &= \kappa_j + \tau^{-1} R_j - \sum_\alpha N_\alpha c_{\alpha j} R_j + \sum_{\alpha k} N_\alpha l_\alpha D_{\alpha k j} c_{\alpha k} R_k.\end{aligned}\tag{S1}$$

In matrix form, this system can be written as:

$$\begin{aligned}\frac{d\mathbf{N}}{dt} &= \mathcal{D}(\mathbf{g} \circ \mathbf{N}) (\mathcal{D}(\mathbb{1} - \mathbf{l}) C \mathbf{R} - \mathbf{z}) \\ \frac{d\mathbf{R}}{dt} &= \boldsymbol{\kappa} - \tau^{-1} \mathbf{R} - \mathcal{D}(\mathbf{R}) C^T \mathbf{N} + \sum_\alpha l_\alpha D_\alpha^T \mathcal{D}(\mathbf{R}) \mathbf{c}_\alpha N_\alpha\end{aligned}$$

Here, \circ denotes an element-wise operation, \mathbf{R} and \mathbf{N} are the vectors of resource concentrations and population abundances respectively, $\mathbb{1}$ is a vector of ones, and the remaining symbols in boldface represent the same quantities in Eq S1 in vector form. C is the matrix of consumers' preferences, and $\mathcal{D}(\mathbf{x})$ is a diagonal matrix with the vector \mathbf{x} as its diagonal.

S1.1 Cost function

Every microbial cell has a maintenance cost, which is the energy required to perform tasks to remain alive such as transportation of metabolites and synthesis of RNA and enzymes for metabolizing substrates. We define this cost to be

$$z_\alpha = \chi_0 (1 + \epsilon_\alpha) (1 - l_\alpha) \sum_j c_{\alpha j}.$$

Here, χ_0 is the average cost of consuming a given resource, the summation represents the total number of resources that species α is able to process, and ϵ_α is a small random fluctuation sampled from a truncated normal distribution ($\epsilon_\alpha \sim N(0, 0.1)$), that introduces variation in the cost for species that have identical preferences. Also, note that the fitness of an organism is determined by both its cost and its metabolic preferences, so we keep the random fluctuations to be small values (relative to the uptake values) in all simulations because if they are too large, fitness of the organism would be purely determined (randomly) by its cost, which would be biologically unrealistic considering the unavoidable feedback between environment and consumers in determining their fitness. As such, setting $\epsilon_\alpha = 0 \forall \alpha$ would not qualitatively change our results about coalescence outcomes.

This cost function entails two key assumptions. The first assumption is that generalist consumer species (which feed on a wide range of resources types) pay a higher maintenance cost (the summed term in the cost function) than specialists (which consume one or few resource types; henceforth, the "generalism cost"). More generalist species necessarily maintain more complex metabolic networks than specialist species, and the upkeep of larger metabolic networks (and thus, larger genomes) incurs in greater maintenance costs (DeLong et al. 2010,

Kempes et al. 2017), leading to a trade-off between resource generalism and cost of maintenance. Here we impose this trade-off such that the maintenance cost is proportional to the sum of all resource preferences (i.e. the number of resources consumed). This assumption is similar to the one made by Tikhonov (2016) (also see Tikhonov & Monasson (2017)).

The second assumption is that the cost of cellular maintenance is also proportional to the fraction of retained resources ($1 - l$), i.e., species that retain greater quantities of resources (those with less leakage, l) have higher maintenance costs (the “efficiency cost”). This assumption stems from the fact that the processing of resources itself incurs a metabolic cost (WIESER 1994). Therefore those species retaining greater quantities (leaking less) of resources must pay a greater metabolic cost of processing those resources. With this cost function, the consumer equation from Eq S1 becomes

$$\frac{dN_\alpha}{dt} = g_\alpha N_\alpha (1 - l_\alpha) \left(\sum_j c_{\alpha j} R_j - \chi_0 (1 + \epsilon_\alpha) \sum_j c_{\alpha j} \right). \quad (\text{S2})$$

Thus, leakage is now a factor that also determines what proportion of the harvested energy is converted to biomass, consistent with the classical notion of biomass conversion or assimilation efficiency (WIESER 1994).

S1.2 Effect of the cost function on community feasibility and dominance

We now explain how the above maintenance cost function has important implications for community feasibility and ultimately, coalescence outcomes. Species α will reach a feasible (positive) equilibrium when its resource surplus term, that is, the amount of energy left when the maintenance cost is subtracted from the initial harvest (terms inside the brackets of the consumer equation in Eq S1) equals 0. That is (after substituting the cost function),

$$\frac{1}{g_\alpha N_\alpha} \frac{dN_\alpha}{dt} = (1 - l_\alpha) \left(\sum_j c_{\alpha j} R_j - \chi_0 (1 + \epsilon_\alpha) \sum_j c_{\alpha j} \right) = 0.$$

Therefore, at steady state,

$$\sum_j c_{\alpha j} R_j^* = \chi_0 (1 + \epsilon_\alpha) \sum_j c_{\alpha j}. \quad (\text{S3})$$

For the whole community, the vector of equilibrium resource abundances is given by,

$$C \mathbf{R}^* = \chi_0 \mathcal{D}(\mathbb{1} + \boldsymbol{\epsilon}) C \mathbb{1}, \\ \text{i.e., } \mathbf{R}^* = \chi_0 C^{-1} \mathcal{D}(\mathbb{1} + \boldsymbol{\epsilon}) C \mathbb{1} \quad (\text{S4})$$

Here, \mathbf{R}^* is the vector of resource concentrations (the $*$ indicating it’s equilibrium state), C is the consumer preferences matrix, $\mathbb{1}$ is a vector of ones, and $\mathcal{D}(\mathbb{1} + \boldsymbol{\epsilon})$ is a diagonal matrix with the vector $\mathbb{1} + \boldsymbol{\epsilon}$ (the vector of random fluctuations in costs) in its diagonal. As such, Eq. (S3) (and (S4)) is a necessary (but not sufficient) condition for the community’s feasibility

(existence of the consumer-resource equilibrium) provided that (i) the preferences matrix C is invertible, and (ii) all the resource concentrations at equilibrium are positive. In the following analysis, we only consider systems where these two conditions are satisfied, because our goal is to establish the conditions for community dominance following coalescence *given* a pair of feasible (and locally asymptotically stable; next section) parent communities. Our simulations show that these results hold even when we coalesce communities with feasible equilibria consisting of different consumer numbers (the measure $S_{1,2}$ is independent of species richness; main text Methods, Step 3).

We now show that the two key assumptions of our cost function—the efficiency cost and the generalism cost—are not just biologically realistic but also crucial for understanding the role of competition and cooperation in community coalescence. To this end, we relax each assumption, and re-calculate the resource abundance vector at equilibrium. When we relax the assumption of an efficiency cost, the equilibrium resource vector becomes

$$\mathbf{R}^* = \mathcal{D}^{-1}(\mathbb{1} - \mathbf{l})\chi_0 C^{-1}\mathcal{D}(\mathbb{1} + \boldsymbol{\epsilon})C\mathbb{1}. \quad (\text{S5})$$

In this case, in contrast to Eq. (S4), the equilibrium resource abundances (i.e., magnitude of the abundance vector \mathbf{R}^*) increases with values of species' leakages (magnitude of \mathbf{l}). Thus, all else (including the supply of different resources) being the same, a “less leaky” community will deplete the given resource pool to a lower concentration than a more leaky one. This means that when two such communities are coalesced, the one with lower leakage will create an environment that does not guarantee feasible coexistence of consumers in the other community (maintenance cost becomes higher than energy harvest, resulting in negative growth rates). That is, upon community-community encounter, more consumer species of the leakier community will be driven extinct relative to those from the less leaky one because the resource environment is not capable of sustaining a positive equilibrium. In other words, in the absence of an efficiency cost, communities with higher leakage will be systematically be displaced or dominated by communities with lower leakage. This can be interpreted as a variant of Tilman’s R^* rule applied to a pair of communities instead of a pair of species (also see Tikhonov (2016)). This analytical result is demonstrated numerically in Fig S1. Furthermore, since leakage is directly correlated with community-wide cooperation level (main text Eq. 4), relaxing this assumption would imply that being more cooperative is detrimental for coalescence success, contradicting current literature which has found that co-operation generally increases structural stability and resistance to invasions (Pascual-García & Bastolla 2017, Kurkjian et al. 2021).

Next, we eliminate the generalism cost, which means that the equilibrium resource vector becomes

$$\mathbf{R}^* = \chi_0 C^{-1}(\mathbb{1} + \boldsymbol{\epsilon}).$$

In this case, magnitude of the resource abundance vector at equilibrium decreases with the number of consumer species’ resource preferences (magnitude of C). This means that generalist consumers are able to deplete resources more efficiently than specialists just because they possess more metabolic pathways, without incurring in an extra cost (i.e., a species able to consume 5 resources would harvest 5 times more energy than a species with just 1 metabolic preference). However, this is unrealistic, since, as mentioned above, the more metabolic pathways present, the higher the probability that any two of them require different

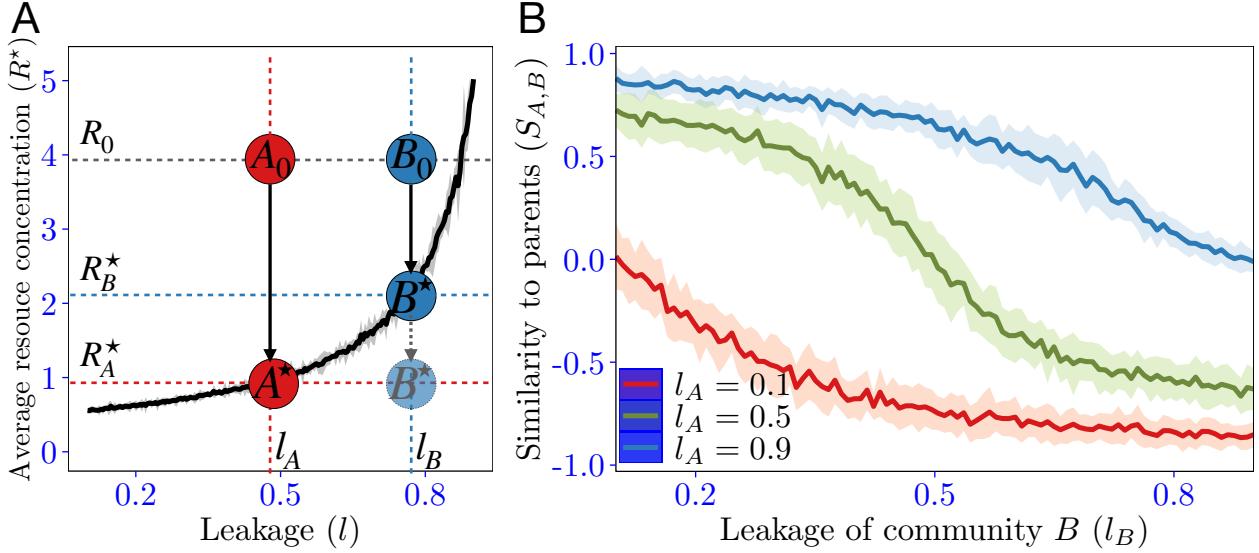


Figure S1: Consequence of eliminating leakage-dependence of the cost function for community dominance in pairwise coalescence events. **A:** Without leakage dependence of the cost function, mean of its resource concentrations (R^*) at equilibrium (black curve) increases with leakage. The black curve is the mean of the mean resource concentrations (shaded region is $\pm \sigma$) reached at equilibrium for many random instances of feasible parent communities. Therefore, when any two communities with different leakage values A ($l_\alpha \forall \alpha = l_A$) and B ($l_\alpha \forall \alpha = l_B$) (all species set to have the same value for simplicity) are assembled in isolation in the same environment R_0 , they will deplete the resources to concentrations R_A^* and R_B^* , respectively. As a result, when the two are coalesced, community A , which can deplete the resources to a lower concentration, will create an environment that does not guarantee feasibility of consumers in community B (semi-transparent blue circle), causing more species from community B to be driven extinct relative to those from A . **B:** Parent community dominance ($S_{A,B}$) after repeated coalescence events between pairs of communities of the types A and B . We use the recursive coalescence simulation procedure (see main text Methods, Step 2 and Section S3), where the leakage (l_B , x-axis) is slightly increased after each iteration. So this simulation is equivalent to that producing main text Fig 4B, the only difference being that here the cost function lacks the leakage term. The result here is opposite to the one in main text Fig 4B, confirming that if the cost function is independent of leakage level, lower leakage (and therefore less cooperation) favors parent community dominance after coalescence. Note the sharper decrease in community dominance of B as its increasing leakage (l_B) approaches the leakage of its competitor (l_A).

cellular machinery to be activated for optimal maintenance and functioning (e.g., through two different cellular compartments), incurring extra cost (Tikhonov & Monasson 2017).

Therefore, both efficiency and generalism costs are necessary for a meaningful analysis of the effect of cooperation vs competition on community coalescence outcomes. Of course, if two communities have the same leakage levels, they will on average have similar levels of cooperation, in which case competition level is the only factor that needs to be considered, and the presence of an efficiency cost does not matter. To confirm this, we ran the ran-

dom coalescence procedure with communities assembled with a cost function lacking leakage dependence (Fig S2), which, as expected, yields qualitatively the same result as for communities assembled *with* the efficiency cost (main text Fig 3) because we are only coalescing communities with the same leakage value.

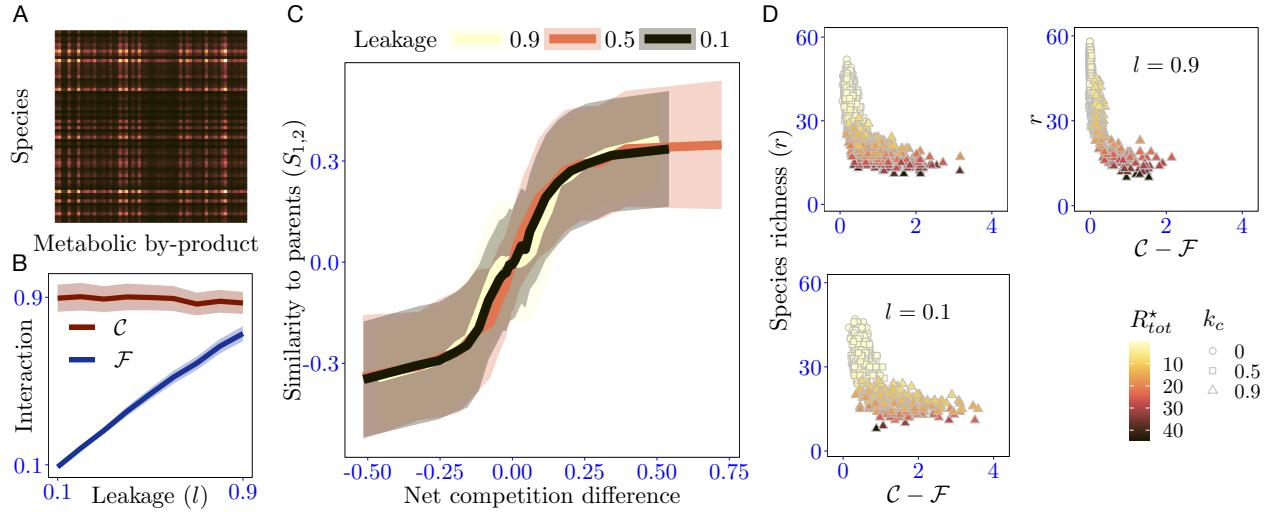


Figure S2: Results of random coalescence procedure without imposing the efficiency cost (leakage dependence). See main text Fig 4 for descriptions of the sub-panels.

S1.3 Stability

After assembly of a feasible parent or coalesced community, we assessed its local stability as follows. We first compute the Jacobian matrix of the system, and evaluate it at the steady state population and resource abundances (the feasible equilibrium, which may or may not be stable). For our system, the Jacobian is the block matrix of the form

$$J\Big|_{\substack{\mathbf{N}=\mathbf{N}^* \\ \mathbf{R}=\mathbf{R}^*}} = \begin{bmatrix} \frac{\partial \dot{\mathbf{n}}}{\partial \mathbf{n}} & \frac{\partial \dot{\mathbf{n}}}{\partial \mathbf{R}} \\ \frac{\partial \dot{\mathbf{R}}}{\partial \mathbf{n}} & \frac{\partial \dot{\mathbf{R}}}{\partial \mathbf{R}} \end{bmatrix} \Bigg|_{\substack{\mathbf{n}=\mathbf{N}^* \\ \mathbf{R}=\mathbf{R}^*}} = \begin{bmatrix} \mathcal{D}(g(\mathcal{D}(\mathbf{R}^* \mathbf{C}(1-l) - \mathbf{z})) & \mathcal{D}(g \circ \mathbf{N}^*)(1-l)\mathbf{C} \\ -\mathcal{D}(\mathbf{R}^*)\mathbf{C}^T + l\mathbf{D}^T\mathcal{D}(\mathbf{R}^*)\mathbf{C}^T & -I\tau^{\circ-1}\mathcal{D}(\mathbf{C}^T \mathbf{N}^*) + l\mathbf{D}^T\mathcal{D}(\mathbf{C}^T \mathbf{N}^*) \end{bmatrix}$$

Here, I is the identity matrix. After each assembly, J was evaluated at the equilibrium abundances and its eigenvalues calculated. We find that the real part of the dominant eigenvalue (right-most on the real axis; e.g., Fig S3) of all assembled parent as well as coalesced communities is negative and real. That is, feasibility guarantees local asymptotic stability in our model systems. Note that here, leakage l is not a vector because during assembly, we assumed that all species had the same community-level leakage.

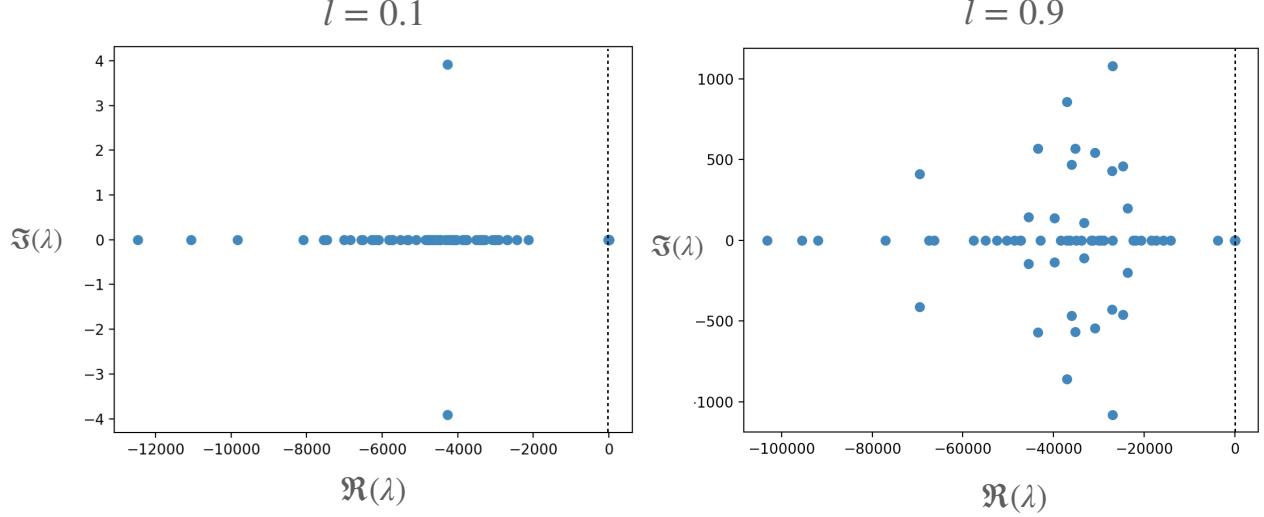


Figure S3: Real (x axis) and complex (y axis) parts of eigenvalues of J when evaluated at equilibrium for two example communities with $l = 0.1$ and $l = 0.9$. In both cases, all eigenvalues are negative (with the leading eigenvalues very close to zero), indicating local stability of the equilibrated communities.

S1.4 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 (maintenance cost)	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. We use a different cost function (further details below) from Marsland et al. (2019) who assume that that maintenance cost z_α is a random fixed quantity for each species.

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. If we take the system of equations in S1 and assume that (i) there is no leakage ($l = 0$) and (ii) the dilution rate is very low, such that $\tau^{-1} \approx 0$, 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):

$$\frac{dN_\alpha}{dt} = g_\alpha N_\alpha \underbrace{\left(\sum_j c_{\alpha j} \frac{\kappa_j}{\sum_\alpha N_\alpha c_{\alpha j}} - z_\alpha \right)}_{\text{Resource surplus } \Delta_\alpha}$$

$$R_j = \frac{\kappa_j}{\sum_\alpha N_\alpha c_{\alpha j}}$$

Furthermore, we use a different cost function (previous section) than the one used by Tikhonov.

S2 Modulating net competition levels

We are interested in generating communities spanning a broad range of competition and facilitation levels, i.e., different competition and facilitation levels ($\mathcal{C} - \mathcal{F}$); Main text Eqs 3–6). To this end we sample the $c_{\alpha j}$ and D_{jk} values for each community with certain constraints (Fig S4).

S2.1 Modulating competition level

First, consider the problem of increasing the competition level (\mathcal{C}) of the community. The 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 ($c_{\alpha j} = 0$). We sample the resource preference vector of each consumer sequentially (that is, one species at a time), in a way that allows us to modify the niche similarity between them. For each consumer α , the sampling probability of each resource j , $p_{\alpha j}$, is re-evaluated (as in a preferential attachment process, Barabási & Albert (1999)) such that those resources

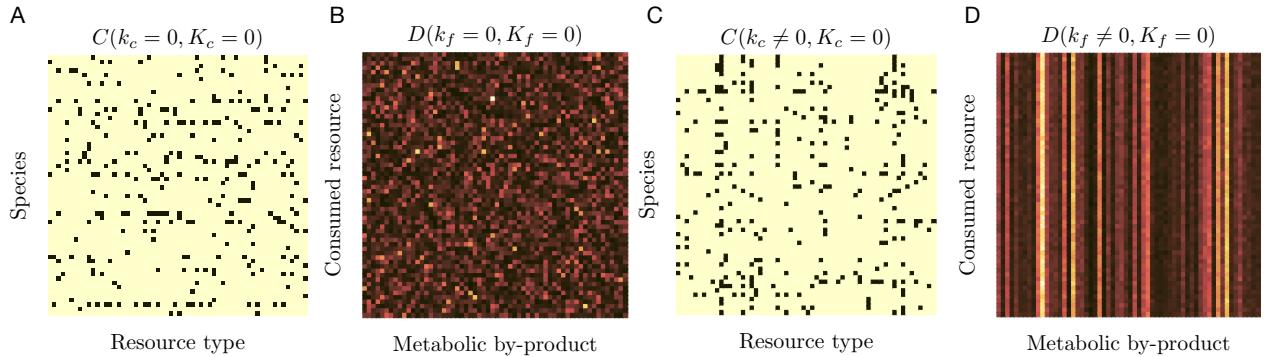


Figure S4: Examples of differently-structured preference (C) and metabolic (D) matrices. These have been generated with different combinations of the competition and facilitation factors k_c , k_f in systems of 60 resource types and 60 consumer species. Here, K_c and K_f are set to zero. Increasing these values would move the regime towards more structured resource use (See section S2). **A & B:** Uniformly random matrices, where all four parameters are 0. **C & D:** As k_c and k_f are increased the regime moves towards greater preferential feeding, where more demanded resources are more likely to be consumed (increase of k_c), but also secreted at higher fractions (increase of k_f). In the metabolic matrices (B & D), lighter colours indicate higher values of resource fractions secreted.

that have been frequently sampled by previous species receive a higher probability in the current species.

More specifically, we construct the the resource preference matrix C by sampling these rows \vec{c}_α 's as follows. First, we 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 $d_{\alpha j}$, we then compute the probability that species α is assigned resource j as one of its preferences

$$p_{\alpha j} = (1 - k_c) \frac{1}{m} + k_c \frac{d_{\alpha-1j}}{\sum_j d_{\alpha-1j}}, \quad (\text{S6})$$

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 ,

$$\tilde{d}_{\alpha-1j} = \frac{d_{\alpha-1j}}{\sum_j d_{\alpha-1j}}.$$

The strength with which $p_{\alpha j}$ deviates from a uniform distribution is given by the parameter $k_c \in [0, 1)$ that is, how much consumers prefer highly-demanded resources, such that when $k_c = 0$ the sampling is uniformly random (Fig S4A); and as $k_c \rightarrow 1$ the feeding becomes increasingly preferential (Fig S4C). Pseudocode for the metabolic preferences sampling procedure is given in **Procedure 1**.

Procedure 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
```

S2.2 Modulating facilitation level

Effective competition can alternatively be reduced through the indirect positive effect of facilitation, that is, by increasing \mathcal{F}/l . This can be achieved if resources that are highly demanded are also secreted in larger fractions (Fig S4E). For this we need to modulate the structure of the the 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 $\sum_j D_{jk} = 1$. Thus, we sample each row of the metabolic matrix D (Fig S4D) from a Dirichlet distribution with a specific concentration parameters q_{jk} . The elements of q_{jk} increase proportionally with the demand of each resource d_j and the cooperation factor k_f as

$$q_{jk} = \frac{1}{u}(1 + k_f d_j). \quad (\text{S7})$$

Thus, k_f sets the degree 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 equi-probable fractions. As $k_f \rightarrow 1$ the structure of D becomes fully determined by the resource demands of the community, so that more demanded resources are released at higher fractions. The factor u in Eq S7 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 above two methods for sampling the elements of C and D share similarities, they are conceptually different. First, the sampling of C 's elements is fully random, in the sense that a vector of probabilities is constructed first, and then preferences randomly sampled from it. On the other hand, the sampling of D 's elements has a random term.

S3 Further details of the community coalescence simulations

S3.1 Coalescence simulation procedures

Recursive coalescence

Given two microbial communities A and B , and two community leakage values l_A and l_B , such that we implement the following computational procedure:

1. Assemble community A with leakage value l_{Ai} .
2. Assemble community B with leakage values l_{Bj} .
3. Perform coalescence event between A and B and record parent community dominance (defined below)
4. Change the value $l_{Bi} \rightarrow l_{Bi+1}$, and repeat steps 2 and 3 until vector l_B has been fully traversed.
5. Replicate steps 1-4 in order to gain statistical power.
6. Change value of $l_{Aj} \rightarrow l_{Aj+1}$ and repeat steps 1-5 until vector l_A has been fully traversed.

Pseudocode for the this coalescence simulation is shown in **Procedure 2**. Note that our

Procedure 2: Recursive coalescence

```
begin
  for 1 ≤ i ≤ length( $\vec{l}_A$ ) do
    while better statistical power is required do
      for 1 ≤ j ≤ length( $\vec{l}_B$ ) do
        Assemble community A with leakage value  $l_{Ai}$ 
        Assemble community B with leakage value  $l_{Bj}$ 
        Coalesce communities A & B
        Record parent community dominance
```

recursive coalescence procedure does not guarantee that every time parent community B is re-assembled (step 2), the present species and their abundances will remain the same as those in the community with the previous leakage value. Thus, at each re-assembly of community B, we check that changing the leakage does not affect significantly the community composition. With this purpose, we calculate the auto-correlation between community abundance vectors of consecutive iterations, and find that it remains close to 1 along all the studied leakage range (see Fig S5).

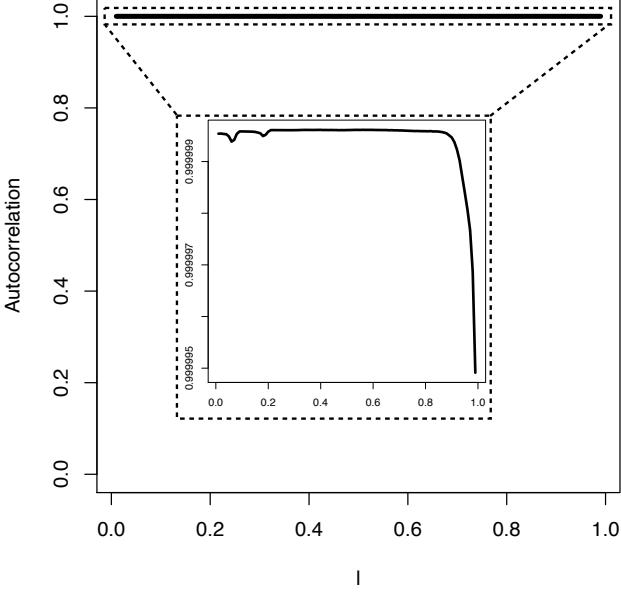


Figure S5: For recursive coalescence, auto-correlation of vector of species abundances in community B for consecutive re-assemblies of this community, along the studied leakage range. (inset) Zoom in to show that, even though the auto-correlation remains consistently ≈ 1 , it starts decreasing from a certain leakage value. While the reason for this decrease is interesting and a potential line of future work, it does not affect our results because of the very small change of magnitude in autocorrelation values over this range.

Serial coalescence

Given two sets of *target* and *invader* communities (each of the latter with different leakage values) we implement the following computational steps:

1. Assemble target community T_0 with leakage l_{T_i} .
2. Assemble a random invader community W with leakage l_{W_j} .
3. Coalesce $T_0 + W = T_1$, and record properties of community T_1 .
4. Repeat steps 2 and 3 until community $T_n = T_{n-1} + W_n$ is reached.
5. Change leakage of the W communities $l_{W_j} \rightarrow l_{W_{j+1}}$.
6. Repeat steps 2-5 until leakage vector \vec{l}_W has been fully traversed.
7. Change leakage of T communities $l_{T_i} \rightarrow l_{T_{i+1}}$.
8. Repeat steps 1-7 until vector \vec{l}_T has been fully traversed.
9. Replicate steps 1-8 in order to gain statistical power.

Pseudocode for the serial coalescence simulation is shown in **Procedure 3**.

Procedure 3: Serial coalescence

```

Procedure begin
  while better statistical power is required do
    for  $1 \leq i \leq \text{length}(\vec{l}_T)$  do
      Assemble target community  $T_0$  with leakage  $\vec{l}_{T_i}$ 
      for  $1 \leq j \leq \text{length}(\vec{l}_W)$  do
        for  $1 \leq k \leq n$  do
          Assemble invader community  $W_k$  with leakage  $\vec{l}_{W_j}$ 
          Coalesce  $T_{k-1}$  and  $W_k$  to form target community  $T_k$ 
          Record properties of target community  $T_k$ 

```

S4 Adding consumer guild structure

Recent empirical studies suggests that microbial species tend to form guilds with similar metabolic capabilities, thus introducing some degree of functional redundancy in the communities they form Louca et al. (2018), Enke et al. (2019). Theoretical studies support these observations Goldford et al. (2018), Marsland et al. (2020), Fant et al. (2021). We therefore add further structure to the matrices C and D , by partitioning resources into classes, and constraining consumers to feed on a preferred class, but leak to any other, forming consumer guilds. Adding this structure yields two interaction layers (imagine superimposing Figs S6B and 2C with Figs S6A and 2D): inter-guild facilitation and competition between consumers preferring distinct resource classes, and intra-guild facilitation and competition, which stems from the previously-imposed preferential feeding, yielding an effective secretion matrix like the one plotted in Fig S6C.

Resource preferences in this scenario are assigned similarly to the unstructured preferential feeding described in Section S2, except that the probability that species α_A (which feeds preferentially on resource class A) samples resource j , is now weighted up or down depending on whether j belongs in guild A , or not, respectively (Fig S6B). To this end, we define the form of this function to be:

$$p_{\alpha j}^A = \begin{cases} M \left((1 - k_c) \frac{1}{m} + k_c \frac{d_{\alpha-1j}}{\sum_j d_{\alpha-1j}} \right) (1 + K_c) & \text{if } j \in A \\ \frac{N}{m - n_c} (1 - K_c) & \text{otherwise,} \end{cases} \quad (\text{S8})$$

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 constraints on each piece of Eq S8

$$p_\alpha^1 = \sum_{C(j) \in T} p_{\alpha j} = \frac{n_c}{m} (1 - K_c) + K_c \quad (\text{S9})$$

and

$$p_\alpha^0 = \sum_{C(j) \notin T} p_{\alpha j} = \left(1 - \frac{n_c}{m}\right) (1 - K_c). \quad (\text{S10})$$

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 S9 and S10 using the definition in expression S8 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

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

The metabolic matrix D (Fig S6A) is constructed such that the fraction of leaked by-product k is lower if it belongs to the same class as the consumed resource j (elements within block-diagonals of D), and higher otherwise (off-block diagonal elements of D). The prominence of this structure in the matrix is given by the inter-guild facilitation factor K_f . Therefore, we sample each row of D 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 S7). 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

$$q_{jk} = \begin{cases} \frac{(1 + k_f d_j)}{u M_{C(j)}} (1 - K_f) & \text{if } A(j) = A(k) \\ \frac{(1 + k_f d_j)}{u(M - M_{C(j)})} (1 + K_f) & \text{otherwise.} \end{cases} \quad (\text{S12})$$

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 S11 and S12 if we make $K_c = K_f = 0$, all the resources belong to the same class, and we recover equations S6 and S7 from the previous version.

S4.1 Coalescence between communities with guild structure

When coalescence is simulated between pairs of communities having guild structure, we find that in the low leakage regime, where significant competition is present, the same result (yellow and red lines in Fig S6D) as in the case of random coalescence with only preferential feeding (Fig S2C) is recovered. We find a positive correlation even in the high leakage regime where facilitation is of the same order of magnitude as competition, indicating that our results qualitatively hold for structured communities as well. Note that competition between guilds, and facilitation within guilds, are both very weak. Therefore, here we calculated community level competition as the average over the block diagonal elements, and community level facilitation as the average over the off-block diagonal elements.

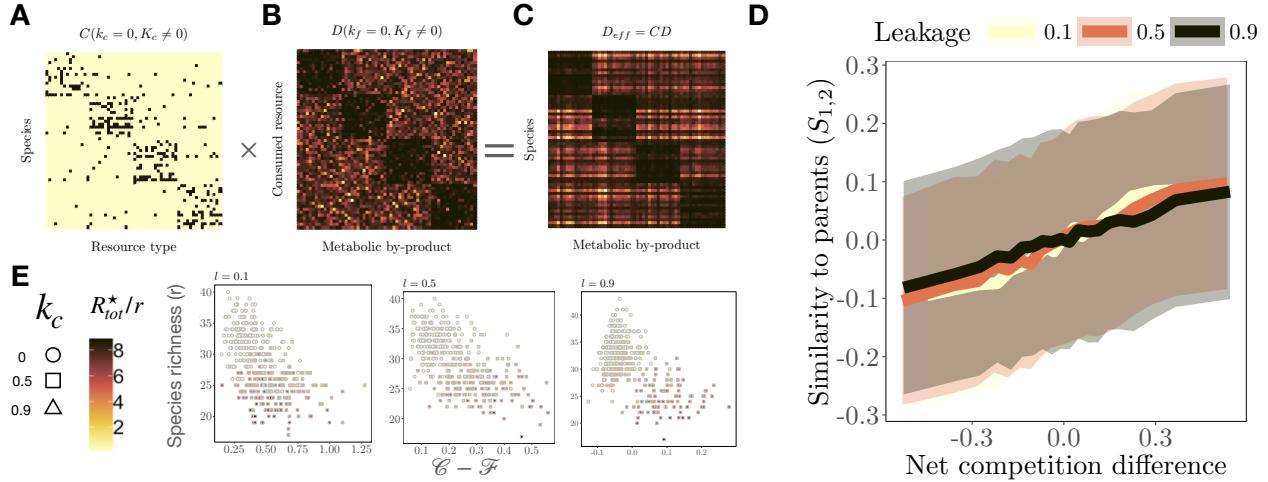


Figure S6: Community coalescence with consumer guilds present. **A-C:** Example of a metabolic matrix, preference matrix, and effective secretion matrix, with consumer guild structure. **D:** Similarity to the parent community as function of the binned mean (20 bins) over parent communities with similar difference in competition levels $\mathcal{C}_1 - \mathcal{C}_2$ (solid line) ± 1 standard deviation (shaded) for the three leakage values. The post-coalescence community is more similar to its less competitive parent. **E:** Species richness (r) as a function of community-level net competition ($\mathcal{C} - \mathcal{F}$), coloured by total resource concentration reached at steady state (R_{tot}^*). The observed negative correlation for all values of leakage confirms that less competitive communities are more species-rich, and deplete resources more efficiently (brighter colours, which correspond with lower levels of R_{tot}^*) are scattered towards the top left of the plot.

References

- Barabási, A.-L. & Albert, R. (1999), ‘Emergence of Scaling in Random Networks’, *Science* **286**(5439).
- DeLong, J. P., Okie, J. G., Moses, M. E., Sibly, R. M. & Brown, J. H. (2010), ‘Shifts in metabolic scaling, production, and efficiency across major evolutionary transitions of life’, *Proceedings of the National Academy of Sciences* **107**(29).
- Enke, T. N., Datta, M. S., Schwartzman, J., Cermak, N., Schmitz, D., Barrere, J., Pascual-García, A. & Cordero, O. X. (2019), ‘Modular Assembly of Polysaccharide-Degrading Marine Microbial Communities’, *Current Biology* **29**(9), 1528–1535.
- Fant, L., Macocco, I. & Grilli, J. (2021), ‘Eco-evolutionary dynamics lead to functionally robust and redundant communities’, *bioRxiv* p. 2021.04.02.438173.
- URL:** <http://biorexiv.org/content/early/2021/04/02/2021.04.02.438173.abstract>
- 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* **361**(6401), 469–474.

- Kempes, C. P., van Bodegom, P. M., Wolpert, D., Libby, E., Amend, J. & Hoehler, T. (2017), 'Drivers of Bacterial Maintenance and Minimal Energy Requirements', *Frontiers in Microbiology* **8**.
- Kurkjian, H. M., Javad Akbari, M. & Momeni, B. (2021), 'The impact of interactions on invasion and colonization resistance in microbial communities', *PLoS Computational Biology* **17**(1), 1–18.
URL: <http://dx.doi.org/10.1371/journal.pcbi.1008643>
- Louca, S., Polz, M. F., Mazel, F., Albright, M. B., Huber, J. A., O'Connor, M. I., Ackermann, M., Hahn, A. S., Srivastava, D. S., Crowe, S. A., Doebeli, M. & Parfrey, L. W. (2018), 'Function and functional redundancy in microbial systems', *Nature Ecology and Evolution* **2**(6), 936–943.
URL: <http://dx.doi.org/10.1038/s41559-018-0519-1>
- 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.
- Marsland, R., Cui, W. & Mehta, P. (2020), 'A minimal model for microbial biodiversity can reproduce experimentally observed ecological patterns', *Scientific Reports* **10**(1), 1–17.
URL: <http://dx.doi.org/10.1038/s41598-020-60130-2>
- Pascual-García, A. & Bastolla, U. (2017), 'Mutualism supports biodiversity when the direct competition is weak', *Nature Communications* **8**.
- 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.
- WIESER, W. (1994), 'COST OF GROWTH IN CELLS AND ORGANISMS: GENERAL RULES AND COMPARATIVE ASPECTS', *Biological Reviews* **69**(1).



Click here to access/download
LaTeX Source File (TEX file)
main_review_edits.tex