nature research

Peer Review Information

Journal: Nature Ecology & Evolution
Manuscript Title: Polarization of microbial communities between competitive and cooperative metabolism
Corresponding author name(s): Kiran R. Patil

Editorial Notes:

Reviewer Comments & Decisions:

Decision Letter, initial version:

8th April 2020

*Please ensure you delete the link to your author homepage in this e-mail if you wish to forward it to your co-authors.

Dear Dr Patil,

Your manuscript entitled "Polarization of microbial communities between competitive and cooperative metabolism" has now been seen by 3 reviewers, whose comments are attached. The reviewers have raised a number of concerns which will need to be addressed before we can offer publication in Nature Ecology & Evolution. We will therefore need to see your responses to the criticisms raised and to some editorial concerns, along with a revised manuscript, before we can reach a final decision regarding publication.

We therefore invite you to revise your manuscript taking into account all reviewer and editor comments. Please highlight all changes in the manuscript text file.

We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

When revising your manuscript:

* Include a “Response to reviewers” document detailing, point-by-point, how you addressed each reviewer comment. If no action was taken to address a point, you must provide a compelling

argument. This response will be sent back to the reviewers along with the revised manuscript.

* If you have not done so already please begin to revise your manuscript so that it conforms to our Article format instructions at http://www.nature.com/natecolevol/info/final-submission. Refer also to any guidelines provided in this letter.

* Include a revised version of any required reporting checklist. It will be available to referees (and, potentially, statisticians) to aid in their evaluation if the manuscript goes back for peer review. A revised checklist is essential for re-review of the paper.

Please use the link below to submit your revised manuscript and related files:

[REDACTED]

Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

We hope to receive your revised manuscript within four to eight weeks. If you cannot send it within this time, please let us know. We will be happy to consider your revision so long as nothing similar has been accepted for publication at Nature Ecology & Evolution or published elsewhere.

Nature Ecology & Evolution is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' on published papers create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on 'Modify my Springer Nature account'. For more information please visit please visit www.springernature.com/orcid.

Please do not hesitate to contact me if you have any questions or would like to discuss these revisions further.

We look forward to seeing the revised manuscript and thank you for the opportunity to review your work.

[REDACTED]

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

In this paper, the authors identify and analyze communities of co-occuring species that can be found in many different locations on the planet. Their main conclusion is that such communities roughly come in two types: cooperative communities, in which the co-occurring species tend to have different

nutrient requirements but help each other by producing byproducts that others species in the community can consume (cross-feeding), and competitive communities, in which co-occurring species tend to compete for a limited number of nutrients. The authors show that competitive communities mainly occur in soil, whereas cooperative communities can live anywhere, including host-associated microbiomes. They also argue that cooperative communities are less affected by changes in the nutritional environment than competitive communities, but are more prone to invasion by other species.

This is a very interesting and well written paper that will be of interest to a wide audience.

My caveat is that I am not well-versed in the software packages that the authors use for their analyses, so I am unable to verify the results (the devil usually lies in the details...). I have no reason to think that their analysis is wrong, and I think they developed an interesting novel way to actually identify co-occurring communities (rather than just, say co-occurring species pairs).

There is one thing that was missing for me: the authors don't mention bacteriocins at all. It is well known that bacterial species try to poison co-occuring species using many different kinds of toxins, including bacteriocins. I think it would be very interesting to overlay the results in this paper with an analysis of bacteriocin production, which could probably be included in the genomic analysis. One might e.g. expect that competitive communities produce more bacteriocins than cooperative communities. I'm not sure whether it is doable, but I think that an analysis in this direction might strengthen the paper even further.

Minor comment: The authors say somewhere (line numbers would be nice!) that they "computed co-occurring communities using an independent collection of 16S amplicon data compiled from multiple sources by Chaffron and coworkers. Again, we observed a clear trade-off between competition and cooperation (Supp. Fig. 3)." To me, Fig 3S looks quite different from Fig 1, and I wonder why. In particular, why are there fewer competitive communities in Chaffon et al?

Reviewer #2 (Remarks to the Author):

The paper presented by Machado et al. tests the nature and prevalence of metabolic competition and cooperation in microbial communities, with up to 40-member species, across thousands of habitats. The authors assess this using Earth Microbiome Project (EMP) datasets and through simulating genome-scale metabolic modelling. Specifically, this vast dataset coverage allowed the authors to measure the impact of metabolic interactions in the assembly of natural communities. Interestingly, results of different community sizes show a trade-off between competition and cooperation.
In carefully planned simulations that exclude the possibility of any biases in the EMP datasets, authors further computed independent datasets to cross-check the trade-off pattern. The authors presented two different community types, clustering at the opposite ends in the trade-off function between competition and cooperation. One end comprises cooperative types categorized by smaller genomes, several auxotrophies, and which are mainly found in free-living and host-associated environments. On the other end, competitive types are characterized by larger genomes, overlapping nutritional requirements, and occurs mainly in soils. Furthermore, they showed that cooperative communities generally show less resource overlap than competitive types that display high resource overlap. Next, they investigated the role of phylogeny and interspecific metabolic interactions. Remarkably, they observed that inter-phylum interactions seem to be more common than expected by chance. Finally,

the authors simulate community-scale flux in response to perturbations. The resulting data showed that cooperative communities were susceptible to invasion, but robust to abiotic perturbations, while competitive types were more resistant to invasion, albeit sensitive to abiotic perturbations.
These results are exciting and provide clear evidence for a bifurcation between metabolic cooperation and competition in the assembly of natural communities. The simulations have been carefully performed and analysed, the statistical analysis is sound, and the data is clearly presented. The presented data represents a rich source for other studies working on microbiomes and microbial communities in general. Thus, I recommend publication, given that the following minor points are being addressed.

Comments:
(1) The authors used the terms competition and cooperation throughout the manuscript, and occasionally metabolic cooperation is interchanged with metabolic cross-feeding. The authors should include a statement that clearly defines what they mean with either term in the present study. For instance, metabolic cross-feeding is not always cooperative. This distinction should be made clear to avoid any confusion.

(2) The authors computed the metabolic resource overlap (MRO) and the metabolic interaction potential (MIP) for each community. MRO and MIP are used as proxies for competition and cooperation, respectively. In this context, the authors should clarify that, within this definition, species with a "higher cooperation potential" in a certain community, can have a "low cooperation potential" in a different one. The "cooperation potential" (and also the competitive potential) is not an absolute propriety of a given species, but instead is relative to the context given by other species in the community. The authors should highlight this distinction, especially when using terms such as "higher cooperation potential", which could be interpreted as a species with a high potential to operate as a cooperator across different communities. This can be especially confusing when the reader is introduced to these concepts in section 3 of the results.

(3) The metabolic interaction potential (MIP) is given by an estimation for the number of metabolites that can be exchanged among community members. However, not all metabolites are equivalent, and interchanging certain specific metabolites can have a crucial impact on the evolution of the division of labour between community members. This distinction should be at least discussed when describing the scope of MIP as a proxy for cooperation.

(4) The authors suggest that the evolution of metabolic competition and cooperation is expected to be driven by nutrient availability in the habitat. As a consequence, competitive communities are mostly present in free-living environments. On the other hand, cooperative communities are expected to evolve and be stable in host-associated habitats, such as the human microbiome. However, this pattern is in stark contrast with previously published findings. For example, Coyote (Coyote et al. Science 350 (6261 663-6) 2015) found that although cooperating networks of microbes from the human gut can be efficient, they are often unstable. The ecological and evolutionary processes responsible for such contrasting findings should be discussed. The current version of the manuscript lacks an evolutionary interpretation of the mentioned, potentially conflicting results obtained.

In general I find the discussion way too short. The authors should extend this section of their paper by better linking their work to previous findings in the field. This is good practice and will help the reader to evaluate the presented findings and their novelty.

(5) As previously mentioned, cooperative communities are expected to evolve in host-associated habitats, where nutritional richness supports the more "cooperative black queen species." In such a context, microbes will evolve dependencies on molecules that are present in the environment. Such a situation is different from dependencies on the availability of molecules produced by other community members, a process favoured by metabolic complementarity and division of labour. The authors should discuss and distinguish between these two situations since they can produce opposite patterns regarding the expected sensitivity to abiotic perturbations. While the latter will result in robustness to abiotic perturbations, the opposite is true for the former. This distinction should be linked with the discussion on their analysis in section 5 from the results, where they ask whether auxotrophies precede community assembly, or if they are a consequence of species co-evolution.

(6) A potentially critical bias introduced in the present version of the analysis is the way the abiotic perturbation is implemented: by adding nutrients. If the abiotic perturbation is instead implemented by a decrease in nutrients, the pattern could be opposite to the one reported. A decrease in nutrients that auxotrophs rely on (for example, a decrease in the amino acids) will have a clear detrimental effect on the cooperative community, contradicting the current claim made in the paper. The authors should take this into account and at least discuss this possibility.

(7) The authors simulated communities with up to 40 member species across diverse habitats; however, there is no information available about which species they modelled, nor which metabolic network they used to simulate the genome-scale metabolic modelling. This information should be included.

(8) The present paper introduces some statements about the black queen hypothesis (BQH) and the red queen hypothesis (RQH). The authors used the terms in the introduction and discussion, however, they did not discuss it further. Thus, it remains unclear how both hypotheses correlate with the results obtained in this manuscript.
a. The authors should cross-check the cited references throughout the manuscript; for instance, the authors cite references 37 and 38 for the red queen hypothesis and the black queen hypothesis. However, these references are not the right to show both the ideas.
b. The RQH focuses on antagonistic co-evolution, while the BQH provides interpretations regarding the evolution of dependencies through adaptive gene loss in free-living organisms. The authors should make a clear argument on how the metabolic competition and cooperation trade-off landscape, shown in the manuscript, illuminate both evolutionary conflict or stability in natural communities.
c. I have never seen anyone referring to the red queen hypothesis in the context of competitive interactions. Even though this is clearly an example of an antagonistic interaction, the red queen has been formulated for the specific case of host parasite interactions. Here, selection pressures and population dynamics are very different from a competitive interaction. For examples, there is no gene-by-gene coevolution going on in cases of competitive interactions and also no oscillating population dynamics. Thus, I would consider removing this analogy completely from the manuscript and reword the corresponding sections.

(9) The reference number 7 seems mistakenly cited.

(10) Statistical or data analysis section: it would be important that the authors describe how they have tested the assumptions of the statistical tests. This should be included as a part of the method sections.

(11) The letters to denote statistical differences are very confusing. It is generally unclear what has been compared with what. The figure legends should explain, which datasets have been compared, the number of samples, P values, degree of freedom, statistical test used.

For example, in the results part discussing Fig. 5, the authors talk about trends and differences, yet neither in the figure nor in the text do they show the result of a statistical test. This should be carefully checked throughout the whole manuscript and corrected accordingly.

(12) It would be essential to make some statements about how authors defined the minimal medium composition in the case of such diverse community types and how this information is integrated into the simulation. If it is the case that previous research work assisted on this, it has to be cited.

(13) I find the introduction lacks a clear hypothesis/ research question. I think the manuscript would greatly benefit by better linking the current study into the existing literature and clearly describing what the motivation and research question was.

Reviewer #3 (Remarks to the Author):

Referee report for: Machado D. et al, Polarization of microbial communities in the competition-cooperation trade-off landscape

Summary:
Machado et. al report an analysis of metabolic models constructed from co-occurring natural microbial consortia, where the authors predict metrics associated with upper bounds on metabolic competition and facilitation. The authors observe an interesting trade-off in these two metrics for large communities, where co-occurring microbial consortia tend to cluster towards either a (1) high (low) or a (2) low (high) competition (cooperation) metric. The authors then analyze properties of the community members in these consortia types, and then show that type (1) consortia are more often found in free-living communities, while type (2) consortia are more often found within host-associated communities. The authors then analyze key differences between cooperative and competitive consortia, including differences in genomic/metabolic features of community members and their interactions, and differences in robustness to abiotic and biotic perturbations.

Overall, the paper presents a novel and interesting approach to study the higher-order structure of naturally co-occurring groups of microbes, but would benefit from an analysis of potentially several technical confounding factors that make it challenging to interpret their results. Furthermore, additional controls and analysis that investigate how media influences their results would greatly strengthen the paper. Below is a more thorough discussion of the main comments.

Major concerns/questions:
1. Are biases in model reconstruction for members within competitive and cooperative consortia driving the polarization presented in Figure. 1?
One could hypothesize that the two types of microbial consortia (cooperative and competitive) that emerge in this study are caused by biases in annotation quality and model reconstruction. If microbes are mapped to genomes with poorer annotation quality, then you would predict the species would have more auxotrophies. I'm concerned that the key results presented in this paper are an artifact of

genome annotation or model reconstruction biases for specific taxa found primarily in cooperative consortia (which appear to be mainly found in host-associated environments). Thus, a useful control would be to see whether taxa in cooperative consortia have lower quality genome annotation (and model reconstruction) than taxa more often observed in competitive consortia.

Figure 4 shows that the members within the competitive and cooperative communities are broadly distributed across the phylogenetic tree, but this does not show the frequency of each taxon in either competitive or cooperative consortia, which would show whether there is a quantitative bias between taxon frequency (i.e. how often a taxon is observed in a consortia type) and the consortia type.

However, there appears to be evidence in the data presented in Figure 4b that there is enrichment for Firmicutes in the cooperative communities, and Proteobacteria for the competitive communities. Based on the previous publication describing CarveMe (Machado et al. 2018), the authors report that the algorithm is much better at predicting the nutrient uptake capabilities for Escherichia coli or Pseudomonas aeruginosa (Proteobacteria) compared to Bacillus subtilis (Firmicutes), due to challenges in annotating transporters for B. subtilis (Figure 2a vs. Figure 2b-c (Machado et al. 2018)). Figure 4b in this manuscript shows that Firmicutes spp. are highly abundant in cooperative communities.

Altogether, this suggests the inclusion of a few key taxa (most likely Frimicutes spp.) into the cooperative communities could drive the polarization observed in this study. Thus, this referee would like to see whether the competitive and cooperative consortia are enriched with particular taxa at various taxonomic levels, and a more thorough assessment of how differences in model quality may affect their results.

2. MIP might have nothing to do with metabolic facilitation or cross-feeding, but rather might reflect the degree of dependencies on abiotically supplied nutrients.

From this referee's understanding, the MIP is designed to measure the upper bound on cross-feeding interactions, given the minimal nutrient environment required to sustain growth of each member as described in the original publication (Zelezniak et al. 2015) . However, microbial communities used in this study are sampled from environments with a large degree of environmental variability. The metabolic outputs from flux balance models can be particularly sensitive to media choices, making it reasonable to assume that the MIP and MRO metrics are also sensitive to the choice of the minimal media compositions. In cases where consortia predominantly reside in a nutrient rich environment, the MIP may not reflect the extent of metabolic facilitation, but rather simply indicate that a large proportion of the nutrients are supplied abiotically or via a host.

In this referee's opinion, it would be highly informative to explore the relationship between the metrics computed in this study, controlling for different environmental contexts. Specifically, exploring the relationship between computed MIP and the nutrient availability would be very informative as well. For instance, it might be worth highlighting any communities with high MIP scores from nutrient poor, free-living communities, which would indicate MIP may actually serve as a metric capturing the degree of cross-feeding, rather than simply just abiotic nutrient dependence.

3. Unclear motivation for using a test for compositional stability as validation of cooperation consortia

The assumption that cooperative groups of taxa should be more stable in composition than competitive consortia lacks clear rationale. If anything, one could argue that interactions between taxa

predicted to be cooperating should display positive correlation in similar habitats, while members in a competitive consortium should be weakly or even anti-correlated in similar habitats.

4. Robustness of cooperative consortia to abiotic perturbations.

Since there may be biases in model reconstruction for specific taxa within cooperative consortia (see point 1), the observed robustness of cooperative consortia to abiotic perturbations might simply reflect that models within these communities contain fewer nutrient transporters. Thus, adding additional media components might not be metabolizable by community members. It would be informative to know whether the number (and types) of annotated transporters is associated with the robustness to abiotic perturbations.

5. Investigating robustness of SMETANA scores to environmental variation.

The results presented in Figure 5 show the predicted robustness of consortia to variation in abiotic or biotic perturbations under a single environmental condition. It would be very informative to see the same analysis done in a variety of environments, as it would give a sense of whether the results presented in the paper are generalizable across multiple environments. Specifically, the original paper describing the method (Zelezniak et al. 2015) performed the simulations in anaerobic and aerobic conditions. It would be very interesting for the authors to explore this factor in their manuscript, as this is likely a key environmental variable that shapes the structure of natural microbial communities.

6. Lack of controls used in the study make it challenging to interpret results

The authors used a very interesting metabolic modeling approach to characterize consortia of microbial communities, using only knowledge of metabolic models and co-occurrence patterns in natural environments. However, since there are no consortia included in the analysis with known competitive or cross-feeding interactions, it remains challenging to interpret the MIP and MRO metric. Thus, it would be very informative to include an analysis of a “control” consortia with known cross-feeding interactions, such as the engineered community of E. coli amino acid auxotrophs (Mee et al. 2014), or natural microbial consortia with inferred cross-feeding interactions (Embree et al. 2015). Additionally, an analysis that would be very helpful for the readers would be to compute the MIP and MRO metrics for communities of monocultures of size N, where communities are constructed using N copies of the same metabolic model.

Minor comments:

7. In Figure 1, what is the difference between green/orange communities? I assume these correspond to competitive/cooperative communities, but there doesn’t appear to be an explanation for how communities are classified.

8. The authors should cite a related paper: Frielich et al, Competitive and cooperative metabolic interactions in bacterial communities, Nature Communications 2011. (Freilich et al. 2011)

9. The citation number for the validation dataset used (Chaffron et al, ref 21) appears inconsistent.

References
Embree, Mallory, Joanne K. Liu, Mahmoud M. Al-Bassam, and Karsten Zengler. 2015. “Networks of Energetic and Metabolic Interactions Define Dynamics in Microbial Communities.” Proceedings of the National Academy of Sciences 112 (50): 201506034.
Freilich, Shiri, Raphy Zarecki, Omer Eilam, Ella Shtifman Segal, Christopher S. Henry, Martin Kupiec, Uri Gophna, Roded Sharan, and Eytan Ruppin. 2011. “Competitive and Cooperative Metabolic Interactions in Bacterial Communities.” Nature Communications 2 (December): 589.
Machado, Daniel, Sergej Andrejev, Melanie Tramontano, and Kiran Raosaheb Patil. 2018. “Fast Automated Reconstruction of Genome-Scale Metabolic Models for Microbial Species and Communities.” Nucleic Acids Research 46 (15): 7542–53.
Mee, M. T., G. M. Church, J. J. Collins, and H. H. Wang. 2014. “Syntrophic Exchange in Synthetic Microbial Communities.” Proceedings of the National Academy of Sciences of the United States of America 111 (20): E2149–56.
Zelezniak, Aleksej, Sergej Andrejev, Olga Ponomarova, Daniel R. Mende, Peer Bork, and Kiran Raosaheb Patil. 2015. “Metabolic Dependencies Drive Species Co-Occurrence in Diverse Microbial Communities.” Proceedings of the National Academy of Sciences of the United States of America 112 (20): 6449–54.

**********END**********

Author Rebuttal to Initial comments

Response to reviewers’ comments

We thank the reviewers for their constructive and helpful comments, which very much helped us to improve the clarity and to strengthen the conclusions. The point-by-point response is provided below. The corresponding changes in the manuscript are copied where needed. We also edited the manuscript for clarity; the main changes in the text are highlighted in the revised manuscript file. We have also included a separate document marking all changes.

Reviewer #1 (Remarks to the Author):

In this paper, the authors identify and analyze communities of co-occurring species that can be found in many different locations on the planet. Their main conclusion is that such communities roughly come in

two types: cooperative communities, in which the co-occurring species tend to have different nutrient requirements but help each other by producing byproducts that others species in the community can consume (cross-feeding), and competitive communities, in which co-occurring species tend to compete for a limited number of nutrients. The authors show that competitive communities mainly occur in soil, whereas cooperative communities can live anywhere, including host-associated microbiomes. They also argue that cooperative communities are less affected by changes in the nutritional environment than competitive communities, but are more prone to invasion by other species.

This is a very interesting and well written paper that will be of interest to a wide audience.

→ We thank the reviewer for their positive assessment and constructive comments.

My caveat is that I am not well-versed in the software packages that the authors use for their analyses, so I am unable to verify the results (the devil usually lies in the details...). I have no reason to think that their analysis is wrong, and I think they developed an interesting novel way to actually identify co-occurring communities (rather than just, say co-occurring species pairs).

→ All the algorithms and codes used in this study are available in the public domain (https://github.com/cdanielmachado/cooccurrence). Furthermore, our results from metabolic modelling are supported in various ways such as genomic patterns (e.g. network size), phylogenetic analysis, concordance of key results like amino acid exchange with previous experimental studies.

There is one thing that was missing for me: the authors don't mention bacteriocins at all. It is well known that bacterial species try to poison co-occurring species using many different kinds of toxins, including bacteriocins. I think it would be very interesting to overlay the results in this paper with an analysis of bacteriocin production, which could probably be included in the genomic analysis. One might e.g. expect that competitive communities produce more bacteriocins than cooperative communities. I'm not sure whether it is doable, but I think that an analysis in this direction might strengthen the paper even further.

→ We had investigated the enrichment of biosynthetic gene clusters (BGCs), but did not include the results in the previous version of the manuscript as we believe these to be a subject for a dedicated study. Nevertheless, we now include some results here in support of our main topic (i.e. metabolic competition and cooperation amongst co-occurring groups). When systematically searching for individual classes of BGCs (including bacteriocins), we find an enrichment for BGCs annotated as lanthipeptides (a subclass of bacteriocins). We have now added the following paragraph to the manuscript.

"Competition for nutrients would be expected to be linked with other, more direct, modes of competition such as production of antimicrobial compounds[31]. To test this, we annotated all species considered in this study with biosynthetic gene clusters from the antiSMASH database32. Supporting the hypothesis, a higher-than-expected number of genes were found to be associated with lanthipeptide production in species comprising competitive communities (odds ratio 1.7; $p < 0.001$, hypergeometric test and Benjamini–Hochberg correction for multiple testing). We also find that, across all species in the EMP dataset, the number of biosynthetic gene clusters encoded in the genome is correlated with the genome size (Spearman's $r = 0.67$, $p < 0.001$) and with the number of metabolic genes (Spearman's $r = 0.57$, $p < 0.001$). This supports the notion that the competition for resources and active antagonism are closely linked."

Minor comment: The authors say somewhere (line numbers would be nice!) that they "computed co-occurring communities using an independent collection of 16S amplicon data compiled from multiple sources by Chaffron and coworkers. Again, we observed a clear trade-off between competition and cooperation (Supp. Fig. 3)." To me, Fig 3S looks quite different from Fig 1, and I wonder why. In particular, why are there fewer competitive communities in Chaffon et al?

→ The trade-off is still evident for randomly-assembled communities, but now the co-occurring communities are all on the cooperative side. This is related to the quality of this particular dataset (a higher abundance cut-off was applied, which removed low abundance species). Since we only show later in the manuscript that abundance is one of the distinguishing features between cooperative and competitive communities, we avoided forward referencing at that stage. Nonetheless, we agree with the reviewer that not giving any explanation could make it difficult to understand the results. To address this, we rewrote the paragraph as follows:

"To check if the observed trade-off pattern results from any biases in the EMP data (such as the habitats covered, experimental protocols, or data processing pipelines), we computed co-occurring communities using an independent collection of 16S amplicon data compiled from multiple sources by Chaffron and co-workers[30]. This analysis also showed a clear trade-off between competition and cooperation with the randomly-assembled communities distributed along the spectrum (Supp. Fig. 4). However, in this case, the co-occurring communities are all located in the cooperative pole. This can be explained by the higher abundance cutoff used in this dataset which – as we discuss in the next section – is one of the distinguishing features between the cooperative and competitive communities."

Reviewer #2 (Remarks to the Author):

The paper presented by Machado et al. tests the nature and prevalence of metabolic competition and cooperation in microbial communities, with up to 40-member species, across thousands of habitats. The authors assess this using Earth Microbiome Project (EMP) datasets and through simulating genome-scale metabolic modelling. Specifically, this vast dataset coverage allowed the authors to measure the impact of metabolic interactions in the assembly of natural communities. Interestingly, results of different community sizes show a trade-off between competition and cooperation.

In carefully planned simulations that exclude the possibility of any biases in the EMP datasets, authors further computed independent datasets to cross-check the trade-off pattern. The authors presented two different community types, clustering at the opposite ends in the trade-off function between competition and cooperation. One end comprises cooperative types categorized by smaller genomes, several auxotrophies, and which are mainly found in free-living and host-associated environments. On the other end, competitive types are characterized by larger genomes, overlapping nutritional requirements, and occurs mainly in soils. Furthermore, they showed that cooperative communities generally show less resource overlap than competitive types that display high resource overlap. Next, they investigated the role of phylogeny and interspecific metabolic interactions. Remarkably, they observed that inter-phylum interactions seem to be more common than expected by chance. Finally, the authors simulate community-scale flux in response to perturbations. The resulting data showed that cooperative communities were susceptible to invasion, but robust to abiotic perturbations, while competitive types were more resistant to invasion, albeit sensitive to abiotic perturbations.

These results are exciting and provide clear evidence for a bifurcation between metabolic cooperation and competition in the assembly of natural communities. The simulations have been carefully performed and analysed, the statistical analysis is sound, and the data is clearly presented. The presented data represents a rich source for other studies working on microbiomes and microbial

communities in general. Thus, I recommend publication, given that the following minor points are being addressed.

→ We are grateful to the reviewer for the positive assessment and excellent suggestions for improving the manuscript.

Comments:

(1) The authors used the terms competition and cooperation throughout the manuscript, and occasionally metabolic cooperation is interchanged with metabolic cross-feeding. The authors should include a statement that clearly defines what they mean with either term in the present study. For instance, metabolic cross-feeding is not always cooperative. This distinction should be made clear to avoid any confusion.

We agree, and apologize for not making this clearer. Indeed, the impact of cross-feeding interactions in the fitness of the respective species is context-dependent and might not always be reflective of cooperation. We now make this clearer when introducing the definition of MIP and MRO:

"While the MRO quantifies the similarity of the nutritional requirements between all species in a community, reflecting the intra-community risk for resource competition, the MIP indicates the number of metabolites that can be exchanged among the community members to decrease their dependency on the abiotic environment. MRO and MIP estimate the respective interaction metrics at their theoretical limit and thus do not require the information regarding the resources actually available in the habitat. The operating degree of competition and cooperation in a given community will be thus habitat dependent. The context-independent nature of MRO and MIP makes these suitable for application to co-occurring communities spanning multiple and diverse habitats."

(2) The authors computed the metabolic resource overlap (MRO) and the metabolic interaction potential (MIP) for each community. MRO and MIP are used as proxies for competition and cooperation, respectively. In this context, the authors should clarify that, within this definition, species with a "higher cooperation potential" in a certain community, can have a "low cooperation potential" in a different one. The "cooperation potential" (and also the competitive potential) is not an absolute propriety of a given species, but instead is relative to the context given by other species in the community. The authors should highlight this distinction, especially when using terms such as "higher cooperation potential",

which could be interpreted as a species with a high potential to operate as a cooperator across different communities. This can be especially confusing when the reader is introduced to these concepts in section 3 of the results.

(3) The metabolic interaction potential (MIP) is given by an estimation for the number of metabolites that can be exchanged among community members. However, not all metabolites are equivalent, and interchanging certain specific metabolites can have a crucial impact on the evolution of the division of labour between community members. This distinction should be at least discussed when describing the scope of MIP as a proxy for cooperation.

→ We believe the clarification in the revised paragraph quoted in response to point (1) also addresses (2) and (3).

(4) The authors suggest that the evolution of metabolic competition and cooperation is expected to be driven by nutrient availability in the habitat. As a consequence, competitive communities are mostly present in free-living environments. On the other hand, cooperative communities are expected to evolve and be stable in host-associated habitats, such as the human microbiome. However, this pattern is in stark contrast with previously published findings. For example, Coyote (Coyote et al. Science 350 (6261 663-6) 2015) found that although cooperating networks of microbes from the human gut can be efficient, they are often unstable. The ecological and evolutionary processes responsible for such contrasting findings should be discussed. The current version of the manuscript lacks an evolutionary interpretation of the mentioned, potentially conflicting results obtained.

→ We note that the cooperative communities have a broader habitat range, including both host-associated and free-living environments; competitive communities, on the other hand, seem to be largely restricted to free-living environment (Figure 3). This is consistent with the higher MIP score for cooperative groups as this makes them more independent of the environment and, thus, they can more easily thrive in different habitats.

The apparent contrast between our results and those of Coyte et al. stems from the fact that we are analyzing communities of co-occurring species, while their study considers complete communities. The groups of co-occurring species would still be competing with other species present in the respective samples/environments. In fact, according to the predictions from Coyte et al. (fig 3 in their paper), the introduction of competitive species can help to stabilize a network of cooperators. This means that our consortia of cooperative species can be stabilized by other competitors present in the environment, and it can explain why we observe the abundance of these cooperative communities to be stable over time. The stability and cooperation is indeed an important point and we have included the following paragraph to elaborate on this:

"Previous theoretical work on community stability has shown that purely cooperative communities should be unstable and competitive interactions are required to re-establish stability[40]. Given that communities of co-occurring species exist as part of larger microbiomes, it is likely that competitive interactions with other species play a role in their stabilization. Moreover, as all environment would ultimately be nutrient limited (at least periodically), there would be stabilizing negative feedback even for the cooperative groups."

In general I find the discussion way too short. The authors should extend this section of their paper by better linking their work to previous findings in the field. This is good practice and will help the reader to evaluate the presented findings and their novelty.

→ We agree that the discussion could be further expanded. However, given the limit on permissible manuscript length, we have tried to do this in a concise manner. Two citations had been accidentally omitted while revising the introduction and are now restored (Freilich et al, 2011; Levy and Borenstein 2013).

(5) As previously mentioned, cooperative communities are expected to evolve in host-associated habitats, where nutritional richness supports the more "cooperative black queen species." In such a context, microbes will evolve dependencies on molecules that are present in the environment. Such a situation is different from dependencies on the availability of molecules produced by other community members, a process favoured by metabolic complementarity and division of labour. The authors should discuss and distinguish between these two situations since they can produce opposite patterns regarding the expected sensitivity to abiotic perturbations. While the latter will result in robustness to abiotic perturbations, the opposite is true for the former. This distinction should be linked with the

discussion on their analysis in section 5 from the results, where they ask whether auxotrophies precede community assembly, or if they are a consequence of species co-evolution.

→ As we mention above in response to (4), the cooperative groups have a broader habitat range and are not restricted to host-associated microbiomes (Figure 3). The adaptive gene loss in these communities, therefore, need not originate in host-associated context. We rather see these groups as 'modules' that can move together from habitat to habitat. We also note that host-association is not necessarily equivalent to nutritional richness. For example, the degree of nutrient availability considerably varies in space and time in the human GI tract.

Our investigation into the origin of auxotrophies led to the conclusion that both gene loss and assembly of species with prior auxotrophies are at play. Thus, indeed we cannot rule out that some of the auxotrophies are acquired due to availability of the nutrient in question from abiotic sources or from the host (i.e. not from other microbes). We have now modified the corresponding text as follows:

"While a majority of amino acid auxotrophies (~90%) seem to have been inherited (Supp. Fig. 10), we also observe a few cases (12 in total) indicative of recent auxotrophy acquisition. The former implies pre-existing auxotrophies that were acquired (or retained) due to the availability of the corresponding nutrients from either abiotic or biotic environment. The latter suggests adaptive gene loss and were found to be most frequent for proline (G. haemolysans, L. hominis, L. inners) and methionine (A. tetradius, M. luteus, R. dentocariosa). Both, the assembly of species with pre-existing auxotrophies, and adaptive gene loss, thus appear to have contributed to the establishment of natural communities."

To further address whether the amino acid auxotrophies in cooperative groups could be due to availability from host or abiotic sources, we analyzed the complementarity of amino acid auxotrophies with respect to the secretion by other community members (Supp. Fig. 11) and observe that this complementarity is higher than expected by chance in cooperative communities (i.e., there is a higher probability for species to be auxotrophic for the amino acids secreted by other community members), which is a strong indication for division of labor. We extended the respective results section with the following paragraph:

"We further explored whether autroxophy retention or acquisition was mainly driven by the environmental availability or by the secretion from co-occurring species. We measured the fraction of

amino acid auxotrophies within a community that can be fulfilled through other community members (Supp. Fig. 11) and observed that this fraction follows a saturation curve as a function of community size (i.e. for a sufficiently large community size, eventually all auxotrophies can be fulfilled by other members). Furthermore, we observe that this curve saturates earlier for cooperative communities compared to competitive and randomly-assembled communities. Therefore, the amino acid auxotrophies between species present in cooperative communities are more complementary than expected by chance, indicating division of labor."

Further, we find that the members of the cooperating groups maintain stable (relative abundance) which supports that the predicted inter-dependencies are likely operational in situ.

(6) A potentially critical bias introduced in the present version of the analysis is the way the abiotic perturbation is implemented: by adding nutrients. If the abiotic perturbation is instead implemented by a decrease in nutrients, the pattern could be opposite to the one reported. A decrease in nutrients that auxotrophs rely on (for example, a decrease in the amino acids) will have a clear detrimental effect on the cooperative community, contradicting the current claim made in the paper. The authors should take this into account and at least discuss this possibility.

→ In our simulations, cooperative communities do respond to abiotic perturbations, but not more than random controls, and less than competitive communities. This can be explained by the complementarity of their metabolic networks, allowing them to more easily switch between obtaining a resource from the environment or from the fellow community members. Following the reviewer's suggestion, we now also tested the effect of removal of compounds (Supp. Fig. 12), and observe the same pattern as before, i.e., competitive communities are more sensitive to abiotic perturbations than random controls and cooperative communities. Also, following the suggestion from reviewer #3, we repeated all simulations in aerobic and anaerobic environments. We observe a lower sensitivity to both abiotic and biotic perturbations in the anaerobic environment, but the overall pattern is still the same (i.e. competitive communities more sensitive to abiotic perturbations and cooperative more sensitive to biotic perturbations).

(7) The authors simulated communities with up to 40 member species across diverse habitats; however, there is no information available about which species they modelled, nor which metabolic network they used to simulate the genome-scale metabolic modelling. This information should be included.

→ The list of species in each environment and the respective genome-scale metabolic models can be found in the supplementary github repository that is mentioned at the end of the manuscript (https://github.com/cdanielmachado/cooccurrence). We also have all the code available, hoping not only to make this study reproducible, but also to make all the tools and models used in this study a useful resource to the community.

(8) The present paper introduces some statements about the black queen hypothesis (BQH) and the red queen hypothesis (RQH). The authors used the terms in the introduction and discussion, however, they did not discuss it further. Thus, it remains unclear how both hypotheses correlate with the results obtained in this manuscript.

a. The authors should cross-check the cited references throughout the manuscript; for instance, the authors cite references 37 and 38 for the red queen hypothesis and the black queen hypothesis. However, these references are not the right to show both the ideas.

→ We thank the reviewer for spotting this error, it was a problem with the reference management software and is now corrected.

b. The RQH focuses on antagonistic co-evolution, while the BQH provides interpretations regarding the evolution of dependencies through adaptive gene loss in free-living organisms. The authors should make a clear argument on how the metabolic competition and cooperation trade-off landscape, shown in the manuscript, illuminate both evolutionary conflict or stability in natural communities.

→ This is a very good suggestion for improving the clarity of our manuscript. We have appended the Discussion section as follows:

"*In the context of metabolism, a scenario consistent with the red queen hypothesis is retention of diverse metabolic capabilities to exploit the available nutrients, which indirectly antagonizes competitors, and to reduce dependencies on other species. The members of the competitive communities also harbor more potential for antimicrobial compounds indicating active antagonism. In cooperative communities, falling under the black queen scenario, our phylogenetic analysis provides evidence for adaptive gene loss in metabolic networks. Consistent with adaptive process, auxotrophies for amino acids with high biosynthetic costs are more common (Supp. Fig. 9). Further, the cooperative groups harbor complementary auxotrophies and exhibit stable proportions across habitats, in line with inter-species*

dependencies. The advantage of the interdependencies in this group is reflected in their high relative abundance (Figure 2d). Collectively, metabolic capabilities, antimicrobial production potential, phylogenetic analysis, and differences in habitat preference and relative abundances, bring forward evolutionary conflict and cooperation in the two co-occurring groups identified in our study."

c. I have never seen anyone referring to the red queen hypothesis in the context of competitive interactions. Even though this is clearly an example of an antagonistic interaction, the red queen has been formulated for the specific case of host parasite interactions. Here, selection pressures and population dynamics are very different from a competitive interaction. For examples, there is no gene-by-gene coevolution going on in cases of competitive interactions and also no oscillating population dynamics. Thus, I would consider removing this analogy completely from the manuscript and reword the corresponding sections.

→ Although formulated in the case of host-parasite interactions, the original definition by Van Valen (Valen LV, Evol. Theory, 1973) is sufficiently generic to include other situations of competition, as supported by the following paragraph taken from the original manuscript:

"It is selectively advantageous for a competitor for resources in short supply (food and space in the broadest senses, and sometimes also externally supplied adjuncts to reproduction or dispersal) both to increase its own effect on its competitors and to decrease the effect of its competitors on itself (33). Every species does the best it can in the face of these pressures. Probably all species are affected importantly by them at least over intervals of a few generations."

The Valen paper also suggests examples of competition in the animal kingdom such as the replacement of Odd-toed ungulates (Perissodactyla) with even-toed ungulates (Artiodactyla). Therefore, we do not believe it is inappropriate to use the Red Queen analogy in the context of metabolic competition between bacteria.

(9) The reference number 7 seems mistakenly cited.

→ Corrected.

(10) Statistical or data analysis section: it would be important that the authors describe how they have tested the assumptions of the statistical tests. This should be included as a part of the method sections.

→ Most of the statistical tests used throughout the paper were used to compare the multiple bell-curve distributions presented in Figure 2, Supp Fig 1, and Supp Fig 7, namely Cohen's d to evaluate effect sizes and Welch's t-test for statistical significance of difference between the means. Since these are parametric tests and require the assumption of normally distributed variables, we tested all these data using the Shapiro-Wilk normality test and observed that despite their gaussian-shaped curves they failed the normality test. Therefore, we replaced the use of Cohen's d with fold-changes to report effect sizes, and Welch's t-test with the non-parametric Mann-Whitney U-test for significance. Apart from that, two other tests we used in the paper were Spearman's correlation and Wilcoxon signed-rank test, which are also non-parametric tests. The used tests are mentioned throughput the manuscript when reporting p-values and effect sizes. We also added just the following mention in the methods section: *"All statistical tests used in this manuscript were performed with SciPy version 1.2.1."*

(11) The letters to denote statistical differences are very confusing. It is generally unclear what has been compared with what. The figure legends should explain, which datasets have been compared, the number of samples, P values, degree of freedom, statistical test used. For example, in the results part discussing Fig. 5, the authors talk about trends and differences, yet neither in the figure nor in the text do they show the result of a statistical test. This should be carefully checked throughout the whole manuscript and corrected accordingly.

→ All the tests used are non-parametric and they are now mentioned in the main text together with the results. Regarding Figure 5, it was indeed missing statistical testing as we believed the reported trends could be easily observed by visual inspection. Nevertheless, we agree with the reviewer that statistical testing would be appropriate in this case, and we now report both the fold-changes between cooperative/competitive communities and controls and the respective statistical significance using Mann-Whitney U-test.

(12) It would be essential to make some statements about how authors defined the minimal medium composition in the case of such diverse community types and how this information is integrated into the simulation. If it is the case that previous research work assisted on this, it has to be cited.

→ The minimal medium composition is calculated as the minimal number of compounds that enables biomass production for all community members (among all compounds that the community members would, in theory, be able to uptake). During simulation, the uptake rate for any compounds that are not part of the medium formulation is set to zero. The calculation of the minimal media is formulated as a MILP minimization problem, and is implemented as part of the SMETANA simulation tool. We re-wrote part of the methods section to make this clearer:

"All community simulations and calculations of minimal media composition were performed using SMETANA v1.0. The mathematical formulation, as well as the description of the different scores, such as MIP and MRO, are described in the original publication (23). The tool is implemented as a standalone python package, openly available at https://github.com/cdanielmachado/smetana."

(13) I find the introduction lacks a clear hypothesis/ research question. I think the manuscript would greatly benefit by better linking the current study into the existing literature and clearly describing what the motivation and research question was.

→ We have now revised the introduction text to better present and contextualize our hypothesis and goals. The length of introduction had to be limited to keep the overall manuscript length within the permitted limit. We believe that the revised text, though concise, now more clearly outlines the motivation and the research question.

Reviewer #3 (Remarks to the Author):

Referee report for: Machado D. et al, Polarization of microbial communities in the competition-cooperation trade-off landscape

Summary:

Machado et. al report an analysis of metabolic models constructed from co-occurring natural microbial consortia, where the authors predict metrics associated with upper bounds on metabolic competition and facilitation. The authors observe an interesting trade-off in these two metrics for large communities,

where co-occurring microbial consortia tend to cluster towards either a (1) high (low) or a (2) low (high) competition (cooperation) metric. The authors then analyze properties of the community members in these consortia types, and then show that type (1) consortia are more often found in free-living communities, while type (2) consortia are more often found within host-associated communities. The authors then analyze key differences between cooperative and competitive consortia, including differences in genomic/metabolic features of community members and their interactions, and differences in robustness to abiotic and biotic perturbations.

Overall, the paper presents a novel and interesting approach to study the higher-order structure of naturally co-occurring groups of microbes, but would benefit from an analysis of potentially several technical confounding factors that make it challenging to interpret their results. Furthermore, additional controls and analysis that investigate how media influences their results would greatly strengthen the paper. Below is a more thorough discussion of the main comments.

→ We thank the reviewer for their encouraging comments and suggestions towards strengthening our conclusions.

Major concerns/questions:

1. Are biases in model reconstruction for members within competitive and cooperative consortia driving the polarization presented in Figure. 1?

One could hypothesize that the two types of microbial consortia (cooperative and competitive) that emerge in this study are caused by biases in annotation quality and model reconstruction. If microbes are mapped to genomes with poorer annotation quality, then you would predict the species would have more auxotrophies. I'm concerned that the key results presented in this paper are an artifact of genome annotation or model reconstruction biases for specific taxa found primarily in cooperative consortia (which appear to be mainly found in host-associated environments). Thus, a useful control would be to see whether taxa in cooperative consortia have lower quality genome annotation (and model reconstruction) than taxa more often observed in competitive consortia.

Figure 4 shows that the members within the competitive and cooperative communities are broadly distributed across the phylogenetic tree, but this does not show the frequency of each taxon in either competitive or cooperative consortia, which would show whether there is a quantitative bias between taxon frequency (i.e. how often a taxon is observed in a consortia type) and the consortia type.

However, there appears to be evidence in the data presented in Figure 4b that there is enrichment for Firmicutes in the cooperative communities, and Proteobacteria for the competitive communities. Based on the previous publication describing CarveMe (Machado et al. 2018), the authors report that the algorithm is much better at predicting the nutrient uptake capabilities for Escherichia coli or Pseudomonas aeruginosa (Proteobacteria) compared to Bacillus subtilis (Firmicutes), due to challenges in annotating transporters for B. subtilis (Figure 2a vs. Figure 2b-c (Machado et al. 2018)). Figure 4b in this manuscript shows that Firmicutes spp. are highly abundant in cooperative communities.

Altogether, this suggests the inclusion of a few key taxa (most likely Frimicutes spp.) into the cooperative communities could drive the polarization observed in this study. Thus, this referee would like to see whether the competitive and cooperative consortia are enriched with particular taxa at various taxonomic levels, and a more thorough assessment of how differences in model quality may affect their results.

→ We included a new supplementary figure (Supp. Fig. 8) with the taxonomic distribution of species present in cooperative and competitive communities at all taxonomic levels (from Phylum to Genus). We observe that, for both types of communities, the species are widely distributed across different taxa.

Regarding model quality, we observe that the number of reactions in each model is highly correlated with the genome size of the respective organism (Spearman's $r=0.763$, $p < 0.001$), which indicates that the metabolic functionality of the species we analyzed in this study is mainly driven by genetic evidence rather than the metabolic reconstruction process (e.g. gap-filling). Furthermore, any systematic biases regarding gene annotation would affect all species in a similar way, and would not explain why the metabolism of cooperative species is more dissimilar than expected by chance.

To further validate that the predicted auxotrophies are biologically meaningful and not merely an artifact introduced by the model reconstruction, we compared the frequency of amino acid auxotrophies across species present in cooperative communities and the respective fitness cost of each amino acid (using a compilation of AA production costs calculated through different methods by Barton et al), and observe a high correlation between these two variables. This, together with the observed stable relative proportions of cooperating group members, strongly points towards a biological meaning to the presence of these auxotrophies, and provides additional support for the lack of bias in the model reconstruction for these species.

2. MIP might have nothing to do with metabolic facilitation or cross-feeding, but rather might reflect the degree of dependencies on abiotically supplied nutrients.

From this referee's understanding, the MIP is designed to measure the upper bound on cross-feeding interactions, given the minimal nutrient environment required to sustain growth of each member as described in the original publication (Zelezniak et al. 2015). However, microbial communities used in this study are sampled from environments with a large degree of environmental variability. The metabolic outputs from flux balance models can be particularly sensitive to media choices, making it reasonable to assume that the MIP and MRO metrics are also sensitive to the choice of the minimal media compositions. In cases where consortia predominantly reside in a nutrient rich environment, the MIP may not reflect the extent of metabolic facilitation, but rather simply indicate that a large proportion of the nutrients are supplied abiotically or via a host.

In this referee's opinion, it would be highly informative to explore the relationship between the metrics computed in this study, controlling for different environmental contexts. Specifically, exploring the relationship between computed MIP and the nutrient availability would be very informative as well. For instance, it might be worth highlighting any communities with high MIP scores from nutrient poor, free-living communities, which would indicate MIP may actually serve as a metric capturing the degree of cross-feeding, rather than simply just abiotic nutrient dependence.

→ Following the reviewer's suggestion, we searched for communities in free-living environments with high MIP values. Indeed, we found several cases, mostly in soil and seawater samples, including grassland and grapevine soils (MIP up to 27), Arctic seawater (MIP up to 23) and near-shore marine sediments (MIP up to 30). These values are on the high-end of the MIP range (approximately 0-40 when considering all environments). We note that the cooperative communities have a broad habitat range, including both host-associated and free-living environments (Figure 3). This is consistent with the higher MIP score for cooperative groups as this makes them more independent from the environment and thus they can more easily thrive in different habitats.

We further show that amino acid auxotrophies within cooperative communities are more easily fulfilled by other community members than expected by chance (Supp. Fig. 11), and that the relative proportions of the cooperative community members are generally stable across habitats. These signals would not be seen if nutrient requirements were mainly driven by abiotic dependencies.

3. Unclear motivation for using a test for compositional stability as validation of cooperation consortia

The assumption that cooperative groups of taxa should be more stable in composition than competitive consortia lacks clear rationale. If anything, one could argue that interactions between taxa predicted to be cooperating should display positive correlation in similar habitats, while members in a competitive consortium should be weakly or even anti-correlated in similar habitats.

→ We note that our compositional stability analysis (and the only addressable using meta-genomics data) is in the space of relative abundance. If considering absolute abundances, indeed one would expect a positive correlation among mutually dependent partners – which implies that the relative proportions remain the same. This follows from the fundamental stoichiometric constraints on species compositional ratio. In accord, our analysis of the EMP data shows that the relative composition of the members in cooperative consortia is stable. This proportional stability is also a strong support to the modelled interdependency being generally operational in situ. In the case of competitors, we indeed observe no consistent trend as would be expected for weak correlation.

4. Robustness of cooperative consortia to abiotic perturbations.

Since there may be biases in model reconstruction for specific taxa within cooperative consortia (see point 1), the observed robustness of cooperative consortia to abiotic perturbations might simply reflect that models within these communities contain fewer nutrient transporters. Thus, adding additional media components might not be metabolizable by community members. It would be informative to know whether the number (and types) of annotated transporters is associated with the robustness to abiotic perturbations.

→ To address this, we calculated the number of compounds that each species can uptake, accounting for the presence, directionality and promiscuity of the respective transporters. The number is indeed slightly lower for species in cooperative communities (94 compounds on average) compared to all species in the EMP dataset (100 compounds on average), though the difference is only marginally significant (Mann-Whitney U-test: p-value = 0.086). Also, we referred to cooperative communities as robust to abiotic perturbations in comparison to competitive communities. However, their sensitivity is approximately the same as that of the randomly-assembled communities (i.e., they are neither more

nor less robust than expected by chance). Therefore, to be more correct, we now rephrased the expression "robust to abiotic perturbations" with "not too sensitive to abiotic perturbations".

5. Investigating robustness of SMETANA scores to environmental variation.

The results presented in Figure 5 show the predicted robustness of consortia to variation in abiotic or biotic perturbations under a single environmental condition. It would be very informative to see the same analysis done in a variety of environments, as it would give a sense of whether the results presented in the paper are generalizable across multiple environments. Specifically, the original paper describing the method (Zelezniak et al. 2015) performed the simulations in anaerobic and aerobic conditions. It would be very interesting for the authors to explore this factor in their manuscript, as this is likely a key environmental variable that shapes the structure of natural microbial communities.

→ Following the reviewer's suggestion, we repeated the simulations for the abiotic and biotic perturbations in aerobic and anaerobic environments (Supp. Fig. 12). In general, we observe the same pattern as before (competitive communities are more sensitive to abiotic perturbations, whereas cooperative communities are more sensitive to biotic perturbations). Nonetheless, we do observe differences regarding the magnitude of the sensitivity. Competitive communities become less sensitive to abiotic perturbations in anaerobic environments. After analyzing the results in more detail, we could observe an increase in cross-feeding interactions when competitive communities grow anaerobically, which makes them slightly more cooperative. We also observe a decrease in sensitivity for biotic interactions (in all types of communities), most likely because some of the invading species are unable to grow in the anaerobic environment.

6. Lack of controls used in the study make it challenging to interpret results

The authors used a very interesting metabolic modeling approach to characterize consortia of microbial communities, using only knowledge of metabolic models and co-occurrence patterns in natural environments. However, since there are no consortia included in the analysis with known competitive or cross-feeding interactions, it remains challenging to interpret the MIP and MRO metric. Thus, it would be very informative to include an analysis of a “control” consortia with known cross-feeding interactions, such as the engineered community of E. coli amino acid auxotrophs (Mee et al. 2014), or natural microbial consortia with inferred cross-feeding interactions (Embree et al. 2015).

→ We thank the reviewer for the suggestion. We opted to use the experiment by Mee et al as our case-study (regarding the other reference, not only are the interactions inferred, as the reviewer mentions, but also we could not obtain the models in SBML format for simulation). We calculated MIP and MRO for all co-culture experiments. For MIP, the value ranges from 0 to 2, reflecting the number of amino acids that can be exchanged, and MRO varies between 0 and 2/3, depending on the requirement for an additional carbon source (it seems that some mutants can use the respective amino acid as a carbon source, whereas others require an additional carbon source). Although MIP and MRO are calculated in a context-independent manner (i.e. without providing any growth medium information), we observe that MIP is significantly associated with the co-culture yields observed in glucose minimal medium. We added a new supplementary figure (Supp. Fig. 3) with these results as well as illustrative diagrams that explain the MIP / MRO calculation. We also added the following paragraph in the main text:

"We also evaluated the ability of MIP to capture biologically meaningful interactions against the data from a systematic screen of pairwise co-cultures of E. coli mutants with engineered amino acid auxotrophies[28]. We observe that the MIP score is positively associated with higher co-culture growth yields (Supp. Fig. 3). This result, together with the previous comparisons with experimental data[15,29], support the relevance of SMETANA simulations."

Additionally, an analysis that would be very helpful for the readers would be to compute the MIP and MRO metrics for communities of monocultures of size N, where communities are constructed using N copies of the same metabolic model.

→ In this case, the MRO is 1.0, i.e. complete resource overlap (since all species can utilize exactly the same set of metabolites) and MIP is 0, i.e. no interaction potential (although the species can exchange metabolites, this does not result in a decrease of dependencies from the environment). Though these results are clear in simulations, we did not include this in the manuscript since they merely reflect the theoretical limits set by the definitions of MRP and MIP and the mass balance constraints intrinsic to the genome-scale metabolic models.

Minor comments:

7. In Figure 1, what is the difference between green/orange communities? I assume these correspond to competitive/cooperative communities, but there doesn't appear to be an explanation for how communities are classified.

→ We apologize that it was not clear before. This explanation is given in the following paragraph in the main text:

"The co-occurring communities thus segregate into highly competitive (green, Figure 1b-h) and highly cooperative (orange, Figure 1b-h) groups, which also coincide with the two main clusters observed based on species composition (Supp. Fig. 2). This polarization between competitive and cooperative groups suggests adaptation of opposite metabolic strategies by the respective community members."

8. The authors should cite a related paper: Frielich et al, Competitive and cooperative metabolic interactions in bacterial communities, Nature Communications 2011. (Freilich et al. 2011)

→ We mentioned the paper in the introduction, but the reference was accidentally deleted. We have now included the reference again.

9. The citation number for the validation dataset used (Chaffron et al, ref 21) appears inconsistent.

→ This was a problem with the reference management software and has been corrected now, we thank the reviewer for spotting the mistake.

References

Embree, Mallory, Joanne K. Liu, Mahmoud M. Al-Bassam, and Karsten Zengler. 2015. "Networks of Energetic and Metabolic Interactions Define Dynamics in Microbial Communities." Proceedings of the National Academy of Sciences 112 (50): 201506034.

Freilich, Shiri, Raphy Zarecki, Omer Eilam, Ella Shtifman Segal, Christopher S. Henry, Martin Kupiec, Uri Gophna, Roded Sharan, and Eytan Ruppin. 2011. “Competitive and Cooperative Metabolic Interactions in Bacterial Communities.” Nature Communications 2 (December): 589.

Machado, Daniel, Sergej Andrejev, Melanie Tramontano, and Kiran Raosaheb Patil. 2018. “Fast Automated Reconstruction of Genome-Scale Metabolic Models for Microbial Species and Communities.” Nucleic Acids Research 46 (15): 7542–53.

Mee, M. T., G. M. Church, J. J. Collins, and H. H. Wang. 2014. “Syntrophic Exchange in Synthetic Microbial Communities.” Proceedings of the National Academy of Sciences of the United States of America 111 (20): E2149–56.

Zelezniak, Aleksej, Sergej Andrejev, Olga Ponomarova, Daniel R. Mende, Peer Bork, and Kiran Raosaheb Patil. 2015. “Metabolic Dependencies Drive Species Co-Occurrence in Diverse Microbial Communities.” Proceedings of the National Academy of Sciences of the United States of America 112 (20): 6449–54.

Decision Letter, first revision:

29th July 2020

*Please ensure you delete the link to your author homepage in this e-mail if you wish to forward it to your co-authors.

Dear Dr Patil,

Your manuscript entitled "Polarization of microbial communities between competitive and cooperative metabolism" has now been seen again by our 3 reviewers, whose comments are attached. While they are mostly satisifed by the revisions, there are still a few concerns, notably from Reviewer 3, that will need to be addressed before we can make a final decision on publication.

We therefore invite you to revise your manuscript taking into account all reviewer comments.

We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

When revising your manuscript:

* Include a “Response to reviewers” document detailing, point-by-point, how you addressed each reviewer comment. If no action was taken to address a point, you must provide a compelling argument. This response will be sent back to the reviewers along with the revised manuscript.

* If you have not done so already please begin to revise your manuscript so that it conforms to our Article format instructions at http://www.nature.com/natecolevol/info/final-submission. Refer also to any guidelines provided in this letter.

* Include a revised version of any required reporting checklist. It will be available to referees (and, potentially, statisticians) to aid in their evaluation if the manuscript goes back for peer review. A revised checklist is essential for re-review of the paper.

Please use the link below to submit your revised manuscript and related files:

[REDACTED]

Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

We hope to receive your revised manuscript within four to eight weeks. If you cannot send it within this time, please let us know. We will be happy to consider your revision so long as nothing similar has been accepted for publication at Nature Ecology & Evolution or published elsewhere.

Nature Ecology & Evolution is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as ‘corresponding author’ on published papers create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on ‘Modify my Springer Nature account’. For more information please visit please visit www.springernature.com/orcid.

Please do not hesitate to contact me if you have any questions or would like to discuss these revisions further.

We look forward to seeing the revised manuscript and thank you for the opportunity to review your work.

[REDACTED]

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

The authors have addressed my concerns in a way that is acceptable to me. Therefore, I am in favour of publication of this paper.

Reviewer #2 (Remarks to the Author):

The authors have done a great job revising the manuscript in response to all the referee's comments. I appreciate the efforts undertaken by the authors to carefully address and implement all comments and suggestions. The manuscript has greatly improved – particularly due to the incorporation of additional analyses (e.g. Supplemental Figures 11 and 12). The visual presentation of the data is very appealing and the text reads well. The presented results are of a high quality and the reported findings are very exciting. I have no doubts that this paper will attract a lot of attention, since it represents a new benchmark of yet unprecedented depth. Thus, I fully recommend publication of this manuscript in Nature Ecology & Evolution.

Below I have some minor comments that should be viewed as suggestions to further improve the manuscript.

(1) Statistics:
It is good scientific practice to always mention all details of a statistical test, so that the reader can understand and reproduce what has been done. This includes a mention of the type of test used, the P-value, sample size (n or df), as well as t, F, U, Z-values (depending on the statistics used). In the manuscript, the authors frequently just mention verbally that there is a difference (e.g. lines 71, 176, 290/291) without performing a test at all or report a test without mentioning the required statistical details (e.g. lines 158, 160, 166, 224, 225, 245, 246, 304, 309, 315, 316). Please correct this.

(2) Use of the red and black queen hypothesis: The authors use these two theories in a completely wrong context. The BQH is not about cooperation (it is about adaptive gene loss) and the RQH is not about competition (it is exclusively about host-parasite interactions). Even if in van Valen's original paper is "sufficient generic", subsequently the concept is exclusively used to denote interactions, in which organisms need to "continuously change in order to stay at the same place". This is not what is expected from competitive interactions at all. Thus, I would strongly suggest to completely remove a mention of at least the Red Queen. This is absolutely wrong. Also the way the Black Queen is referred to is not 100% correct. Please make sure you do not link it to cooperative interactions, but only to situations, in which gene loss can be adaptive. I specifically refer to the discussion section (line 343-354)(line 364).

(3) The term "cooperation" implies that a behaviour is costly and has evolved to benefit another individual. Something like this cannot be inferred from analysing metabolic networks. Thus, I would strongly recommend to remove the term and replace it with e.g. "synergism" throughout the manuscript or at least define carefully what the authors mean by it.

Minor comments:

- Use of the word "trade-off landscape" (e.g. line 16): I think this an unfortunate term. Please consider rewording it.

Line 10: insert hyphen between "habitat" and "dependent"

Line 144: "member species" sounds odd. Consider revising.

Lines 202, 205, 207, 362: reword “free-living environment”. There are no free-living environments.

Line 185: “harbour larger potential”: Consider rewording. This sounds strange.

Lines 230/231: not all environments are nutrient-limiting. Please include a reference to illustrate what you mean and tone down the statement.

Lines 231: It is unclear what is meant by “stabilizing negative feedback”. Please clarify.

Line 245: include “more” after “phylogenetically”

Line 260: remove “the” after “among” and “of”.

Line 275/276: It is unclear, how many cases are being discussed. Please mention the total number (12 of how many? 90% of how many?).

Line 278: From this data it cannot be inferred that the observed gene loss was adaptive. Please correct this here and throughout the text.

Line 283: remove “the”

Line 284: remove “by”.

Line 288: replace “fulfilled” with “compensated for”

Line 291: replace “indicating” with “likely suggesting”

Line 295: remove “the” before cooperative”

Line 296: reword “manifest” (e.g. by using “caused a”)

Line 303: Remove “the”

Line 314: Insert “to be” before “too”

Line 321: replace “wherein” with “in which”

Line 321/322: reword this sentence.

Lines 333/334: Please tone down this statement. Someone might have observed a trade-off between competition and cooperation in microbial communities before.

Line 357: please reword “bring forward”

Line 370: reword this line.

Line 374: please include the reference you are referring to.

Lines 374/375: 2x “malleable”. Please reword once.

Reviewer #3 (Remarks to the Author):

The authors have addressed the majority of the other reviewer’s concerns, and have added a significant number of new analysis, appropriate changes to the main text, and supplemental figures to accommodate reviewer suggestions. However, I am not completely convinced by the authors’ attempt to address the possible biases introduced by annotation quality and model reconstruction, which could be a critical confounding variable that is unaccounted for in this paper (see more details below). That being said, I am in favor of publication, contingent on the authors addressing this issue in two ways.

First, the authors should provide a supplementary table containing key data and metrics for each genome and model used in the study, including (but not limited to):

Taxonomic classification at various levels.
Annotation metrics, including the numbers of open reading frames, genes with unknown function, and annotated genes with known function
Model reconstruction metrics, including the number metabolic reactions, the number of gap-filled reactions and the number of exchange reactions.
The frequency with which each model occurred in cooperative and competitive consortia, respectively.

Providing these metrics and raw data will aid the research community in fully understanding the results presented in this paper.

Second, the authors should include a statement that more broadly describes the potential limitations of using metabolic models generated at this scale (as opposed to only using metabolic models for well characterized species). Specifically, the authors should
indicate that model reconstruction may be more accurate for some taxa and less for others, as highlighted in their previous paper (Machado et al. 2018), and these known biases are a factor that might influence the interpretation of their results.

In the authors rebuttal, they suggest that modeling biases are likely not a factor in this study because of the following reasons:

-Genome size correlates with the number of metabolic reactions

Some degree of (rank) correlation between genome size and metabolic reactions does not support the claim that gap-filling biases or annotation quality biases are not potentially responsible for the auxotrophic enrichment within cooperative consortia. In fact, recent work indicates that genome annotation quality is also correlated with genome size (Lobb et al. 2020), so this claim at the very least supports further investigation of this potential confounding variable.

-Biases in gene annotation would affect all species in a similar way, and thus would not explain why the metabolism of cooperative species is more dissimilar than expected by chance

This is not necessarily true, since annotations for classes of genes carrying out the same enzymatic function might be more challenging to annotate within a specific group of related taxa compared to taxa more closely related to well-characterized species.
Supp. Fig 8 confirms the suspicion that cooperative communities are heavily enriched with Firmicutes compared to competitive consortia, and the fact that taxa in cooperative consortia contain fewer transporters is also consistent with biases in model reconstruction. Additionally, if cooperative consortia are dominated by Firmicutes-Proteobacteria (inter-phyla) interactions, as the authors suggest, then both metabolic dissimilarity and rescue of auxotrophies (i.e. Proteobacteria spp. supplying amino acids to Firmicutes spp.) would be higher in cooperative consortia than expected by chance, since "random" consortia are likely enriched for intra-phyla interactions. Thus, this argument does not support the claim that annotation biases are not relevant in this study.

-AA production costs correlate with frequency of amino acid auxotrophies in cooperative communities

Finally, the authors argue that the correlation between amino acid production cost and frequency of predicted auxotrophies support the claim that annotation biases might not be a factor. However, this result could also be associated with annotation bias. For instance, correlation between amino acid production cost and frequency of predicted auxotrophies could be explained by the fact that amino acids with high production costs require more enzymes/genes for their production, and thus have a higher likelihood of being assigned auxotrophic merely by chance.

References:

Lobb, Briallen, Benjamin Jean-Marie Tremblay, Gabriel Moreno-Hagelsieb, and Andrew C. Doxey. 2020. "An Assessment of Genome Annotation Coverage across the Bacterial Tree of Life." Microbial Genomics 6 (3). https://doi.org/10.1099/mgen.0.000341.
Machado, Daniel, Sergej Andrejev, Melanie Tramontano, and Kiran Raosaheb Patil. 2018. "Fast Automated Reconstruction of Genome-Scale Metabolic Models for Microbial Species and Communities." Nucleic Acids Research 46 (15): 7542–53.

**********END**********

Author Rebuttal, first revision:

Response to reviewers' comments

We would like to thank all the reviewers for their constructive feedback. Please find below our response to the comments on the revised manuscript.

Reviewer #1 (Remarks to the Author):

The authors have addressed my concerns in a way that is acceptable to me. Therefore, I am in favour of publication of this paper.

→ We thank the reviewer for their positive feedback.

Reviewer #2 (Remarks to the Author):

The authors have done a great job revising the manuscript in response to all the referee's comments. I appreciate the efforts undertaken by the authors to carefully address and implement all comments and suggestions. The manuscript has greatly improved – particularly due to the incorporation of additional analyses (e.g. Supplemental Figures 11 and 12). The visual presentation of the data is very appealing and the text reads well. The presented results are of a high quality and the reported findings are very exciting. I have no doubts that this paper will attract a lot of attention, since it represents a new benchmark of yet unprecedented depth. Thus, I fully recommend publication of this manuscript in Nature Ecology & Evolution.

Below I have some minor comments that should be viewed as suggestions to further improve the manuscript.

(1) Statistics:

It is good scientific practice to always mention all details of a statistical test, so that the reader can understand and reproduce what has been done. This includes a mention of the type of test used, the P-value, sample size (n or df), as well as t, F, U, Z-values (depending on the statistics used). In the manuscript, the authors frequently just mention verbally that there is a difference (e.g. lines 71, 176, 290/291) without performing a test at all or report a test without mentioning the required statistical details (e.g. lines 158, 160, 166, 224, 225, 245, 246, 304, 309, 315, 316). Please correct this.

→ Thank you for bringing these to our notice. We now report test statistic (U-value, z-score), and p-value for all comparisons where statistical significance is maintained / implied. The sample sizes (n) are described in the respective figures and/or in the Methods section.

(2) Use of the red and black queen hypothesis: The authors use these two theories in a completely wrong context. The BQH is not about cooperation (it is about adaptive gene loss) and the RQH is not about competition (it is exclusively about host-parasite interactions). Even if in van Valen's original paper is "sufficient generic", subsequently the concept is exclusively used to denote interactions, in which organisms need to "continuously change in order to stay at the same place". This is not what is expected from competitive interactions at all. Thus, I would strongly suggest to completely remove a mention of at least the Red Queen. This is absolutely wrong. Also the way the Black Queen is referred to is not 100% correct. Please make sure you do not link it to cooperative interactions, but only to situations, in which gene loss can be adaptive. I specifically refer to the discussion section (line 343-354)(line 364).

→ Although, in our opinion, the Red and Black queen analogies hold sufficient flexibility, we agree that this may be misaligned with their conventional usage in the field. Therefore, we removed references to the Red Queen and Black Queen hypotheses from the Abstract. These are now referred to only in one paragraph in the Discussion, and in a toned-down manner:

"This dichotomy between competition and cooperation is in certain ways analogous to that between the red queen and the black queen hypotheses[51,52]. The former is reflected in competitive species as they tend to retain most biosynthetic capabilities and harbor genes useful for active antagonism; the latter is reflected in gene loss in cooperative species leading to dependencies on fellow community members. Our results provide evidence that both theories are operating in natura as two extremes in a metabolic trade-off between competition and cooperation."

(3) The term "cooperation" implies that a behaviour is costly and has evolved to benefit another individual. Something like this cannot be inferred from analysing metabolic networks. Thus, I would strongly recommend to remove the term and replace it with e.g. "synergism" throughout the manuscript or at least define carefully what the authors mean by it.

→ We have clarified the usage of cooperation as follows:

"While the MRO quantifies the similarity of the nutritional requirements between all species in a community, reflecting the intra-community risk for resource competition, the MIP indicates the number of metabolites that can be exchanged among the community members to decrease their dependency on the abiotic environment. In this work, we use the MIP value as a proxy measure for cooperative metabolism."

'Cooperative', in contrast to 'cooperation', implies a tendency or potential and not active operation. We further note that metabolite secretion is always costly to the secretor. This follows from the fundamental mass and energy balance considerations (metabolite formation costs material and energy), and also from the perspective of cellular resource allocation (such as enzymatic and transport capacity.; substrate uptake above that is needed for cell growth, and is diverted to secretion, requires additional investment in transport and enzymatic capacity). There is, of course, a possible (or even likely) situation that this cost is offset by the benefits received; but that does not make the secretion costless. We are aware of the works that deem metabolite secretion as 'costless'; these, however, do not appropriately account for the above-mentioned fundamental constraints, or perhaps use this term inadvertently.

Minor comments:

- Use of the word "trade-off landscape" (e.g. line 16): I think this an unfortunate term. Please consider rewording it.

→ Changed.

Line 10: insert hyphen between "habitat" and "dependent"

→ Changed.

Line 144: “member species” sounds odd. Consider revising.

→ Changed.

Lines 202, 205, 207, 362: reword “free-living environment”. There are no free-living environments.

→ Though we agree, this is the nomenclature used in the ontology of the EMP dataset, so for the sake of consistency we would prefer to keep it.

Line 185: “harbour larger potential”: Consider rewording. This sounds strange.

→ Changed.

Lines 230/231: not all environments are nutrient-limiting. Please include a reference to illustrate what you mean and tone down the statement.

→ We opted to remove the statement in question as it is not central to our point (thus additional explanation would cost us space), and modified the previous statement as follows:

“*Given that communities of co-occurring species exist as part of larger microbiomes, it is likely that competitive interactions with other species and/or resource limitations play a role in their stabilization.*”

Lines 231: It is unclear what is meant by “stabilizing negative feedback”. Please clarify.

→ As explanation on this would take more space and is not central to our point, we have now removed this sentence and modified the previous one as noted above.

Line 245: include “more” after “phylogenetically”

→ Changed.

Line 260: remove “the” after “among” and “of”.

→ Changed.

Line 275/276: It is unclear, how many cases are being discussed. Please mention the total number (12 of how many? 90% of how many?).

→ Rephrased to include the total number (117).

Line 278: From this data it cannot be inferred that the observed gene loss was adaptive. Please correct this here and throughout the text.

→ We agree that this data alone cannot show that gene loss was adaptive. When combined with the observations that these species have a higher and more stable overall abundance across environments (when compared to other species), it provides to be a clear indication of an increased fitness advantage. We have modified the sentence as: “*The latter, together with the higher abundance of cooperative species (Fig. 2d), suggests adaptive gene loss and…*”

Line 283: remove “the”

→ Changed.

Line 284: remove “by”.

→ Changed.

Line 288: replace “fulfilled” with “compensated for”

→ Changed.

Line 291: replace “indicating” with “likely suggesting”

→ Changed.

Line 295: remove “the” before cooperative”

→ Changed.

Line 296: reword “manifest” (e.g. by using “caused a”)

→ Changed.

Line 303: Remove “the”

→ Changed.

Line 314: Insert “to be” before “too”

→ Changed.

Line 321: replace “wherein” with “in which”

→ Changed.

Line 321/322: reword this sentence.

→ Changed.

Lines 333/334: Please tone down this statement. Someone might have observed a trade-off between competition and cooperation in microbial communities before.

→ We replaced “observed” with “reported”.

Line 357: please reword “bring forward”

→ Changed to “highlight”.

Line 370: reword this line.

→ Changed.

Line 374: please include the reference you are referring to.

→ Included.

Lines 374/375: 2x “malleable”. Please reword once.

→ Changed.

Reviewer #3 (Remarks to the Author):

The authors have addressed the majority of the other reviewer’s concerns, and have added a significant number of new analysis, appropriate changes to the main text, and supplemental figures to accommodate reviewer suggestions. However, I am not completely convinced by the authors’ attempt to address the possible biases introduced by annotation quality and model reconstruction, which could be a critical confounding variable that is unaccounted for in this paper (see more details below). That being said, I am in favor of publication, contingent on the authors addressing this issue in two ways.

First, the authors should provide a supplementary table containing key data and metrics for each genome and model used in the study, including (but not limited to):

Taxonomic classification at various levels.

Annotation metrics, including the numbers of open reading frames, genes with unknown function, and annotated genes with known function

Model reconstruction metrics, including the number metabolic reactions, the number of gap-filled reactions and the number of exchange reactions.

The frequency with which each model occurred in cooperative and competitive consortia, respectively.

Providing these metrics and raw data will aid the research community in fully understanding the results presented in this paper.

→ We agree and have accordingly added a new supplementary table (Supplementary Table 1) with the following fields: assembly accession, strain, species, genus, family, order, class, phylum, genome length, number of ORFs, annotated metabolic genes, enzymatic reactions, transport reactions, gap-filled reactions, internal metabolites, external metabolites, and occurrence frequency in cooperative and competitive communities.

Second, the authors should include a statement that more broadly describes the potential limitations of using metabolic models generated at this scale (as opposed to only using metabolic models for well characterized species). Specifically, the authors should indicate that model reconstruction may be more accurate for some taxa and less for others, as highlighted in their previous paper (Machado et al. 2018), and these known biases are a factor that might influence the interpretation of their results.

→ As suggested by the reviewer, we added the following statement alerting the reader to the potential limitations:

"However, our results are still limited and subject to biases due to the exclusion of species without a reference genome assembly, and due to variable quality of gene annotations - on which the models are based - across different species."

In the authors rebuttal, they suggest that modeling biases are likely not a factor in this study because of the following reasons:

-Genome size correlates with the number of metabolic reactions

Some degree of (rank) correlation between genome size and metabolic reactions does not support the claim that gap-filling biases or annotation quality biases are not potentially responsible for the auxotrophic enrichment within cooperative consortia. In fact, recent work indicates that genome

annotation quality is also correlated with genome size (Lobb et al. 2020), so this claim at the very least supports further investigation of this potential confounding variable.

→ The Lobb et al. study reports negative correlation between genome size and annotation quality; the situation with the metabolic genes, as we find in our set of species, is opposite. Further, there is no requirement *per se* to have a positive association between genome size and number of metabolic genes. What we meant is that the number of gap-filled reactions does not necessarily disproportionately increase for larger genomes.

-Biases in gene annotation would affect all species in a similar way, and thus would not explain why the metabolism of cooperative species is more dissimilar than expected by chance

This is not necessarily true, since annotations for classes of genes carrying out the same enzymatic function might be more challenging to annotate within a specific group of related taxa compared to taxa more closely related to well-characterized species.

Supp. Fig 8 confirms the suspicion that cooperative communities are heavily enriched with Firmicutes compared to competitive consortia, and the fact that taxa in cooperative consortia contain fewer transporters is also consistent with biases in model reconstruction. Additionally, if cooperative consortia are dominated by Firmicutes-Proteobacteria (inter-phyla) interactions, as the authors suggest, then both metabolic dissimilarity and rescue of auxotrophies (i.e. Proteobacteria spp. supplying amino acids to Firmicutes spp.) would be higher in cooperative consortia than expected by chance, since “random” consortia are likely enriched for intra-phyla interactions. Thus, this argument does not support the claim that annotation biases are not relevant in this study.

→ We agree that the auxotrophies *per se* could be biased, though complementarity of these auxotrophies is less likely to arise just by chance and is thus indicative of selection.

-AA production costs correlate with frequency of amino acid auxotrophies in cooperative communities

Finally, the authors argue that the correlation between amino acid production cost and frequency of predicted auxotrophies support the claim that annotation biases might not be a factor. However, this result could also be associated with annotation bias. For instance, correlation between amino acid

production cost and frequency of predicted auxotrophies could be explained by the fact that amino acids with high production costs require more enzymes/genes for their production, and thus have a higher likelihood of being assigned auxotrophic merely by chance.

→ Even if we consider that the longer pathways are associated with more false positives for auxotrophies, this would not explain the complementarity. Nevertheless, we agree that, as is case for all bioinformatics work based on the current state-of-the-art on annotations, we cannot claim that the results are not subject to biases. We believe that our disclaimer included in the revised manuscript (quoted above) and the new supplementary table make it transparent to the readers, and we hope that these will encourage further studies as new genomes and new functional genomics analyses become available.

References:

Lobb, Briallen, Benjamin Jean-Marie Tremblay, Gabriel Moreno-Hagelsieb, and Andrew C. Doxey. 2020. "An Assessment of Genome Annotation Coverage across the Bacterial Tree of Life." Microbial Genomics 6 (3). https://doi.org/10.1099/mgen.0.000341.

Machado, Daniel, Sergej Andrejev, Melanie Tramontano, and Kiran Raosaheb Patil. 2018. "Fast Automated Reconstruction of Genome-Scale Metabolic Models for Microbial Species and Communities." Nucleic Acids Research 46 (15): 7542–53.

Decision Letter, second revision:

25th September 2020

*Please ensure you delete the link to your author homepage in this e-mail if you wish to forward it to your co-authors.

Dear Dr Patil,

Your manuscript entitled "Polarization of microbial communities between competitive and cooperative metabolism" has now been seen by our reviewers, and in the light of their advice I am delighted to say that we can in principle offer to publish it. First, however, we would like you to revise your paper to ensure that it is as brief as possible and complies with our Guide to Authors at http://www.nature.com/natecolevol/info/final-submission.

TRANSPARENT PEER REVIEW
Nature Ecology & Evolution offers a transparent peer review option for new original research manuscripts submitted from 1st December 2019. We encourage increased transparency in peer review by publishing the reviewer comments, author rebuttal letters and editorial decision letters if the authors agree. Such peer review material is made available as a supplementary peer review file. Please state in the cover letter 'I wish to participate in transparent peer review' if you want to opt in, or 'I do not wish to participate in transparent peer review' if you don't. Failure to state your preference will result in delays in accepting your manuscript for publication.
Please note: we allow redactions to authors' rebuttal and reviewer comments in the interest of confidentiality. If you are concerned about the release of confidential data, please let us know specifically what information you would like to have removed. Please note that we cannot incorporate redactions for any other reasons. Reviewer names will be published in the peer review files if the reviewer signed the comments to authors, or if reviewers explicitly agree to release their name. For more information, please refer to our FAQ page.

SPECIFIC POINTS:
In particular, while checking through the manuscript and associated files, we noticed the following specific points which we will need you to address:

1. A brief editorial summary of the paper will appear on the journal homepage with the link to the paper. This is our proposed summary: 'Analysing data from thousands of microbial communities, the authors show that they cluster at different ends of the spectrum between resource competition and metabolic cooperation. Cooperative communities tend to have smaller genomes and multiple auxotrophies, whereas competitive communities have large genomes, overlapping niches and high antimicrobial activity". Please let us know of any factual inaccuracies.
2. Please note that we have recently moved from having figures in the supplementary information to having them as Extended Data items, which are linked directly from the main text in the html version of the paper. Please see below for further details of how to submit supporting files.
3. Please ensure error bars are defined in all relevant figure legends.
4. Please complete the Editorial policy checklist and the new version of the Reporting Summary (links below) and upload them with your revised manuscript. We will publish the latter along with the paper. Please note that these forms are dynamic 'smart pdfs' and must therefore be downloaded and completed in Adobe Reader. Please also ensure that "Final Submission" box is checked.
a. Editorial policy checklist: https://www.nature.com/authors/policies/Policy.pdf
b. Reporting summary: https://www.nature.com/authors/policies/ReportingSummary.pdf

GENERAL POINTS:
We will also need you to check through all of the following general points when preparing the final version of your manuscript:

The main manuscript file should include the abstract, main text, methods, author contribution, data availability, code availability and competing interests statements, acknowledgements, references, and figure legends. Figures should be submitted separately as individual files. For details on other supporting material, please see below.

Title & Abstract:
Titles should give an idea of the main finding of the paper and ideally not exceed 90 characters

(including spaces). We discourage the use of active verbs and do not allow punctuation.

The paper's abstract (about 150-200 words; no references) should serve both as a general introduction to the topic, and as a brief, non-technical summary of your main results and their implications. It should start by outlining the background to your work (why the topic is important) and the main question you have addressed (the specific problem that initiated your research), before going on to describe your new observations, main conclusions and their general implications. Because we hope that scientists across the wider ecology and evolution community will be interested in your work, the abstract should be as accessible as possible, explaining essential but specialised terms concisely. We suggest you show your abstract to colleagues in other fields to uncover any problematic concepts.

Figures:
Choosing the right electronic format for your figures at this stage will speed up the processing of your paper. We would like the figures to be supplied as vector files - EPS, PDF, AI or postscript (PS) file formats (not raster or bitmap files), preferably generated with vector-graphics software (Adobe Illustrator for example). Please try to ensure that all figures are non-flattened and fully editable. All images should be at least 300 dpi resolution (when figures are scaled to approximately the size that they are to be printed at) and in RGB colour format. Please do not submit Jpeg or flattened TIFF files. Please see our guidelines https://www.nature.com/documents/NRJs-guide-to-preparing-final-artwork.pdf for more details, and also our image policies http://www.nature.com/authors/editorial_policies/image.html.

We will edit your figures/tables electronically so they conform to Nature Ecology & Evolution style. If necessary, we will re-size figures to fit single or double column width. If your figures contain several parts, the parts should be labelled lower case a, b, and so on, and form a neat rectangle when assembled.

Figure legends must provide a brief description of the figure and the symbols used, within 350 words. This must include definitions of any error bars employed in the figures.

Should your Article contain any items (figures, tables, images, videos or text boxes) that are the same as (or are adaptations of) items that have previously been published elsewhere and/or are owned by a third party, please note that it is your responsibility to obtain the right to use such items and to give proper attribution to the copyright holder. This includes pictures taken by professional photographers and images downloaded from the internet. If you do not hold the copyright for any such item (in whole or part) that is included in your paper, please complete and return this Third Party Rights Table, and attach any grant of rights that you have collected.

Please check the PDF of the whole paper and figures (on our manuscript tracking system) VERY CAREFULLY when you submit the revised manuscript. This will be used as the 'reference copy' to make sure no details (such as Greek letters or symbols) have gone missing during file-transfer/conversion and re-drawing.

Supporting Information:
All Supporting Information must be submitted in accordance with the instructions in the attached Inventory of Supporting Information, and should fit into one of two categories:

1. EXTENDED DATA: Extended Data are an integral part of the paper and only data that directly contribute to the main message should be presented. These figures will be integrated into the full-text HTML version of your paper and will be appended to the online PDF. There is a limit of 10 Extended Data figures, and each must be referred to in the main text, cited as Extended Data 1, Extended Data 2, etc. Each Extended Data figure should be of the same quality as the main figures, and should be supplied at a size that will allow both the figure and legend to be presented on a single A4 page. Each figure should be submitted as an individual .jpg, .tif or .eps file with a maximum size of 10 MB each. All Extended Data figure legends must be provided in the attached Inventory of Supporting Information, not in the figure files themselves.

2. SUPPLEMENTARY INFORMATION: Supplementary Information is material that is essential background to the study but which is not practical to include in the printed version of the paper (for example, video files, large data sets and calculations). Each item must be detailed in the attached Inventory of Supplementary Information. Tables containing large data sets should be in Excel format, with the table number and title included within the body of the table. All textual information and any additional Supplementary Figures (which should be presented with the legends directly below each figure) should be provided as a single, combined PDF. Please note that we cannot accept resupplies of Supplementary Information after the paper has been formally accepted unless there has been a critical scientific error.

Additional Supplementary Figures and other items are not required to be referred to in your manuscript text (though they can be), but should be numbered as Supplementary Figure 1, not SI1, etc.

Methods & Notes:
Please include references for the Methods in the same list as those for the main text, following on sequentially after the main text references. Any citations in the Supplementary Information will need inclusion in a separate SI reference list.

Please include a data availability statement as a separate section after Methods but before references, under the heading "Data Availability". This section should inform readers about the availability of the data used to support the conclusions of your study. This information includes accession codes to public repositories (data banks for protein, DNA or RNA sequences, microarray, proteomics data etc…), references to source data published alongside the paper, unique identifiers such as URLs to data repository entries, or data set DOIs, and any other statement about data availability. All data that support the findings of the study must be made available. If DOIs are provided, we also strongly encourage including these in the Reference list (authors, title, publisher (repository name), identifier, year). For more guidance on how to write this section please see:
http://www.nature.com/authors/policies/data/data-availability-statements-data-citations.pdf

Nature Research policies (https://www.nature.com/authors/policies/availability.html#data) include a strong preference for research data to be archived in public repositories and in some cases this is mandatory. If you need help complying with this policy, or need help depositing and curating your research data (including raw and processed data, text, video, audio and images) you should consider:

Contacting Springer Nature's Research Data Helpdesk
(https://www.springernature.com/gp/authors/research-data-policy/helpdesk/12327114) for advice.

Finding a suitable data repository (https://www.springernature.com/gp/authors/research-data-policy/repositories/12327124) for your data.
Uploading your data to Springer Nature's Research Data Support service (https://springernaturedata.typeform.com/to/UeGGKT). Please note there are fees (https://www.springernature.com/gp/authors/research-data-policy/pricing/15499842) for using Springer Nature's Research Data Support service.

Finally, we require authors to include a statement of their individual contributions to the paper, such as experimental work, project planning, data analysis, etc., immediately after the acknowledgements. The statement should be short, and refer to authors by their initials. For details please see the Authorship section of our joint Editorial policies at http://www.nature.com/authors/editorial_policies/authorship.html

We will not send your revised paper for further review if, in the editors' judgement, the referees' comments on the present version have been addressed. If the revised paper is in Nature Ecology & Evolution format, in accessible style and of appropriate length, we shall accept it for publication immediately.

Please resubmit electronically

* the final version of the text (not including the figures) in either Word or Latex.

* publication-quality figures. For more details, please refer to our Figure Guidelines, which is available here: https://www.nature.com/documents/NRJs-guide-to-preparing-final-artwork.pdf .

* any Extended Data and Supplementary Information, as per instructed, with the associated Inventory document.

* copies of our reporting and editorial policy checklists even if they have not changed since the previous round of revision.

* a point-by-point response to any issues raised by our reviewers and to any editorial suggestions.

* any suggestions for cover illustrations, which should be provided at high resolution as electronic files. Please note that such pictures should be selected more for their aesthetic appeal than for their scientific content. I am sure you will understand that we cannot make any promise as to whether any of your suggestions might be selected for the cover of Nature Ecology & Evolution.

Please use the following link to access your home page:

[REDACTED]

*This url links to your confidential homepage and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this e-mail to co-authors, please delete this link to your homepage first.

Please also send the following forms as a hand-signed PDF by email to ecoevo@nature.com.

*Please sign and return the Licence to Publish form

Or, if the corresponding author is a Crown government employee (including Great Britain and Northern Ireland, Canada and Australia), please sign and return the Licence to Publish form for Crown government employees , or the Licence to Publish form for US government employees

For more information on our licence policy, please consult http://npg.nature.com/authors.

AUTHORSHIP

CONSORTIA -- For papers containing one or more consortia, all members of the consortium who contributed to the paper must be listed in the paper (i.e., print/online PDF). If necessary, individual authors can be listed in both the main author list and as a member of a consortium listed at the end of the paper. When submitting your revised manuscript via the online submission system, the consortium name should be entered as an author, together with the contact details of a nominated consortium representative. See https://www.nature.com/authors/policies/authorship.html for our authorship policy and https://www.nature.com/documents/nr-consortia-formatting.pdf for further consortia formatting guidelines, which should be adhered to prior to acceptance.

ORCID

Nature Ecology & Evolution is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as ‘corresponding author’ create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS) prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. For more information please visit http://www.springernature.com/orcid

For all corresponding authors listed on the manuscript, please follow the instructions in the link below to link your ORCID to your account on our MTS before submitting the final version of the manuscript. If you do not yet have an ORCID you will be able to create one in minutes. https://www.springernature.com/gp/researchers/orcid/orcid-for-nature-research

IMPORTANT: All authors identified as ‘corresponding author’ on the manuscript must follow these instructions. Non-corresponding authors do not have to link their ORCIDs but are encouraged to do so. Please note that it will not be possible to add/modify ORCIDs at proof. Thus, if they wish to have their ORCID added to the paper they must also follow the above procedure prior to acceptance.

To support ORCID's aims, we only allow a single ORCID identifier to be attached to one account. If you have any issues attaching an ORCID identifier to your MTS account, please contact the Platform Support Helpdesk.

We hope that you will support this initiative and supply the required information. Should you have any query or comments, please do not hesitate to contact me.

Nature Research journals encourage authors to share their step-by-step experimental protocols on a protocol sharing platform of their choice. Nature Research's Protocol Exchange is a free-to-use and open resource for protocols; protocols deposited in Protocol Exchange are citable and can be linked from the published article. More details can found at www.nature.com/protocolexchange/about.

We hope to hear from you within two weeks; please let us know if the revision process is likely to take longer.

[REDACTED]

****

Reviewer Comments:

Reviewer #3 (Remarks to the Author):

The authors have addressed the reviewers' suggestions, and inclusion of the new supplemental table and changes to the main text are very helpful for interpreting the authors' findings. The data provided in the new table suggests that gap-filling biases are not strongly associated with cooperative taxa, which greatly strengthens the paper. However, it is debatable that the "complementarity" (i.e. auxotrophic rescue) the author's observe could be indicative of selection. In this referee's opinion, this point should be discussed by the broader scientific community, and thus should not delay publication any further.

Overall, the authors have a done a great job with the revisions and I can recommend publication in Nature Ecology and Evolution.

**********END**********

Final Decision Letter:

21st October 2020

Dear Dr Patil,

We are pleased to inform you that your Article entitled "Polarization of microbial communities between competitive and cooperative metabolism", has now been accepted for publication in Nature Ecology & Evolution.

Before your manuscript is typeset, we will edit the text to ensure it is intelligible to our wide readership and conforms to house style. We look particularly carefully at the titles of all papers to ensure that they are relatively brief and understandable.

The subeditor may send you the edited text for your approval. Once your manuscript is typeset you will receive a link to your electronic proof via email within 20 working days, with a request to make any corrections within 48 hours. If you have queries at any point during the production process then please contact the production team at rjsproduction@springernature.com. Once your paper has been scheduled for online publication, the Nature press office will be in touch to confirm the details.

Acceptance of your manuscript is conditional on all authors' agreement with our publication policies (see www.nature.com/authors/policies/index.html). In particular your manuscript must not be published elsewhere and there must be no announcement of the work to any media outlet until the publication date (the day on which it is uploaded onto our web site).

The Author's Accepted Manuscript (the accepted version of the manuscript as submitted by the author) may only be posted 6 months after the paper is published, consistent with our self-archiving embargo. Please note that the Author's Accepted Manuscript may not be released under a Creative Commons license. For Nature Research Terms of Reuse of archived manuscripts please see: http://www.nature.com/authors/policies/license.html#terms
If you have posted a preprint on any preprint server, please ensure that the preprint details are updated with a publication reference, including the DOI and a URL to the published version of the article on the journal website.

An online order form for reprints of your paper is available at https://www.nature.com/reprints/author-reprints.html. All co-authors, authors' institutions and authors' funding agencies can order reprints using the form appropriate to their geographical region.

We welcome the submission of potential cover material (including a short caption of around 40 words) related to your manuscript; suggestions should be sent to Nature Ecology & Evolution as electronic files (the image should be 300 dpi at 210 x 297 mm in either TIFF or JPEG format). Please note that such pictures should be selected more for their aesthetic appeal than for their scientific content, and that colour images work better than black and white or grayscale images. Please do not try to design a cover with the Nature Ecology & Evolution logo etc., and please do not submit composites of images related to your work. I am sure you will understand that we cannot make any promise as to whether any of your suggestions might be selected for the cover of the journal.

You can now use a single sign-on for all your accounts, view the status of all your manuscript submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity for the Nature journals.

To assist our authors in disseminating their research to the broader community, our SharedIt initiative provides you with a unique shareable link that will allow anyone (with or without a subscription) to read the published article. Recipients of the link with a subscription will also be able to download and

print the PDF.

You can generate the link yourself when you receive your article DOI by entering it here: <a href="http://authors.springernature.com/share">http://authors.springernature.com/share<a>.

[REDACTED]

P.S. Click on the following link if you would like to recommend Nature Ecology & Evolution to your librarian http://www.nature.com/subscriptions/recommend.html#forms

** Visit the Springer Nature Editorial and Publishing website at www.springernature.com/editorial-and-publishing-jobs for more information about our career opportunities. If you have any questions please click here.**