Peer Review File

Manuscript Title: Multi-kingdom ecological drivers of microbiota assembly in preterm infants

Editorial Notes:

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referee #1 (Remarks to the Author):

Summary. Understanding how the human gut microbiome is established is important to address infant health. In this paper, the authors explore the role of inter- and intra-kingdom microbial interactions on the colonization and development of the pre-term infant gut microbiome. The authors inferred intra-/inter-species interactions by fitting a generalized Lotka-Volterra model to absolute abundance data, as measured by MK-SpikeSeq. Then, they validated a number of these interactions experimentally, either in vitro or in mouse models.

This work demonstrates:
• a reasonable method for absolute abundance quantification of individual taxa within microbial community samples;
• that ecological interactions within complex natural communities can be inferred from data;
• that some of the key inferred interactions can be validated experimentally (although difficult to validate overall accuracy);
• that these interactions are consistent with the successional patterns characteristic of the developing infant gut microbiome.
Overall, I was quite impressed with this paper. The questions are important, the data are solid (but see below), and the validation with laboratory experiments encouraging. I believe that the paper will be of broad interest and should be published in Nature if the authors can address the following questions.

Major comments

1. I would like to see more visualizations of the different trajectories for different patients. Only representation in Fig. 2C for a “typical” patient, where it looks to me that at the end of the trace that there are equal amounts of Escherichia, Enterococcus, and Kleb, rather than just one dominating (as indicated in the text, line 112). Which “path” did this typical patient take?

2. Having inferred interactions from the data doesn’t meant that the patterns in the data are driven by the interactions. This statement is true even if the interactions are validated in the lab with isolates. I say this despite the fact that I think that it is likely (hopefully?) true, and that I was impressed by the validation. I don’t think that any new measurement is necessary, but the authors may need to slightly weaken the text (lines 161, 207, 236 for example).

3. Do the authors think that the inferred interaction network sufficient to recapitulate the observed final three states? I am ok if the answer is no, but it would be nice to know the authors’ thoughts.

4. I felt the need to learn more about the similarities and differences between pre-term and full-term microbiome assembly. Also, in this paper, the authors studied early-life microbiome assembly using mostly samples from infants who were born via Caesarean-section (131/178). It’s known that the mode of baby delivery is a significant factor that affects the composition of the gut microbiota throughout the neonatal period and into infancy. Also, C-section babies are missing key microbes. (https://doi.org/10.1038/s41586-019-1560-1). On a related note, lines 42-53 may not accurately

capture what the field has already learned about the role of the different factors that were mentioned.

Minor comments

1. In the MK-SpikeSeq method the authors add cells instead of DNA as the standard for absolute quantification. It seems to me that depending upon the nature of the error in DNA extraction (ie whether variation in DNA extraction efficiency is correlated among the different species from sample to sample) then the authors' approach could be either better or worse than adding bare DNA. Is this true? I would have appreciated a bit more of a justification for why the authors think that it is better to add cells.
2. The data used to validate the MK-SpikeSeq method are plotted in a way that makes it difficult to evaluate the method.
a. Fig. S2A and S2B don't have axis labels, the axes aren't the same length, and the distance between tick labels is not the same.
b. In Fig. S2A, S2B, and S2C, the authors should report the slope and offset of the best-fit line, not just the r. Otherwise, it's not clear that there is a 1:1 mapping between observed and expected data.
c. Could be nice to quantify the accuracy of absolute abundance measurements for individual taxa, since that's the object of interest for all subsequent analyses.

3. The conclusion from the linear mixed model (lines 127-131) seems very strong, given the analysis presented.
a. Residuals? Is this model a good fit to the data?
b. How was the confidence interval calculated?
c. Plot of bacterial load vs. fungal load?
d. Effect size seems small, but also difficult to evaluate, given normalization.
e. Negative correlation is only barely statistically significant

4. Fig. S6: For the fungal ITS, it seems like the data were fit on a linear scale, thereby biasing the best-fit line towards high-abundance samples. Probably would be best to log-transform the data before fitting.

5. I found Fig. 1A more confusing than illuminating. Despite already being familiar with the effect, it took me some time to figure out what all of the bars were supposed to be telling me.

6. More generally, are any of the primary conclusions dependent upon whether the data is analyzed on a linear or log-data?

7. Fig. S7: Statistical test? P-value?

8. I would have appreciated a sentence somewhere discussing similarities and differences between the pattern of colonization and succession between pre-term and full-term infants.

9. Line 165: Given the regularization that was done, we should not be surprised that many interactions contain a 0 (commensalism or amensalism).

10. Line 230: "occur" should be deleted

11. Line 235: "context-independent" should read "context-dependent"

12. Line 717: The finding of Cryptococcus spp. as a dominant microbial genus in the studied mycobiome contradicts the consensus of the mycobiome community regarding the absence of cryptococcus from the normal human gut microbiome. Could the authors elaborate? (There was a reference to a table that I could not access)

13. It's surprising that Saccharomyces spp. wasn't found in the gut mycobiome since it's the most abundant yeast in the human gut microbiome. (see: https://doi.org/10.1186/s40168-017-0373-4)

14. Lines 160-161: The species should perhaps always be mentioned. “Candida inhibited both Klebsiella and Escherichia, but was itself inhibited by Staphylococcus.”

15. There are data showing that E. coli produces a fungicidal factor against C. albicans. (doi: 10.15698/mic2018.05.631.)

16. There is some evidence that the interactions between C. albicans and S. epidermidis is synergistic (https://doi.org/10.1186/1471-2180-13-257 and https://doi.org/10.1099/0022-1317-51-4-344)

17. Line 544: In vivo co-colonization validation. A reference for the antibiotic pre-treatment and the ampicillin drinking water should be added. Was it necessary to add antibiotic treatment during the co-colonization? It may change bacterial growth in monoculture colonization/co-colonization. Also, bacteria may respond differently as a response to the antibiotic. More generally, it seems that it would be good if the animal experiments mimicked the real conditions as much as possible.

18. I couldn’t find important information that would be needed to reproduce the experiment, such as an anesthesia protocol and inoculum concentrations. It’s not clear when the authors say: “lower inoculation was used for K. pneumoniae than others as this bacterium more readily and rapidly established colonization in the mouse gut”. What were the actual doses used? What is the background supporting this statement?

19. Line 805: Figure S2. What is the label “others” in A?

20. It would have been nice to see Cryptococcus in a mock community, but I will not insist on this.

Referee #2 (Remarks to the Author):

The authors develop and use to great effect a technique of converting relative abundances into absolute abundances, use this to assess the temporal changes in microbial gut populations, and infer ecological relationships between different microbes in the developing infant gut. The authors support their inferred relationships by isolating representatives of their genera of interest, then showing how in vitro coculturing replicates most of the interaction dynamics suggested by their model. Interestingly, the in vitro dynamics that didn’t follow the model predictions could be injected into a SPF mouse model system to mimic the model predictions, implying that these interactions were dependent on the gut environment. The key findings of this paper – that specific pairwise microbial interactions drive the pattern of succession in the infant gut, only some of which are context-dependent on the gut environment, are valuable discoveries for understanding the assembly of this microbiome.

Further, the authors show that inference from relative abundance entirely misrepresented the dynamics that could be concluded from absolute abundances; that the dominant species of later succession (which microbe dominates after the pioneer Staphylococcus) seems to be related to the outcome of multiple pairwise interactions between dominant microbes, and that fungal and bacterial populations are negatively correlated, perhaps due to inhibitory effects of yeast species like Candida albicans.

The combination of approaches used in this work is rarely accomplished in microbiome work, especially the experimental validation of microbial interactions. In addition, the contribution of fungal and archaeal species in the human microbiome is an underexplored area, so it is nice to see that the authors applied their method broadly. This paper represents a big step in a much needed direction for the microbiome field.

Major comments:

I found the description of MK-SpikeSeq to be lacking in the text and confusingly depicted in Fig 1. It was difficult to piece together all of the information from the figure, methods and supp info to get a clear understanding of both how it works and the various validations used with mock communities. It might

be more effective to move it all to the supplementary material and combine the associated data together.

Figure 1C shows how the absolute abundance of spike-in microbes can be back-calculated from their known dilution factor. In lines 87-88, the authors claim that this approach is “robust to variable DNA extraction efficiency,” but does not address bias in extraction efficiency between, e.g., gram positive bacteria versus their gram-negative spike-in standard or yeast species compared to filamentous fungi. Further, in Supp figure S2, some of the filamentous fungal species appear to be present at very low levels (if at all). Is this due to differences in DNA extraction efficiency, lower input, PCR bias, or something else. My concern is that there is systematic over- or under-estimation of certain kinds of microbes biased by the selection of spike-in species. It would be nice to see whether or not the species selection has a significant effect on the calculation of absolute abundance of other microbes.

In Figure 3D and F, why are the input levels for certain species so different between alone and interacting conditions? How do we know that this doesn’t affect the growth outcome?

Minor comments:
• Within the abstract (lines 18-19), the authors claim that the “predictable assembly [of the infant gut] is vital to health,” yet the cited papers do not support this exact claim, rather just that gut composition is important to health status.
• Authors could clarify that the antibiotic they are referring to in line 146 are not microbiome-produced, but rather clinical antibiotics.
• Figure 1a might be streamlined by demonstrating how the same relative abundance scenario could be represented by two different absolute abundance scenarios, instead of convoluting the modeled scenario with different cumulative populations trajectories.
• I would like to see multiple infant successions, even if in supplementary, to evaluate how representative the one shown in Figure 2C is.
• Figure S2a and S2b: the scatter plots are missing axis text – important to know how much expected versus observed abundances only correlate, as mentioned in the text, versus how close they match.

Referee #3 (Remarks to the Author):

This manuscript by Rao et al introduces a new method for absolute quantitation of bacteria, archaea and fungi present in the intestinal microbiota. They apply this method for analysis of longitudinal samples from a cohort of 178 preterm infants, and use the results to model and predict intra- and inter-kingdom interactions, which are finally validiated by in vitro experiments as well as in animals.

The methodology is original and highly relevant, since the many limitations of focusing only on relative bacterial abundances within the microbiome are becoming increasingly clear to the field.
The results obtained are novel and important for the understanding of the establishment of the preterm microbiome, as well as for the understanding of microbial ecology in general.

I have the following comments and suggestions:

1) The microbiome of preterm infants is significantly different from that of infants born within term, the latter being characterized by a high abundance of bifidobacteria and a massive influence of breastfeeding. I therefore think it is important to make it clear already in the title, that the present study deals with preterm infants. Perhaps also emphasize it a bit more in the abstract since the facultative organisms selected by the aerobic preterm gut environment differ drastically from those of the normal term infant microbiome.

2) The researchers spike the samples with bacterial, fungal and archaeal DNA, and can quantify the absolute abundances within these kingdoms. However, the manuscript and figures are focused only on bacteria and fungi, with no mentioning of the results obtained for archaea. These would be interesting to know. Even if they are not conclusive, it would be good to include them.

3) l. 118 - The blooming of Klebsiella is not really masked in the relative data.

4) l. 125 - Similarly, I think that '..uncovered dramatic blooms and collapses in fungal genera masked by relative abundances..' is a bit overstated, since these blooms and collapses are largely also seen in the relative abundance data.

5) l. 165-166 - I think that this is remarkable and surprising. It there a risk that the applied model might be biased against the detection of symmetric (+/+ and -/-) interactions?

6) l. 186-190: It would be relevant to know which distinct species within each genus were captured. Also, it is unclear whether each of these species were tested seperately against each of the species from another genus? Or were they tested as a mix of all species from one genus? Are only representative/affirmative data shown in Fig 3c? Were there no species-species interactions which did not fit the model prediction?

7) l. 193-194: Note that the environment (gut versus lab medium) may have a huge impact on e.g. competition for nutrients.

8) l. 196: You state here that no positive effect of Staphylococcus on Klebsiella was observed under in vitro conditions, but still this is marked with a big blue arrow in Fig 3c?

9) 228: I think a new sub-headline would be feasible here.

Comments to figures:
Fig 1b - I don't understand the purpose of having two fecal samples/ two soil samples with an arrow between them in the left side of the figure?

Fig 3a - What is the green round dot in this figure representing?

Fig 3c - see comment 8). Also, do the dots represent replicate experiments or different species? (If both, it would be good to label the difference as for Candida).

Fig S6 - Something went wrong with the panel letter references in the legend to this figure.

Author Rebuttals to Initial Comments:

We thank all three referees for their enthusiasm for our work and for their helpful comments, which have allowed us to substantially improve our manuscript. Please see below for our point-by-point responses.

Referees' comments:

Referee #1 (Remarks to the Author):

Summary. Understanding how the human gut microbiome is established is important to address infant health. In this paper, the authors explore the role of inter- and intra-kingdom microbial interactions on the colonization and development of the pre-term infant gut microbiome. The authors inferred intra-/inter-species interactions by fitting a generalized Lotka-Volterra model to absolute abundance data, as measured by MK-SpikeSeq. Then, they validated a number of these interactions experimentally, either in vitro or in mouse models.

This work demonstrates:
• a reasonable method for absolute abundance quantification of individual taxa within microbial community samples;
• that ecological interactions within complex natural communities can be inferred from data;
• that some of the key inferred interactions can be validated experimentally (although difficult to validate overall accuracy);
• that these interactions are consistent with the successional patterns characteristic of the developing infant gut microbiome.
Overall, I was quite impressed with this paper. The questions are important, the data are solid (but see below), and

the validation with laboratory experiments encouraging. I believe that the paper will be of broad interest and should be published in Nature if the authors can address the following questions.

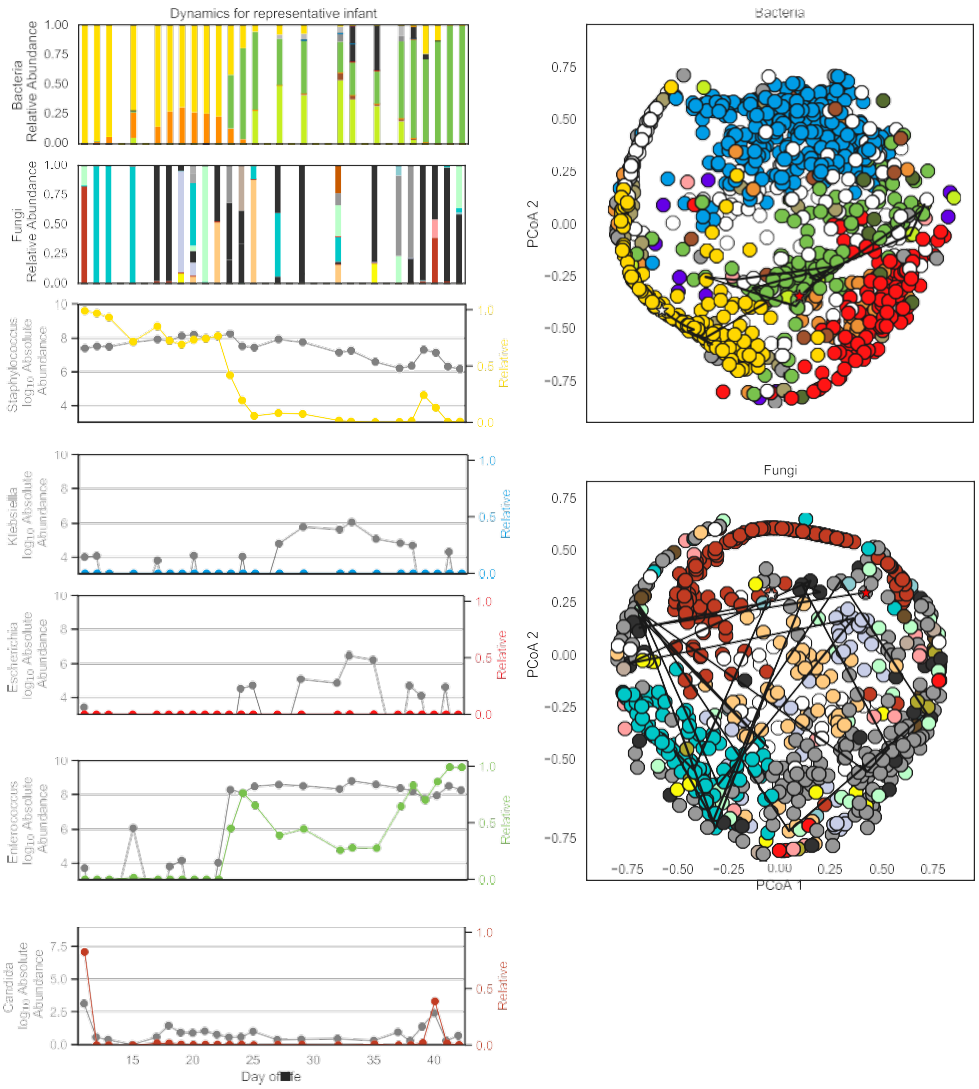
Major comments

1. I would like to see more visualizations of the different trajectories for different patients. Only representation in Fig. 2C for a "typical" patient, where it looks to me that at the end of the trace that there are equal amounts of Escherichia, Enterococcus, and Kleb, rather than just one dominating (as indicated in the text, line 112). Which "path" did this typical patient take?

Thank you for your suggestion, we've now included supplementary figures illustrating the changes in bacterial and fungal community composition and abundance for all of the 178 infants in our cohort (Extended Data Figs 8,9,24,25). For the 13 infants for whom we sequenced daily samples we've also included a new visualization that illustrates how each infant's bacterial and fungal communities change over time, relative to all other infants in our cohort (Extended Data Figs 10-22, see below for an example). As the reviewer can hopefully see, together these additional figures highlight how conserved the trajectories are for the infants' bacterial communities, and how unpredictable their fungal dynamics are.

We suspect the apparent equal distributions are due to the log scale on which we've plotted the total abundances, which obscures that *Klebsiella* is roughly an order of magnitude higher in abundance than the other taxa (we admit, this wasn't very clear in our original figure). We've now added gridlines to Fig 2c and the equivalent supplementary figures, which we hope makes these differences more apparent. More fundamentally though, the reviewer has hit on an important point - that even though *Klebsiella* dominates in terms of relative abundances, the other major taxa are still present in appreciable levels - a phenomenon shared by several of the other infants. This again highlights the importance of quantifying absolute abundances when studying microbiome communities, and we've added additional text highlighting this point in the manuscript:

"Such comparisons also indicated that though the bacterial communities within our cohort are typically dominated by just one genus, in many cases the other major genera remain at high levels within the preterm infant gut (Extended Data Figs 10-22)."



Extended Data Figure 15. Bacterial and fungal dynamics for Infant 6. left, *Time series illustrating the relative abundances of bacteria and fungi within Infant 6, demonstrating overall microbiota dynamics and comparing relative and absolute dynamics for major taxa.* ***right,*** *Black line indicates the trajectories of the bacterial (top) and fungal (bottom) communities within Infant 6 in relation to all other infant samples. White star indicates the first sample collected and red star indicates the final sample taken.*

2. Having inferred interactions from the data doesn't meant that the patterns in the data are driven by the interactions. This statement is true even if the interactions are validated in the lab with isolates. I say this despite the fact that I think that it is likely (hopefully?) true, and that I was impressed by the validation. I don't think that any new measurement is necessary, but the authors may need to slightly weaken the text (lines 161, 207, 236 for example).

We agree with the reviewer, though our inference and validation suggest the interactions do influence community dynamics, we were probably too strong in our wording. As such, we've softened our text accordingly in the following sections:

Line 161, now line 173:

"These results suggested that not only do preterm infants harbor diverse fungal communities, but that members of these communities ***may play a role in*** ~~*are critically*~~ *influencing larger scale community dynamics."*

Line 207, now line 223:

*"These in vivo data recapitulated the dynamics observed in infant assembly (Fig 2c) and ~~demonstrated~~ **suggested** that the predictable patterns in infant microbiome assembly ~~are~~ **may indeed be** due to exploitation of an early pioneer species by a late colonizer."*

Line 236, now line 256:
*"Our findings ~~demonstrate~~ **suggest** a common mechanism of assembly between the infant microbiota and macroscopic ecological succession. Just as in macroscopic ecosystems[15–17], microbes **may** exploit one another to establish within the infant gut, and direct interactions between kingdoms **appear to** play a central role in community dynamics"*

Line 244:

*"Together, our data demonstrated a novel species-specific cross-kingdom interaction that **appears to shapes** ~~the~~ preterm infant gut microbiota."*

Abstract:

*"We infer computationally and demonstrate experimentally in vitro and in vivo that predictable assembly dynamics ~~are~~ **may be** driven by directed, context-dependent interactions between specific microbes"*

*"Remarkably, we find that interactions between kingdoms ~~drive~~ **can influence** assembly"*

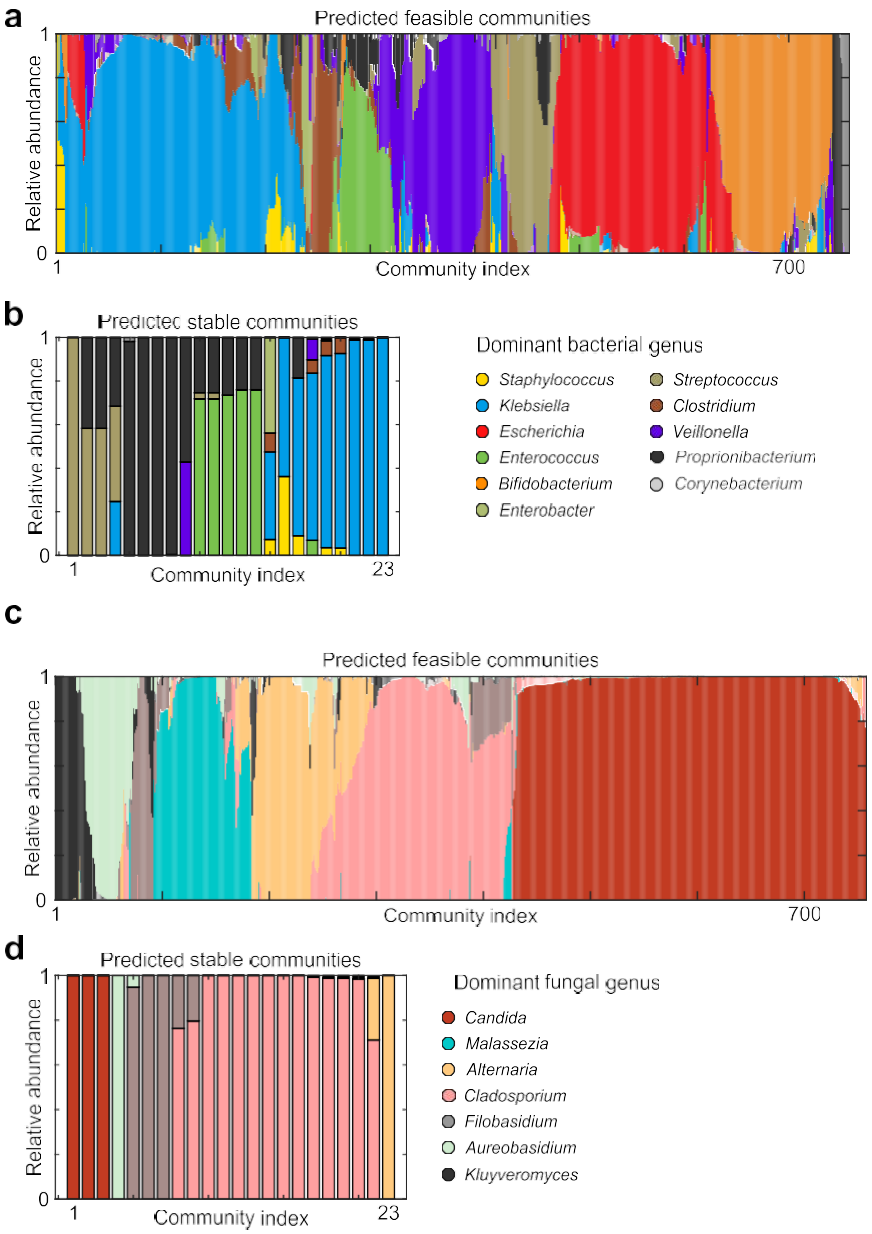
3. Do the authors think that the inferred interaction network sufficient to recapitulate the observed final three states? I am ok if the answer is no, but it would be nice to know the authors' thoughts.

This is a really interesting question from the reviewer. To try and answer this we took our 18 prevalent genera and calculated all of the possible subcommunities composed of one or more genera (yielding 262143 possible subcommunities). Using our inferred interaction matrix we then calculated which of these subcommunities had feasible steady states (states where all constituent species had non-zero abundances) – yielding 763 possible subcommunities. Finally, we examined the community composition of these possible steady states and checked which of these subcommunities were linearly asymptotically stable - yielding 26 possible communities, of which 3 had no bacteria present so were discounted.

Interestingly, we found that the majority of our feasible steady states were dominated by either *Klebsiella* or *Escherichia*, with the remainder dominated by either *Enterococcus, Bifidobacterium* or *Veillonella*. Perhaps more remarkably still, the majority of the few stable communities observed were characterized by either high *Klebsiella* abundance (9/23) or high *Enterococcus* abundance (5/23). To a certain extent, therefore, our model does seem to recapitulate the observed final states. That is, our model suggests that domination by *Klebsiella* or *Enterococcus* are common stable community states, and thus we might expect infants' microbiomes to converge towards one of these two states over time. Further, the model suggests that while domination by *Escherichia* is feasible, this state is not stable and thus likely to be transient - something that indeed we see relatively frequently in our cohort, where infants that are dominated by *Escherichia* often transition to domination by *Klebsiella* or *Enterococcus*. Finally, the model suggests *Candida* domination is the most common feasible fungal community state, but that such communities are not often stable – perhaps unsurprising given the lack of stable fungal colonization within our cohort.

However, while interesting, we are wary about (over)interpreting these results. Though feasible and stable community states are tractable methods of assessing likely model outcomes, in reality we imagine infant guts are unlikely to be at steady state (as we know from longer term studies of preterm infants, their microbial communities will change further). Moreover, our final state predictions are performed agnostic of antibiotic perturbations, as doing so would require additional assumptions about dosing and timing of antibiotics, but in reality, these will likely also be shaping the final community states.

In the light of this, we've now included our model predictions (see below) and a conservative discussion of our findings in the supplement (Supplemental Text Section 5) as we think they'll be of interest to readers.



Extended Data Figure 29. Feasible and stable bacterial and fungal communities, as predicted by inferred interaction network. ***a,*** *Steady-state bacterial relative abundances of those subcommunities predicted to have a feasible equilibrium, wherein all constituent genera have non-zero abundances.* ***b,*** *Steady-state bacterial relative abundances of those feasible subcommunities predicted to be linearly asymptotically stable.* ***c,*** *Steady-state fungal relative abundances of those subcommunities predicted to have a feasible equilibrium.* ***d,*** *Steady-state fungal relative abundances of those feasible subcommunities predicted to be linearly asymptotically stable.*

4. I felt the need to learn more about the similarities and differences between pre-term and full-term microbiome assembly. Also, in this paper, the authors studied early-life microbiome assembly using mostly samples from infants who were born via Caesarean-section (131/178). It's known that the mode of baby delivery is a significant factor that affects the composition of the gut microbiota throughout the neonatal period and into infancy. Also, C-section babies are missing key microbes. (https://doi.org/10.1038/s41586-019-1560-1). On a related note, lines 42-53 may not accurately capture what the field has already learned about the role of the different factors that were mentioned.

Thank you for this comment, indeed all three reviewers suggested we include further details on differences in microbiota development between delivery modes and gestational age. We have therefore edited our text accordingly (including citing the paper referenced above and others supporting the impact of different factors on infant gut microbiome composition):

"Yet despite the apparent importance of the infant microbiota, we do not understand what drives initial microbiota development. ***Premenstrual age (term vs. preterm), mode of delivery (Caesarian section vs. vaginal), host epithelial and immune ontogeny, diet (formula vs. mother's milk), antibiotics, and the interactions between individual microbes have been demonstrated[2,18–22,28–30] or have the potential to influence microbiota composition****. But with so many moving parts, the specific role of any individual factor in mediating microbiota development and assembly has remained unclear."*

"Specifically, we assembled a prospective cohort of 178 preterm infants from a tertiary-care neonatal intensive care unit (NICU). ***The assembly of the preterm microbiota differs substantially from that of term infants. Most preterm infants are born via C-section and as such are seeded with skin and hospital-associated microbes and devoid of key maternal derived bacteria[7,21,29]. In addition, the preterm microbiota displays "delayed" maturity with a prolonged membership of facultative anaerobic bacteria compared to that of a predominantly strict anaerobic community of term infants[7,21,29]****. We focused on preterm infants due to their clinical relevance and because they are amenable to high-frequency longitudinal sampling with readily available clinical and dietary metadata."*

"Consistent with previous preterm gut studies, we observed that preterm infant gut bacterial communities cluster primarily into four distinct community states, characterized by the domination of one of four genera: Staphylococcus, Klebsiella, Escherichia or Enterococcus (Fig 2a). ***In contrast to the microbiome of full-term infants, these community clusters were independent of diet or delivery mode (Extended Data Fig 7)****."*

Minor comments

1. In the MK-SpikeSeq method the authors add cells instead of DNA as the standard for absolute quantification. It seems to me that depending upon the nature of the error in DNA extraction (ie whether variation in DNA extraction efficiency is correlated among the different species from sample to sample) then the authors' approach could be either better or worse than adding bare DNA. Is this true? I would have appreciated a bit more of a justification for why the authors think that it is better to add cells.

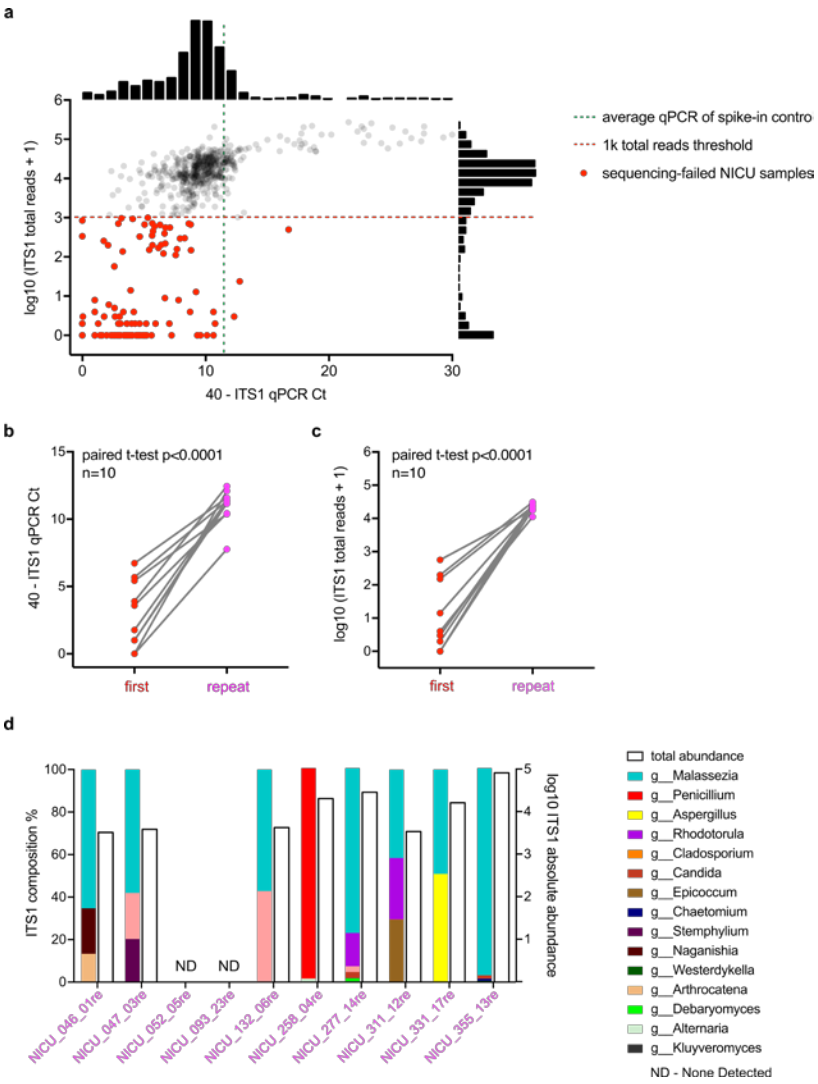
The reviewer raises an important point that we did not make clear enough in our manuscript. We chose to add spiked cells specifically because it helps control against errors in DNA extraction. That is, by spiking in cells rather than bare DNA we are able to identify cases where DNA extraction has failed, which would be indicated by seeing no or little DNA from the spike-in cells. This means, unlike when one spikes in bare DNA, we can distinguish between cases where extraction has failed and cases where there is simply very little microbial community present in the sample. This also means we can normalize between samples with variable DNA extraction, such as between samples from different environments, or different experimental conditions.

As demonstrated in the figure below, we found the ability to distinguish between these two cases was particularly important for the fungal communities within our cohort. Here we observed several examples that our rDNA sequencing criteria identified as having failed during processing (<1k total reads), likely due to poor DNA extractions as indicated by low qPCR signals (red dots in Extended Data Fig. 6a; these failures were largely unrelated to limitation of sequencing depth, as a separate re-sequencing only rescued six of these samples).

Using a bare DNA spike-in would overcome the poor extraction efficiency of the spike-in signal (most reads would come from bare DNA) and suggest the total fungal load in these samples was at or near zero. However, when we selected 10 of these samples and reprocessed them to achieve better DNA extraction (Extended Data Fig. 6b, c), we found that 8 had total fungal loads in line with the average fungal loads in successful samples (log10 (ITS1) > 3), with only 2 samples showing no detectable fungal signal (Extended Data Fig. 6d).

Our motivation for the use of a cell-based spike-in and its robustness to extraction efficiency variability are all vital points that we did not make clear enough in our original manuscript. As such, we have now included an additional supplementary section, "***Section 2 – MK-SpikeSeq validation".*** Here we discuss in more detail how MK-SpikeSeq compares to existing methods, demonstrate its ability to detect failures in extraction, and illustrate how MK-

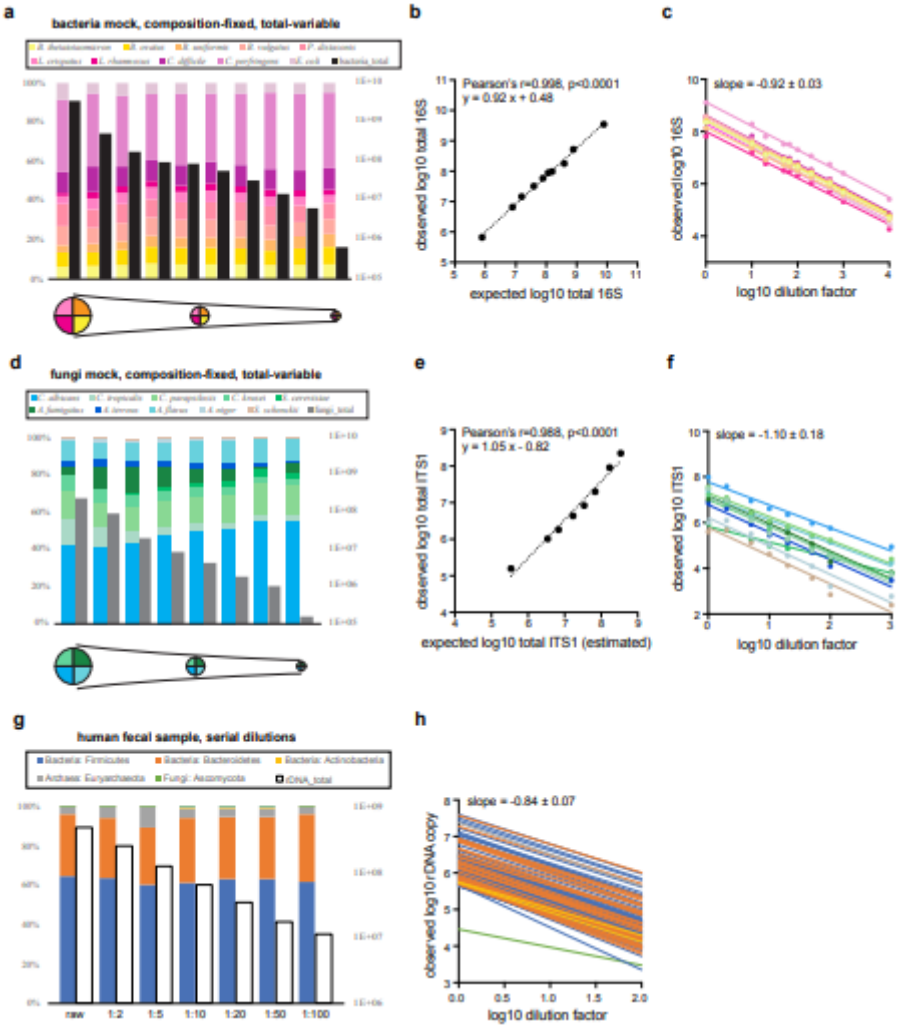
SpikeSeq can accurately detect changes in absolute abundances of individual taxa. We have also added several supplementary figures (Extended Data Figs. 2-6) to support this discussion.



Extended Data Figure 6. MK-SpikeSeq identifies errors in sample processing of fungal communities. a, *In our first phase of NICU sequencing (see Supplemental Text), we identified a number of samples, highlighted in red dots, that failed to yield >1k ITS1 reads per sample post quality filtering (red dashed line). Many of these sequencing-failed samples showed much lower (>5 deltaCt) ITS1 qPCR signals than the spike-in control (green dashed line), indicating poor DNA extractions of fungi in these samples. Shown next to the axes are frequency histograms of measurements.* ***b,*** *Reprocessing of 10 of these sequencing-failed samples led to increased ITS1 qPCR signals, indicating better DNA extractions.* ***c,*** *These reprocessed samples also yielded normal >10k ITS1 reads, passing our rDNA sequencing criteria.* ***d,*** *Eight of the reprocessed samples showed non-zero fungal communities, and only two had no detectable fungal signal. Shown are the composition (colored bars) and total abundance (empty bars) of fungal communities in these reprocessed samples.*

2. The data used to validate the MK-SpikeSeq method are plotted in a way that makes it difficult to evaluate the method.
a. Fig. S2A and S2B don't have axis labels, the axes aren't the same length, and the distance between tick labels is not the same.
b. In Fig. S2A, S2B, and S2C, the authors should report the slope and offset of the best-fit line, not just the r. Otherwise, it's not clear that there is a 1:1 mapping between observed and expected data.
c. Could be nice to quantify the accuracy of absolute abundance measurements for individual taxa, since that's the object of interest for all subsequent analyses.

Thank you for catching this, we've now adjusted the axis in Extended Data Fig 2 accordingly and reported the slope and offset as suggested. We've also now added additional supplementary figures illustrating that MK-SpikeSeq accurately captures changes in absolute abundance for a range of individual taxa within both mock communities and human fecal samples:

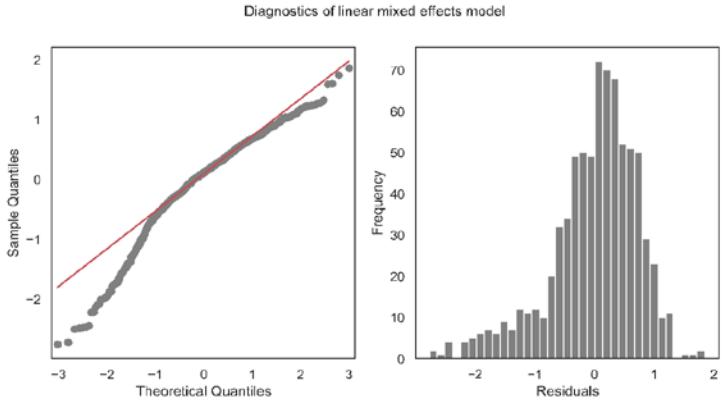


Extended Data Figure 2. MK-SpikeSeq reliably measures absolute abundances across kingdoms. *A set of single-kingdom mock communities with a fixed composition of 10 bacterial (**a**) or 10 fungal (**d**) species and variable total microbial loads (indicated by the pie chart schematics underneath), were quantified using MK-SpikeSeq for relative composition (colored bars) and absolute abundance (black/grey bars). **b**, **e**, Correlations between expected (based on initial microbial densities and known dilution factors) and MK-SpikeSeq-measured total absolute abundances show that MK-SpikeSeq reliably detects absolute abundances of bacteria and fungi. Note that for **e**, as exact rDNA copy numbers per fungal cell are undefined, the expected total ITS1 abundances are only estimates (here using 200 rDNA copies per fungal cell). **c**, **f**, Absolute abundance changes for individual members (color coded same as **a**, **d**) in the bacterial and fungal mock communities are largely consistent with known dilution factors. **g**, A set of serial dilutions of a human fecal sample was quantified using MK-SpikeSeq for relative composition (colored bars, shown are the phylum level taxa) and absolute abundance (empty bars). **h**, Absolute abundance changes for individual OTUs (color coded in phyla same as **g**) across kingdoms are largely consistent with known dilution factors.*

3. The conclusion from the linear mixed model (lines 127-131) seems very strong, given the analysis presented.
a. Residuals? Is this model a good fit to the data?
b. How was the confidence interval calculated?
c. Plot of bacterial load vs. fungal load?
d. Effect size seems small, but also difficult to evaluate, given normalization.
e. Negative correlation is only barely statistically significant

We agree with the reviewer that in hindsight our original discussion of the linear mixed model was too strong given the weakness of the effect, and the skew in our residuals (see below, now included in the extended data). We have therefore toned down our discussion of the inverse correlation observed, both in the main text and the abstract, and provided details on how confidence intervals were calculated:

*"However, though the fungal dynamics were themselves unpredictable, a linear mixed-effects model that accounted for infant age, anti-bacterials and anti-fungals, uncovered a ~~significant~~ **weak** negative correlation between bacterial and fungal loads (normalized effect size: -0.060, 95% Wald CI: [-0.119, -0.001], Fig 2f). That is, when accounting for all clinical covariates, samples with higher fungal loads tended to have lower bacterial loads. This inverse relationship ~~between bacterial and fungal communities suggested~~ **led us to wonder whether** cross-kingdom interactions might be influencing preterm microbiota dynamics."*



Extended Data Figure 27d, *Diagnostics for linear mixed effects model.*

4. Fig. S6: For the fungal ITS, it seems like the data were fit on a linear scale, thereby biasing the best-fit line towards high-abundance samples. Probably would be best to log-transform the data before fitting.

Thank you for the suggestion, we have now fit the data on a log scale and adjusted the figure (now Extended Data Fig. 27d) accordingly.

5. I found Fig. 1A more confusing than illuminating. Despite already being familiar with the effect, it took me some time to figure out what all of the bars were supposed to be telling me.

All three reviewers found Fig. 1a more confusing that illuminating. In the light of this, we've now moved the mock community data into the supplement and replaced Fig. 1a with a simplified schematic demonstrating the general principle of how relative abundances can obscure different absolute abundance trajectories. We hope this will now be clearer for our readers.

6. More generally, are any of the primary conclusions dependent upon whether the data is analyzed on a linear or log-data?

By definition, our major analysis - the interaction inference - can only be performed on logged data as this is necessary to transform the gLV equations to a linear regression form. The only conclusion that is impacted when we change between a log or linear scale is the linear mixed effects model. When this is fit on non-logged data the correlations with day of life and antibiotic administration remain, however we find that the inverse correlation between total bacterial and fungal loads is no longer significant (Wald 95% CI: [-0.123, 0.003], Satterthwaite's p-value: 0.06). Importantly though, the LME model fits more poorly to the non-logged data, supporting our choice to

analyze on a logged scale. That said, for this reason, alongside the others mentioned above, we have toned down our discussion of the linear mixed effects model results.

7. Fig. S7: Statistical test? P-value?

Thank you for this suggestion. We tested to see whether the observations that anti-bacterials primarily inhibit bacteria and antifungals fungi was significant and found this not to be the case, so have softened our language in the figure legend to reflect this:

"Antibacterials primarily inhibit bacteria, and antifungals primarily inhibit fungi, ***however there is not a significant bias in the likelihood of either antimicrobial inhibiting their target kingdom (exact binomial tests, H0: P(Inhibition) = 0.5, p>0.05)"***

8. I would have appreciated a sentence somewhere discussing similarities and differences between the pattern of colonization and succession between pre-term and full-term infants.

As discussed above (major comment 4), we've now added additional text highlighting the differences between full-term and pre-term infants.

9. Line 165: Given the regularization that was done, we should not be surprised that many interactions contain a 0 (commensalism or amensalism).

The reviewer is completely right, the prevalence of commensal and amensal interactions is not surprising given the regularization used, we were unclear regarding two key points when trying to discuss the pitfalls of correlational analysis. We've now edited this paragraph to focus on the prevalence of exploitative interactions, which we find to be surprisingly common and that cannot be detected using correlational analysis. We've also edited our initial discussion of the regularization approach to make clear this method identifies only those interactions playing a strong role in shaping community dynamics, and will thus filter out weak interactions:

"Together this yields a highly parameterized model of community dynamics, which we fit to our data using a highly conservative regularization framework. By doing so, we were able to identify those microbe-microbe or microbe- antibiotic interactions playing a strong and consistent role in shaping community dynamics – while avoiding overfitting and filtering out weak interactions that do not influence overall community dynamics."

"Notably, we discovered that a substantial proportion of the interactions shaping preterm infant assembly are exploitative (+/-), with these asymmetric interactions comprising over 20% of inferred microbe-microbe interactions (Extended Data Fig 28c)."

10. Line 230: "occur" should be deleted

Thank you for catching this, we've now edited accordingly.

11. Line 235: "context-independent" should read "context-dependent"

Thank you for also catching this, we've now edited accordingly.

12. Line 717: The finding of Cryptococcus spp. as a dominant microbial genus in the studied mycobiome contradicts the consensus of the mycobiome community regarding the absence of cryptococcus from the normal human gut microbiome. Could the authors elaborate? (There was a reference to a table that I could not access)

The reviewer raises an interesting point, as indeed *Cryptococcus* is normally thought to be rare in the human gut microbiome. It's important to note, the *Cryptococcus* species present were *C. frias*, *C. uniguttulatus* and *C. saitoi,* none of which are typically considered pathogens, and we did not observe any incidence of the pathogens *C. neoformans or C. gattei.* Given this divergence between infant and adult mycobiome communities we've added some discussion of this (and the rarity of *Saccharomyces*) to our main text:

"Of note, though rare in adults[30], Cryptococcus was the dominant fungal genus in approximately 5% of all samples, while despite being a common inhabitant of the adult gut, Saccharomyces species[30] were found in only 5 infants."

13. It's surprising that Saccharomyces spp. wasn't found in the gut mycobiome since it's the most abundant yeast in the human gut microbiome. (see: https://doi.org/10.1186/s40168-017-0373-4)

As above, we thank the reviewer for pointing out this interesting difference between our cohort and what's known of the adult mycobiome. Indeed, we only found *Saccharomyces* spp in 5 of the infants we sampled. As mentioned above, we've now added a brief discussion of the differences in mycobiome composition we observe.

14. Lines 160-161: The species should perhaps always be mentioned. "Candida inhibited both Klebsiella and Escherichia, but was itself inhibited by Staphylococcus."

We apologize for being unclear; here we were referring to the results of the inference which we conducted at the genus level. We've now edited our text to hopefully make this clearer:

*"Our ecological inference predicted that strong intra- and inter-kingdom interactions **between specific microbial genera** play a pivotal role in shaping infant gut assembly, independently of antibiotic perturbations"*

We've also now added a supplementary table (Supplemental Table 14a) outlining exactly which species pairs were used for all of the *in vitro* interaction validation experiments.

15. There are data showing that E. coli produces a fungicidal factor against C. albicans. (doi: 10.15698/mic2018.05.631.)

Thank you for bringing this to our attention, we wonder if this may be an example of an interaction that is either context-dependent so does not occur in the gut (for example, the fungicide appears to require magnesium limitation) or is relatively weak, so is dropped in the regularization. This touches on an important point also mentioned by Reviewer 3 that context can play a key role in determining how taxa interact with one another, and we've now added a short sentence discussing this point more explicitly in our main text:

"These data also underlined the vital importance of context when studying microbiota interactions; illustrating how taxa may interact differently in vitro versus within a host-associated microbiota."

16. There is some evidence that the interactions between C. albicans and S. epidermidis is synergistic (https://doi.org/10.1186/1471-2180-13-257 and https://doi.org/10.1099/0022-1317-51-4-344)

Thank you for bringing this to our attention, we were interested to see that Pammi *et al*'s paper suggests *Candida* does worse in mixed species biofilms than in single species ones, adding further evidence to the cross-kingdom interactions we infer. We also find it interesting that both papers observe a positive effect of *Candida* on *Staphylococcus*, which we don't infer. As with the fungicide mentioned above, this might be because the positive interaction is weak and doesn't affect overall community dynamics (so is regularized out), or because this interaction is specific to the biofilm nature of the experiments referenced. Either way, this again underlines the importance of context on microbial interactions, and we've now cited the Pammi paper in the main text (note, we have not referenced the Adam *et al* paper as it focuses on antibiotic susceptibility, as opposed to direct growth inhibition / promotion):

*"As predicted, Candida members caused on average a ~100-1000-fold inhibition of each Enterobacteriaceae and experienced a ~10-100-fold reduction in growth when co-cultured with Staphylococcus, **consistent with previous observations from mixed species biofilms[48].**"*

17. Line 544: In vivo co-colonization validation. A reference for the antibiotic pre-treatment and the ampicillin drinking water should be added. Was it necessary to add antibiotic treatment during the co-colonization? It may change bacterial growth in monoculture colonization/co-colonization. Also, bacteria may respond differently as a response to the antibiotic. More generally, it seems that it would be good if the animal experiments mimicked the real conditions as much as possible.

Thank you for bringing this to our attention. We've now added references to support both the pre-treatment and antibiotic drinking water. Specifically, Jung *et al* (2019) used vancomycin and metronidazole treatment to overcome

colonization resistance against *K. pneumoniae* and Koh *et al* (2008) used penicillin drinking water to achieve a higher colonization level of *C. albicans*. We implemented these treatments following pilot experiments that found without the antibiotic treatment *K. pneumoniae* was rapidly lost from the mouse gut following inoculation (data not shown), likely due to colonization resistance from endogenous gut commensals. Importantly though, all of the NICU isolates used in our colonization experiments were resistant *in vitro* to ampicillin, suggesting that the antibiotic treatment should not directly affect any interactions.

It would of course have been ideal to directly mimic the exact antibiotic conditions experienced by the infants in our cohort in our mouse model. Unfortunately, this was not possible as the dosing and frequency of antibiotic administration varied substantially infant to infant. Notably though, all 13 of the frequently sampled infants received ampicillin during their first week of life and 6/13 received vancomycin, thus we believe the antibiotic treatments in the mouse model do capture this aspect of real preterm conditions.

18. I couldn't find important information that would be needed to reproduce the experiment, such as an anesthesia protocol and inoculum concentrations. It's not clear when the authors say: "lower inoculation was used for K. pneumoniae than others as this bacterium more readily and rapidly established colonization in the mouse gut". What were the actual doses used? What is the background supporting this statement?

We apologize for the lack of information here. The details regarding the timeline and inoculations were in Supplemental Table 15, and we've now clarified them in the supplementary text:

"For inoculation into each mouse, we used 10^9 CFU for S. epidermidis, 10^8 CFU for Candida and 10^6 CFU for K. pneumoniae, each in 200 uL sterile PBS and injected intragastrically using a 1 ml syringe bearing a 22G ball tip needle through the pharynx. No anesthesia was performed in the gavage, but gentle handling of mice was ensured. Note that we used a lower inoculation for K. pneumoniae because a 10^8 CFU oral inoculation of K. pneumoniae was shown to achieve rapid colonization in the mouse gut (Jung, H. J. et al. MBio 2019) and here we aimed to measure the colonization dynamics of K. pneumoniae."

19. Line 805: Figure S2. What is the label "others" in A?

We apologize for the confusion. "Others" in the panel were spurious OTU calls from the sequencing reads, likely due to contamination. As they only represent <0.1% of the population, for simplicity we have now removed these from the figure.

20. It would have been nice to see Cryptococcus in a mock community, but I will not insist on this.

We agree with the reviewer that including *Cryptococcus* in our mock community would be a nice addition. Unfortunately owing to the ongoing situation we are currently somewhat limited in our access to the lab and our ability to source new strains, therefore we hope the reviewer will not mind if we do not conduct further mock community tests. Our feeling is that our current mock community is sufficient to both demonstrate the general principles of the spike-in method and validate that our method can accurately recapitulate known absolute abundances.

Referee #2 (Remarks to the Author):

The authors develop and use to great effect a technique of converting relative abundances into absolute abundances, use this to assess the temporal changes in microbial gut populations, and infer ecological relationships between different microbes in the developing infant gut. The authors support their inferred relationships by isolating representatives of their genera of interest, then showing how in vitro coculturing replicates most of the interaction dynamics suggested by their model. Interestingly, the in vitro dynamics that didn't follow the model predictions could be injected into a SPF mouse model system to mimic the model predictions, implying that these interactions were dependent on the gut environment. The key findings of this paper – that specific pairwise microbial interactions drive the pattern of succession in the infant gut, only some of which are context-dependent on the gut environment, are valuable discoveries for understanding the assembly of this microbiome.

Further, the authors show that inference from relative abundance entirely misrepresented the dynamics that could

be concluded from absolute abundances; that the dominant species of later succession (which microbe dominates after the pioneer Staphylococcus) seems to be related to the outcome of multiple pairwise interactions between dominant microbes, and that fungal and bacterial populations are negatively correlated, perhaps due to inhibitory effects of yeast species like Candida albicans.

The combination of approaches used in this work is rarely accomplished in microbiome work, especially the experimental validation of microbial interactions. In addition, the contribution of fungal and archaeal species in the human microbiome is an underexplored area, so it is nice to see that the authors applied their method broadly. This paper represents a big step in a much needed direction for the microbiome field.

Major comments:

I found the description of MK-SpikeSeq to be lacking in the text and confusingly depicted in Fig 1. It was difficult to piece together all of the information from the figure, methods and supp info to get a clear understanding of both how it works and the various validations used with mock communities. It might be more effective to move it all to the supplementary material and combine the associated data together.

All three reviewers suggested that we clarify our description of MK-SpikeSeq both in the text and in Fig 1. We have therefore completely revamped Fig 1 (see below) and edited our main text and supplement to present a clearer understanding. Specifically, as suggested we've now moved our mock community validation experiments from Fig 1 into the supplement and replaced them with simpler schematics outlining how relative abundances can obscure community dynamics and describing in more detail how the MK-SpikeSeq method works. We've also substantially simplified our discussion of method validation in the main text and expanded our discussion of method validation/ comparisons with existing methods in the supplement.

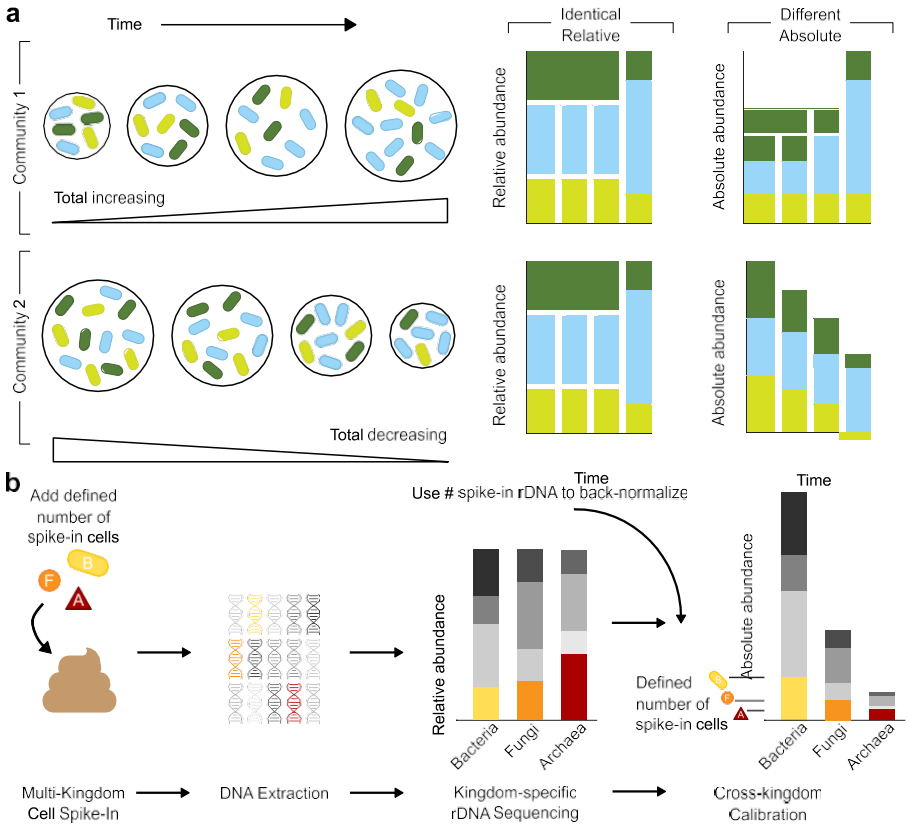


Figure 1. Multiple Kingdom SpikeSeq (MK-SpikeSeq) enables robust quantitation of absolute abundances. a, *Schematic illustrating how relative abundance data can mask underlying community dynamics, rendering it challenging to distinguish different ecological scenarios.* ***b,*** *Overview of the MK-SpikeSeq pipeline. Prior to DNA extraction, defined amounts of each spike-in cell (bacteria (B) , fungi (F) and archaea (A)) are added to each microbiome sample. Relative abundances of each microbial kingdom are then quantified using standard kingdom-*

specific rDNA amplicon sequencing. As the absolute abundances of each spike-in cell's rDNA are known, these quantities can be used as back-normalization factors to calculate the absolute abundances of all other organisms present in each sample. The spike-in cells also serve as internal controls for the entire sample processing procedure, rendering the absolute quantification robust to factors such as sample-to-sample variable DNA extraction efficiency.

Figure 1C shows how the absolute abundance of spike-in microbes can be back-calculated from their known dilution factor. In lines 87-88, the authors claim that this approach is "robust to variable DNA extraction efficiency," but does not address bias in extraction efficiency between, e.g., gram positive bacteria versus their gram-negative spike-in standard or yeast species compared to filamentous fungi. Further, in Supp figure S2, some of the filamentous fungal species appear to be present at very low levels (if at all). Is this due to differences in DNA extraction efficiency, lower input, PCR bias, or something else. My concern is that there is systematic over- or under-estimation of certain kinds of microbes biased by the selection of spike-in species. It would be nice to see whether or not the species selection has a significant effect on the calculation of absolute abundance of other microbes.

The reviewer raises a very important point. Compared with existing methods of quantifying absolute abundances from microbiome samples, our method is more robust to confounding factors like host cells that affect overall microbial DNA extraction efficiency between samples (see response to Reviewer 1). However, it cannot solve the problem of variable extraction efficiency between taxa. This problem is universal to all sequencing studies and is a function of the DNA extraction methods used (Greathouse *et al* Genome Biol 2019). To minimize this issue we used a standard commercial kit that includes the bead-beating mechanical lysis, a practice shown to improve the DNA extraction performance (Costea, P., *et al*. Nat Biotechnol, 2017), however, we cannot overcome this problem entirely.

Crucially though, the variability in extraction efficiency between taxa should be conserved sample to sample. As such, this variability should not impact comparisons in microbial load between samples, and thus will not affect any of our inference analyses. To show this conservation of extraction efficiency, we've now added additional supplementary figures (Extended Data Figure 2) illustrating that MK-SpikeSeq accurately captures changes in absolute abundances for individual taxa, including Gram-negative and Gram-positive bacteria and a range of fungal species including filamentous fungi, within both mock communities and human fecal samples. We've also expanded our Supplemental Text (Section 2 – MK-SpikeSeq validation) to discuss in more detail the differences between MK-SpikeSeq and existing abundance quantification methods.

Finally, and perhaps most importantly, the choice of spike-in species will not affect the abundance measurements as the spike in cells simply serve as a normalization factor during the sequencing process – thus they will not impact either the underlying community composition or the extraction efficiency of other taxa. To demonstrate this, we have performed rDNA amplicon sequencing on a set of samples with and without spike-in, which confirm that our spike-in does not alter rDNA relative abundance profiles (Extended Data Fig 4d).

In Figure 3D and F, why are the input levels for certain species so different between alone and interacting conditions? How do we know that this doesn't affect the growth outcome?

We apologize for being unclear here. The first data points shown in Fig. 3D and 3F are not the input levels but rather the first measurement taken (i.e. the first stool sample collected 12 hrs post gavage). The same input levels were used for both the mono- and co-colonization experiments. We have now edited Fig 3d/f, the accompanying figure legend and our supplementary text to make this much clearer:

"For inoculation into each mouse, we used 10^9 CFU for S. epidermidis, 10^8 CFU for Candida and 10^6 CFU for K. pneumoniae, each in 200 uL sterile PBS and injected intragastrically using a 1 ml syringe bearing a 22G ball tip needle through the pharynx. No anesthesia was performed in the gavage, but gentle handling of mice was ensured. Note that we used a lower inoculation for K. pneumoniae because a 10^8 CFU oral inoculation of K. pneumoniae was shown to achieve rapid colonization in the mouse gut (Jung, H. J. et al. MBio 2019) and here we aimed to measure the colonization dynamics of K. pneumoniae."

"CFU counts quantifying microbial fitness in vivo in a specific pathogen free (SPF) mouse model, reproducing the predicted exploitation of Staphylococcus by Klebsiella. ***Gavage 2 indicates day of inoculation with K. pneumoniae****"*

Minor comments:

• Within the abstract (lines 18-19), the authors claim that the "predictable assembly [of the infant gut] is vital to health," yet the cited papers do not support this exact claim, rather just that gut composition is important to health status.

We apologize for being unclear here, we have now altered this sentence and another in the main text to better reflect the references cited:

"The gut microbiota is vital to preterm infant health[8,9] yet the forces underlying these predictable dynamics remain unknown."

"Remarkably, this developmental process occurs in a strikingly predictable manner, with specific bacterial taxa establishing at distinct points in infant life[18–22]. The early life microbiome is critical to infant and later-life health, with microbiota composition linked to a range of disease outcomes, morbidity and mortality, particularly within vulnerable preterm infants"

• Authors could clarify that the antibiotic they are referring to in line 146 are not microbiome-produced, but rather clinical antibiotics.

Thank you for bringing this to our attention, we've now clarified this:

*"The model also allows each **clinically-administered** antimicrobial agent to inhibit, promote or have no effect on each community member."*

• Figure 1a might be streamlined by demonstrating how the same relative abundance scenario could be represented by two different absolute abundance scenarios, instead of convoluting the modeled scenario with different cumulative populations trajectories.

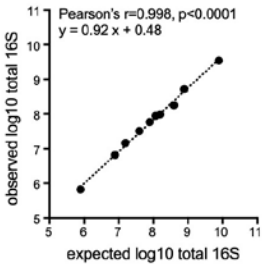
All three reviewers found Fig. 1a more confusing that illuminating, as discussed above (major comment 1), we've now moved the mock community data into the supplement and replaced Fig. 1a with a simplified schematic demonstrating the general principle of how relative abundances can obscure different absolute abundance trajectories.

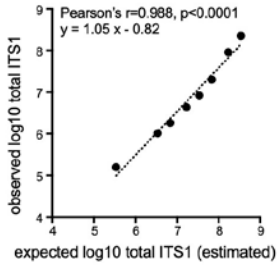
• I would like to see multiple infant successions, even if in supplementary, to evaluate how representative the one shown in Figure 2C is.

Thank you for your suggestion, we've now included additional figures plotting the succession dynamics for all of the infants in our cohort (178 in total). We've also included new pathway visualization figures that plot how individual infants' microbiomes and mycobiomes transition over time relative to all the infants in our cohort, which we've plotted for the 13 infants for whom we have very high frequency samples (see response to Reviewer 1).

• Figure S2a and S2b: the scatter plots are missing axis text – important to know how much expected versus observed abundances only correlate, as mentioned in the text, versus how close they match.

Our apologies, we have now edited these figures (now Extended Data Figure 2) accordingly, adding axis text and information on the statistical fits:





Referee #3 (Remarks to the Author):

This manuscript by Rao et al introduces a new method for absolute quantitation of bacteria, archaea and fungi present in the intestinal microbiota. They apply this method for analysis of longitudinal samples from a cohort of 178 preterm infants, and use the results to model and predict intra- and inter-kingdom interactions, which are finally validiated by in vitro experiments as well as in animals.

The methodology is original and highly relevant, since the many limitations of focusing only on relative bacterial abundances within the microbiome are becoming increasingly clear to the field.
The results obtained are novel and important for the understanding of the establishment of the preterm microbiome, as well as for the understanding of microbial ecology in general.

I have the following comments and suggestions:

1) The microbiome of preterm infants is significantly different from that of infants born within term, the latter being characterized by a high abundance of bifidobacteria and a massive influence of breastfeeding. I therefore think it is important to make it clear already in the title, that the present study deals with preterm infants. Perhaps also emphasize it a bit more in the abstract since the facultative organisms selected by the aerobic preterm gut environment differ drastically from those of the normal term infant microbiome.

The reviewer makes a very good point, and we've now changed our title and abstract accordingly. We've also included more discussion of how the preterm microbiome differs from that of full term infants in our main text:

Title:
*Multi-kingdom quantitation reveals distinct ecological drivers of predictable ~~early-life~~ microbiome assembly **in preterm infants***

Abstract:
*"The **preterm** infant gut microbiota develops remarkably predictably, with pioneer species colonizing after birth, followed by an ordered succession of other microbes. **The gut microbiota is vital to preterm infant health yet the forces underlying these predictable dynamics remain unknown**."*

Main text:
*"Yet despite the apparent importance of the infant microbiota, we do not understand what drives initial microbiota development. **Premenstrual age (term vs. preterm), mode of delivery (Caesarian section vs. vaginal), host epithelial and immune ontogeny, diet (formula vs. mother's milk), antibiotics, and the interactions between***

***individual microbes have been demonstrated or have the potential to influence microbiota composition**. But with so many moving parts, the specific role of any individual factor in mediating microbiota development and assembly has remained unclear."*

*"Specifically, we assembled a prospective cohort of 178 preterm infants from a tertiary-care neonatal intensive care unit (NICU). **The assembly of the preterm microbiota differs substantially from that of term infants. Most preterm infants are born via C-section and as such are seeded with skin and hospital-associated microbes and devoid of key maternal derived bacteria[7,21,38]. In addition, the pre-term microbiota displays "delayed" maturity with a prolonged membership of facultative anaerobic bacteria compared to that of a strict anaerobic community of term infants[7,21,38]**. We focused on preterm infants due to their clinical relevance and because they are amenable to high-frequency longitudinal sampling with readily available clinical and dietary metadata."*

2) The researchers spike the samples with bacterial, fungal and archaeal DNA, and can quantify the absolute abundances within these kingdoms. However, the manuscript and figures are focused only on bacteria and fungi, with no mentioning of the results obtained for archaea. These would be interesting to know. Even if they are not conclusive, it would be good to include them.

Thank you for the suggestion, we've now moved our plots of archaeal community development from the supplement into Fig 2. Specifically, we now show how total archaeal abundances change over time, and illustrate for each week of life the proportion of samples in which archaea were / were not detected (see below). We've also edited the main text to discuss this in more detail:

"Archaea were notably rare within our cohort, with most samples showing no archaeal signal. However, we did detect a weak positive trend in both the frequency of archaeal detection (Chi Squared test for trend, p=0.002), and total archaeal load over time (Spearman's R=0.13, p=0.002), with higher archaeal abundances generally detected in later weeks of life (Fig 2g,h)"

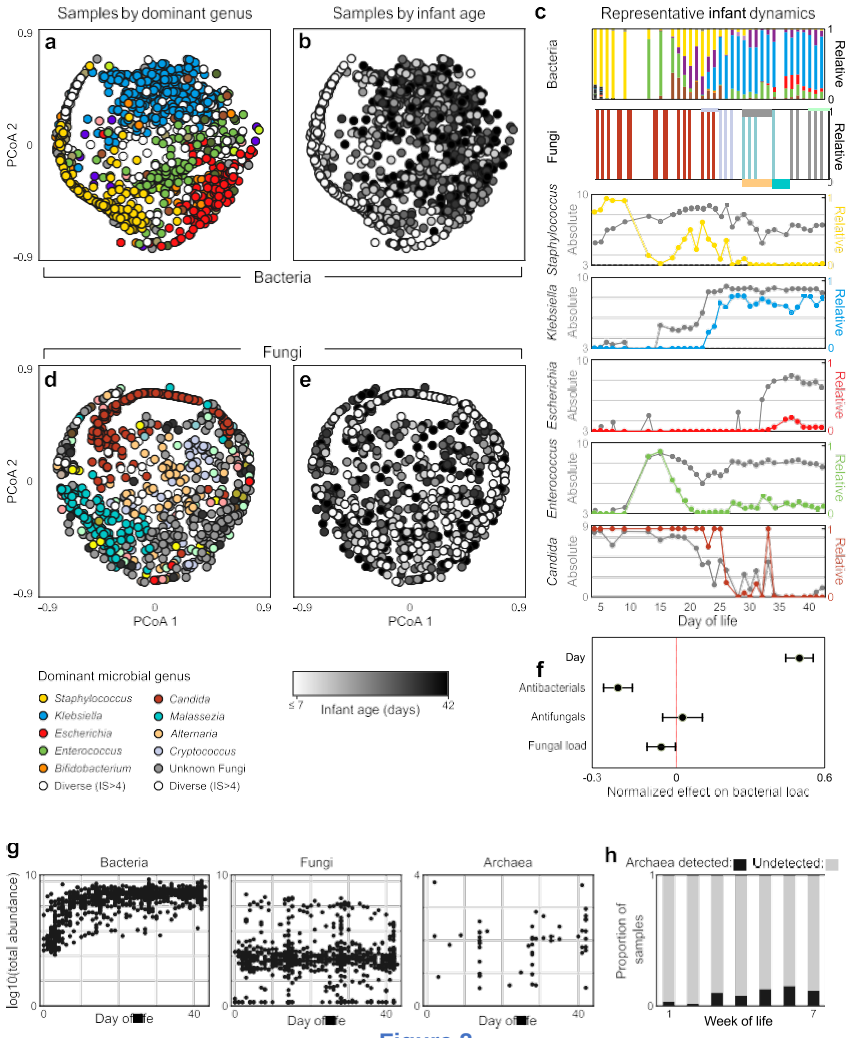


Figure 2

3) l. 118 - The blooming of Klebsiella is not really masked in the relative data.

Our apologies for being unclear, here we were referring to the fact that absolute abundances show *Klebsiella* begins to bloom within the infant gut several days before this rise is evident in the relative abundance measurements. We've now adjusted our text slightly to clarify this point, and referred to other infants where we see similar discrepancies between the relative and absolute abundance dynamics of *Klebsiella*:

"In several infants, relative abundances masked ***the timings of*** *blooms in Klebsiella and Escherichia (Fig. 2c, Extended Data Figs. 10, 15, 16, 20), and showed Staphylococcus and Enterococcus collapsing in the community when their abundances were, in fact, comparatively stable (Fig. 2c,* ***Extended Data Figs. 10-20****)."*

4) l. 125 - Similarly, I think that '..uncovered dramatic blooms and collapses in fungal genera masked by relative abundances..' is a bit overstated, since these blooms and collapses are largely also seen in the relative abundance data.

Again we apologize for being unclear. We've now toned this statement down somewhat, but also added additional text and figures to clarify the dynamics we were referring to:

"As with bacterial communities, MK-SpikeSeq uncovered ~~dramatic~~ blooms and collapses in fungal genera masked by relative abundances. ***For example, in several infants Candida stably maintained a high relative abundance within the mycobiome, despite dropping by multiple orders of magnitude in absolute load over time (Fig. 2c, Extended Data Figs. 10, 14, 19).***"

5) l. 165-166 - I think that this is remarkable and surprising. It there a risk that the applied model might be biased against the detection of symmetric (+/+ and -/-) interactions?

This is a very good question and we were not sufficiently clear in our manuscript: our model is somewhat biased against +/+ and -/- interactions as the regularization will filter out weak interactions that do occur within the gut, but do not play a major role in shaping community dynamics. In other words, the commensal/amensal interactions observed may in fact represent +/+ or -/- interactions, but where one partner is only very weakly affected by the other. What our results do imply, however, is that strong symmetric interactions are either very rare within the microbiome, or do not frequently impact community dynamics.

Notably, there should not be any bias in favor of strong +/- interactions, yet these are common in our inference. This is an important observation, as such exploitative interactions cannot be detected by correlational analyses. We've now edited several sections of our text to make these points (and the biases in our model) clearer:

"Together this yields a highly parameterized model of community dynamics, which we fit to our data using a highly conservative regularization framework. By doing so, we were able to identify those microbe-microbe or clinical antibiotic-microbe interactions playing a strong and consistent role in shaping community dynamics ***– while avoiding overfitting and filtering out weak interactions that do not influence overall community dynamics****."*

"Notably, we discovered that a substantial proportion of the interactions shaping preterm infant assembly are exploitative (+/-), with these asymmetric interactions comprising over 20% of inferred microbe-microbe interactions (Extended Data Fig 28c)."

6) l. 186-190: It would be relevant to know which distinct species within each genus were captured. Also, it is unclear whether each of these species were tested seperately against each of the species from another genus? Or were they tested as a mix of all species from one genus? Are only representative/affirmative data shown in Fig 3c? Were there no species-species interactions which did not fit the model prediction?

Thank you for bringing up these important points, we've now included supplementary tables outlining more clearly which distinct species within each genus were isolated (Supplemental Table 13) and which species/strains were co-cultured with which for the interaction validation experiments (Supplemental Table 14a). For the validation experiments each isolated species was tested independently against multiple strains/species of the partner genus (not a mix of all species from one genus). Testing each species independently was important as it allowed us to demonstrate that in most cases interactions were conserved at the genus level – we have now edited the legend of Fig 3 (see below) to make this clearer.

Regarding the reviewer's final question, Figs 3c-f demonstrate all the interactions we tested, together illustrating that all the strong interactions between dominant taxa predicted by our inference could be recapitulated *in vitro*, or in the case of *Staphylococcus-Klebsiella*, *in vivo*. In other words, we did not find any false positives (nor in the case of *Klebsiella-Enterococcus*, a false negative). As noted above though, our inference will only identify those interactions that strongly and consistently influence community dynamics. Other interactions between taxa may be biologically possible, but our model suggests these either do not occur *in vivo*, or do not have a strong influence upon overall community dynamics. We've now edited our manuscript to make this clearer as we think it is an important point, and thank the reviewer for bringing it up (see response to comment 5 above)

7) l. 193-194: Note that the environment (gut versus lab medium) may have a huge impact on e.g. competition for nutrients.

The reviewer raises a very important point that context can play a major role in shaping inter-taxa interactions. We have now called this out more explicitly in the paragraph following lines 193-194, as we think it is most relevant to the section of the manuscript where we discuss differences in interactions *in vitro* vs *in vivo*:

"That is, we hypothesized that, due to differing environments in vitro versus in vivo, such as differing nutritional or spatial conditions, Klebsiella might only gain a benefit from Staphylococcus within the gut."

"These data also underlined the vital importance of context when studying microbiota interactions; illustrating how taxa may interact differently in vitro versus within a host-associated microbiota."

8) l. 196: You state here that no positive effect of Staphylococcus on Klebsiella was observed under in vitro conditions, but still this is marked with a big blue arrow in Fig 3c?

We apologize for being unclear, the purpose of the arrows in Fig 3c was to indicate which predicted interactions were being tested in the corresponding figure, not those validated *in vitro*. We have now edited the legend and added additional labels to Fig 3 to make this clearer:

Legend:
"in vitro growth effects of infant isolates upon one another using monoculture and pairwise co-culture, ***testing the predicted within-kingdom interactions indicated by schematic****."*

9) 228: I think a new sub-headline would be feasible here.

Thank you for this suggestion. We have now added a "Discussion" sub-headline here.

Comments to figures:
Fig 1b - I don't understand the purpose of having two fecal samples/ two soil samples with an arrow between them in the left side of the figure?

We apologize for being unclear, our goal here was to illustrate the longitudinal nature of our sampling. As discussed above, we've now substantially revamped Fig. 1 to make this clearer.

Fig 3a - What is the green round dot in this figure representing?
This was supposed to indicate that our interaction network captures within- and between-kingdom interactions (ie the circle is supposed to be a fungus) we have now edited our figure legend to make this clearer and altered Fig. 1 so the use of a circle to represent fungi is consistent throughout:

"a. Schematic illustrating the generalized Lotka-Volterra (gLV) model used to identify causative drivers of ~~microbiota~~

bacterial (purple ovals) and fungal (green circle) *dynamics* ***within the infant gut****."*

Fig 3c - see comment 8). Also, do the dots represent replicate experiments or different species? (If both, it would be good to label the difference as for Candida).

Thank you for pointing this out. As discussed above, we have now added a new supplementary table (Supplemental Table 14a) to indicate each of the species combinations used for each experiment. We have not labelled the differences between species in the figure as we did for *Candida* because we found this made the figure harder to understand (while we only had two *Candida* species, we had multiple representatives of most of the other genera). However, we have now edited our figure legend to make clear to the reader the dots refer to distinct isolate combinations:

"Figure 3c. *in vitro growth effects of infant isolates upon one another using monoculture and pairwise co-culture, testing the predicted within-kingdom interactions indicated by schematic.* ***Each dot represents a unique combination of an individual isolate from each genus (Supplemental Table 14a)****."*

Fig S6 - Something went wrong with the panel letter references in the legend to this figure. Thank you for catching this, we have now edited the figure legend accordingly.

Reviewer Reports on the First Revision:

Referee #1 (Remarks to the Author):

The authors have done a good job addressing the questions raised by me and the other referees. I believe that this is an excellent study that can be published as is.

Jeff Gore

Referee #2 (Remarks to the Author):

Thank you for providing detailed responses, and for making the suggested changes. I do not have any further comments or concerns.

Referee #3 (Remarks to the Author):

The authors have answered all the issues that I raised, and I have no further comments.
I thank the authours for their thorough rebuttal, which made it easy for the reviewers to identify and assess the changes that were made.