nature portfolio

Peer Review File

Deep longitudinal lower respiratory tract microbiome profiling reveals genome-resolved functional and evolutionary dynamics in critical illness

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

Cheng et al investigated the lower respiratory tract microbiome of 157 critically ill patients monitored longitudinally. They developed a method that allows them to recover near-complete microbial genomes using a low-biomass microbial enrichment method. They characterized the lower respiratory microbiota of patients admitted in 3 different hospitals in China and compared the differences in antibiotic resistance, dominant species, temporal dynamics, associations with clinical variables and more. In general, the authors have applied state-of-the-art computational pipelines to analyze their data, however, there are no in vitro validations for any of the proposed conclusions. A few more specific comments:

1) In the Results section "The development of CMEM enabled the deep-sequencing of LRT microbiomes" the authors write "…the microbiome alpha diversity in samples from pneumonia patients was significantly lower than those without pneumonia (Supplementary Fig. 1g). Furthermore, we noted a significantly higher total abundance of ARGs in samples from pneumonia patients". The minor comments here is that "significantly lower" and "significantly higher" statements should be accompanied with the statistical test and p values. Now if I look at the Suppl. Table 1 the patients with diagnosed pneumonia are coming only from Hospital B and C since this information is missing in Hospital A. Out of the 51 patients in Hospital C all were treated with antibiotics and 50 out of 51 had pneumonia. In Hospital B there are 69 patients with pneumonia and 17 without pneumonia and 50% of all the patients are taking antibiotics but the authors do not specify which are the patients treated with antibiotics. So do the authors in their statistical comparisons correct for antibiotics usage? Are the conclusions here and elsewhere that the difference between patients with pneumonia and patients without pneumonia is simply a comparison between patients treated with or without antibiotics? If you make in Hospital B a subgroup for comparison of patients with pneumonia treated vs not treated with antibiotics, how many of the conclusions still stand? I am not sure what the authors' main message is in the paper that antibiotics induce changes in the microbiome and resistome (which is well documented in the literature) or that pneumonia is a factor shaping the microbiome and resistome?

2) In the Results section “Associations between clinical variables and the LRT resistomes” the authors discuss the impact of prolonged stay in ICU, especially above 28 days. In the Excel file of the demographics the range of days in ICU for all 3 hospitals is less than 27 days. Have I misunderstood something?

3) In the last section of their paper the authors study the strain transmission events within and between wards. They use as examples pairs of patients with overlapping hospitalization patients to conclude the dissemination of strains. However, for this part I am not convinced. The authors have longitudinal samples, do they observe changes in the dominant strains of the pathogens during the stay in the hospital? Do they see cases where upon admission in the ICU a strain is not found in a particular patient but it is detected in the subsequent sampling days? The authors should take advantage of their longitudinal sampling to provide strong evidence of transmission from the hospital setting or other patients of strains otherwise this section is particularly weak.

Reviewer #2 (Remarks to the Author):

In this study, Cheng et al., used 453 lower airway (tracheal aspirate) longitudinal samples from 157 critically ill intubated patients to first attempt to reduce mammalian DNA prior to sequencing and then to identify metagenomic signatures of interest in these critically ill patients. The paper deals with the important problem of obtaining microbiome data beyond just taxonomy (amplicon based commonly used) using WGS approaches in low biomass samples. Of note, I have reviewed this paper in now two prior submissions to different journals. For this submission, I did a new fresh look without reviewing my comments from before and again I see a significant change in the approach to describing their data including some newer data and analyses.

Regarding the improvements: a) some of the overstated statements were removed; b) More background controls were included though with some problems as described below; c) they added a comparison between culture results and sequencing results, though with some issues.

I still see several issues that seem rather unacceptable to me:

• They still try to claim that they have accomplished a better approach of mammalian DNA depletion leading to better quality of metagenome data.

• The lack of well described and planned head to head comparison is a major issue for this claim.

• There is still very significant lack of clarity of methods for sequencing and analyses (e.g., how they calculated oxygenation index).

• They are not making their data/analytical codes publically available which raises concerns about data reproducibility (internal).

Thus, I am still left with the impression that for a methodological "advancement" claim, it seems to fall short on their design. Previously I recommended the authors to try to decide where their focus should be, methodological advancement or just limit to describe their findings and aknowledge their limitation properly.

To be clear, I have tremendous appreciation for the efforts put into this paper: the large sample size and longitudinal data is certainly a strength, there are several analyses that seem well done and several observations are quite interesting specially related to mobile elements and evolution of MAGs. I would like to see this published because I see that it could serve the scientific community and I'm not making any recommendation based on where this journal is submitted (that for the editors to decide based on their priorities). Given that this is the third time I carefully review this paper (believe me I do not want to keep reviewing it) I'm going to take the liberty to give the authors my opinion of how I would deal with the issues raised below. I think regarding the methodological advancement issue, I would make a recommendation despite that I do not know, nor have access to the data or analytical codes: It seems to me that their design has not been prepared to address this and they should just acknowledge this and refraing from claims of superiority compared to other methods that have not been tested. I think the authors decided on this approach hoping to get higher depth of coverage for a MAGs assembly approach, which would make sense. There is no clear well design head-to-head comparison. They should be clear about the methodological analyses and data for the background controls and I can understand that there might be few MAGs assembled though there is probably multiple short reads annotated to microbes in those samples. The justification of going with a MAG approach is clear and would be appreciate it by the analyses perform including the evolutionary and mobile elements. They should limit to describe their findings based on this data and address the possibility of different findings with different methods (not necessarily better or

worse). Their limitation paragraph should be extended to include all of these. They should also mention that this data is limited by the complexity of the cohort, multiple varibles that are difficult to capture (e.g, treatment received, comorbidities, etc), type of samples (these are tracheal aspirate and different sample types, including BAL or brushings may lead to different results) among others. I really hope this helps the authors.

Below are my point-by-point comments.

1. Introduction: “Conventionally, researchers have predominantly relied on culture-based methods for genomic and functional characterization of LRT microbiota due to the low microbial biomass and high level of host contamination”. Not sure that I agree with this. I think there is clear role for culture-based method to identify the presence of certain bacteria but I do not think that anyone thinks that it can be used to profile the LRT microbiota. Then they state: “The culture-based isolation of these opportunistic pathogen species has provided profound insights into the mechanisms of pathogenesis, antibiotic resistance profiles, and virulence factors”. Which “opportunistic pathogen species”? Where does this come from? It seems disconnected.

2. Putting together what was done is quite difficult. For Sup Fig S1, they compared, as per legend: “‘NBT+PowerWater’ refers to a combination of the previously reported method for host DNA depletion, followed by Qiagen PowerWater for the extraction of the remaining total DNA; (2) ‘CMEM’ represents our newly developed chelex100-based low-biomass microbial-enrichment method. See the Methods for detailed information.”. Then I go to the methods and I get that: “We first attempted reported host DNA removal methods for 26 of our samples. However, the low microbial biomass resulted in undetectable DNA concentrations (measured by Qubit) for approximately half of the samples, making it impractical for metagenomic sequencing.” How were those 26 samples selected? Fig S1 does not show a head to head comparison between the two methods. I would expect to see 26 samples in both Fig S1a and Fig S1b. What threshold was used to define successful DNA extraction? For attempted reported host methods use for their ‘NBT+PowerWater’ they cite the following paper: “Charalampous, T. et al. Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. Nat Biotechnol 37, 783–792 (2019)”. Why was this selected as the comparison? That paper also uses a completely different technology for sequencing. Why was it decided that it was a good standard to compare to? The comparisons need to be clearly stated and the reader should not work as hard to figure out what the comparisons are.

3. In reference to the above, in line 119: “To assess the impact of our method on the LRT microbial community, we collected 32 endotracheal aspirate (ETA) samples and divided each sample into two aliquots: one assigned to the control group and the other to the treatment group.”. It is still not clear what the “other treatment group” is. Also, remove the “to the” in “the other to the treatment group”

4. Line 108: “CMEM removed up to 97% of human reads, leading to a maximum of 97% microbial reads in the sequencing results (median 13%, interquartile range 10%-16%; Supplementary Fig. 1c)”. While from the strictly data stand point is accurate the emphasis is on the outliers with their statement “97% of human reads, leading to a maximum of 97% microbial reads”. Also, where is the head to head comparisons with other method to show that CMEM removed up to 97% of human reads? Without a comparison, it could be possible that in a sample not treated with a CMEM approach all it were there would be microbial DNA, especially in these intubated critically ill patients.

5. Line 121: “The CMEM introduced small variances in the relative abundance of individual taxa between some pairs, but we observed no systematic biases between the two groups ($p > 0.05$ in Supplementary Fig. 1d and Supplementary Table 2)” None of the shown data is sufficient to test this. Supp fig 1d is a beta diversity plot and cannot evaluate for individual taxa and the lack of difference in PERMANOVA (I assume that’s what the p value is based on) is a high bar for this. Then, Supplementary Table 2 is not a differential enrichment analyse but a table with I think relative abudnace expressed as % (though numbers do not necessary add up to 100% in all samples). Also, in that table there are only 90 taxa. Is that all there is using this metagenome approach? This is clearly much less taxa than a lot of metagenome studies using respiratory samples (see PMID 38168095, 35544065, 37009900 and 34465900), which are mainly based on short reads. Finally, what is the control? Is it the NBT+PowerWater method that is also not clear as mentioned above. I recommend to make this easy to find for readers and reviewers.

6. Line 124: “the more abundant taxa were impacted less, underscoring the accuracy of CMEM in identifying opportunistic pathogen species in LRT samples (Supplementary Table 2)”. What’s the criteria used to define “opportunistic pathogen specie”? The fact that a taxa is found in culture independent data does not make it opportunistic nor pathogen. Also, what analyses was used to support that “the more abundant taxa were impacted less”

7. Line 129: “Consistent with previous findings, our results revealed that the microbiome alpha diversity in samples from pneumonia patients was significantly lower than those without pneumonia (Supplementary Fig. 1g)” I would expect some description of the cohort and sample selection to conduct this analyses.

8. Line 134: “To further validate our approach, we observed a 92.2% consistency between microbial species detected via clinical culture and CMEM for samples with available clinical microbiology data (Supplementary Table 5)”. If the data annotated on the metagenome approach involves only ~90 taxa as depicted in Supplementary table 2, then it is not possible that the culture data and the metagenome data overlaps that well. For example, Citrobacter koseri is seen in culture data but I cannot see that listed among the taxa in Supplementary table 2. I do see it in supplementary Table 6 but it is not clear why they are getting different taxa (probably due to different samples used but this is not clear) and why so few (probably because of the MAG approach rather than short reads).

9. Then in line 160: “De-novo assembly of microbial reads and binning at the individual level resulted in 433 metagenome-assembled genomes (MAGs), of which 120 MAGs were high-quality (Methods; Fig. 1d).” This should probably go before the sections above that utilizes the 120 MAGs for tables such as Supp table 2, 5 and 6. What is the criteria used to define high-quality MAGs. The way it is presenting is asking the reader, and reviewer to work really hard to figure things out.

10. Figure 1C depicts Percentage of Detected samples (%). Can the authors explain what 150% means for C. striatum and all other ones. I know what they did but having it as stacked bars do not make sense. They want to have parallel bars with each site having their own % which should not be > 100%

11. Line 221: “We further compared the oxygenation index (OI) and the percentage of neutrophils between patients with a consistent dominant species and dynamically changing dominant species over time. Our findings showed that patients with changing dominant species exhibited a significantly lower OI ($p=0.016$)”. How was OI calculated.

12. Line 242: “the most predominant ARG types were significantly more abundant in Hospital C than in Hospital B (Fig. 2f and Supplementary Fig. 3b), reflecting the increased antibiotic usage in Hospital C”. Where is the data that show that Hospital C has increased antibiotic usage? How was that measured? From supplementary table 1, I see that Hospital A and C had similar Antibiotics at ICU admission. Again, the authors should make sure that every statement is adequately supported with objective data

13. The limitation paragraph is really short and several of the comments above could be dealt with a more extensive discussion in this paragraph.

14. Negative controls. There are much more samples described as negative controls than in prior versions of the paper that I review in other journals. I appreciate the attempt to improve this. However, in Line 670 they state: “For each hospital, ten negative environmental sampling control samples were collected. Sterile saline solution was

collected by aspiration through the sputum aspirator as a negative sampling control. Ten reagent controls (DNA extraction blanks) were collected at the laboratory handling study samples. All thirty negative sampling controls and ten reagent controls were included for library preparation and sequencing. Bacterial species identified in more than two negative controls with relative abundance >0.001 were considered as potential contaminants and excluded from downstream analyses (Supplementary Table 3 and 4)" what "Sterile saline solution" was used, which "sputum aspirator" was used. These have to have specific terms and make information. Then Supp table 3 and 4 are tables of relative abundance. What is the sequence depth? Are these MAGs. Why are there so few? Then, if those taxa were removed from samples, what was the relative abudnace of the total removed data from the biological samples (453 lower airway).

15. Data availability: in line 799: "Correspondence and requests should be addressed to C.J. (jiang_chao@zju.edu.cn). All human-removed metagenomic sequencing data analyzed in this study are available in the European Nucleotide Archive (https://www.ebi.ac.uk/ena/browser/home) under the project accession number PRJEB64676." Availability upon request is not acceptable. Downstream data and analytical codes should be made available to ensure internal reproducibility.

We are grateful for the opportunity to revise our manuscript and sincerely appreciate the insightful comments and suggestions provided by the reviewers. We have carefully studied and addressed each point raised, which has significantly improved the quality and clarity of our paper. Below, please find our point-by-point response.

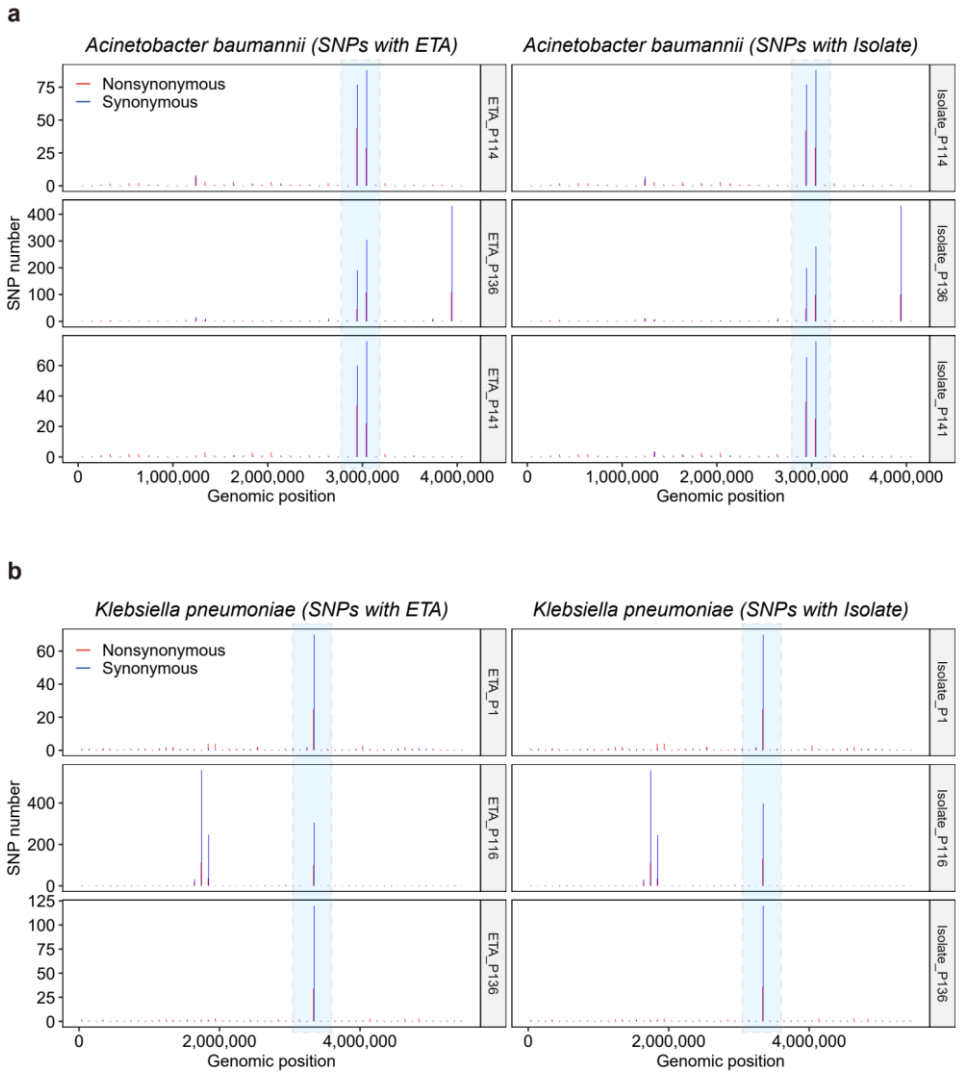
Reviewer #1 (Remarks to the Author):

General comment: Cheng et al investigated the lower respiratory tract microbiome of 157 critically ill patients monitored longitudinally. They developed a method that allows them to recover near-complete microbial genomes using a low-biomass microbial enrichment method. They characterized the lower respiratory microbiota of patients admitted in 3 different hospitals in China and compared the differences in antibiotic resistance, dominant species, temporal dynamics, associations with clinical variables and more. In general, the authors have applied state-of-the-art computational pipelines to analyze their data, however, there are no in vitro validations for any of the proposed conclusions. A few more specific comments:

Response to general comment: We thank the reviewer for the positive remarks and agree with the reviewer on the limitations of this study. We greatly appreciate the feedback, as it allows us to strengthen our work further. We have conducted additional in vitro validation experiments to provide more evidence supporting the reliability of our metagenome-assembled genome (MAG)-based analyses, which is the main focus of our manuscript.

Specifically, we have attempted to culture diverse dominant microbial species from nearly 100 clinical samples, successfully recovering 47 isolates of 7 species. 33 isolates have paired high-quality MAGs from the same ETA samples. The new genomic data are uploaded with the revised manuscript. With the genomes of cultured isolates, we have further validated our results by conducting the following analyses:

- To validate the reliability and quality of our MAGs recovered directly from ETA samples, we performed pair-wise comparisons with in vitro cultured isolates. This analysis revealed an average high genome-level similarity of 99.945% in the 33 MAG-isolate pair comparisons (Supplementary Table 8). The high similarity between the two independent approaches demonstrates the reliability of our methods in reconstructing high-quality MAGs directly from complex respiratory samples.
- Through evolutionary analyses, the SNP and recombination hotspots identified in *Acinetobacter baumannii* and *Klebsiella pneumoniae* were also consistently observed in the cultured isolates, further validating our approach (Author Response Fig. 1).



Author Response Fig. 1. Distributions of synonymous (blue) and nonsynonymous (red) SNPs in coding regions of *Acinetobacter baumannii* (a) and *Klebsiella pneumoniae* (b) identified with metagenomic sequencing data of ETA samples (left) and cultured isolates (right).

- To further validate the putative strain transmission events identified through our MAG analyses, we repeated the snp-based strain-tracking analysis using the corresponding patient-derived cultured isolates. We observed a consistently high level of genome-wide similarity between isolates cultured from samples in the putative transmission chains, validating the quality of our MAG and lending strong support to the putative strain transmission events (Author response table 1, as detailed in the response to Comment 3).

In the Discussion section, we have also provided a detailed acknowledgment of the limitation of the lack of sufficient in vitro validations. Conducting additional *in vitro* and *in vivo* experiments to validate our findings will continue to be a priority for our future research endeavors.

Discussion (Line 600-605): "*Finally, our primary findings were predominantly derived from the metagenomic sequencing data and MAGs, with several findings validated with the cultured isolate approach. However, some results, such as the observed associations between various temporal dynamics patterns of LRT microbiomes and pulmonary functions, require more detailed mechanistic experiments to verify the potential causal relationships.*"

Comment 1: In the Results section "The development of CMEM enabled the deep-sequencing of LRT microbiomes" the authors write "…the microbiome alpha diversity in samples from pneumonia patients was significantly lower than those without pneumonia (Supplementary Fig. 1g). Furthermore, we noted a significantly higher total abundance of ARGs in samples from pneumonia patients".

Comment 1.1: The minor comments here is that "significantly lower" and "significantly higher" statements should be accompanied with the statistical test and p values.

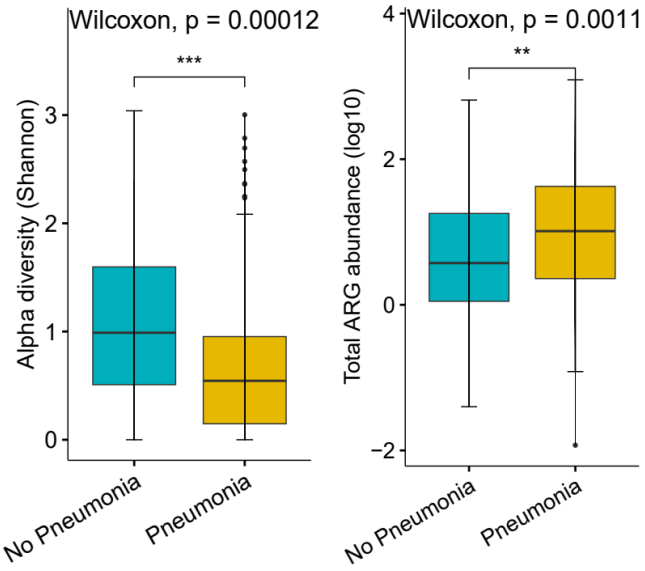
Response 1.1: Thank you for this comment. We have added the statistical test and p values in the revised manuscript.

Results (Line 150-156): "*Consistent with previous findings, our results revealed that the microbiome alpha diversity in samples from pneumonia patients was significantly lower than those without pneumonia (Wilcoxon, p=0.00012, Supplementary Fig. 1g). Furthermore, we noted a significantly higher total abundance of ARGs in samples from pneumonia patients (Wilcoxon, p=0.0011, Supplementary Fig. 1h).*"

Comment 1.2: Now if I look at the Suppl. Table 1 the patients with diagnosed pneumonia are coming only from Hospital B and C since this information is missing in Hospital A. Out of the 51 patients in Hospital C all were treated with antibiotics and 50 out of 51 had pneumonia. In Hospital B there are 69 patients with pneumonia and 17 without pneumonia and 50% of all the patients are taking antibiotics but the authors do not specify which are the patients treated with antibiotics. So do the authors in their statistical comparisons correct for antibiotics usage? Are the conclusions here and elsewhere that the difference between patients with pneumonia and patients without pneumonia is simply a comparison between patients treated with or without antibiotics?

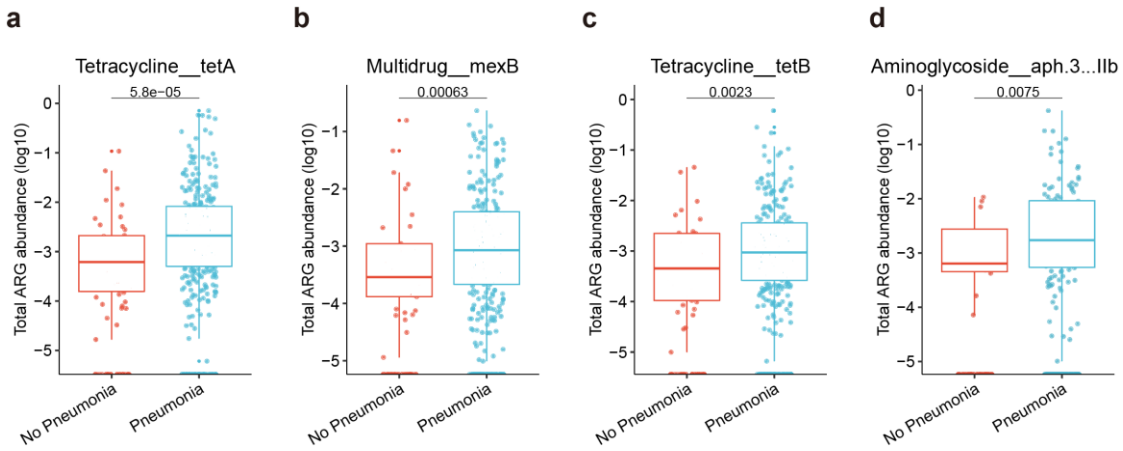
Response 1.2: We thank the reviewer for raising an important point about the issue of correcting antibiotic usage for included patients. Regarding the patients with diagnosed pneumonia listed in Supplementary Table 1, which were solely from Hospitals B and C, we apologize for the omission of information from Hospital A in our initial submission. We have now obtained the relevant information from Hospital A, with 14 patients with diagnosed pneumonia. We have updated the Supplementary Table 1 accordingly. Importantly, all patients received antibiotics treatment during the ICU stay, regardless of pneumonia status (as explained in detail below). We accordingly updated the results in the following corresponding sections:

- Our results regarding the comparisons of microbiome alpha diversity and total ARG abundance remain valid (Author Response Fig. 2).



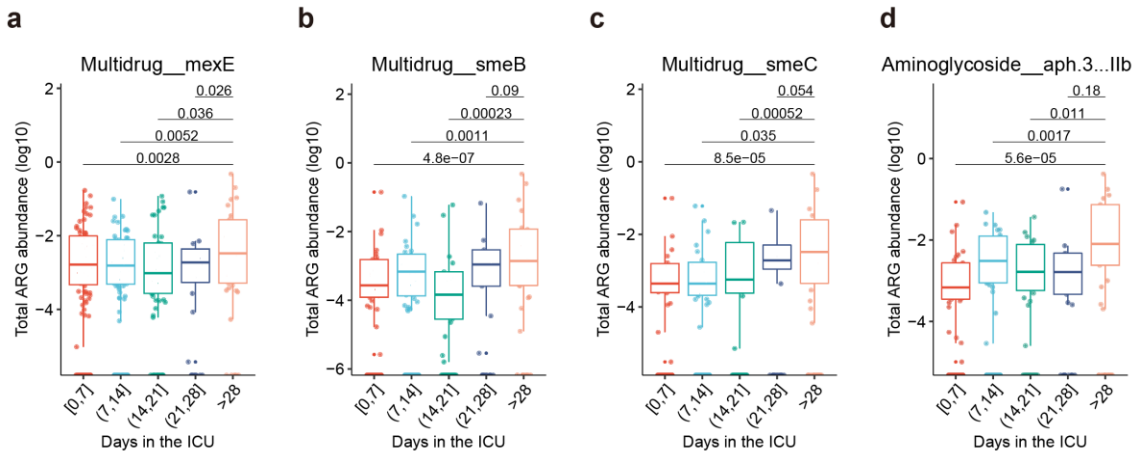
Author Response Fig. 2 (Manuscript Fig. S1h,i)**.** The Shannon microbiome alpha diversity (left) and total abundance of ARGs (right) in the samples collected from the pneumonia patient group and non-pneumonia patient group. Significance was determined using the two-sided Wilcoxon rank sum test.

- The ARGs *tetA*, *tetB*, *mexB,* and *aph(3')-IIb* were still significantly higher in pneumonia patients (Author Response Fig. 3).



Author Response Fig. 3 (Manuscript Fig. 2g,h and S3g)**.** The abundance of (a) *tetA*, (b) *mexB*, (c) *tetB*, and (d) *aph(3')-IIb* genes in patients without pneumonia and patients with pneumonia. Significance was determined using the Wilcoxon rank sum test (Mann–Whitney U test).

- The finding of significantly higher abundances of several ARGs in patients staying over 28 days remains valid (Author Response Fig. 4).



Author Response Fig. 4 (Manuscript Fig. 2i,j and S3h)**.** The abundance of (a) *mexE* and (b) *smeB* , (c) *SmeC*, and (d) *aph(3')-IIb* genes with days in the ICU. Significance was determined using the Wilcoxon rank sum test (Mann–Whitney U test).

Regarding the numbers showing that only 50% of patients at Hospital B were treated with antibiotics in Supplementary Table S1, we apologize for the confusing description "Antibiotics at ICU admission". This column was intended to indicate whether the patients were receiving antibiotics **at the time of admission to the ICU**, not their overall antibiotic treatment status **during the ICU stay.** To clarify, **all patients included in our study received antibiotic treatment during their ICU stay,** which is almost always required for pneumonia patients and most patients admitted to ICU.

We have added a new row titled "Antibiotic treatment during ICU stay" in the updated Supplementary Table 1 to explicitly indicate that all the patients received antibiotic treatment during their ICU stay. We have also added the description for the criteria for samples included in the comparisons between pneumonia and non-pneumonia groups in this revised manuscript.

Methods (Line 641-643): "*To mitigate the potential influence of antibiotic usage, all patients included in the comparisons between the pneumonia and non-pneumonia groups were treated with antibiotics during the ICU stay*."

Comment 1.3: If you make in Hospital B a subgroup for comparison of patients with pneumonia treated vs not treated with antibiotics, how many of the conclusions still stand? I am not sure what the authors' main message is in the paper that antibiotics induce changes in the microbiome and resistome (which is well documented in the literature) or that pneumonia is a factor shaping the microbiome and resistome?

Response 1.3: We thank the reviewer for this suggestion. As described in the response to Comment 1.2, all patients were treated with antibiotics during the ICU stay. As a result, we are unable to make Hospital B a subgroup for the comparison of pneumonia patients treated with antibiotics versus those not treated with antibiotics. We would like to clarify that the main message of our manuscript was not to indicate that antibiotics or pneumonia are a factor shaping the microbiome

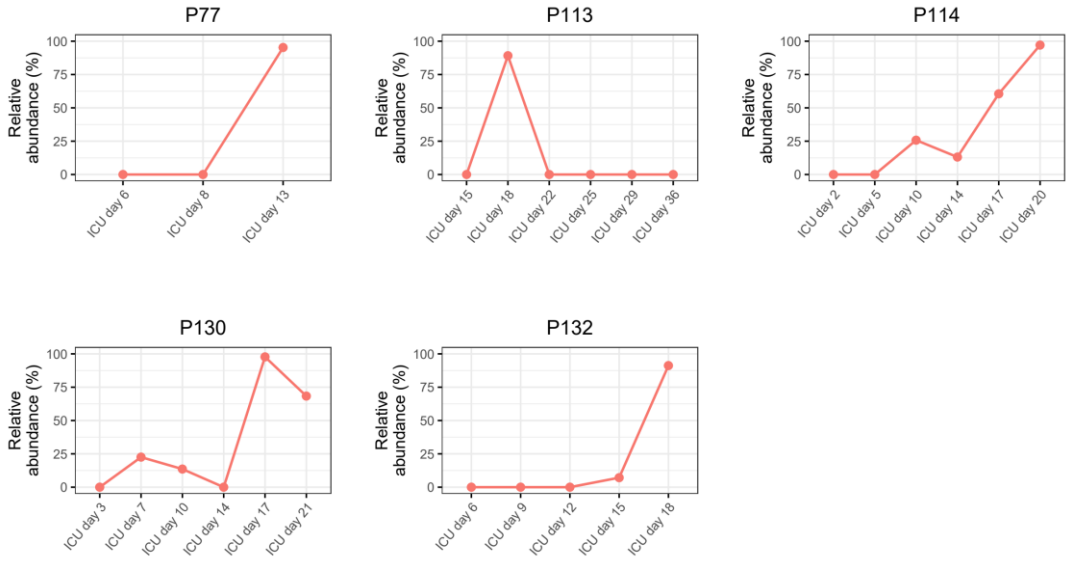
and resistome, as they have been demonstrated before. We included the analyses to validate previously reported conclusions[1–3] and demonstrate the reliability of our CMEM method in processing the respiratory samples (ETAs).

Comment 2: In the Results section "Associations between clinical variables and the LRT resistomes" the authors discuss the impact of prolonged stay in ICU, especially above 28 days. In the Excel file of the demographics the range of days in ICU for all 3 hospitals is less than 27 days. Have I misunderstood something?

Response 2: We thank the reviewer for this comment. We apologize for the confusing and misleading identifier "median (range)" in the first column of Supplementary Table 1. To clarify, the continuous variables listed in Supplementary Table 1 are presented in the format of "median (interquartile range)", as stated in the additional comment at the bottom of the table. We confirm that our study samples included patients with ICU stays exceeding 27 days. We have revised the identifiers into "median (interquartile range)" in the updated Supplementary Table 1 to accurately reflect this information.

Comment 3: In the last section of their paper the authors study the strain transmission events within and between wards. They use as examples pairs of patients with overlapping hospitalization patients to conclude the dissemination of strains. However, for this part I am not convinced. The authors have longitudinal samples, do they observe changes in the dominant strains of the pathogens during the stay in the hospital? Do they see cases where upon admission in the ICU a strain is not found in a particular patient but it is detected in the subsequent sampling days? The authors should take advantage of their longitudinal sampling to provide strong evidence of transmission from the hospital setting or other patients of strains otherwise this section is particularly weak.

Response 3: We appreciate the reviewer's suggestion to leverage longitudinal data to strengthen the evidence for the strain transmission events section. To improve the precision of our descriptions, we have added the term 'putative' before every instance of 'strain transmission events' in the last result section of the revised manuscript. Of the 12 putative transmission events involving 15 patients, longitudinal sampling data were available for 10 patients. To further substantiate the putative transmission events, we investigated the temporal changes in the relative abundance of transmitted species. For at least 5 patients, the species of interest was undetected upon ICU admission but identified on subsequent sampling days (Author Response Fig. 5). The remaining patients acquired the transmitted species before the first sampling point, indicating they were possibly infected earlier. This temporal pattern provides additional evidence supporting the putative transmission events. Additionally, we used in vitro cultured isolates recovered from the corresponding samples to validate the genome-wide average nucleotide identity (ANI) for 8 out of 12 putative strain transmission events (Author Response Table 1). The results are concordant with MAG-based analysis, further supporting the putative strain transmission events between patients.



Author Response Fig. 5 (Manuscript Fig. S7)**.** Temporal changes of the abundance of *Acinetobacter baumannii* during intubation.

Patientid_1	Collection date#	Patientid_2	Collection date#	Coverage_overlap	Compared_bases_count	SNPs	ANI	Sampling site
P114	2021/10/12	P132	2021/10/21	0.999978716	3946573	0	1	Hospital C
P25	2022/2/26	P77	2022/3/15	0.999339295	3794950	0	1	Hospital B
P113	2022/7/7	P130	2022/7/21	0.997383474	3838545	1	0.999999739	Hospital C
P113	2022/7/7	P118	2022/7/17	0.999299541	3840501	1	0.99999974	Hospital C
P123	2022/2/28	P143	2021/12/13	0.999594188	3842584	4	0.999998959	Hospital C
P130	2022/7/21	P118	2022/7/17	0.997650891	3839228	11	0.999997135	Hospital C
P141	2021/11/25	P132	2021/10/21	0.997782414	3937882	19	0.999995175	Hospital C
P114	2021/10/12	P141	2021/11/25	0.99779152	3937889	20	0.999994921	Hospital C

Author Response Table 1 (Manuscript Table. S13)**.** Genome-wide comparison for the cultured isolates (*Acinetobacter baumannii* strains).

The column "Collection date" refers to the collection date of the sample used for culture.

Reviewer #2 (Remarks to the Author):

General comment: In this study, Cheng et al., used 453 lower airway (tracheal aspirate) longitudinal samples from 157 critically ill intubated patients to first attempt to reduce mammalian DNA prior to sequencing and then to identify metagenomic signatures of interest in these critically ill patients. The paper deals with the important problem of obtaining microbiome data beyond just taxonomy (amplicon based commonly used) using WGS approaches in low biomass samples. Of note, I have reviewed this paper in now two prior submissions to different

journals. For this submission, I did a new fresh look without reviewing my comments from before and again I see a significant change in the approach to describing their data including some newer data and analyses.
Regarding the improvements: a) some of the overstated statements were removed; b) More background controls were included though with some problems as described below; c) they added a comparison between culture results and sequencing results, though with some issues.
I still see several issues that seem rather unacceptable to me:
• They still try to claim that they have accomplished a better approach of mammalian DNA depletion leading to better quality of metagenome data.
• The lack of well described and planned head to head comparison is a major issue for this claim.
• There is still very significant lack of clarity of methods for sequencing and analyses (e.g., how they calculated oxygenation index).
• They are not making their data/analytical codes publically available which raises concerns about data reproducibility (internal).

Thus, I am still left with the impression that for a methodological "advancement" claim, it seems to fall short on their design. Previously I recommended the authors to try to decide where their focus should be, methodological advancement or just limit to describe their findings and aknowledge their limitation properly.

Response to general comment: We appreciate the reviewer's positive remarks and insightful issues raised to further strengthen our manuscript. We have carefully studied each of the comments. Please find below our point-by-point responses to the comments.
Regarding the remaining issues:

- We recognize that our initial claim for achieving a better approach of depleting mammalian DNA was a bit overreaching and misleading. We have removed all relevant statements claiming improved performance in host DNA depletion in the manuscript to avoid any ambiguity. The revised manuscript now specifically focuses on elucidating the efficiency of our method in recovering low-biomass microbial DNA after host DNA depletion, enabling metagenome-assembled genome (MAG) reconstruction (Line 103-114).

- We have added head-to-head comparisons with 34 samples (with replicates) to demonstrate the efficiency of our method in recovering the low-biomass microbial DNA after depleting host DNA.

- We have added additional descriptions about the adjudication of LRTI status, criteria for sample selection in the comparison between pneumonia and non-pneumonia cohorts, how we calculated the oxygenation index, and specific details about the negative controls (as detailed in the Responses to Comments 7, 11, and 14) in this revised manuscript.

- We have made our data/analytical codes publicly available (as detailed in the response to Comment 15). The GitHub repository link (https://github.com/processbys/LRT_microbiomes_critical_illness) is now provided directly in the "Data Availability" and "Code Availability" sections of this revised manuscript.

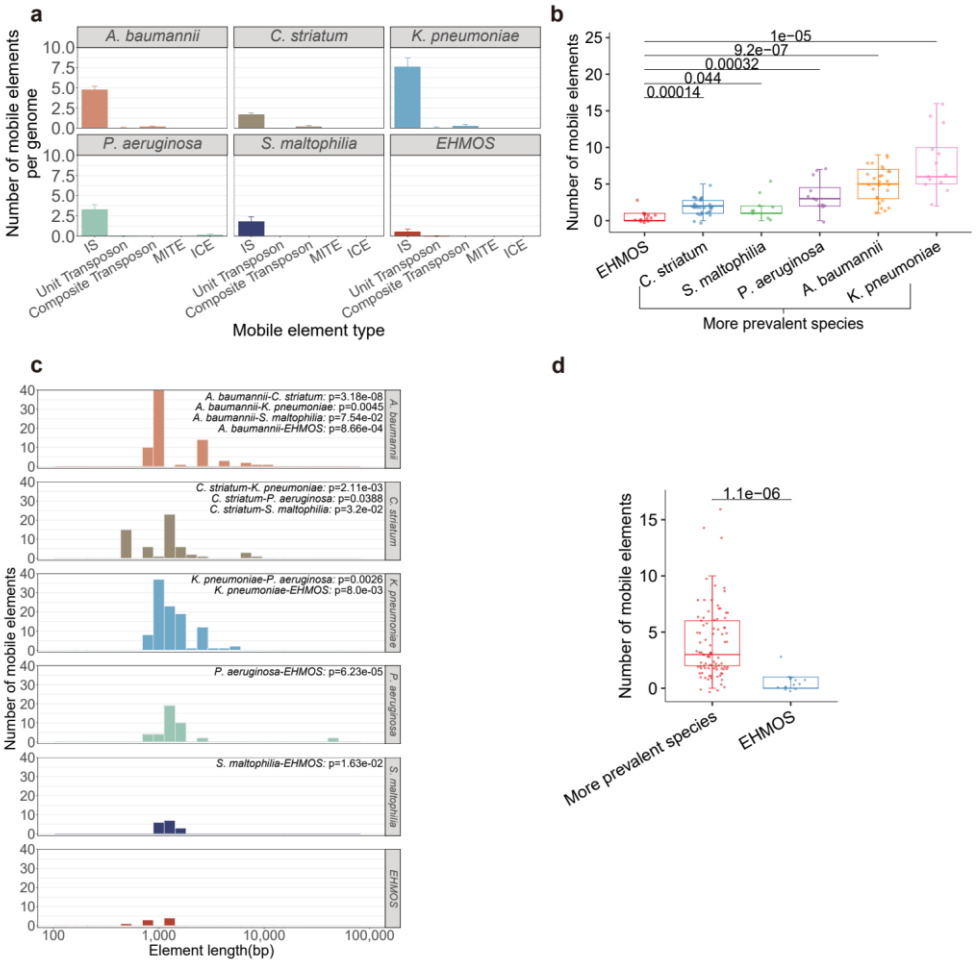
We thank the reviewer's suggestion to define the focus of this study. We would like to clarify that our main focus is on the findings facilitated by the CMEM method. The CMEM is the basis of our described findings, enabling us to achieve the goal of conducting genome-resolved functional and evolutionary analyses. Consequently, we now simply want to introduce the CMEM method as the methodological basis of our study. Additionally, the limitations of CMEM and our study have been thoroughly revised as the reviewer suggested (as detailed in the response to Comment 13).

To be clear, I have tremendous appreciation for the efforts put into this paper: the large sample size and longitudinal data is certainly a strength, there are several analyses that seem well done and several observations are quite interesting specially related to mobile elements and evolution of MAGs. I would like to see this published because I see that it could serve the scientific community and I'm not making any recommendation based on where this journal is submitted (that for the editors to decide based on their priorities). Given that this is the third time I carefully review this paper (believe me I do not want to keep reviewing it) I'm going to take the liberty to give the authors my opinion of how I would deal with the issues raised below. I think regarding the methodological advancement issue, I would make a recommendation despite that I do not know, nor have access to the data or analytical codes: It seems to me that their design has not been prepared to address this and they should just acknowledge this and refraing from claims of superiority compared to other methods that have not been tested. I think the authors decided on this approach hoping to get higher depth of coverage for a MAGs assembly approach, which would make sense. There is no clear well design head-to-head comparison. They should be clear about the methodological analyses and data for the background controls and I can understand that there might be few MAGs assembled though there is probably multiple short reads annotated to microbes in those samples. The justification of going with a MAG approach is clear and would be appreciate it by the analyses perform including the evolutionary and mobile elements. They should limit to describe their findings based on this data and address the possibility of different findings with different methods (not necessarily better or worse). Their limitation paragraph should be extended to include all of these. They should also mention that this data is limited by the complexity of the cohort, multiple varibles that are difficult to capture (e.g, treatment received, comorbidities, etc), type of samples (these are tracheal aspirate and different sample types, including BAL or brushings may lead to different results) among others. I really hope this helps the authors. Below are my point-by-point comments.

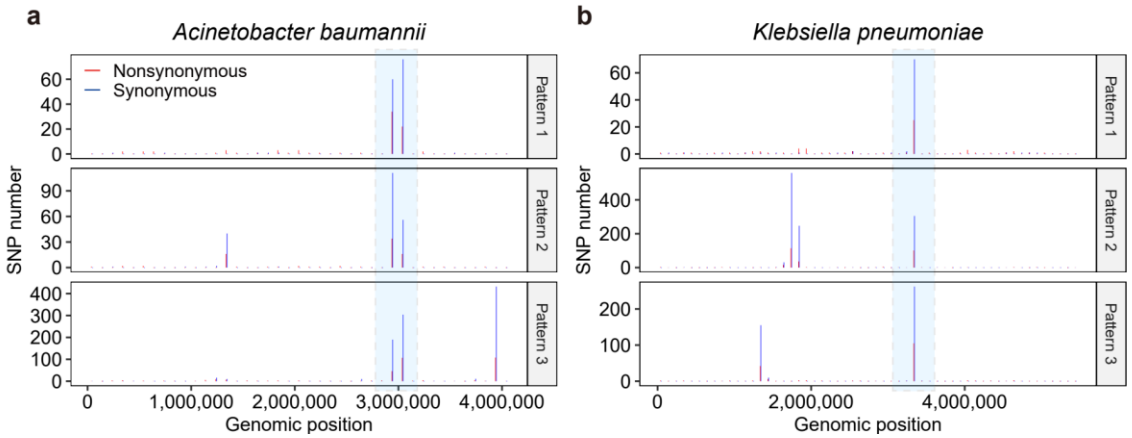
Response to general comment: We thank the reviewer for the positive remarks and invaluable suggestions to improve our manuscript. We have carefully studied each of the suggestions. Regarding the reviewer's insightful recommendations, we have revised the manuscript from the following aspects:

- We have now removed the descriptions claiming improved performance in host DNA depletion. We have added additional head-to-head comparison experiments to reflect the advantage of our approach in recovering low-biomass microbial DNA, refraining from making broad claims about host DNA depletion.

- To further address the concerns regarding the potential influence of negative controls on the mobile elements and evolutionary analyses, we implemented an additional filtering step where sequencing data were first remapped to the MAGs recovered from negative control samples, and all mapped reads were excluded from subsequent analysis. This approach ensures that the analyses are based on high-quality and contaminant-free data. The results related to the mobile elements and evolutionary analyses remain valid (Author Response Fig. 6-7). We have updated the figures and texts accordingly in this revised manuscript.



Author Response Fig. 6 (Manuscript Fig. 4a-c and S5e)**.** Analyses of mobile genetic elements (MGEs) utilizing sequencing data justified by a metagenome-assembled genome (MAG) approach to exclude potential influences from negative controls.



Author Response Fig. 7 (Manuscript Fig. 5a-b)**.** Distributions of synonymous (blue) and nonsynonymous (red) single-nucleotide polymorphisms (SNPs) in coding regions of *Acinetobacter baumannii* (a) and *Klebsiella pneumoniae* (b) with sequencing data justified by a metagenome-assembled genome (MAG) approach to exclude potential influences from negative controls.

- We have removed the potentially overstated descriptions in this revised manuscript (as detailed in the responses to Comments 4-6, and 12). The limitation paragraph in the Discussion section has been thoroughly reformatted and extended to fully describe the limitations of our study, including acknowledging the complexity of the cohort, the potential influence of multiple clinical variables, the sample types used, and the need for additional in vitro validation experiments, as the reviewer suggested (as detailed in the response to Comment 13).

Comment 1: Introduction: "Conventionally, researchers have predominantly relied on culture-based methods for genomic and functional characterization of LRT microbiota due to the low microbial biomass and high level of host contamination". Not sure that I agree with this. I think there is clear role for culture-based method to identify the presence of certain bacteria but I do not think that anyone thinks that it can be used to profile the LRT microbiota. Then they state: "The culture-based isolation of these opportunistic pathogen species has provided profound insights into the mechanisms of pathogenesis, antibiotic resistance profiles, and virulence factors". Which "opportunistic pathogen species"? Where does this come from? It seems disconnected.

Response 1: Thank you for this comment. We agree with the reviewer that culture-based isolation cannot be used to comprehensively profile the LRT microbiota, as it is limited to recovering selected microbial species, with useful diagnostic applications nonetheless. We have modified the description to clarify and accurately emphasize the role of culturing isolates in detecting certain microbial species.

Introduction (Line 71-73): "*Researchers have traditionally relied on culture-based methods for diagnosis, whole-genome information, and functional characterization of specific LRT microbial species.*"

As the reviewer suggested, we have removed the previously included disconnected sentence, "*The culture-based isolation of these opportunistic pathogen species has provided profound insights into the mechanisms of pathogenesis, antibiotic resistance profiles, and virulence factors.*" from this revised manuscript.

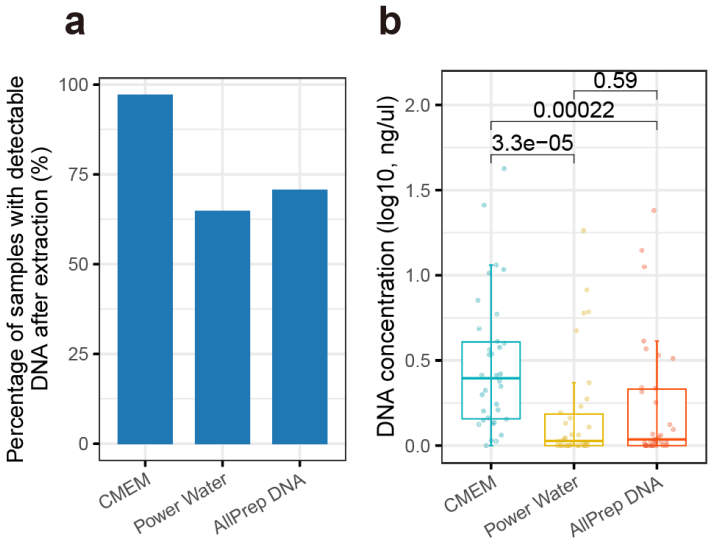
Comment 2: Putting together what was done is quite difficult. For Sup Fig S1, they compared, as per legend: "'NBT+PowerWater' refers to a combination of the previously reported method for host DNA depletion, followed by Qiagen PowerWater for the extraction of the remaining total DNA; (2) 'CMEM' represents our newly developed chelex100-based low-biomass microbial-enrichment method. See the Methods for detailed information.". Then I go to the methods and I get that: "We first attempted reported host DNA removal methods for 26 of our samples. However, the low microbial biomass resulted in undetectable DNA concentrations (measured by Qubit) for approximately half of the samples, making it impractical for metagenomic sequencing."

Comment 2.1: How were those 26 samples selected? Fig S1 does not show a head to head comparison between the two methods. I would expect to see 26 samples in both Fig S1a and Fig S1b.

Response 2.1: We agree with the reviewer's concern and appreciate the opportunity to address this important point. We acknowledge that our previous claim of the enhanced efficiency of CMEM in recovering low-biomass microbial DNA was inappropriate without direct head-to-head comparisons. For the 26 samples included for evaluating the reported host DNA removal methods in our initial manuscript, they were not selected based on specific criteria. These samples were collected during the initial stage of this project, and we utilized them to assess the host DNA depletion method to get higher coverage of microbial data for our lower respiratory microbiome research. We have now removed them from the manuscript to avoid confusion.

Instead, we have now performed head-to-head comparisons on 34 samples (with replicates) to properly evaluate the efficiency of CMEM relative to other methods for recovering low-biomass microbial DNA. Specifically, we compared the efficiency of our CMEM method against the Qiagen Power Water and Allprep DNA/RNA kits, which are commonly used in microbiome research studies[4,5]. Each of the 34 samples was divided into three aliquots and first underwent the same saponin-based host DNA depletion process, followed by DNA extraction using Chelex-100 (CMEM), Qiagen Power Water, and Allprep DNA/RNA kit, respectively. This head-to-head comparison of CMEM against widely adopted commercial kits demonstrates the efficiency of CMEM in enhancing the final DNA yield from low microbial biomass samples after removing human DNA (Author Response Fig. 8). We have included the

new results from the direct head-to-head comparisons (n=34) in the updated Figures S1a and S1b.



Author Response Fig. 8 (Manuscript Fig. S1a,b). Assessment of the efficiency of CMEM against Qiagen Power Water, and Allprep DNA/RNA kit in recovering low-biomass microbial DNA following host DNA depletion. For the 34 samples, each was divided into three aliquots and first underwent the same saponin-based host DNA depletion process, followed by DNA extraction using Chelex-100 (CMEM), Qiagen Power Water, and Allprep DNA/RNA kit, respectively. (a) Bar plot showing the fraction of samples that yielded detectable DNA (detectable DNA concentrations as measured by Qubit) using the three methods. (b) Box plot showing the final yield of DNA concentration for samples processed using the three listed methods. Significance was determined using the Wilcoxon rank sum test (Mann–Whitney U test).

Comment 2.2: What threshold was used to define successful DNA extraction? For attempted reported host methods use for their 'NBT+PowerWater' they cite the following paper: "Charalampous, T. et al. Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. Nat Biotechnol 37, 783–792 (2019)". Why was this selected as the comparison? That paper also uses a completely different technology for sequencing. Why was it decided that it was a good standard to compare to? The comparisons need to be clearly stated and the reader should not work as hard to figure out what the comparisons are.

Response 2.2: We thank the reviewer for bringing this point to our attention. Regarding the threshold used to define successful DNA extraction in the initial submission, we used the following criteria: Samples with a DNA concentration measured as 'too low' by the Qubit assay were regarded as unsuccessful DNA recovery. Conversely, samples with a detectable DNA concentration, regardless of the specific value, were defined as successful DNA extractions. We recognize that the use of the term "unsuccessful DNA extraction" may have been misleading and confusing. We have modified the statement to " *This substantially improved detectable microbial DNA recovery (as measured by Qubit) and DNA yield*

(Supplementary Fig. 1a,b). " (Line 94-95) in this revised manuscript. Additionally, we have changed the y-axis label in Figure S1a to "Percentage of samples with detectable DNA after extraction (%)" to accurately reflect the data presented.

As noted by the reviewer, the paper by Charalampous et al. (Nat Biotechnol 2019) used a completely different sequencing technology with the primary purpose of clinical diagnostics. We acknowledge that this paper was an inappropriate comparison. We referred to this paper in our initial manuscript to demonstrate that many of the ETA samples failed to yield sufficient DNA after the specific saponin-based host DNA depletion described in the paper. We removed the comparison from the manuscript. Instead, we have now added the head-to-head comparisons to prove the efficiency of CMEM in recovering low-biomass microbial DNA, as the reviewer suggested (as detailed in the response to Comment 2.1). We have also provided additional descriptions for the newly included head-to-head comparisons in the main text and figure legends to demonstrate how the comparisons were designed and implemented.

Results (Line 103-106): *"Head-to-head comparisons with Qiagen Power Water and Qiagen Allprep DNA/RNA kits demonstrated that CMEM produced significantly higher final DNA yield and markedly increased detectable DNA recovery (as measured by Qubit, Supplementary Fig. 1a,b)."*

Figure Legend (Fig. S1a-b): "*Assessment of the efficiency of CMEM against Qiagen Power Water, and Allprep DNA/RNA kit in recovering low-biomass microbial DNA following host DNA depletion. For the 34 samples, each was divided into three aliquots and first underwent the same saponin-based host DNA depletion process, followed by DNA extraction using Chelex-100 (CMEM), Qiagen Power Water, and Allprep DNA/RNA kit, respectively."*

Comment 3: In reference to the above, in line 119: "To assess the impact of our method on the LRT microbial community, we collected 32 endotracheal aspirate (ETA) samples and divided each sample into two aliquots: one assigned to the control group and the other to the treatment group.". It is still not clear what the "other treatment group" is. Also, remove the "to the" in "the other to the treatment group"

Response 3: Thank you for this comment. We have modified the sentence and removed "to the" accordingly.

Results (Line 115-118): "*To assess the impact of our method on the LRT microbial community, we collected 32 endotracheal aspirate (ETA) samples and divided each sample into two aliquots: one assigned to the treatment group (processed with CMEM) and the other control group (processed with the same steps but without the host depletion).*"

Comment 4: Line 108: "CMEM removed up to 97% of human reads, leading to a maximum of 97% microbial reads in the sequencing results (median 13%, interquartile range 10%-16%; Supplementary Fig. 1c)". While from the strictly data

stand point is accurate the emphasis is on the outliers with their statement "97% of human reads, leading to a maximum of 97% microbial reads". Also, where is the head to head comparisons with other method to show that CMEM removed up to 97% of human reads? Without a comparison, it could be possible that in a sample not treated with a CMEM approach all it were there would be microbial DNA, especially in these intubated critically ill patients.

Response 4: We thank the reviewer for bringing this issue to our attention. We apologize for the previous overstatement regarding the ability of CMEM to remove human reads. We acknowledge that describing the extent of human reads depletion without providing comparative data from head-to-head evaluations against other methods would be an unsubstantiated claim. Additionally, we did not intend to imply that our saponin-based host DNA depletion method is superior to other existing approaches for removing host DNA. We now refrain from making any specific claims about the efficiency of host depletion (as detailed in the response to General Comment). We have modified the sentence to simply describe the statistics observed in our sequenced samples.

Results (Line 106-108): "*For the samples treated with CMEM, an average of 14.8% microbial reads were yielded in the sequencing results (median 12.1%, interquartile range 9.7%-16.4%; Supplementary Fig. 1c).* "

Comment 5: Line 121: "The CMEM introduced small variances in the relative abundance of individual taxa between some pairs, but we observed no systematic biases between the two groups ($p > 0.05$ in Supplementary Fig. 1d and Supplementary Table 2)" None of the shown data is sufficient to test this. Supp fig 1d is a beta diversity plot and cannot evaluate for individual taxa and the lack of difference in PERMANOVA (I assume that's what the p value is based on) is a high bar for this. Then, Supplementary Table 2 is not a differential enrichment analyse but a table with I think relative abudnace expressed as % (though numbers do not necessary add up to 100% in all samples). Also, in that table there are only 90 taxa. Is that all there is using this metagenome approach? This is clearly much less taxa than a lot of metagenome studies using respiratory samples (see PMID 38168095, 35544065, 37009900 and 34465900), which are mainly based on short reads. Finally, what is the control? Is it the NBT+PowerWater method that is also not clear as mentioned above. I recommend to make this easy to find for readers and reviewers.

Response 5: We thank the reviewer for this insightful comment. We did not want to claim that the CMEM method has no impact on individual taxa, as all host-depletion methods do. We have modified the statement to reflect the reviewer's suggestions.

Results (Line 118-122): "*The CMEM introduced noticeable variances in the relative abundance of individual taxa, and we observed no significant difference in the overall beta diversity between the two groups (PERMANOVA, $p > 0.05$ in Supplementary Fig. 1d and Supplementary Table 2).*"

Regarding that only 90 taxa were listed in Supplementary Table 2, this table specifically compares the microbial taxa identified **in the 32 ETA sample replicates used for the head-to-head CMEM assessment**, not all ETA samples in the study, which is presented in Supplementary Table 6. We agree with the reviewer that this focused comparison of 32 replicates limits the observed taxa compared to full taxa identified in other metagenome studies using respiratory samples (PMID 38168095, 35544065, 37009900 and 34465900). Additionally, differences in sample types, taxonomic identification pipelines, and sequencing technologies between the referenced papers and our study may also contribute to variations in the number of identified microbial taxa.

To clarify, the treatment group refers to samples processed with CMEM, and the control group refers to samples extracted using the same steps but without the host depletion process. Both groups underwent the same metagenomic next-generation sequencing (mNGS) approach. We have added the description for the treatment and control samples in the title of the updated Supplementary Table 2.

"*Table S2. Comparison of the relative abundance of microbial taxa identified in 32 samples (with replicates) from treatment and control groups (treatment group: samples processed with CMEM; control group: samples processed with the same steps but without the host depletion).*"

Furthermore, to ensure clarity regarding the difference of samples in treatment and control groups, we have modified the relevant descriptions in the main text as mentioned in the response to Comment 3. We have added statements defining the treatment (samples processed with CMEM) and control (samples processed with the same steps but without the host depletion) groups in the corresponding figure legends (the updated Supplementary Fig 1d-f).

Comment 6: Line 124: "the more abundant taxa were impacted less, underscoring the accuracy of CMEM in identifying opportunistic pathogen species in LRT samples (Supplementary Table 2)". What's the criteria used to define "opportunistic pathogen specie"? The fact that a taxa is found in culture independent data does not make it opportunistic nor pathogen. Also, what analyses was used to support that "the more abundant taxa were impacted less"

Response 6: We thank the reviewer for pointing out this inappropriate statement. We intend to convey that the most abundant microbial species (not necessarily opportunistic) maintained their relative dominance. And this should be better discussed elsewhere. Following the reviewer's suggestion in Comment 13, we have removed the statement from the main text and instead discussed this point in the Discussion section.

Discussion (Line 590-592): "*The host-removal step also led to noticeable changes in the relative abundance of microbial taxa. While the most abundant species in the samples maintained their dominance, measurements of very low-abundance taxa may be inaccurate.*"

Comment 7: Line 129: “Consistent with previous findings, our results revealed that the microbiome alpha diversity in samples from pneumonia patients was significantly lower than those without pneumonia (Supplementary Fig. 1g)” I would expect some description of the cohort and sample selection to conduct this analyses.

Response 7: We thank the reviewer for this insightful comment. We have reorganized the order of the description to improve the clarity and flow of the narrative. The analysis regarding alpha diversity in pneumonia and non-pneumonia patients has been relocated to Results section 2, after the introduction of our study cohort (Line 150-156). Additionally, we have added additional details for the adjudication of pneumonia and the criteria for sample selection in the comparison between pneumonia and non-pneumonia cohorts.

Method (Line 634-643): "*The status of LRTI was adjudicated by the clinicians referenced on the previously reported criteria[24,50], classifying the patients into (1) Diagnosed pneumonia, defined as clinically documented pneumonia in the medical record system, with clinical positive microbiological culture results or computed tomography (CT) images (2) Non-pneumonia, defined as no clinically documented pneumonia in the medical records and the lack of microbiological evidence of respiratory infection. (3) Undefined LRTI status, assigned to patients without clinically relevant records to determine LRTI status. To mitigate the potential influence of antibiotic usage, all patients included in the comparisons between the pneumonia and non-pneumonia groups were treated with antibiotics during the ICU stay*."

Comment 8: Line 134: “To further validate our approach, we observed a 92.2% consistency between microbial species detected via clinical culture and CMEM for samples with available clinical microbiology data (Supplementary Table 5)”. If the data annotated on the metagenome approach involves only ~90 taxa as depicted in Supplementary table 2, then it is not possible that the culture data and the metagenome data overlaps that well. For example, Citrobacter koseri is seen in culture data but I cannot see that listed among the taxa in Supplementary table 2. I do see it in supplementary Table 6 but it is not clear why they are getting different taxa (probably due to different samples used but this is not clear) and why so few (probably because of the MAG approach rather than short reads).

Response 8: We appreciate this comment and would like to note that Supplementary Table 2 specifically lists the microbial species identified only in the 32 ETA samples (with replicates) used for the head-to-head assessment of CMEM's influence on microbial taxa abundance. The full repository of microbial taxa identified across our entire study cohort is listed in Supplementary Table 6. Therefore, *Citrobacter koseri*, which was not included in the subset of taxa presented in Supplementary Table 2 for the 32-sample CMEM assessment, can be found among the comprehensive set of taxa reported in Supplementary Table 6. Additionally, the microbial taxa listed in Supplementary Tables 2 and 6 were identified using MetaPhlAn3 with short reads, which used a clade-specific marker-based classification approach.

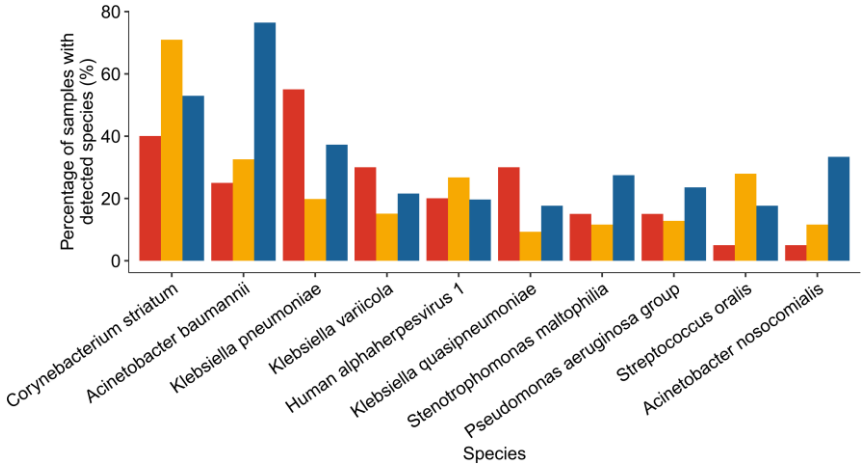
Comment 9: Then in line 160: "De-novo assembly of microbial reads and binning at the individual level resulted in 433 metagenome-assembled genomes (MAGs), of which 120 MAGs were high-quality (Methods; Fig. 1d)." This should probably go before the sections above that utilizes the 120 MAGs for tables such as Supp table 2, 5 and 6. What is the criteria used to define high-quality MAGs. The way it is presenting is asking the reader, and reviewer to work really hard to figure things out.

Response 9: Thank you for this comment. We apologize for the oversight as we included the detailed information in the Methods previously. We would like to clarify that the species listed in Supplementary Tables 2, 5, and 6 were detected by MetaPhlAn3 using short-reads data, but not MAGs. MetaPhlAn3 is based on the exact alignments of unique clade-specific markers (as detailed in the responses to Comments 8 and 13). Therefore, we didn't move the text as the reviewer suggested. We have now mentioned the criteria for high-quality MAGs in the main text.

Results (Line 157-161): "*De-novo assembly of microbial reads and binning at the individual level resulted in 433 metagenome-assembled genomes (MAGs), of which 120 MAGs were high-quality based on the MIMAG (Minimum Information about a Metagenome-Assembled Genome, completeness > 90% and contamination < 5%) standards[26] (Methods; Fig. 1d).*"

Comment 10: Figure 1C depicts Percentage of Detected samples (%). Can the authors explain what 150% means for C. striatum and all other ones. I know what they did but having it as stacked bars do not make sense. They want to have parallel bars with each site having their own % which should not be > 100%

Response 10: We thank the reviewer for this comment and apologize for the inappropriate format of Figure 1C. We have revised Figure 1C to use parallel bars explicitly indicating the percentage of samples with detected species at each sampling site (Author Response Fig. 9) and updated it accordingly to address this issue.



Author Response Fig. 9 (Manuscript Fig. 1c)**.** Percentage of samples detected with the top 10 prevalent microbial species present in the LRT microbiome of patients from Hospital A (red), Hospital B (yellow), and Hospital C (blue), respectively.

Comment 11: Line 221: "We further compared the oxygenation index (OI) and the percentage of neutrophils between patients with a consistent dominant species and dynamically changing dominant species over time. Our findings showed that patients with changing dominant species exhibited a significantly lower OI (p=0.016)". How was OI calculated.

Response 11: Thank you for this comment. We have added the detail of how OI was calculated in the Methods section and refer to the Methods section in the main text accordingly.

Method (Line 643-645): "*The oxygenation index (OI) was calculated by dividing the partial pressure of arterial oxygen (PaO2) by the fraction of inspired oxygen (FiO2).*"

Results (Line 223-225): "*We further compared the oxygenation index (OI, Methods) and the percentage of neutrophils between patients with a consistent dominant species and dynamically changing dominant species over time.*"

Comment 12: Line 242: "the most predominant ARG types were significantly more abundant in Hospital C than in Hospital B (Fig. 2f and Supplementary Fig. 3b), reflecting the increased antibiotic usage in Hospital C". Where is the data that show that Hospital C has increased antibiotic usage? How was that measured? From supplementary table 1, I see that Hospital A and C had similar Antibiotics at ICU admission. Again, the authors should make sure that every statement is adequately supported with objective data.

Response 12: Thank you for this comment. We apologize for the overstated statement and have removed it from this revised manuscript.

Results (Line 242-246): "*We next evaluated the resistome variation among hospitals. Except for Macrolide-Lincosamide-Streptogramin (MLS), the most predominant ARG types were significantly more abundant in Hospital C than in Hospital B (Fig. 2f and Supplementary Fig. 3b).*"

Comment 13: The limitation paragraph is really short and several of the comments above could be dealt with a more extensive discussion in this paragraph.

Response 13: We appreciate the reviewer's comment. The limitation paragraph has been thoroughly revised to provide extensive discussions of the limitations of our study.

Discussion (Line 588-605): "*Our study has several limitations. Firstly, our approach is limited by its potentially reduced ability to characterize viruses from LRT samples with the host-removal step. The host-removal step also led to noticeable changes in the relative abundance of microbial taxa. While the most abundant species in the samples maintained their dominance, measurements of very low-abundance taxa may be*

inaccurate. The performance of CMEM in handling other types of respiratory specimens, such as sputum or BALF, may differ and warrants further evaluation. Secondly, the complexity of our cohorts, such as the heterogeneity of the health conditions of these critically ill patients, the diverse treatment they received, and differences in commodities may have contributed to additional variances in their respiratory microbiomes. Thirdly, our study only included ETA samples. The utilization of alternative respiratory specimen types, such as bronchoalveolar lavage fluid (BALF) or bronchial brushings, may yield varying findings. Finally, our primary findings were predominantly derived from the metagenomic sequencing data and MAGs, with several findings validated with the cultured isolate approach. However, some results, such as the observed associations between various temporal dynamics patterns of LRT microbiomes and pulmonary functions, require more detailed mechanistic experiments to verify the potential causal relationships."

Comment 14: Negative controls. There are much more samples described as negative controls than in prior versions of the paper that I review in other journals. I appreciate the attempt to improve this. However, in Line 670 they state: "For each hospital, ten negative environmental sampling control samples were collected. Sterile saline solution was collected by aspiration through the sputum aspirator as a negative sampling control. Ten reagent controls (DNA extraction blanks) were collected at the laboratory handling study samples. All thirty negative sampling controls and ten reagent controls were included for library preparation and sequencing.

Comment 14.1: Bacterial species identified in more than two negative controls with relative abundance >0.001 were considered as potential contaminants and excluded from downstream analyses (Supplementary Table 3 and 4)" what "Sterile saline solution" was used, which "sputum aspirator" was used. These have to have specific terms and make information.

Response 14.1: We appreciate the reviewer's insightful comment and the opportunity to clarify this important issue. In the Method section, we have added specific details about the 'Sterile saline solution' and 'sputum aspirator' for each hospital.

Method (Line 707-711): "*For each hospital, ten negative environmental sampling control samples were collected. Sterile saline solution (Hospital A: Kelun Pharmaceutical Co., Ltd, H43020456, China; Hospital B: Otsuka Pharmaceutical Co., Ltd, H12020024, China; Hospital C: Kelun Pharmaceutical Co., Ltd, H43020456, China) was collected by aspiration through the sputum aspirator (Hospital A: Shanghai Shangyi Kangge Medical Instruments Co., Ltd, #05366, China; Hospital B: Pacific Pharmaceutical Co.,Ltd, #G05014, China; Hospital C: Medtec Medical Devices Co., Ltd, #3278675, China) as a negative sampling control.*"

Comment 14.2: Then Supp table 3 and 4 are tables of relative abundance. What is the sequence depth? Are these MAGs.

Response 14.2: The average sequencing depths for the negative environmental sampling controls and reagent controls are listed in the table below (Author Response Table 2).

Negative control	Average sequencing depth (G)
Hospital A	14.29
Hospital B	12.53
Hospital C	11.90
Reagent control	14.99

Author Response Table 2 (Manuscript Table. S15). Average sequencing depth of the environmental sampling controls and reagent controls.

We would like to clarify that the microbial taxa listed in Supplementary Tables 3 and 4 were not identified using MAGs. These taxa were identified using short-read sequencing data with MetaPhlAn3.

Comment 14.3: Why are there so few?

Response 14.3: The contamination species in the negative control samples in microbiome studies can be introduced in three phases: 1. at the time of collection, 2. extraction with reagents, and 3. bioinformatics analysis. We discuss the three phases separately to explain the possible reasons for the few species identified in the negative control samples.

1. At the time of collection

The high sterility standards for the solution and equipment used in the ICU would ensure minimal contamination in this phase. Compared to most environmental setups in microbiome studies, the ICU demands very high sterility standards for clinical purposes.

2. Extraction with reagents

Previous studies have reported the important issue of contamination in the reagents of various commercial DNA extraction kits[6]. For our CMEM method, we only involved one Chelex100 reagent in the DNA extraction process, which is generally used in forensic studies and criminal investigations. This is supported by the observation of a noticeably higher number of microbial species identified in a negative control processed with the Qiagen Power Water Kit, which involves multiple reagents (Author Response Table 3). Furthermore, to minimize the potential contaminants introduced during DNA extraction and library construction, all samples and negative controls were processed within a BSL3-level biological safety cabinet in our study, capable of handling infectious pathogens.

Taxonomy	Relative abundance (%)
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Enterococcaceae|g__Enterococcus|s__Enterococcus_faecalis	56.6923
k__Bacteria|p__Bacteroidetes|c__Chitinophagia|o__Chitinophagales|f__Chitinophagaceae|g__Vibrionimonas|s__Vibrionimonas_magnilacihabitans	12.58011
k__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Pseudomonadales|f__Moraxellaceae|g__Acinetobacter|s__Acinetobacter_lwoffii	6.19998
k__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Pseudomonadales|f__Moraxellaceae|g__Acinetobacter|s__Acinetobacter_johnsonii	5.77566
k__Bacteria|p__Proteobacteria|c__Betaproteobacteria|o__Burkholderiales|f__Burkholderiaceae|g__Paraburkholderia|s__Paraburkholderia_fungorum	4.80612
k__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Enterobacterales|f__Enterobacteriaceae|g__Citrobacter|s__Citrobacter_freundii	4.30264
k__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Pseudomonadales|f__Pseudomonadaceae|g__Pseudomonas|s__Pseudomonas_putida_group	2.08768
k__Bacteria|p__Proteobacteria|c__Betaproteobacteria|o__Burkholderiales|f__Comamonadaceae|g__Delftia|s__Delftia_acidovorans	1.41009
k__Bacteria|p__Proteobacteria|c__Alphaproteobacteria|o__Caulobacterales|f__Caulobacteraceae|g__Brevundimonas|s__Brevundimonas_diminuta	1.3837
k__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Pseudomonadales|f__Moraxellaceae|g__Acinetobacter|s__Acinetobacter_guillouiae	0.92463
k__Bacteria|p__Proteobacteria|c__Alphaproteobacteria|o__Rhizobiales|f__Brucellaceae|g__Ochrobactrum|s__Ochrobactrum_anthropi	0.79179
k__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Pseudomonadales|f__Moraxellaceae|g__Acinetobacter|s__Acinetobacter_radioresistens	0.75902
k__Bacteria|p__Proteobacteria|c__Alphaproteobacteria|o__Rhizobiales|f__Bradyrhizobiaceae|g__Bradyrhizobium|s__Bradyrhizobium_sp_BTAi1	0.55087
k__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Enterobacterales|f__Enterobacteriaceae|g__Escherichia|s__Escherichia_coli	0.39197
k__Bacteria|p__Proteobacteria|c__Alphaproteobacteria|o__Sphingomonadales|f__Sphingomonadaceae|g__Sphingomonas|s__Sphingomonas_melonis	0.34373
k__Bacteria|p__Proteobacteria|c__Betaproteobacteria|o__Burkholderiales|f__Oxalobacteraceae|g__Herbaspirillum|s__Herbaspirillum_sp	0.33623
k__Bacteria|p__Proteobacteria|c__Alphaproteobacteria|o__Sphingomonadales|f__Sphingomonadaceae|g__Sphingomonas|s__Sphingomonas_paucimobilis	0.21826
k__Bacteria|p__Proteobacteria|c__Alphaproteobacteria|o__Caulobacterales|f__Caulobacteraceae|g__Brevundimonas|s__Brevundimonas_naejangsanensis	0.20127
k__Bacteria|p__Proteobacteria|c__Betaproteobacteria|o__Burkholderiales|f__Comamonadaceae|g__Variovorax|s__Variovorax_sp_SCN_67_85	0.19397
k__Bacteria|p__Proteobacteria|c__Betaproteobacteria|o__Burkholderiales|f__Comamonadaceae|g__Comamonas|s__Comamonas_testosteroni	0.02924
k__Bacteria|p__Bacteria_unclassified|c__Bacteria_unclassified|o__Bacteria_unclassified|f__Bacteria_unclassified|g__Vermiphilus|s__Vermiphilus_pyriformis	0.01207
k__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Enterobacterales|f__Yersiniaceae|g__Serratia|s__Serratia_nematodiphila	0.00868

Author Response Table 3. The relative abundance of microbial species identified in the negative control sample processed with Qiagen Power Water Kit.

3. Bioinformatics analysis

We employed MetaPhlAn3, a marker-based profiler, to conduct species-level microbial taxonomic identification across all samples. This tool estimates the relative abundance of microbial taxa in a metagenome using the coverage of clade-specific marker genes. The use of these curated marker genes allows MetaPhlAn3 to achieve highly accurate species detection, quantify taxonomic abundance profiles precisely, and maintain a low rate of false positive identifications[7,8]. Compared to

Kmer-based profilers (e.g., Kraken2), the marker-based identification approach minimizes false positive results that can arise from low-abundance bacteria with highly fragmented DNA in the negative controls.

Comment 14.4: Then, if those taxa were removed from samples, what was the relative abudnace of the total removed data from the biological samples (453 lower airway).

Response 14.4: To provide specific details about the abundance of taxa identified in the negative controls and subsequently removed them from the sample data, we added a new table to explicitly demonstrate the average relative abundance of these negative control-affiliated microbial taxa across our samples, using Metaphlan3 (Author Response Table 4). Furthermore, we have conducted a Bowtie2-based mapping analysis to evaluate the influence of microbial species identified in the negative controls on our sequencing results. Specifically, we have remapped the sequencing data from our samples against the MAGs recovered from the negative controls. The average proportion of mapped reads across our samples is 0.31% (median 0.12%, interquartile range 0.06%-0.48%). Both analyses provide reviewers and readers with a comprehensive understanding of the potential very limited influence of microbial species in the negative controls on our data.

Taxonomy	SamplingSite	Average proportion of contaminants in sample data (%)
k__Bacteria|p__Actinobacteria|c__Actinobacteria|o__Propionibacteriales|f__Propionibacteriaceae|g__Cutibacterium|s__Cutibacterium_acnes	HospitalA	0.04
k__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Pseudomonadales|f__Moraxellaceae|g__Acinetobacter|s__Acinetobacter_johnsonii	HospitalA	0
k__Bacteria|p__Firmicutes|c__Bacilli|o__Bacillales|f__Bacillaceae|g__Bacillus|s__Bacillus_cereus_group	HospitalA	0
k__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Pseudomonadales|f__Moraxellaceae|g__Moraxella|s__Moraxella_osloensis	HospitalA	0
Taxonomy	SamplingSite	Average proportion of contaminants in sample data (%)
k__Bacteria|p__Actinobacteria|c__Actinobacteria|o__Propionibacteriales|f__Propionibacteriaceae|g__Cutibacterium|s__Cutibacterium_acnes	HospitalB	0.044

k__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Pseudomonadales|f__Moraxellaceae|g__Acinetobacter|s__Acinetobacter_johnsonii	HospitalB	0
k__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Pseudomonadales|f__Moraxellaceae|g__Acinetobacter|s__Acinetobacter_pittii	HospitalB	0.007
k__Bacteria|p__Firmicutes|c__Bacilli|o__Bacillales|f__Bacillaceae|g__Bacillus|s__Bacillus_cereus_group	HospitalB	0
k__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Pseudomonadales|f__Moraxellaceae|g__Moraxella|s__Moraxella_osloensis	HospitalB	0
Taxonomy	SamplingSite	Average proportion of contaminants in sample data (%)
k__Bacteria|p__Actinobacteria|c__Actinobacteria|o__Propionibacteriales|f__Propionibacteriaceae|g__Cutibacterium|s__Cutibacterium_acnes	HospitalC	0.01
k__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Pseudomonadales|f__Moraxellaceae|g__Acinetobacter|s__Acinetobacter_johnsonii	HospitalC	0
k__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Pseudomonadales|f__Moraxellaceae|g__Acinetobacter|s__Acinetobacter_pittii	HospitalC	0
k__Bacteria|p__Firmicutes|c__Bacilli|o__Bacillales|f__Bacillaceae|g__Bacillus|s__Bacillus_cereus_group	HospitalC	0
k__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Pseudomonadales|f__Moraxellaceae|g__Moraxella|s__Moraxella_osloensis	HospitalC	0
k__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Pseudomonadales|f__Moraxellaceae|g__Acinetobacter|s__Acinetobacter_schindleri	HospitalC	0

Author Response Table 4 (Manuscript Table. S14)**.** The average proportion of contaminants identified in the sample data (ETA samples).

Comment 15: Data availability: in line 799: "Correspondence and requests should be addressed to C.J. (jiang_chao@zju.edu.cn). All human-removed metagenomic sequencing data analyzed in this study are available in the European Nucleotide

Archive (https://www.ebi.ac.uk/ena/browser/home) under the project accession number PRJEB64676." Availability upon request is not acceptable. Downstream data and analytical codes should be made available to ensure internal reproducibility.

Response 15: We appreciate the reviewer's comment regarding code availability. In the revised manuscript, we have addressed this by making our code publicly accessible and providing the GitHub repository link directly in the "Data Availability" and "Code Availability" sections.

Data Availability (Line 852-856): "*All human-removed metagenomic sequencing data and cultured isolate data analyzed in this study are available in the European Nucleotide Archive (https://www.ebi.ac.uk/ena/browser/home) under the project accession number PRJEB64676. The processed data and analytical code are available at: https://github.com/processbys/LRT_microbiomes_critical_illness*"

Code Availability (Line 858-859): "*The analytical code is available at: https://github.com/processbys/LRT_microbiomes_critical_illness*"

References

1. Langelier, C. *et al.* Integrating host response and unbiased microbe detection for lower respiratory tract infection diagnosis in critically ill adults. *Proceedings of the National Academy of Sciences* **115**, E12353–E12362 (2018).
2. Pérez-Cobas, A. E., Ginevra, C., Rusniok, C., Jarraud, S. & Buchrieser, C. The respiratory tract microbiome, the pathogen load, and clinical interventions define severity of bacterial pneumonia. *CR Med* **4**, (2023).
3. Fenn, D. *et al.* Composition and diversity analysis of the lung microbiome in patients with suspected ventilator-associated pneumonia. *Critical Care* **26**, 203 (2022).
4. Li, L. *et al.* Neisseria species as pathobionts in bronchiectasis. *Cell Host & Microbe* **30**, 1311-1327.e8 (2022).
5. Chu, V. T. *et al.* The antibiotic resistance reservoir of the lung microbiome expands with age in a population of critically ill patients. *Nat Commun* **15**, 92 (2024).

6. Salter, S. J. *et al.* Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology* **12**, 87 (2014).
7. Sun, Z. *et al.* Challenges in benchmarking metagenomic profilers. *Nat Methods* **18**, 618–626 (2021).
8. Beghini, F. *et al.* Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *eLife* **10**, e65088 (2021).

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

I would like to thank the authors for their efforts during the revision. I have no more comments.

Reviewer #3 (Remarks to the Author):

I was asked by the editors to specifically deal with the authors responses to reviewer 2 (who was unavailable). The main concern of reviewer 2 was that the authors had claimed they had developed an improved method compared to existing methods to perform microbial metagenomics on LTRI samples. I can see that the authors have removed alot of these claims but there are still claims in the paper to this effect that remain. Specifically:

line 40: "CMEM opens new avenues for genome-resolved analyses" - the authors cannot claim this as they haven't performed a comparison with other metagenomic methods that use host-depletion to show that this indeed is an advance

line 77-78 - "However, applying the MAG approach to LRT microbiota has not yet been feasible" - this is not true. There have been many other methods published for host-depletion of LTRI samples for metagen (e.g https://pubmed.ncbi.nlm.nih.gov/30784601/). They might not have assembled MAGs for the analysis - but this doesn't mean it wasn't feasible.

line 88 - "demonstrating CMEM's advantages in direct deep genome analyses" - advantages over what? The authors haven't compared to other similar approaches

Another concern raised by reviewer 2 was the comparison with two other methods - qiagen power water and qiagen Allprep. Neither of these methods employ host-depletion. The authors provided more information on the DNA extraction metrics and the cutoffs used - but they provide no information on the improvement in microbial enrichment used in the CMEM methods. In line 106 they state that for CMEM an average of 14.6% of reads were microbial. But what about the other two methods? This data is needed to show that there really is an improvement in microbial recovery for the proposed method.

It would also be much more appropriate if they compared their method with other published host-depletion & extraction approaches.

Comment 5 made by reviewer 2 raises the issue of a lack of analysis for individual taxa and a reliance on beta diversity measurements. In my opinion - the authors have failed to deal with this issue in their response - and a much more detailed taxon specific analysis is required.

Upon my own review of the manuscript - I found that the size and breadth of the dataset to be impressive and the longitudinal aspect of sampling to be a particular strength. Some of the analysis has revealed interesting results. But the focus on the novelty of the method lets this manuscript down and I still don't think the authors have completely addressed this issue.

We are very grateful for the opportunity to revise our manuscript and sincerely appreciate the insightful comments and suggestions provided by the reviewers. We have carefully studied and addressed each point raised, which has significantly improved the quality and clarity of our paper. Below, please find our point-by-point response.

Reviewer #1 (Remarks to the Author):

Comment: I would like to thank the authors for their efforts during the revision. I have no more comments.

Response: We greatly appreciate your positive remarks on our efforts in improving the manuscript.

Reviewer #3 (Remarks to the Author):

General Comment (original comment 4): Upon my own review of the manuscript - I found that the size and breadth of the dataset to be impressive and the longitudinal aspect of sampling to be a particular strength. Some of the analysis has revealed interesting results. But the focus on the novelty of the method lets this manuscript down and I still don't think the authors have completely addressed this issue.

Response to general comment: We thank you for the positive remarks on the design and analysis of the manuscript, and we value your insightful suggestions. We would like to clarify that the CMEM method is not the primary focus of our manuscript, as indicated by the title and the emphasis on MAG-analysis-related sections, but it is an important method to enable high-quality sequencing data from the ETA samples. The prominence of the CMEM method in our previous response was due to addressing the specific concerns raised by Reviewer #2 regarding this method. Before we get into specific comments, we want to provide some background on why we decided to introduce CMEM in this manuscript.

At the beginning of the project, we tested several commercial and established host DNA depletion-extraction protocols. However, these methods either resulted in unsatisfactorily low DNA yields or did not sufficiently enrich microbial reads in the sequencing results. Of course, in theory, extensive sequencing of any library could yield sufficient microbial reads, but the associated costs are prohibitive for large sample sizes.

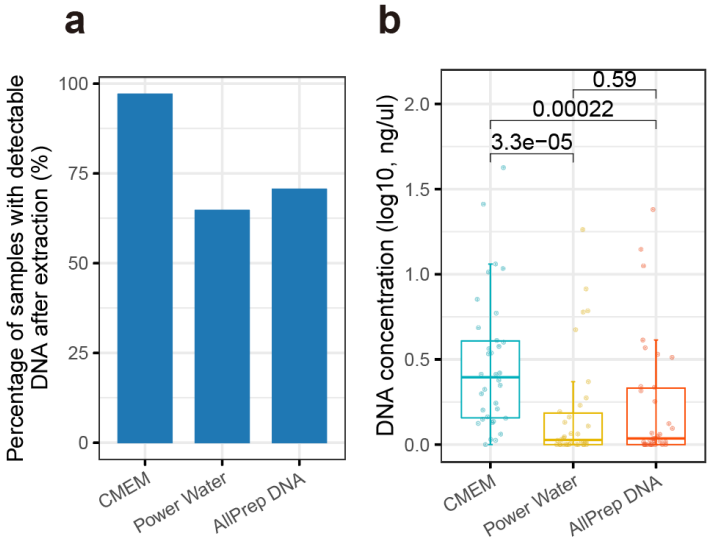
We explored different types of host-depletion protocols. After trying the saponin-based host DNA depletion method, we found that this approach produced better results, which was also confirmed by the head-to-head comparisons (Author Response Fig. 1). However, this method often resulted in undetectable final DNA yields. Specifically, as mentioned in the Methods section, we initially applied the saponin-based host DNA removal method to 26 of our samples, and approximately

half of the samples yielded undetectable DNA concentrations (as measured by Qubit).

This challenge led us to explore alternative extraction protocols, leading to the adoption of the Chelex100 method. Commonly used in forensic science and low-biomass scenarios, Chelex100 enhanced our final DNA recovery significantly, as shown in Figure S1a,b. Therefore, we integrated the saponin-based host DNA depletion step with the Chelex100 extraction to construct the CMEM pipeline. **We do not want to claim originality for the saponin-based depletion technique; however, integrating it with Chelex100 extraction represents a substantial adaptation for our research context.** This integration has notably increased final DNA yields, allowing the successful inclusion of most clinical LRT samples in our study, reducing the number of PCR cycles needed during library construction, improving the library quality, and ultimately saving on sequencing costs due to higher proportions of microbial reads.

To recap, the CMEM is primarily composed of a host-depletion step and a Chelex100-based extraction step. **While we take no credit for the host-depletion step, we do consider the integration of the Chelex100 into this pipeline, at least in the context of human respiratory microbiome research, an important contribution to the field.** Through the revisions, we have performed three sets of head-to-head comparisons on the CMEM method, using more than 90 ETA samples with replicates.

However, we do acknowledge that there are remaining overstated claims that were inappropriate, and we thank the reviewer for bringing this to our attention. We have modified all the corresponding descriptions of the CMEM method accordingly. These changes are visible in the tracked version of the revised manuscript.



Manuscript Fig. S1a-b. Assessment of the efficiency of CMEM against Qiagen Power Water and Allprep DNA/RNA kits in recovering low-biomass microbial DNA following host DNA depletion. For the 34 samples, each was divided into three aliquots and first underwent the same saponin-

based host DNA depletion process, followed by DNA extraction using Chelex-100 (CMEM), Qiagen Power Water, and Allprep DNA/RNA kit, respectively. (a) Bar plot showing the fraction of samples that yielded detectable DNA (detectable DNA concentrations as measured by Qubit) using the three methods. (b) Box plot showing the final yield of DNA concentration for samples processed using the three listed methods. Significance was determined using the Wilcoxon rank sum test (Mann–Whitney U test).

Comment 1: I was asked by the editors to specifically deal with the authors responses to reviewer 2 (who was unavailable). The main concern of reviewer 2 was that the authors had claimed they had developed an improved method compared to existing methods to perform microbial metagenomics on LTRI samples. I can see that the authors have removed alot of these claims but there are still claims in the paper to this effect that remain. Specifically:

Response 1: We deeply appreciate your willingness to address the comments in the absence of Reviewer 2. We thank the reviewer for bringing the remaining inappropriate claims to our attention. We have addressed these remaining claims as detailed in the responses to Comments 1.1-1.3.

Comment 1.1: line 40: "CMEM opens new avenues for genome-resolved analyses" - the authors cannot claim this as they haven't performed a comparison with other metagenomic methods that use host-depletion to show that this indeed is an advance

Response 1.1: Thank you for your constructive feedback. We have updated the description to more accurately depict the role of CMEM in our research. The revised sentence now emphasizes its effectiveness as a robust tool for genome-resolved analyses.

Abstract (Line 40-41): "*Moreover, CMEM can serve as a robust tool for genome-resolved analyses without laborious culturing.*"

Comment 1.2: line 77-78 - "However, applying the MAG approach to LRT microbiota has not yet been feasible" - this is not true. There have been many other methods published for host-depletion of LTRI samples for metagen (e.g https://pubmed.ncbi.nlm.nih.gov/30784601/). They might not have assembled MAGs for the analysis - but this doesn't mean it wasn't feasible.

Response 1.2: Thank you for this comment. We apologize for the inappropriate claims. We have modified the description accordingly to more accurately reflect the current state of research in this area.

Introduction (Line 75-76): "*However, the MAG-based studies of the LRT microbiota were limited due to the aforementioned difficulties.*"

Comment 1.3: line 88 - "demonstrating CMEM's advantages in direct deep genome analyses" - advantages over what? The authors haven't compared to other similar approaches

Response 1.3: We thank you for pointing out this inappropriate statement. We have removed the statement accordingly in this revised manuscript.

Introduction (Line 84-86): "*Further analyses of MAGs revealed the genome-resolved functional, evolutionary, and transmission landscapes in the ICU LRT microbiome.*"

Comment 2: Another concern raised by reviewer 2 was the comparison with two other methods - qiagen power water and qiagen Allprep. Neither of these methods employ host-depletion. The authors provided more information on the DNA extraction metrics and the cutoffs used - but they provide no information on the improvement in microbial enrichment used in the CMEM methods.

Response 2: Thank you for your comment and we appreciate the opportunity to clarify our procedures and intentions. **We would like to clarify that all samples, including those processed by the Qiagen Power Water and Qiagen Allprep DNA kits, were initially subjected to the same saponin-based host DNA depletion process.** The purpose of our comparisons was to illustrate the superior efficiency of the CMEM method in recovering DNA from low-biomass samples, which addresses the issue of undetectable DNA yields following host DNA depletion. Additionally, we have conducted further comparisons with other host DNA depletion methods to demonstrate the efficiency of CMEM in yielding microbial reads data, as detailed in response to Comment 2.1.

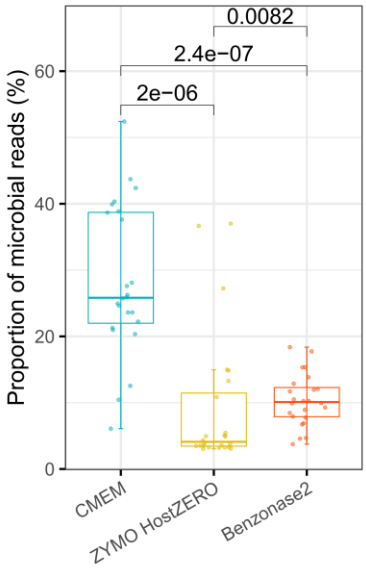
Comment 2.1: In line 106 they state that for CMEM an average of 14.6% of reads were microbial. But what about the other two methods? This data is needed to show that there really is an improvement in microbial recovery for the proposed method. It would also be much more appropriate if they compared their method with other published host-depletion & extraction approaches.

Response 2.1: We appreciate your suggestion to demonstrate the efficiency of the CMEM method in recovering microbial reads data through comparisons with other host DNA depletion approaches. Regarding the statement that CMEM yielded an average of 14.6% microbial reads, we would like to clarify that this number represents the average percentage of microbial reads across all sequenced endotracheal aspirate (ETA) samples (n=442) in this study rather than the testing samples used in the head-to-head comparisons. We apologize for any confusion caused by our previous wording. We have moved this statement to the next result section to better reflect this fact.

Results (Line 130-132): *"An average of 14.8% microbial reads were obtained in the sequencing data (median 12.1%, interquartile range 9.7%-16.4%; Supplementary Fig. 1g)."*

Additionally, following the reviewer's suggestion, we have conducted additional head-to-head comparisons with other host DNA depletion methods, utilizing the ZYMO HostZERO Microbial DNA Kit[1] and the published Benzonase2 method[2]. Our analysis revealed that the CMEM method recovered a significantly higher proportion of microbial reads from ETA samples (N =24; Author Response Figure

1). Furthermore, in response to comments from Reviewer 2 during the 1st revision, we recognize that the saponin-based host DNA depletion approach used in the CMEM approach is not our original contribution. We have updated the manuscript accordingly to avoid inappropriate statements regarding the performance of host DNA depletion.

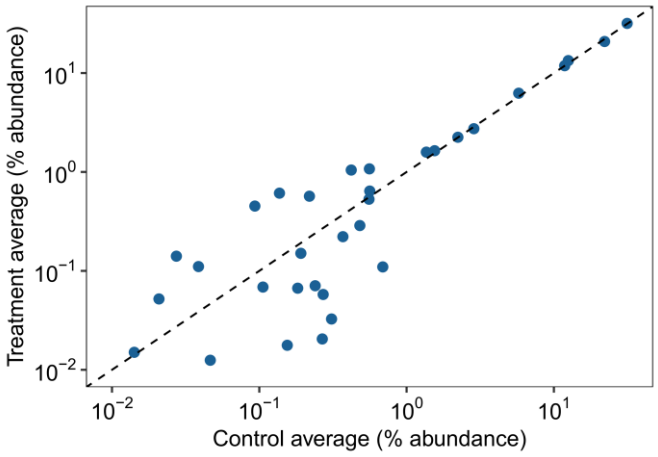


Author Response Fig. 1. Assessment of the efficiency of CMEM in recovering microbial reads data from ETA samples (n=24). The 24 ETA samples were divided into three aliquots and processed with CMEM, ZYMO HostZERO Microbial DNA Kit, and the published Benzonase2 method, respectively. Box plot showing the percentage of microbial reads data for samples processed using the three listed methods. Significance was determined using the Wilcoxon rank sum test (Mann–Whitney U test).

Comment 3: Comment 5 made by reviewer 2 raises the issue of a lack of analysis for individual taxa and a reliance on beta diversity measurements. In my opinion - the authors have failed to deal with this issue in their response - and a much more detailed taxon specific analysis is required.

Response 3: We thank you for this insightful comment. Firstly, we wish to clarify that we did not claim that the host DNA depletion step in our protocol, or any host-depletion process in general, does not affect the relative abundance of individual taxa. To offer a more comprehensive understanding of how host DNA depletion might influence the relative abundance of individual taxa, we conducted additional taxon-specific analyses. These analyses demonstrated that the host DNA depletion process does not significantly alter the overall microbiome composition. We compared the results from 32 samples (with replicates) between the treatment and control groups (treatment group: samples processed with CMEM; control group: samples processed identically but without host depletion). Our analysis showed a high correlation in the relative abundance of individual taxa between the two groups (Author Response Fig. 2, Supplementary Table 3; Pearson coefficient of determination, R^2 = 0.74; **with R^2 = 0.98 for taxa above 0.5% average relative**

abundance). We noted that the correlation was less pronounced for the low-abundance taxa (average relative abundance < 0.5%). We have updated the relevant sections in the results and limitations paragraph to highlight the influence of the host DNA depletion process on low-abundance taxa.



Author Response Fig. 2 (Manuscript Fig. s1c)**.** Species-level taxon relative abundances were plotted for treatment and control samples (N=32). Taxa with average relative abundance < 0.01% were filtered. The dashed line shows a 1:1 correlation.

Results (Line 112-116): *"The host-depletion introduced noticeable variances in the relative abundance of individual taxa, and we observed no significant difference in the overall beta diversity between the two groups (PERMANOVA, p > 0.05 in Supplementary Fig. 1 c,d and Supplementary Table 2 and 3)."*

Discussion (Line 583-586): *"The host-removal step also led to noticeable changes in the relative abundance of microbial taxa. While the most abundant species in the samples maintained their dominance, measurements of very low-abundance taxa (average relative abundance < 0.5%) may be inaccurate."*

References

1. Heravi, F. S., Zakrzewski, M., Vickery, K. & Hu, H. Host DNA depletion efficiency of microbiome DNA enrichment methods in infected tissue samples. *J. Microbiol. Methods* **170**, 105856 (2020).

2. Nelson, M. T. *et al.* Human and Extracellular DNA Depletion for Metagenomic Analysis of Complex Clinical Infection Samples Yields Optimized Viable Microbiome Profiles. *Cell Reports* **26**, 2227-2240.e5 (2019).

REVIEWERS' COMMENTS

Reviewer #3 (Remarks to the Author):

Thankyou to the authors for their response to my concerns and the changes they have made. I have no further comments