

Peer Review Information

Journal: Nature Microbiology

Manuscript Title: Microbiome-metabolite linkages drive greenhouse gas dynamics over a permafrost thaw gradient

Corresponding author name(s): Dr Malak Tfaily

Reviewer Comments & Decisions:

Decision Letter, initial version:

Message: 7th February 2024

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

Dear Malak,

I hope you've been well. Thank you for your patience while your manuscript "Microbial-Metabolite Dynamics Illuminate Permafrost Ecosystem Complexity at the Feature Scale" was under peer-review at Nature Microbiology. It has now been seen by 3 referees, whose expertise and comments you will find at the end of this email. Although they find your work of considerable potential interest, they have also raised a number of concerns that will need to be addressed before we can consider publication of the work in Nature Microbiology.

The referees do agree that the work is broadly interesting and that the focus on permafrost through the lens of metabolomics is important. However, Reviewer #3 raised a number of questions and concerns about the metabolomics methods, which will need to be addressed. Additionally, though this reviewer was overall positive (by our interpretation!), Reviewer #2 raised a number of concerns that were more editorial in nature. These concerns boil down to the fact that they felt the manuscript did not yet gel into a cohesive and clear take home message (other than, "it's complicated")--in their comments, they urge you to dig a bit deeper into the interpretation of your results, putting the pieces together into a more cohesive message. We do agree that this would increase the broader appeal of your paper. I do also concede that the concerns of Reviewer #2 are a bit nebulous, and so I want to highlight that I am at your service and happy to discuss editorial changes that you might make to your manuscript to address those concerns.

Overall, should further experimental data allow you to address these criticisms, we would be happy to look at a revised manuscript.

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

We strongly support public availability of data. Please place the data used in your paper into a public data repository, if one exists, or alternatively, present the data as Source Data or Supplementary Information. If data can only be shared on request, please explain why in your Data Availability Statement, and also in the correspondence with your editor. For

2



Open Access This file is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In the cases where the authors are anonymous, such as is the case for the reports of anonymous peer reviewers, author attribution should be to 'Anonymous Referee' followed by a clear attribution to the source work. The images or other third party material in this file are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

some data types, deposition in a public repository is mandatory - more information on our data deposition policies and available repositories can be found at <https://www.nature.com/nature-research/editorial-policies/reporting-standards#availability-of-data>.

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

If revising your manuscript:

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

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

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

When submitting the revised version of your manuscript, please pay close attention to our [href="https://www.nature.com/nature-portfolio/editorial-policies/image-integrity">Digital Image Integrity Guidelines](https://www.nature.com/nature-portfolio/editorial-policies/image-integrity). and to the following points below:

- that unprocessed scans are clearly labelled and match the gels and western blots presented in figures.
- that control panels for gels and western blots are appropriately described as loading on sample processing controls
- all images in the paper are checked for duplication of panels and for splicing of gel lanes.



Finally, please ensure that you retain unprocessed data and metadata files after publication, ideally archiving data in perpetuity, as these may be requested during the peer review and production process or after publication if any issues arise.

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

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

If you wish to submit a suitably revised manuscript we would hope to receive it within 6 months. If you cannot send it within this time, please let us know. We will be happy to consider your revision, even if a similar study has been accepted for publication at Nature Microbiology or published elsewhere (up to a maximum of 6 months).

In the meantime we hope that you find our referees' comments helpful.

Yours sincerely,

Reviewer Expertise:

Referee #1: permafrost microbiology, metagenomics

Referee #2: microbial ecology, permafrost/polar microbiology, metagenomics

Referee #3: metabolomics

Reviewer Comments:

Reviewer #1 (Remarks to the Author):

Zapata et al. delves into the challenging field of connecting microbiome diversity with metabolome diversity and ecosystem processes driven by microbes in permafrost. The site has been studied before and there are previous publications, such as Woodcroft et al and Emerson et al, that explored the role of microbial diversity and interactions with ecosystem processes. The present study attempts to expand previous studies by testing well structured hypotheses using high resolution data (metagenomics and FTICR-MS) and robust methods to explore microbial and metabolite assembly processes and connect both. To me this is the outstanding feature. The methods used and results presented may be of interest to researchers in several different areas. More clarification is needed in some parts and a few limitations of the study also need to be addressed. Below I share specific

3



Open Access This file is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In the cases where the authors are anonymous, such as is the case for the reports of anonymous peer reviewers, author attribution should be to 'Anonymous Referee' followed by a clear attribution to the source work. The images or other third party material in this file are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

comments to provide constructive feedback and highlight areas for improvement.

L33: could remove 'understanding'?

L135: 882 metagenomes collected between 2011-2017, which contained 13,290 [unpublished] MAGs; but the analyses were focused on samples from 2012. Please clarify this part. Why 2012? How many metagenomes comprehend the 2012 dataset? Supp Table 2 has 85 samples.

Also, give more information to the reader about the sampling - e.g., depth and monthly scales (first mentioned in L168) and point to Supp Table 2. Include the SRA IDs of metagenomes used in the Supp Table.

L136: 'in review' - the fasta files of the MAGs analysed in this study need to be available to readers as supplementary data or in a public repository.

L309 - wouldn't it favour metabolite dispersion too?

L476-477: give more details to reader on how the score was obtained. The wording is not much clear.

L486: provide details about CoverM parameters used and briefly write the relative abundance strategy (as in ref. 3).

L488: provide the details of SingleM parameters used and clarify why it was used. i.e., explain the reader where and why relative abundance data was used and where and why rarefied [MAG-derived?] OTU counts data was used.

L669 - mapping of reads against MAGs has been described in a previous section and is being described again [a bit differently] in this section?

Limitations of the study that need to be considered in the Discussion/Conclusion:
- microbial activity was not measured; it is possible that more microbe-metabolite correlations would be captured if metatranscriptomics had been included. This discussion is important for future studies.

- How representative are the studied MAGs in terms of the expected whole prokaryotic community diversity in these habitats? Rare taxa tend to be underrepresented in MAGs. It is important to give the readers evidence about if the MAGs reported are dominant in the system. Outline as well the limitations of missing rare taxa in the analyses.

L Braga

Reviewer #2 (Remarks to the Author):

I hoped I could offer high praise, but regret it must also be qualified quite substantially.

4



Open Access This file is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In the cases where the authors are anonymous, such as is the case for the reports of anonymous peer reviewers, author attribution should be to 'Anonymous Referee' followed by a clear attribution to the source work. The images or other third party material in this file are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

The authors apply concepts from metacommunity ecology to reveal the intricate interactions between vast datasets of metagenomes and metabolomes from Arctic Sweden. This is intellectually adventurous and applied to an experimental system of global significance. There is potential relevance here for microbial ecology, metagenomics, metabolomics, systems biology, Arctic science, terrestrial biogeochemistry, and climate change. However the manuscript as it stands would benefit from substantive revision to bring out these links as relevant to a broader audience. I hope the recommendations below are helpful in refining the paper's appeal to different audiences:

For Arctic science, terrestrial biogeochemistry, and climate change:
The key take-home I have from the paper is that "it's complicated" and the closer we look, the more complicated it becomes. To that end, the paper is better at complicating our understanding of greenhouse gas emissions from permafrost rather than clarifying it. This can be OK, as it is sometimes unavoidable and possibly essential depending on the intrinsic complexity of the story. However it is not an approach to take lightly when seeking to communicate effectively with the broad readership of a journal like *Nature Microbiology*. Foremost I would advise the authors to streamline how they narrate their story: let the feast of data not be at the expense of famine of knowledge. Indeed, with an eye to the consequential biogeochemical processes within the thawing permafrost of the Arctic, the core insight I derive is that the cycling of organosulfur compounds, apparently by *Acidobacteria* (L219-234) is implied as an influence on potential methane emission (L351). The identification of these metabolites and taxa is useful, however the fundamentals of redox chemistry tell us that sulfate reduction (L222) is energetically more favourable than methanogenesis. This is readily understood by biogeochemists as a first-order control on methane release in regions of the Arctic where sulfur is more available. In a revision it would be advisable to make as explicit case as possible for the expensive and intensive analyses of the kind reported here which hint at such processes as opposed to the simpler quantification of these biogeochemical species. How, specifically, do the conjoined analyses of metabolomes and metagenomes advance our ability to forecast greenhouse gas trajectories from thawing permafrost, as claimed (L45-46). It is a bold promise which this impressive manuscript doesn't quite deliver upon in the present form. Moreover, while the team has been highly productive within the Stordalen mire region, what might the results mean, or how might the methodology apply, to the broader question of change within Arctic lands?

For microbial ecology, metagenomics, metabolomics, systems biology:
Turnbaugh and Gordon (DOI:<https://doi.org/10.1016/j.cell.2008.08.025>) invited us to "the marriage of metagenomics and metabolomics" in 2008. This study, at the forward edge of the present state of the art, showcases at least the comfortable co-habitation (or perhaps the sambo of metagenomics and metabolomics, noting the paper's Swedish locus) of both disciplines. Yet it remains the case that it is very difficult to make the leap from sophisticated correlative analyses such as those ably performed by the authors into convincing mechanistic insights. This is not to dispute the excellence of the present manuscript, but a cautionary, conservative note must be offered when considering the promising state of this science. In essence the authors treat pools of metabolites within metabolomes like microbial OTUs within communities. I found this stimulating and innovative. Having dabbled with both methodologies in similar settings I greatly enjoyed seeing the authors' approach. Their concepts and methodology appear portable to a diverse range of systems, and has the potential to underpin advances in other fields, which would



make this work of broad relevance, but requires further thought to crack the challenges of elucidating causal relationships and mechanistic insights. I would advise the authors to reflect on this and perhaps see how far their methodological framework can support this kind of progress, or at the least provide some caveats within the discussion as to its limitations. The dangers of large scale application of correlational analyses at pace are particularly apparent when the authors disclose that two Actinobacteria genera (L261) correlate both with carbon dioxide and methane concentration, and in all three actinobacterial and one acidobacterial lineage correlate with methane. The significance and magnitude of the positive correlations of each is not reported (it should be; Sup Table 6). Is it not surprising to the authors that methanogenic Archaea do not show correlations? If not why not? Moreover, given that the authors have access to the MAGs of these correlating lineages and the tools with which to explore their functional potential, what might be the biological mechanism by which two different bacterial phyla appear to contribute to methanogenesis within this system? In a revised manuscript it would be helpful if the authors devoted attention to properly exploring this and other conclusions of this kind. Some biology might be lurking, or at least validation of the methodology.

For science generally:

On a related note, while it is pleasing that the authors seek to frame their study's goals so prominently within the hypothetico-deductive framework (L94-L130) through a series of explicit hypotheses, the hypotheses posed, while valuable, do require refinement. Most of all, their presentation needs further thought. In places the hypotheses are not declarative but probabilistic (e.g. L102 ...conditions likely contribute...) and in many cases the predictions are implicit or at least qualitative. Ensuring a hypothesis has an explicit statement, defined methodology and testable prediction is always a powerful way of making sure the hypothesis is apt. Meanwhile, since this paper reports the outcomes of work started over a decade ago, it is best to use the past tense throughout the tracts of text dealing with hypotheses.

It should be noted that the manuscript depends heavily on data from another manuscript which is currently under review (McGivern et al. L136). Given the brief turnaround for reviewing in this round it hasn't been feasible to ask for it to be provided. In a revision, unless this other paper is publicly available it would be helpful to include as supplemental data for review purposes only. At this stage I am taking for granted that the McGivern manuscript does not materially affect the interpretation of this manuscript or its significance. Asynchrony (L66) vs microscale synchrony (L352) vs micro/fine-scale synchrony (L359): I take asynchrony to mean events happening with differing temporal patterns and as the antonym of contemporaneous sequences of events. Resolving the relevant scales of variation is important: The authors concede their temporal and spatial scales of are only revealed at monthly timescales in some instances (L168) – this is still quite a synoptic timescale! Given the transience of some metabolic interactions, can the authors justify this? Capturing time-lagged events remains problematic within this methodology. No methodology is perfect, but if the maximal temporal resolution is month-by-month, the authors need to be cautious. I really liked Table 1 which made the definition of terms and their cognate metrics explicit. It would be helpful if the authors could be similarly explicit about their thresholds for asynchrony, synchrony, and micro/fine-scale synchrony, and whether microsynchrony or fine-scale synchrony are distinctive.

Minor comments:



L186: please provide a reference for "previous work"

L206: hydrocarbon like

Table 1: is "more" missing from (ecological similar communities than expected by chance)

Reviewer #3 (Remarks to the Author):

Thank you for giving me an opportunity to review this work and thank you for your patience with the review.

Overall I think this is a well-written and carefully performed study (as one would expect for something submitted to Nature Microbiology). I think the work is an interesting approach to soil metabolomics and microbial ecology albeit similar to some recent previous studies that are not cited (see minor comments section). I have some major and some minor comments as set out below.

I am not a microbiologist or an ecologist, so I have focussed mainly on the metabolomics in the paper.

Major comments

This could perhaps be missing something, but I am a little concerned with the extraction method used for metabolites in this study as it is not a standard metabolomic method. As far as I can see only water was used to extract metabolites from the soil, but this potentially means any large molecules like fatty acids, lipids, and many larger secondary metabolites would likely not have been extracted. This in turn means a lot of information may have been missed and the models might be incomplete. It is standard in metabolomics to use an organic solvent as well as water. The almost field standard method is methanol-chloroform-water extraction. Why was this not used here? Have I missed something? If there is a reason, I think it needs to be explained.

A similar question is why were the samples only analyzed in negative ionization mode and not positive mode? Some compounds analyze better using positive ionization so why not use both to get the best coverage? Using only one ionization mode could again mean metabolites were not seen.

Also, can you please add details of the quality control methods used for the metabolite analysis? How did you confirm metabolite ID for example and did you follow the recommended protocols for Standard reporting requirements for biological samples in metabolomics experiments in the biological or environmental context (1)?

Again, perhaps I missed it but I am also not clear on how the raw mass spectral data was processed. A peak or m/z value in a chromatogram does not necessarily equal a metabolite, it could be contamination in the instrument or solvents, etc so out of the 14432 peaks that you found how many were actually metabolites/compounds of interest and how did you get down to 6763 and then to 6526?. Even 6526 seems a much higher number of metabolites than many other studies and if potentially non-relevant data were used to build



models of relationships between metabolites and other factors it could introduce error and uncertainty.

Line 422 to 433. Why were samples removed from the dataset just because they had fewer peaks? I don't feel this is a legitimate reason to exclude samples. They may just have had fewer metabolites in that sample. Please expand on the reasoning here. Later on there is also discussion only including features if they are in >50% of samples. This again seems a bit extreme, perhaps there is an interesting reason some features are in <50% of samples but by not looking at them the chance to investigate is lost?

Minor comments

Line 58, my wish to give a short definition of metabolomics where the term is first used for those that are not familiar with it.

Line 62, give a reference for the statement on metabolites here.

Line 63, I think the definition of soil community metabolome was first given in Jones et al (2) not in the papers cited here.

Authors seem to be missing out on previous work in soil community metabolite and function analysis e.g. (3, 4). These papers have essentially applied the same principles and techniques - examining microbial communities and metabolomes together just applied to slightly different problems. NB this reviewer is not the author of those studies, and I am not saying the papers have to be cited just asking the authors to consider if they are relevant and warrant inclusion.

How did you ensure the push corers used to take samples did not contain microbes from outside the study area? Were they sterilized before use?

Lines 393 to 400. I think websites should be cited as proper references not just pasted into the text.

Line 413, while authors have cited technical replicates what about actual sample replicants, e.g. two or three samples from the same areas being separately extracted and analyzed?

Line 471-472, seems like something is missing here or there is odd spacing.

Line 540, the Bonferroni method is generally though to be too extreme a method in metabolomics in that it might get rid of data that was actually useful and make a type 2 error. Is there a specific reason this method was used over others.

Similar to the above why was a Mann-Whitney U test used over parametric alternatives

Line 602, what do you mean "approach described by 46" this is not correct English. Unless this is a journal specific issue I suggest changing to something like - ...using a previously described approach 46...



Line 605 why the closed brackets after the words daisy and nun? I know it refers to the name of the program/package but are they actually needed? Fine if they are, I am just curious.

Thank you for the comprehensive details of what code was used and where the code and data can be downloaded from. It is nice to see such good practice in data sharing.

References

1. N. Morrison et al., Standard reporting requirements for biological samples in metabolomics experiments: environmental context. *Metabolomics* 3, 203-210 (2007).
2. O. A. H. Jones et al., Metabolomic analysis of soil communities can be used for pollution assessment. *Environmental Toxicology and Chemistry* 33, 61-64 (2014).
3. R. M. Shah et al., Functional analysis of pristine estuarine marine sediments. *Science of The Total Environment* 781, 146526 (2021).
4. R. M. Shah et al., Omics-based ecosurveillance uncovers the influence of estuarine macrophytes on sediment microbial function and metabolic redundancy in a tropical ecosystem. *Science of The Total Environment* 809, 151175 (2022).

Author Rebuttal to Initial comments

Reviewer Expertise:

Referee #1: permafrost microbiology, metagenomics

Referee #2: microbial ecology, permafrost/polar microbiology, metagenomics

Referee #3: metabolomics

Reviewer Comments:

Reviewer #1 (Remarks to the Author):

Freire-Zapata et al. delves into the challenging field of connecting microbiome diversity with metabolome diversity and ecosystem processes driven by microbes in permafrost. The site has been studied before and there are previous publications, such as Woodcroft et al and Emerson

9



Open Access This file is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In the cases where the authors are anonymous, such as is the case for the reports of anonymous peer reviewers, author attribution should be to 'Anonymous Referee' followed by a clear attribution to the source work. The images or other third party material in this file are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

et al, that explored the role of microbial diversity and interactions with ecosystem processes. The present study attempts to expand previous studies by testing well structured hypotheses using high resolution data (metagenomics and FTICR-MS) and robust methods to explore microbial and metabolite assembly processes and connect both. To me this is the outstanding feature. The methods used and results presented may be of interest to researchers in several different areas. More clarification is needed in some parts and a few limitations of the study also need to be addressed. Below I share specific comments to provide constructive feedback and highlight areas for improvement.

Author's response: We thank the reviewer for their insightful and constructive comments on our manuscript. We are also grateful that the reviewer acknowledges our attempt to integrate multi-omics to elucidate the assembly processes connecting microbial and metabolite diversity in permafrost ecosystems. Please find below our detailed responses to the specific comments. We believe that we have addressed all the comments and points that were brought up.

L33: could remove 'understanding'?

Author's response: Thank you for this suggestion, 'understanding' has been removed.

L135: 882 metagenomes collected between 2011-2017, which contained 13,290 [unpublished] MAGs; but the analyses were focused on samples from 2012. Please clarify this part.

Why 2012? How many metagenomes comprehend the 2012 dataset? Supp Table 2 has 85 samples.

Author's response: Thank you for the opportunity to clarify. We focused on 2012 since it provided matching metagenomic and metabolomic data, for 67 samples, and the desired temporal resolution. To maximize interpretation of these data, rather than just using MAGs derived from 2012, we leveraged an existing published MAG database containing 13,290 MAGs that were built from samples collected at the Stordalen Mire from 2010-2017¹. We mapped metagenomic reads from the 67 samples collected in 2012 (Supplementary Table 2, Metagenome_reads_Accesion tab) against the full MAG database¹ and kept only those MAGs that reported any hits. To help clarify these points in the paper, we have now included clarification in the Materials and Methods section to better explain our rationale for focusing on the 2012 integrated dataset. We have also added more information in **L166** noting that of



the 13,290 MAGs in the full database, 1,402 MAGs matched reads from the metagenomes collected in 2012, which are listed in Supplementary Table 2 (MAGs tab).

The new text reads as follows: **L618** "To maximize genome-resolved inferences from these data, we queried the 2012 metagenomes with the site-specific metagenome-assembled genome (MAG) database spanning samples from 2011-2017 of 13,290 MAGs (1,806 when dereplicated to species level; BioProject PRJNA386568, McGivern et al., accepted). Read mapping with CoverM to the dereplicated dataset resulted in an abundance table of 1,402 species-dereplicated MAGs that were present in 2012 samples and were used for subsequent analyses."

Also, give more information to the reader about the sampling – e.g., depth and monthly scales (first mentioned in L168) and point to Supp Table 2.

Author's response: We now have included a description of the sampling strategy in **L147**. Also, information of specific collection dates and depths are now included in Supplementary table 2, Sample metadata and Metagenome reads accession tabs.

The new text reads as follows: **L147** "Our study leveraged paired genome-resolved metagenomics and high-resolution Fourier Transform Ion Cyclotron Resonance (FTICR) mass spectrometry via direct injection to characterize peatland microbial communities and their metabolomes from samples collected across three habitats (palsa, bog and fen) in June, July and August 2012 at different depths (shallow, middle and deep) at the Stordalen Mire (**Supplementary Table 2**)."

Include the SRA IDs of metagenomes used in the Supp Table.

Author's response: SRA IDs of metagenomes used in this study are now included in Supplementary Table 2, Metagenome reads Accession tab.

L136: 'in review' - the fasta files of the MAGs analysed in this study need to be available to readers as supplementary data or in a public repository.

Author's response: The MAGs used in our work were first released in a manuscript in review (McGivern et al.) at the time of submission of our manuscript, but McGivern et al. (2024) paper is now accepted in Nature Microbiology. Also, the metagenome-assembled genomes files (.fna) and SRA IDs that are part of the Stordalen Mire database used in this study are publicly



available at <https://zenodo.org/records/7596016> and published in ENA under BioProject PRJNA386568. We have cited these resources in our manuscript and specified their use in the Materials and Methods and Data Availability sections. However, to further improve clarity, we have now also included this citation in **L153** alongside the citation for the recently accepted McGivern et al. (2024) paper in Nature Microbiology that utilized this same database. Additionally, we have added Supplementary Table 2 MAGs tab, which contains information on the 1,402 specific MAGs that mapped to the 2012 reads analyzed in our study.

The new text reads as follows: **L614** "To maximize genome-resolved inferences from these data, we queried the 2012 metagenomes with the site-specific metagenome-assembled genome (MAG) database spanning samples from 2011-2017 of 13,290 MAGs (1,806 when dereplicated to species level; BioProject PRJNA386568, McGivern et al., accepted). Read mapping with CoverM to the dereplicated dataset resulted in an abundance table of 1,402 species-dereplicated MAGs that were present in 2012 samples and were used for subsequent analyses."

L309 - wouldn't it favour metabolite dispersion too?

Author's response: Yes, we believe that the increase of water availability with thaw might contribute to microbial and metabolite movement and potentially increased interactions. While we focused the paper on microbial-metabolite interactions only, we did indeed observe an increase of microbe-microbe, microbe-metabolite and metabolite-metabolite interactions along with the thaw gradient (from palsa, to bog, to fen) (Supplementary Fig. 4b).

The new text reads as follows: **L338** "Conversely, increased water content in the bog and fen was associated with greater microbial and metabolite dispersal, likely contributing to increased microbe-microbe, microbe-metabolite and metabolite-metabolite interactions (**Supplementary Fig. 4b**)."

L476-477: give more details to reader on how the score was obtained. The wording is not much clear.

Author's response: Thank you for highlighting that more details were needed. We have now specified how the quality score was estimated and used to select a candidate bin set with the highest completion and lowest contamination. The reason why this score was implemented was



to select one candidate bin set per sample because multiple pipelines were used to generate bin sets.

The new section reads as follows: **L591** " Following the second round of ensemble binning, completeness and contamination statistics of the resulting ensemble bins were assessed via CheckM (v1.0.12) ¹ lineage workflow. For each of the three ensemble bin sets, bins with 70% completion and less than 10% contamination were used to calculate a quality score: completeness - (5 x contamination). A candidate bin set for each sample was chosen based on the ensemble binning tool that had yielded the bin set with the highest quality score. The bins in each candidate bin set were then further refined with RefineM (v0.0.24) ² and manually examined in anvi'o (v.5.2) ³."

L486: provide details about CoverM parameters used and briefly write the relative abundance strategy (as in ref. 3).

Author's response: To address this we have included CoverM parameters and the strategy for relative abundance calculation.

The new text reads as follows: **L605** "The relative abundance of each MAG was assessed by read-mapping with CoverM ⁴ (<https://github.com/wwood/CoverM>), using the same relative abundance calculation strategy as in ⁵ (i.e. using CoverM genome option, with the following parameters: --min-read-percent-identity 0.95, --min-read-aligned-percent 0.75, -coupled, -m trimmed_mean (**Supplementary Table 2**). For all statistical analyses, the relative abundance of each species in each sample was calculated by dividing its coverage by the total coverage of all species in the dereplicated set."

L488: provide the details of SingleM parameters used and clarify why it was used. i.e., explain the reader where and why relative abundance data was used and where and why rarefied [MAG-derived?] OTU counts data was used.

Author's response: The sentence stating SingleM was used corresponds to a previous version of the manuscript and was an oversight on our end. All of the analyses in the current manuscript use MAG-based relative abundance. We have removed the reference to SingleM to correct this inaccuracy in the methods and appreciate the reviewer's noting this inconsistency.



L669 - mapping of reads against MAGs has been described in a previous section and is being described again [a bit differently] in this section?

Author's response: This section describes how the pipeline METABOLIC determines the coverage of the reads of the networking bacteria. METABOLIC pipeline was used with the default parameters in community mode and the description in line L669 intended to briefly explain the steps of the METABOLIC pipeline. To avoid confusion to the reader this sentence has been removed and the reader is referred to the paper describing METABOLIC pipeline for specifics on how this pipeline works.

Limitations of the study that need to be considered in the Discussion/Conclusion:

- microbial activity was not measured; it is possible that more microbe-metabolite correlations would be captured if metatranscriptomics had been included. This discussion is important for future studies.

Author's response: Thank you for this suggestion. We have addressed these comments in three ways.

1) We have added new analyses using metatranscriptomics data. We had not originally used these data as functional links between transcripts and metabolites are not optimized well yet and curation related to this was the focus of another EMERGE manuscript that is not yet submitted. However, we brought those data into this manuscript to specifically validate the correlations between metagenomic and metabolite data. To address the potential for missing actively expressed genes, we investigated the metatranscriptomic profiles of metagenome-assembled genomes (MAGs) that correlated with specific metabolites. By leveraging metatranscriptomics, we have now provided evidence for the activity of MAGs identified through metagenomics in the context of metabolite production, despite the wider range of MAGs obtained from metagenomic data.

The new text reads as follows: L363 “Further, metatranscriptomics analyses were conducted to investigate the potential microbial activity associated with the microbial-metabolite networks. The analyses revealed networked Acidobacteria are key degraders of palsa and bog peat organic matter, highly expressing diverse carbohydrate-active enzymes (CAZymes) for plant-derived polysaccharides (e.g., those in the glycoside hydrolases and polysaccharide lyases families) and polyphenolics compounds degradation (e.g., enzymes in the auxiliary activities



family with ligninolytic activity⁸¹ such those classified into AA1 and AA2 families) (Supplementary Fig. 5), consistent with prior meta-omics data at this site³. The latter finding further supports the correlation of the networked bacteria with lignin-like compound nodes, especially in the bog (Supplementary Fig. 4e), highlighting the use of microbial-metabolite networks for surveying potential microbial activity.

Metatranscriptomics revealed expression of genes for fatty acid (i.e., butanoate, propionate) and fermentation end products (i.e., glycerol, ethanol, lactate) utilization. Transcripts for propionate fermentation was dominated by Acidobacteria in palsa and bog. For butanoate, Acidobacteria led in palsa while sharing with Actinobacteria in bog. Glycerol/ethanol fermentation transcripts were mostly Acidobacteria in palsa. In bog, networked Acidobacteria highly expressed lactate/ethanol fermentation genes. Bog networked Acidobacteria, Verrucomicrobia and Halobacteria expressed acetogenesis genes compared to a broader range of genomes in the fen (Supplementary Fig. 6).

Metatranscriptomics also revealed the involvement of networked genomes in core redox reactions. Bog/fen Halobacteria and Methanobacteriota expressed hydrogenotrophic methanogenesis genes. Bog Methanobacteriota also expressed methylotrophic methanogenesis genes⁸¹, while in the fen, only Methanobacteriota genomes were responsible for this process. Dissimilatory sulfur oxidation, assimilatory/dissimilatory sulfur reduction transcripts were in bog/fen networked bacteria but only assimilatory in palsa. These expression patterns across carbon degradation, fermentation, methanogenesis, sulfur cycling linked to co-occurrence network topology shifts highlight integrating metatranscriptomics to directly measure active processes across the different thaw environments (palsa, bog, and fen). It is important to note that metagenomics data, while informative, doesn't guarantee gene expression. Integrating metatranscriptomics data directly into the correlation networks offers a more direct measure of active processes, highlighting the value of a combined approach in future studies.”

2) Also, methods related to these new analyses are added as follows: **L825** “For MAGs that showed the strongest correlations with specific metabolites, we further analyzed their corresponding metatranscriptomic data. This analysis validated the expression of genes involved in the biosynthesis pathways of these metabolites, reinforcing the functional connection between the identified microbial populations and the observed metabolite profiles.”



3) We also added an additional statement regarding the limitation of using metagenomes for correlation rather than metatranscriptomics: **L392** “It is important to note that metagenomics data, while informative, doesn't guarantee gene expression. Integrating metatranscriptomics data directly into the correlation networks offers a more direct measure of active processes, highlighting the value of a combined approach in future studies.”

- How representative are the studied MAGs in terms of the expected whole prokaryotic community diversity in these habitats? Rare taxa tend to be underrepresented in MAGs. It is important to give the readers evidence about if the MAGs reported are dominant in the system. Outline as well the limitations of missing rare taxa in the analyses.

Author's response: We thank the reviewer for bringing attention to this question. In response to this comment, we utilized SingleM (version 0.15.1) appraise tool, employing the imperfect option with a sequence identity threshold set at 0.86, to relate the 2012 raw reads and the 1,402 MAGs identified as present therein. This clarified that the MAGs collectively represent 71.8% of the bacterial community and 62.9% of the archaeal community at the genus level within the 2012 samples.

We have now included this in the material and methods section. The text reads as follows: **L625** “In order to quantify the portion of the microbial community diversity present in the 2012 samples that was represented by the 1,402 MAGs, we applied the ‘appraise’ tool from SingleM¹¹² (version 0.15.1), setting the sequence identity threshold to 0.86 (genus-level divergence). Our analysis revealed that the 1402 MAGs collectively accounted for 71.8% of the bacterial community and 62.9% of the archaeal community at the genus level within the 2012 samples.”

L Braga

Reviewer #2 (Remarks to the Author):

I hoped I could offer high praise, but regret it must also be qualified quite substantially. The authors apply concepts from metacommunity ecology to reveal the intricate interactions between vast datasets of metagenomes and metabolomes from Arctic Sweden. This is

16



Open Access This file is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In the cases where the authors are anonymous, such as is the case for the reports of anonymous peer reviewers, author attribution should be to 'Anonymous Referee' followed by a clear attribution to the source work. The images or other third party material in this file are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

intellectually adventurous and applied to an experimental system of global significance. There is potential relevance here for microbial ecology, metagenomics, metabolomics, systems biology, Arctic science, terrestrial biogeochemistry, and climate change. However the manuscript as it stands would benefit from substantive revision to bring out these links as relevant to a broader audience. I hope the recommendations below are helpful in refining the paper's appeal to different audiences:

Author's response: We appreciate the reviewer's recognition of the intellectual novelty and potential impact of our work on various scientific fields including microbial ecology, metagenomics, metabolomics, systems biology, Arctic science, terrestrial biogeochemistry, and climate change. We also appreciate the reviewer's seeming willingness to allow us to bring substantive revisions. Given this, we have carefully considered the reviewer's recommendations and addressed them wherever possible to refine the paper's appeal and ensure its key messages resonate with readers from diverse backgrounds. The reviewer's feedback has been invaluable in helping us enhance the quality and impact of our manuscript; thank you.

Please find below our point-by-point responses, organized into bullet points for clarity and ease of reference.

For Arctic science, terrestrial biogeochemistry, and climate change:

The key take-home I have from the paper is that "it's complicated" and the closer we look, the more complicated it becomes. To that end, the paper is better at complicating our understanding of greenhouse gas emissions from permafrost rather than clarifying it. This can be OK, as it is sometimes unavoidable and possibly essential depending on the intrinsic complexity of the story. However it is not an approach to take lightly when seeking to communicate effectively with the broad readership of a journal like Nature Microbiology. Foremost I would advise the authors to streamline how they narrate their story: let the feast of data not be at the expense of famine of knowledge.

Author's response: We appreciate the reviewer's thoughtful comments regarding the balance between complexity and clarity in our manuscript. To address this, we have implemented significant revisions throughout the paper. These changes extend beyond the technical questions raised by the reviewer and encompass a complete overhaul of the narrative structure



and key message, including rewriting the abstract and conclusions to emphasize the actionable insights gained from this study.

For example, the revised [abstract](#) reads as follows: **L32** “Microbiomes and their metabolites play a crucial role in ecosystems, yet their coordination in driving greenhouse gas emissions during ecosystem change remains unclear. Here we analyzed microbial and metabolite composition across a permafrost thaw gradient in Stordalen Mire, Sweden, treating metabolites analogously to biological communities to separately infer drivers of microbial and metabolite assembly and test if microbes and metabolites exhibit concordant responses to changing drivers. We found divergence between inferred microbial versus metabolite assembly processes, suggesting distinct responses to the same selective pressures across the ecosystem. This contradicts common assumptions in trait-based microbial models, highlighting limitations of community-level data alone. Feature-scale analysis illuminated specific microbe-metabolite lineages traceable to observed CO₂/CH₄ porewater variations. Our work suggests a new approach using feature-level data to identify key microbe-metabolite interactions for better understanding and modeling the metabolic processes driving greenhouse gas emissions during ecosystem changes.”

The revised [conclusion](#) reads as follows: **L392** “ This study employed an integrative metagenomic and metabolomic approach to investigate the complexity of permafrost peatland ecosystems. While our initial hypothesis centered on the congruence between microbial communities and their associated metabolites, the observed divergence in assembly patterns across the permafrost thaw gradient challenged this assumption. This highlights the need for a more nuanced framework that captures the complex interplay between microbes and metabolites within these ecosystems. Our analyses employed a multi-pronged approach to robustly integrate the multi-omics data. First, we utilized a null modeling framework to quantify the ecological processes governing the assembly of microbial communities and metabolomes. This statistically null model approach uses phylogenetic metrics to determine if the observed data deviate significantly from a null expectation, thereby identifying signatures of processes like selection and dispersal limitation ^{6, 7, 8, 9, 10}. Unlike correlation methods, null modeling provides quantitative estimates of the relative influences different ecological drivers have on community and metabolome assembly patterns.

While null modeling elucidated key assembly processes, we further integrated the datasets through correlation-based analyses. Correlations alone cannot definitively establish mechanisms, but when combined with null model inferences, they offer a valuable tool for



hypothesis generation about potential microbe-metabolite interdependencies. Notably, we identified the potential influence of organosulfur compounds, potentially originating from *Sphagnum* mosses, as pivotal factors influencing both metabolome composition and microbial community assembly, thereby impacting greenhouse gas concentrations in porewater. This finding underscores the importance of organosulfur compounds in sulfur and carbon cycling dynamics in peatlands, as sulfur cycling exerts an important control on organic carbon degradation and greenhouse gas emissions^{11, 12, 13}. Consequently, further exploration of organosulfur compounds in *Sphagnum*-dominated peatlands is crucial for ecosystem understanding and management.

This integrated multi-omics approach provides a focused view of the metabolic processes driving greenhouse gas emissions from thawing permafrost. Moving forward, our ongoing work aims to elucidate the causal mechanisms underlying these identified correlations through complementary approaches like mapping multi-omics data into metabolic pathways and microbial isolation studies. Incorporating longer timescales and expanded spatial representation will further refine our understanding of the most relevant microbe-metabolite interactions. By integrating feature-level data (microbial lineages and metabolite formulas) into future mechanistic models of greenhouse gas emissions as those summarized here⁸⁶, we can gain a focused picture of the key interactions and processes driving emissions across Arctic permafrost ecosystems, particularly given the importance of plant-derived organosulfur compounds and *Sphagnum*'s role in shaping microbial and metabolic dynamics.”

In addition to the changes mentioned above, we have also:

- 1) Addressed the lack of correlation with archaea as requested by the reviewer.
 - 2) Enhanced the S study section for clarity.
 - 3) Refined our initial hypothesis.
- Indeed, with an eye to the consequential biogeochemical processes within the thawing permafrost of the Arctic, the core insight I derive is that the cycling of organosulfur compounds, apparently by Acidobacteria (L219-234) is implied as an influence on potential methane emission (L351). The identification of these metabolites and taxa is useful, however the fundamentals of redox chemistry tell us that sulfate reduction (L222) is energetically more favorable than methanogenesis. This is readily understood by biogeochemists as a first-order control on methane release in regions of the Arctic where sulfur is more available. In a revision it would be advisable to make as explicit



case as possible for the expensive and intensive analyses of the kind reported here which hint at such processes as opposed to the simpler quantification of these biogeochemical species.

Author's response: We appreciate the reviewer's insightful comment highlighting the importance of understanding the cycling of organosulfur compounds and their potential influence on methane emissions in thawing permafrost regions of the Arctic. We acknowledge that sulfate reduction is generally considered the more energetically favorable pathway compared to methanogenesis, as per the fundamentals of redox chemistry. However, our research goes beyond just measuring total sulfate and other electron acceptors. Bulk measurements might miss the complexities of methane production, especially at smaller scales. We looked closely at specific organosulfur compounds and the microbes that interact with them. This detailed analysis revealed new information. We found Acidobacteria with the potential to use sulfate in various ways, suggesting their role in both the makeup of the local molecules (metabolome) and methane production in the Stordalen Mire permafrost. Additionally, by comparing the dominant moss species (*Sphagnum*) with the peat, we found evidence that these mosses might release sulfur compounds that specific microbes then use¹⁴. This new approach highlights the importance of considering specific organosulfur compounds, not just total sulfate levels. While acknowledging that sulfate reduction is generally more favorable energetically, our research shows how specific sulfur compounds and the microbes that interact with them can influence methane production. By considering the energy limitations and the variations in oxygen levels at different depths (as in¹⁵), our study offers a more complete picture of the metabolic processes that affect methane emissions. This detailed approach, although resource-intensive, provides insights that bulk measurements cannot. We believe this method is essential for understanding the intricate biogeochemical processes that control greenhouse gas emissions in thawing permafrost. Ultimately, this will improve our ability to model and predict these emissions as the environment changes.

We have now included a detailed explanation of the likely interactions of 55 metagenome assembled genomes (MAGs) with sulfur-containing metabolites in the bog habitat (cluster 1), allowing us to explore potential interactions beyond simple sulfate concentration quantification. We have identified Acidobacteria with the functional potential and gene expression for assimilatory and dissimilatory sulfate reduction pathways. These findings suggest that the identified sulfur-containing metabolites (now described further for chemical



properties) could act as substrates for these pathways or serve as sinks for H₂S produced via dissimilation. Furthermore, we hypothesize that these compounds are derived from *Sphagnum* mosses, based on comparative analysis of the dominant vegetation and the peat. This supports a potential link between plant-derived sulfur compounds and their subsequent utilization by specific microbial groups^{16, 17}.

Further, we characterized the potential interactions of sulfur-reducing and methane-cycling microbes, accounting for the energetic constraints imposed by the presence/absence of different terminal electron acceptors in the bog with depth and variable, perched water table. Our findings highlight organic sulfur compounds as contributors to bog metabolome variation (divergence) and identify key microbial members that are correlated with those compounds likely participating in their cycling. Beyond the quantification of sulfate species within the bog, and in peatlands in general, our study identifies specific organosulfur metabolites and the microbial taxa potentially involved in their cycling. This detailed approach allows us to connect the observed metabolome divergence to specific microbial activities and their potential influence on methane production.

The associated new text in the revised paper, reads as follows: **L232** “Cluster 1 metabolites exhibited characteristics of recalcitrance, potentially *Sphagnum* mosses- derived compounds. They had a high proportion of aromatic rings (AI_mod > 0.5 and AI_mod > 0.67), were highly unsaturated (DBE_O > 0), had a low nominal oxidation state of carbon (NOSC < 0), suggesting low bioavailability¹⁸, and were enriched in sulfur-containing condensed hydrocarbon and lignin-like compounds (**Figure 2d, Supplementary Fig. 1c**). Previous studies have proposed that *Sphagnum* litter is the primary source of organic sulfur in peat, serving as the main reservoir for recycling^{16, 17}. These sulfur-containing metabolites likely play a key role in soil organic matter (SOM) cycling within bog habitats. Comparing metabolites across plant species and peat soil from the three habitats showed: (1) *Sphagnum* had lower NOSC than other plants; (2) S- containing compounds were highly consumed in bogs ; and (3) newly observed lignin-like compounds (compounds not observed in the plant extracts or the peat) were abundant in the bog and fen¹⁴. These findings suggest that bog microbial communities co-evolved to efficiently utilize these metabolites consistent with prior substrate addition experiments¹⁹. Moreover, given low sulfur and nitrogen levels in bogs, these metabolites may control SOM decomposition rates in the Stordalen Mire.



Organic sulfur compounds, like those in cluster 1, may serve as reservoirs for microbial sulfate mobilization ²⁰ via assimilatory or dissimilatory sulfate reduction. Supporting this mechanism in the bog, 55 MAGs across 9 phyla, mostly Acidobacteria, Verrucomicrobia and Proteobacteria, strongly correlated with cluster 1 metabolites, and encoded expressed genes for sulfur metabolism pathways (**Supplementary Fig. 2, Supplementary Note 2**).

Sulfur-containing compounds can also result from sulfide (H₂S) incorporation into organic matter by sulfate-reducing microorganisms (SRM) ¹² during dissimilatory sulfate reduction. SOM degradation in peatlands depends on available terminal electron acceptors, forming a redox gradient ²¹. As sulfate reduction is more thermodynamically favorable than fermentation/methanogenesis, SRM can outcompete methanogens for substrates, mitigating methane flux ^{5, 22, 23}. SRM have versatile substrates from H₂, fatty acids, alcohols to aromatics ¹¹. Syntrophic SRM-methanogen interactions occur via H₂, formate, acetate ²⁴, methyl sulfides ²⁵ or direct electron transfer ²⁶, further influencing greenhouse gas production.

Therefore, bacteria involved in organosulfur transformations and methanogen/methanotroph interactions may be key functional species in the bog ecosystem. Our analysis identifies these potentially crucial bacteria and suggests their role in driving metabolome divergence. Correlations between bacterial genomes and sulfur-containing metabolite clusters hint at their possible use as terminal electron acceptors (TEAs), aligning with previous proposals ²². Going beyond simply quantifying inorganic sulfur species, our study identifies specific organosulfur metabolites and associated taxa elucidates complex bog biogeochemical cycling.”

- How, specifically, do the conjoined analyses of metabolomes and metagenomes advance our ability to forecast greenhouse gas trajectories from thawing permafrost, as claimed (L45-46). It is a bold promise which this impressive manuscript doesn't quite deliver upon in the present form.

Author's response: We appreciate the reviewer's perspective here. We view the manuscript's advance as setting the stage for improved forecasting by highlighting the limitations of current methods and proposing a more nuanced framework. Current approaches often rely on broad categories like microbial biomass or bulk soil properties. While these methods provide a starting point, they miss crucial details about the specific metabolic processes at play. Our study



takes a different approach, utilizing feature-level analysis to identify key interactions between specific microbial groups and their metabolites. This detailed analysis offers a more focused understanding of the metabolic processes driving GHG emissions. For instance, we identified the potential influence of lignin-like compounds with nitrogen and phosphorus heteroatoms across habitats, as well as the role of sulfur-containing compounds in driving metabolic divergence, particularly in bog environments. This framework can be further refined by employing complementary approaches, such as mapping different omics data into pathways of interest and isolating specific microbes. Additionally, incorporating data from longer time periods and a broader spatial representation will be crucial to comprehensively understand the most relevant microbe-metabolite interactions within the broader ecosystem context. The key advantage of our approach lies in its ability to integrate feature-level data into molecular-resolved models. These models explicitly represent both microbial and substrate chemistry, allowing researchers to capture the specific interactions between microbes and metabolites that directly influence GHG production. By identifying the key players driving emissions and the types of metabolites present, alongside how they interact, we can build more comprehensive models. This paper provides a summary of current models, their advantages, disadvantages and future challenges ²⁷.

Our analysis identified specific microbial groups and metabolites potentially driving GHG emissions, particularly in the bog habitat. This information can be directly integrated into models, allowing them to focus on the most relevant microbes and the specific metabolic pathways they influence. Crucially, our feature-level analysis identifies these critical interactions, providing putative mechanistic connections that cannot be observed through bulk or community-level analyses. These connections can then be leveraged within molecularly-resolved models that employ both microbially and substrate-explicit representations of biological and chemical processes ²⁷. Ultimately, this approach allows for a more comprehensive understanding of biogeochemical reaction networks, leading to more accurate predictions of biogeochemical rates like GHG fluxes.

In essence, this study demonstrates the power of feature-level analysis in identifying key microbe-metabolite interactions driving GHG emissions. This approach paves the way for developing more comprehensive models to forecast future GHG emissions from thawing permafrost, aiding in crucial climate change predictions.



To make this point clearer, we make significant changes to the abstract and the conclusion (as discussed previously) to reflect our response above. We additionally made changes throughout to make sure our take home message is clear.

- Moreover, while the team has been highly productive within the Stordalen mire region, what might the results mean, or how might the methodology apply, to the broader question of change within Arctic lands?

Author's response: We appreciate the reviewer's emphasis on the broader applicability of our findings to understand change across Arctic landscapes. While our study focused on the Stordalen Mire, the methodology and key findings can be applied to other Arctic permafrost ecosystems with some considerations. Factors like vegetation type, permafrost age, and thaw stage may influence the specific microbial communities and metabolite profiles observed. However, our approach of combining metagenomics and metabolomics with feature-level analysis offers a flexible framework that can be adapted to different environments. For example, future research can explore the transferability of these findings across a wider range of Arctic permafrost regions. This could involve applying our methodology to locations with varying characteristics, allowing for a more comprehensive understanding of how microbial communities and metabolite interactions influence greenhouse gas emissions across the Arctic.

Regarding the transferability of specific results, the importance of plant-derived organosulfur compounds in Arctic bogs could potentially extend beyond the Stordalen Mire. These compounds have been shown to play crucial roles in regulating greenhouse gas emissions and microbial community dynamics in other peatland ecosystems. Furthermore, the keystone role of *Sphagnum* mosses in shaping Arctic bog environments is well-documented in the literature. Our findings on the influence of *Sphagnum*-derived metabolites on microbial processes could have broader relevance, as these mosses are widely distributed across Arctic regions.

To address this comment, we added the following statement to the conclusion section: **L408** "Notably, we identified the potential influence of organosulfur compounds, potentially originating from *Sphagnum* mosses, as pivotal factors influencing both metabolome composition and microbial community assembly, thereby impacting greenhouse gas concentrations in porewater. This finding underscores the importance of organosulfur



compounds in sulfur and carbon cycling dynamics in peatlands, as sulfur cycling exerts an important control on organic carbon degradation and greenhouse gas emissions^{11, 12, 13}. Consequently, further exploration of organosulfur compounds in *Sphagnum*-dominated peatlands is crucial for ecosystem understanding and management.

This integrated multi-omics approach provides a focused view of the metabolic processes driving greenhouse gas emissions from thawing permafrost. Moving forward, our ongoing work aims to elucidate the causal mechanisms underlying these identified correlations through complementary approaches like mapping multi-omics data into metabolic pathways and microbial isolation studies. Incorporating longer timescales and expanded spatial representation will further refine our understanding of the most relevant microbe-metabolite interactions. By integrating feature-level data into future mechanistic models of greenhouse gas emissions, we can gain a focused picture of the key interactions and processes driving emissions across Arctic permafrost ecosystems, particularly given the importance of plant-derived organosulfur compounds and *Sphagnum*'s role in shaping microbial and metabolic dynamics.”

For microbial ecology, metagenomics, metabolomics, systems biology:

Turnbaugh and Gordon (DOI:<https://doi.org/10.1016/j.cell.2008.08.025>) invited us to “the marriage of metagenomics and metabolomics” in 2008. This study, at the forward edge of the present state of the art, showcases at least the comfortable co-habitation (or perhaps the sambo of metagenomics and metabolomics, noting the paper’s Swedish locus) of both disciplines. Yet it remains the case that it is very difficult to make the leap from sophisticated correlative analyses such as those ably performed by the authors into convincing mechanistic insights. This is not to dispute the excellence of the present manuscript, but a cautionary, conservative note must be offered when considering the promising state of this science. In essence the authors treat pools of metabolites within metabolomes like microbial OTUs within communities. I found this stimulating and innovative. Having dabbled with both methodologies in similar settings I greatly enjoyed seeing the authors’ approach. Their concepts and methodology appear portable to a diverse range of systems, and has the potential to underpin advances in other fields, which would make this work of broad relevance, but requires further thought to crack the challenges of elucidating causal relationships and mechanistic insights. I would advise the authors to reflect on this and perhaps see how far their methodological framework can support this kind of progress, or at the least provide some caveats within the discussion as to its limitations.



Author's response: We deeply appreciate the reviewer's insightful comments, particularly their recognition of the innovative approach employed in our study and its potential applicability to diverse systems. The reviewer raises a crucial point regarding the need to move beyond correlation towards mechanistic understanding to establish causal relationships for a more comprehensive grasp of the underlying processes especially when trying to integrate meta omics data. We acknowledge the limitations of correlation-based analyses and agree that establishing causal mechanisms is vital for a complete picture. Our study primarily utilized statistical and correlation-based **integration** (and not mechanistic) due to the limitations of the metabolomic data generated via FTICR MS. FTICR MS allows us to capture the full diversity of detected peaks, potentially reflecting broader metabolic processes, but unlike LC MS/MS doesn't provide structural information which is needed for mechanistic investigation. The latter however (LC MS/MS) only provide annotations for a limited number of metabolites, hence challenging some of the mechanistic investigations.

However, it is important to note that our main analyses (aside from metabolome-metatranscriptome-metagenome integration) relied on a null modeling-based operational approach that uses phylogenetic metrics to provide a quantitative estimation of the ecological processes influencing microbial and metabolome assembly processes ^{6, 7, 8, 10}. This statistical approach does not depend on correlation analysis, and the identified signatures of assembly processes have provided important insights into microbial ecology processes in different environments ²⁸.

While correlation-based analyses have their limitations, we believe they provide valuable insights when combined with other lines of evidence, such as the null model mentioned above and metatranscriptomics data. The inclusion of metatranscriptomics data in our study allows for a stronger characterization of expressed microbial metabolisms, going beyond the limits of metagenomic inferences of metabolic potential. As the reviewer suggests, combining correlation with mechanistic approaches like pathway-based integration is ideal. This is precisely the focus of another ongoing project in our lab, where we will be applying a similar framework to ten years of data alongside mechanistic analysis. This future study builds upon the foundation laid here, aiming to move beyond correlations and elucidate underlying mechanisms. By integrating identified connections/interactions into molecularly resolved substrate and microbially explicit models with mechanistic representations of reaction



networks, we can generate new mechanistic hypotheses amenable to targeted experimental evaluation.

In our revised manuscript, we strengthened the discussion on the limitations of correlation-based analyses and the importance of complementary approaches, both through additional data and mechanistic investigations. We highlighted the robustness of the null modeling approach in providing inferences beyond correlation, as well as the value of metatranscriptomics data in characterizing expressed microbial metabolisms. By addressing these limitations and highlighting future research directions, we believe our study paves the way for a more comprehensive understanding of the interplay between microbial communities, metabolic profiles, and ecosystem processes.

This text reads as follows and has been added to the conclusion section: **L397** “Our analyses employed a multi-pronged approach to robustly integrate the multi-omics data. First, we utilized a null modeling framework to quantify the ecological processes governing the assembly of microbial communities and metabolomes. This statistically null model approach uses phylogenetic metrics to determine if the observed data deviate significantly from a null expectation, thereby identifying signatures of processes like selection and dispersal limitation ^{6, 7, 8, 9, 10}. Unlike correlation methods, null modeling provides quantitative estimates of the relative influences different ecological drivers have on community and metabolome assembly patterns. While null modeling elucidated key assembly processes, we further integrated the datasets through correlation-based analyses. Correlations alone cannot definitively establish mechanisms, but when combined with null model inferences, they offer a valuable tool for hypothesis generation about potential microbe-metabolite interdependencies.”

- The dangers of large scale application of correlational analyses at pace are particularly apparent when the authors disclose that two Actinobacteria genera (L261) correlate both with carbon dioxide and methane concentration, and in all three actinobacterial and one acidobacterial lineage correlate with methane. The significance and magnitude of the positive correlations of each is not reported (it should be; Sup Table 6). Is it not surprising to the authors that methanogenic Archaea do not show correlations? If not why not?



Author's response: Thanks for your comment, and for identifying the missing correlation metrics, which we have added to Supplementary Table 6.

There are various reasons why we did not observe correlations with Archaea genomes (described below).

Our study uses community (i.e., microbiome and metabolome) assembly analysis to understand (a) if the microbiome and metabolome are influenced by similar or different ecological processes; (b) members within the community (specific microorganisms and metabolite molecular formulas) play key roles and likely contribute to how the community assembles with thaw progression; (c) if those key microbes and key metabolites are related to greenhouse gas production in the bog (with the aim to connect to environmental function). In other words, in (c), we only examined MAGs that showed a significant contribution to assembly processes, for potential correlation to CO₂ and CH₄ porewater concentrations. Archaeal genomes did not contribute significantly to assembly processes, so were not assessed for correlations to porewater gas.

To address the reviewer's concerns, we repeated the correlation using all the MAGs that showed a significant correlation with bog metabolite clusters (85 MAGs, 2 of them were Archaea). We found 45 MAGs with significant correlation with CO₂ porewater concentration, including one archaeal MAG, and 29 bacterial MAGs that correlated with CH₄ porewater concentration (Supplementary Table 6, Bog CO₂ and Bog CH₄ correlations tabs). The lack of archaeal MAG correlation to porewater CH₄ could have several explanations. First, we looked at those MAGs correlated to metabolite clusters because of the metabolome anchoring of this manuscript, and thus only two archaea were included. Second, an examination of the expressed genes of these two archaeal genomes revealed that both PLTK01 and PMGM01 possess genes involved in hydrogenotrophic and methylotrophic metabolisms, but not complete pathways were expressed. This is in concordance with a prior examination of these two archaeal MAGs² that showed that only PMGM01 encodes all the genes necessary for methanogenesis, potentially explaining the lack of correlation with CH₄ for PLTK01. Third, our analysis uses the relative abundance of MAGs to correlate with concentrations of CO₂ and CH₄ in porewater. It is possible that the use of other omics, for example gene expression, would show a better correlation with concentration of greenhouse gases.



- Moreover, given that the authors have access to the MAGs of these correlating lineages and the tools with which to explore their functional potential, what might be the biological mechanism by which two different bacterial phyla appear to contribute to methanogenesis within this system? In a revised manuscript it would be helpful if the authors devoted attention to properly exploring this and other conclusions of this kind. Some biology might be lurking, or at least validation of the methodology.

Author's response: We thank the reviewer for this comment. We believe that the main reason why Acidobacteria and Actinobacteria genomes contributed to methanogenesis in the bog is because they provide methanogens with the metabolites necessary to produce CH₄ (discussed below). To address this and to improve clarity, we included a supplementary note that describes the predicted and expressed metabolism of these 5 genomes that correlated with CO₂ and CH₄ concentrations. Beyond these 5 genomes, we also described the predicted potential and expression of microorganisms that form part of the microbial-metabolite networks.

The supplementary note 3 reads as follow: "Genomes that significantly correlated with β NTI feature-derived metabolite clusters (see section 2) were correlated with CO₂ and CH₄ concentrations in the bogs and fen pore water, with the aim to find connections between metabolome ecological processes, the microbiome, and ecosystem functions.

This note describes the metabolic potential and expression of 5 metagenomes that showed significant correlations with porewater greenhouse gasses in the bog, were significantly correlated with metabolite clusters, and that also contributed to microbiome assembly.

Two Acidobacteria genomes of genus Terracidiphilus (Terracidiphilus sp003138365 and Terracidiphilus sp003142015) significantly correlated with CO₂ and CH₄ concentrations (Supplementary Table 6, Bog_greenhouse_correlation sheet). An Acidobacteria genome of genus Holophaga (Holophaga sp003165415) only correlated with CO₂. Additionally, two Actinobacteria genomes (RAAP-2 sp003160115 and Fen-455 sp003140595) both significantly correlated with CO₂ and CH₄ concentrations (Supplementary Table 6, Bog_greenhouse_correlation sheet). The metabolism of these genomes was predicted based on MAG annotation, the MAG completeness of these 5 genomes was higher than 95% with less than 3% of contamination (Supplementary Table 2- MAGs sheet). Furthermore, the activity of these genomes was validated with metatranscriptomics data.



Acidobacteria genomes

These genomes encoded genes for alpha-glucosidase, beta-galactosidase and beta-glucuronidase, mannose-6-phosphate isomerase, L-itol-2-dehydrogenase, xylulokinase and polyphenol oxidase for polyphenol cleavage (the latter only in the *Holophaga*). These genomes, especially *Terracidiphilus* sp003138365 and *Terracidiphilus* sp003142015, expressed a wide set of carbohydrate active enzymes from different families, including glycoside hydrolases, carbohydrate esterases, auxiliary activities and polysaccharide lyases (last family only expressed in *Terracidiphilus* sp003138365) for the degradation and utilization of diverse polysaccharides (see expression in Supplementary Fig. 3). *Terracidiphilus* sp003142015 encodes genes for hydrocarbon degradation (i.e., Homoprotocatechuate and protocatechuate) while the *Holophaga* genome encoded genes for Catechol ortho-cleavage and Homoprotocatechuate degradation.

Terracidiphilus sp003138365, *Terracidiphilus* sp003142015, and *Holophaga* sp003165415 each encoded a complete central carbon metabolism. All glycolysis genes were present as well as an almost complete Entner-Doudoroff pathway. Enzymes for the aerobic (pyruvate dehydrogenase complex) and anaerobic transformation (pyruvate ferredoxin oxidoreductase [por]) of pyruvate to acetyl CoA were observed as well. Genes for pyruvate oxidation to acetyl CoA were expressed in all three genomes, especially in *Terracidiphilus* sp003138365. The 2-oxoglutarate/2-oxoacid ferredoxin oxidoreductase subunits A (korA), B (korB) and D (korD) that are involved in the transformation of succinyl-CoA to 2-oxoglutarate and CO₂ and 2H⁺ were expressed as well.

Also, *Terracidiphilus* sp003138365 and *Holophaga* genomes has the potential to transform acetyl CoA to acetate (a substrate of acetoclastic methanogens) and CO₂ using phosphate acetyltransferase (pta) and acetate kinase (ackA) enzymes while *Terracidiphilus* sp003142015 only encoded for acetate kinase enzyme. These genes pta and ackA were expressed according to metatranscriptomics analysis. Acetyl CoA can also enter the citrate cycle, whose genes were complete according to the metagenome predictions. Also, *Holophaga* and *Terracidiphilus* genomes had a complete pentose phosphate cycle, non-oxidative and oxidative phase, but only genes in the non-oxidative phase were expressed.

These genomes also have the potential of carbon fixation using different pathways, a fairly complete reductive citrate cycle (Arnon-Buchanan cycle) was predicted and almost all genes in



the incomplete reductive citrate cycle. Few genes within the reductive citrate cycle (Wood-Ljungdahl pathway) were predicted in *Terracidiphilus* and *Holophaga* genomes but only the methylenetetrahydrofolate dehydrogenase (NADP+) / methenyltetrahydrofolate cyclohydrolase (fold) gene were found expressed only in *Terracidiphilus* sp003142015 and *Holophaga* genomes.

Genes for the propanoyl-CoA metabolism were only predicted in the *Holophaga* genome, while expression of these genes were mostly expressed in *Terracidiphilus* sp003142015.

Potential for sulfur metabolism was predicted as well. Enzymes for assimilatory sulfate reduction were predicted in these genomes (phosphoadenosine phosphosulfate reductase, sulfite reductase (only these two enzymes were expressed), sulfate adenylyltransferases).

Within the methane metabolism pathway, *Terracidiphilus* sp003142015 encoded genes for heterodisulfide reductase 2 enzyme (hdrA2) and F420-non-reducing hydrogenase iron-sulfur subunit but were not found expressed. Meanwhile, *Holophaga* expressed hdrA2, hdrB2 and hdrC2 enzymes. Also, *Terracidiphilus* sp003138365 expressed a formaldehyde dehydrogenase (fdhA), which catalyze the transformation of formaldehyde to formate (a classical substrate of methanogenesis) and a formylmethanofuran dehydrogenase subunit E (fwdE) which is involved in the transformation of formylmethanofuran to formate and methanofuran. *Holophaga* genome expressed a formate dehydrogenase iron-sulfur subunit (fdoH) which is involved in the transformation of formate to CO₂.

These 3 Acidobacteria genomes genes for amino acid metabolism including genes for the biosynthesis of cysteine, methionine, threonine, valine, leucine and isoleucine, lysine (succinyl-DAP pathway), arginine, ornithine, histidine, tryptophan and phenylalanine.

Also, these 3 genomes have the potential to degrade short chain fatty acids and alcohols as they encode genes for the butyrate kinase (*Terracidiphilus* sp003142015 and *Holophaga* genome), propionate CoA-transferase (*Holophaga* genome), phosphate butyryltransferase (*Terracidiphilus* sp003142015) and alcohol dehydrogenase (expressed genes Supplementary Table 8, metaT_bog_correlating_MAGS). Genes within the butanoate metabolism were also expressed specially in *Holophaga* genome, for example the acetyl-CoA C-acetyltransferase (ACAT) that transform Acetyl CoA to Acetoacetyl-CoA and the acetate CoA/acetoacetate CoA-transferase alpha and beta subunits (atoD and atoA) that are involved in the transformation of Acetoacetyl-CoA and Acetate to Acetoacetate and Acetyl-CoA.



Hydrogenases observed in these genomes included: NAD-reducing hydrogenase large subunit, bidirectional [NiFe] hydrogenase diaphorase subunit, F420-non-reducing hydrogenase iron-sulfur subunit, [NiFe] hydrogenase diaphorase moiety large subunit and NADP-reducing hydrogenase subunit HndD.

Actinobacteria genomes

Actinobacteria encoded genes for several central carbon metabolism pathways, which together with their reported function in xylose degradation and lactate fermentation ⁵ may explain their role as potential keystone species that contribute to core functions of the microbial communities.

RAAP-2 sp003160115 encoded several genes involved in polyphenol degradation (i.e., catechol 1,2-dioxygenase, 3-phenylpropionate/cinnamic acid dioxygenase). Enzymes involved in alcohol degradation were present in these genomes as well (i.e., alcohol dehydrogenase (both) and lactic acid dehydrogenase (only in RAAP-2 sp003160115)). Alpha-glucosidase was encoded in these genomes as well.

The two Actinobacteria genomes (RAAP-2 sp003160115 and Fen-455 sp003140595) had an almost complete glycolysis cycle, enzymes for pyruvate oxidation to acetyl-CoA and a complete citrate cycle were predicted as well. Genes for the pentose phosphate pathway were predicted as well, where only genes for the non-oxidative phase were expressed. The Entner-Doudoroff pathway was complete, but only genes for the semi-phosphorylative pathway were expressed in Fen-455 sp003140595. Genes for propanoyl-CoA metabolism were also predicted and expressed in both genomes.

These genomes have the potential to fix carbon using different pathways, genes for the reductive citrate cycle, the incomplete reductive citrate cycle and the reductive acetyl-CoA pathways were predicted but pathways were incomplete.

Within the methane metabolism pathway, Fen-455 sp003140595 encoded genes for heterodisulfide reductase 2 enzyme (hdrA2), hdrB2, and hdrC2 enzymes. It also expressed the glutathione-independent formaldehyde dehydrogenase, that is involved in the transformation of formaldehyde to formate. Both genomes showed genes involved in coenzyme F420 biosynthesis (LPPG:FO 2-phospho-L-lactate transferase [cofD] and coenzyme F420-0:L-



glutamate ligase / coenzyme F420-1:gamma-L-glutamate ligase [cofE]) but they were not expressed.

Similar to the Acidobacteria genomes, these Actinobacteria genomes encoded genes for the biosynthesis of different amino acids, serine, threonine, methionine, valine, isoleucine, leucine, lysine (succinyl-DAP pathway), arginine, proline, histidine, tryptophan.

Genes involved in the assimilatory sulfate reduction pathway were predicted as well, these genes were expressed only in Fen-455 sp003140595.

Genes encoding hydrogenases were observed in Fen-455 sp003140595 (i.e., NAD-reducing hydrogenase large and small subunits, F420-non-reducing hydrogenase iron-sulfur subunit, bidirectional [NiFe] hydrogenase diaphorase subunit and coenzyme F420 hydrogenase subunit beta)."

For science generally:

- On a related note, while it is pleasing that the authors seek to frame their study's goals so prominently within the hypothetico-deductive framework (L94-L130) through a series of explicit hypotheses, the hypotheses posed, while valuable, do require refinement. Most of all, their presentation needs further thought.
- In places the hypotheses are not declarative but probabilistic (e.g. L102 ...conditions likely contribute...) and in many cases the predictions are implicit or at least qualitative. Ensuring a hypothesis has an explicit statement, defined methodology and testable prediction is always a powerful way of making sure the hypothesis is apt.
- Meanwhile, since this paper reports the outcomes of work started over a decade ago, it is best to use the past tense throughout the tracts of text dealing with hypotheses.

Author's response: We would like to thank the reviewer for these excellent points which we completely agree with. We have converted all verbs related to the hypotheses to the past tense, and adjusted the hypotheses as follows: (i) Each is presented as a clear and declarative statement, explicitly outlining the predicted outcome for each ecosystem. (ii) the methodology that will be used to test these hypotheses is stated at the beginning of the section. We believe these revisions address your concerns by ensuring a more transparent and testable hypothesis



framework. For example, in the original hypothesis regarding microbial community assembly in palsa environments, we stated: "We anticipate that selection pressures will yield homogenous microbial communities...". This has been revised to: ""We hypothesized that selective pressures would dominate, leading to a predominance of homogenous microbial communities adapted to exploit seasonal thawing events (Ref. 33, 34). Limited dispersal, further encouraged by the dry and ombrotrophic nature of palsa sites, would further contribute to this homogeneity ..."" Also building up on additional comments from the reviewer regarding the main take home message of the paper, we also added one additional statement to the end of the hypotheses section that connects back to how this study will inform enhanced model predictions. The rest of the new text reads as follows:

"Hypotheses on key drivers of microbial community assembly

Our study aimed to unravel the key drivers shaping microbial community assembly across a thaw gradient encompassing palsa, bog, and fen ecosystems. We hypothesized that these drivers would exhibit a dynamic shift in response to the unique environmental signatures of each habitat. To rigorously test these predictions, we leveraged the power of shotgun metagenomics data analysis, guided by principles from community assembly theory to decipher the composition and structure of microbial communities across the thaw gradient. In palsa environments, we hypothesized that selection pressures would have dominated assembly, leading to a predominance of homogenous microbial communities adapted to capitalize on seasonal thawing events^{29, 30}. Limited dispersal, further encouraged by the dry ombrotrophic conditions of palsa sites, would have contributed to this homogeneity. Bog ecosystems present a contrasting scenario. Although exposed to a strong selective filter from *Sphagnum* metabolites and acidity, the bog microbiota likely experienced some local dispersal due to fluctuating water table levels, potentially mitigating the dominance of homogenous selection. Finally, for fen environments, we hypothesized that minerotrophic conditions and full inundation promoted microbial dispersal outweighing selective pressure signal^{29, 31}.

Hypotheses on key drivers of metabolite assembly

We further anticipated variations in metabolite composition across the thaw gradient due to differences in environmental conditions and microbial activity. To investigate this, we employed high-resolution metabolomics techniques to profile the metabolite pools in each ecosystem and we leveraged on assembly theory (treating metabolites as analogous to microbial communities) to investigate ecological processes contributing to metabolome variability.



In *palsa* environments, the high chemical lability of litter and the presence of oxygen likely promoted rapid microbial degradation compared to the anoxic conditions of bogs and fens¹⁴. Steep environmental gradients further suggested the dominance of deterministic factors, particularly variable selection, in shaping the *palsa* metabolome. Additionally, minimal dispersal was expected due to the dry and ombrotrophic nature of *palsa* sites. In contrast, *Sphagnum* moss significantly influenced the bog metabolome through litter inputs and bioactive metabolites. Here, we predicted homogenization of the metabolome due to the accumulation of recalcitrant metabolites, altered enzymatic activity, and anoxic conditions. Dispersal processes might have played a more significant role compared to *palsa* sites due to water table fluctuation^{32, 33}. Finally, for fen environments, we hypothesized a more variable metabolome likely shaped by the highly diverse and active microbial community present. The expected presence of microbially-derived compounds with low aromaticity^{22, 32, 34} further supported this prediction. Additionally, environmental conditions like minerotrophic features, runoff from surrounding areas, and water mixing might have facilitated some level of metabolite dispersal³¹.

Community-scale or feature-scale congruence: what to expect in permafrost ecosystems?

Building on the above hypotheses across the thaw gradient, we further examined the congruence between microbial communities and their metabolomes at the community and feature-scale. Understanding this relationship is important for elucidating ecosystem functioning, informing predictive modeling efforts, and guiding conservation efforts. We hypothesized that synchrony between ecological assembly processes shaping microbial communities and metabolite profiles at the community level signifies a more stable and coordinated ecosystem response within permafrost systems. Conversely, alignment at the feature-scale would suggest the presence of specialized, finely tuned microbial activities and metabolite patterns within these habitats. This scenario emphasizes the significance of fine/feature scale processes in shaping ecosystem function across the thaw gradient. “

- It should be noted that the manuscript depends heavily on data from another manuscript which is currently under review (McGivern et al. L136). Given the brief turnaround for reviewing in this round it hasn't been feasible to ask for it to be provided. In a revision, unless this other paper is publicly available it would be helpful to include as supplemental data for review purposes only. At this stage I am taking for granted that the McGivern manuscript does not materially affect the interpretation of this manuscript or its significance.



Author's response: Thanks for your comment, the paper that the reviewer is referring to has now been accepted in Nature Microbiology. Also, all the data that this manuscript uses is currently deposited in repositories as described in the data availability section. Moreover, to improve accessibility to the reader, the main datasets used in the manuscripts are included as supplementary tables.

- Asynchrony (L66) vs microscale synchrony (L352) vs micro/fine-scale synchrony (L359): I take asynchrony to mean events happening with differing temporal patterns and as the antonym of contemporaneous sequences of events. Resolving the relevant scales of variation is important: The authors concede their temporal and spatial scales of are only revealed at monthly timescales in some instances (L168) – this is still quite a synoptic timescale! Given the transience of some metabolic interactions, can the authors justify this? Capturing time-lagged events remains problematic within this methodology. No methodology is perfect, but if the maximal temporal resolution is month-by-month, the authors need to be cautious.

Author's response: Thank you for raising the important issue of terminology and temporal resolution in our manuscript. We acknowledge the potential for confusion surrounding the terms "asynchrony," "microscale synchrony," and "micro/fine-scale synchrony."

With respect to “terminology”: We have revised the terminology throughout the manuscript to ensure consistency. We will replace "microscale synchrony" and "micro/fine-scale synchrony" with a more precise term like "congruence at the feature-scale" or "alignment at the metabolite level." This clarifies that we are not referring to synchronized events across time but rather to a consistent relationship between specific microbial activities and their corresponding metabolites. It is important to note that while our study focuses on identifying congruence at the feature-scale (i.e., consistent relationships between specific microbial activities and their metabolites) (rather than community level interactions), we acknowledge the possibility that time-dependent processes might also contribute to this observed pattern. However, it is hard to test with the data that we have.

With respect to “temporal resolution”: The reviewer is correct in pointing out that our study's temporal resolution, with data collected at monthly intervals (in this particular study), might not be ideal for capturing highly transient metabolic interactions. We recognize this limitation and have addressed it in the following ways: We acknowledged this limitation in the conclusion,



emphasizing the need for future studies to incorporate higher temporal resolution techniques to fully understand the dynamics of microbial community-metabolome interactions.

We believe that despite the limitations in temporal resolution, our study offers valuable insights into the potential relationship between microbial communities and metabolomes across permafrost habitats at least monthly.

As the reviewer mentions though, tracking changes with time is hard especially in systems like permafrost systems coupled with the high cost of doing omics on all these time points and samples with different spatial resolution. Personally, I think the development of in situ sensors capable of real-time monitoring of specific metabolites or microbial populations could be a valuable tool for future studies. These sensors could be deployed in permafrost environments to provide continuous data on the temporal dynamics of the microbial community-metabolome relationship.

- I really liked Table 1 which made the definition of terms and their cognate metrics explicit. It would be helpful if the authors could be similarly explicit about their thresholds for asynchrony, synchrony, and micro/fine-scale synchrony, and whether microsynchrony or fine-scale synchrony are distinctive.

Author's response: Please see our comment above.

Minor comments:

L186: please provide a reference for “previous work”

Author's response: Thanks, we resolved this comment.

L206: hydrocarbon like

Author's response: Thanks, we resolved this comment.

Table 1: is “more” missing from (ecological similar communities than expected by chance).

Author's response: Thanks, we resolved this comment.

Reviewer #3 (Remarks to the Author):

Thank you for giving me an opportunity to review this work and thank you for your patience

37



Open Access This file is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In the cases where the authors are anonymous, such as is the case for the reports of anonymous peer reviewers, author attribution should be to 'Anonymous Referee' followed by a clear attribution to the source work. The images or other third party material in this file are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

with the review.

Overall I think this is a well-written and carefully performed study (as one would expect for something submitted to Nature Microbiology). I think the work is an interesting approach to soil metabolomics and microbial ecology albeit similar to some recent previous studies that are not cited (see minor comments section). I have some major and some minor comments as set out below.

I am not a microbiologist or an ecologist, so I have focused mainly on the metabolomics in the paper.

Author's response: We sincerely appreciate you taking the time to thoroughly review our manuscript and provide constructive feedback. We are grateful for the opportunity you gave us to strengthen the work, especially regarding the metabolomics methods and analyses. We have also addressed each of the specific major and minor comments through revisions to the text, methods, and/or analyses. We believe these changes have improved the clarity, context, and rigor of the work, particularly with respect to the metabolomic protocols and interpretation.

Before addressing the specific major and minor comments, we would like to clarify that our metabolomics analysis utilized Direct Infusion Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (DI FT-ICR MS). Thus, chromatographic techniques for compound separation, identification, and quantification were not performed. We leveraged the high resolving power, ultra-high mass accuracy and sensitivity of FT-ICR MS³⁵, along with the rapid data acquisition of DI-MS, to gain a comprehensive overview of the soil organic matter profile across the thawing permafrost gradient (i.e. comprehensive metabolome characterization). As natural organic matter is a complex and dynamic mixture, FT-ICR MS is a powerful tool to detect and resolve individual soil organic matter molecules³⁶.

While we acknowledge the advantages of LC-MS/MS for accurate structural elucidation, LC-MS/MS can be biased towards highly abundant compounds with MS/MS spectra. Thus, many low abundance peaks that may play major roles in biogeochemical cycling could be lost. Considering that our goal here was to examine assembly processes, we were more interested in capturing as many metabolites as possible without focusing too much on structural annotation beyond biochemical classification. We have a follow-up study examining samples from the past decade where we employ LC-MS/MS since we discuss microbial metabolism in greater detail.

38



Open Access This file is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In the cases where the authors are anonymous, such as is the case for the reports of anonymous peer reviewers, author attribution should be to 'Anonymous Referee' followed by a clear attribution to the source work. The images or other third party material in this file are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

However, the limitations of relying solely on DI-FTICR-MS for metabolite identification are clear, including the inability to differentiate isomers, determine molecular structures, or identify functional groups, as well as issues with signal suppression or enhancement^{37, 38}. Moving forward, implementing this technique alongside chromatographic techniques could provide better understanding of the complex processes in natural ecosystems. To address this, we have added clarification in the materials and methods about the limitations of DI FT-ICR MS and the importance of coupling it with LC-MS/MS approaches.

The new text reads as follows: **L482** “As soil organic matter comprises a complex and dynamic mixture of metabolites, DI-FTICR-MS detects and resolves a wide range of individual molecules for characterization of molecular composition and transformation. While chromatography-based methods like LC-MS/MS enable more accurate structural elucidation, they can be biased towards highly abundant compounds. Since the objective here was to capture assembly processes across the soil metabolome, we favored a high-throughput approach providing wider detection of low abundance metabolites without focusing extensively on structural annotation beyond biochemical classification.

In this study, we used only water as an extractant to focus on the bioavailable compounds actively cycling with microbes in this dynamic thawing environment. Our goal was to mimic the natural thaw conditions as closely as possible, without introducing any additional compounds that would not naturally be present”

This could perhaps be missing something, but I am a little concerned with the extraction method used for metabolites in this study as it is not a standard metabolomic method. As far as I can see only water was used to extract metabolites from the soil, but this potentially means any large molecules like fatty acids, lipids, and many larger secondary metabolites would likely not have been extracted. This in turn means a lot of information may have been missed and the models might be incomplete. It is standard in metabolomics to use an organic solvent as well as water. The almost field standard method is methanol-chloroform-water extraction. Why was this not used here? Have I missed something? If there is a reason, I think it needs to be explained.

Author's response: The reviewer makes an excellent point. A core challenge in soil metabolomics is obtaining a comprehensive view of the diverse range of metabolites. Our lab employs a variety of extraction approaches spanning from water alone to water/methanol mixtures to more non-polar organic Folch or MPLEx extractions using



chloroform/methanol/water. We agree with the reviewer that adding organic solvents can help extract additional non-polar and semi-polar metabolites from the sample. However, in this specific study our rationale for using only water was to focus on the bioavailable compounds actively cycling with microbes in this dynamic thawing system ^{36, 39}.

We aimed to mimic the natural thaw environment as closely as possible. By doing so, we are covering a wide range of compounds, from carbohydrate-like, to amino-sugar-like to polyphenolic compounds ³⁹. Nonetheless, we should have articulated our justification for selecting water as the extractant to enhance clarity. Most certainly, combining multiple extraction approaches can provide wider coverage of the soil metabolome. As the reviewer suggests, employing methods targeting a broader range of metabolites with different polarities could reveal additional insights into the impacts of permafrost thaw on soil organic matter transformations. To address this comment, we added additional justifications to the materials and methods section of the paper.

The new text reads as follows: **L490**“In this study, we used only water as an extractant to focus on the bioavailable compounds actively cycling with microbes in this dynamic thawing environment. Our goal was to mimic the natural thaw conditions as closely as possible, without introducing any additional compounds that would not naturally be present.”

A similar question is why were the samples only analyzed in negative ionization mode and not positive mode? Some compounds analyze better using positive ionization so why not use both to get the best coverage? Using only one ionization mode could again mean metabolites were not seen.

Author's response: The reviewer raises an excellent point. Ideally, both positive and negative ionization modes would be used to capture a diverse range of metabolites. Previous work has revealed that organic matter is dominated by oxygen-containing compounds, as demonstrated by tandem mass spectrometry on organic matter compounds - the majority lost a fragment ion of m/z 44, suggesting carboxylic acid-rich organic matter. Considering this, we opted for negative mode only, expecting that the majority of organic matter would ionize better in negative mode due to the prevalence of acidic functional groups ⁴⁰. However, we acknowledge that we may be missing some metabolites that preferentially ionize in positive mode. Even so, most compounds, even those with nitrogen, often also contain oxygen based on research on



organic matter composition. Hence, we believe that while they may ionize better in positive mode, they would still ionize in negative mode. Since the focus of this analysis was on presence/absence, we do not think using only negative mode will impact the data significantly. However, some compounds that exclusively ionize in positive mode may be missing. Considering all samples were analyzed the same way, comparisons remain valid. However, we have added a sentence to the Materials and Methods section acknowledging the potential limitations of using only negative ion mode.

The new sentence reads as follows: **L502** “Negative ionization mode was used for all metabolomics analyses in this study, as previous work has shown that organic matter, which makes up a large proportion of the sample matrix ⁴⁰, is predominantly composed of oxygen-containing compounds like carboxylic acids that ionize best in negative mode. While positive ionization could provide additional coverage of some nitrogen-containing compounds, the focus of this analysis was presence/absence and relative comparisons between samples, rather than capturing all possible metabolites. Since all samples were analyzed under the same parameters, comparisons should still be valid. However, the use of only negative ion mode may preclude detection of some metabolites that ionize exclusively in positive mode.”

Also, can you please add details of the quality control methods used for the metabolite analysis?

Author’s response: Thanks for your question. We included the following quality control methods: instrument stability was tested using a Suwannee River Fulvic Acid standard (SRFA) that was run on a daily basis. To minimize the potential or carry-over between different samples the instrument was flushed with a solution of milliQ water and HPLC grade methanol between samples injection. Also, HPLC grade methanol blanks were injected and collected through the process to keep an eye on any contamination. We included details about these methods in the material and methods section.

The new section reads as follows: **L511** “To ensure the instrument stability, a Suwannee River Fulvic Acid (SRFA) standard obtained from the International Humic Substance Society (IHCC) was injected at the beginning of the run. To monitor potential carry-over between samples, HPLC grade methanol blanks were injected throughout the process. Additionally, the instrument was flushed between samples with a solution of milliQ water and HPLC grade methanol. Variations in carbon concentration from different samples were controlled by



modulating the ion accumulation time that was adjusted for each sample and ranged between 0.1 to 0.3 seconds.”

How did you confirm metabolite ID for example?

Author’s response: The reviewer makes a good point - direct infusion Fourier transform ion cyclotron resonance mass spectrometry (DI-FTICR MS) does not allow identification of specific molecular structures, as it does not utilize chromatographic separation or fragmentation techniques. However, the ultra-high mass resolution and accuracy provided by FTICR MS enables detection of individual SOM molecules differing by a single electron mass unit across a wide m/z range. To ensure accurate mass assignment, each spectrum was calibrated post-data collection using a homologous series of common organic compounds. Following calibration, features in each sample spectrum were aligned to the same features across all sample spectra. Molecular formulae were then assigned within 1 ppm mass accuracy taking into consideration the seven golden rules for molecular formula assignment^{36, 41, 42}. Details on calibration, peak alignment, and molecular formula assignment have been added to the Materials and Methods section. While molecular structures cannot be definitively identified, this approach provides an in-depth view of the individual components of complex SOM samples.

The new text reads as follows: **L522** “Raw spectra collected from each sample, was converted to a list of m/z values using the BrukerDaltonik version 4.2, FT-MS peak picking module using signal to noise ratio of 7 and absolute intensity threshold of 100 (default). Out of the 85 samples analyzed, 17 were technical replicates where the mass spectrometry data was collected twice through independent sample injections. These replicate samples allowed assessment of technical variability in the metabolomics data pipeline. The remaining 68 samples corresponded to the biological experimental conditions of interest across different sites, depths, and time points; these samples constitute 3 replicates per peat core sampled within each habitat and month, unless otherwise specified (**Supplementary Table 2**). Chemical formula assignment was performed with Formularity⁴¹ using parameters described in⁴³. Compounds with m/z values outside 200 to 900 m/z and isotopic peaks (¹³C-peaks) were filtered out for downstream analysis.”

And did you follow the recommended protocols for Standard reporting requirements for biological samples in metabolomics experiments in the biological or environmental context (1)?

Author’s response: Thanks for pointing this out. Details about sampling date, time of sample collection, GPS location, depth of coring, air and soil temperature, active layer depth, and notes

42



Open Access This file is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In the cases where the authors are anonymous, such as is the case for the reports of anonymous peer reviewers, author attribution should be to 'Anonymous Referee' followed by a clear attribution to the source work. The images or other third party material in this file are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

related to the sampling can be found at the EMERGE database <https://emerge-db.asc.ohio-state.edu/datasources/12>. The link to this information was provided in the Material and methods section and the Data availability section of the manuscript. For further clarity, we included this information in Supplementary 2, sample metadata tab. This table includes a description for samples and the environment involved in the study. Also, materials and methods give detailed descriptions of the process of sampling. The authors believe that this data is in concordance with the requirements described in ⁴⁴, as the reviewer suggested.

Again, perhaps I missed it but I am also not clear on how the raw mass spectral data was processed. A peak or m/z value in a chromatogram does not necessarily equal a metabolite, it could be contamination in the instrument or solvents, etc so out of the 14432 peaks that you found how many were actually metabolites/compounds of interest and how did you get down to 6763 and then to 6526?. Even 6526 seems a much higher number of metabolites than many other studies and if potentially non-relevant data were used to build models of relationships between metabolites and other factors it could introduce error and uncertainty.

Author's response: Thanks for your questions. Analysis of organic matter using Direct Injection FT ICR MS allows the detection of many thousands of peaks in a m/z range from 150 to 1000 m/z ³⁶. This technique does not allow to “identify” metabolites of interest but rather to produce an organic matter profile of the samples based on the intensity and the assigned molecular formula of each m/z. For this type of analysis, raw spectra collected from each sample was converted to a list of m/z values using the BrukerDaltonik version 4.2, where the FT-MS peak finding algorithm will pick peaks based on parameters such as a signal to noise ratio and absolute intensity threshold. Then a calibration and alignment step of the mass spectra was applied using an organic homologous series separated by 14 Da (CH₂), followed by estimation of molecular formulas for each m/z within a mass error range of 5 ppm. For this step we used Formularity ⁴¹, a widely used automated pipeline following parameters described in ⁴³.

A total of 42514 m/z were initially detected, representing the full suite of masses that passed the signal to noise ratio and intensity threshold. Of them, Formularity was able to assign a molecular formula to 10 601 m/z following the parameters described in ⁴³. This resulted in a table containing the intensities of each m/z per sample, which was the input data for the manuscript analysis. This table was previously published in ¹⁴, and it was also uploaded to this



manuscript GitHub repository. Moreover, to improve clarity, this table has now been included as a Supplementary Table 2, FT-ICR matrix tab.

Then, following the workflow described in ⁸, processing and filtering of the data (detailed in the R scripts of this manuscript GitHub repository) resulted in 14432 m/z of which 6763 were assigned a molecular formula. The final number of peaks that the reviewer mentioned (6526), corresponds to further downstream analysis where m/z with and without formula assignment were used to build a dendrogram following ⁸.

Line 422 to 433. Why were samples removed from the dataset just because they had fewer peaks? I don't feel this is a legitimate reason to exclude samples. They may just have had fewer metabolites in that sample. Please expand on the reasoning here.

Author's response: Thanks for this question. We eliminated these samples because the number of m/z detected in them were very low in comparison to the rest of the samples, less than 200 m/z (and in some cases close to zero m/z) compared to the more than 2000 m/z detected in most of the samples, suggesting technical issues during data collection. These differences had the potential to introduce bias in the analysis. Furthermore, the number of m/z detected in the other replicates and technical replicates (when available) of these samples were similar to the rest of the dataset suggesting that the eliminated samples might have suffered from issues during the m/z collection process.

Later on there is also discussion only including features if they are in >50% of samples. This again seems a bit extreme, perhaps there is an interesting reason some features are in <50% of samples but by not looking at them the chance to investigate is lost?

Author's response: Thank you for your inquiry. To construct microbe-metabolite networks, we specifically retained only those metabolites and metagenomes assembled genomes (MAGs) that are present in at least 50% of the samples. However, for all other analyses, we utilized the entirety of the datasets available to us. The networks themselves were built on Spearman correlations derived from centered log-ratio transformed microbial and metabolite abundances. Additionally, we employed an RMT-based approach to establish the correlation cut-off threshold for network construction. This selective inclusion of the most representative microbes and metabolites was essential to ensure the accuracy and reliability of our correlation calculations ^{45, 46}. Another reason is the limitation of the number of features that the online tool



used in this analysis (MENAP) can handle at a time. We acknowledge that using a lower filter can likely allow the identification of more correlation patterns by including less prevalent metabolites and microbes and it will be taken into consideration for future analysis. Nonetheless, patterns using our current filter are representative of the correlations that occur among the core microbial and metabolite features that can be found across all habitats.

Minor comments

Line 58, my wish to give a short definition of metabolomics where the term is first used for those that are not familiar with it.

Author's response: Thanks for this comment, we included the concept of metabolomics.

Line 62, give a reference for the statement on metabolites here.

Author's response: Thanks for this comment, we have included references to this statement, the new text reads as follows: “Metabolites, produced by microbial reactions in response to physiological or environmental changes⁹, provide direct insight into in situ microbial function⁴⁷”

Line 63, I think the definition of soil community metabolome was first given in Jones et al (2) not in the papers cited here.

Author's response: Thanks for this comment, we included the citation for ecological metabolome first suggested by¹⁷ and community metabolome¹⁸ that the reviewer kindly suggested.

Authors seem to be missing out on previous work in soil community metabolite and function analysis e.g. (3, 4). These papers have essentially applied the same principles and techniques - examining microbial communities and metabolomes together just applied to slightly different problems. NB this reviewer is not the author of those studies, and I am not saying the papers have to be cited just asking the authors to consider if they are relevant and warrant inclusion.



Author's response: Thank you so much for bringing these to our attention. We have read these papers. We agree that certain aspects of this work should be referenced here. We have now included these citations in the introduction of the manuscript.

How did you ensure the push corers used to take samples did not contain microbes from outside the study area? Were they sterilized before use?

Author's response: Thanks for your comment. Core sampler and equipment used in the sampling were rinsed with DI water between cores. To avoid microbial contamination, 1 cm of the outermost core edge was avoided.

Lines 393 to 400. I think websites should be cited as proper references not just pasted into the text.

Author's response: Thanks for pointing this out. We reviewed submission guidelines, and it states that "URLs for web sites should be cited parenthetically in the main text, not in the reference list."

Line 413, while authors have cited technical replicates what about actual sample replicants, e.g. two or three samples from the same areas being separately extracted and analyzed?

Author's response: Thank you for bringing up this important point, the sampling strategy involved sampling 3 peat cores per habitat (palsa, bog and fen) within each month (June, July and August). We clarified the sampling strategy in L458. However, some replicates were excluded and/or lost during sample processing and collection. To improve clarity of the reader we included a sheet in Supplementary Table 2, sample replicates tab, that summarizes the number of replicates per habitat and moth. Also, we provided an explanation of the number of replicates and directed the reader to Supplementary Table 2, Sample_replication sheet.

Line 471-472, seems like something is missing here or there is odd spacing.

Author's response: Thanks, we resolved this comment.

Line 540, the Bonferroni method is generally though to be too extreme a method in



metabolomics in that it might get rid of data that was actually useful and make a type 2 error. Is there a specific reason this method was used over others.

Author's response: Thanks for your comment. The Bonferroni method was not directly used in metabolomics data (i.e., m/z abundances), it was used as the multiple testing correction during the calculation of differences between the estimated diversity index (β - nearest taxon index) derived from metabolite features and metagenome assembled genomes. Also, it was used in the analysis that calculated significant differences between the same diversity index and environmental variables. We think that using the Bonferroni test for this analysis is a valid method, however given the reviewer comment we performed the same analysis but using the FDR method within the `adjust_pvalue` function of the `rstatix` R package. We obtained the same results, with exception of the correlation of the mantel test between the β - nearest taxon index derived from metabolites in the palsa and changes in soil temperature. Here the Bonferroni method found a non-significant p-value while the FDR method found it significant.

Regarding how the Bonferroni method is different from the FDR approach and why we want a more conserved approach, we believe that the Bonferroni method is focused on controlling the false positive controlling rate (Family wise error rate = probability (FalsePositive $\geq$ 1))⁴⁸. It is a strict method that penalizes false positives but potentially increases the false negative rate. On the other hand, the FDR method controls the false discovery rate (FalseDiscoveryRate (FDR)=Expected (FalsePositive/ (FalsePositive+TruePositive)))⁴⁸. Choosing which method to use will depend on the number of comparisons being tested and the tolerance for false positives which will depend on specific questions being asked by the researcher⁴⁹. Since we were interested in finding potential environmental drivers of metabolite assembly processes, we decided to use a strict approach to avoid potential false positives mediated by causal interactions. However, as mentioned above, the main finding of our correlations was supported by both methods.

Similar to the above why was a Mann–Whitney U test used over parametric alternatives

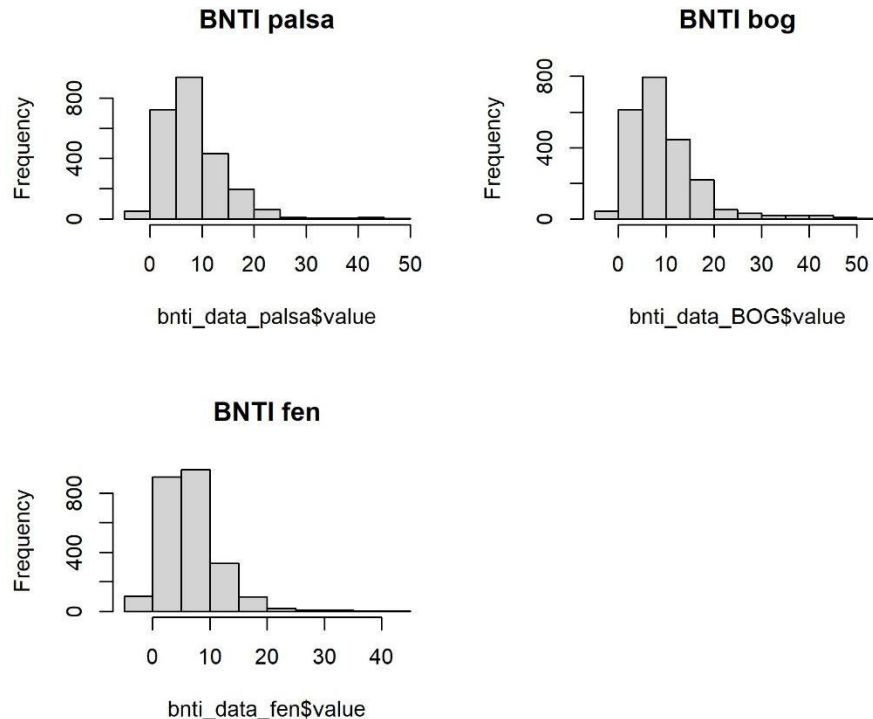
Author's response: Thanks for your comment. The estimated diversity index (β - nearest taxon index) derived from metabolite features and metagenome assembled genomes follows a non-normal distribution which is the main reason to use a non-parametric test over parametric alternatives. We tested the distribution of this index using the Shapiro-Wilk Test of Normality whose p- value was lower than 0.05 indicating that the data significantly deviated from a

47



Open Access This file is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In the cases where the authors are anonymous, such as is the case for the reports of anonymous peer reviewers, author attribution should be to 'Anonymous Referee' followed by a clear attribution to the source work. The images or other third party material in this file are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

normal distribution. The histogram below represents the skewed distribution of metabolite derived β - nearest taxon index from palsa, bog and fen habitats.



Line 602, what do you mean “approach described by 46” this is not correct English. Unless this is a journal specific issue I suggest changing to something like - ...using a previously described approach 46...

Author’s response: We apologize for this oversight. We have now corrected the sentence.

The new sentence reads as follows: **L750** “ A hierarchical cluster analysis of metabolites (palsa n = 29, bog n = 27, fen n = 29) was performed in R using the package *cluster* version 2.1.3, using a modification of a previously described approach⁵⁰ but using β BNTI_{-feature} value estimated in previous section instead of metabolite abundances. ”

Line 605 why the closed brackets after the words daisy and nun? I know it refers to the name of the program/package but are they actually needed? Fine if they are, I am just curious.



Author's response: Thanks, we resolved this comment. The parentheses are not needed. We deleted them.

Thank you for the comprehensive details of what code was used and where the code and data can be downloaded from. It is nice to see such good practice in data sharing.

Author's response: Thank you so much for this comment! We strive to adhere to transparent data protocols diligently.

References

1. N. Morrison et al., Standard reporting requirements for biological samples in metabolomics experiments: environmental context. *Metabolomics* 3, 203-210 (2007).
2. O. A. H. Jones et al., Metabolomic analysis of soil communities can be used for pollution assessment. *Environmental Toxicology and Chemistry* 33, 61-64 (2014).
3. R. M. Shah et al., Functional analysis of pristine estuarine marine sediments. *Science of The Total Environment* 781, 146526 (2021).
4. R. M. Shah et al., Omics-based ecosurveillance uncovers the influence of estuarine macrophytes on sediment microbial function and metabolic redundancy in a tropical ecosystem. *Science of The Total Environment* 809, 151175 (2022).

References used in the response to reviewer document

1. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**, 1043-1055 (2015).
2. Parks DH, *et al.* Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nature Microbiology* **2**, 1533-1542 (2017).



3. Eren AM, *et al.* Community-led, integrated, reproducible multi-omics with anvi'o. *Nature Microbiology* **6**, 3-6 (2021).
4. Aroney STN, Newell, R. J. P., Nissen, J., Camargo, A. P., Tyson, G. W., & Woodcroft, B. J. CoverM: Read coverage calculator for metagenomics (v0.7.0). Zenodo (2024).
5. Woodcroft BJ, *et al.* Genome-centric view of carbon processing in thawing permafrost. *Nature* **560**, 49-54 (2018).
6. Stegen JC, Lin X, Fredrickson JK, Konopka AE. Estimating and mapping ecological processes influencing microbial community assembly. *Frontiers in Microbiology* **6**, (2015).
7. Stegen JC, Lin X, Konopka AE, Fredrickson JK. Stochastic and deterministic assembly processes in subsurface microbial communities. *The ISME journal* **6**, 1653-1664 (2012).
8. Danczak RE, *et al.* Using metacommunity ecology to understand environmental metabolomes. *Nature communications* **11**, 1-16 (2020).
9. Zhou J, *et al.* Stochasticity, succession, and environmental perturbations in a fluidic ecosystem. *Proceedings of the National Academy of Sciences* **111**, E836-E845 (2014).
10. Stegen JC, *et al.* Quantifying community assembly processes and identifying features that impose them. *The ISME journal* **7**, 2069-2079 (2013).
11. Pester M, Knorr K-H, Friedrich MW, Wagner M, Loy A. Sulfate-reducing microorganisms in wetlands – fameless actors in carbon cycling and climate change. *Frontiers in Microbiology* **3**, (2012).
12. Blodau C, Mayer B, Peiffer S, Moore TR. Support for an anaerobic sulfur cycle in two Canadian peatland soils. *Journal of Geophysical Research: Biogeosciences* **112**, (2007).



13. Bottrell SH, Hatfield D, Bartlett R, Spence MJ, Bartle KD, Mortimer RJG. Concentrations, sulfur isotopic compositions and origin of organosulfur compounds in pore waters of a highly polluted raised peatland. *Org Geochem* **41**, 55-62 (2010).
14. Wilson RM, *et al.* Plant organic matter inputs exert a strong control on soil organic matter decomposition in a thawing permafrost peatland. *Sci Total Environ* **820**, 152757 (2022).
15. Angle JC, *et al.* Methanogenesis in oxygenated soils is a substantial fraction of wetland methane emissions. *Nature Communications* **8**, 1567 (2017).
16. Urban NR, Eisenreich SJ, Grigal DF. Sulfur cycling in a forested *Sphagnum* bog in northern Minnesota. *Biogeochemistry* **7**, 81-109 (1989).
17. Herndon E, Richardson J, Carrell AA, Pierce E, Weston D. Sulfur speciation in *Sphagnum* peat moss modified by mutualistic interactions with cyanobacteria. *New Phytologist* **241**, 1998-2008 (2024).
18. Wilson RM, Tfaily MM. Advanced molecular techniques provide new rigorous tools for characterizing organic matter quality in complex systems. *Journal of Geophysical Research: Biogeosciences* **123**, 1790-1795 (2018).
19. Fofana A, *et al.* Mapping substrate use across a permafrost thaw gradient. *Soil Biol Biochem*, 108809 (2022).
20. Fakhraee M, Li J, Katsev S. Significant role of organic sulfur in supporting sedimentary sulfate reduction in low-sulfate environments. *Geochim Cosmochim Acta* **213**, 502-516 (2017).
21. Candry P, Abrahamson B, Stahl DA, Winkler M-KH. Microbially mediated climate feedbacks from wetland ecosystems. *Global Change Biol* **29**, 5169-5183 (2023).
22. AminiTabrizi R, *et al.* Controls on soil organic matter degradation and subsequent greenhouse gas emissions across a permafrost thaw gradient in Northern Sweden. *Frontiers in Earth Science* **8**, (2020).



23. Gauci V, *et al.* Sulfur pollution suppression of the wetland methane source in the 20th and 21st centuries. *Proceedings of the National Academy of Sciences* **101**, 12583-12587 (2004).
24. Alperin MJ, Hoehler TM. Anaerobic methane oxidation by archaea/sulfate-reducing bacteria aggregates: 1. Thermodynamic and physical constraints. *Am J Sci* **309**, 869-957 (2009).
25. Moran JJ, Beal EJ, Vrentas JM, Orphan VJ, Freeman KH, House CH. Methyl sulfides as intermediates in the anaerobic oxidation of methane. *Environ Microbiol* **10**, 162-173 (2008).
26. McGlynn SE, Chadwick GL, Kempes CP, Orphan VJ. Single cell activity reveals direct electron transfer in methanotrophic consortia. *Nature* **526**, 531-535 (2015).
27. Graham EB, Hofmockel KS. Ecological stoichiometry as a foundation for omics-enabled biogeochemical models of soil organic matter decomposition. *Biogeochemistry* **157**, 31-50 (2022).
28. Zhou J, Ning D. Stochastic community assembly: does it matter in microbial ecology? *Microbiol Mol Biol Rev* **81**, e00002-00017 (2017).
29. Dini-Andreote F, Stegen JC, Van Elsas JD, Salles JF. Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. *Proceedings of the National Academy of Sciences* **112**, E1326-E1332 (2015).
30. Doherty SJ, *et al.* The Transition From Stochastic to Deterministic Bacterial Community Assembly During Permafrost Thaw Succession. *Frontiers in Microbiology* **11**, (2020).
31. Mondav R, *et al.* Microbial network, phylogenetic diversity and community membership in the active layer across a permafrost thaw gradient. *Environmental microbiology* **19**, 3201-3218 (2017).
32. Hodgkins SB, *et al.* Changes in peat chemistry associated with permafrost thaw increase greenhouse gas production. *Proceedings of the National Academy of Sciences* **111**, 5819-5824 (2014).



33. Schellekens J, *et al.* Preferential degradation of polyphenols from *Sphagnum*–4-Isopropenylphenol as a proxy for past hydrological conditions in *Sphagnum*-dominated peat. *Geochim Cosmochim Acta* **150**, 74-89 (2015).
34. Hodgkins SB, *et al.* Elemental composition and optical properties reveal changes in dissolved organic matter along a permafrost thaw chronosequence in a subarctic peatland. *Geochim Cosmochim Acta* **187**, 123-140 (2016).
35. Barrow MP, Burkitt WI, Derrick PJ. Principles of Fourier transform ion cyclotron resonance mass spectrometry and its application in structural biology. *Analyst* **130**, 18-28 (2005).
36. Bahureksa W, *et al.* Soil Organic Matter Characterization by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FTICR MS): A Critical Review of Sample Preparation, Analysis, and Data Interpretation. *Environ Sci Technol* **55**, 9637-9656 (2021).
37. González-Domínguez R, Sayago A, Fernández-Recamales Á. Direct infusion mass spectrometry for metabolomic phenotyping of diseases. *Bioanalysis* **9**, 131-148 (2017).
38. Koch BP, Dittmar T, Witt M, Kattner G. Fundamentals of Molecular Formula Assignment to Ultrahigh Resolution Mass Data of Natural Organic Matter. *Analytical Chemistry* **79**, 1758-1763 (2007).
39. Tfaily MM, *et al.* Sequential extraction protocol for organic matter from soils and sediments using high resolution mass spectrometry. *Anal Chim Acta* **972**, 54-61 (2017).
40. Stenson AC, Landing WM, Marshall AG, Cooper WT. Ionization and Fragmentation of Humic Substances in Electrospray Ionization Fourier Transform-Ion Cyclotron Resonance Mass Spectrometry. *Analytical Chemistry* **74**, 4397-4409 (2002).
41. Tolić N, *et al.* Formularity: software for automated formula assignment of natural and other organic matter from ultrahigh-resolution mass spectra. *Analytical chemistry* **89**, 12659-12665 (2017).



42. Kind T, Fiehn O. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics* **8**, 105 (2007).
43. Tfaily MM, Hess NJ, Koyama A, Evans RD. Elevated [CO₂] changes soil organic matter composition and substrate diversity in an arid ecosystem. *Geoderma* **330**, 1-8 (2018).
44. Morrison N, *et al.* Standard reporting requirements for biological samples in metabolomics experiments: environmental context. *Metabolomics* **3**, 203-210 (2007).
45. Yuan MM, *et al.* Climate warming enhances microbial network complexity and stability. *Nature Climate Change* **11**, 343-348 (2021).
46. Zhou J, Deng Y, Luo F, He Z, Tu Q, Zhi X. Functional molecular ecological networks. *MBio* **1**, e00169-00110 (2010).
47. Bauermeister A, Mannocho-Russo H, Costa-Lotufo LV, Jarmusch AK, Dorrestein PC. Mass spectrometry-based metabolomics in microbiome investigations. *Nat Rev Microbiol* **20**, 143-160 (2022).
48. Jafari M, Ansari-Pour N. Why, When and How to Adjust Your P Values? *Cell J* **20**, 604-607 (2019).
49. Krzywinski M, Altman N. Comparing samples—part II. *Nature Methods* **11**, 355-356 (2014).
50. Merder J, *et al.* Dissolved organic compounds with synchronous dynamics share chemical properties and origin. *Limnol Oceanogr* **66**, 4001-4016 (2021).

Decision Letter, first revision:



Open Access This file is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In the cases where the authors are anonymous, such as is the case for the reports of anonymous peer reviewers, author attribution should be to 'Anonymous Referee' followed by a clear attribution to the source work. The images or other third party material in this file are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

Message: Our ref: NMICROBIOL-23123579A

18th June 2024

Dear Malak,

Thank you for your patience as we've prepared the guidelines for final submission of your Nature Microbiology manuscript, "Multi-Omics Unveils Feature-Level Microbe-Metabolite Linkages Driving Permafrost Greenhouse Gas Dynamics" (NMICROBIOL-23123579A). Please carefully follow the step-by-step instructions provided in the attached file, and add a response in each row of the table to indicate the changes that you have made. Please also check and comment on any additional marked-up edits we have proposed within the text. Ensuring that each point is addressed will help to ensure that your revised manuscript can be swiftly handed over to our production team.

We would like to start working on your revised paper, with all of the requested files and forms, as soon as possible (preferably within two weeks). Please get in contact with us if you anticipate delays.

When you upload your final materials, please include a point-by-point response to any remaining reviewer comments.

If you have not done so already, please alert us to any related manuscripts from your group that are under consideration or in press at other journals, or are being written up for submission to other journals (see: <https://www.nature.com/nature-research/editorial-policies/plagiarism#policy-on-duplicate-publication> for details).

In recognition of the time and expertise our reviewers provide to Nature Microbiology's editorial process, we would like to formally acknowledge their contribution to the external peer review of your manuscript entitled "Multi-Omics Unveils Feature-Level Microbe-Metabolite Linkages Driving Permafrost Greenhouse Gas Dynamics". For those reviewers who give their assent, we will be publishing their names alongside the published article.

Nature Microbiology offers a Transparent Peer Review option for new original research manuscripts submitted after December 1st, 2019. As part of this initiative, we encourage our authors to support increased transparency into the peer review process by agreeing to have the reviewer comments, author rebuttal letters, and editorial decision letters published as a Supplementary item. When you submit your final files please clearly state in your cover letter whether or not you would like to participate in this initiative. Please note that failure to state your preference will result in delays in accepting your manuscript for publication.

Cover suggestions

55



Open Access This file is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In the cases where the authors are anonymous, such as is the case for the reports of anonymous peer reviewers, author attribution should be to 'Anonymous Referee' followed by a clear attribution to the source work. The images or other third party material in this file are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

COVER ARTWORK: We welcome submissions of artwork for consideration for our cover. For more information, please see our [guide for cover artwork](#).

Nature Microbiology has now transitioned to a unified Rights Collection system which will allow our Author Services team to quickly and easily collect the rights and permissions required to publish your work. Approximately 10 days after your paper is formally accepted, you will receive an email in providing you with a link to complete the grant of rights. If your paper is eligible for Open Access, our Author Services team will also be in touch regarding any additional information that may be required to arrange payment for your article.

Please note that you will not receive your proofs until the publishing agreement has been received through our system.

Please note that *Nature Microbiology* is a Transformative Journal (TJ). Authors may publish their research with us through the traditional subscription access route or make their paper immediately open access through payment of an article-processing charge (APC). Authors will not be required to make a final decision about access to their article until it has been accepted. [Find out more about Transformative Journals](#)

Authors may need to take specific actions to achieve [compliance](#) with funder and institutional open access mandates. If your research is supported by a funder that requires immediate open access (e.g. according to [Plan S principles](#)) then you should select the gold OA route, and we will direct you to the compliant route where possible. For authors selecting the subscription publication route, the journal's standard licensing terms will need to be accepted, including [self-archiving policies](#). Those licensing terms will supersede any other terms that the author or any third party may assert apply to any version of the manuscript.

For information regarding our different publishing models please see our [Transformative Journals](#) page. If you have any questions about costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com.

Reviewer #1:
Remarks to the Author:
The authors have addressed all my comments.



Reviewer #2:

Remarks to the Author:

It was a pleasure to read the revision and rebuttal for this manuscript. The authors have worked to carefully address the concerns I posed (in all but one respect, below) to significantly improve the clarity of their paper's contribution. There are meaningful additions (e.g. the role of sphagnum derived organosulfur compounds) which help explain some of the patterns observed, and there are thoughtful comments and revision. In my view this paper will make a very helpful contribution to the research landscape, bringing the links between microbes, metabolites, and Arctic change into clear focus.

Minor issue: I previously highlighted missing data within supplementary table 6. The authors offer assurance that this has been addressed, but ST6 in the merged manuscript comprises just a description of the tabs of their spreadsheet. This is probably a compilation error, but it would need to be addressed.

Reviewer #3:

Remarks to the Author:

I thank the authors for their comprehensive response to my queries and those of the other reviewers. The paper is very interesting.

I think the author have made a great deal of effort in response to my comments and they have addressed all my concerns except one which is that, in my view and experience, it is not true that LC-MS based methods can be biased towards highly abundant compounds with MS/MS spectra. Any bias is down to the analytical and statistical methods used in a study, not the instrument which is quite capable of picking up huge numbers of compounds. While it is true that an LC-QQQ-MS has better sensitivity towards compounds that you know are there, an LC-QTOF-MS is better for working out structures and identities of unknown compounds. That is why LC-QTOF-MS is so widely used in environmental science, for example to detect unknown toxicants. You can also use tools such as MS Mine for detailed data processing of the raw files. As I do not think it is correct to say that 'LC-MS might not see low abundance peaks' I'd ask the authors to consider rephrasing. Perhaps to something like..."can be biased towards highly abundant compounds if appropriate care is not taken and this increases the analysis time" (or something similar).
--

I also have a few small comments/suggestions for consideration.

- Perhaps when talking about the quality control and metadata the authors could state in the manuscript that they believe this fulfils the standard reporting requirements for biological samples in metabolomics experiments in the biological or environmental context (<https://link.springer.com/article/10.1007/s11306-007-0067-1>)

57



Open Access This file is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In the cases where the authors are anonymous, such as is the case for the reports of anonymous peer reviewers, author attribution should be to 'Anonymous Referee' followed by a clear attribution to the source work. The images or other third party material in this file are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

- Line L490 - I don't think it is correct to say that using organic solvents would introduce additional compounds that would not naturally be present. Any compounds there would be already present naturally, they just might not be extracted by water. Can I suggest just saying "Our goal was to mimic the natural thaw conditions as closely as possible". And leaving it at that?

- I'd suggest maybe a little more discussion on the potential disadvantages of DI-FTICR MS.

Author Rebuttal, first revision:

Reviewer #1:

Remarks to the Author:

The authors have addressed all my comments.

Reviewer #2:

Remarks to the Author:

It was a pleasure to read the revision and rebuttal for this manuscript. The authors have worked to carefully address the concerns I posed (in all but one respect, below) to significantly improve the clarity of their paper's contribution. There are meaningful additions (e.g. the role of sphagnum derived organosulfur compounds) which help explain some of the patterns observed, and there are thoughtful comments and revision. In my view this paper will make a very helpful contribution to the research landscape, bringing the links between microbes, metabolites, and Arctic change into clear focus.

Minor issue: I previously highlighted missing data within supplementary table 6. The authors offer assurance that this has been addressed, but ST6 in the merged manuscript comprises just a description of the tabs of their spreadsheet. This is probably a compilation error, but it would need to be addressed.

Author's response: We thank the reviewer for their support and for taking the time to read the revision. We highly appreciated that. With respect to supplementary Table 6 (now Supplementary table 5), the file in excel format includes the Spearman rho value, p value and p value adjusted of

58



Open Access This file is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In the cases where the authors are anonymous, such as is the case for the reports of anonymous peer reviewers, author attribution should be to 'Anonymous Referee' followed by a clear attribution to the source work. The images or other third party material in this file are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

each genome that correlated with greenhouse gases (CO₂ and CH₄). This table also includes the taxonomy corresponding to each genome and the results of the feature-level analysis. However, it seems that the error occurred when transforming the excel file to PDF which influenced the merged file. We will make sure to correct this compilation error.

Reviewer #3:

Remarks to the Author:

I thank the authors for their comprehensive response to my queries and those of the other reviewers. The paper is very interesting.

I think the author have made a great deal of effort in response to my comments and they have addressed all my concerns except one which is that, in my view and experience, it is not true that LC-MS based methods can be biased towards highly abundant compounds with MS/MS spectra. Any bias is down to the analytical and statistical methods used in a study, not the instrument which is quite capable of picking up huge numbers of compounds. While it is true that an LC-QQQ-MS has better sensitivity towards compounds that you know are there, an LC-QTOF-MS is better for working out structures and identities of unknown compounds. That is why LC-QTOF-MS is so widely used in environmental science, for example to detect unknown toxicants. You can also use tools such as MS Mine for detailed data processing of the raw files. As I do not think it is correct to say that 'LC-MS might not see low abundance peaks' I'd ask the authors to consider rephrasing. Perhaps to something like..."can be biased towards highly abundant compounds if appropriate care is not taken and this increases the analysis time" (or something similar).

Author's response: We thank the reviewer for their comments on our manuscript. We have included the author's suggestions in L416. The new text reads as follows:

"While chromatography-based methods like LC-MS/MS enable more accurate structural elucidation, they can be biased towards highly abundant compounds if appropriate care is not taken which increases the analysis time."

I also have a few small comments/suggestions for consideration.

- Perhaps when talking about the quality control and metadata the authors could state in the manuscript that they believe this fulfils the standard reporting requirements for biological samples in metabolomics experiments in the biological or environmental context

(<https://link.springer.com/article/10.1007/s11306-007-0067-1><<https://link.springer.com/article/10.1007/s11306-007-0067-1>>)



Author's response: We thank the reviewer for their comments on our manuscript. We have included the author's suggestions in L399. The new text reads as follows:

"Details about sampling date, time of sample collection, GPS location, air and soil temperature, active layer depth, and notes related to the sampling can be found at the EMERGE database <https://emerge-db.asc.ohio-state.edu/datasources/12>. For further clarity, we included this information in Supplementary 2, sample metadata tab. The authors believe that this data is in concordance with the requirements described in ⁹⁰"

- Line L490 - I don't think it is correct to say that using organic solvents would introduce additional compounds that would not naturally be present. Any compounds there would be already present naturally, they just might not be extracted by water. Can I suggest just saying "Our goal was to mimic the natural thaw conditions as closely as possible". And leaving it at that?

Author's response: We agree with the author's comment, and we eliminated the corresponding sentence. The new text reads as follows (L423):

"In this study, we used only water as an extractant to focus on the bioavailable compounds actively cycling with microbes in this dynamic thawing environment. Our goal was to mimic the natural thaw conditions as closely as possible"

- I'd suggest maybe a little more discussion on the potential disadvantages of DI-FTICR MS.

Author's response: We thank the reviewer for this comment. We moved some of our responses on the potential disadvantages of DI-FT-ICR MS from our previous rebuttal to the main text in the current revised draft. The new text reads as follows (L346).

"Future work aims to elucidate causal mechanisms through complementary methods like mapping multi-omics data into metabolic pathways using chromatographic techniques (e.g. LC-MS/MS) and microbial isolation studies. We acknowledge the limitations of relying solely on DI-FTICR-MS for metabolite identification, including the inability to differentiate isomers, or determine molecular structures, as well as issues with signal suppression or enhancement ^{80, 81}. Moving forward, implementing this technique alongside chromatographic techniques could provide better understanding of the complex processes in natural ecosystems."

Other than the above, since the authors have addressed all my comments and the journal has asked for a very quick turnaround I think the paper can go forward.

Author's response: Thank you so much for your time. We highly appreciated all your feedback.



Final Decision Letter:

Message: 30th July 2024

Dear Malak,

I am pleased to accept your Article "Microbiome-metabolite linkages drive greenhouse gas dynamics over a permafrost thaw gradient" for publication in Nature Microbiology. Thank you for having chosen to submit your work to us and many congratulations.

Over the next few weeks, your paper will be copyedited to ensure that it conforms to Nature Microbiology style. We look particularly carefully at the titles of all papers to ensure that they are relatively brief and understandable.

Once your paper is typeset, you will receive an email with a link to choose the appropriate publishing options for your paper and our Author Services team will be in touch regarding any additional information that may be required. Once your paper has been scheduled for online publication, the Nature press office will be in touch to confirm the details.

You may wish to make your media relations office aware of your accepted publication, in case they consider it appropriate to organize some internal or external publicity. Once your paper has been scheduled you will receive an email confirming the publication details. This is normally 3-4 working days in advance of publication. If you need additional notice of the date and time of publication, please let the production team know when you receive the proof of your article to ensure there is sufficient time to coordinate. Further information on our embargo policies can be found here:
<https://www.nature.com/authors/policies/embargo.html>

After the grant of rights is completed, you will receive a link to your electronic proof via email with a request to make any corrections within 48 hours. If, when you receive your proof, you cannot meet this deadline, please inform us at rjsproduction@springernature.com immediately. You will not receive your proofs until the publishing agreement has been received through our system

Due to the importance of these deadlines, we ask you please us know now whether you will be difficult to contact over the next month. If this is the case, we ask you provide us with the contact information (email, phone and fax) of someone who will be able to check the proofs on your behalf, and who will be available to address any last-minute problems.

Acceptance of your manuscript is conditional on all authors' agreement with our publication policies (see <https://www.nature.com/nmicrobiol/editorial-policies>). In particular your manuscript must not be published elsewhere.

61



Open Access This file is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In the cases where the authors are anonymous, such as is the case for the reports of anonymous peer reviewers, author attribution should be to 'Anonymous Referee' followed by a clear attribution to the source work. The images or other third party material in this file are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

Please note that *Nature Microbiology* is a Transformative Journal (TJ). Authors may publish their research with us through the traditional subscription access route or make their paper immediately open access through payment of an article-processing charge (APC). Authors will not be required to make a final decision about access to their article until it has been accepted. [Find out more about Transformative Journals](#)

Authors may need to take specific actions to achieve [compliance](#) with funder and institutional open access mandates. If your research is supported by a funder that requires immediate open access (e.g. according to [Plan S principles](#)) then you should select the gold OA route, and we will direct you to the compliant route where possible. For authors selecting the subscription publication route, the journal's standard licensing terms will need to be accepted, including [self-archiving policies](#). Those licensing terms will supersede any other terms that the author or any third party may assert apply to any version of the manuscript.

If you have any questions about our publishing options, costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com

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

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

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

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



As soon as your article is published, you will receive an automated email with your shareable link.

With kind regards,



Open Access This file is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In the cases where the authors are anonymous, such as is the case for the reports of anonymous peer reviewers, author attribution should be to 'Anonymous Referee' followed by a clear attribution to the source work. The images or other third party material in this file are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.