MetFish: A Metabolomics Pipeline for Studying Microbial Communities in Chemically Extreme Environments

Thomas Metz, Chengdong Xu, Sneha Couvillion, Ryan Sontag, Nancy Isern, Yukari Maezato, Stephen Lindemann, Taniya Roy Chowdhury, Rui Zhao, Beau Morton, Rosalie Chu, Ronald Moore, Janet Jansson, Vanessa BAiley, Paula J. Mouser, Margie Romine, and James Fredrickson

Corresponding Author(s): Thomas Metz, Pacific Northwest National Laboratory

Review Timeline:

Submission Date:	October 15, 2020
Editorial Decision:	December 23, 2020
Revision Received:	April 2, 2021
Accepted:	April 16, 2021

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

December 23, 2020

Dr. Thomas Metz
Pacific Northwest National Laboratory
Richland

Re: mSystems01058-20 (MetFish: A Metabolomics Platform for Studying Microbial Communities in Chemically Extreme Environments)

Dear Dr. Thomas Metz:

Your manuscript has been reviewed by experts in the field and their recommendation is minor modifications

Below you will find the comments of the reviewers.

To submit your modified manuscript, log onto the eJP submission site at https://msystems.msubmit.net/cgi-bin/main.plex. If you cannot remember your password, click the "Can't remember your password?" link and follow the instructions on the screen. Go to Author Tasks and click the appropriate manuscript title to begin the resubmission process. The information that you entered when you first submitted the paper will be displayed. Please update the information as necessary. Provide (1) point-by-point responses to the issues raised by the reviewers as file type "Response to Reviewers," not in your cover letter, and (2) a PDF file that indicates the changes from the original submission (by highlighting or underlining the changes) as file type "Marked Up Manuscript - For Review Only."

Due to the SARS-CoV-2 pandemic, our typical 60 day deadline for revisions will not be applied. I hope that you will be able to submit a revised manuscript soon, but want to reassure you that the journal will be flexible in terms of timing, particularly if experimental revisions are needed. When you are ready to resubmit, please know that our staff and Editors are working remotely and handling submissions without delay. If you do not wish to modify the manuscript and prefer to submit it to another journal, please notify me of your decision immediately so that the manuscript may be formally withdrawn from consideration by mSystems.

If your manuscript is accepted for publication, you will be contacted separately about payment when the proofs are issued; please follow the instructions in that e-mail. Arrangements for payment must be made before your article is published. For a complete list of **Publication Fees**, including supplemental material costs, please visit our website.

Corresponding authors may join or renew ASM membership to obtain discounts on publication fees. Need to upgrade your membership level? Please contact Customer Service at Service@asmusa.org.

Thank you for submitting your paper to mSystems.

Sincerely,

Marcelino Gutierrez

Editor, mSystems

Journals Department
American Society for Microbiology
1752 N St., NW
Washington, DC 20036
E-mail: peerreview@asmusa.org
Phone: 1-202-942-9338

Reviewer comments:

Reviewer #1 (Comments for the Author):

The manuscript by Xu et al details the development and validation of MetFish. MetFish is a pipeline for chemical derivatization of specific functional groups, namely, amines, carboxylates, carbonyls, and alcohols with dansyl based reagents from high salt concentration biological samples. This derivatization allows for increased detection capabilities after the samples have passed through an inline SPE prior to LC-MS(/MS) analysis. Overall this was an exciting paper that could have broader implications for metabolomics from high salinity or even non-high salinity samples. The premise of being able to incorporate heavy labels for metabolomics is interesting as TMT and iTRAQ have become so popular in proteomics, I would like to thank the authors for drawing this comparison so clearly within the text. Overall I think this paper will greatly appeal to the readership of mSystems and this was a very well written manuscript with broadly applicable applications. Line 335, thank you for including this not about reagents not being used in combination, this manuscript was very explicit about certain precautions which will aid in its adoptability from the community. Thank you for depositing the data to MassIVE! Below please find a list of major and minor critiques:
Major
1. Line 28, use of 'first time' while I appreciate that there aren't many studies dealing with high salinity samples for LC-MS/MS there have been a few studies dealing with dissolved organic matter (DOM) analyses from the ocean. This seems like an overstatement when one takes these other studies into account, which definitely had alternative approach, but still made the claim to be measuring exometabolites from high salt samples, specifically doi.org/10.3389/fmars.2017.00405 and https://doi.org/10.26434/chemrxiv.9817133.v3 Along these lines, line 92, do the authors have further literature to cite for this statement about hyper saline? What defines these levels and how have people tested this?
2. Could the authors expand on the carbonyl chemistry? Does this include ketones, aldehydes, esters? Or specifically just ketones, R groups were used in Figure 1 and it might help to further clarify for the non-chemistry based audience of mSystems.
3. Line 99, Use of the E coli Metabolome and Plantcyc may be extremely limited considering the authors mention natural products. E. coli is not known to produce much in the way of natural products, this is likely beyond the scope of the paper but this reviewer would be curious to know how these numbers compare using a database like NP Atlas (https://www.npatlas.org/joomla/)
4. Line 167, 234 is mentioned on line 169 but not 167? Which also doesn't seem to match Table S2. The numbers should be consistent for recommendations for others to follow the protocol/analysis.
5. Table S1, how does one recover more than 100% in the eluent, there are at least three values that are very high?

6. Why was Cys absent from the amino acid analysis? This seems like a ubiquitous amino acid to not attempt to quantify as well.

7. Could the authors comment on labelling efficiency? Name are amino acids, like Lys in the case of amine labelling labeled twice? Could this complicate the subsequent quantification of this amino acid? How might a user of this platform anticipate or control the Rxn. This seems like an important caveat to mention.

8. Regarding Figure 4A, were the other 5 amino acids below the limit of detection? Why were only 14 analyzed, could the authors please explain this a little further. Along these lines, were these biological or technical triplicates, not totally clear from the figure legend.

9. Line 314, Were the authors able to find any other metabolites that were labelled, perhaps using something like MS2LDA (http://ms2lda.org/) might be an interesting tool to leverage for untargeted discovery given how consistently these dansyl probes fragment and would be largely unnatural to a system. This might be a creative way to probe for unexpected biological results.

Minor

1. The importance statement was almost verbatim the abstract, these should be two separate statements.

2. Define QDA at first use in main text

3. Figure 1, the fishing hook and compound would make a cute TOC graphic but is not needed withing the main figure

4. Line 164, Regarding Leu and Ile, the Hicks lab has published a nice review that does touch on how these two amino acids could be differentiated in the context of antimicrobial peptides: https://doi.org/10.1039/D0NP00046A

5. Could the authors please clarify or explicitly state 2-C13 for Gly? Or the numbers of carbon 13 labels for other amino acids.

6. Figure 3, are these transitions for metabolite quantification? It is unclear what exactly the authors have listed on what I assume are EICs.

7. Line 240, µM or nM not nm?

8. Figure 5, Line 345, please consider using Precursor- product ion, rather than parent- product ion. Parent and Daughter ion terminology is largely outdated.

9. Supplemental Table S9 might be cut off? Which of these ions ended up as identified, could this information please be included in the table.

10. The discussion read as more of a conclusion, consider relabeling the sections of the manuscript.

Reviewer #2 (Comments for the Author):

The manuscript entitled "MetFish: A Metabolomics Platform for Studying Microbial Communities in Chemically Extreme Environments" by Xu and collaborators, submitted as original research presents an LC-MS-based method for analysis of diversity of samples with high concentration of dissolved salts that cannot analyzed by conventional approaches.

The approach, as presented, shows potential for application to a diverse of samples due to its versatility and affordability. Overall, the approach is well supported based on the data and it is totally pertinent and useful for the field of metabolomics and others.

I might have a few comments and suggestions that might help to improve the current version if the authors want to consider them:

Title: After reading the manuscript and the SI material, I consider this work as a pipeline or a method, more than an actual platform. The whole concept of MetFish is applying tagging and extraction protocols suitable for LC-MS acquisition targeting specific functionalities from metabolite

moieties (functional groups). What do the authors have to further support this work to offer it as a platform?

Line 104: The following item "1) could be used by researchers with diverse skill sets studying myriad sample types" basically covers "everything" and I don't see how this sentence is relevant. If "1)" is removed, the following "2)" and "3)" justifications are enough to support the author's point about the reagents used in MetFish. Another option, instead of removing "1)" might be rephrasing it to something like "1) could be used to study diverse sample types based on the functional groups of interest";

Line 109: Correct to "three types of samples", as based on the results the study was performed on three sample types with proper replicates for each of them, even being three replicates these are not actually just "three samples" here;

Line 131: Clarify what the "QDA" abbreviation stands for and cite again the paper from where this reference comes from, since QDA (N-[2-(Aminooxy)ethyl]-N,N-dimethyl-1-dodecylammonium iodide) is only mentioned in the SI;

Line 135: Check punctuation "(ᵤ) groups. . Recognizing (ᵤ)";

Line 164-165: In the following "To illustrate this (ᵤ)", I feel this paragraph is not connected to the previous sentence, which talks about Leu and Ile. Maybe use another connector instead of "To illustrate this (ᵤ)", unless the authors want to actually illustrate the lack of unique fragments produced by CID of Leu and Ile vs Gly;

Line 192: "(ᵤ) and is available for study": This seems kind of obvious to me. What do the authors mean here, that the microbial mat was easy to study? Everything is available for study, if the efforts are properly set for that. Please, clarify the idea;

Line 212 (Figure 2): I suggest to include the y-axis (intensity) in all the chromatograms, as well as clarify that (b) are Extracted Ion Chromatograms, if I assume correctly based on the figure;

Line 227 (Figure 3): considering these are also EIC (or XIC, extracted ion chromatograms), do the authors have any comment on the additional ions observed, e.g. butanetriol and peak shape, e.g 2-deoxy-D-ribose? Maybe include the m/z range used for selecting the EIC. It would be also worthy to clarify these are EIC from SRM;

Line 278-281: Can the authors elaborate a little bit more here to discuss or hypothesize the reason behind Ser and Pro results? Since these were unexpected results, as the authors suggest, it would be valuable for the manuscript to include this discussion;

Line 328 (Figure 5): Can a title for this figure be included in this figure legend, for consistency? currently, only the (a) and (b) description of the figs are shown;

Line 349: In "(ᵤ) samples ranging from 86 and 154 d post-injection", are these from 86 and 154 days?

Line 360 (Figure 6): I suggest to expand a little bit the description of this figure. It is not clearly specified the statistical tools used to generate these plots. Are there somewhere else? In Fig 5a, is this a hierarchical clustering? What the distance metric used for these plots, etc. Please, provide details either here or in the methods section. These are currently missing;

Line 382-385: I would encourage the authors to include and discuss the limitations before ending the manuscript. The "Discussion" section starts with summary. Please restructure this. I would suggest to discuss the limitations before making the summary;

Also related to Lines 382-385: Additionally, GC-MS results are not discussed in any place from the main text although the results are included in the SI material (e.g. Table S8 shows identifications of metabolites from GC-MS libraries). I think that info is missing in the manuscript and should be discussed or at least briefly mentioned, as I don't see a clear reason behind including this in the SI if not used.

Line 433: "(ᵤ) as described previously" where? Please, mention where and/or provide reference to a published work if available;

Line 472: "(ᵤ) performed at speed 7 (ᵤ)", please Clarify what kind of device from the instrumentation

was used for this step;

Line 520: "(...) in-house tool MASIC", It would be worthy that the settings for these processing steps be included in the methods section;

From the SI material, I have the following comments:

Figure S1: Can the two pipelines be clearly highlighted here? Which is target vs untargeted? At least that is not clear here based on the description from the main text.

Figure S2: It might be useful to label the identified compounds from Table S8 directly in the chromatograms from Fig S2

See my previous comment regarding to "GC-MS analysis of samples with high salt content"

The manuscript by Xu et al details the development and validation of MetFish. MetFish is a pipeline for chemical derivatization of specific functional groups, namely, amines, carboxylates, carbonyls, and alcohols with dansyl based reagents from high salt concentration biological samples. This derivatization allows for increased detection capabilities after the samples have passed through an inline SPE prior to LC-MS(/MS) analysis. Overall this was an exciting paper that could have broader implications for metabolomics from high salinity or even non-high salinity samples. The premise of being able to incorporate heavy labels for metabolomics is interesting as TMT and iTRAQ have become so popular in proteomics, I would like to thank the authors for drawing this comparison so clearly within the text. Overall I think this paper will greatly appeal to the readership of mSystems and this was a very well written manuscript with broadly applicable applications. Line 335, thank you for including this not about reagents not being used in combination, this manuscript was very explicit about certain precautions which will aid in its adoptability from the community. Thank you for depositing the data to MassIVE! Below please find a list of major and minor critiques:

Major

1. Line 28, use of 'first time' while I appreciate that there aren't many studies dealing with high salinity samples for LC-MS/MS there have been a few studies dealing with dissolved organic matter (DOM) analyses from the ocean. This seems like an overstatement when one takes these other studies into account, which definitely had alternative approach, but still made the claim to be measuring exometabolites from high salt samples, specifically doi.org/10.3389/fmars.2017.00405 and https://doi.org/10.26434/chemrxiv.9817133.v3 Along these lines, line 92, do the authors have further literature to cite for this statement about hyper saline? What defines these levels and how have people tested this?

2. Could the authors expand on the carbonyl chemistry? Does this include ketones, aldehydes, esters? Or specifically just ketones, R groups were used in Figure 1 and it might help to further clarify for the non-chemistry based audience of mSystems.

3. Line 99, Use of the E coli Metabolome and Plantcyc may be extremely limited considering the authors mention natural products. E. coli is not known to produce much in the way of natural products, this is likely beyond the scope of the paper but this reviewer would be curious to know how these numbers compare using a database like NP Atlas (https://www.npatlas.org/joomla/)

4. Line 167, 234 is mentioned on line 169 but not 167? Which also doesn't seem to match Table S2. The numbers should be consistent for recommendations for others to follow the protocol/analysis.

5. Table S1, how does one recover more than 100% in the eluent, there are at least three values that are very high?

6. Why was Cys absent from the amino acid analysis? This seems like a ubiquitous amino acid to not attempt to quantify as well.

7. Could the authors comment on labelling efficiency? Name are amino acids, like Lys in the case of amine labelling labeled twice? Could this complicate the subsequent quantification of this amino acid? How might a user of this platform anticipate or control the Rxn. This seems like an important caveat to mention.

8. Regarding Figure 4A, were the other 5 amino acids below the limit of detection? Why were only 14 analyzed, could the authors please explain this a little further. Along these lines, were these biological or technical triplicates, not totally clear from the figure legend.

9. Line 314, Were the authors able to find any other metabolites that were labelled, perhaps using something like MS2LDA (http://ms2lda.org/) might be an interesting tool to leverage for untargeted discovery given how consistently these dansyl probes fragment and would be largely unnatural to a system. This might be a creative way to probe for unexpected biological results.

Minor

1. The importance statement was almost verbatim the abstract, these should be two separate statements.

2. Define QDA at first use in main text

3. Figure 1, the fishing hook and compound would make a cute TOC graphic but is not needed withing the main figure

4. Line 164, Regarding Leu and Ile, the Hicks lab has published a nice review that does touch on how these two amino acids could be differentiated in the context of antimicrobial peptides: https://doi.org/10.1039/D0NP00046A

5. Could the authors please clarify or explicitly state 2-C13 for Gly? Or the numbers of carbon 13 labels for other amino acids.

6. Figure 3, are these transitions for metabolite quantification? It is unclear what exactly the authors have listed on what I assume are EICs.

7. Line 240, µM or nM not nm?

8. Figure 5, Line 345, please consider using Precursor- product ion, rather than parent- product ion. Parent and Daughter ion terminology is largely outdated.

9. Supplemental Table S9 might be cut off? Which of these ions ended up as identified, could this information please be included in the table.

10. The discussion read as more of a conclusion, consider relabeling the sections of the manuscript.

Msystems Reviewer Comments

Reviewer #1 (Comments for the Author):

The manuscript by Xu et al details the development and validation of MetFish. MetFish is a pipeline for chemical derivatization of specific functional groups, namely, amines, carboxylates, carbonyls, and alcohols with dansyl based reagents from high salt concentration biological samples. This derivatization allows for increased detection capabilities after the samples have passed through an inline SPE prior to LC-MS(/MS) analysis. Overall this was an exciting paper that could have broader implications for metabolomics from high salinity or even non-high salinity samples. The premise of being able to incorporate heavy labels for metabolomics is interesting as TMT and iTRAQ have become so popular in proteomics, I would like to thank the authors for drawing this comparison so clearly within the text. Overall I think this paper will greatly appeal to the readership of mSystems and this was a very well written manuscript with broadly applicable applications. Line 335, thank you for including this not about reagents not being used in combination, this manuscript was very explicit about certain precautions which will aid in its adoptability from the community. Thank you for depositing the data to MassIVE! Below please find a list of major and minor critiques:

We thank Reviewer #1 for a positive and encouraging review and very helpful comments that have improved the manuscript. Please see our responses to each of the comments below.

Major
1. Line 28, use of 'first time' while I appreciate that there aren't many studies dealing with high salinity samples for LC-MS/MS there have been a few studies dealing with dissolved organic matter (DOM) analyses from the ocean. This seems like an overstatement when one takes these other studies into account, which definitely had alternative approach, but still made the claim to be measuring exometabolites from high salt samples, specifically doi.org/10.3389/fmars.2017.00405 and https://doi.org/10.26434/chemrxiv.9817133.v3 Along these lines, line 92, do the authors have further literature to cite for this statement about hyper saline? What defines these levels and how have people tested this?

We appreciate the reviewers comment here. "Hypersaline" refers to environments that have much higher salinity than typical seawater (see references: doi.org/10.1016/B978-0-12-394626-3.00006-5, doi.org/10.1016/B0-12-369396-9/00308-7). While we agree with the reviewer that there have been studies measuring metabolites in seawater, including the two examples provided by the reviewer, our method targets hypersaline samples containing considerably higher salinity than average seawater (35 PSU or

g/kg). We have shown that MetFish can be applied to samples containing 400 mM to 2 M salt which corresponds to 48.14 -240.73 g/L MgSO$_4$.
An example of a recent study involving hypersaline samples where high salt levels hindered analysis of polar metabolites doi.org/10.3389/fmars.2020.00250 is an illustration of the challenge of performing metabolomics in these types of samples.

As we have indicated in the paper, to our knowledge no method has yet been developed for enabling metabolomics analyses in such hypersaline matrices. Thus, we prefer to retain the wording of "for the first time" in the paper.

2. Could the authors expand on the carbonyl chemistry? Does this include ketones, aldehydes, esters? Or specifically just ketones, R groups were used in Figure 1 and it might help to further clarify for the non-chemistry based audience of mSystems.

We have included the following text under the chemical tagging methods for the carbonyl reagent:

"Dansylhydrazine (DNSH) is a derivatization agent used for carbonyl bonds, including ketones and aldehydes (66). It is less reactive with ester bonds (CO2R') because the carbonyl carbon of this functional group has decreased electrophilicity due to resonance stabilization."

3. Line 99, Use of the E coli Metabolome and Plantcyc may be extremely limited considering the authors mention natural products. E. coli is not known to produce much in the way of natural products, this is likely beyond the scope of the paper but this reviewer would be curious to know how these numbers compare using a database like NP Atlas (https://www.npatlas.org/joomla/)

We appreciate the reviewer bringing this database to our attention. We have used MolDB6 (https://homepage.univie.ac.at/norbert.haider/cheminf/moldb6doc.html#appendix_c) to carry out functional group searches in the databases. In the NP Atlas database, we found that 97.5% of unique database entries belong to one of the 4 functional groups targeted by MetFish. We have modified the text in the manuscript to include the results for the NP Atlas database:
"The four functional groups targeted by MetFish represent over 97%, 89% and 83% of the metabolites contained in the Natural Products Atlas, *E. coli* Metabolome and Plantcyc databases, respectively."

4. Line 167, 234 is mentioned on line 169 but not 167? Which also doesn't seem to match Table S2. The numbers should be consistent for recommendations for others to follow the protocol/analysis.

We have edited the text in the main manuscript and the contents of **Supplemental Table S2** for consistency:
"Fragment ions due only to the dansyl moiety are e.g. *m/z* 157, 170, and 252, whereas fragment ions due to dansyl-glycine are *m/z* 263 and 294. ……All metabolites that have been tagged using the dansyl chloride reagent will generate the same fragment ions (e.g. *m/z* 157, 170, and 252), providing confidence in detection of an appropriately tagged amine-containing metabolite."

5. Table S1, how does one recover more than 100% in the eluent, there are at least three values that are very high?

We thank the reviewer for noticing this. We were not able to identify the reason for why the values for acetate are greater than 100% but suspect that contamination by acetate in a buffer or solvent may have contributed. We have removed that row from Table S1.

6. Why was Cys absent from the amino acid analysis? This seems like a ubiquitous amino acid to not attempt to quantify as well.

We attempted to analyze cysteine using the amine tagging reagent but the expected product *m/z* was not observed. Though the reason is not clear, we suspect that it was due to the formation of oxidative side reactions during dansylation of SH- group of Cys.

7. Could the authors comment on labelling efficiency? Name are amino acids, like Lys in the case of amine labelling labeled twice? Could this complicate the subsequent quantification of this amino acid? How might a user of this platform anticipate or control the Rxn. This seems like an important caveat to mention.

We did not determine the labeling efficiencies of the MetFish reagents with the various metabolites demonstrated in the paper, since these efficiencies will depend on several factors, including the chemistry of the metabolite being labeled, the specific MetFish reagent, and the nature of the matrix in which the reaction is occurring. Thus, the labeling efficiencies could be very specific to the metabolite in question and fall within a broad range.

To compensate for possible variation in labeling efficiencies that may affect accurate quantification in targeted applications of MetFish, we recommend using stable-isotope-labeled internal standards, as we have demonstrated. For example, if the targeted analyte has a labeling efficiency of 70%, then the internal standard, which is added to the sample before tagging, should also have the same labeling efficiency (70%). Therefore, any offset in labeling efficiency will be corrected by the internal standard calibration approach. For applications of MetFish in untargeted analyses, relative quantification is used, where the abundances of tagged molecules are normalized across samples, e.g. via median centering. As long as the molecular composition of the samples being compared does not change dramatically when performing untargeted analyses, then labeling efficiency should be similar for the same metabolites found in different samples.

For molecules such as lysine that have more than one reactive group, they can be either labeled with 1 or 2 dansylations, and we observed both mono- and di-substituted ions in the MS. We selected the more intense dansylated di-cation for quantification. In addition, in order to get the best reproducibility of the method, we use amino acids as an example to optimize the reaction conditions (pH, reaction time, etc), and to ensure the reaction reached is completed and the yield is maximum (thus the reaction variation from time to time will be minimal, **Figure R1-R6** shown below).

Finally, we have validated the method in terms of accuracy, precision, etc. And this method demonstrated good validation parameters (**Supplemental Table S3**).

Per the reviewer's additional suggestion in Minor comments, we have renamed the Results section to Results and Discussion, and have changed the Discussion section to Conclusion. In the Conclusion section, we open with statements on the limitations of the MetFish approach, namely the possibility for incompletely derivatized molecules, or multiple forms of derivatized molecules.
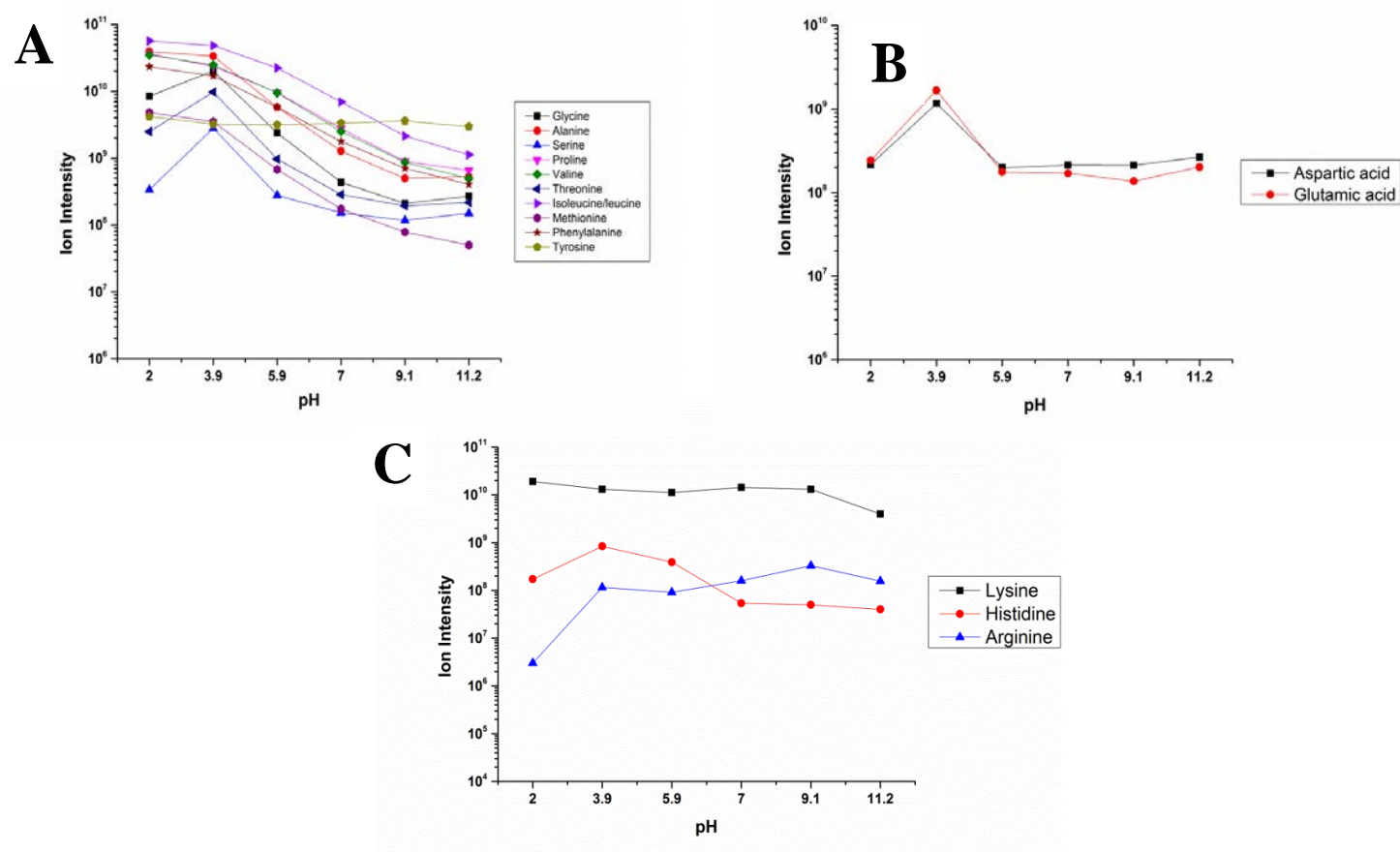

Figure R1. Effect of pH on extraction efficiency. (**A**) ESI response of neutral amino acids at different extraction pH. (**B**) ESI response of acidic amino acids at different extraction pH. (**C**) ESI response of basic amino acids at different extraction pH.
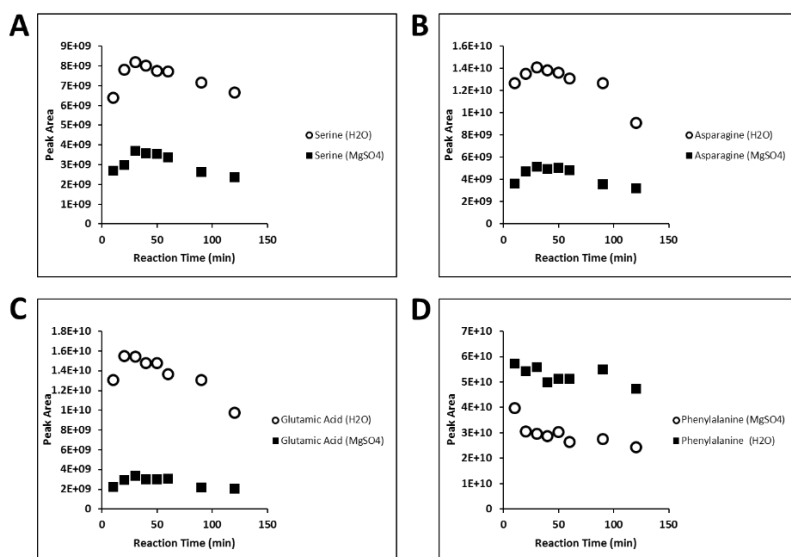
Figure R2. Comparison of ion intensities of amine metabolites derivatized in the neat reaction media and salty reaction media as the reaction time increased from 10 min to 150 min. (**A**) Serine (**B**) Asparagine (**C**) Glutamic Acid (**D**) Phenylalanine

Figure R3. Comparison of ion intensities of carboxylic metabolites derivatized in the neat reaction media and salty reaction media as the reaction time increased from 10 min to 180 min. (A)) Succinic Acid (**B**) Propionic Acid (**C**) Pyruvic Acid (**D**) Lactic Acid
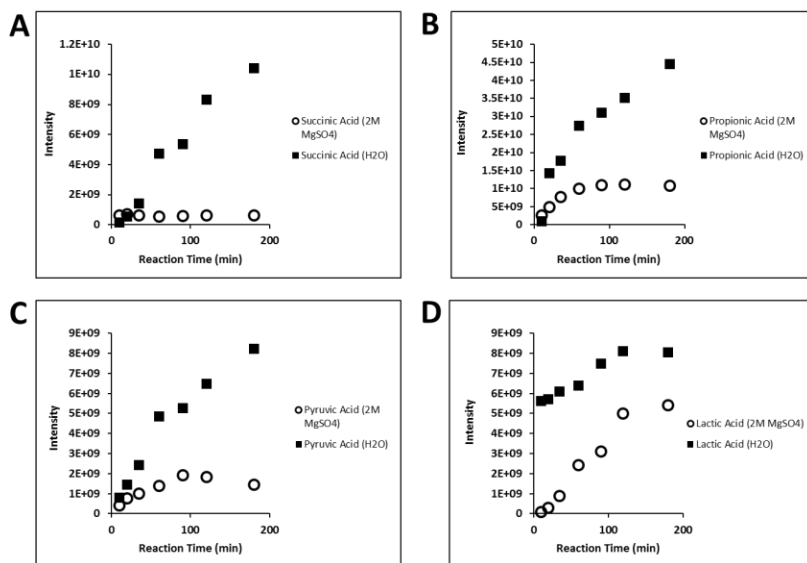
Figure R4. Comparison of ion intensities of carbonyl metabolites derivatized in the neat reaction media and salty reaction media as the reaction time increased from 2 min to 90 min. (**A**) Pyridoxal (**B**) 2-oxobutyrate (**C**) hydroxyphenylpyruvic acid (**D**) acetone

Figure R5. Comparison of ion intensities of sugar metabolites derivatized in the neat reaction media and salty reaction media as the reaction time increased from 2 min to 90 min. (**A**) maltose (**B**)) glucose (**C**) deoxyribose (**D**) ribose
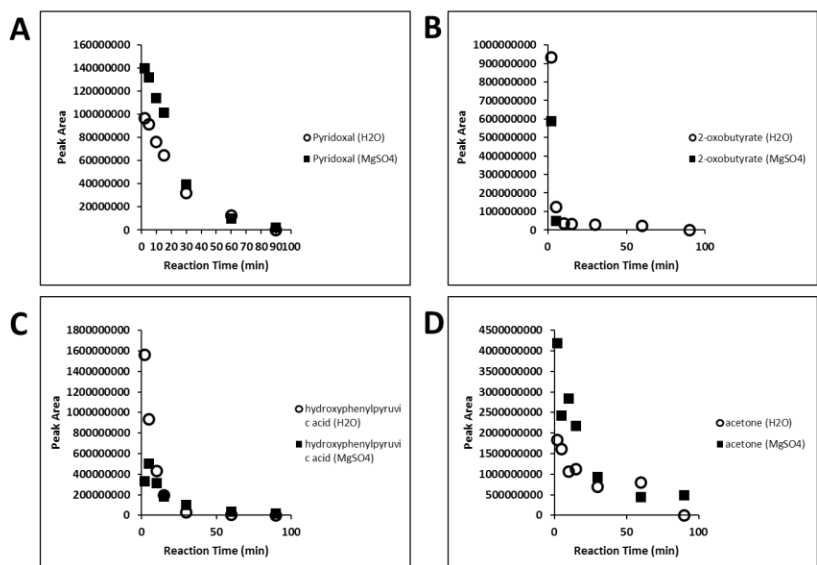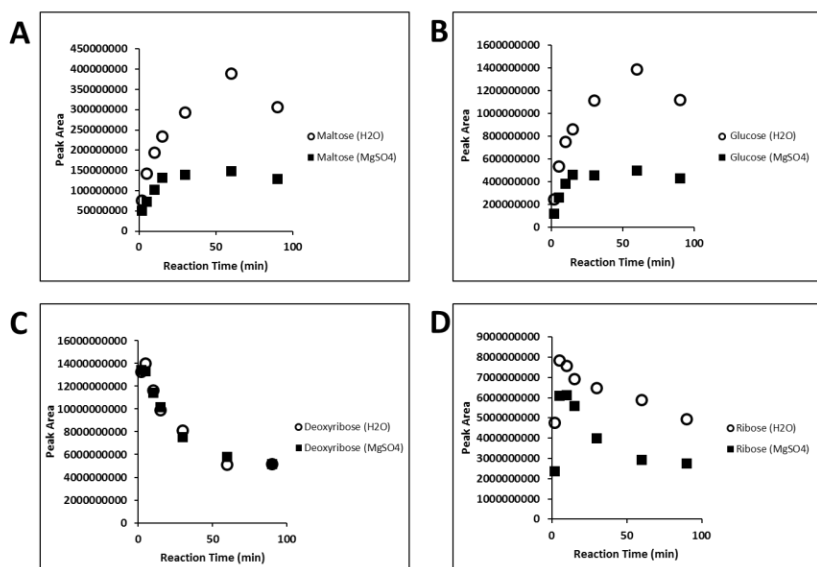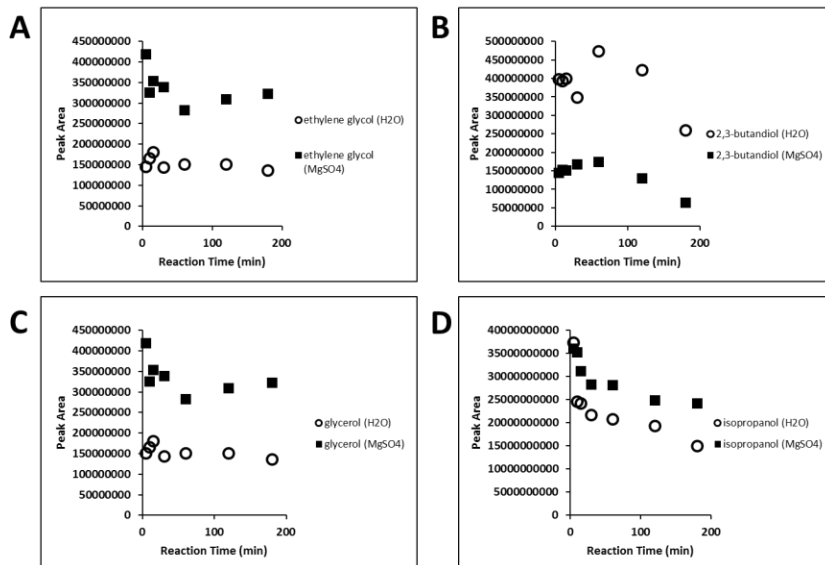
Figure R6. Comparison of ion intensities of hydroxyl metabolites derivatized in the neat reaction media and salty reaction media as the reaction time increased from 5 min to 180 min. (**A**) ethylene glycol (**B**) 2,3-butandiol (**C**) glycerol (**D**) isopropanol

8. Regarding Figure 4A, were the other 5 amino acids below the limit of detection? Why were only 14 analyzed, could the authors please explain this a little further. Along these lines, were these biological or technical triplicates, not totally clear from the figure legend.

Response: For Figure 4a (Quantification of amino acids during phototrophic microbial community succession): The other 5 amino acids were not detected (below the limit of detection). These are biological triplicates.

We have added the following sentence in the main text: 'The remaining 5 amino acids were below the limit of detection.' We have added the term 'biological' in the figure legend:
"…from analysis of 3 **biological** replicate succession experiments."

9. **Line 314, Were the authors able to find any other metabolites that were labelled, perhaps using something like MS2LDA (http://ms2lda.org/) might be an interesting tool to leverage for untargeted discovery given how consistently these dansyl probes fragment and would be largely unnatural to a system. This might be a creative way to probe for unexpected biological results.**

Response: For an untargeted, discovery-based approach suggested by the reviewer, we processed the fracking fluid datasets -- which were collected using data-dependent MS/MS and thus amenable to comprehensive untargeted analysis unlike the remaining datasets which were collected in SRM mode -- by searching the MS/MS spectra against GNPS (Global Natural Products Social Molecular Networking https://gnps.ucsd.edu/) spectral reference libraries. A table of library matches with a cosine score above 0.7 are

shown in **Supplemental Table S6**. A total of one hundred and sixty (160) unique compounds were observed. We have added a statement to the Results section of the manuscript in the discussion of the fracking fluid sample data to describe the GNPS search results.

Minor
1. The importance statement was almost verbatim the abstract, these should be two separate statements.

We have re-written the importance statement as shown below:

Importance: The identification and accurate quantification of metabolites using electrospray ionization-mass spectrometry (ESI-MS) in hypersaline samples is a challenge due to matrix effects. Clean-up and desalting strategies that typically work well for samples with lower salt concentrations are often ineffective in hypersaline samples. To address this gap, we developed and demonstrated a simple yet sensitive and accurate method – MetFish – using chemical derivatization to enable mass-spectrometry based metabolomics in a variety of hypersaline samples from varied ecosystems, containing up to 2 M dissolved salts.

2. Define QDA at first use in main text

QDA has been defined in the main text as suggested by the reviewer.

3. Figure 1, the fishing hook and compound would make a cute TOC graphic but is not needed withing the main figure

We have removed the fishing hook graphic from Fig 1.

4. Line 164, Regarding Leu and Ile, the Hicks lab has published a nice review that does touch on how these two amino acids could be differentiated in the context of antimicrobial peptides: https://doi.org/10.1039/D0NP00046A

We thank the reviewer for bringing this review paper to our attention.

5. Could the authors please clarify or explicitly state 2-C13 for Gly? Or the numbers of carbon 13 labels for other amino acids.

We used amino acid standards that were uniformly labeled with ^{13}C and ^{15}N. The Fig.2 legend and text in the main manuscript have been edited to explicitly state this as shown below:

"Figure 2 | Validation of the MetFish method using amino acids. (a) Tandem mass spectra (EIC) from analysis of a mixture of unlabeled (black spectrum) and ^{13}C and ^{15}N-**uniformly labeled** glycine (red spectrum), both derivatized with dansyl chloride…"

"As shown in Fig. 2a, dansylated-**uniformly labeled** ^{13}C and ^{15}N-glycine produces fragment ions specific to the dansyl-glycine complex and with mass shifts proportional to the degree and type of isotope labeling...... The samples were spiked with ^{13}C and ^{15}N-**uniformly** labeled amino acid standards, and endogenous amino acids in the media were quantified using isotope dilution MS."

6. Figure 3, are these transitions for metabolite quantification? It is unclear what exactly the authors have listed on what I assume are EICs.

Yes, these transitions are for metabolite quantification and are EICs. We have edited the figure caption to include these details: "Shown are **extracted ion chromatograms with the transitions for metabolite quantification** from application of MetFish in measurement…"

7. Line 240, µM or nM not nm?

Thanks for pointing this out, we have corrected it to 'nM'.

8. Figure 5, Line 345, please consider using Precursor- product ion, rather than parent- product ion. Parent and Daughter ion terminology is largely outdated.

We have replaced the term 'parent' with 'precursor' in the text and in Fig 5 as suggested by the reviewer.

9. Supplemental Table S9 might be cut off? Which of these ions ended up as identified, could this information please be included in the table.

Note: Supplemental Table S9 has been renamed Supplemental Table S5.

Since this is the precursor ion scan mode, we used the presence of three reporter ions (170, 172, 252) to track the "precursor ion". The precursor ion was the ion identified. Then we deducted the tag from the precursor ion to get the actual mass of the metabolite. We have now included **Supplemental Table S5** in an excel sheet. The table legend for S5 has been modified to include these details:

Table S5. Features detected in the untargeted metabolomics analysis by use of each of the four MetFish tags. Data has been processed to remove low abundant ions (noise), reduce false identifications and remove duplicate features. The precursor ions that generated all 3 reporter ions (170, 172, 252) were considered as potential features and the tag mass was deducted from the precursor mass to obtain the metabolite mass.

10. The discussion read as more of a conclusion, consider relabeling the sections of the manuscript.

Thank you for pointing this out, we agree with the reviewer. The 'Results' section has been relabeled 'Results and Discussion' and the 'Discussion' section has been

relabeled 'Conclusion'.

Reviewer #2 (Comments for the Author):

The manuscript entitled "MetFish: A Metabolomics Platform for Studying Microbial Communities in Chemically Extreme Environments" by Xu and collaborators, submitted as original research presents an LC-MS-based method for analysis of diversity of samples with high concentration of dissolved salts that cannot analyzed by conventional approaches.
The approach, as presented, shows potential for application to a diverse of samples due to its versatility and affordability. Overall, the approach is well supported based on the data and it is totally pertinent and useful for the field of metabolomics and others.

We thank the reviewer for this positive feedback and find it encouraging that the reviewer agrees that the approach has high relevance and utility. We are also thankful to the reviewer for providing comments and suggestions to help improve the manuscript.

I might have a few comments and suggestions that might help to improve the current version if the authors want to consider them:

1.Title: After reading the manuscript and the SI material, I consider this work as a pipeline or a method, more than an actual platform. The whole concept of MetFish is applying tagging and extraction protocols suitable for LC-MS acquisition targeting specific functionalities from metabolite moieties (functional groups). What do the authors have to further support this work to offer it as a platform?

We appreciate this input from the reviewer. The title has been changed to:
'MetFish: A Metabolomics **Pipeline** for Studying Microbial Communities in Chemically Extreme Environments'

2. Line 104: The following item "1) could be used by researchers with diverse skill sets studying myriad sample types" basically covers "everything" and I don't see how this sentence is relevant. If "1)" is removed, the following "2)" and "3)" justifications are enough to support the author's point about the reagents used in MetFish. Another option, instead of removing "1)" might be rephrasing it to something like "1) could be used to study diverse sample types based on the functional groups of interest";

Following the reviewer's suggestion, the sentence has been rephrased to:
MetFish uses low cost, commercially-available reagents that 1) **could be used to study diverse sample types based on the functional groups of interest**; 2) facilitate physical separation of metabolites from salt, mineral and other matrix components that interfere with quantitative LC-MS-based analysis; and 3) can be deployed in situ to minimize sample manipulation.

3. Line 109: Correct to "three types of samples", as based on the results the study was performed on three sample types with proper replicates for each of them, even being three replicates these are not actually just "three samples" here;

The sentence has been corrected to incorporate the reviewer's suggestion:
We demonstrate the utility and simplicity of MetFish in LC-MS-based exo-metabolomics analyses of three **types of** samples containing or derived from microbial communities from diverse ecosystems: a hypersaline aquatic microbial community, a prairie soil, and fluids injected into and produced from a hydraulically fractured well, each consisting of or derived from hypersaline (i.e. from 400 mM to 2 M) sample matrices.

4. Line 131: Clarify what the "QDA" abbreviation stands for and cite again the paper from where this reference comes from, since QDA (N-[2-(Aminooxy)ethyl]-N,N-dimethyl-1-dodecylammonium iodide) is only mentioned in the SI;

We have included the full form of QDA and cited the paper again.

5. Line 135: Check punctuation "(...) groups. . Recognizing (...)";

Thank you for noticing this and pointing this out, we have corrected this.

6. Line 164-165: In the following "To illustrate this (...)", I feel this paragraph is not connected to the previous sentence, which talks about Leu and Ile. Maybe use another connector instead of "To illustrate this (...)", unless the authors want to actually illustrate the lack of unique fragments produced by CID of Leu and Ile vs Gly;

To address this, we have added the sentence that refers to Leu and Ile to the previous paragraph and then started a new paragraph with a rephrased sentence as follows:

To illustrate **metabolite identification using unique fragment ions**, the fragmentation spectrum for dansylated glycine is shown in Fig. 2a.

7. Line 192: "(...) and is available for study": This seems kind of obvious to me. What do the authors mean here, that the microbial mat was easy to study? Everything is available for study, if the efforts are properly set for that. Please, clarify the idea;

We have edited the sentence as follows:

MgSO4 was chosen as it is a major salt component of Hot Lake, located in Oroville, WA, where a photoautotrophic microbial mat community resides.

8. Line 212 (Figure 2): I suggest to include the y-axis (intensity) in all the

chromatograms, as well as clarify that (b) are Extracted Ion Chromatograms, if I assume correctly based on the figure;

We thank the reviewer for these suggestions.

To clarify, Figure 2a is not a chromatogram, but a MS/MS spectrum. The intensities of the fragment ion peaks in this spectrum have been normalized such that the most intense peak in the spectrum (m/z 156.72) has been set to 100%. The heights of the remaining peaks in the spectrum have been set as a ratio to the maximum intensity, and thus the y-axis ranges from 0-100%.

For Figure 2b, yes, these are all extracted ion chromatograms, and we have modified the figure legend to include this information. Similar to Figure 2a, the extracted ion chromatograms for the upper and lower plots have been overlayed and the y-axes normalized to a range of 0-100% based on the most intense peak in the overlays. The middle plot is a total ion chromatogram whose y-axis has also been normalized from 0-100%.

Thus, for the overlayed extracted ion chromatograms, it is difficult to present the y-axes with actual intensity values. The absolute peak intensities of the target analyte peaks for each portion of the figure are shown in the **Table R1** shown below.

Table R1: Peak intensities for the EICs of amino acids shown in the upper and lower chromatograms of Fig 2(b).

Amino Acid	LC-MS intensity	
	MetFish (25 pmol amino acids in 2M MgSO4)	HILIC (25 pmol amino acids in water)
Glycine	4.43E+07	2.20E+05
Alanine	4.49E+07	1.05E+06
Serine	9.51E+06	1.27E+06
Proline	7.83E+08	5.12E+05
Valine	1.94E+08	1.93E+06
Threonine	3.15E+07	1.77E+05
Isoleucine	2.83E+08	1.80E+06
Leucine	2.40E+08	2.46E+06
Asparagine	1.25E+07	7.60E+05
Aspartic acid	1.58E+07	5.51E+05
Glutamine	2.00E+07	2.72E+06
Lysine	1.24E+06	1.80E+06
Glutamic acid	5.79E+06	1.77E+05
Methionine	4.00E+08	2.20E+05
Histidine	1.71E+08	2.85E+05

Phenylalanine	4.13E+08	2.68E+06
Arginine	1.31E+08	1.32E+06
Tyrosine	7.81E+06	1.62E+05
Tryptophan	3.16E+08	2.06E+06

9. Line 227 (Figure 3): considering these are also EIC (or XIC, extracted ion chromatograms), do the authors have any comment on the additional ions observed, e.g. butanetriol and peak shape, e.g 2-deoxy-D-ribose? Maybe include the m/z range used for selecting the EIC. It would be also worthy to clarify these are EIC from SRM;

Although we were unable to do an in-depth investigation into the peak shape, chromatographic separation and detection of specific metabolites mentioned by the reviewer, we have summarized all the transitions and fragment ions used for quantifying the representative metabolites in **Tables R2-R4** shown below. We have edited the Fig 3 legend to clarify that the EICs are from SRM:

Table R2. Parent ions, quantification ions, and confirmation ions of carboxylic metabolites used in selected reaction monitoring mass spectrometry.

Metabolite	Parent Ion	Quantification Ion	Conformation Ions		
			Fragment Ion 1	Fragment Ion 2	Fragment Ion 3
Butyric Acid	406.1	335.8	370.1	391.0	170.1
Propionic Acid	392.1	335.8	355.7	376.7	170.1
Acetic Acid	378.0	335.8	341.9	363.0	170.1
Salicylic Acid	456.2	335.8	392.1	438.0	170.1
Shikimic acid	492.1	335.8	455.7	173.8	170.1
Lactic Acid	408.1	335.8	362.1	389.6	170.1
Pyruvic Acid	406.0	335.8	86.1	345.6	170.1
Oxalic Acid	408.0	335.8	362.0	389.7	170.1
Malic Acid	385.0	335.8	518.2	536.2	170.1
Succinic Acid	377.0	335.8	418.0	520.2	170.1

Table R3. Parent ions, quantification ions, and confirmation ions of carbonyl metabolites used in selected reaction monitoring mass spectrometry.

Analyte	Parent Ion	Quantification Ion	Conformation Ions		
			Fragment Ion 1	Fragment Ion 2	Fragment Ion 3
6-oxoheptanoate	392.1	235.8	156.5	356.0	374.0
Hydroxyphenylpyruvate	428.1	235.8	320.0	382.0	409.7
Pyridoxal	415.1	235.8	108.9	151.7	397.0
Acetone	306.1	235.8	261.4	271.2	290.4
Sodium 2-oxobutyrate	350.0	235.8	279.1	291.3	307.9
Ketoglutamic acid	394.1	235.8	345.8	364.0	375.9
Glucose	428.1	235.8	331.8	374.1	409.8
Maltose	590.3	235.8	392.0	427.8	572.0
Ribose	398.1	235.8	307.8	361.3	380.2
2-deoxy-D-ribose	382.1	235.8	291.5	345.6	364.0

Table R4. Parent ions, quantification ions, and confirmation ions of hydroxyl metabolites used in selected reaction monitoring mass spectrometry.

Analyte	Parent Ion	Quantification Ion	Conformation Ions		
			Fragment Ion 1	Fragment Ion 2	Fragment Ion 3
ethylene glycol	210.1	150.7	133.8	192.1	194.4
glycerol	240.1	150.7	197.2	208.2	224.7
1,2,4-butanetriol	254.1	150.7	70.5	236.1	238.9
2,3-butandiol	238.1	150.7	191.2	220.3	222.6
butanol	222.1	150.7	133	193.1	206.7
isopropanol	208.1	150.7	150.6	179.1	192.2
Phenol	242.1	150.7	178.5	213	226.1
m-cresol	256.1	150.7	205.7	219.7	240.9

"Figure 3 | MetFish is applicable to measuring metabolites with a broad range of functional groups in challenging sample matrices. Shown are extracted ion chromatograms with the transitions obtained in SRM mode for metabolite quantification from application of MetFish in measurement of (**a**) amine metabolites, (**b**) carboxyl metabolites, (**c**) carbonyl metabolites, (**d**) hydroxyl metabolites as sugars, and (**e**) hydroxyl metabolites as alcohols. In all cases, MetFish was deployed *in situ* in metabolite-salt mixtures containing 2 M $MgSO_4$."

10. Line 278-281: Can the authors elaborate a little bit more here to discuss or hypothesize the reason behind Ser and Pro results? Since these were unexpected results, as the authors suggest, it would be valuable for the manuscript to include this discussion;

Unfortunately, we are unable to formulate a plausible biochemical hypothesis for why serine and proline have concentration maxima later than the other amino acids in this incubation.

11. Line 328 (Figure 5): Can a title for this figure be included in this figure legend, for consistency? currently, only the (a) and (b) description of the figs are shown;

The figure legend has been modified:
Figure 5 | Untargeted metabolomics using MetFish (a) Workflow for untargeted metabolomics analysis using MetFish. **(b)** Global amine tag based metabolite profile of a produced fluid sample. The size of the circle is proportional to the ion intensity, and putatively identified metabolites are labeled.

12. Line 349: In "(...) samples ranging from 86 and 154 d post-injection", are these from 86 and 154 days?

The produced fluid samples were collected at 7 time points between day 86 and day 154. The text has been modified to make this clearer:
"..applied MetFish in a targeted exo-metabolomics analysis to confirm molecular identities in a time series of produced fluid samples **collected between 86 and 154 days** post-injection.."

13. Line 360 (Figure 6): I suggest to expand a little bit the description of this figure. It is not clearly specified the statistical tools used to generate these plots. Are there somewhere else? In Fig 5a, is this a hierarchical clustering? What the distance metric used for these plots, etc. Please, provide details either here or in the methods section. These are currently missing;

The figure legend has been expanded to include the information suggested by the reviewer:

"**Figure 6** | **Targeted metabolomics analysis using MetFish in injected and produced fluids from hydraulic fracturing** (**a**) Quantification of 37 metabolites identified by targeted MetFish (**b**) Comparison of metabolite levels in the input material and spent fracking fluid. Data shown are from replicate analysis (n = 3) of each fluid sample. Heatmaps were generated using MetaboAnalyst 3.0 (64) using Pearson's distance measure and average linkage for the clustering algorithm."

14. Line 382-385: I would encourage the authors to include and discuss the limitations before ending the manuscript. The "Discussion" section starts with summary. Please restructure this. I would suggest to discuss the limitations before making the summary;

We have relabeled the 'Discussion' section as 'Conclusion' and have started this section with a discussion of the limitations and then included the summary as suggested by the reviewer.

15. Also related to Lines 382-385: Additionally, GC-MS results are not discussed in any place from the main text although the results are included in the SI material (e.g. Table S8 shows identifications of metabolites from GC-MS libraries). I think that info is missing in the manuscript and should be discussed or at least briefly mentioned, as I don't see a clear reason behind including this in the SI if not used.

Note: Supplemental Table S8 has been renamed Supplemental Table S4.

We have included the following text in the main manuscript's 'Results and Discussion' section:
*"We also tested the feasibility of using GC-MS to detect the presence of amino-acid standards from high salt matrices and found that the presence of salt (both 400 mM and 2 M total dissolved salts) severely affected the measurement and no analyte peaks were observed in the chromatograms (**Supplemental Figure S2 and Supplemental Table S4**)."*

16. Line 433: "(...) as described previously" where? Please, mention where and/or provide reference to a published work if available;

We have included an appropriate reference for that sentence in the text.

17. Line 472: "(...) performed at speed 7 (...)", please Clarify what kind of device from the instrumentation was used for this step;

We have included the instrument information in the text:
A scoop of stainless steel beads and garnet beads were added into the centrifuge tube. 1 mL of water was then added and bead beating was performed at speed 7 in a **Bullet Blender Tissue Homogenizer (Next Advance, NY, USA)** for 4 min at 4 °C to lyse microbial cells.

18. Line 520: "(...) in-house tool MASIC", It would be worthy that the settings for these processing steps be included in the methods section;

We have included details on the parameters used for MASIC processing of the data in the methods section, and we have also provided a citation to the relevant paper.

19. From the SI material, I have the following comments:

Figure S1: Can the two pipelines be clearly highlighted here? Which is target vs untargeted? At least that is not clear here based on the description from the main text.

We apologize for the confusion. What is shown in Figure S1 is not a schematic of two pipelines but a schematic of a dual-column LC system. We apologize for the confusing use of the term "pipeline" and have modified the figure and legend to be more accurate.

20. Figure S2: It might be useful to label the identified compounds from Table S8 directly in the chromatograms from Fig S2. See my previous comment regarding to "GC-MS analysis of samples with high salt content"

We have labelled the peaks directly in the chromatogram in Fig S2A, as suggested by the reviewer. The reviewer's previous comment has also been addressed and we have included text in the main manuscript that refers to the G-CMS experiment and supplementary data.

April 16, 2021

Dr. Thomas Metz
Pacific Northwest National Laboratory
Richland

Re: mSystems01058-20R1 (MetFish: A Metabolomics Pipeline for Studying Microbial Communities in Chemically Extreme Environments)

Dear Dr. Thomas Metz:

I am glad to inform you that your manuscript has been accepted, and I am forwarding it to the ASM Journals Department for publication. For your reference, ASM Journals' address is given below. Before it can be scheduled for publication, your manuscript will be checked by the mSystems senior production editor, Ellie Ghatineh, to make sure that all elements meet the technical requirements for publication. She will contact you if anything needs to be revised before copyediting and production can begin. Otherwise, you will be notified when your proofs are ready to be viewed.

As an open-access publication, mSystems receives no financial support from paid subscriptions and depends on authors' prompt payment of publication fees as soon as their articles are accepted. You will be contacted separately about payment when the proofs are issued; please follow the instructions in that e-mail. Arrangements for payment must be made before your article is published. For a complete list of **Publication Fees**, including supplemental material costs, please visit our [website](#).

Corresponding authors may [join or renew ASM membership](#)to obtain discounts on publication fees. Need to upgrade your membership level? Please contact Customer Service at Service@asmusa.org.

For mSystems research articles, you are welcome to submit a short author video for your recently accepted paper. Videos are normally 1 minute long and are a great opportunity for junior authors to get greater exposure. Importantly, this video will not hold up the publication of your paper, and you can submit it at any time.

Details of the video are:

· Minimum resolution of 1280 x 720
· .mov or .mp4. video format
· Provide video in the highest quality possible, but do not exceed 1080p
· Provide a still/profile picture that is 640 (w) x 720 (h) max

We recognize that the video files can become quite large, and so to avoid quality loss ASM suggests sending the video file via https://www.wetransfer.com/. When you have a final version of the video and the still ready to share, please send it to Ellie Ghatineh at eghatineh@asmusa.org.

Thank you for submitting your paper to mSystems.

Sincerely,

Marcelino Gutierrez
Editor, mSystems

Journals Department
American Society for Microbiology
1752 N St., NW
Washington, DC 20036
E-mail: peerreview@asmusa.org
Phone: 1-202-942-9338

Supplemental Figure S3: Accept
Supplemental Table S2: Accept
Supplemental Figure S1: Accept
Supplemental Figure S2: Accept
Supplemental Table S4: Accept
Supplemental Table S6: Accept
Supplemental Table S1: Accept
Supplemental Table S5: Accept
Supplemental Table S3: Accept
Supplemental Information: Accept