nature portfolio

Peer Review File

Decomposing the molecular complexity and transformation of dissolved organic matter for innovative anaerobic bioprocessing

Corresponding Author: Professor Xian-Wei Liu

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Attachments originally included by the reviewers as part of their assessment can be found at the end of this file.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
This paper shared a detailed and technical passage on anaerobic bioprocessing of organic waste, focusing on dissolved organic matter (DOM) analysis using FT-ICR MS. The study explores the molecular complexities and transformations in high-solid anaerobic digestion across food waste treatment plants in China. Key findings include the diverse molecular compositions of DOM, shifts in molecular diversity from primary to secondary digestion stages, and biochemical reactions involved in lipid, lignin, and protein degradation. However, we can not recommend for publication in its current form for several reasons.
First, the manuscript overly relies on FT-ICR MS data, with all results based on interpretations of molecular composition data to infer structural changes, rather than directly measuring a subset of structures, especially the core molecular community as proposed by the authors.
Second, FT-ICR MS results are largely qualitative, leading to speculative conclusions, and there is a lack of direct measurement data on microbial changes.
Therefore, while the research content and findings are interesting, the heavy reliance on molecular composition data reduces the credibility of the results. We strongly recommend supplementing the manuscript with microbial data as mentioned and structural data on molecular composition to enhance the credibility of the findings.

Reviewer #2

(Remarks to the Author)
AD is widely used for the treatment of organic wastes, and the microbial communities were well characterized previously. However, the transformation of the complex organics was not well studied. This is an interesting study that investigated the DOM transformation in AD process by FT-ICR MS analysis with a novel analytical framework. The molecular properties of DOM and their transformation in AD were well characterized, which provided new insights for the organic conversion in AD. The manuscript was well written and can be accepted for publication after considering the following comments.
1, Line 42, what kind of variables? Please specify.
2. Line 57, what are simplified systems? Please add more information. There are also studies using FT-ICR MS to investigate the DOM in AD treating complex substrates like sludge. Please be more accurate.
3. Line 59, “doe” should be “due”
4. Line 124-125, why was the increase of CHONS from lipid accumulation? Please explain.
5. Line 163, why did two samples show atypical compositions? Please explain.
6. Line 204, were the newly produced molecules from the solubilization of solid organics or the degradation of DOM ?
7. Lin 310-312, lipid can be degraded during AD process, why were they still present in the effluent?
8. Line 339-351, the traditional methods were mentioned to increase the DOM conversion. It is better to propose new methods based on the new understanding of the DOM conversion in this study.

Reviewer #3

(Remarks to the Author)
See attached pdf file (review.pdf)

Version 1:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
The manuscript has been revised according to reviewers' comments. It could be accepted for publication as it is.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
[Note from the Editor: Reviewer #3 was asked to also assess the responses given to Reviewer #1.]

Review for revison of “Decomposing the molecular complexity and transformation of dissolved organic matter for innovative anaerobic bioprocessing”, by Jun Hu, Chuan-Guo Liu, Wen-Kai Zhang, Xue-Wen Liu, Bin Dong, Zhan-Dong Wang, Yuan-Guo Xie, Zheng-Shuang Hua and Xian-Wei Liu* resubmitted to Nature Communications.

Jun Hu, Xian-Wei Liu and colleagues revised their initial manuscript. The authors did a good job, making the study much more robust and comprehensible. I think that all reviewer’s points were addressed adequately. Improvements in regard of reviewer #1’s criticism should be discussed more clearly in the discussion section (Figure S4) which is right now not connected well with the rest of the study (DOM data). From my side, there remain some adjustments, largely intended to improve the comprehensibility further. If those comments and suggestions are considered, the paper should be acceptable for publication.

Edit on Feb 27th, 2025: I was asked to also review responses to Reviewer #1, which is why I added to general comments #3, #6 and #7. I think the authors did a good job responding, they should make sure that all the given information/ reasoning is also added to the manuscript (or SI), see the general comments for more details.

Decision: Minor revisions

Yours sincerely, Carsten Simon

General comments and last suggestions

1) New “reactomics” approach not introduced properly despite some considerable improvements.
In lines 262 – 264 this becomes very obvious when you discuss the interpretation of your index values – what do these numbers mean? – why can they become negative? – why are they so high if you quantify individual transformations between formula pairs in one sample only? Figure 4 looks pretty but does not really help me to grasp what is going on. For example, in line 255f, you discuss that DMI is negative against expectation. Could this mean that long-chain molecules remain while short-chain ones vanish, in a relative sense? Thus, DMI would appear to indicate “production” of long-chain molecules, while in reality, long-chain molecules persist and short-chain ones vanish? Here, it really becomes obvious that the reader would benefit from a better documentation of the reaction calculations… I appreciate the effort you invested to improve this, but it is not detailed enough in Supplementary Method 2 – you essentially propose a new data analysis approach, so please (!) introduce it properly. See also comment on “data and code availability”.

2) NOSC shift should be shown more clearly in Figs 3 and 5.
In line 205 you state, “The redox states, indicated by the NOSC, shifted to a more reduced state (Fig. 3c).” I cannot see this significant shift in the density plots in Figure 3c, it rather seems that PD is more reduced.... To improve visualization, add a plot in the style of panel 3d to visualize the NOSC distribution across PD and SD and add lines for the range of NOSC(IWA) in both sample sets (n=7 for PD and SD). It just now occurs to me that you have many distribution plots of other indices in the manuscript, but not many showing the NOSC. Fig. 3 would be a good place to do so. In Figure 5a, it becomes obvious too that the IWA of the compound groups remains similar, and that a shift to more lipids likely explains the lower NOSC. In Figure 5c it is not clear which panel shows which sample (top, PD; bottom SD?). In Figure 5, please denote clearly which data the formulas rely on. In the caption, cite Gunina, A., & Kuzyakov, Y. (2022, From energy to (soil organic) matter. Glob Change Biol, 28, 2169–2182. https://doi.org/10.1111/gcb.16071) if you decide to add info to your figure that is 1:1 adapted from their paper (“Reduced substrates: Energetically rich, but require more energy for breakdown”).

3) Also in line with Reviewer 1 comments: Supplementary Figure S4 is not referenced and Supplementary Note 1 is not referenced in the discussion.
Include a reference for Fig S4 and Note S1; e.g. in line 377 where you discuss sludge microbiota. You should add some

statements regarding the analyses shown in Fig S4 as well… Probably the figure should also be shifted to a later position in the SI (Numbers of figures will change).

4) Use of terms “recalcitrance”/ “recalcitrant”. It is now widely accepted that there are no inherently recalcitrant compounds but that compounds merely “persist” under certain environmental conditions. I would advise to critically examine the occurrence of the term “recalcitrant” and “recalcitrance” and evaluate whether it could be replaced by the more general term “persistent” and “persistence”. An alternative strategy would be to use a wording in the lines of “more/ less recalcitrant” or “higher/ lower recalcitrance” to stress its relative character. Occurrences: main text (lines 26; 133; 159; 173; 201; 310; 333; 345; 356; 373; 383; 393; 410; caption of Fig 6), text in figures (Fig 6) and supplement (lines 160; 171; 253).

5) Data and code availability.
I thank the authors for adding information to the SI, but I still think that the paper could be made more accessible. Keep in mind that your work will be cited more if your data and code can be used openly and easily. As you state, “A grand challenge in DOM research is to visualize the potential biochemical processes that DOM undergoes throughout its lifetime.” All the more important it would be to allow other researchers to rerun your code using your raw data…

5a) I checked the github link and found the info there to be unchanged to last time I would suggest to add proper documentation to the code structure (the code is not commented right now). It is not obvious what the code does because there is no example input files or data to test the scripts... I would suggest to add simple test data from your study. Think of your paper not only as a story, but also as a data and code publication.

5b) I noted that the data you provide is only the data to reproduce figures, but no raw data has been made available so far. I would suggest you make the raw data publicly available alongside your publication to allow peers to use your data. This would be in line with your “global” approach to compare the AD process across sites – allow others to make use of your data to continue your work.

6) Also in response to reviewer #1 comments: Include the reasoning for using ultrahigh resolution mass spectormetry (Response 2 to reveiwer #1) as a Supplemental Note and reference it in the main text (introduction). You made a good job arguing for using the chosen approach. Please transfer the text you responded to the SI for other readers.

7) Also in response to reviewer #1 comments: Make sure that the reasons that your study stands out (response 4 to reviewer #1) are mentioned in the abstract and conclusion of your manuscript, especially pint 1) Study design.

Specific comments and corrections

Line 90 “in the DOM” <- “of DOM”

Line 101/102 Please add that the AD system is a relatively “closed” system as opposed to natural aqueous systems which are open. Replace “static” by “stagnant”.

Line 103 Again, indicate that this is a “closed” system.

Line 108 “To quantify chemical diversity, we defined the number of formulae as richness and intensity-weighted formulae as abundance.” I don’t understand the second part of the sentence. How can “intensity-weighted formulae” equal “abundance”? Not clear. Description of calculation of data underlying Fig 1c and 1d is also missing in methods.

Line 189 “The” spelled wrong

Line 214f “These features are significantly more pronounced than those found in smaller, more bioavailable molecules like cellobiose, glucose, and acetate (Fig. 3f-h)” Unclear. Cellobiose, glucose and acetate are not shown in the panels.

Line 232 Please add right at the start that you infer “theoretical” or “potential” reactions here by comparing two samples (“before” vs “after”) but that this indication is indirect… also, mention again that molecules in the “lignin region” are not necessarily lignin monomers… I cannot stress this enough – the interpretation of molecular structure from O/C vs H/C is very crude. I provide a reality check here: https://pubs.acs.org/doi/full/10.1021/acs.est.2c01332

Line 301 NOSC is not "nitrogen oxidation state". Check again throughout manuscript

Line 333 The use of the word recalcitrant appears most wrong here. These compounds likely can be decomposed, they just persist during the conditions of the decomposition process…

Line 354 What is “mass size”? Write mass, that is the property you measured

Fig S24 Chemodiversity on y axis spelled wrong

Table S8 In the caption it reads “compound classes”, I think “compounds” would be more suited.

(Remarks on code availability)
I reviewed the code but did not run it (I do not use PERL).

As pointed out above, I find the code is not documented well enough. There are no comments added, and no test data provided to run the code with data from the paper. A schematic workflow of all calculations involved would help to comprehend the processing better. What is the input, what is the output?

The description of the approach in the Supplement has improved due to the revision but is still not clear enough. This can be seen in the results passage of the paper, where the authors themselves seem to be unsure how to interpret their new indices (see line 255 and following); Figure 4 looks nice but does not help to understand the approach either.

It is unclear to me how these indices are calculated, why they can become negative, what the numbers actually say (can they be compared to each other?)... By all means, I think that the analysis is sound and the approach innovative, but the authors need to make a better effort to properly describe their approach so that readers understand it and hence, reuse (cite) it!

Version 2:

Reviewer comments:

Reviewer #3

(Remarks to the Author)
Review for 2nd revison of “Decomposing the molecular complexity and transformation of dissolved organic matter for innovative anaerobic bioprocessing”, by Jun Hu, Chuan-Guo Liu, Wen-Kai Zhang, Xue-Wen Liu, Bin Dong, Zhan-Dong Wang, Yuan-Guo Xie, Zheng-Shuang Hua and Xian-Wei Liu* resubmitted (2nd time) to Nature Communications.

Jun Hu, Xian-Wei Liu and colleagues revised their initial manuscript in a second round. All my concerns were addressed properly.

The paper and the github code are now much better to understand, and I thank the authors for accepting my suggestions. The reactomics perspective is very interesting and (apart from its reliance on van Krevelen “domains”...) reveals new insights.

I have read the whole paper once again and stumbled across some things that may still be unclear. Please understand these minor comments as an opportunity to clarify your work better.

Suggestion: Accept after minor revisions, no further round of revision required.

Yours sincerely, Carsten Simon

Last comments

Line 59 “these studies” – add citations to be more specific

Line 84 How is the pH set to 7.77? Is a buffer added?

Data underlying Fig 1c/d Why does the abundance not add up to a 100%? Maybe clarify

Line 110 “[…] with abundance of DOM” I think you mean to say “with abundance of DOM molecular groups”? Abundance of DOM is unclear

Line 124 “chemical footprint” I think “chemical fingerprint” is more suitable

Line 131 Maybe add a statement where you think the P goes. Into microbial biomass?

Line 142 “[…] inherent resistance to microbial degradation” What about production of lipids? They are constituents of cell membranes also. I think this could also contribute to rising lipid levels in DOM?

Line 168 “richer” aromatic rings maybe use “more” instead of “richer”

Line 175 I think it may be helpful to add for clarity that these insights are based only on composition and do not take into account the potential removal of DOC (unfortunately you do not supply DOC data in Fig 1a).

Fig. 2h Unclear how you calculated the C-1-DBE metric and what insight it gives (applies to Fig 3f as well). Please add info to line 498 (after DBE) in the methods

Fig 2 caption, line 190 Unclear which “boxplots” you refer to here. Maybe they were removed in the reviewing process?

Line 219 “[…] double bonds (C-C and aromatic […]” write “C=C” instead of C-C

Line 226 … however, looking at Fig 3b, the core community only makes up a small proportion based on formula number. Maybe add an intensity-based figure (pie diagrams) as well usually, common formulas are also high in intensity, meaning that the core formulas might be small in number but important in terms of intensity (information).

Fig 3 caption Panel c – please give more information how the molecular groups (Carb, Lignin, …) are linked between samples and that they are also shown separately for elemental compositions (CHO, …). Maybe consider adding the median of the NOSC distributions (as a red line or so) to highlight that there is a shift to lower NOSC

Line 237 Specify again that this is not only based on core formulas (that were discussed before). Also, be more clear – formulas unique to PD means that you averaged all PD mass spectra and compared this to the average SD spectrum? More info would help to clearly separate it from the section before where you looked for common molecules between PD and SD.

Line 240 “broken down” and “newly produced” – consider rewriting to “disappearing” and “newly detected”.

Line 250 “or” missing

Line 254 At the end of the paragraph, you may want to add a note that this analysis focuses on all putative reactions between all potential formula pairs in PD and SD, and does not actually provide evidence of these reactions taking place – it is an explorative analysis (similar to this one: https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/rcm.7719)

Line 262 Why does decarboxylation facilitate “the interconversion among lignin-, protein- and lipid-like molecules”? You state: “Our analysis demonstrates a notable lignin-to-lipid transition with DCI (lignin→lipid) = 26688” – Be more specific here: Knowing your paper, I would interpret this statement the following way: You found decarboxylation reactions among formulas, and a high proportion of those formula couples were classified as “lignin” in PD and classified as “lipid” in SD. If so this would seem contradictory because one only removes C and O (CO_2) but lipid-like molecules are defined by high H/C and low O/C, but it makes sense if one considers that the H/C ratio of the remaining molecule increases due to removal of C, and O/C decreases as well because more O is removed than C. Maybe it would make sense to explain it in detail at this stage. I was struggling a bit to follow your argumentation in the reactomics section of the results but I think I understand what you mean. Considering that I have spend longer time with your manuscript than most other readers will, it may help to explain the result, at least for one index (e.g., decarboxylation of lignin to lipid) in easier words.

Line 278 You state that there was a “demethylation pathway […] followed by decarboxylation” – maybe explain how you come to this conclusion based on your data. If I understood correctly, you calculated each index separately. This statement sounds as if you calculated them in sequence or in parallel…?

Line 308 “prediction bond” – do you mean “confidence interval” or sth. similar? Maybe prediction range sounds better

Line 310 “Most compound classes showed a shift” – From looking at Figure 5a, I’m not fully convinced, the change seems marginal. Maybe you could show the difference in ΔH and NOSC instead of the absolute values to make this point more clear…

Line 319 change to “[…] in the sludges, including lignin, lipid […]”

Line 343 What are “insoluble liquids” – do you mean oils? Please specify or use other wording

Line 356 “despite the removal of large amounts of DOM” this is important, can you add an estimate how much DOM is removed?

Line 357 “20% of the DOM molecules”; this contradicts with the fraction of core formulas that make up only 5-8% (Fig 3b). Maybe revise and be more specific.

Line 363 “matrix effects” unclear what you mean. Matrix effects can mean everything. Why not just write “is largely governed by its structural properties”?

Line 389 “Narrow product state” – unclear what you refer to here.

Line 393 Why is acidogenesis an efficiency-limiting process? Please add more info here, it may be interesting for readers.

Line 415 “In summary, our study […] decomposed the […]” use another word than decompose. For example, you could use “elucidated” or “dissected” or “uncovered”… decomposed has a double meaning because you also deal with DOM decomposition in the literal sense.

Supplementary Method 2
I suggest to add some more info to Supplementary Method 2 (basically including your response #3 in the text). You stated in your response that “(1) the broad mass range distribution of DOM molecules, (2) the large number of unique molecular formulas in each sample, and (3) the indices being calculated as the cumulative sum across seven FWTPs, each with

diverse DOM composition" are reasons for high index numbers. Please add this information to the text and explain why this is especially for reasons (1) and (2) – I assume the higher the number of formulas (and hence mass range), the more potential formula pairs are there to calculate your mass difference. What I don't understand is how unique formulas would affect this.

Table S1 Please add a cautionary statement to the caption like "Note that comparison of numbers of formulas ("richness") may be impeded by differences in sample number, ionization conditions, instrumental settings (including resolution power), formula assignment rules, and post-processing rules (filtering steps)."

(Remarks on code availability)
Code has improved and is much more accessible now.

Response to Reviewer 1's Comments

This paper shared a detailed and technical passage on anaerobic bioprocessing of organic waste, focusing on dissolved organic matter (DOM) analysis using FT-ICR MS. The study explores the molecular complexities and transformations in high-solid anaerobic digestion across food waste treatment plants in China. Key findings include the diverse molecular compositions of DOM, shifts in molecular diversity from primary to secondary digestion stages, and biochemical reactions involved in lipid, lignin, and protein degradation. However, we cannot recommend for publication in its current form for several reasons.

Response: We sincerely thank the reviewer for their thoughtful and constructive feedback. We appreciate the opportunity to improve our manuscript and have addressed your concerns point by point below.

First, the manuscript overly relies on FT-ICR MS data, with all results based on interpretations of molecular composition data to infer structural changes, rather than directly measuring a subset of structures, especially the core molecular community as proposed by the authors.

Response: We acknowledge the reviewer's concern regarding the reliance on FT-ICR MS data to infer structural changes in DOM. However, we would like to clarify and address the following points regarding our methodological approach:

(1) Challenges in directly determining molecular structures of DOM.

- **Inherent complexity of DOM molecular composition**. DOM from complex environmental systems comprises an enormous range of molecular sizes, functional groups, and chemical compositions. This diversity makes it difficult to characterize specific structures of individual molecules. Even with advanced techniques like MS/MS or NMR, only a fraction of such complexity can be definitively characterized. Achieving an exhaustive direct structural analysis across all compounds is therefore technically unfeasible for complex matrices like food waste digestate.
- **Balancing broad compound detectability with in-depth structural characterization**. FT-ICR MS is the only MS technique offering sub-ppm mass accuracy for various peaks in a complex mixture, making it the most advanced tool for unraveling molecular complexity with ultra-high resolution. However, FT-ICR MS alone cannot directly measure specific subsets of DOM structures without more specialized techniques (e.g., 2D-NMR), which, however, may not capture the full complexity of DOM. Therefore, it's difficult to balance broad compound detectability with in-depth structural characterization for DOM analysis.
- **Technical constraints for Integrating different MS techniques**. Integrating different MS techniques for comprehensive interpretation is extremely challenging due to differences in resolution, mass accuracy, analyte detection, ionization methods, and

sample preparation. Each technique provides unique insights but also generates data in different formats, often capturing distinct subsets of a sample. These inherent differences make it difficult to achieve a unified, coherent interpretation of results from multiple MS platforms, requiring specialized workflows, considerable expertise, and often significant manual effort to align the datasets.

(2) Inferring structural features from molecular formulas is a well-established and reliable approach.

We have used widely accepted computational frameworks to infer structural features, such as aromaticity index (AI), double-bond equivalents (DBE), nominal oxidation state of carbon (NOSC), and compound classification. The computational methods used to infer structural features from molecular formulas are reliable and widely accepted in the scientific community, such as Van Krevelen Diagram (*Kim et al., Anal. Chem 2003, 75, 5336-5344*), Formularity (*Tolic et al., Anal. Chem, 2017, 89: 12659−12665*), UltraMassExplorer (*Rapid Communications in Mass Spectrometry, 2019, 2: 193-202*), ftmsRanalysis (*Bramer et al., PLoS Comput Biol, 2020, 3, e1007654*), PyKrev (*Kitson et al., Journal of the American Society for Mass Spectrometry, 2021, 5, 1263-1267*) and the Gaussian-based Alignment Algorithm (*Fu et al., Anal. Chem, 2023, 95, 2796-2803*). These approaches have been used extensively to calculate molecular structural features. Such computational frameworks are standard in DOM analysis and has been applied in numerous high-impact studies, including those published in Nature family journals, where specific subset structures for DOM molecules are not provided.

Examples include:

Kellerman et al., Nature Geoscience, 2015, 6: 454-457 (AI, NOSC, etc.)

McDonough et al., Nature Communications, 2022, 1: 2513 (DBE, AI, NOSC, etc.)

Kellerman et al., Nature Communications, 2014, 1: 3804 (AI, etc.)

Hu et al., Nature Communications, 2022, 1: 3600 (AI, DBE, NOSC, etc.)

Osterholz et al., Nature Communications, 2015, 1: 7422 (DBE, etc.)

Hu et al., Nature Communications, 2023, 1: 3681.(AI, DBE, NOSC, etc.)

Zhou et al., Nature Communications, 2024, 1: 6356 (AI, DBE, NOSC, etc.)

Osterholz et al., Nature Communications, 2022, 1: 4837 (AI, etc.)

Hu et al., Nature Communications, 2024, 1: 576 (AI, DBE, NOSC, etc.)

Zark et al., Nature Communications, 2018, 1: 3178 (AI, etc.)

Danczak et al., Nature Communications, 2020, 1: 6369 (AI, etc.)

By integrating advanced computational analyses (AI, DBE, NOSC, etc.) on ultra-high-resolution MS data, we provide new insights into the pathways of organic matter transformation during AD. These insights go beyond raw formula assignments; rather, we cluster and categorize compounds to illustrate overarching reactivity trends, which is especially relevant for

understanding and optimizing bioprocesses in food-waste-derived digestates.

In summary, while we appreciate that more extensive structural verification using complementary techniques (e.g., MS/MS, 2D-NMR) can further validate individual molecular assignments, the logistical challenges make comprehensive direct structural characterization unfeasible for the scope of this study. FT-ICR MS remains the most robust technique available for mapping the immense molecular diversity in DOM, and our reliance on standard computational frameworks and cross-validation with known transformations ensures that our conclusions are both methodologically sound and widely accepted within the scientific community.

Second, FT-ICR MS results are largely qualitative, leading to speculative conclusions, and there is a lack of direct measurement data on microbial changes.

Response: We have supplemented our revised manuscript with microbial compositional and functional analyses for AD processes across seven food waste treatment plants (FWTPs). Microbial diversity showed a significant increase from primary digestion (PD) to secondary digestion (SD) (**Fig. R1**a), reflecting an expanded functional capacity for transforming diverse DOM molecules during the SD stage. This increase in diversity is often associated with greater metabolic versatility, facilitating the breakdown of complex organic matter.

Among the DOM compound classes, lipids were the only group to exhibit a significant increase from PD to SD, indicating lipid accumulation (**Fig. 2**b and **Fig. 3**c). This finding aligns with microbial functional analyses, which revealed a slight reduction in lipid metabolism during SD compared to PD (**Fig. R1**b). Notably, despite the more favorable fermentation conditions in SD (pH 7.77 vs. pH 3.87 in PD), lipid metabolism remained inactive and was even significantly inhibited, contributing to the observed lipid accumulation. Lignin-like compounds, the most abundant compound class and characterized by their richness in aromatic rings, showed a significant decrease. This reduction corresponds to the enhanced microbial metabolism of aromatic compounds, including cofactors, vitamins, terpenoids, and polyketides. Proteins, the third most abundant class, showed minimal changes, likely due to ammonia-induced inhibition of microbial activity. This explanation is supported by the active metabolism of amino acids to ammonia observed during SD (**Fig. R1**b). These findings provide a nuanced understanding of how microbial community dynamics and metabolic functions drive DOM transformations in AD systems.

Additionally, Mantel analysis confirmed a strong correlation between microbial composition changes and shifts in DOM compound classes such as amino sugars, carbohydrates, and lipids, as well as elemental combinations like CHO and CHONS. Molecular structure evolution, including DBE changes, was also closely linked to microbial functions (**Fig. R1**c). These results demonstrate a clear alignment between microbial metabolism and DOM molecular

transformations, thereby enhancing the validity and robustness of our findings.

We have included the discussion in the revised *Supplementary Information* (**Supplementary Note 1** and **Supplementary Fig. 4**).

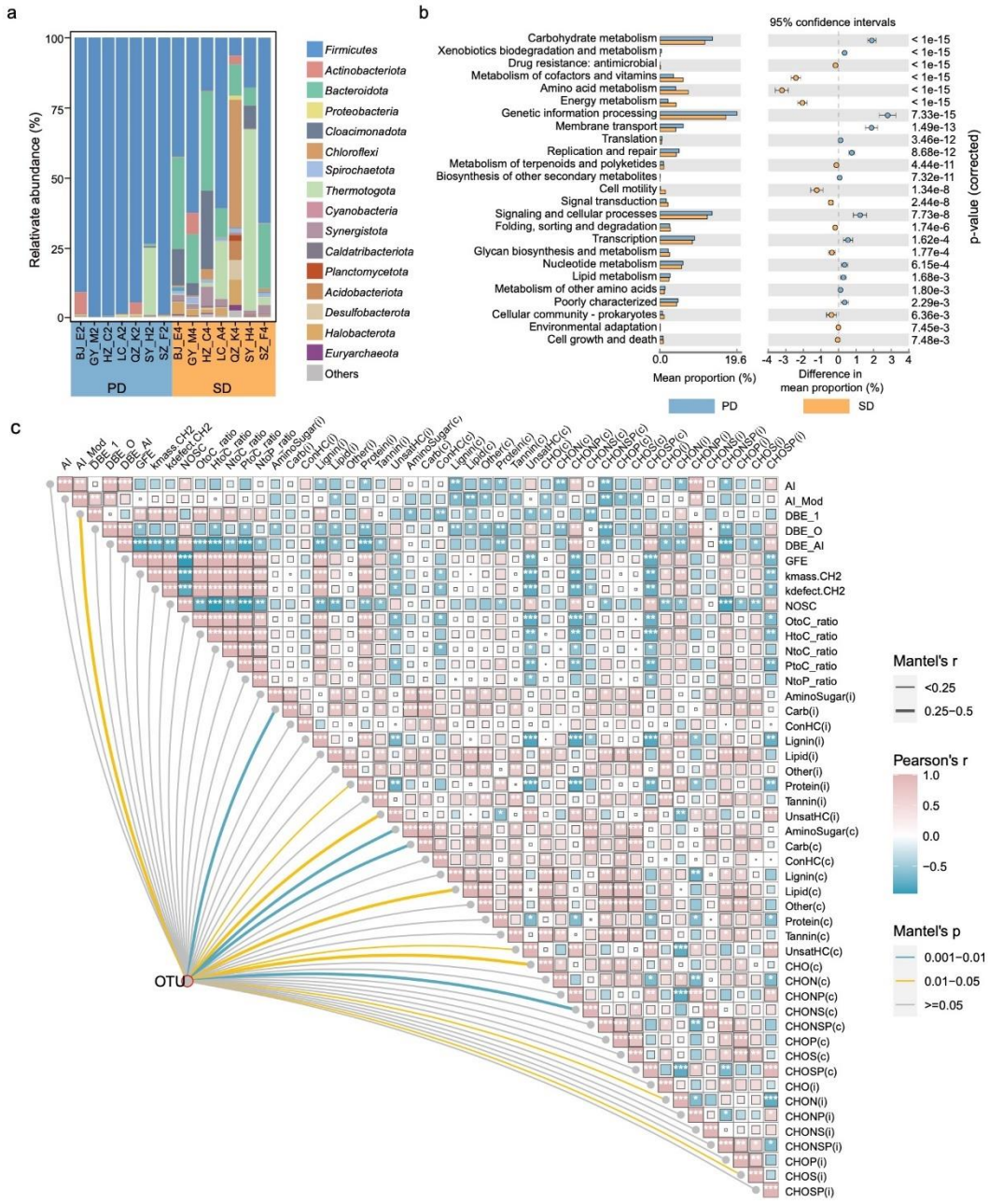

Fig. R1 Microbial community composition and functional analysis in the anaerobic digestion (AD) of seven food waste treatment plants (FWTPs). (**a**) Relative abundance of the major microbial groups and (**b**) functional analysis in PD-SD two-stage AD processes based on amplicon sequencing of the 16S rRNA gene; (**c**) Pairwise comparisons of DOM molecular characteristics are shown, with a color gradient denoting Spearman's correlation coefficients. Taxonomic community composition was related to each molecular parameters by Mantel tests. Edge width corresponds to the Mantel's r statistic for the

corresponding distance correlations, and edge color denotes the statistical significance. The “i” and “c” in brackets represent “molecular intensity composition” and “molecular count composition”, respectively.

Therefore, while the research content and findings are interesting, the heavy reliance on molecular composition data reduces the credibility of the results. We strongly recommend supplementing the manuscript with microbial data as mentioned and structural data on molecular composition to enhance the credibility of the findings.

Response: We sincerely appreciate the reviewer’s recommendation to supplement our work with microbial data and additional structural insights. As detailed in our responses above, we have added microbial compositional and functional analyses to our revised manuscript. This addition allows us to directly link microbial community dynamics and metabolic functions to the observed DOM molecular transformations.

While our FT-ICR MS-based framework cannot directly confirm molecular structures, it is grounded in rigorous experimental protocols and provides reliable insights for DOM transformations in complex AD systems. Key clarifications include:

- **Scientific experimental design**. Unlike many studies that focused on samples from a single site or a limited number of samples (typically two: pre- and post-fermentation) in controlled laboratory settings, our study collected samples from widely distributed FWTPs across diverse geographic regions. This approach allowed us to identify more "global" trends and derive more reliable common patterns based on average or core molecular data.
- **Reliable quantitative analysis**. By employing standardized FT-ICR MS protocols with both internal and external calibrations, we were able to use intensity changes for a more accurate relative quantification of DOM transformations. This method offers greater scientific rigor compared to many previous studies, which relied mostly on molecular count changes.
- **Transparent acknowledgment of method limitations.** We explicitly clarify in our revised manuscript that our approach is based on molecular formula assignments rather than direct structural proof:

  “*These classes are a priori assignments that do not rely on any measured proof of identification, such as structural confirmation by MS/MS, and are used solely to categorize the data for easier interpretation*”. (Line 478-480 in our revised manuscript)

  Moreover, we discuss the challenges of detecting smaller molecules such as VFAs, which may be underrepresented in FT-ICR MS analyses but significantly impact AD processes:

  “*The DOM pool experiences a smaller reduction compared to COD and SCOD during the SD process (Fig. 1a and Fig. 2b), suggesting that the inflow of DOM molecules from*

organic solid solubilization outweighs their breakdown. Another possible reason is that FT-MS–undetected molecules are notably higher in PD than in SD, given that smaller molecules such as VFAs are considerably more abundant in PD than in SD (Fig 1a and Supplementary Table 2)." (Line 338-343 in our revised manuscript)

- **Rigorous discussion and future directions**. In the revised manuscript and *Supplementary Information*, we discuss how factors such as molecular composition, structural features, and microbial metabolism collectively influence the composition and transformation of DOM throughout AD, supported by relevant references. While we acknowledge that combining multiple MS platforms, NMR, and multi-omics approaches could further refine the structural characterization of DOM, such expansions require substantial method and instrumentation developments. We posit that our high-resolution, quantitative FT-ICR MS data, coupled with new microbial insights, already provides valuable guidance for optimizing organic waste treatment processes.

In summary, we have integrated new microbial data and clarified the strengths, limitations, and broader applicability of our FT-ICR MS-based approach. We believe these additions significantly enhance the credibility of our findings and offer a comprehensive understanding of DOM molecular transformations in food waste AD systems. We are grateful for the reviewer's constructive feedback and hope these revisions sufficiently address all concerns.

Response to Reviewer 2's Comments:

AD is widely used for the treatment of organic wastes, and the microbial communities were well characterized previously. However, the transformation of the complex organics was not well studied. This is an interesting study that investigated the DOM transformation in AD process by FT-ICR MS analysis with a novel analytical framework. The molecular properties of DOM and their transformation in AD were well characterized, which provided new insights for the organic conversion in AD. The manuscript was well written and can be accepted for publication after considering the following comments.

Response： Thank you for your positive feedback and thoughtful comments. We have carefully revised our manuscript in response to the comments received. Below, we address each comment point by point and have made the corresponding revisions to the manuscript.

1, Line 42, what kind of variables? Please specify.

Response: The term "variables" refers to factors that impact waste/wastewater treatment processes, such as pH, temperature, and organic loading. We have revised the sentence in the manuscript to specify these variables for clearer understanding (Lines 42 in revised manuscript).

2. Line 57, what are simplified systems? Please add more information. There are also studies using FT-ICR MS to investigate the DOM in AD treating complex substrates like sludge. Please be more accurate.

Response: Thank you for your comments. Here for "simplified systems", we meant lab-scale experiments using model solutions or controlled environments that do not fully replicate the complexity of industrial AD systems treating diverse organic wastes. We acknowledge that studies have used FT-ICR MS to investigate DOM in AD processes treating complex substrates such as sludge. However, many of these studies focus on lab-scale setups where sludge is used as initial inoculum under control conditions.

To enhance accuracy and avoid misunderstanding, we have revised the terminology to "lab settings" (line 60 in revised manuscript).

3. Line 59, "doe" should be "due"

Response: Thank you for pointing out this typographical error. Corrected (Line 63 in revised manuscript).

4. Line 124-125, why was the increase of CHONS from lipid accumulation? Please explain.

Response: We apologize for the confusion. Our initial analysis inadvertently misassigned some lipid compounds as CHONS due to an error in formula assignment. After carefully rechecking and correcting these assignments, we found that CHONS compounds did not actually increase during the AD process. The observed lipid accumulation primarily arose from CHO compounds, rather than CHONS molecules, which remained relatively unchanged across all samples (**Fig. 2**a).

5. Line 163, why did two samples show atypical compositions? Please explain.

Response: We appreciate the reviewer's concern about atypical compositions in two samples. After carefully reanalyzing all samples in our revised manuscript, we confirm that no atypical compositions were detected. We have now incorporated all samples into the core molecular community analysis, ensuring a robust and comprehensive representation of the DOM composition across the study.

6. Line 204, were the newly produced molecules from the solubilization of solid organics or the degradation of DOM ?

Response: Thank you for the insightful question. Our analysis indicates that the newly produced molecules in the SD phase are primarily the result of solubilization of solid organics rather than degradation of existing DOM. As shown in **Fig. 4**a, 32.1% of total DOM molecules originated from the PD and were completely broken down in the SD. Newly produced molecules in the SD accounted for 48.5% of the total. The high concentration of organic solids in food waste sludge (**Fig. 1**a) serves as the primary source for these new DOM molecules through solubilization processes. While microbial biosynthesis using small DOM molecules and biogas does contribute, its impact is significantly smaller compared to the solubilization of solid organics and DOM degradation (**Fig. 6**a). We have clarified this point in the revised manuscript: *"Organic solids are a primary source of DOM molecules, especially for high-solid AD. The DOM pool experiences a smaller reduction compared to COD and SCOD from PD to SD (Fig. 1a and Fig. 2b), suggesting that the inflow of DOM molecules from organic solid solubilization outweighs their breakdown. Another possible reason is that FT-MS–undetected molecules are notably higher in PD than in SD, given that smaller molecules such as VFAs are considerably more abundant in PD than in SD (**Fig 1**a and **Supplementary Table 2**). The molecules unique to SD samples are more diverse than those of PD samples, as well as those common to PD and SD samples (**Fig. 4**a). This indicates that recalcitrant DOM from biological solubilization is the main component of the final effluent, which is evidenced by the increasing dominance of lipid-like molecules and the persistence of lignin-like molecules (**Fig. 2**a and **Supplementary Figs 20-21**), despite the removal of large amounts of DOM through the biogas."* (Line 337-347 in revised manuscript)

7. Lin 310-312, lipid can be degraded during AD process, why were they still present in the effluent?

Response: We acknowledge that lipids can be degraded during AD process. However, owing to the structural features of lipids, their degradation requires significantly more energy and occurs at a much slower rate compared to readily degradable compounds such as carbohydrates, amino sugars, and proteins (**Fig. 2, 3, 5**). As a result, the microorganisms

responsible for lipid degradation in AD processes typically do not occupy the core ecological niches needed to effectively degrade these lipid molecules. Furthermore, the structural complexity of lipids, especially long-chain fatty acids, makes them more resistant to biodegradation. We have expanded the discussion in the revise manuscript:

"Lipid-like molecules showed a significant increase in both richness and abundance, primarily due to their inherent resistance to microbial degradation. Additionally, biological hydrolysis and conversion processes likely contributed to the further accumulation of soluble lipid molecules." (Line 139-142)

"Lipids, which have a higher proportion of C-C single bonds, accumulated significantly during the SD process. In contrast, lignin, with fewer C-C single bonds, was more readily degraded." (Line 164-167)"

"The primary DOM compounds in the sludges include lipid, lignin and unsaturated hydrocarbons pose greater thermodynamic challenges for microbial breakdown due to their higher energy demands and less favorable redox conditions." (Line 315-318 in revised manuscript)

8. Line 339-351, the traditional methods were mentioned to increase the DOM conversion. It is better to propose new methods based on the new understanding of the DOM conversion in this study.

Response: We appreciate this suggestion to propose new methods informed by our findings. The purpose of lines 339-351 is to support the co-treatment strategy we propose, which combines the treatment of wastewater (recalcitrant DOM molecules) and organic solids in one reactor (**Fig. 6**b). To advance this strategy, we propose the following methods for targeted degradation:

- **Addressing lignin persistence**: Given that lignin is a major component of the persistent DOM and rich in benzene ring structure, we can use dioxygenase-catalyzed ring-opening and O-demethylation strategy, like using enzymatic treatment or advanced oxidation processes to initiate lignin depolymerization.
- **Enhancing lipid degradation**: Lipids, composed primarily of long-chain molecules with C-C single bonds, require strategies to enhance β-oxidation pathways. This could involve introducing or enriching specific microbial strains capable of efficient lipid degradation or modifying existing microbial communities to enhance their lipid-degrading capabilities.

More detailed information on integrating these methods into the AD process is provided in the Discussion section of the main text (lines 379-398).

Response to Reviewer 3's Comments:

Review for "Decomposing the molecular complexity and transformation of dissolved organic matter for innovative anaerobic bioprocessing", by Jun Hu, Chuan-Guo Liu, Wen-Kai Zhang, Xue-Wen Liu, Bin Dong, Zhan-Dong Wang, Yuan-Guo Xie, Zheng-Shuang Hua and Xian-Wei Liu* submitted to Nature Communications.

General comments

Jun Hu, Xian-Wei Liu and colleagues study the compositional change of DOM in anaerobic digestion of food waste of seven treatment plants in China and compare the two different stages (primary and secondary) of the process. They show that across sites, DOM changes in a similar way and provide ultrahigh resolution mass spectrometric data to make their case. Compared to other papers in the field that I found in a quick search, their paper peaks out by the replication of the process in seven different sites, whereas most other papers show data from one site only. This highlights why their paper could be of interest to the readership of Nature Communications, as it starts to look for more "global" trends in waste treatment and valorization.

Despite this positive outlook, I have my problems with the data that is presented and the innovation that is claimed. This is mostly related to the molecular formula data that forms the basis of the submission. Problem 1) is the high number of CHONSP assignments that I did not find well reflected in the literature and that seem unjustified at a resolving power of 80.000 only. Problem 2) is the lack of information regarding how the molecular formulas were obtained, and how the data was treated to ensure high quality, most importantly if blanks and replicate measurements were used to filter out false positives, and how multiple assignments were treated. Problem 3) relates to the innovation of the paper. A central figure of the paper (Fig 5, especially 5c) seems to be a 1:1 version of similar figures in recent papers (Gunina & Kuzyakov, 2022; Wang & Kuzyakov, 2023) including the text shown in the figure, but all this without proper referencing. It also is unclear to me how the ΔH data was generated in the first place. There are other issues, but these are the three main ones in my view. Find below my comments in detail; I added a → to indicate what I deem important in the revision.

Response: Thank you for your detailed and constructive feedback. We appreciate the time and effort you have invested in evaluating our work. Your insights have been invaluable in helping us improve the quality and clarity of our study. Below, we address each of your comments in detail, including both the general concerns and specific points you have raised:

1) **High Number of CHONSP Assignments (Problem 1)**:

We discovered a significant oversight in our initial analysis that positive-ion-mode formula assignments were inadvertently applied to FT-ICR MS data acquired in negative-ion mode, resulting in an inflated ratio of CHONSP formulas. After carefully reanalyzing the raw data using the correct ion mode assignments, **CHONSP molecules account for 4.4% for PD samples**

and 4.7 for SD samples (Table R3 and **Fig. 2**a in the revised manuscript), aligning more closely with values reported in the literature.

Furthermore, we would like to clarify that our FT-ICR MS instrument operated at a resolving power of >320,000 at m/z 400 (not 80,000 as previously stated), with a mass accuracy of ⩽ 0.5 ppm. These parameters are more than sufficient for reliable molecular formula assignments in complex environmental samples.

We have updated the manuscript to include these corrections and clarifications. We sincerely apologize for the confusion caused by the initial mismatch of ion modes.

2) **Methodology for DOM Detection and Analysis (Problem 2)**

We recognize the importance of detailing our molecular formula assignments and data quality assurance. Our methodology followed the protocol established by Quan Shi lab (*He et al., ACS omega, 2020, 20: 11730-11736*), with data analysis using Formularity (*Tolic et al., Anal. Chem, 2017, 89: 12659−12665*) and ftmsRanalysis (*Bramer et al., PLoS Comput Biol, 2020, 3, e1007654*), developed by Pacific Northwest National Laboratory (PNNL). These well-established methods ensure the reliability and traceability of our DOM analysis. Our formula assignments followed the "golden rules" (*Kind T. and Fiehn O., BMC bioinformatics, 2007, 8: 1-20*, incorporating strict elemental limitations to reduce the likelihood of improbable assignments. These details, including a summary of the results in **Table R1**, have been further elaborated in responses to Specific Comments 1 and 2.

To address concerns about reliability, we have updated the manuscript to include a comprehensive description of our FT-MS testing methodology and rigorous quality control measures (Lines 444-498 in the revised manuscript and **Supplementary Methods 1-2**).

3) **Innovation and Figure 5 Concerns (Problem 3)**:

Quantitative models of organic matter degradation are essential for understanding the chemical state and evolution of natural and engineering organic systems. A pioneering study by LaRiwe et al. developed the Gibbs energies-NOSC relationship for estimating the energetic potential of compounds based on major elements (C, H, O, N, P, S) ratios, which enables us evaluating the energetics of oxidative degradation of a large number of naturally occurring organic compounds (*Giochimica et al., Cosmochimica Acta 2011, 8: 2030-42*). Gunina and Kuzyakov simplified the calculation of oxidative reactions and used combustion enthalpy per carbon to evaluate the energy potential of soil organic matter, achieving a linear energy-NOSC relationship of soil organic matter (*Gunina, A. and Kuzyakov Y., Global Change Biology 2022, 7: 2169-2182*). Subsequently, Wang and Kuzyakov updated the relationship between energy content and NOSC of compounds by considering additional 52 compounds to obtain a more precise equation (*Wang C. and Kuzyakov Y., Global Change Biology 2023, 22: 6170-87*).

In this study, we applied the framework developed by Gunina and Kuzyakov and also cited the key reference (ref. 45 in revised manuscript). We acknowledge that the need to cite additional references, including Giochimica et al., 2011 (ref. 44) and Wang & Kuzyakov, 2023 (ref. 46). Our study, however, differs from Kuzyakov's work on soil organic matter in several key aspects:

- **Scope and representativeness.** While Kuzyakov et al. focused on soil organic matter with compounds largely sourced from textbooks, our study centers on DOM in food waste sludge. Our selected compounds, with NOSC values primarily between -2 and 1, are more representative of food waste sludge (**Fig. 5b**), and we have provided the raw data for our analysis in **Supplementary Table 8** for transparency.
- **Data and analysis.** Unlike Kuzyakov et al., who did not use actual data from soil organic matter, our results are based on direct FT-ICR MS measurements of DOM in food waste sludges (**Fig. 5a)**. In **Fig. 5b,** we compare actual DOM data from PD and SD stages to theoretical predictions, while **Fig. 5c** plots density-weighted Δ*H*-NOSC relationships that more accurately, offering a more accurate and comprehensive comparison of bioavailability of DOM form PD to SD. Additionally, the observed relationships in real DOM samples yields lower R^2 values (0.52 for PD samples and 0.54 for SD samples) than both our theoretical relationship (R^2 = 0.72) and Kuzyakov et al.'s reported 0.98, suggesting that Δ*H*-NOSC relationships in real-world organic environments are less linear, particularly for DOM molecules containing N, P, and S elements.

In summary, by applying the Δ*H*-NOSC framework to food waste sludge DOM, our study provides a more accurate assessment of DOM degradability in real anaerobic digestion systems. This approach can guide better operational strategies for anaerobic digesters, complementing the soil-focused models in the literature.

Table R1 Formula assignment information for all the DOM samples in our study

	Golden Rules		User Defined	BJ_PD	BJ_SD	GY_PD	GY_SD	HZ_PD	HZ_SD	LC_PD	LC_SD	QZ_PD	QZ_SD	SY_PD	SY_SD	SZ_PD	SZ_SD
Mass Range (m/z)	For mass < 1000 m/z		100-1000	117-999	114-999	101-1000	149-994	115-999	105-999	101-1000	112-1000	101-1000	115-992	111-1000	103-998	115-1000	109-999
Element Limits	C	C_{max}=78	Not restricted	4-75	4-73	3-75	4-71	4-74	4-75	3-76	3-75	3-74	4-63	3-75	4-75	4-75	6-71
	H	H_{max}=126	Not restricted	4-118	4-109	4-122	5-121	4-122	4-124	3-125	2-119	4-125	4-125	4-123	4-119	3-125	6-119
	O	O_{max}=27	O_{min}=1	1-24	1-25	4-26	1-26	1-25	1-26	1-26	1-26	1-26	1-23	1-26	1-26	1-26	1-26
	N	N_{max}=25	N_{max}=3	0-3	0-3	0-3	0-3	0-3	0-3	0-3	0-3	0-3	0-3	0-3	0-3	0-3	0-3
	S	S_{max}=14	S_{max}=2	0-2	0-2	0-2	0-2	0-2	0-2	0-2	0-2	0-2	0-2	0-2	0-2	0-2	0-2
	P	P_{max}=9	P_{max}=1	0-1	0-1	0-1	0-1	0-1	0-1	0-1	0-1	0-1	0-1	0-1	0-1	0-1	0-1
	H_{max} = C*2 + 2 + N + P			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Element ratios	H/C	0.2-3.1	0.3-3	0.3-2.8	0.3-2.5	0.3-2.8	0.3-2.9	0.3-2.5	0.3-2.8	0.3-2.9	0.3-2.8	0.3-2.9	0.3-2.5	0.3-2.9	0.3-2.9	0.3-2.9	0.3-2.5
	O/C	0-1.2		0.0-1.2	0.0-1.2	0.0-1.2	0.0-1.2	0.0-1.2	0.0-1.2	0.0-1.2	0.0-1.2	0.0-1.2	0.0-1.2	0.0-1.2	0.0-1.2	0.0-1.2	0.0-1.2
	N/C	0-1.3		0.0-0.4	0.0-0.5	0.0-0.8	0.0-0.5	0.0-0.4	0.0-0.6	0.0-0.8	0.0-0.7	0.0-0.8	0.0-0.4	0.0-0.8	0.0-0.6	0.0-0.5	0.0-0.4
	S/C	0-0.8		0.0-0.5	0.0-0.4	0.0-0.7	0.0-0.4	0.0-0.5	0.0-0.5	0.0-0.7	0.0-0.7	0.0-0.7	0.0-0.4	0.0-0.7	0.0-0.4	0.0-0.5	0.0-0.3
	P/C	0-0.3		0.0-0.2	0.0-0.2	0.0-0.2	0.0-0.2	0.0-0.2	0.0-0.3	0.0-0.3	0.0-0.3	0.0-0.3	0.0-0.3	0.0-0.3	0.0-0.2	0.0-0.2	0.0-0.2
DBE	integer & >=0			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
S/N			>7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Error			0.5 ppm	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Lastly, I would advise that the revised version of the paper builds more upon the work of Liu et al. (2023), Water Research and Zhao et al. (2023) Water Research. Zhao et al. found that compositional analyses of sludge DOM are clearly dependent of the DOM isolation and ionization modes, and may affect the changes one can observe between initial wastewater sludge and anaerobic digests; this is an aspect that is missing from your discussion so far. Zhao et al. showed that O/C and NOSC largely explained the functional response of molecules during wastewater treatment (albeit partly after pretreatment of the sludge) and that CHNOS (specifically, S-containing) molecules, although small in number, were potential keystone molecules explaining treatment performance; although your results seem to agree with literature reports regarding dominance of lipids, proteins and lignin-related molecules, they tend to disagree when it comes to the relevance of the P-containing formulas; these aspects of comparability need to be discussed and potential systematic errors in data treatment (over-assignment of P formulas) need to be disproven by the authors before the manuscript is acceptable.

Response: Thank you for your insightful comments and for directing us to the studies by Liu et al. (2023) and Zhao et al. (2023). We have carefully reviewed these works and agree that incorporating their findings will strengthen our manuscript.

We acknowledge that compositional analyses of sludge DOM are influenced by sample preparation methods and ionization modes, as noted by Zhao et al. (2023). They demonstrated that the proportion of P-containing formulas can vary widely (from less than 5% to 39%), influenced by various factors such as the DOM isolation and ionization modes (summarized in **Table R2**). Our findings show that lignin, lipids and proteins are the dominant compound classes, which are consistent with Liu et al. (2023) and Zhao et al. (2023). In our revised DOM analysis, P-containing formulas account for 16.5% for PD samples and 14% for SD samples, somewhat higher than the average values of reports in sewage sludge studies. This may result from the inherent characteristics of food waste sludges and the analytical methods we employed (See the responses to General Comment (Problem 2), Specific comment 2 and **Fig. R3**).

We have updated the manuscript to include a detailed discussion of these points and to acknowledge the influence of sample preparation and ionization modes on DOM compositional analyses. (Line 130-134 in revised manuscript).

Specific comments

1) Line 401: If your mass resolution was, as you state, 80.000 @ m/z 400 I seriously doubt your CHONSP assignments, especially because they are also that high in number – CHONSP peaks are by far the most abundant but will be hard to resolve under these

instrumental conditions.

a. Did you use ^{13}C and ^{34}S isotopes to provide additional support of the most abundant of these assignments? Show some of your mass spectra and specific mass spectral features to prove your point, and include them in the SI (see for example, Figure 1 in Shakeri Yekta et al. 2012) (Shakeri Yekta et al., 2012).
b. Can you provide other references that have shown similar molecular class (CHO, CHNO, etc.) distributions for food waste DOM? It would help if you could support your findings here. E.g., Liu et al found that CHNOSP formulas only accounted for <5% of their molecular count data while CHNO and CHOS formulas were more important (Liu et al., 2023). Xiao et al. state that proteins account for 50 w% of sludge DOM, while lipids only account for up to 5 w% (Xiao et al., 2020). Hu et, al. showed that lipids, proteins and carbohydrates were formed in sludge processing and contained N (Huet al., 2018). Li et al. showed that sludge was mainly composed of CHNO and CHNOS formulas (Li et al., 2023). Shakeri Yekta et al. found that CHNOS class was only 1-9% of their molecular formula data (Shakeri Yekta et al., 2012). Zhao et al. report that wastewater sludge and anaerobic digest sludge are rich in protein mainly with pointing out the effect of SPE as well; in their data it becomes evident that anaerobic sludge DOM contains only little P formulas, with highest contribution from CHNO and CHOS, and is composed of 40% lignin-, 25% protein-, 30% lipid-like formulas (measured by ESI- after SPE isolation, Fig 2b); i.e., their data are in line with your findings in terms of molecular groups but not formula classes and furthermore strongly depending on the analytical method you chose (Zhao et al., 2023). provide critical comparison of formula classes (CHO, CHNO, etc.) and molecular groups (lipids, lignin, protein) to other studies in your early discussion section, and mention potential drawbacks by citing other refs.
c. I look at the data from the environmental DOM and soil DOM standpoint, where the CHO class is by far the most abundant, followed by CHNO and CHOS; especially assignments with P are dealt with extra care because P is usually low in our samples but ecologically highly important. This may be different in your samples, where P could be more abundant (as may be N and S). P assignments can also be due to mis-assigned ^{35}Cl atoms which could be ruled out by including ^{35}Cl into the assignment.
d. Can you back up the high N, S, P content with complementary data, e.g. from elemental analysis of DOM or biosolids (i.e., element totals or organic fractions of C, H, O, N, S, P)?
e. Usually there's a few CHO-only molecular formulas as similarly appropriate candidate within a 0.5 ppm tolerance limit. Please report more specifically how the FTICRMS

was externally and internally calibrated. Please provide the formulas of the ions you used ("Bruker tuning mix"). If your data lacks CHO formula suggestions, it may be you're your data was not calibrated correctly.

f. Did you measure duplicate spectra of each sample to rule out false positives (i.e., by only using doubly detected peaks)?

Response: We appreciate your critical observation regarding the mass resolution and its impact on the reliability of our CHONSP assignment. We apologize for the confusion caused by a typographical error in our original manuscript. The correct mass resolution for our instrument is >320,000 at m/z 400, not 80,000 as previously stated. A resolving power of >320,000 at m/z 400 provides sufficient mass resolution and accuracy to confidently assign molecular formulas, including those with multiple heteroatoms such as N, S, and P, in complex mixtures like DOM from food waste sludge.

We have corrected this information in the revised manuscript (Methods section, line 444-480 in revised manuscript, **Supplementary Fig. 27**) and ensure that all instrument parameters are accurately reported.

a. In this study, we did not include ^{13}C and ^{34}S isotopic data in our previous tests. However, we have provided mass spectra of the PD and SD DOM samples from Shenzhen (SZ) to illustrate specific spectral features that support our assignments (**Fig. R2**).

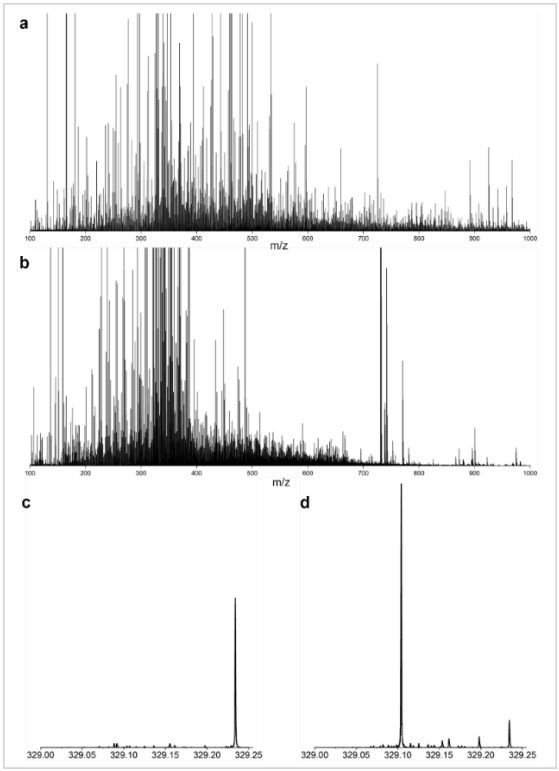

Fig. R2 ESI negative FT-ICR mass spectra of DOM from (**a**) PD and (**b**) SD of Shenzhen (SZ) sludge samples. The zoomed areas of (**c**) PD and (**d**) SD at nominal mass 329 are shown to underline the

differences between samples and resolution power.

b. We appreciate the opportunity to contextualize our findings within broader literature. We have made a thorough comparison with relevant studies.

Several studies have investigated the molecular composition of DOM in wastewater sludge and anaerobic digestion processes, often reporting different distributions of elemental combinations and compound classes. Notably, Liu et al. (2023) and Zhao et al. (2023) reported that the ratios of P-containing formulas can exceed 30%, with lipids, lignin, and proteins being the three most abundant compound classes in wastewater sludge. These findings are consistent with our results. In contrast, Li et al. (2023) and Shakeri Yekta et al. (2012) did not analyze P-containing formulas. In our revised DOM analysis, P-containing formulas represent 16.5% in PD samples and 14% in SD samples, specifically, **CHONSP compounds account for 4.7%**. These values fall within the expected ranges, consistent with findings from the referenced studies and prior research.

As food waste treatment in China is transitioning from traditional landfill methods to anaerobic bioprocessing, there are few studies focusing on DOM of food waste AD. To the best of our knowledge, only one recent paper has reported on this topic (Yang et al., Microbiome, 2024, 12:16). We have systematically compared Yang et al.'s data with our results, as summarized in **Tables R2 and R3**. Yang et al. identified CHO, CHON, and CHONS as the top three elemental combinations, which is in agreement with our results. Yang et al. identified lignin, lipids, and unsaturated hydrocarbons as the three most abundant compound classes, which aligns closely with our findings, where lignin, lipids, and proteins rank as the top three, followed by unsaturated hydrocarbons in fourth place. Additionally, their reported P-containing formulas, accounting for 18%, are consistent with our data, which show 16.5% for PD samples and 14% for SD samples.

Table R2 Typical Studies on DOM in Anaerobic Digesters

Feedstock	Waste processing	SPE treatment	Elementary combinations	Compound compositions	References
Dewatered sludge	Persulfate-pretreatment + AD	the Agilent Bond Elute PPL cartridges (500 mg/6 mL)	Top three: CHON, CHOS, CHO; **CHONSP<5%**	Top three: **Lignin, proteins**, **lipids**	Liu et al., 2023
Sludge	Ultrasonication and acidification treatments	NA, Spectrometry methods	NA	Protein 40-50w% and lipids 3-5w%	Xiao et al., 2020
Wastewater	Active sludge process	The functionalized styrene-divinyl-benzene polymer resin (PPL), Supelco	NM	Top three of **DON:** Proteins/amino sugars, **lipids**, carbohydrates	Hu et al., 2018
Dewatered sludge	AD	The PPL column	Top three: CHO, CHON, CHONS; P-DOM was excluded	Top two of DOM: Condensed aromatic structures, proteins	Li et al., 2023

Municipal solid waste, dry fodder, manure, fat, etc.	full-scale AD	the Agilent Bond Elut solid-phase extraction cartridges	Top three: CHO, CHOS, CHON; CHONS, 1.3-8.6%; P-DOM was excluded	NA	Shakeri Yekta et al., 2012
Wastewater sludge	Activated sludge (AS) process/AD process	Styrene divinyl benzene polymer (PPL), Octadecyl (C18) Electrodialysis (ED)	TOP tree: (AS/AD) 1st: **CHOS/CHON**, 2nd: CHO/CHO 3rd: CHON/CHOS **P-containing:** (AS/AD) **ED+: 39%**/<11% ED-: <12%/<5% PPL+: 12%/**26%** PPL-: <5%/<5% C18+: 12%/<11% C18-: <5%/<5%	Top one (AS/AD): ED+: **Lipids**/Lignin ED-: **Lipids**/Lignin PPL+:**Lipids**/Proteins PPL-: **Lipids**/Lignin C18+: **Lipids/Lipids** C18-: **Lipids/Lipids**	Zhao et al., 2023
Food waste	AD	the Bond Elut-PPL straight barrel (1 g, 6 mL Agilent, USA)	Top three: CHO, CHON and CHONS **P-containing DOM: 18%**	Top three: lignin, **lipids**, unsaturated hydrocarbons	Yang et al., 2024
Food waste	AD in FWTPs	HyperSep™ retain PEP, Thermo Scientific	Top three: CHO, CHON and CHONS **P-containing DOM: 16.5%/14%** **CHONSP:4.4%/4.7%**	Top three: lignin, lipids and proteins	This study

Table R3 Comparison of elemental combinations of DOM in food waste sludge from Yang et al. (data derived from **Fig. S1** of Yang et al., Microbiome, 2024) and our study

Elementary combinations	**N**-containing DOM				**P**-containing DOM				**S**-containing DOM			
	CHON	CHONS	CHONP	CHONSP	CHONP	CHONSP	CHOP	CHOSP	CHONS	CHONSP	CHOS	CHOSP
Yang et al.	24.0%	15.8%	8.7%	0	8.7%	0	6.0%	3.3%	15.8%	0	13.8%	3.3%
PD (average)	32.4%	17.7%	6.0%	4.4%	6.0%	4.4%	2.3%	3.8%	17.7%	4.4%	9.6%	3.8%
SD (average)	34.2%	19.4%	4.0%	4.7%	4.0%	4.7%	2.5%	2.8%	19.4%	4.7%	10.3%	2.8%
Yang et al. (sum)	48.5%				18.0%				32.9%			
PD (sum)	60.5%				16.5%				35.5%			
SD (sum)	62.3%				14.0%				37.2%			

c. We acknowledge that CHO compounds often dominate environmental and soil DOM studies. In our revised analysis, the intensity-weighted ratio of CHO compounds reached 41.0% (PD) and 52.3% (SD), both the highest among all compound types. However, in terms of molecular counts, CHO accounted for 23.7% (PD) and 22.1% (SD), ranking second after CHON compounds. This discrepancy indicates that CHO compounds exhibit relatively lower molecular diversity but higher molecular abundance.

Additionally, we used the Isotopic Pattern Algorithm (IPA) function of Formularity (*Tolic et al., Anal. Chem, 2017, 89: 12659−12665*) to assess the potential impact of misassigned 35Cl atoms on phosphorus (P) assignments. To ensure the reliability of our formula assignments, we excluded all formulas containing P, as identified by the Compound

Identification Algorithm (CIA) function, that were also assigned as ^{35}Cl-containing formulas by the IPA function.

d. We have provided additional elemental analysis (C, H, N and S) of food waste sludges and sludge solids using an elemental analyzer (**Table R4**). The results indicate higher ratios of N and S to C in the sludge compared to the solids, suggesting an abundance of N- and S-containing molecules in the liquid phase (DOM). Additionally, we compared the N, S, and P content of our samples with values reported in previous studies that used standard analysis methods (**Table R5**). Our results show that the N and P content in our DOM samples is obviously higher than those reported in wastewater sludge studies. This may partly explain the slightly higher proportions of P-containing formulas (16.5% for PD; 14.0% for SD) and N-containing formulas (60.5% for PD; 62.3% for SD) (**Table R2**). Notably, a significant portion of the N and P in the DOM exists as NH_4^+/NO_3^-, and PO_4^{3-}, which is undatable by FT-ICR MS.

Table R4 Elementary analysis of food waste sludges (g%)

Elements		C	H	N	S	P	C/N_w	C/S_w
PD	Sludge	2.71	10.21	0.24	0.16	/	11.3	16.9
	Solid	46.5	6.7	4.7	0.76		9.9	61.2
SD	Sludge	0.22	10.46	0.25	0.13	/	0.9	1.7
	Solid	25.01	4.58	4.76	0.75		5.3	33.3

Notes: ① The "sludge" refers to the sludge samples that were directly used for elemental analysis; The "solid" refers to sludge samples that were centrifuged to remove the supernatant and then freeze-dried before performing the elemental analysis.

② The elemental composition of sludge and solid samples was analyzed using an elemental analyzer (Vario MACRO cube, Elementar, Germany).

③ Phosphorus (P) cannot be measured using an elemental analyzer.

Table R5 N, S and P content in the DOM of AD

Studies	N	S	P
Liu et al., 2023	NH_3-N: 218-279 mg/L	/	PO_4^{3-}P: 2.2-47 mg/L
Hu et al., 2018	26.5 mg/L	/	TP: 2.9 mg/L
Li et al., 2023	~45.3 mg/L	/	/
Yang et al., 2024	356.7-1120 mg/L	/	/
This study	**PD: ~1400 mg/L** **SD: ~1780 mg/L**	**PD: 143 mg/L** **SD: 36.5 mg/L**	**PD: ~500 mg/L** **SD: ~90 mg/L**

Note: N, S, and P here refer to total dissolved nitrogen, sulfur, and phosphorus. N and P were analyzed

using standard methods. S was specifically measured in our study using ICP-MS (NexION 5000G, PerkinElmer, USA).

e. The calibration and testing of FT-ICR MS were performed meticulously to ensure data reliability and accuracy. Sodium trifluoroacetate (NaTFA) was used as a quality control sample with the mass range of 112-2968 m/z. Additionally, internal calibration was used using specific calibrants (listed in **Table R6**). The analytical procedure include:

i. Verifying instrument performance.
ii. Cleaning the sample inlet system with methanol.
iii. Running solvent bank using methanol to check for interference.
iv. Optimizing the operating parameters using quality control sample.
v. Running sample test.
vi. Assigning formulas and analyzing data.

f. All samples were analyzed in duplicates, and only peaks consistently detected in both replicates were considered in the subsequent analysis.

Above analytical details have been included in the Methods and Supplementary Methods in the revised manuscript.

Table R6 List of (internal) calibrants used for FT-ICR MS

Neutral Molecular Formula	m/z (H-)	Neutral Molecular Formula	m/z (H-)
$C_2HF_3O_2$	112.985587	$C_{16}H_{24}O_8$	343.139841
$C_7H_{14}O_2$	129.092103	$C_{19}H_{22}O_6$	345.134363
$C_9H_{18}O_2$	157.123403	$C_{20}H_{34}O_3S$	353.215589
$C_{11}H_{22}O_2$	185.154704	$C_{23}H_{46}O_2$	353.342504
$C_{10}H_8O_5$	207.029898	$C_{20}H_{24}O_6$	359.150013
$C_{10}H_{14}O_3S$	213.059089	$C_{21}H_{36}O_3S$	367.231239
$C_{13}H_{26}O_2$	213.186004	$C_{24}H_{48}O_2$	367.358154
$C_{12}H_{14}O_4$	221.081933	$C_{21}H_{26}O_6$	373.165663
$C_{11}H_{16}O_3S$	227.074739	$C_{24}H_{18}O_3N_2$	381.124466
$C_{14}H_{28}O_2$	227.201654	$C_{22}H_{38}O_3S$	381.246889
$C_{10}H_{16}O_6$	231.087412	$C_{22}H_{28}O_6$	387.181313
$C_{13}H_{16}O_4$	235.097583	$C_{19}H_{24}O_9$	395.134756
$C_{12}H_{18}O_3S$	241.090389	$C_{23}H_{40}O_3S$	395.262539
$C_{15}H_{30}O_2$	241.217304	$C_{26}H_{52}O_2$	395.389454
$C_{14}H_{18}O_4$	249.113233	$C_{23}H_{30}O_6$	401.196963
$C_{13}H_{20}O_3S$	255.106039	$C_{16}H_{12}O_{13}$	411.020518

$C_{16}H_{32}O_2$	255.232954	$C_{17}H_{14}O_{13}$	425.036168
$C_{10}H_{21}O_5N_3$	262.140844	$C_{21}H_{26}O_{10}$	437.145321
$C_{15}H_{22}O_4$	265.144533	$C_{18}H_{16}O_{13}$	439.051818
$C_{14}H_{22}O_3S$	269.121689	$C_{30}H_{60}O_2$	451.452055
$C_{17}H_{34}O_2$	269.248604	$C_{19}H_{18}O_{13}$	453.067468
$C_{13}H_{20}O_6$	271.118712	$C_{25}H_{32}O_8$	459.202442
$C_{16}H_{24}O_4$	279.160183	$C_{20}H_{20}O_{13}$	467.083118
$C_{15}H_{24}O_3S$	283.137339	$C_{21}H_{22}O_{13}$	481.098768
$C_{18}H_{36}O_2$	283.264254	$C_{22}H_{24}O_{13}$	495.114418
$C_{17}H_{26}O_4$	293.175833	$C_{21}H_{24}O_{14}$	499.109329
$C_{16}H_{26}O_3S$	297.152989	$C_{23}H_{26}O_{13}$	509.130068
$C_{19}H_{38}O_2$	297.279904	$C_{24}H_{28}O_{13}$	523.145718
$C_{18}H_{28}O_4$	307.191483	$C_{25}H_{30}O_{13}$	537.161368
$C_{13}H_{10}O_9$	309.025206	$C_{27}H_{26}O_{12}$	541.135150
$C_{17}H_{28}O_3S$	311.168639	$C_{26}H_{32}O_{13}$	551.177018
$C_{20}H_{40}O_2$	311.295554	$C_{27}H_{34}O_{13}$	565.192668
$C_{19}H_{30}O_4$	321.207133	$C_{28}H_{36}O_{13}$	579.208318
$C_{18}H_{30}O_3S$	325.184289	$C_{29}H_{38}O_{13}$	593.223968
$C_{21}H_{42}O_2$	325.311204	$C_{28}H_{24}O_{15}$	599.104244
$C_{20}H_{30}O_4$	333.207133	$C_{26}H_{52}O_{15}$	603.323345
$C_{19}H_{32}O_3S$	339.199939	$C_{30}H_{40}O_{13}$	607.239618
$C_{22}H_{44}O_2$	339.326854	$C_{31}H_{42}O_{13}$	621.255268

2) Missing information in FTICRMS acquisition and formula assignment.

   a. In Fig. 1b and Table S1, you demonstrate that anaerobic digest sludge has most molecular formulas. This is an interesting overview, but it may be misleading here. The sheer number of molecular formulas depends on instrument parameters and acquisition settings, which may not be comparable among studies. For example, you applied a (Line 401) 54-20000 m/z scan range which surpasses most of the upper range limits of studies in know in environmental science, where 1000 m/z is often chosen as the upper limit. This alone would explain why you detect so many more formulas. To compare: We detect 30.000 formulas in oxic/anoxic incubations of suspended substrates running for one year, but we only achieved these high numbers because we measured by LC and in both ionization modes at an S/N of 4 and mass range up to 1000.

   b. You do not report the ranges of elements in your assignment. What was the minimum

and maximum number of atoms allowed for each element? Looking at tables that you provided there's a maximum of 106 C atoms (minimum, 2) and 179 H atoms (minimum, 4); The number of C atoms only allows to cover a minute fraction of your mass range (106 x 12Da = 1272 Da), even if you add H and O atoms. O goes from 1-37, N up to 9, S up to 5. Reporting these values is necessary to being able to understand your analysis.

c. How did you deal with multiple assignments if you allowed mass range to extend until 20000 Da? There must have been a few … In environmental DOM analysis, compounds with multiple heteroatoms except O are usually limited to a few N, S and P atoms, for example max. 4 N, 2 S and 1 P; improbable assignments are then ruled out based on a rule allowing only three of those heteroatoms per molecule, except for N which can have 4N. A molecule featuring the combination CxHyOzN4S2P would be ruled out. Provide evidence that wider limits are appropriate in your case, see also point 1. Explain how you dealt with multiple assignments

d. Did you apply other rules to filter out improbable formulas, such as O/C < 1.2, or 0.3 >H/C < 3, or -10 > DBE-O < 10?

Response: Thank you for your insightful observations regarding our FT-ICR MS acquisition settings and formula assignment process. We recognize that some of your questions are interconnected, and we appreciate the opportunity to provide a comprehensive response.

(1) Clarification of mass range

While our FT-ICR MS instrument detected ions over a broader range of 54-2000 m/z, we restricted the formula assignment and subsequent calculations to the 100-1000 m/z range. This range is consistent with many studies in environmental science and ensures better comparability. This focus is confirmed in several figures in our study, such as **Fig. 2**c, **Fig.3**d and **Supplementary Fig. 4**.

(2) Formula assignment

As discussed in the response to General Comments and detailed in **Table R1**, our formula assignments followed the "golden rules" (*Kind T. and Fiehn O., BMC bioinformatics, 2007, 8: 1-20*, while also imposing strict element limitations. Furthermore, we prioritized formulas with the fewest heteroatoms (N, S, P) and the lowest mass error, reducing the likelihood of improbable assignments.

(3) Reasons for the High Number of Detected Molecules

The average number of formulas for PD is ~8639 and for SD is ~9650. Owing to the significantly differences (see the PCA analysis in **Fig. 3**a) among DOM of 7 FWTPs (14 total samples), the total unique formulas is up to 46327. Several factors contribute to the relatively high number of detected molecular formulas in our study:

- **Culinary Complexity**: Chinese cuisine is exceptionally diverse, featuring thousands of ingredients and a wide array of cooking techniques. This results in food waste with a highly complex organic composition (**Fig. R3**). As can be seen from Yang et al.'s study, they detected 11,086 formulas using **simulated** food waste composed of **13 ingredients in a single laboratory reactor**. In contrast, our study analyzed actual food waste, sourced from food leftovers containing **thousands of ingredients**, (**Table R6** and **Supplementary Fig 1**). This broader scope naturally leads to a much higher molecular complexity.
- **Broad geographic distribution of FWTPs**: Food sludge samples were collected from seven FWTPs across a broad geographic range, including three in North China and four in South China (**Supplementary Fig. 1**). This wide coverage reflects the diverse cultural influences of China's 56 ethnic groups, marked by distinct culinary traditions and dietary habits across regions. The molecular composition of food waste is further shaped by regional dietary staples, such as the preference for wheat-based dishes in the north and rice-based meals in the south, as well as the complex food supply chains and waste management practices in densely populated urban areas.
- **The treatment process of food waste:** Diverse foods are often prepared using various high-temperature cooking methods, which can lead to the release of more lipid molecules into the soups and broths, much of which ends up food waste. The management of this food waste involves a series of treatment processes—chemical and physical, as well as aerobic and anaerobic bioprocesses—within long-term operational FWTPs. These complex processes inevitably contribute to an increase in molecular diversity in the resulting waste.
- **Other reasons:** The choice of SPE material can influence the types of compounds retained and detected. Our use of HyperSep™ Retain PEP cartridges, which balance retention of polar and non-polar analytes, may enhance the detection of a wider range of molecules.

(4) The Chemodiversity Comparison

A major challenge in understanding molecular complexity is determining the number of distinct molecules within a system. We compared the number of DOM molecules in various natural and engineered environmental systems to assess the chemical diversity of DOM in AD processes. Our intention is not to claim that the chemodiversity observed in our study is the highest, but rather to emphasize that the chemical composition of DOM in AD systems is extremely complex. In our study, the number of molecular formulas detected in AD systems reached approximately ~10^4, which is higher than those typically observed in natural water environments (rivers, lakes, seas, etc., usually around ~10^3).

We acknowledge that the number of molecular formulas detected can vary depending on instrumental parameters and acquisition settings. To improve the reliability and accuracy of our comparisons, we have added more relevant parameters and settings that could influence the detection of molecular formulas in **Supplementary Table 1**. We hope these efforts will enhance the reliability of our comparisons.

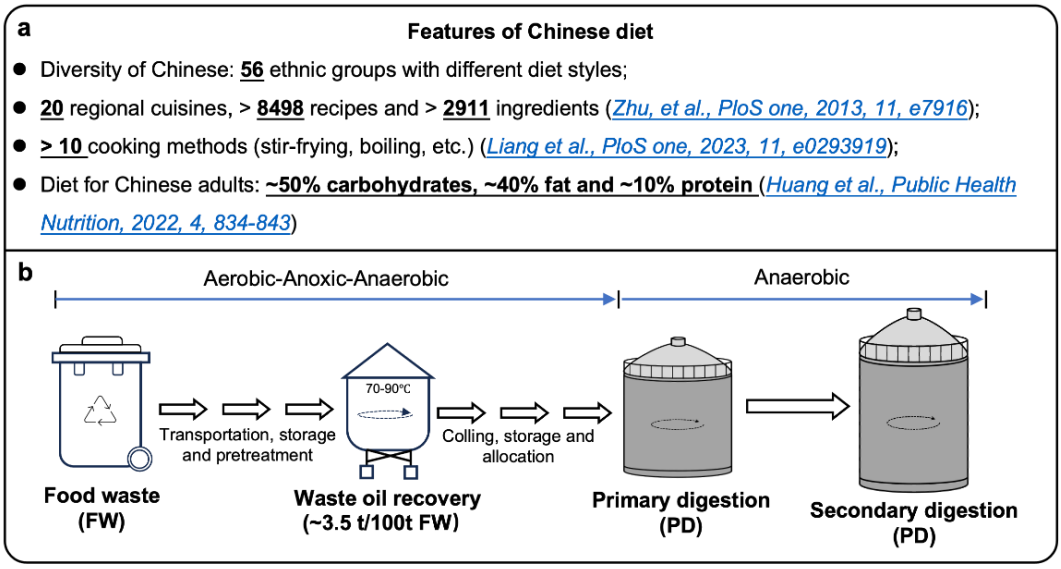

Fig. R3 | Typical features of Chinese food and food waste treatment process. (**a**) Diversity and features of Chinese cuisines. Chinese cuisines are known for their incredible diversity, shaped by regional and cultural variations across the country. High carbohydrate and fat characteristics are evident in various regional dishes. (**b**) Overview of the Food Waste Treatment Process. Food waste is collected and transported to full-scale FWTPs. After undergoing pretreatment processes such as centrifugation, filtration, and solid-phase extraction, the waste progresses to oil recovery and subsequently to anaerobic digestion. The entire treatment cycle includes aerobic, anoxic, and anaerobic stages. The varied food diets and the complex treatment processes result in exceptionally high diversity of DOM.

Table R7 Comparative analysis for food waste AD processes of Yang et al.'s study and our study.

Key processes	Yang et al., Microbiome, 2024	Our study (FWTPs)
FW composition	Simulated, containing **13** ingredients	**>1000** ingredients [a]
Food processing	NO	Various cooking processes: Stir-frying, deep-frying, steaming, Boiling, Braising, etc.
Pretreatment	NO	Food waste collecting, transportation,

		sorting, storage, etc.
Oil recovery	NO	Yes
AD process	Single-stage direct AD	PD-SD two-stage AD

[a] **Note**: The diversity of Chinese cuisines and enriched cooking techniques can be found in the following three references:① *Zhu, et al., PloS one, 2013, 11, e7916*; ② *Chen, J. & Fang, X. The Cookbook China (Sichuan People's Publishing House, 2021)* and ③ *Chen, T. A Bite of China (The Chinese Overseas Publishing House, 2018)*.

3) Line 382: Please add the gravitational force that the samples were objected to, 7000 rpm is specific to your centrifuge system. High g force stress can affect the release of DOM from sludge particles that would otherwise remain non-soluble, e.g., EPS bound to particles (Xiao et al., 2020) → add the g force value.

Response: The 7000 rpm corresponds to 5752 rcf in our centrifuge system. We have updated the centrifuge conditions in the revised manuscript accordingly. (Line 434 in revised manuscript).

4) SPE description (Supplementary Method 1): I do not see the use of step 3 ("washed using 4mL 95% methanol-deionized water"). I would wash with water if I aim to elute in methanol as you then do in step 4. I'm surprised that you still recover ~50% of DOC this way because I'd assume most of your DOC elutes already instep 3 if you use 95% MeOH. Please explain

Response: Thank you for your observation. We sincerely apologize for the oversight in the description of step 3. The solution used in this step is actually 5% methanol in deionized water, not 95% methanol in deionized water. The purpose of this wash is to remove salts and other hydrophilic impurities while ensuring that majority of the DOC remains bound to the SPE sorbent.

We have revised the number in the manuscript. (**Supplementary Method 1** in revised *Supplementary Information*)

5) Fig 3c. Please explain better how you infer the links between molecules in PD and SD. The lines connecting the both samples seem like extrapolation because you have no actual samples between both. Explain what this analysis is based on in the caption, otherwise this remains meaningless to the reader

Response: Thank you for your comment. The intent of **Fig. 3**c is to serve as a visualization method, enabling a better comparison of the core DOM compositions in the PD and SD samples. The lines connecting molecules between samples are not meant to imply direct chemical transformations or physical links; rather, they provide a framework to visually compare the molecular compositions and highlight potential trends and patterns in DOM composition and redox states across the treatment process.

We have revised the **Fig. 3** caption accordingly to clarify that its primary goal is to facilitate comparison.

6) Discussion of biochemical classes. I have problems with the way you interpret and discuss your compound class data. Looking at the formula assignments, most compounds are N-, S- and P-containing.
   a. If you frame your discussion based on lipids, proteins and lignin you should discuss how you think these biochemical classes of compounds harbor this amount of heteroatoms (e.g., lignin does not contain either of the above elements)
   b. Strictly discuss these annotations as a priori without structural proof
   c. These classes are often used in environmental sciences where N, S, and P- containing compounds are rare compared to CHO-only compounds. It makes sense to use a biochemical classification that mainly divides compounds into CHO classes (except amino acids and sugars that also contain N). These classes may however not be appropriate to interpret food waste DOM if the abundance of N, S and P is as high as you report; additional constraints using N/C, S/C and P/C may be required.
   d. Reference of parameters to differentiate molecular groups in Table S11 is missing, I believe these are compound classes from Formularity? Why is protein/ aminosugar not constrained by N/C?

Response: Thank you for your detailed feedback. We appreciate your insight and agree that further clarification is needed regarding the interpretation of our compound class data.

a. In this study, we primarily focus on the biochemical reactions involving carbon (C), hydrogen (H), oxygen (O), and nitrogen (N), as these elements play more significant roles in organic matter metabolism. In our analysis, we assume that the P and S groups remain structurally stable and do not undergo significant transformation. Therefore, the core biochemical reactions occur primarily within the C/H/O/N framework. Given the complexity of S and P metabolism and limited references in DOM studies, an in-depth discussion is currently challenging. Therefore, we believe our approach is reasonable, and we have clarified these points to enhance the understanding of our interpretation. (**Fig. 4** caption in the revised manuscript and **Supplementary Method 2** in *Supplementary Information*)

b. We recognize that the compound classifications and biochemical reaction predictions are based on molecular formulas, without direct structural confirmation. To clarify, we have updated the manuscript to clearly state the limitations of our methods (**Supplementary Method 2**, Method section line 478-480 in revised manuscript). Additionally, we have carefully discussed the results and incorporated more references to strengthen our conclusions.

c. We acknowledge that CHO-only classifications are commonly employed in environmental sciences, where N-, S-, and P-containing compounds are relatively rare. In our revised analysis, heteroatom-containing compounds declined markedly, though they remained a litter higher than those reported for sewage sludge DOM. This discrepancy likely arises from compositional differences between food waste and sewage sludge, as well as other factors such as sample preparation (e.g., SPE treatment) and DOM analysis (e.g., ionization modes) (**Table R1**). Given the relatively low proportion of heteroatoms in the DOM, our focus on the C/H/O/N framework—assuming P and S groups remain stable without major transformations—is reasonable, though not without limitations.

d. Our compound classification is based on O/C and H/C, following on the key work by Bailey et al. (*Soil Biology & Biochemistry, 2017, 107: 133-143*), which we have cited in the Methods section of our original manuscript. In our study, we employed the Formularity/Formultitude software (*Tolic et al., Ana. Chem, 2017, 89: 12659−12665*) for formula assignment, following this classification method. Furthermore, the R package ftmsRanalysis (*Bramer et al., PLoS Comput Biol, 2020, 3: e1007654*) also adopts the classification method (boundary_set = "bs2"). Another widely used classification method from Kim et al. (*Ana. Chem 2003, 75: 5336-5344*) also relies solely on O/C and H/C ratios, although its specific elemental ratio limits differ from ours. Therefore, we believe that compound classification based on O/C and H/C ratios is a reliable and well-referenced approach. We have revised **Supplementary Table 13** to include the missing references and detail the parameters used to differentiate molecular groups.

7) It remains unclear to me how the ΔH data was derived. (e.g. Table S8, Figure S14, Fig. 5, Line 257; see my comments for each of those items as well). Figure 5c is obviously inspired form recent review papers, this should be referenced then accordingly (Gunina & Kuzyakov, 2022; Wang & Kuzyakov, 2023) and clearly indicated what data was plotted here in comparison to their figures. Please also provide info on how you derived this data for each of the compound classes – independent measurements of ΔH or using Kuzyakov et al's relation ΔH = 108*NOSC-454 (?). How does this help to understand how PD and SD processes differ and act in concert, it just shows what Kuzyakov et al have already shown, that different compounds differ in NOSC and that this relates linearly to ΔH (and analogous to what has been shown earlier by La Rowe and Van Capellen (LaRowe & Van Cappellen, 2011)). Is your aim to provide proof of this relationship? Maybe explain in more detail in the

text

Response: Thank you for your insightful comments. We appreciate the opportunity to clarify the derivation of the Δ*H* data and to elaborate on how our analysis enhances the understanding of PD and SD processes. We apologize for the lack of clarity regarding the calculation of the standard molar enthalpy of its complete combustion (Δ*H*). Δ*H* values are calculated based on the Patel-Erickson equation (*Patel, S.A. and Erickson, L.E., Biotechnology and Bioengineering, 1981, 9: 2051-2067*, *Battley, E. H. Thermochimica Acta, 1998, 1-2: 17-37*), where the standard molar enthalpy of combustion of organic matter is proportional to the number of electrons that it transfers to oxygen during combustion. This proportionality is expressed through the equation Δ*H* = -114.14 *(4C+H-2O+5P+6S) (*Marko Popovic, Heliyon 2019, 5, e01905*). Therefore, evaluating the Δ*H* of DOM compounds, along with their NOSC values, can help predict the compounds' biodegradability and energy yield during microbial processes. This information is essential in understanding the efficiency of AD processes and how different compounds within DOM pools contribute to overall energy flow and biotransformation potential in waste treatment systems.

We appreciate your observation that **Fig. 5**c (Fig. 5b in the revised manuscript) is inspired by recent review papers. We have revised the manuscript to include appropriate references to Gunina & Kuzyakov (2022) and Wang & Kuzyakov (2023), ensuring we acknowledge their contributions. In **Fig. 5**c (**Fig. 5**b in the revised manuscript), we plotted the calculated Δ*H* values against the NOSC for each DOM compound class identified in our study. This visualization allows us to compare the energy content and redox states of different compound classes, highlighting how they may influence microbial degradation processes. We have also clarified in the figure caption and the main text what data was plotted and how it relates to the figures presented in the referenced papers, facilitating a clearer comparison for the reader.

Our aim is to apply relationship between NOSC and Δ*H* to better understand the interplay between PD and SD processes in AD systems. By analyzing the Δ*H* and NOSC of DOM compounds, we can infer their relative biodegradability and the energy yield potential during microbial degradation. Compounds with lower NOSC values are generally more reduced and can release more energy upon oxidation, making them more susceptible to microbial degradation during PD. In contrast, compounds with higher NOSC values are more oxidized and may persist in the SD stage, where different microbial communities and pathways facilitate their breakdown. By examining the distribution of Δ*H* and NOSC across different compound classes, we can elucidate how PD and SD processes act in concert to transform the DOM pool. Our study extends the application of the NOSC-Δ*H* relationship by integrating it with the specific context of AD processes, providing insights into the efficiency of energy recovery and the transformation pathways of organic matter in waste treatment systems.

We have expanded the discussion in the manuscript to explain in more detail how the analysis of Δ*H* and NOSC contributes to understanding the differences and synergistic actions of PD and SD processes (Line 307-318). Furthermore, we reference the work of LaRowe & Van Cappellen (2011) (Ref. 45) to acknowledge their contributions to understanding the thermodynamics of organic matter degradation in our revised manuscript.

8) Shifts in relative abundance of molecular classes (CHO, etc.) and groups (amino acid-like, etc.) are discussed at length. I miss inclusion of changes in DOC concentration, i.e., is the SD molecular composition seemingly more lipidic BUT represents a much smaller DOM pool compared to PD? Fig 1a reports a decrease of 79 to 22 g/L TS, but no change in DOC concentration is reported. Neither are volumes of reference to estimate whether the total load in solids or DOC has decreased significantly during the SD process.

Response: Thank you for your valuable observation. While we did not directly measure DOC concentrations, we assessed the soluble COD (SCOD of DOM), which serves as an alternative evaluation of DOM concentration. As reported in **Fig. 1**a and **Table R8**, the SCOD decreased significantly from PD to SD, indicating a reduction in the concentration of DOM. Additionally, we observed a notable decrease in the total lipid content, which dropped by 62.3% between PD and SD (from 9.91 ± 2.02 g/L to 3.74 ± 0.67 g/L). These findings confirm that the DOM pool not only changes in composition but also reduces in overall concentration during the SD process.

Despite the significant decrease in SCOD, the intensity-weighted DOM detected by FT-ICR MS did not decrease as drastically. This discrepancy can be attributed to the following factors:

- **Higher ratios of undetectable DOM in PD**: The PD phase contained significantly higher levels of easily bioavailable DOM components, such as VFAs, lactate, and ethanol (**Table R7**). These compounds often have masses below 100 m/z, falling outside the detection range of FT-ICR MS. Moreover, these molecules are highly hydrophilic and may be lost during the SPE process used for sample preparation prior to FT-ICR MS analysis.
- **Increased recalcitrant DOM in SD**: The PD stage primarily involves the degradation of easily degradable molecules like starch, resulting in minimal production of recalcitrant DOM. In contrast, the SD stage sees a significant increase in recalcitrant DOM production, driven by the higher pH (7.77 in SD vs. 3.87 in PD), which facilitates the solubilization of larger, more complex molecules. Consequently, while SCOD and TCOD decrease from PD to SD, the remaining DOM pool consists of larger, more recalcitrant molecules, which are more readily detected by FT-ICR MS.

We have clarified this distinction in the manuscript and revised the discussion to better explain the changes in the DOM pool from PD to SD, referencing the decrease in SCOD and total lipid content, along with the composition shift towards recalcitrant molecules (Lines 338-

342).

Table R8 Key parameters of sludge samples in FWTPs

Parameters	PD (g/L)	SD (g/L)
Soluble COD (SCOD)	64.56 ± 11.52	5.81± 5.48
VFAs (C2-C8)	4.81 ± 1.69	1.05 ± 0.25
Lactate	13.62 ± 7.63	0.49 ± 0.89
Ethanol	2.98 ± 1.89	0
Lipids	9.91 ± 2.02	3.74 ± 0.67

9) Line 209: Refer to section Supplementary Methods 2 here. Provide an example of how you detect a specific reaction in the MS data, pick out a pair of molecular formulas and how you deal with their intensity information during the index calculation, and how you derive the index for a whole sample; the data you provide in the Supplement and on github is not transparent enough to understand it. This is not described in detail enough. I like the approach however.

Response: Thank you for your constructive feedback and your appreciation of our approach. We recognize that our original explanation in *Supplemental Information* and on GitHub may not have provided enough transparency to fully understand this process. In response to your comments, we have thoroughly revised **Supplementary Method 2** and updated our GitHub to include a detailed step-by-step workflow and example, ensuring that the calculation and its application are clearer for readers.

10) Line 329 vs. 334: Discuss the implications of this mismatch further. Why is there a dominance of reduced compounds and what is their origin? Is the DOM overall more microbial in character compared to initial stages and PD when food waste character still dominates? Related to that, I think that in Fig 6a there's a solubilization arrow missing from "Cell biomass" to "Effluent DOM". Such a pathway would explain why your DOM becomes more reduced because the decomposer biomass becomes ever more dominant during the PD -> SD process.

Response: Thank you for your insightful observation and for highlighting the need to further discuss the implications of the dominance of reduced compounds in our samples. The increased abundance of reduced compounds in the SD phase can be attributed to several factors related to both the nature of the substrates and the microbial processes occurring during anaerobic fermentation:

- **Limited oxidation under anaerobic conditions**: In stand-alone digesters treating food waste, the anaerobic conditions limit the extent of oxidation of organic molecules. This limitation results in the accumulation of reduced compounds, as oxidative pathways are

less active. In contrast, digesters at water resource recovery facilities, such as those treating excess sludges from processes like A2O (Anaerobic-Anoxic-Oxic), handle materials that have already undergone partial oxidation during the wastewater treatment process. This difference in the treatment conditions results in a greater proportion of reduced molecules in FWTPs, as these compounds are not sufficiently oxidized during the AD process. The reduced state of the molecules in the effluent DOM in FWTPs is therefore expected.

- **Persistence of recalcitrant compounds**: Lignin and lipids are the two most abundant components in the effluent DOM. Lignin is primarily degraded through oxidative processes, such as dioxygenase-catalyzed ring-opening, while lipids undergo degradation through β -oxidation. These degradation pathways are generally less thermodynamically favorable compared to the breakdown of carbohydrates and other easily degradable substrates. As a result, lipids and lignin under reduced states tend to persist in the organic-rich environment of the AD process.
- **Shift in DOM source from food waste to microbial biomass**: We also agree that the solubilization of microbial cell biomass is a significant contributor to the effluent DOM pool, particularly in the later stages of the digestion process. The shift toward more reduced DOM in secondary digestion (SD) is further influenced by the decomposition of microbial biomass, which is more resistant to degradation than more bioavailable food waste. As a result, it tends to persist as recalcitrant DOM in the final effluent. We have added a solubilization arrow missing from "Cell biomass" to "Effluent DOM" in **Fig. 6**a and included above discussion in **Supplementary Note 3** of revised *Supplementary Information*.

Technical corrections

Abstract

Line 17 processes

Response: Thank you for your suggestions. We have correct "process" to "processes".

Line 19 "its biochemical processes" of the underlying biochemical processes

Response: We have revised "its biochemical processes" to "the underlying biochemical processes".

Introduction

Line 45 mention that biosolids need to be transformed into DOM; and that this is a most critical step (Xiao et al., 2020)

Response: Thank you for your valuable comments. We have clarified the statement regarding the transformation of biosolids into DOM as follows: "*This transformation primarily involves converting biosolids into dissolved organic matter (DOM), followed by further*

conversion for pollutant removal and resource recovery (Xiao et al., Water Research, 2020, 187, 116441)".

Line 48 poses

Response: The "pose" has been corrected for "poses".

Line 53 mention that decomposition and stabilization processes are key in environmental studies. This is not so different from the questions you're interested in

Response: We have added the following statement: "*The decomposition and stabilization of organic matter are key in environmental studies, especially in engineered systems that demand greater insights into DOM transformation processes rather than just molecular composition analysis*". (Line 56-59)

Line 59 "due" written wrong

Response: Corrected.

Results

Line 80 "fractions" rather than "parts"

Response: We have changed "parts" to "fractions".

Line 82 add "further" in front of "substrate degradation"

Response: Added as suggested.

Line 83 specify what "volatile solids" are, this sounds confusing to a general reader

Response: Here volatile solids (VS) refer to the portion of total solids in a sample that can be vaporized or burned off when heated to a high temperature, typically around 550 °C. Volatile solids are primarily composed of organic matter, such as proteins, carbohydrates, fats, and other organic compounds, which are combustible. VS are used as an indicator of organic content in samples such as wastewater sludge, food waste, or other biosolids. We have included this explanation in *Supplementary Information*.

Line 122 instead of "reduction", use "decrease"

Response: Changed, as suggested.

Line 123 Lignin contains no N, be more specific

Response: In our revised manuscript, we have removed this statement due to changes in the analysis data. However, in our compound classification, lignin contains N (**Fig. 3**), consistent with findings from other studies (e.g., Yang et al., Microbiome, 2024; *Hu et al., Nature Communications, 2022*). This can be attributed to two factors: first, some lignin derivatives do contain nitrogen; second, current compound classification methods are not entirely precise and require further refinement to achieve greater accuracy.

Line 125 Be more specific, which lipids contain N and S

Response: Lipids containing N: phospholipids, sphingolipids, glycolipids, and ceramides. Lipids containing S: sulfolipids and sphingolipids. However, we have removed this statement

due to data changes in our revised analysis.

Line 128 Instead of “rise of” write “increase in”

Response: We have changed “rise of” to “increase in”.

Line 129 Delete “the intensity-weight of”

Response: Deleted, as requested

Line 130 Delete “with the ratios”

Response: Deleted, thanks.

Line 138 Instead of “a growth” write “a relative increase”

Response: We have changed “a growth” to “a relative increase”.

Line 140, 145 What are “ring-rich double bonds”? You mean aromatic rings

Response: We have changed “ring-rich double bonds” to “aromatic rings” for clarity.

Line 141 “Largely determined” “affected”

Response: “Largely determined” has been changed to “affected”.

Line 144 delete “the” before “previous reports”

Response: The word “the” has been deleted

Line 145 “Less unsaturated” → “more saturated”

Response: We have changed “less unsaturated” to “more saturated”.

Line 147 delete “more” in front of “resistant structures”

Response: The word “more” has been deleted.

Line 148 “or” instead of “and” in front of “ammonia-induced”; “knowledge” instead of “knowledges”

Response: We have changed “and” to “or” and “knowledges” to “knowledge”.

Line 149ff delete “may”; provide → provides; the whole sentence reads a little general. Maybe you can be more specific. Which molecules seem to be the most problematic in your view, looking at the data you presented so far?

Response: The extra word “may” has been removed and “provide” has been corrected to “provides”. We agree that the statement could be more specific. We have revised the text to reflect this specific observation as follows:

“*In summary, these findings highlight how the structural characteristics of lipids, lignin, and proteins influence their degradation and persistence in the AD process. This deeper understanding is crucial for developing more effective strategies for the targeted removal and recycling of recalcitrant DOM compounds.*” (Line 170-174)

Fig 2 caption Line 157, presents → represents

Response: We have corrected “presents” to “represents”.

Line 180 Instead of “echoed”, write “also visible”

Response: We have changed “echoed” to “also visible”.

Line 181 What does this tell you? That there is a simultaneous buildup and decay? Maybe mention that then

Response: This indicates that the breakdown of larger molecules into smaller ones occurs within the same compound classes, as observed in key groups like lignin, proteins, and lipids (Supplementary Fig. 13). We have included the above discussion in the revised manuscript. *"In the SD stage, the distribution of DOM at larger molecular sizes (450–600 m/z) showed a significant decrease (Fig. 3d). This pattern was also visible in the distribution of homologous compounds (Fig. 3e), suggesting that the breakdown of larger molecules into smaller ones occurs within the same compound classes, such as lignin, proteins, and lipids (Supplementary Fig. 13)."* (Line 207-211 in revised manuscript)

Line 198ff Rewrite this sentence, it sounds clumsy (too many commas)

Response: We have rewritten the sentence as follows:

"*We plotted all the assigned formulas in the van Krevelen diagram, categorizing them into (i) formulas unique to PD samples, (ii) formulas unique to SD samples, and (iii) formulas common to both PD and SD samples. This allows us to visualize the distribution of DOM molecules based on H/C and O/C ratios, as well as the pathways of key chemical reactions (**Fig. 4**a).*" (Line 233-237 in our manuscript)

Line 206 behind "… the digestion process", add ", in line with anaerobic conditions."

Response: We have added "in line with the fact that reduction reactions dominate under anaerobic conditions" after "the digestion process" for clarity.

Line 209 "introduce" instead of "introduced"; behind " … (DAI)", change sentence to ", together allowing to quantify the prevalence of each reaction …"

Response: "Introduced" has been corrected to "introduce" and " ··· (DAI)" has been changed to ", together allowing to quantify the prevalence of each reaction ···".

Line 229 "one nitrogen atom" instead of "a nitrogen atom"

Response: We have changed "one nitrogen atom" instead of "a nitrogen atom".

Line 230 I personally cannot see what you mean when you write " … an increase in CHO compounds (Fig. 2a …" be more specific with regard to the information shown

Response: CHON compounds can undergo oxidative deamination, converting them into CHO compounds. As shown in **Fig. 2**a, CHON compounds significantly decrease from PD to SD, while CHO compounds show a slight, but not significant, increase. This change may be partially attributed to oxidative deamination, which removes a nitrogen atom and adds an oxygen atom for each reaction. We have included this description in the revised manuscript as follows:

"*Oxidative deamination reactions are notably prominent in protein-like compounds (**Fig. 4**f and **Supplementary Table 7**), leading to shifts in DOM composition toward higher oxygen*

*and lower hydrogen and nitrogen contents. For example, activated CHON compounds—particularly those containing only one nitrogen atom—undergo oxidative deamination, resulting in an increase in CHO compounds (**Fig. 2**a).*" (Line 267-271)

Line 235 Add references for the respective biochemical reactions if you infer they are plausible pathways to explain your data

Response: The five key biochemical reactions are referenced in the textbook by Zhu and Xu (*Biochemistry [M], 4th Edition. Beijing: Higher Education Press, 2017*).

Line 239 Add reference if you state a fact ("typically")

Response: We have added the reference (*Zhu and Xu, Biochemistry [M], 4th Edition. Beijing: Higher Education Press, 2017*) regarding the dehydration of proteins in the revised manuscript.

Line 242 delete "the" ("to DOM's chemical diversity")

Response: Deleted.

Fig 4a The oxidation/ reduction line should be labelled "Oxygenation – Deoxygenation"

Response: We agree that "oxygenation/deoxygenation" is a more precise and rigorous term than "oxidation/reduction." As a result, we have updated our text to reflect this change, even though the original study we referenced used the latter term (*Anal. Chem 2003, 75: 5336-5344*).

Fig 4e I think there is 3H2 produced from the reaction producing an aromatic ring (not just one H2)

Response: We agree with your observation and have revised our figure to reflect that $3H_2$ molecules are produced from the reaction that forms an aromatic ring, rather than just one. Thank you for bringing this to our attention and we have revised the figure.

Fig 4 caption Specify which molecules were used for this analysis. Only the core molecules? Or all of them? Refer to Supplementary Methods 2 for further detail.

Response: In the revised caption for **Fig. 4**, we have clarified that all molecules were used for this analysis. Additional details are provided in **Supplementary Method 2**.

Line 255 add "required" behind "energy input" Line 256, 261 KJ → kJ

Response: This word "required" has been added and "KJ" has been changed to "kJ".

Line 260 delete "units"

Response: This extra word has been removed.

Fig 5a Where do the ΔH data come from? Was this derived from the NOSC data itself or is it an independent estimate? You need to add these details to the caption, otherwise this is meaningless

Response: The ΔH data of Fig. 5a are from the calculation based on FT-ICR MS detected formulas in our study. The details regarding ΔH and its calculation have been addressed in

responses to General Comments on **Fig. 5** and Specific Comment 7. We have also added additional details to the **Fig. 5** caption for improved clarity.

Fig 5b These distributions do not really make the change visible. The distributions remain in the same spot, only lipids are more reduced. This is already obvious from Fig 5a

Response: We agree with you and remove the density plot of ΔH-NOSC that focuses on the most abundant compound class (lipids), which is already obvious from Fig. 5a.

Fig 5 caption "presented" → "present"

Response: We have corrected "presented" to "present".

Discussion

Line 288 "the framework" →"this"

Response: Changed.

Line 289 delete "have systematically"

Response: Deleted.

Line 292 "solids and insoluble liquids" → "molecules"

Response: We use the phrase "solids and insoluble liquids" to specifically highlight the transformation process from solids to liquids through biological solubilization. This terminology is more precise than simply using the term "molecules".

Line 292 what do you mean by "inter-transformation"? reword

Response: In this context, "inter-transformation" refers to the conversion of one form of DOM into another. We have revised "inter-transformation" to "transformation between different forms of DOM" for clarity.

Line 293 "based-biosynthesis" →"production"

Response: We have replaced "based-biosynthesis" with "production".

Line 294f Rephrase, I don't understand what you mean to say, specify why

Response: As shown in **Fig. 1**a, and **Fig. 2**b, the DOM pool experiences a smaller reduction compared to COD and SCOD from PD to SD. This increase is likely due to the more favorable fermentation conditions (pH 7.77 in SD vs. pH 3.87 in PD), where organic solid solubilization to DOM is more pronounced than DOM breakdown into biogas. Based on a comprehensive analysis of DOM evolution, it can be inferred that the generation of DOM molecules from organic solid solubilization surpasses their loss through breakdown into biogas during the SD process (**Fig. 6**a). We have revised this statement to:

"*The DOM pool experiences a smaller reduction compared to COD and SCOD during the SD process (Fig. 1a and Fig. 2b), suggesting that the inflow of DOM molecules from organic solid solubilization outweighs their breakdown. Another possible reason is that FT-ICR MS–undetected molecules are notably higher in PD than in SD, given that smaller molecules such*

as VFAs are considerably more abundant in PD than in SD" (Line 338-342 in revise manuscript)

Line 297 "higher" →"more diverse"

Response: "Higher" has been replaced with "more diverse" to convey a more accurate description.

Line 299 "rise" →"increasing dominance"

Response: We have changed "rise" to "increasing dominance".

Line 301 amounts

Response: The "amount" has been corrected for "amounts".

Line 302f Ok, so this means the process is not efficient? Relate this to the decrease in Carbon contained in solids and DOC after the SD process by calculating the absolute amount of C coming from the PD system into the SD system, and how much C leaves the SD system as effluent DOC/ remains as solid in the SD system. In short, calculate a simple C balance. Although DOC becomes more recalcitrant I believe the process still removes large amounts of C entering; it would be good to know what fraction the effluent C represents to judge the impact of the DOM composition shift on the whole processes' efficiency.

Response: The statement "Nearly 20% of the DOM molecules in PD are also detected in the SD effluents" does not directly imply that the process is inefficient. Rather, it indicates that nearly 20% of the DOM molecules are not fully transformed. This could be due to two main factors: (1) FT-ICR MS, as an extremely high-resolution MS, can detect even trace amounts of molecules; (2) FWTPs operate in continuous mode with the rapid transfer of food waste sludge, leading to the simultaneous production and consumption of certain DOM. Supplementary elemental analysis (**Table R4**, and **R5**) and key parameters of DOM (**Table R8**) suggest the content of FT-ICR MS-detected intensity-weighed DOM did not decrease as drastically as the changes observed in actual dissolved carbon from PD to SD. The possible reasons have been discussed in response to Specific Comment 8.

Line 312 "ring-dominated double bonds" →"aromatic rings"

Response: We have changed"ring-dominated double bonds" to "aromatic rings".

Line 313 add "including" behind "oxidative processes"

Response: The word "including" has been added.

Line 335 "indiscriminate"? what do you mean, rephrase

Response: The "indiscriminate removal" here refers to a broad removal process using non-targeted approaches. Recalcitrant DOM typically remains in the wastewater from food waste AD, which is generally treated using activated sludge processes, followed by advanced treatment methods such as advanced oxidation or membrane processes. These treatments aim for the non-targeted removal of DOM molecules, achieving as low effluent COD as

possible. We have replaced "indiscriminate removal" with "non-targeted removal" to enhance clarity.

Line 345 Would it be an option to repeat the PD/ SD process? Would it be an option to add simple, easily metabolizable substrates (like sugars) to fuel co-metabolization (aka "priming")?

Response: Thank you for your insightful suggestion. Repeating the PD/SD process could potentially enhance the breakdown of complex DOM molecules that are resistant to degradation in a single cycle. By subjecting partially transformed DOM to additional rounds of solubilization and microbial digestion, we may promote further degradation and reduce recalcitrant DOM. However, it is essential to consider the practical implications of implementing multiple PD/SD cycles. Factors such as increased processing time, additional energy inputs, and operational costs need to be evaluated to determine the economic feasibility of this strategy. Moreover, the potential impacts on microbial community dynamics and system stability would require careful assessment to avoid unintended consequences.

Adding simple, easily metabolizable substrates, such as sugars, could indeed serve as a co-metabolization or "priming" strategy. This approach would stimulate microbial activity, potentially enhancing the degradation of otherwise resistant compounds by providing a readily available energy source. However, given the complexity of DOM-microbe interactions, identifying the optimal type, timing, and method for adding these co-substrates in AD systems for maximum performance remains a significant challenge.

Both strategies offer potential avenues for improving the degradation of recalcitrant DOM in AD systems. Implementing these approaches would require a deeper understanding of the complex interactions between DOM, microbial communities, and process parameters within the AD environment. Further research is needed to "uncover the black box" of these complex dynamics. These two strategies may ultimately support the full utilization of DOM molecules, optimizing the AD process.

We appreciate your valuable input and have incorporated this discussion into the manuscript to highlight possible strategies for enhancing the AD process (Line 391-393).

Methods

Line 390 Add a reference for the GC method parameters or add details

Response: We have added the reference (*Raposo et al., Journal of Chromatography A, 2015, 1413, 94-106*) for VFAs analysis using the GC method and explain in Methods (Line 440-443).

Line 393 Thermo Scientific

Response: We have changed "Thermo scientific" to "Thermo Scientific".

Line 396 I wonder why you chose a sediment DOM method while there's abundant papers out that describe methods for activated sludge, etc. especially considering the high N content

of those materials.

Response: We referenced the testing method from Chen et al.'s study (*Chen et al., Environment International, 2023, 178, 108080*) for several reasons. As all the DOM samples have undergone SPE pretreatment, we believe that the testing method for the FT-ICR MS equipment is not significantly affected by the type of samples. Additionally, the composition of food waste sludges is notably different from that of activated sludges, indicating that testing methods tailored for activated sludge may not be the most appropriate for our samples. The tests were conducted in the laboratory of Quan Shi at the State Key Laboratory of Heavy Oil Processing, China University of Petroleum. Shi's team is among the earliest and most accomplished groups in China to develop FT-ICR MS analytical methods, which lends credibility to the test methods they have developed. Chen et al.'s study provides detailed procedures for FT-ICR MS testing, making it a useful reference for our own testing protocols. To ensure rigor and transparency of our methodology, we have now supplemented our reference with the original method described by He et al. (*ACS omega, 2020, 20: 11730-11736*) , which also presents comprehensive procedures for FT-ICR MS analysis applicable to our samples.

Line 398 voltages→voltage

Response: The "voltages" has been corrected for "voltage".

Line 401 "range with" → "range and"; "resolution was" → "resolution set to"

Response: We have changed "range with" to "range and" and "resolution was" to "resolution set to".

Line 402 Please add details on the ions in the ICR tuning mix. Was this used for external calibration only? You should also internally recalibrate the data before formula assignment

Response: We have updated the details on external calibration (quality control) and internal calibration in the revised Methods section (Line 454-458, 461-463).

Line 402 "the mass weight error was kept to" → "the mass accuracy was ≤ 0.5 ppm"

Response: We revised "the mass weight error was kept to" to "Assignments adhered to a mass accuracy threshold of ≤0.5 ppm, within the mass range of 100-1000 m/z" for better clarity and fluency in the context..

Line 407 "mass measurement error less than" →"mass accuracy ≤"

Response: We have changed "mass measurement error less than" to "mass accuracy ≤".

Line 408 Which min and max numbers of atoms CHONSP were allowed?

Response: We filtered formulas using the elemental boundaries detailed in **Table R1**, prioritizing those with the lowest error and minimizing the sum of (N + S + P) for formula assignment.

Line 409 Twelve combinations. I only count eight mentioned here. Are they exclusive, i.e., a

CHOS formula cannot be counted as a CHONS formula and vice versa?

Response: Thank you for pointing out the inconsistency in the count of combinations. You are correct that there are only eight combinations mentioned, not twelve, and we apologize for the oversight. To clarify, these combinations are exclusive, meaning a CHOS formula is counted separately from a CHONS formula and cannot overlap. We have revised the text in the manuscript to reflect the correct count of eight combinations and to emphasize the exclusivity of these classifications.

Line 412 “and other, are” “and all molecules unassigned to the prior categories (“Other”), were”

Response: We have changed “and other, are” to “and all molecules unassigned to the prior categories (“Other”), were”.

Line 414 Please add a statement that these classes are a priori assignments that do not rely on any measured proof of identification (structure, e.g., by MS/MS) but are solely used to categorize the data for easier interpretation. Also, add a reference to the classification you used. I wonder how protein or amino sugars can be identified without the use of N/C as an additional constraint.

Response: Thank you for your suggestion. We have included a statement clarifying that these classes are a priori assignments that do not depend on any measured proof of identification, such as structural confirmation by MS/MS, and are used solely to categorize the data for easier interpretation. (Line 464-466 in our revised manuscript)

Our compound classification based on O/C and H/C ratios built on the key work by Bailey et al. (*Soil Biology & Biochemistry, 2017, 107: 133-143*), which we have cited in the Methods section. In our study, we employed the Formularity/Formultitude software (*Tolic et al., Ana. Chem, 2017, 89: 12659−12665*) for formula assignment, following this classification method. Furthermore, the widely used R package ftmsRanalysis (*Bramer et al., PLoS Comput Biol, 2020, 3: e1007654*) also adopts the classification method (boundary_set = "bs2"). Therefore, we believe that compound classification only based on O/C and H/C is a reliable and well-referenced approach, although it does require further refinement.

Line 415ff Add references for the indices you used. AImod, DBE. Also NOSC is missing here. How did you calculate it?

Response: References and calculations for AI_{mod}, DBE, and NOSC have been added in the revised Methods section.

Line 424 This requires much more detail, please provide this in Supplementary Method 2

Response: The **Supplementary Method 2** has been updated to include additional details regarding the biochemical reaction analysis.

Supplementary information

Supplementary method 2 Provide refs for LCFA metabolism, amino acid metabolism, and TCA metabolism; the readme info on the github page does not provide actual insight into what data is used, how you deal with intensities and detection limits, etc. Add such info and provide an example with some actually measured molecular formulas and how the index is derived. Provide info on how the index for the whole sample is derived, and if PD and SD samples are compared or not (is the index only for each mass spectrum or is it for a pair of PD and SD samples?).

Response: Thank you for your thoughtful and detailed comments. We have added references for the metabolism of key organic compounds in the revised **Supplementary Method 2**. Additionally, we have updated the README information on the GitHub page and **Supplementary Method 2** to include more comprehensive details, such as the data used, how intensities are managed, and detection limits. We have also provided examples of measured molecular formulas, along with an explanation of how the index is derived, to ensure clarity regarding our methodologies.

Fig S2 “Redox evolution” it seems more appropriate to label this “redox properties”.

Response: Agree and have changed to “Redox properties”.

Fig S3 “most abundant molecules” → “most abundant compound classes”; please add significance info as done in Fig 2b.

Response: We have changed “most abundant molecules” to “most abundant compound classes” and the significance information has been added to **Supplementary Fig. 6** (Original Supplementary Fig. 2)**.**

Fig S4 “DOM molecules” → “DOM compound classes”; “compounds” → “compound classes”

Response: We have changed “DOM molecules” to “DOM compound classes” and “compounds” to “compound classes” (**Supplementary Fig. 8** in revised Supplementary Information**).**

Fig S5 Provide reference for DBE_O. Is this DBE-O? (Herzsprung et al., 2014)

Response: The DBE_O refers to DBE minus oxygen and its calculation is based the package ftmsRanalysis (*Bramer et al., PLoS Comput Biol, 2020, 3: e1007654*). In this R package, DBE_O has advantages in minimizing program errors compared to DBE-O. However, this figure has been removed owing to data changes in our revised analysis.

Fig S7 compound classes

Response: Changed.

Fig S9 I do not understand the message of this figure. The sentence is not grammatically correct as well. Provide more info what this figure shows and explain the takeaway in the caption

Response: This figure illustrates the frequency of each sample appearing in shared formulas by 6, 5, 4, 3, and 2 samples, respectively. It highlights patterns of molecular formula occurrence among samples, emphasizing the unique characteristics. The figure supports the findings presented in **Fig. 3**a and explain why BJ_PD and QZ_SD were excluded from the core molecule community analysis. However, this figure has been removed owing to data changes in our revised analysis.

Fig S10 “bigger” → “higher”; Add density plots to the sides to adjust for the overplotting in Fig S10b

Response: The “bigger” has been removed, as the trend is not evident in our revised analysis. The density plots have been added to clearly illustrate the molecular distribution. (**Supplementary Fig. 12** in revised *Supplementary Information***)**.

Fig S11 Are those approximations or where they made by a specific function? Provide the details if so. The caption text is not helping to understand how this was used to quantify reactions

Response: This figure presents density plots derived from the data in **Fig. 4**a of the main text, showing only density levels greater than 50%. These plots allow us to identify the density centers of molecular distributions unique to PD (red) and SD (blue). The straight line between the two centers along different reaction directions (such as hydrogenation/dehydrogenation and oxidation/reduction) results in distance values that can be interpreted as the degree of the corresponding chemical reactions. However, we have removed this figure as it does not represent a precise or scientifically robust method.

Fig S12 Reference? Color the 5 reactions differently and mark them accordingly in the plot

Response: We have added some references for key biochemical reactions and updated the figure with marked metabolic reactions. (**Supplementary Fig. 18** in revised *Supplementary Information*)

Fig S13 Revise the caption, it’s grammatically incorrect. Provide info on how the ΔH and NOSC data was derived. The shift in NOSC is not obvious to me …

Response: This figure has been removed as the shift in NOSC was not clearly evident and was deemed unnecessary in the context of our revised analysis.

Fig S14 Add more details to the caption, it is too short for such a complex figure. What do you mean by energy density? Is this ΔH? Provide info on how the ΔH and NOSC data was derived and what this tells you. From what I’ve seen, NOSC has been used to derive estimates of ΔH based on actual info of both variables. I haven’t seen that you determined ΔH independently

Response: Energy density (ΔH) means standard molar combustion enthalpy per mol carbon. To avoid confusion, we have changed "energy" to ΔH for clarity throughout the figure and

caption. The concerns regarding ΔH and its calculation have been addressed in responses to General Comments on **Fig. 5** and Specific Comment 7, and we have revised the Methods section accordingly (Line 488-493). Figure captions were also refined to enhance readability and comprehension (**Supplementary Fig. 19** in revised *Supplementary Information*).

Figs S15/16/, S17/18, S19/20 Captions are always too short. I do not understand what these plots want to tell me. Why is it important, and how are they linked? How do you determine isomers? What is the takeaway in each figure?

Response: Thank you for your feedback. We have revised the captions and content for Figures S15/16, S17/18, and S19/20 to improve clarity and explain their significance. There figures provide supplementary information for **Fig. 4**a in the main text. These figures offer a detailed breakdown of the compound class composition of DOM molecules: Figures S15/16 (**Supplementary Fig 20** in our revised SI) focus on compounds unique to PD samples, Figures S17/18 (**Supplementary Fig 21** in our revised SI) on those unique to SD samples, and Figures S19/20 (**Supplementary Fig 22** in our revised SI) on compounds common to both PD and SD samples. This information enhances our understanding of the core molecules present in PD and SD, highlighting which compound classes are produced and which are degraded during the PD process. Additionally, it partially explains the global evolution of DOM metabolic flow, as illustrated in **Fig. 6**a. The results regarding isomers did not contribute meaningfully to our findings and have been removed.

Figs S23/ 24 Explain the abbreviations in the caption.

Response: All the abbreviations (VFA, COD, TDP, TDN, TS and VS) have been explained in the caption of **Supplementary Fig. 24** (original Supplementary Fig 23).

Fig S25 Interesting figure, could be appropriate for main text. Please provide more info on the distributions shown on top and on the bottom and how they relate to your data. Add references for the different stages of metabolism in a) and where you obtained NOSC data from; add a reference to metabolism of β-oxidation b) as well.

Response: Thank you for your valuable feedback. This figure aims to directly link the DOM results with the four-stage theory of the AD process. However, the FTMS-detected DOM primarily consists of larger molecules, and although we identified smaller molecules using GC and other methods, integrating these data to provide robust scientific insights is challenging. Despite this, the figure presents a fresh perspective on AD theory through the analysis of DOM data. We believe it would be more appropriate to include it in the *Supplementary Information*.

We have provided references for the four-stage metabolism of AD (*Zeikus et al., Science, 1985 4691: 1167-1173*; *Batstone et al., Water Science and technology, 2002, 10: 65-73*) and the β-oxidation metabolism (*Zhu and Xu, Biochemistry [M], 4th Edition. Beijing: Higher*

Education Press, 2017). The NOSC was calculated on the equation: NOSC=-H/C+3N/C+2O/C-5P/C+2S/C (*LaRowe & VanCappellen, Geochimica et Cosmochimica Acta, 2011, 75, 2030-2042*). These references have been added to our revised manuscript and supplementary information.

Table S1 What about mass range used for detection?

Response: We have added more parameters including mass ranges that may affect the chemodiversity comparison in revised **Supplementary Table 1**.

Table S2 Explain abbreviation VFA

Response: "VFA" is the abbreviation of volatile fatty acids, and we have added the full name in **Table S2**.

Table S3-S7 Explain what the numbers mean and how they were derived in each caption. Refer to Supplementary Methods 2.

Response: We have updated both table captions and **Supplementary Methods 2** to clearly explain what the numbers represent and how they were derived.

Table S8 Is this the base data for your plots in Fig S14? Label "energy" as ΔH if you refer to this. Where did you obtain the ΔH data from – this is missing a reference or you should explain in the methods how you determined ΔH independently.

Response: This data serves as the foundational information for Fig. 5c in the manuscript. The label "energy" has been updated to ΔH for clarity. We have also included the relevant reference (*Marko Popovic, Heliyon 2019, 5, e01905*), and the method for calculating ΔH is now detailed in the Methods section (Line 488-493).

Table S10 Add a reference for the chromatographic method or explain in Methods.

Response: We have added the reference (*Raposo et al., Journal of Chromatography A, 2015, 1413, 94-106*) for VFAs analysis and explain in Methods (Line 440-443).

Table S11 Did you come up with these ranges? If not, add the reference. To me it is unclear why protein and amino sugar do not include a N/C constraint (i.e., N/C > 0)

Response: Thank you for your comment. The ranges used are based on O/C and H/C built on the key work by Bailey et al. (*Soil Biology & Biochemistry, 2017, 107: 133-143*), and we employed the Formularity/Formultitude software (*Tolic et al., Anal. Chem, 2017, 89: 12659−12665*) for formula assignment. Additional details regarding compound classification are provided in response to Specific Comment 7, with references added to the Methods section and **Supplementary Table 13**.

References used in our responses:

[1] Batstone DJ, Keller J, Angelidaki I, Kalyuzhnyi SV, Pavlostathis SG, Rozzi A, Sanders WT, Siegrist HA, Vavilin VA. The IWA anaerobic digestion model no 1 (ADM1). *Water Science*

and technology. 2002 May 1;45(10):65-73.

[2] Battley EH. The development of direct and indirect methods for the study of the thermodynamics of microbial growth. *Thermochimica Acta*. 1998 Jan 26;309(1-2):17-37.

[3] Bramer LM, White AM, Stratton KG, Thompson AM, Claborne D, Hofmockel K, McCue LA. ftmsRanalysis: An R package for exploratory data analysis and interactive visualization of FT-MS data. *PLoS computational biology*. 2020 Mar 16;16(3):e1007654.

[4] Chen, J. & Fang, X. *The Cookbook China*. 2021, Sichuan People’s Publishing House

[5] Chen, T. *A Bite of China*. 2018, The Chinese Overseas Publishing House, 2018

[6] Chen X, Cai R, Zhuo X, Chen Q, He C, Sun J, Zhang Y, Zheng Q, Shi Q, Jiao N. Niche differentiation of microbial community shapes vertical distribution of recalcitrant dissolved organic matter in deep-sea sediments. *Environment International*. 2023 Aug 1;178:108080.

[7] Danczak RE, Chu RK, Fansler SJ, Goldman AE, Graham EB, Tfaily MM, Toyoda J, Stegen JC. Using metacommunity ecology to understand environmental metabolomes. *Nature communications*. 2020 Dec 11;11(1):6369.

[8] Fu QL, Fujii M, Ma R. Development of a Gaussian-based alignment algorithm for the ultrahigh-resolution mass spectra of dissolved organic matter. *Analytical Chemistry*. 2023 Jan 23;95(5):2796-803.

[9] Gunina A, Kuzyakov Y. From energy to (soil organic) matter. *Global change biology*. 2022 Apr;28(7):2169-82.

[10] He C, Zhang Y, Li Y, Zhuo X, Li Y, Zhang C, Shi Q. In-house standard method for molecular characterization of dissolved organic matter by FT-ICR mass spectrometry. *ACS omega*. 2020 May 14;5(20):11730-6.

[11] Herzsprung, P., Hertkorn, N., Von Tumpling, W., Harir, M., Friese, K., & Schmitt-Kopplin, P. Understanding molecular formula assignment of Fourier transform ion cyclotron resonance mass spectrometry data of natural organic matter from a chemical point of view. *Anal Bioanal Chem*, 2014, 406, 7977-7987.

[12] Hu A, Choi M, Tanentzap AJ, Liu J, Jang KS, Lennon JT, Liu Y, Soininen J, Lu X, Zhang Y, Shen J. Ecological networks of dissolved organic matter and microorganisms under global change. *Nature communications*. 2022 Jun 23;13(1):3600.

[13] Hu A, Jang KS, Tanentzap AJ, Zhao W, Lennon JT, Liu J, Li M, Stegen J, Choi M, Lu Y, Feng X. Thermal responses of dissolved organic matter under global change. *Nature Communications*. 2024 Jan 17;15(1):576.

[14] Hu, H., Liao, K., Shi, Y., Ding, L., Zhang, Y., & Ren, H. Effect of Solids Retention Time on Effluent Dissolved Organic Nitrogen in the Activated Sludge Process: Studies on Bioavailability, Fluorescent Components, and Molecular Characteristics. *EnvironSci Technol*, 2018, 52, 3449-3455.

[15] Hu J, Kang L, Li Z, Feng X, Liang C, Wu Z, Zhou W, Liu X, Yang Y, Chen L. Photo-produced aromatic compounds stimulate microbial degradation of dissolved organic carbon in thermokarst lakes. *Nature Communications*. 2023 Jun 21;14(1):3681.

[16] Huang L, Wang L, Jiang H, Wang H, Wang Z, Zhang B, Ding G. Trends in dietary carbohydrates, protein and fat intake and diet quality among Chinese adults, 1991–2015: results from the China Health and Nutrition Survey. *Public Health Nutrition*. 2023 Apr;26(4):834-43.

[17] Kellerman AM, Dittmar T, Kothawala DN, Tranvik LJ. Chemodiversity of dissolved organic matter in lakes driven by climate and hydrology. *Nature communications*. 2014 May 2;5(1):3804.

[18] Kellerman AM, Kothawala DN, Dittmar T, Tranvik LJ. Persistence of dissolved organic matter in lakes related to its molecular characteristics. *Nature Geoscience*. 2015 Jun;8(6):454-7.

[19] Kim S, Kramer RW, Hatcher PG. Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the van Krevelen diagram. *Analytical chemistry*. 2003 Oct 15;75(20):5336-44.

[20] Kind T, Fiehn O. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC bioinformatics*. 2007 Dec;8:1-20.

[21] Kitson E, Kew W, Ding W, Bell NG. PyKrev: a python library for the analysis of complex mixture FT-MS data. *Journal of the American Society for Mass Spectrometry*. 2021 Apr 27;32(5):1263-7.

[22] LaRowe DE, Van Cappellen P. Degradation of natural organic matter: a thermodynamic analysis. *Geochimica et Cosmochimica Acta*. 2011 Apr 15;75(8):2030-42.

[23] Leefmann T, Frickenhaus S, Koch BP. UltraMassExplorer: a browser‐based application for the evaluation of high‐resolution mass spectrometric data. *Rapid Communications in Mass Spectrometry*. 2019 Jan 30;33(2):193-202.

[24] Liang CB, Cui B, Wang FR, Peng J, Ma JY, Xu MY, Ke J, Tian Y, Cui ZQ. Promoting healthy cooking patterns in China: Analysis of consumer clusters and the evolution of cooking pattern trends. *Plos one*. 2023 Nov 15;18(11):e0293919.

[25] Li, X., Huang, X., Zhao, C., Wang, X., Dong, B., Goonetilleke, A., & Kim, K. H. Characterizing molecular transformation of dissolved organic matter during high-solid anaerobic digestion of dewatered sludge using ESI FT-ICR MS. *Chemosphere*, 2023, 320, 138101.

[26] Liu, J., Wang, C., Hao, Z., Kondo, G., Fujii, M., Fu, Q. L., & Wei, Y. Comprehensive understanding of DOM reactivity in anaerobic fermentation of persulfate-pretreated sewage sludge via FT-ICR mass spectrometry and reactomics analysis. *Wat Res*, 2023,

229, 119488.

[27] López-Bascón MA, De Castro ML. Soxhlet extraction. In *Liquid-phase extraction* 2020 Jan 1 (pp. 327-354). Elsevier.

[28] McDonough LK, Andersen MS, Behnke MI, Rutlidge H, Oudone P, Meredith K, O'Carroll DM, Santos IR, Marjo CE, Spencer RG, McKenna AM. A new conceptual framework for the transformation of groundwater dissolved organic matter. *Nature Communications*. 2022 Apr 20;13(1):2153.

[29] Osterholz H, Niggemann J, Giebel HA, Simon M, Dittmar T. Inefficient microbial production of refractory dissolved organic matter in the ocean. *Nature communications*. 2015 Jun 18;6(1):7422.

[30] Osterholz H, Turner S, Alakangas LJ, Tullborg EL, Dittmar T, Kalinowski BE, Dopson M. Terrigenous dissolved organic matter persists in the energy-limited deep groundwaters of the Fennoscandian Shield. *Nature Communications*. 2022 Aug 17;13(1):4837.

[31] Patel SA, Erickson LE. Estimation of heats of combustion of biomass from elemental analysis using available electron concepts. *Biotechnology and Bioengineering*. 1981 Sep;23(9):2051-67.

[32] Popovic M. Thermodynamic properties of microorganisms: determination and analysis of enthalpy, entropy, and Gibbs free energy of biomass, cells and colonies of 32 microorganism species. *Heliyon*. 2019 Jun 1;5(6).

[33] Raposo F, Borja R, Cacho JA, Mumme J, Mohedano ÁF, Battimelli A, Bolzonella D, Schuit AD, Noguerol-Arias J, Frigon JC, Peñuela GA. Harmonization of the quantitative determination of volatile fatty acids profile in aqueous matrix samples by direct injection using gas chromatography and high-performance liquid chromatography techniques: multi-laboratory validation study. *Journal of Chromatography A*. 2015 Sep 25;1413:94-106.

[34] Shakeri Yekta, S., Gonsior, M., Schmitt-Kopplin, P., & Svensson, B. H. Characterization of dissolved organic matter in full scale continuous stirred tank biogas reactors using ultrahigh resolution mass spectrometry: A qualitative overview. *Environ Sci Technol*, 2012, (Vol. 46, pp. 12711-12719).

[35] Tolić N, Liu Y, Liyu A, Shen Y, Tfaily MM, Kujawinski EB, Longnecker K, Kuo LJ, Robinson EW, Paša-Tolić L, Hess NJ. Formularity: software for automated formula assignment of natural and other organic matter from ultrahigh-resolution mass spectra. *Analytical chemistry*. 2017 Dec 5;89(23):12659-65.

[36] Vesper H, Schmelz EM, Nikolova-Karakashian MN, Dillehay DL, Lynch DV, Merrill Jr AH. Sphingolipids in food and the emerging importance of sphingolipids to nutrition. *The Journal of Nutrition*. 1999 Jul 1;129(7):1239-50.

[37] Wang C, Kuzyakov Y. Energy use efficiency of soil microorganisms: Driven by carbon

recycling and reduction. *Global Change Biology*. 2023 Nov;29(22):6170-87.

[38] Xiao, K., Abbt-Braun, G., & Horn, H. Changes in the characteristics of dissolved organic matter during sludge treatment: A critical review. *Wat Res*, 2020, (Vol. 187, p. 116441). Elsevier Ltd.

[39] Yang X, Feng K, Wang S, Yuan MM, Peng X, He Q, Wang D, Shen W, Zhao B, Du X, Wang Y. Unveiling the deterministic dynamics of microbial meta-metabolism: a multi-omics investigation of anaerobic biodegradation. *Microbiome*. 2024 Sep 7;12(1):166.

[40] Zark M, Dittmar T. Universal molecular structures in natural dissolved organic matter. *Nature communications*. 2018 Aug 9;9(1):3178.

[41] Zeikus JG, Kerby R, Krzycki JA. Single-carbon chemistry of acetogenic and methanogenic bacteria. *Science*. 1985 Mar 8;227(4691):1167-73.

[42] Zhao, P., Du, Z., Fu, Q., Ai, J., Hu, A., Wang, D., & Zhang, W. Molecular composition and chemodiversity of dissolved organic matter in wastewater sludge via Fourier transform ion cyclotron resonance mass spectrometry: Effects of extraction methods and electrospray ionization modes. *Wat Res*, 2023, 232, 119687.

[43] Zhou P, Tian L, Graham N, Song S, Zhao R, Siddique MS, Hu Y, Cao X, Lu Y, Elimelech M, Yu W. Spatial patterns and environmental functions of dissolved organic matter in grassland soils of China. *Nature Communications*. 2024 Jul 28;15(1):6356.

[44] Zhu and Xu, *Biochemistry*, 4th Edition, 2017, Higher Education Press, Beijing.

[45] Zhu YX, Huang J, Zhang ZK, Zhang QM, Zhou T, Ahn YY. Geography and similarity of regional cuisines in China. *PloS one*. 2013 Nov 18;8(11):e79161.

Response to Reviewer 2's Comments

The manuscript has been revised according to reviewers' comments. It could be accepted for publication as it is.

Response: Thank you for your positive feedback and recommendation. We appreciate your time and effort in reviewing our manuscript.

Response to Reviewer 3's Comments

Jun Hu, Xian-Wei Liu and colleagues revised their initial manuscript. The authors did a good job, making the study much more robust and comprehensible. I think that all reviewer's points were addressed adequately. Improvements in regard of reviewer #1's criticism should be discussed more clearly in the discussion section (Figure S4) which is right now not connected well with the rest of the study (DOM data). From my side, there remain some adjustments, largely intended to improve the comprehensibility further. If those comments and suggestions are considered, the paper should be acceptable for publication.

Response: We sincerely appreciate your positive feedback and valuable input in strengthening our work. To better address Reviewer #1's criticism, we have integrated the results of Supplementary Fig. 4 into the Discussion section (Lines 378–381). Additionally, we have further refined the manuscript for clarity and provided a detailed point-by-point response outlining the specific improvements made.

Edit on Feb 27th, 2025: I was asked to also review responses to Reviewer #1, which is why I added to general comments #3, #6 and #7. I think the authors did a good job responding, they should make sure that all the given information/ reasoning is also added to the manuscript (or SI), see the general comments for more details.

Decision: Minor revisions

Response: Thank you for your additional review for carefully evaluating our responses to Reviewer #1. We have ensured that all relevant information and reasoning addressing Reviewer #1's comments are now incorporated into the manuscript and Supplementary Information. For additional details, please refer to our responses to General Comments #3, #6, and #7.

General comments and last suggestions

1) New "reactomics" approach not introduced properly despite some considerable improvements.

In lines 262 – 264 this becomes very obvious when you discuss the interpretation of your index values – what do these numbers mean? – why can they become negative? – why are they so high if you quantify individual transformations between formula pairs in one sample only? Figure 4 looks pretty but does not really help me to grasp what is going on. For example, in line

255f, you discuss that DMI is negative against expectation. Could this mean that long-chain molecules remain while short-chain ones vanish, in a relative sense? Thus, DMI would appear to indicate “production” of long-chain molecules, while in reality, long-chain molecules persist and short-chain ones vanish? Here, it really becomes obvious that the reader would benefit from a better documentation of the reaction calculations… I appreciate the effort you invested to improve this, but it is not detailed enough in Supplementary Method 2 – you essentially propose a new data analysis approach, so please (!) introduce it properly. See also comment on “data and code availability”.

Response: Thank you for your detailed feedback and for acknowledging our efforts to refine the presentation of the ‘reactomics’ approach. We recognize the need for clearer explanations of the derived index values, their interpretation, and the underlying reaction calculations. These concerns have been addressed as follows:

- **What do index numbers mean?**

  Each index represents the number of functional group losses in reactions occurring between PD and SD samples.

- **Why can they become negative?**

  A positive index indicates a net loss of functional groups, while a negative index signifies a net gain of functional groups in biochemical reactions from PD to SD samples. So negative index values mean net number of corresponding functional groups increase from PD to SD.

- **Why are index numbers so high?**

  The high index values result from several factors: (1) the broad mass range distribution of DOM molecules, (2) the large number of unique molecular formulas in each sample, and (3) the indices being calculated as the cumulative sum across seven FWTPs, each with diverse DOM composition.

- **Reason for negative DMI**

  As discussed in Fig. 4 and the Discussion section, a negative DMI indicates the net production of long-chain molecules. Based on DOM metabolic flow analysis (Fig. 6), we believe that this net production is primarily driven by the solubilization of solid waste. While the persistence of long-chain molecules from the initial DOM input may also contribute, it is likely not the dominant factor — particularly given that our study focuses on the (semi-)continuous AD process in full-scale FWTPs.

To further improve clarity, we have expanded the description of our methodology in Fig. 4 caption, Supplementary Method 2, and code documentation on GitHub, providing a more detailed breakdown of how the indices are calculated and interpreted.

2) NOSC shift should be shown more clearly in Figs 3 and 5.

In line 205 you state, “The redox states, indicated by the NOSC, shifted to a more reduced

state (Fig. 3c).” I cannot see this significant shift in the density plots in Figure 3c, it rather seems that PD is more reduced.... To improve visualization, add a plot in the style of panel 3d to visualize the NOSC distribution across PD and SD and add lines for the range of NOSC(IWA) in both sample sets (n=7 for PD and SD). It just now occurs to me that you have many distribution plots of other indices in the manuscript, but not many showing the NOSC. Fig. 3 would be a good place to do so. In Figure 5a, it becomes obvious too that the IWA of the compound groups remains similar, and that a shift to more lipids likely explains the lower NOSC. In Figure 5c it is not clear which panel shows which sample (top, PD; bottom SD?). In Figure 5, please denote clearly which data the formulas rely on. In the caption, cite Gunina, A., & Kuzyakov, Y. (2022, From energy to (soil organic) matter. Glob Change Biol, 28, 2169–2182. https://doi.org/10.1111/gcb.16071) if you decide to add info to your figure that is 1:1 adapted from their paper (“Reduced substrates: Energetically rich, but require more energy for breakdown”).

Response: Thank you for your detailed feedback and suggestions. We acknowledge the need to better visualize the NOSC shift and will implement the following improvements:

- **Enhanced NOSC Visualization in Fig. 3**: Fig. 3c illustrates the relationship between compound classes (e.g., carbohydrates, lipids, and lignin), elemental combinations (e.g., CHO, CHON, and CHONS), and NOSC. Instead of adding a separate panel like Fig. 3d, we have included Supplementary Fig. 13 to explicitly show $NOSC_{(IWA)}$ distributions across PD and SD. To further improve clarity, we have added grid lines in Fig. 3c to better highlight distribution differences. These results clearly demonstrate a shift in NOSC toward a more reduced state.
- **Clarifications in Fig. 5**: In Fig. 5c, the top panel represents PD, and the bottom panel represents SD. We have explicitly labeled these panels for clarity. Additionally, all DOM formulae from PD and SD samples were used for calculations in Fig. 5, and this has been clearly highlighted in the figure captions.
- **Reference Citation**: We have cited this reference (*Glob Change Biol*. 2022, 28, 2169–2182) in the main text, as well as in the Fig. 5 caption of our manuscript.

3) Also in line with Reviewer 1 comments: Supplementary Figure S4 is not referenced and Supplementary Note 1 is not referenced in the discussion. Include a reference for Fig S4 and Note S1; e.g. in line 377 where you discuss sludge microbiota. You should add some statements regarding the analyses shown in Fig S4 as well… Probably the figure should also be shifted to a later position in the SI (Numbers of figures will change).

Response: Thank you for your careful review. While Supplementary Note 1 was cited in line 105, Supplementary Fig. 4 was not referenced in the original manuscript. In the revised version, we have added citations for Supplementary Fig. 4 in lines 105 and 377 and integrated its

analysis for better clarity.

4) Use of terms “recalcitrance”/ “recalcitrant”. It is now widely accepted that there are no inherently recalcitrant compounds but that compounds merely “persist” under certain environmental conditions. I would advise to critically examine the occurrence of the term “recalcitrant” and “recalcitrance” and evaluate whether it could be replaced by the more general term “persistent” and “persistence”. An alternative strategy would be to use a wording in the lines of “more/ less recalcitrant” or “higher/ lower recalcitrance” to stress its relative character. Occurrences: main text (lines 26; 133; 159; 173; 201; 310; 333; 345; 356; 373; 383; 393; 410; caption of Fig 6), text in figures (Fig 6) and supplement (lines 160; 171; 253).

Response: Done as suggested.

5) Data and code availability.

I thank the authors for adding information to the SI, but I still think that the paper could be made more accessible. Keep in mind that your work will be cited more if your data and code can be used openly and easily. As you state, “A grand challenge in DOM research is to visualize the potential biochemical processes that DOM undergoes throughout its lifetime.” All the more important it would be to allow other researchers to rerun your code using your raw data…

5a) I checked the github link and found the info there to be unchanged to last time, I would suggest to add proper documentation to the code structure (the code is not commented right now). It is not obvious what the code does because there is no example input files or data to test the scripts... I would suggest to add simple test data from your study. Think of your paper not only as a story, but also as a data and code publication. 5b) I noted that the data you provide is only the data to reproduce figures, but no raw data has been made available so far. I would suggest you make the raw data publicly available alongside your publication to allow peers to use your data. This would be in line with your “global” approach to compare the AD process across sites – allow others to make use of your data to continue your work.

Response: Thank you for your feedback. We agree that improving data and code accessibility will enhance the impact of our work. To address this, we have made the following improvements:

(1) Enhanced Code Documentation:

- Added detailed comments throughout the code to improve readability and clarify key steps.
- Provided test data aligned with the DOM data used in the study to facilitate validation.
- Developed a schematic workflow illustrating the computational process, including input data, reaction classification, and index calculations.
- Clearly defined the expected input format (DOM molecular formulas from PD and SD samples) and output (calculated reaction indices such as DMI, DCI, DHI, DHGI, and DAI) to ensure reproducibility.

The updated code documentation is available at GitHub:

https://github.com/hzhengsh/DOM_biochemical_reaction

(2) Raw Data Availability:

- Provided a Raw Data File containing all formulas and processed data used in our analysis.
- Added a statement indicating that the raw mass spectrometry data are available from the corresponding author upon reasonable request.

6) Also in response to reviewer #1 comments: Include the reasoning for using ultrahigh resolution mass spectormetry (Response 2 to reveiwer #1) as a Supplemental Note and reference it in the main text (introduction). You made a good job arguing for using the chosen approach. Please transfer the text you responded to the SI for other readers.

Response: Thank you for your suggestion. We have included the reasoning for using ultrahigh-resolution mass spectrometry as Supplementary Note 4 and reference it in our revised manuscript (Line 484-486).

7) Also in response to reviewer #1 comments: Make sure that the reasons that your study stands out (response 4 to reviewer #1) are mentioned in the abstract and conclusion of your manuscript, especially pint 1) Study design.

Response: Thank you for your suggestion. The key aspects that distinguish our study, particularly the study design, have been clearly highlighted in both the Results (Line 194-195) and Discussion (Line 332-337) sections to emphasize its significance.

Specific comments and corrections

1. Line 90 “in the DOM” <- “of DOM”

Response: Thank you for your suggestion. We have revised 'in the DOM' to 'of DOM' accordingly.

2. Line 101/102 Please add that the AD system is a relatively “closed” system as opposed to natural aqueous systems which are open. Replace “static” by “stagnant”.

Response: We have clarified that the AD system is relatively 'closed' and replace 'static' with 'stagnant' as recommended.

3. Line 103 Again, indicate that this is a “closed” system.

Response: Done as suggested.

4. Line 108 “To quantify chemical diversity, we defined the number of formulae as richness and intensity-weighted formulae as abundance.” I don’t understand the second part of the sentence. How can “intensity-weighted formulae” equal “abundance”? Not clear. Description of calculation of data underlying Fig 1c and 1d is also missing in methods.

Response: Thank you for your comment. We referenced 'richness and abundance' in the context of microbial diversity evaluation. Here, molecular richness refers to the number of unique formulae within a defined molecular group, while molecular abundance represents the

weighted contribution of these formulae based on their relative intensity. Detailed description and calculation results have been included in Fig.1 caption and the Raw Data File.

5. Line 189 “The” spelled wrong

Response: Corrected.

6. Line 214f “These features are significantly more pronounced than those found in smaller, more bioavailable molecules like cellobiose, glucose, and acetate (Fig. 3f-h)” Unclear. Cellobiose, glucose and acetate are not shown in the panels.

Response: Thank you for your comment. Cellobiose, glucose, and acetate were provided as examples to compare structural features with DOM molecules, particularly in terms of C–C, C=C, and aromatic ring numbers, which are easier to calculate for these reference compounds. Therefore, they are not directly represented in Fig. 3f–h. To avoid misunderstanding, we have moved the reference to Fig. 3f–h to the previous sentence and revised this statement accordingly.

7. Line 232 Please add right at the start that you infer “theoretical” or “potential” reactions here by comparing two samples (“before” vs “after”) but that this indication is indirect… also, mention again that molecules in the “lignin region” are not necessarily lignin monomers… I cannot stress this enough – the interpretation of molecular structure from O/C vs H/C is very crude. I provide a reality check here: https://pubs.acs.org/doi/full/10.1021/acs.est.2c01332

Response: Thank you for your insightful comments. In this paragraph, we have repeatedly emphasized that we infer 'theoretical' or 'potential' reactions, as indicated by phrases such as 'visualize the potential biochemical processes' (line 234) and 'To quantify these potential biochemical processes' (lines 247–248). Additionally, the section title, 'Biochemical Reaction Prediction,' reflects the predictive nature of our approach rather than actual data calculations, ensuring a more rigorous expression. We also acknowledge the limitations of interpreting molecular structures solely based on O/C vs. H/C ratios and have highlighted these constraints in Supplementary Note 2. Additionally, we have clarified in the Supplementary Method 2 that molecules in the 'lignin region' are not necessarily lignin monomers.

8. Line 301 NOSC is not "nitrogen oxidation state". Check again throughout manuscript

Response: Corrected.

9. Line 333 The use of the word recalcitrant appears most wrong here. These compounds likely can be decomposed, they just persist during the conditions of the decomposition process…

Response: Thank you for your clarification. We agree with you and have revised the sentence as follows: “*We found that lignin-, lipid-, and protein-like compounds, dominated the DOM community in the final effluent.*”

10. Line 354 What is “mass size”? Write mass, that is the property you measured

Response: Corrected.

11. Fig S24 Chemodiversity on y axis spelled wrong

Response: Corrected.

12. Table S8 In the caption it reads “compound classes”, I think “compounds” would be more suited.

Response: Revised.

Reviewer #3 (Remarks on code availability):

I reviewed the code but did not run it (I do not use PERL).

As pointed out above, I find the code is not documented well enough. There are no comments added, and no test data provided to run the code with data from the paper. A schematic workflow of all calculations involved would help to comprehend the processing better. What is the input, what is the output?

Response: Done as suggested. For more details, see our above response on “data and code availability”.

The description of the approach in the Supplement has improved due to the revision but is still not clear enough. This can be seen in the results passage of the paper, where the authors themselves seem to be unsure how to interpret their new indices (see line 255 and following); Figure 4 looks nice but does not help to understand the approach either. It is unclear to me how these indices are calculated, why they can become negative, what the numbers actually say (can they be compared to each other?)... By all means, I think that the analysis is sound and the approach innovative, but the authors need to make a better effort to properly describe their approach so that readers understand it and hence, reuse (cite) it!.

Response: Thank you for your thoughtful feedback and for recognizing the soundness and innovation of our approach. As noted in our response to General Comment 1 on 'reactomics', we have clarified explanations in the Supplementary Information and main text to improve the interpretation and usability of Fig. 4. These improvements include clarifying the calculation methodology, explaining index values and their interpretation, enhancing the description of Fig. 4, and ensuring index comparability. These updates can be found in the Fig. 4 caption, the Methods section of the main text, Supplementary Method 2, and code documentation on GitHub.

Response to Reviewer 3's Comments

Review for 2nd revision of "Decomposing the molecular complexity and transformation of dissolved organic matter for innovative anaerobic bioprocessing", by Jun Hu, Chuan-Guo Liu, Wen-Kai Zhang, Xue-Wen Liu, Bin Dong, Zhan-Dong Wang, Yuan-Guo Xie, Zheng-Shuang Hua and Xian-Wei Liu* resubmitted (2nd time) to Nature Communications. Jun Hu, Xian-Wei Liu and colleagues revised their initial manuscript in a second round. All my concerns were addressed properly. The paper and the Github code are now much better to understand, and I thank the authors for accepting my suggestions. The reactomics perspective is very interesting and (apart from its reliance on van Krevelen "domains"...) reveals new insights. I have read the whole paper once again and stumbled across some things that may still be unclear. Please understand these minor comments as an opportunity to clarify your work better.

Suggestion: Accept after minor revisions, no further round of revision required.

Response: We thank the reviewer for the positive feedback and thoughtful suggestions. We're glad the revisions improved clarity and that the reactomics perspective was found valuable. We appreciate the additional minor comments and have addressed them in the revised manuscript to further enhance clarity.

Last comments

Line 59 "these studies" – add citations to be more specific

Response: Done as suggested.

Line 84 How is the pH set to 7.77? Is a buffer added?

Response: Thank you for the comment. In this study, we did not add any buffers to regulate pH. The reported pH of 7.77 reflects the average in situ value measured from sludge samples collected at full-scale FWTPs during routine operation. While pH control is commonly practiced to prevent acidification and ensure stable anaerobic digestion, often using soda lime as a low-cost buffer to maintain neutral to slightly alkaline conditions, this reflects general industry practice and not a method used in our study.

Data underlying Fig 1c/d Why does the abundance not add up to a 100%? Maybe clarify

Response: The DOM in anaerobic sludges exhibits diverse molecular compositions, categorized by both elemental combinations (e.g., CHO, CHON) and biochemical classes (e.g., amino sugars, carbohydrates). Fig. 1c/d shows the relative distribution within each subclassification, not the total composition—hence the abundances do not sum to 100%.

Line 110 "[…] with abundance of DOM" I think you mean to say "with abundance of DOM molecular groups"? Abundance of DOM is unclear

Response: Done as suggested.

Line 124 "chemical footprint" → I think "chemical fingerprint" is more suitable

Response: Done as suggested.

Line 131 Maybe add a statement where you think the P goes. Into microbial biomass?

Response: Thank you for the suggestion. We agree and have added a statement as follows: "*The removed phosphorus is likely assimilated into microbial biomass or precipitated as inorganic minerals, such as phosphate salts, under anaerobic conditions*." (Lines 123-124 in revised manuscript)

Line 142 "[…] inherent resistance to microbial degradation" What about production of lipids? They are constituents of cell membranes also. I think this could also contribute to rising lipid levels in DOM?

Response: We agree that cell membrane components can contribute to rising lipid levels. In the following sentence, we had already clarified that "biological hydrolysis and conversion" includes the breakdown of lysed microbial cells, which releases membrane-bound lipids and contributes to the accumulation of soluble lipid molecules in DOM.

Line 168 "richer" aromatic rings → maybe use "more" instead of "richer"

Response: Done as suggested.

Line 175 I think it may be helpful to add for clarity that these insights are based only on composition and do not take into account the potential removal of DOC (unfortunately you do not supply DOC data in Fig 1a).

Response: While we did not directly measure DOC concentrations, we assessed soluble COD (SCOD), which serves as a proxy for DOM concentration. As shown in Fig. 1a, SCOD decreased significantly from PD to SD, indicating a reduction in DOM levels. The possible reasons for this trend are discussed in the *Discussion* section (lines 299–304). We believe additional clarification in the main text is not necessary.

Fig. 2h Unclear how you calculated the C-1-DBE metric and what insight it gives (applies to Fig 3f as well). Please add info to line 498 (after DBE) in the methods

Response: Done as suggested. The revised text now reads:

"*To assess carbon–carbon single bonding, we used the C–1–DBE index, which reflects the number of C–C single bonds. Notably, the estimate may be lower than the actual value, as non-C–C double bonds (e.g., C=O, C=N) cannot be accurately distinguished in this calculation.*" (Lines 447-450)

Fig 2 caption, line 190 Unclear which "boxplots" you refer to here. Maybe they were removed in the reviewing process?

Response: Thank you for pointing this out. The reference to box plots was unnecessary and has been removed from the figure caption to avoid confusion.

Line 219 "[…] double bonds (C-C and aromatic […]" → write "C=C" instead of C-C

Response: Done as suggested.

Line 226 … however, looking at Fig 3b, the core community only makes up a small proportion based on formula number. Maybe add an intensity-based figure (pie diagrams) as well → usually, common formulas are also high in intensity, meaning that the core formulas might be

small in number but important in terms of intensity (information).

Response: Thank you for the thoughtful suggestion. The core molecular community was defined based on shared molecular formulas across the seven FWTPs. Due to substantial variability in signal intensities for these formulas among samples, intensity-based averaging was not considered scientifically meaningful. Therefore, we focused on formula occurrence without incorporating intensity.

Fig 3 caption Panel c – please give more information how the molecular groups (Carb, Lignin, …) are linked between samples and that they are also shown separately for elemental compositions (CHO, …). Maybe consider adding the median of the NOSC distributions (as a red line or so) to highlight that there is a shift to lower NOSC

Response: Thank you for the valuable suggestions. In Panel c, each molecular formula is assigned to a compound class (e.g., Carb, Lignin), an elemental composition group (e.g., CHO, CHON), and a corresponding NOSC value. The parallel plot is used to simultaneously compare these changes in these molecular characteristics between PD and SD samples. To more clearly demonstrate the shift in oxidation state, we have included Supplementary Fig. 13, which shows the intensity weighted $NOSC_{(IWA)}$ distributions for PD and SD samples, highlighting a clear trend toward more reduced compounds in SD.

Line 237 Specify again that this is not only based on core formulas (that were discussed before). Also, be more clear – formulas unique to PD means that you averaged all PD mass spectra and compared this to the average SD spectrum? More info would help to clearly separate it from the section before where you looked for common molecules between PD and SD.

Response: Thank you for the comment. To clarify that the reaction analysis is not limited to core formulas, we have added the following sentence to the figure caption:

"*All molecules detected in PD and SD were included in these reaction analyses, and quantitative data on chemical transformations across various compound classes is provided in Supplementary Tables 3–7.*" (Lines 699-701)

Line 240 "broken down" and "newly produced" – consider rewriting to "disappearing" and "newly detected".

Response: Done as suggested.

Line 250 "or" missing

Response: Added.

Line 254 At the end of the paragraph, you may want to add a note that this analysis focuses on all putative reactions between all potential formula pairs in PD and SD, and does not actually provide evidence of these reactions taking place – it is an explorative analysis (similar to this one: https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/rcm.7719)

Response: Thank you for the suggestion. We have added the following note to Supplementary Method 2:

"*It is important to note that this analysis is based on all theoretically possible reactions between molecular formula pairs detected in PD and SD, and does not imply that these reactions actually occurred. Additionally, both compound classifications and biochemical reaction predictions are based on molecular formulas, without direct structural confirmation.*" (Supplementary Method 2).

We believe this clarification makes it clear to readers that the results reflect theoretical inferences rather than experimentally confirmed reaction pathways.

Line 262 Why does decarboxylation facilitate "the interconversion among lignin-, protein- and lipid-like molecules"? You state: "Our analysis demonstrates a notable lignin-to-lipid transition with DCI (lignin→lipid) = 26688" – Be more specific here: Knowing your paper, I would interpret this statement the following way: You found decarboxylation reactions among formulas, and a high proportion of those formula couples were classified as "lignin" in PD and classified as "lipid" in SD. If so this would seem contradictory because one only removes C and O (CO2) but lipid-like molecules are defined by high H/C and low O/C, but it makes sense if one considers that the H/C ratio of the remaining molecule increases due to removal of C, and O/C decreases as well because more O is removed than C. Maybe it would make sense to explain it in detail at this stage. →I was struggling a bit to follow your argumentation in the reactomics section of the results but I think I understand what you mean. Considering that I have spend longer time with your manuscript than most other readers will, it may help to explain the result, at least for one index (e.g., decarboxylation of lignin to lipid) in easier words.

Response: Thank you for this thoughtful comment and for engaging deeply with the reactomics section. Our analysis quantifies putative biochemical reactions by identifying mass differences between molecular formula pairs, such as the loss of CO_2 in decarboxylation reactions. We agree that the term "facilitate" may be misleading in this context and have revised the text as follows:

"*Furthermore, decarboxylation, another prevalent reaction, is associated with compositional shifts among lignin-, protein- and lipid-like molecules.*" (Lines 231-232)

Specifically, decarboxylation reactions were inferred between molecular formulas classified as lignin-, protein-, and lipid-like compounds, based on changes observed from PD to SD. For example, a substantial number of formula pairs showed a transition from lignin-like compounds in PD to lipid-like compounds in SD, with mass (formula) differences consistent with the loss of CO_2. This is what we refer to with the statement: "*Our analysis demonstrates a notable lignin-to-lipid transition with DCI $_{(lignin\rightarrow lipid)}$ = 26688, suggesting a likely shift from lignin-like to lipid-like compounds.*"

Although decarboxylation involves only the removal of CO_2 (C and O atoms), this transformation alters the elemental ratios of the remaining molecule—decreasing the O/C ratio and increasing the H/C ratio—which can result in a shift in classification within van Krevelen

space. This explains how a formula initially classified as lignin-like could be reclassified as lipid-like after decarboxylation, even though the molecular structure itself is not directly resolved.

Additionally, it is important to emphasize that both compound classification and reaction inference are based solely on molecular formulas, without direct structural confirmation. For instance, molecules classified as "lignin-like" occupy the lignin region in van Krevelen space but are not necessarily lignin monomers. To aid interpretation, we have updated a more detailed, explanation in Supplementary Method 2, and additional guidance is provided in the code documentation on GitHub.

Line 278 You state that there was a "demethylation pathway [⋯] followed by decarboxylation" – maybe explain how you come to this conclusion based on your data. If I understood correctly, you calculated each index separately. This statement sounds as if you calculated them in sequence or in parallel⋯?

Response: Thank you for the comment. The statement regarding the "demethylation pathway followed by decarboxylation" is based on established theoretical pathways of lipid metabolism, as illustrated in Supplementary Fig. 19 and described in Supplementary Method 2. These pathways are drawn from the Textbook of Biochemistry (Zhu and Xu, 4th Edition, 2017). We did not calculate these transformations in sequence; rather, the indices were assessed independently, and the proposed sequence reflects known biochemical mechanisms rather than a direct computational result.

Line 308 "prediction bond" – do you mean "confidence interval" or sth. similar? Maybe prediction range sounds better

Response: Done as suggested.

Line 310 "Most compound classes showed a shift" – From looking at Figure 5a, I'm not fully convinced, the change seems marginal. Maybe you could show the difference in ΔH and NOSC instead of the absolute values to make this point more clear⋯

Response: Thank you for the insightful comment. While the absolute differences in Fig. 5a may appear small, they represent consistent shifts in average ΔH and NOSC values across hundreds to tens of thousands of molecules, which are meaningful at the compound class level. To support this, we have provided the averaged data in the Raw Data File, along with Fig. 5b and Supplementary Fig. 20, which further confirm these trends.

Line 319 change to "[⋯] in the sludges, including lignin, lipid [⋯]"

Response: Done as suggested.

Line 343 What are "insoluble liquids" – do you mean oils? Please specify or use other wording

Response: Thank you for the comment. We agree that the term "insoluble liquids" could be clearer. In this context, it refers to water-insoluble liquids, such as oils and other hydrophobic organic compounds, which are distinct from solid-phase organic matter.

Line 356 “despite the removal of large amounts of DOM” → this is important, can you add an estimate how much DOM is removed?

Response: Thank you for the comment. It is difficult to precisely quantify how much DOM was removed through biogas production and microbial growth, as this would require specific measurements of gas yield and microbial biomass, which are beyond the scope of our study. However, since the samples were collected from well-operated full-scale FWTPs with active sludge microbiomes and stable biogas production, it is reasonable to infer that a substantial portion of DOM was effectively utilized.

Line 357 “20% of the DOM molecules”; this contradicts with the fraction of core formulas that make up only 5-8% (Fig 3b). Maybe revise and be more specific.

Response: Thank you for the comment. The 5–8% shown in Fig. 3b refers to core molecules shared within SD samples, whereas the ~20% mentioned in Fig. 4a represents the proportion of shared molecules between all PD and SD samples. The datasets supporting both figures are provided in the Raw Data File.

Line 363 “matrix effects” → unclear what you mean. Matrix effects can mean everything. Why not just write “is largely governed by its structural properties”?

Response: Done as suggested.

Line 389 “Narrow product state” – unclear what you refer to here.

Response: Thank you for pointing this out. To improve clarity, we have revised the sentence to:

“As a result, the chemical composition shifts toward a more uniform redox state, with NOSC values transitioning from a broad range to values concentrated near zero.” (Lines 341-343)

Line 393 Why is acidogenesis an efficiency-limiting process? Please add more info here, it may be interesting for readers.

Response: Thank you for the thoughtful comment. Most DOM molecules fall within a relatively large mass range (200–800 m/z) and must first be broken down into smaller, more bioavailable compounds through hydrolysis and acidogenesis. Among the four key AD steps—hydrolysis, acidogenesis, acetogenesis, and methanogenesis—acidogenesis plays a crucial role in converting complex organics into volatile fatty acids. When inefficient, it can become a rate-limiting step by impeding downstream processes like acetogenesis and methanogenesis. This is supported by the data presented in Supplementary Figs. 19 and 27, as well as Supplementary Note 3.

Line 415 “In summary, our study […] decomposed the […]” → use another word than decompose. For example, you could use “elucidated” or “dissected” or “uncovered”… decomposed has a double meaning because you also deal with DOM decomposition in the literal sense.

Response: We have replaced “decomposed” with “elucidated” in the sentence.

Supplementary Method 2

I suggest to add some more info to Supplementary Method 2 (basically including your response #3 in the text). You stated in your response that "(1) the broad mass range distribution of DOM molecules, (2) the large number of unique molecular formulas in each sample, and (3) the indices being calculated as the cumulative sum across seven FWTPs, each with diverse DOM composition" are reasons for high index numbers. Please add this information to the text and explain why this is especially for reasons (1) and (2) – I assume the higher the number of formulas (and hence mass range), the more potential formula pairs are there to calculate your mass difference. What I don't understand is how unique formulas would affect this.

Response: Thank you for the thoughtful suggestion. We agree that the wording could be clarified to avoid misunderstanding. The term "unique" was not intended to imply exclusivity across samples, but rather to highlight the molecular diversity and the large number of detected formulas in each sample. To prevent confusion, we have removed "unique" and revised the text in Supplementary Method 2 to explain that (1) the broad mass range of DOM molecules and (2) the large number of molecular formulas per sample contribute to a higher number of potential formula pairs and thus higher index values. Additionally, the cumulative calculation across seven FWTPs with diverse DOM composition further amplifies the total index values. This clarification has been added to Supplementary Method 2 as suggested.

Table S1 Please add a cautionary statement to the caption like "Note that comparison of numbers of formulas ("richness") may be impeded by differences in sample number, ionization conditions, instrumental settings (including resolution power), formula assignment rules, and post-processing rules (filtering steps)."

Response: Done as suggested.

Reviewer #3 (Remarks on code availability):

Code has improved and is much more accessible now.

Response: Thank you! We're glad to hear that the improvements have made the code more accessible.