Response to Reviewer #2's Comments:

Reviewer #2 comments are in grey

Responses are in black and indented

The pygcms data for peat and plants of several tropical peatlands are very interesting and certainly a valuable addition to the existing literature. However, and unfortunately, the MS is not yet mature, in many aspects.

We thank the reviewer for the meticulous and extensive review. We will address all comments below.

1)Figures and text are far too much, it can be divided into two manuscripts, or reduced to about 50%.

Agreed, we will elaborate the chosen omissions in two comments below.

Interpretations are often poorly connected to the data. On the one hand details that are relevant to support main interpretations are not shown (or not explored), for example the discrepancies between FTIR and PyGCMS. On the other hand, many details are highlighted, but these are often not well demonstrated by own data, and therefore for me not convincing/consistent, and likely misleading to a more neutral reader.

We addressed the minor comments on the data interpretation and resolved this point of feedback.

These details cause a loss of focus. I recommend to make a choice, focus on comparing FTIR and PyGCMS of the peat samples, or focus on comparing peat and plant composition (PyGCMS), this could be two manuscripts. It is very courageous to include so many plant samples, but if you do, it needs a more critical data analysis.

I understand the addition of the **FTIR data** to enable comparison with other studies on tropical peatlands using aromaticity as a proxy, but would it be an idea to exclude the FTIR data to improve the focus? If you do include it, could you try and find support for the discrepancies between FTIR and py data (and thus interpret the data in more detail)? The interpretation of silica is now very quickly made, but can you check this with the TOC data, for example? The depth trends are the same for the peatland with large difference between the two methods, perhaps it helps to improve the presentation of the figure with the depth records, to visualize the similarities and differences better. And consult the literature on each of the peatlands to find out more about the composition of the inorganic part.

We agree with the reviewer. The FTIR vs Py-GC-MS will require more detailed analysis than the focus and length of this manuscript allows. Reviewer #1 also highlighted that more detailed analysis necessary here. Therefore, we have decided to omit the FTIR and EA data, and that will be included in a future paper that can do justice to the comparison and allow more detailed discussion of the mismatches, especially in the plant material, while maintaining the narrative and focus of this manuscript. We will take all of Reviewer #2's feedback into the other manuscript and thank the reviewer for the encouragement.

Removed the following lines: 27-28, 95:", complemented by FTIR", 211-226; 269-331; 832

Section 3.3.3 can be omitted and/or combined with other sections. l770-805 can be omitted, there is no connection to the data. l806-829 content on future work can be omitted too, and the remaining can be combined with conclusions and/or discussion.

Reviewer #1 found this section overextended as well. We omitted lines 771-806 as suggested here. Line 806-829 places our study in the current literature landscape and is essential to reach a broader audience and signposts future work. We adapted this according to Reviewer #1's suggestions.

See also below more suggestions to reduce the ms in length

We look forwards to the suggestions.

2) Can you reconsider and homogenise the term **aromatics**?

Agreed, we define aromatics on line 267 as a term describing all pyrolysates containing an aromatic ring (P, G, S, Ca and Ar), which is how it is approached in wider techniques (e.g. FTIR and NMR). However, later we refer it to describe the pyrolysate group Ar (line 344). We adjusted this line to "*Furthermore, we detect* ***non-functionalised*** *aromatic pyrolysates (Ar, n = 17) from miscellaneous aromatic macromolecules*". Same adjustments made in line 381, 387, 398. Any confusion with the FTIR-related use of this term have been omitted earlier.

Its use is very confusing, and often not correct. Aromatics is used for i) everything else then carbohydrates, but aliphatics and for example aromatics are very different,

We define aromatics as all aromatic bond containing pyrolysates in line 267. This does **not** include aliphatic and N-containing pyrolysates and the manuscript is consistent with that.

ii) as overarching term for MAH and PAH pyrolysates, and

Yes, we include both MAH and PAH pyrolysates under aromatic, as both are pyrolysis products of aromatic-rich macromolecules (e.g. lignin).

iii) as overarching term for MAH, PAH, phenol, and lignin pyrolysates (but these have a different environmental meaning). Avoid mixing the meaning of this general term 'aromatics'.

This is our approach as intended. We agree there is nuance between these groups, but we choose to focus on the collective effect of these aromatic bonds as their resistance to degradation is well described (and consistent in our dataset), unlike the effect of these individual groups. Furthermore, this grouping is the closest parallel to other bulk OM characterisation techniques, such as FTIR/NMR. Future researchers can elucidate these subtleties based on our extensive SI, while we maintain focus in this manuscript.

Best avoid using the term aromaticity for interpreting the own data.

Agreed; we mean to communicate: "proportion of aromatics", but this is not the definition in organic chemistry of this word. We replaced all mentions of "*aromaticity*" with "***proportion of aromatics***" Lines: 34, 808, 840

If referring to FTIR data (or literature that uses FTIR data), best use 'aromatic band' consistently (and provide the band e.g. 1506 cm^{-1}). Furthermore, be aware when reading the literature, how 'lignin' or 'aromatic' is measured. Could you improve the terminology, and specify and correct where needed.

Following the reviewers' advice, we omitted the FTIR data and will take this feedback into a future manuscript.

Also, I would not mix 'vegetation type' and 'chemical groups', a woody peat is of course more 'aromatic' than a moss peat. Thus if you compare northern with tropical peat, and one is moss and the other wood, why not name it moss and wood. Decomposition of wood or moss, and thus the resulting more stable fraction, is not following the same pathway.

We agree with the reviewer to not mix vegetation type and chemical groups, and we do not believe we have done so. A peat with woody vegetation or even a "woody" peat does not directly imply an increase in aromatic macromolecules because the aromatic proportion might be altered by degradation during early deposition. Especially with tropical peat, there are literature paradigms that imply some plant material is preserved preferentially (e.g. roots preferentially preserve, (Hoyos-Santillan et al., 2015)). In short, plant cover OM composition does not equal peat OM composition, which we also show through the increase in aliphatics and reflection of different plant tissues in the peat (e.g. Mpologoma shows a Papyrus leaf signal at depth, while roots at the surface). It is also hypothesized that this early degradation regime is governed by other factors, such as high temperature and humidity in the tropical sites; therefore we prefer to handle the site terminology by their tropical vs temperate assignment. Most importantly, in all figure legends we have already sorted and labeled the sites by dominant vegetation type as the reviewer prefers.

No changes have been made.

3. Pyrolysis data

- The hemicellulose assignment is not correct, pyrolysis of cellulose results in many of these products.

We did not intend to imply that cellulose does not produce other polysaccharide pyrolysis products. The pyrolysis of cellulose heavily favours production of levoglucosan above all other products, while hemicelluloses have a more spread-out distribution, including levoglucosan, levomannosan etc. We think we handled this subtlety well in Section 3.3.1 but were not as careful in Section 3.3.2. Therefore, we replace all mention of hemicellulose in 3.3.2 with polysaccharides.

Line 394: "*Polysaccharide-related compounds include several oxygen-containing heterocycles such as furans, likely originating from pentoses (e.g. xylose)* ***primarily*** *found in hemicelluloses (Moers et al., 1990)*"

Replace "*hemicellulose*" with *"****polysaccharides***" in line 438, 476, 481, and 484

Table S1- identification

- Why not order to RT within each group?

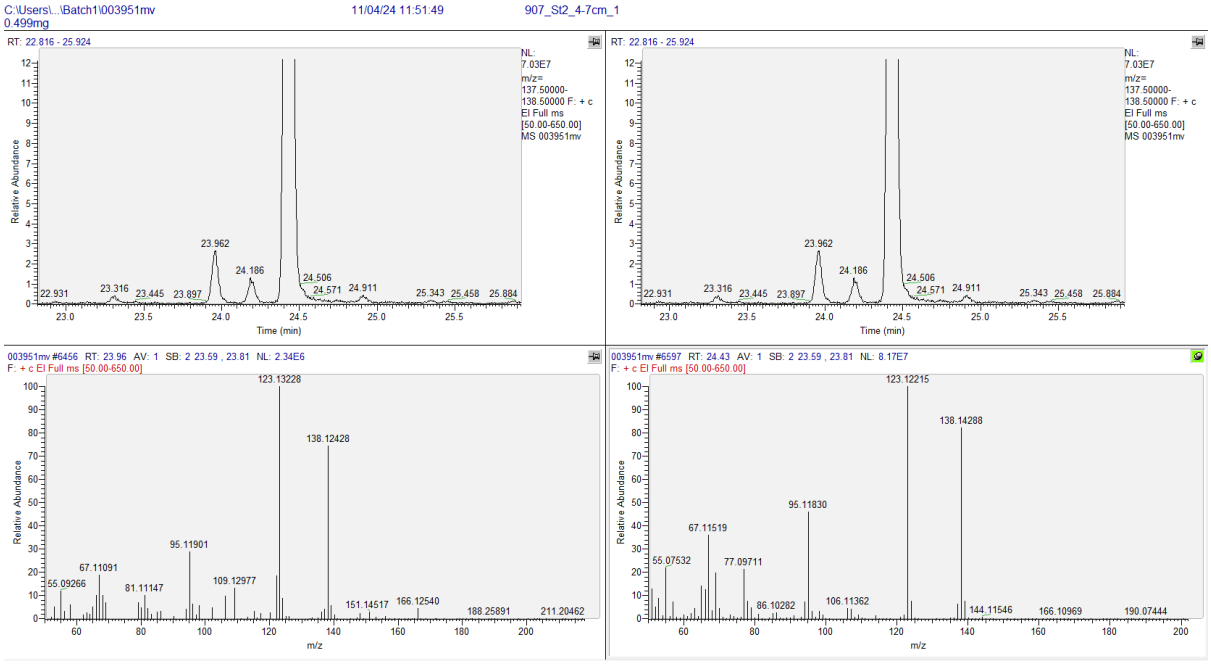
Good suggestion and implemented.

- what is m-guaiacol? Can you be consistent in the naming of products?

We utilise the NIST Chemistry Webbook to standardise the naming of our pyrolysates. We prefer grouping the pyrolysates functionally in the groups (m-guaiacol being named “guaiacol” instead of 3-methoxy-phenol). We noticed that pyrolysate naming can be confusing in most studies and have therefore provided a InChl, which can be inserted into Chemdraw directly for a structure or any search engine to refer to the NIST Chemistry Webbook. This eliminates any confusion.

- G012 not correct. the m/z suggests methylguaiacol, but not sure without seeing the spectrum

G012 is also a methylguaiacol, however it does not have a methylguaiacol-consistent name in the NIST Webbook, therefore is named: 2-Methoxy-6-methylphenol. See below the chromatogram of the ubiquitous 4-methylguaiacol G002 (peak at 24.5 min, MS on right panel), and G012 (peak at 24 min, MS on left panel). The spectra are similar, but 4-methylguaiacol features a larger 95 (loss of methoxy and methyl group), which implies that the methyl group is further from the OH, which is also consistent with the inhouse library spectra of “6-methylguaiacol” (but named according to NIST webbook to 2-Methoxy-6-methylphenol). We are confident in this identification but will introduce guaiacol-consistent naming regardless of the clash with the NIST Webbook (see two comments below).



- P16 name is not correct. But it is a phenol indeed

The name is consistent with the alternative names in the NIST Webbook, but we agree that the several different naming conventions for the phenol groups might be confusing. We renamed as follows:

P001 Phenols Phenol (P1)

P012 Phenols 2-Methylphenol

P002 Phenols 4-Methylphenol (P2)

P013 Phenols 2,3-Dimethylphenol

P014 Phenols 3,4-Dimethylphenol

P006 Phenols 4-Ethylphenol

P015 Phenols 2,4-Dimethylphenol

P003 Phenols 4-Vinylphenol (P3)

P016 Phenols 2-Ethyl-5-methylphenol

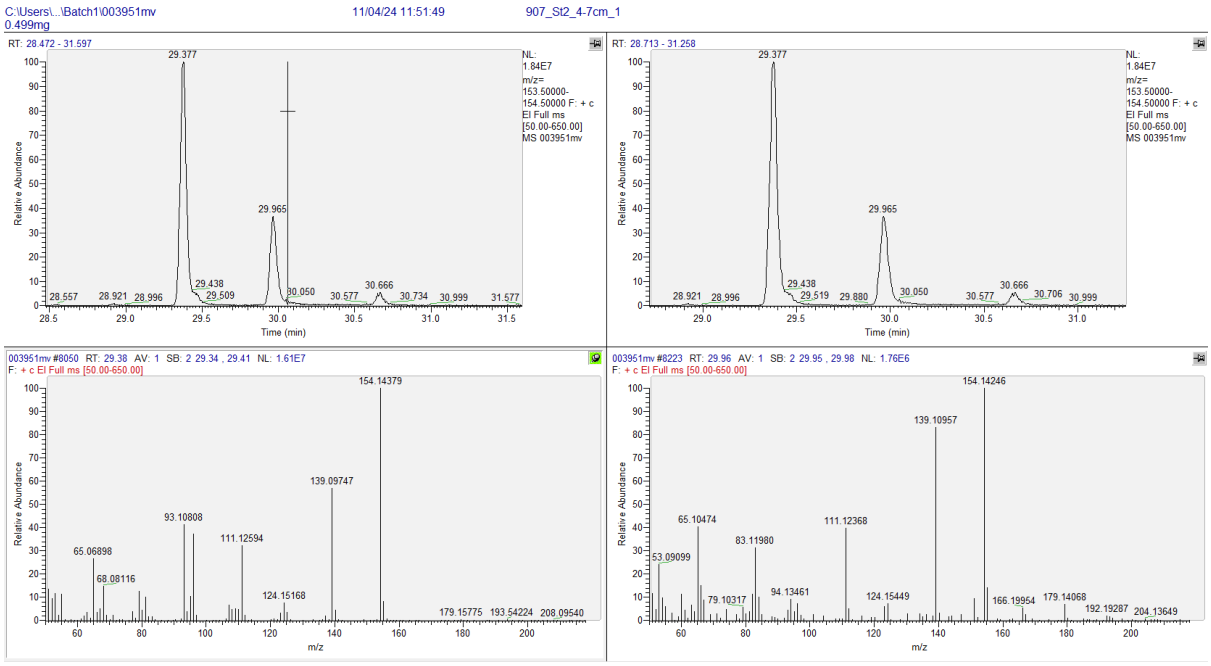
P017 Phenols p-Isopropenylphenol

P018 Phenols 3-Methoxy-5-methylphenol

P019 Phenols 3,4-Dimethoxyphenol

- P19 this product is not common in NOM. Without more information on the spectrum I canmot judge it, but I suppose this is syringol based on RT and MW.

The retention time of the component identified as P019 (30 min) is distinctly different from syringol (29.4 min, left MS) and the MS is similar to that of P019 (right MS) with a more abundant m/z 139. This implies that the OH group is further from the methoxy groups, which is also consistent with our inhouse library spectra. This inhouse library is maintained for lignocellulosic materials (i.e. NOM). We are confident in this identification.



- G012, G014 and G017. If these are guaiacols, can you use the nomenclature of guaiacols?

This clashes with our standard approach using the NIST webbook, but we agree. Renamed as follows:

G012 6-Methylguaiacol

142 G014 This is a methoxylated guaiacol, which we infer to originate from guaiacol-rich macromolecules. We
143 will stick to the NIST naming here: 3,4-Dimethoxytoluene

144 G017 4-(2-carboxyethyl)guaiacol

145 - Catechols why naming some pyrocatechols and some not?

146 That is a useful suggestion. We will adjust all naming to catechols

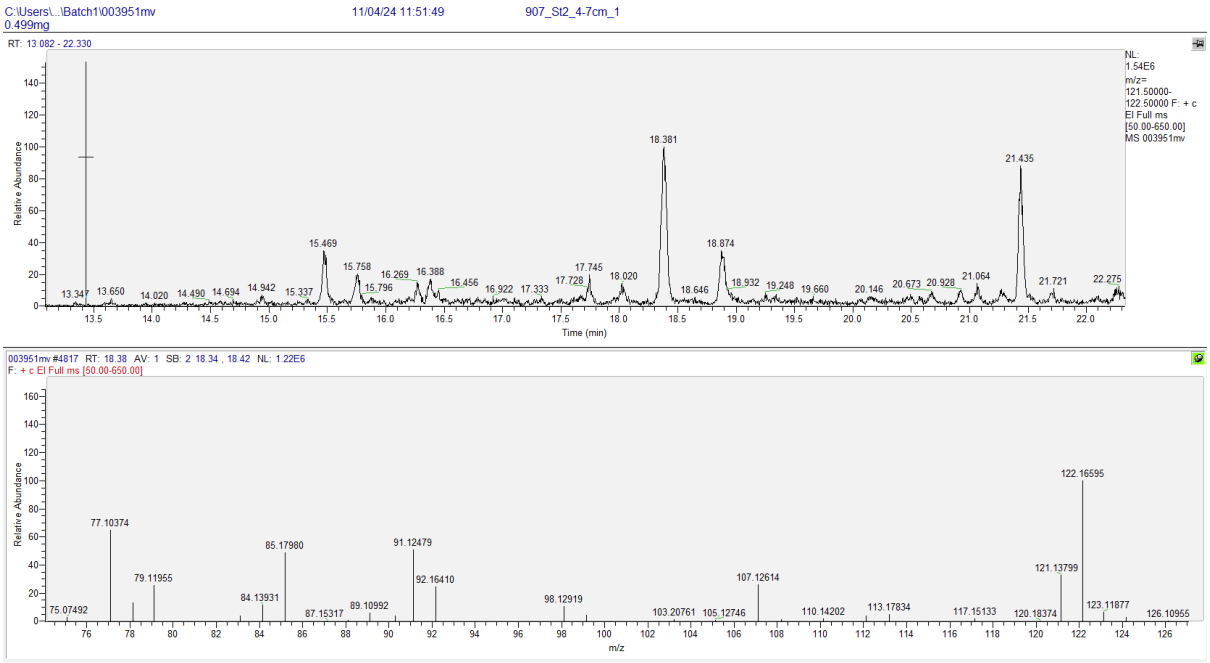
147 D002 Catechols 3-Methoxycatechol

148 D003 Catechols 3-Methylcatechol

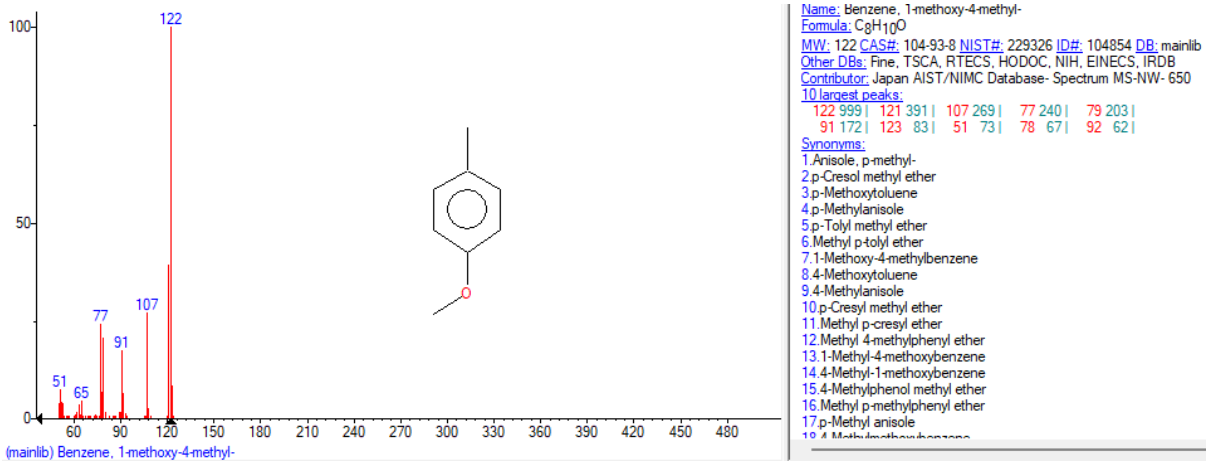
149 - B009 not correct

150 See mass spectra below. The m/z 122, 107, 91 and 77 and their relative distribution are consistent with 4-
151 Methylanisole. We are confident in this identification.

152



153



154 - H019 this is not a carbohydrate but a terpene

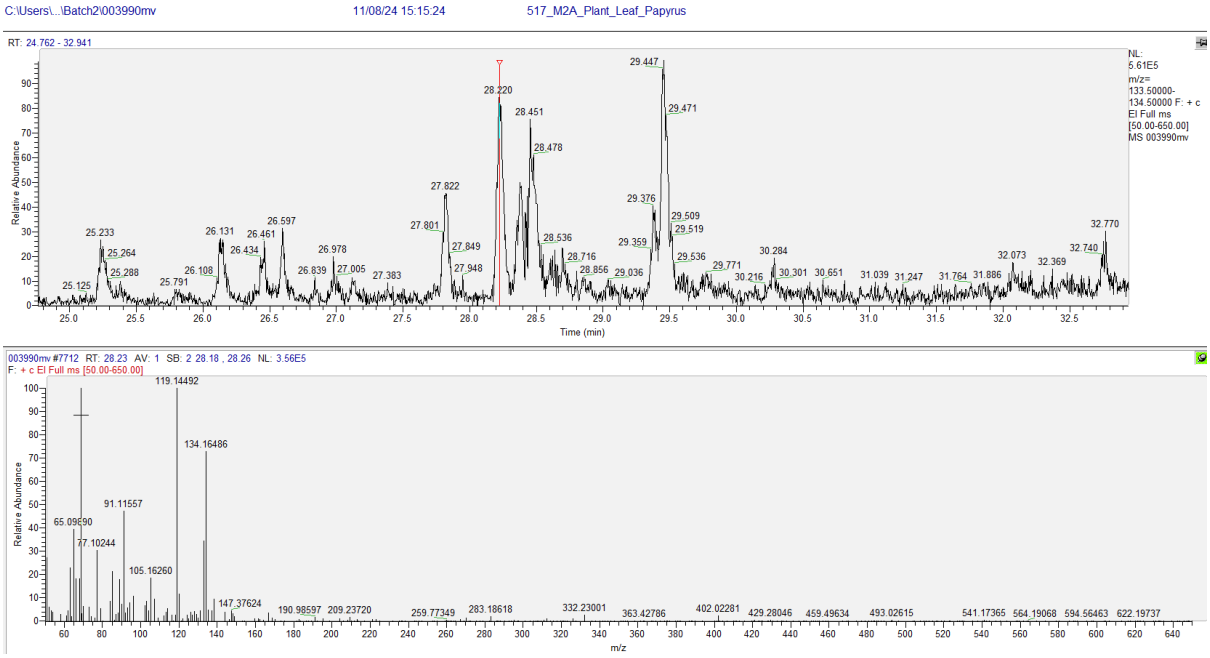
Good catch, limonene is a terpene theorised to derive from the isoprenoidal part of chlorophyll. This makes this compound out of the scope of our polysaccharide grouping. We will omit limonene from our analysis. Since limonene is a minor component of the total number (1/141) and abundance of compounds (< 0.0001), none of the results/discussions of this manuscript have changed.

Table S6

- I did not and cannot check the quantificion of the data, but I checked p-isopropenylphenol, and it seems unlikely that this products, specific for sphagnum, is abundant in the plants that were analysed, for example in trunk and root (row 1283-1287) or rhizome, leaves, wood (rows19589-19591).

p-Isopropenylphenol is not abundant in other plant material. In fact, in 49 plant samples we did not detect it. Of the plant samples we analysed, only *Sphagnum papillosum* contained it in high abundance (3%). But it is indeed present in other plant tissues. The second highest were the Papyrus leaves (see MS below, it is consistent) with an abundance of 0.1%. This is an order of magnitude lower, with other tissues being in the 0.01-0.001% range. Pyrolysis creates a myriad of products and often most products are present in different abundances such that exclusive biomarkers are rare. Especially the Papyrus tissues appear to produce p-Isopropenylphenol on pyrolysis in very minor abundance that does not invalidate the use of p-Isopropenylphenol as a *Sphagnum* indicator or our analysis. However, we do note that previous work by our group has highlighted p-Isopropenylphenol as diagnostic for Sphagnum; our findings are consistent with that but now with a few very minor exceptions, and that is worth an additional sentence in the manuscript (and will prevent similar concerns by future readers).

Added to line 378: *"... lignin and two diagnostic polyphenol pyrolysates (**e.g. 3% p-isopropenylphenol in S. papillosum**) universal among Sphagnum (Van Der Heijden et al., 1997**), which are also found in minor (< 0.01%) abundance in tissues from Papyrus as an exception**."*



- on the other hand, you mention 4-vinylphenol to be a biomarker for some peatland plants in the MS, but as you can see in your data it was present in all plant tissues and plant species. It may raise doubts on the correctness of the data. Thus, I would recommend to check the quantifications, and the interpretations related to it.

4-Vinylphenol is a common pyrolysis product of any polyphenol-containing material. We and previous literature have observed that the proportion of 4-Vinylphenol is higher in sedge material. That does not mean it is not a "biomarker" (sterols, hopanoids, carotenoids are all biomarkers even if they have varying degrees of source specificity). Nonetheless, for clarity we replaced all uses of the term "biomarker" with "indicator".

Replaced "*biomarker*" with "***indicator***" in lines: 408, 450, and 848

Table S3.

- You include the phenols into the lignin pool, this is not correct, and biasing the statistics. P lignin (or H-lignin) is p-hydroxyphenyl not phenols

Good catch. We handle this correctly everywhere else, but in Table S3, this is wrong.

We removed the column *"%Lignin"* and renamed the column *"%P-lignin"* to ***"%Phenols"***

- what are the 3 last columns? Lignin and polysaccharides differ from the columns with the same name. and what do you consider as cellulose?

These columns reflect the data used in the ternary diagrams (Line: 263-265), which would be unclear from the Table independently.

We renamed the columns to: **Ternary Cellulose (%), Ternary Polysaccharides (%), Ternary Lignin (G+S).**

4. external standards -Section 3.3.1

L557-565 I do not understand this.

Based on this listing of misunderstandings, we have clearly not written this processing step in the methods and discussion well. We rewrote and restructured 2.3.3 and 3.3.1. We will clarify specific concerns in this rewritten section. See at the end of this major comment.

Could you include the pygcms results of these standards in supplement?

Well spotted.

We included the results of the external standards in Table S6

If you have a cellulose standard, then you should have observed that this did not result in levoglucosan alone, but included many products that are now 8abelled hemicellulose.

We did indeed observe most polysaccharide pyrolysis products in the cellulose standard, but levoglucosan dominates the relative abundance as we discuss above in response to comment 3.

It is not clear how the peat signal is compared with that of these standards, is it expressed as proportion of the sam total of products?

We exclusively use the external standards for the ternary diagram workflow in order to provide a visual context for interpreting those data. The ternary diagram interprets and pools the levoglucosan as Cellulose, polysaccharides as Hemicellulose, and guaiacols and syringols as Lignin (G+S) prior to the second-order fractional abundance that is represented in the ternary diagrams. This second-order fractional abundance also compensates for the other (hemicellulose) assigned components found in the cellulose standard.

We rewrote and restructured 2.3.3 and 3.3.1. See at the end of this comment.

Furthermore, the idea of including external standards is nice, but you should be very cautious and demonstrate that you understand your analytical method, your samples, and the effects of environmental processes.

We thank the reviewers for appreciating our external standard approach, and we have clearly not written this processing step in the methods and discussion well.

We rewrote and restructured 2.3.3 and 3.3.1.

But even if the peat carbohydrates were exclusively composed of cellulose and hemicellulose and would not share products, the composition of (pyrolysis products of) hemicellulose differ among plants.

This is correct and this is why we caution that %cellulose in our peat or plants is not a true compositional percentage but only a relative comparison to this particular reference. Only in section 3.3.1 do we generalise, including grouping these pyrolysates as hemicellulose, but this is only to provide context not mathematical assignments . Future workers are able to utilize our dataset and can reinterpret the assignment as they see fit.

Also, the 'lignin cellulose hemicellulose space', should be defined better, it must be clear which products are in there. Also, the lignin standard is a wood extract. If you have the lignin composition of your plant tissues and species available, does it make any sense to use this standard?

The lignin standard has a distinct lignin pyrolysate distribution from any plant/peat, however, we group all the lignin-interpreted pyrolysates (Line 265).

We rewrote and restructured 2.3.3 and 3.3.1 so that it is clear exactly which products have been included in these figures.

And if it does, why would it help in interpreting the peat lignin composition if you have the lignin composition of the input materials?

Because pyrolysis-GC-MS is notoriously non-quantitative, the standard mixes allow contextualization of differences, i.e. a semi-quantitative link (noting caveats above) between the distribution of a given peat pyrolysate and *potential* biomolecular precursors. With the standards and their mixes, we show that the pyrolysate abundances changes are profound and significant in Section 3.3.1.

We rewrote and restructured 2.3.3 and 3.3.1.

The alkali used to prepare the standard also changes the lignin structure.

Correct, hence we pool all lignin-pyrolysates and refer to the percentages as *% akin to lignin standard*.

We rewrote and restructured 2.3.3 and 3.3.1.

And also for the aliphatics, if you have the results of all these plant species, why would a standard of Agave be used to represent cutan and suberan sources in the peat?

Extracted *A. americana* leaf is a known aliphatic-dominated (cutan-dominated) plant tissue. Analogous to discussions above about the imperfection of a cellulose or lignin standard, *A. americana* is not meant to represent all cutans but to provide an established reference for a cutan-rich plant. Presenting the data akin to *A. americana* confirms that the pyrolysate abundances changes are significant (i.e. plants have very little of this component, but some peat does).

We rewrote and restructured 2.3.3 and 3.3.1.

It is also confusing to use different groupings of the standards, it would help in clarity to be consistent with these groups in the MS.

This is a misunderstanding; the grouping and pyrolysates of the external standards are identical to all plant/peat data. This clearly needs clarification.

We rewrote and restructured 2.3.3 and 3.3.1.

Using levoglucosan to reflect cellulose seems reasonable, but it represents the major part of the 'polysaccharides'.

We define in our manuscript all pyrolysates that originate from polysaccharides other than cellulose (Table S1; line 343). Cellulose is also a polysaccharide, but we are confident that our usage is clearly explained.

In figure 5 you use the labels hemicellulose and cellulose, and also polysaccharides, which is confusing and unclear. Be consistent in labeling pyrolysis products and/or its interpretation, thus levoglucosan instead of cellulose (etc), (the other labels are also not microbial material, cutan, etc. bult aliphatics and N-compounds).

We agree with the reviewer that this is unclear, and in Figure 5 we have not yet jumped to the interpretation of hemicellulose.

We replaced all interpretation terms in this figure with the pyrolysate groups ("hemicellulose" with "polysaccharides")

But if you chose to present interpreted terms (hemicellulose), then make sure that it is clarified which products include this label. Same for aromatics, phenols, lignin, P3 (not in the caption, what this is?)

We will adjust according to the prior comment, but we did not intend confusion with included products. All are defined in Section 2.3.3.

We included all labels in the figure caption and rewrite and restructured 2.3.3.

We decided to change the term "*analytical standards*" to "***reference materials***" to better reflect its use in lines 229, 552, 561, 591, and 592

Rewritten sections: 2.3.3:

Added the following lines after line 263: *"Ternary diagrams were constructed by taking a second order fractional abundance over the complete min-max normalized data including only groups of interest, such as lignin (G+S),* ***Hemicellulose*** *(Ps), Cellulose (Ce), and/or Aliphatics (Al).* ***The ternary diagrams include reference materials in order to contextualize the percentages*** *of biomacromolecules inferred* ***from peat and plant pyrolysates. Each is associated with caveats, i.e. there is no single 'lignin' macromolecule,*** *and consequently we do not use them to calculate, for example, lignin contents of peat; instead, they* ***represent a measure of similarity (i.e. % akin to a cellulose reference). such that the normalization step compensates for pyrolysates that are found in both locust bean gum (Hemicellulose reference) and the cellulose standard (e.g. the levoglucosan found in the hemicellulose reference material is deducted from the relative abundance in the cellulose reference material).***"*

Rewritten Section 3.3.1:

Replaces Lines 557-565: "*Most carbon in peatlands originates from (dead) vegetation which on a molecular level is built primarily from three widespread biopolymers that give structure to the cell walls: cellulose, hemicellulose, and lignin (Scheller and Ulvskov, 2010). We pooled* ***and interpreted*** *pyrolysates associated with these biopolymers* ***(Lignin (G+S), Hemicellulose (Ps), and Cellulose (Ce))*** *and normalized their proportion with reference materials* ***(Section 2.3.3)****. This enables* ***interpretive comparison*** *of the proportion of these biopolymers among the peat-forming vegetation, associated peat, and downcore changes in the peat profile. The percentages arising from these assignments are not quantitative and should be considered as a percentage relative to a purified standard. For example, a 30% proportion of hemicellulose does not suggest 30% weight percentage of the biopolymer pool is made up from hemicellulose, rather that the* ***pyrolysis product distribution*** *is 30% similar to hemicellulose. For the remainder of this section, macromolecule percentages refer to the percentage in lignin/cellulose/hemicellulose space, noting these caveats."*

5) PCA.

The figures are difficult to read. Can you label the products that do not behave similar to their assigned group in PCA figures, and/or provide the loading in table format as supplement? In the text is often mentioned 'some specific phenols' or 'some specific carbohydrates', but why not name them? The thin sample samples and lables are often not visible behind the bold arrows.

We appreciate that the text in the figures can be challenging to read. We agree that the PCA loadings can have value in the supplements.

We have improved visibility of Figure 4 and included the PCA loadings in the supplements (New Table S6 (samples) and S7 (loadings). We also specified the ambiguities: Line 381-382: "*PC3 primarily reflects the behaviour of specific polysaccharides* ***(Table S1; H006, H011-014, and H017)****, phenols and non-*

*functionalised aromatic compounds as a positive loading, and specific phenols **(Table S1; P016 and P019)** and aliphatics as negative loadings (Fig. A3)*."

6, More technical suggestions (not complete)

Abstract

L25 Perhaps mention the range of max depths? This is a strength of the data.

Great idea. Added depth ranges: *"…and across peat depth profiles **(0–8 m)** using a…"*

L30 could you specify which groups, in addition to lignin, are in the aromatic pyrolysates?

This is to avoid confusion, lignin pyrolysis products are indeed aromatic according to the chemical nomenclature, but lignin and aromatic pyrolysates often have distinct environmental meaning. If it is everything else than polysaccharides, perhaps limit to mentioning a loss of the sugars?

Agreed. Adjusted to: *"…in aromatic **ring containing** pyrolysates, such as from lignin, vs polysaccharide-**associated** pyrolysates."*

L31 enrichment and decline is a bit confusing for me.

Agreed. Adjusted to: *"…follows a logarithmic **function**, especially…"*

L34 I do not see a relation between the loss of sugars and assigned aromaticity of leaf litter.

We clarified what we meant: *"The **high starting proportions of aromatic macromolecules** of the leaf litter and the consistent preferential **enrichment of aromatic macromolecules** at depth in the tropical sites can aid peatland accumulation modelling and enable more accurate predictions of peatland dynamics under future climate change."*

Introduction

L67 can you clarify what is meant with reactive?

Added on line 66: "*...which imparts a significant control on OM reactivity (**i.e. respiration rate of the OM;** Wright et al., 2011)."*

L65-80 I feel this part can be condensed considerably, since you focus on tropical peat perhaps there is no need to compare northern with tropical peat so detailed? In my opinion it can be reduced to two sentences.

We think it is important to establish the current paradigms around OM composition to bring the more neutral reader into the subtleties of tropical peatlands. We do include two temperate sites, especially for this comparison, so these paradigms are important to highlight here to refer back to in the discussion. This paragraph connects best to the work conducted and is essential.

L92-94 palm hardwood etc can be ombrotropic as well. I recommend using the same classification for the tropical and temperate sites.

Agreed. Adjusted to: *"…include a **Sphagnum/grass** bog and a **sedge-dominated** fen…"*

L96-97 so you assume that no changes in plant species composition occurs when redox conditions change? I would use the word explore also here, instead of identify. I understand that differences between the peatlands are likely much larger, but it can be mentioned.

We understand the confusion around the word "plant biopolymers". We did not intend to suggest that we study exclusively live plants in the peat profile. We will avoid the word "identify" as well as suggested.

We rewrote the sentence for clarity: "We first **characterize** major differences in the OM composition of peat-forming vegetation and explore **plant-derived** biopolymer transformations throughout the peat profile."

Methods

201-202 could you clarify the composition of the plant samples in more detail? To me it reads as if leaves and roots are mixed. I propose to extend column J in Table S2, instead of 'tissues included', provide a column for each sample to make clear these are not mixed.

Agreed. We changed line 201 to *"Plant material is **split on tissue type in the field**..."*. We also adjusted Table S2 to each row representing a pyrolyzed tissue sample as suggested by the reviewer.

L222 that depends on the pH and is not valid for peatlands in general, in minerotrophic peat there can be lateral input. Is there information on pH available from previous studies perhaps?

We omitted this Section as part of major comment of both reviewers. However, we will carry this point of feedback into the future FTIR vs Py-GC-MS paper.

l230-235 can you add these data to the Supplement? It would help to understand what has been done

We added all these data to Table S6 as requested by the reviewer.

L241-242 if two different columns were used, could you indicate in supplement which samples were with Restek and which with the other one? To avoid that this separates the samples as a different column may contribute to the signal.

These columns are identical in their chromatographic performance. Replicates 9 and 10 of the analytical standard (Table S3) are from the DB-1ms and show identical results. Any offset would be covered by the uncertainties reported in Table S3. We will specify which samples were ran on the DB-1ms in the methods:

We added to line 242: *"... or an Agilent DB-1ms column (60 m x 0.32 mm i.d.; 0.25 µm df; fused silica**; for any peat samples from UKCB, UKTR_2, and replicates 9 and 10 of SSPS2 4-7cm**)"*.

L248 ion fragment instead of ion chromatogram

We intend to say "*extracted ion chromatogram*" here, as was written. This term is defined that we are looking at the chromatogram from a specific ion fragment. The text remains unchanged.

L255. Could you explain how products were selected? Is it based on dominance, how?

Added line 248: "***These extracted ion chromatograms were selected to optimize peak integrity by choosing the most abundant, non-coeluting ion fragment occurring in the pyrolysate***"

Do you have a homogeneous dataset, i.e. all selected products were quantified for all samples? Or did you quantify each sample based on the abundance in that sample?

Adjusted line 250: "*All automated integrations and identifications* ***were conducted over all samples for all identified pyrolysates,*** *and checked manually…"*

What is meant with peak area of all? Please make sure tha no misunderstandings can occur in how data were obtained.

Adjusted line 255: *"… by the total peak area of all* ***pyrolysates****"*

L265 what is meant with cellulose? Pyrolysis of cellulose results in many products that cannot on beforehand linked to cellulose since they are now unique for cellulose.

This was discussed in major comment 3. See above.

L266-267 I would not name this grouping functionally. It is more very strictly (but out of context) chemical structure. Please define the full names of abbriviations P G S etc. it is confusing to use aromatics for benzenes and also for overarching aromatics.

This line refers to section 3.3.1, which interprets the pyrolysate origins into likely biomacromolecule precursors, which both allows links between vegetation precursors and connect to a broader suite of readers. It is also a very common approach in this literature. Such assignments do come with caveats, which we now make more clear. In any case, we are transparent in our interpretation, and future workers can use our SI to group differently as they see fit. We changed the aromatics as discussed in major comment 2.

R&D

L270 bands instead of carbon?

We omitted this Section as part of major comment 1 of both reviewers. However, we will carry this point of feedback into the future FTIR vs Py-GC-MS paper.

L273-274 how do you know it reflects vegetation and not oxidation?

We omitted this Section as part of major comment 1 of both reviewers. However, we will carry this point of feedback into the future FTIR vs Py-GC-MS paper.

Fig 2 seems that some records continue to deeper sections, why are these not included?

We will include Fig 2a in this revised manuscript, therefore we must address this point. The reviewer is correct that we have deeper data from some sites. The deeper samples exhibit less change and this common depth axis for all sites better shows the steeper changes in composition that occur in shallow peat. The same and complete data is fully represented in Fig. 9 on a logarithmic scale.

L274 I suppose everglades is USWCA1. Can you provide the abbreviations that are used in the caption also in the text?

We omitted this Section as part of major comment 1 of both reviewers. However, we will carry this point of feedback into the future FTIR vs Py-GC-MS paper.

L274-275 why is here the low sugar content interpreted as decomposition and above as plant characteristic? Perhaps it would help to be consistent in mentioning the vegetation type.

We omitted this Section as part of major comment 1 of both reviewers. However, we will carry this point of feedback into the future FTIR vs Py-GC-MS paper.

Section 3.1 it is very difficult to follow the reasoning if site abbreviations and names are not used consistently

We omitted this Section as part of major comment 1 of both reviewers. However, we will carry this point of feedback into the future FTIR vs Py-GC-MS paper.

L297-300 It may also related to mineral interference in pyrolysis-GCMS? Catalyst reactions can increase aliphatic and aromatic products (not silica, but metals, carbonates, and clay minerals)

We omitted this Section as part of major comment 1 of both reviewers. However, we will carry this point of feedback into the future FTIR vs Py-GC-MS paper.

309-322 I think this is not really relevant. And can be excluded completely. Perhaps parts of it can be moved to the discussion below, if relevant there.

We omitted this Section as part of major comment 1 of both reviewers. However, we will carry this point of feedback into the future FTIR vs Py-GC-MS paper.

L324-331 Caution if minerals other than silica are involved. 324-326 can be excluded I think. The semi quantitative aspect of the data is well-known and no need to prove that again. I would suggest to compare with TOC and CN. The matrix effect can be relevant and is not limited to the free aliphatic extracted fraction. The last sentence is also out of context.

We omitted this Section as part of major comment 1 of both reviewers. However, we will carry this point of feedback into the future FTIR vs Py-GC-MS paper.

Section 3.2

L338 needs more explanation. Refer to fig A1?

Adjusted line 338: "Specifically, **141 distinct** pyrolysates **are identified**, integrated, and assigned into structurally distinctive groups with similar origin **(Fig. A1)**."

L339 phenols in your list are not p-hydroxyphenol lignin, except for 4-vinylphenol

Most pyrolysates in the list can be derived from p-hydroxyphenol lignin, as we observe in our lignin standard (although the purity of isolated alkali lignin can be disputed), which is why we group them as *p-coumaryl lignin-derived.* We were clear in the text that these can derive from other sources as well: "*p-coumaryl lignin monomers or other polyphenol sources*".

L347 23 alkenes?

Good catch! Adjusted to: "(Al; *n*-alkanes, n = 23 and *n*-alkenes, n = 2**3**)"

L346 proteins from fungal/microbial biomass instead of i.e. proteins. Plants also contain proteins

Good suggestion. Implemented: "mostly attributed to **proteins from** microbial/fungal biomass"

L349-350 phenols do not represent lignin

Agreed. Adjusted to "…*represent* ***aromatic macromolecules such as*** *lignin include…*"

L349-351 I find it difficult to read this sentence. Could you simplify or clarify.

Agreed. We simplified the sentence, for specific structures the reader can refer to the caption: *"…the phenols (e.g.* ***P1/P2/P3****), guaiacols (e.g.* ***G1/G2/G3****), and syringols (e.g.* ***S1/S2/S3****)."*

L354 so you quantified the products found in the peat also in the plant pyrograms? I suppose the plant were very very different? it is a good approach, but it is not clearly explained

Agreed, we clarify in line 349 to highlight this: "*similar* ***origin in peat, and the same pyrolysates were integrated in the plant samples****"*

L357 their assigned macromolecular origin. But I do not see that in fig A2, aromatics from wood and sugars from peat with other vegetation type? So you likely mean the chemical structure instead of origin, it also demonstrates the phenols being very variable.

The confusion is related to the interpretation of the word origin. We will avoid this word in this line and signpost the phenol variability (which is properly discussed later): *"Pyrolysates grouped into their respective macromolecular assignments (aliphatic, lignin and polysaccharide) at similar loading angles,* ***besides the phenols, which will be discussed later****.* ***This grouping of the loading angles independently confirms their likely parent macromolecule****"*

L 365 it is not macromolecular assignment but chemical structure assignment.

We agree that the chemical structure can be deduced from the pyrolysate structure alone, but their biomacromolecular assignment cannot. This section, however, argues that the specific groups of pyrolysates (aliphatics etc.) show the same behaviour across the whole dataset and concludes that their parent macromolecule is likely to be similar.

The text remains unchanged.

The title of 3.2 is not clear for me, it seems that you mean vegetation type. Instead of shifts in plant species distribution. Please clarify, because this is something substantially different. If you mena plant species composition, then you should take into account changes with time/depth, and how you decipher this from decomposition. Or mention that you assume a homogenous peat (in terms of plant composition) with depth

We intended to keep the titles short and to the point, but nuance got lost here. We adjusted it to vegetation type as suggested by the reviewer: *"3.2 Vegetation* ***type*** *leaves a signature in the peat OM, even when decomposition transforms the OM significantly"*

L357 Be cconsistent in the grouping, in the figure caption it is different.

Agreed. Changed to *"...**(e.g. guaiacols/n-alkanes)**..."*

Fig 4a. is the symbol of the plant sample referring to the site it was sampled? Would it make sense to aseparate them based on family and or tissue? The symbols of samples and their name are not readable or difficult to read

Yes, the symbols (and therefore both legends) are consistent with both panel a and b. We prefer to keep the symbols consistent with the peat biplot, so that readers can easily cross-check the peatlands major vegetation with the peat in panel b (this information is not accessible without utilizing the symbols). We enhanced the visibility of Figure 4 as mentioned in major comment 5.

We clarified in the caption that: *"pyrolysate vectors coloured by macromolecular assignment and shape by site of origin **in both panels**"*

L370 could you visualize this?

See reply to line 384 below.

L371. Was the trunk sample without bark? If so that is worth mentinoning as it would explain this observation

The trunk samples do include bark, but the majority of the material is indeed non-bark. We added this detail in the methods line 202: *"trunk samples are from a single individual **with bark included**"*

L374 can you calculate whether the 29 and 31 indeed had a higher contribution to the alkane group in these samples, and to which extend? If not sufficient, exclude this explanation on residual plant waxes? You have the data, so no need for speculation and unclarity.

We dedicate a whole section (3.2.3) to elucidating this, so we have signposted that to the reader here: Line 375: *"...leaf cuticle, **which will be explored in Section 3.2.3**"*

I cannot read the samples of plant and litter

We assume this is in reference to Figure 4, which was dealt with in major comment 5

L377-379 I see two phenols with low negative loading, I would not name that a variety, and your data seem to indicate that this is due to polysaccharides and cellulose in particular

Agreed. We changed the wording to: *"lignin and **two** diagnostic polyphenol pyrolysates"*

379 combine with lines 371-372

Our approach is to structure the paragraphs by most important variability (PC1) downwards. This section explains behaviour on PC2, therefore this line is better suited here.

No changes made to the text.

PC3 aliphatics and two phenols negative, and some sugars and aromatics positive along with one or two phenols.

This comment was in major comment 5, which we discussed and resolved there.

L383-385 again you come up with explanations that are not visible in the data. if you see a separation in P and S lignin, then label the compounds so that this can be verified by the reader (given your compound list, I suppose that the phenols with negative loading on PC3 is 4-vinylphenol?) Ah now I see that the aromatics may be syringols, could you use a more contrasting colour, I hardly see a difference. What about the G lignin negative on PC4, is this not worth to mention, seems that PC3-4 separates the different lignin moeites.

This section is about PC3 specifically. With the included loadings in the SI, the interested reader is able to discern the exact pyrolysate loading angles. We improved the overall readability and visibility of Fig 4 (see reply to major comment 5), which makes the separation between P and S lignin more clear. The scree plot in figure 4 shows an elbow at PC3, where significance drops (discussed in line 368-369). Therefore, we chose to not discuss PC4 in the text, hence the omission of guaiacols, since it does not vary on PC3.

L384-385 it is not clear for me, the figures must be improved. What do you mean with plant type, species, or family, tissue? Main controls on plant composition, this sentence seems self evident?

We clarified what is meant by plant type by referring to specific plant characteristics discussed earlier in the text: *"…indicates that tissue **(i.e. trunk/root/leaf)** and plant type **(i.e. softwood/hardwood/sedge/moss etc.)** are the"*

Perhaps consider moving the whole plant PCA to the supplement, and summarize in a table the main findings of PC1-PC3, regarding tissue and species. If the individual products behave according to chemical grouping, why not provide a Table with the average abundance of groups for familily and tissue. or somehow visualize the differences between plant samples better

We agree that we should improve the clarity of the discussion of plant samples better. Given related comments by reviewer 1, we will expand and clarify the distinguishing features of the main taxa. Due to manuscript length we had only a limited discussion on the composition of plant tissues in the original submission. That already confirmed or revealed high level differences, such as the lack of lignin-associated components in Sphagnum and the greater proportions of S-lignin in trunks, P3 in sedges, and phenolic moieties in palms (Fig. 5). We also note that the inclusion of all of our raw data in the Supplementary files is a rich resource for future researchers to examine and interpret at their discretion. However, we agree that we can explore this in greater depth. Given the encouragement of both reviewers and the space afforded by removing the FTIR comparison, we can now expand that discussion. We do not think it is appropriate to have in an in depth discussion on every species or plant type and will focus on the major taxonomic groups of peat-forming plants as examples (i.e. Sphagnum, sedges and palms*), and especially those features that are retained in peat and play a key role in distinguishing peat types. The suggestion to include a table that summarises the key products (or some other visualisation) fits perfectly and we will incorporate that.

L389-391 again a statement from the literature that is not connected to the data. you have the data, so if diketodipyrole is quickly degraded it should be higher in plant than peat. From your N compounds the others are not commonly associated with plant proteins in SOM.

Excellent suggestion. The proportion of N-compounds is significantly higher in peat than plant material, which supports the microbial biomass mechanism.

We rewrote as follows: Line 391: *"can act as a proxy for microbial/fungal necromass in peat (Kaal et al., 2017),* ***which is supported by their significantly higher proportion in peat over plant material (p < 0.05).*** *"*

L392 are catechols derived from G and S lignin, is that from the reference you provide? and.. catechols were likely present in your sphagnum samples as well. Did you have them in you alkali lignin standard?

Good catch, while the catechols do vary with the G and S groups (G and S covered by the reference, not catechols), the sentence suggests a pure lignin origin, which is incorrect. It can have wider polyphenolic sources, which was established earlier in the paper (line 343). We also find methyl catechols in both the alkali lignin standard and *Sphagnum* tissues.

We omitted catechols from the sentence.

386-391 I see the reverse for alkanes etc: positive. Can you also provide an overall interpretation for this combination of products, N, aliphatics and aromatics? It may also point towards algae?

Good catch, the order of positive and negative is reversed here. After rerunning the PCA in later stages of the manuscript, the polarity flipped, which is common and the axis polarity in PCA has little interpretive value. It is useful in the text to draw the reader towards certain effects, so we will change the text. We further elaborate in the origin of aliphatics in section 3.2.3.

We swapped positive and negative in the sentence: *"is described by* ***negative*** *loadings of polysaccharide pyrolysates and* ***positive*** *loadings…"*

L394 not relevant if they are common here, but where they plot in PCA, and how they differ from the furans etc., and this is not visible from the figures.

The sentence before links the term "heterocycles" to the polysaccharides, which all have similar loading angles as the sentence dictates. We interpret the levomannosan/galactosan as polysaccharides and levoglucosan as cellulose, which has different colour arrows.

We have improved Figure 4 readability.

L398 do you mean aromatics or aromatics?

Highlighted and resolved in major comment 2

L401 more complex than pyrogenic OM and polyphenols? It is not clear what you want to say here, can you indicate what type of molecules

We added the common examples to the line: *"..of more complex aromatic-hosting compounds* ***(e.g. lignin)*** *"*

L401-402 I see negative on PC3 particularly aromatics and N

Good catch, the order of positive and negative is reversed here.

We swapped positive and negative in line 398 and line 401.

L403 I see that the catechols are particular high on PC4

This is interesting to point out, so we added a new sentence: "***PC4 also features a strong negative loading from the catechols, which hints at a distinct macromolecular origin detached from other aromatic-hosting compounds***."

L408-409 p-coumaryl = 4-vinylphenol

4-vinylphenol is one of many pyrolysis products of p-coumaryl lignin as is communicated here. They are not identical.

L410-411 yes seems logical , but that can be said on beforehand with the ID of your selected phenols.

This paragraph explains the diversity found in phenols based on literature (and therefore only the ID), therefore uses the same ratio as the reviewer suggests.

Tor Royal etc is not indicated in the figure or its caption, please provide the abbreviations also in the text.

We like to refer to site names at times to keep the text engaging (the acronyms are introduced in Fig. 1 and section 2.1., but we know that readers will not always remember these assignments). We appreciate that this leads to confusion. We added all acronyms to every mention of the site names throughout the manuscript.

"Tor Royal" to *"Tor Royal* ***(UKTR)"*** on lines 376, 412, 447, 666

"Everglades" to *"Everglades* ***(USWCA1)"*** on lines 412, 447, 495, 534, 580

"Cananguchal" to "Cananguchal ***(CAC)"*** on lines 473, 490, 611

"Pantano" to "Pantano ***(CAP)"*** on lines 490, 595

"Inirida" to "Inirida ***(CIH)"*** on lines 490, 604

"Mpologoma" to "Mpologoma ***(UMp_PA2)"***on lines 454, 481, 495, 538, 645, 653

"Panama" to "Panama ***(SSPS)"*** on lines 495

"Crymlyn Bog" to "Crymlyn Bog" on lines 490, 656,

L416 aliphatic are positive on PC1

Swapped positive and negative in text:

"*Tor Royal (UKTR) and the Everglades (USWCA1) are the only sites that group relatively tightly in the PCA1–2 plot. Tor Royal is characterised by positive loadings on PC1* ***and negative on PC2****, similar to the polysaccharides. However, the deeper Tor Royal samples exhibit a* ***decrease*** *in PC1 likely driven by aliphatics. The Everglades has positive loadings on PC1* ***and PC2****, similar to the aliphatic and lignin-derived components."*

L417 I see UMP all with negative scores on PC2 and positive ones on PC1. I see USWCA positive on PC1 and PC2.

Swapped positive and negative in text.

L417-418 very difficult to extract from the figure. And particularly a loss of sugars, instead of enrichment of lignin and aliph?

We improved the visibility of Figure 4. In this PCA, it is easiest to see the enrichment of lignin and aliphatics. We discuss the mechanistic implications of this later in the manuscript (3.3.2) (it is likely a preferential loss of sugars that causes this enrichment.)

L418 it is not peat decomposition at depth but more decomposed peat at depth

We understand the confusion, we updated the sentence for clarity: *"This likely represents* more decomposed peat *at depth"*

L420 what is common, the loss of sugars with depth? Please be clear

We will be more specific with what is meant with "transformation" instead:

"…*shows that* ***the enrichment of aliphatic and aromatic macromolecules*** *is remarkably*…"

420 thus assuming that the vegetation was similar with time

We present data-driven evidence that vegetation change cannot solely drive this enrichment in Section 3.3.1, therefore we feel comfortable making this statement here. This section explores this thoroughly.

L425 not clear what you mean. Why stronger preferential decomposition (of what) (with depth?) at tropical sites.

We signpost later Sections here too early. We removed the part of the sentence that caused the confusion: *"~~with a stronger preferential degradation signal in the tropical sites~~"*

Section 3.2.2

L430 is there a difference in peat composition and OM characteristics?

Peat itself has constituents that are not in the OM. We do not intend to communicate that we can understand the material "peat" holistically at any point in the manuscript.

Nonetheless, we will reiterate that we are looking at pyrolysates in this line 430 "*Overall, the PCA on* ***vegetation pyrolysis products*** *shows a dominant axis of variance based on the lignin and aliphatic macromolecules, whereas the variance in polysaccharides drives the peat ordination*"

Fig 6 If you could plot alkanes and alkanes below each other it would be easier to read this figure. What is the meaning of fractional abundances? Percentage of aliphatics? Alkanes as percentage of alkanes?

One of the main objectives to communicate with this graph is how linked the alkane/alkene abundances are (i.e. the ratio). Plotting them next to each other gives the most instinctive representation. Fractional

abundance is defined as the abundance as a fraction (20% = 0.2) and the caption specifies it is calculated within the aliphatics group (alkanes+alkenes). No changes have been made.

Section 3.2.3 can be reduced to a few sentences in my opinion. There is too much detail that is not connected to the data, and or the data are not relevant in the context of the ms (e.g. fig 6b)

Discussing the mechanisms that drive the enrichment of aliphatics in peat samples is important in the context of OM recalcitrance as it makes up a significant portion of the OM (See also the comments from Reviewer 1). Other techniques have also highlighted this (FTIR/NMR) and interest is developing in incorporating this unique OM pool into our peatland understanding. We believe our data enables a unique opportunity to discuss the mechanisms that drive this OM pool.

Reviewer #1 highlighted that the structure was confusing, which we adjusted. We are confident that this section is relevant and well-structured in the revised manuscript.

The ternary diagrams can perhaps be moved to supplement as well? To me it is not helping to get the message.

The ternary diagrams are central in the discussion where we move from observations (PCA/HCA of pyrolysates) to interpretation (lignin, hemicellulose etc.), and eventually link to wider implications (logarithmic decline & microbes). We are confident the ternary diagrams help the message.

I stopped with reading in detail here. I recommend to first improve the focus.

We thank the reviewer for the meticulous read of previous sections and the careful checking of the data and supplements. We appreciate the incredible amount of work put in this review. However, from this point in the paper, the focus is narrowed again and we move to interpretations that are more impactful. We would have really appreciated the same level of detailed-oriented review in this section in addition to Reviewer #1's.