

Thank you for the opportunity to submit our revised manuscript for publication in Biogeosciences. We revised the manuscript according to the suggestions by the reviewers. We accordingly implemented our proposed changes, see the discussion of the manuscript on egusphere. In addition, we list our responses to the editorial decision in bold. According to these changes, the Figures 2, 7 and 9 have been slightly modified.

Thank you for posting your author replies to the two review reports. Both reviewers are overall supportive of your work, but they also flag concerns which should be addressed in a revision. After reading through the review reports and your replies, I believe that this manuscript can become acceptable after major revisions. Especially Reviewer 1 raises some critical points, many of which you have replied to satisfactorily and some of which rest on some confusion regarding the methodology. Regarding the latter, I believe the manuscript would likely benefit from some clarification of the methodology and an attempt to make it more accessible for a somewhat broader audience (as per the comments and replies to the report of Reviewer 1).

We have clarified the methodology, with explanatory comments what the model parameters represent. Importantly, we realised that part of the confusion of Reviewer 1 was likely caused by a missing paragraph and a title of a subsection of the methodology, which somehow got lost during the copying process from a Word version to a LaTeX version. Unfortunately, this missing part covered critical information regarding the morphological data used in the study. Furthermore, this error led to the integration of the rest of the morphology-based phylogeny section into the section on the geochemical data collection. We added the missing part and apologise for this small but consequential mistake. We made sure that the revised version is complete.

In addition to the changes suggested, I would suggest that you carefully consider the following aspects in your revision:

In their first point, Reviewer 1 poses the question whether the data underlying the study are sufficient to draw conclusions about the taxonomy. I think this question was wrongly interpreted in the Author Comment: The reviewer does not actually ask for more (new) data to be added to the study. Instead, if I read the question correctly, they would like to see a demonstration that the structure of the dataset (with a large part of the data coming from a small number of specimens) does not negatively affect the conclusions concerning the phylogeny. It seems that this could be achieved by doing a statistical test, and the authors should consider this. In its simplest form, the authors could average (or take a median value of) all the data per taxon before the analysis to check if that gives the same outcome in terms of clearly separate chemical compositions for different taxa. I believe a test like this would be a good addition to the manuscript.

As explained in our responses, we did not draw any conclusions about the taxonomy based on geochemical data (this should now be clear, see previous point). The proposed statistical test (average of all data per taxon) is essentially already present in the manuscript, see Figure 2, where the distributions of all element ratios for each taxon are shown. It would be possible to use a statistical test (e.g., Kruskal-Wallis test) but the separation between different taxa is already so clearly visible that we do not see an added benefit here. Nevertheless, if the editorial team insists on such a test, we are happy to include it in a final version.

Similarly, in reply to comment 5 of Reviewer 1, it would be useful to add a “sensitivity test” to demonstrate that including the potential diagenetic data from Sørensen et al. does not significantly affect the conclusions, as stated by the authors in their reply.

Thank you for this suggestion. Since the potential diagenetic data from Sørensen et al. make up less than 5% of the total dataset and affect only a single taxon, we anticipate that this would change very little in the overall result. Instead, we performed a sensitivity test by randomly altering the input values (i.e., element ratios) of all taxa, on average by about 10% (see updated Methodology for details), and repeated this process 100 times. Note that this is a change to the median value, so depending on the sample size of the taxon, it would take quite a lot of data in reality. This would represent a hypothetical, severely biased dataset compared to a slight change in a single taxon. Nevertheless, and to our surprise, the results appear to be quite robust against such alterations, which strengthens our conclusions.

In reply to the third main comment by Reviewer 1, I agree with the authors that it is a good idea to highlight the applications of the presented approach a bit more clearly in the discussion. In addition, the authors may consider reviewing the methodology section to make sure that their approach for linking chemical data to phylogeny is clear, as the reviewer was apparently confused by this.

We considerably extended the discussion to outline potential applications. As for the second part of the comment, see above.

Re the comment of Reviewer 1 to line 166, I would agree with the reviewer that the median value would probably be a more meaningful statistic for describing these data. If it is not too much effort, I would suggest using the medians as descriptive statistics.

We checked the values again and noticed that the mean value in the manuscript was actually the result of a typo. Thus, we had used the median value all along,

presumably for the same reasons. Apologies for this oversight – it was calculated at a very early stage of the study and then apparently incorrectly noted.

Finally, please carefully consider the general point by Reviewer 2. It seems that a more detailed discussion of the value of Phylogeochimistry including recommendations for future studies to flesh out this new research niche are warranted.

This was also highlighted by Reviewer 1, so the discussion is now considerably extended and we provide examples and recommendations for future studies.