Reviewing article 95688

TITLE: Contribution of metabolomics to the taxonomy and systematics of Alcyonaceans: case study in the Tropical Eastern Pacific

1. General comments

In the present manuscript, the authors applied an integrative taxonomy method to identify and classify 12 Octocoral species (5 genera) of the REMAPE (Marine Protected Area El Pelado) from the Tropical Eastern Pacific near Ecuador. For that purpose, morphologic and DNA barcoding analyses were implemented along with metabolomic analyses on a total of 71 collected samples. The objective was to assess the potential of comparative metabolomics combined with multivariate analyses to support complementarily the discrimination and identification of different Alcyonacean samples down to the species level. Metabolomic analyses were shown to contribute to species discrimination with identification of a new species when other analyses such as genomic and morphologic analyses provided unclear results. The benefit of such integrated taxonomy method was illustrated in the manuscript with Muricea species and with *Leptogorgia cf. alba*, respectively. This is an interesting manuscript showcasing a taxonomic toolbox for the description and identification of such keystone marine species. The transparency and rigorous raw data sharing will increase biological and chemical knowledge on this Octocoral species.

This research article is within the Aims and Scope of the journal. The structure of the article follows the PeerJ guidelines. All raw data were provided and are freely available. The article is written in correct English, nevertheless a few sentences will require clarification and will need to be reformulated. The experimental design from sampling strategy down to the acquisition of metabolomic data were done rigorously. All the raw data are freely available. The table compiling sample codes and their description allow researchers to trace back the available raw MS-data down to their original species. This was done rigorously and is very important in the context of raw data sharing and research transparency.

In general, the interpretation of metabolomics results through multivariate data analyses lacks precisions, and in a few instances inappropriate vocabulary was used. The figures could be improved. A figure illustrating the complementarity of metabolomics with other taxonomic methods would bring more depth to the presented results. Suggestions are provided below.

Major revisions are required as per the following detailed comments.

2. Abstract:

The abstract should also summarize key results showcasing in what aspects metabolomics was found to be a complementary or useful tool. A sentence such as the one written on lines 428-429 would help.

The authors wrote: "Additionally, this approach allowed a quick selection of putative bioactive compounds for further MS - guided isolation work among the metabolome of species of interest." This has not been illustrated through the presented results or sufficiently exemplified.

3. Introduction/ state of the art

Research hypothesis is stated in the introduction (lines 103-104): metabolomics analyses can help in the systematics of Octocorals from REMAPE.
The objective of the study is clearly stated in the introduction (lines 82-83, and lines 109-111). The objective is to assess the potential of metabolomics to complementarily discriminate Alcyonacean taxa down to the species level for Muricea using an integrative taxonomy approach.

Line 109: The authors mentioned '' the first objective…'' but in fact there is no ''second objective'' so the word first could be removed.

In the paragraph starting at line 84. The author could give the definition of what integrative taxonomy is (as opposed to chemo-taxonomy). Such explanation would also give more strength on the originality of the current study with the use of metabolomics. There are other research articles pertaining to the metabolomic study of octocorals to cite in the introduction. A few references are missing as suggested below:

- Gerhart, Donald J. "The Chemical Systematics of Colonial Marine Animals: An Estimated Phylogeny for the Order Gorgonacea Based on Terpenoid Characters." The Biological Bulletin 164, no. 1 (February 1983): 71–81.
- Coll, John C. "The Chemistry and Chemical Ecology of Octocorals (Coelenterata, Anthozoa, Octocorallia)." Chemical Reviews 92, no. 4 (June 1992): 613–31.
- Imbs, A. B., and T. N. Dautova. "Use of Lipids for Chemotaxonomy of Octocorals (Cnidaria: Alcyonaria)." Russian Journal of Marine Biology 34, no. 3 (May 1, 2008): 174–78.
- Kessel, Gustav M, Philip Alderslade, Jaret P Bilewitch, Kareen E Schnabel, and Jonathan P A Gardner. "The Use of Integrative Taxonomy in Octocorallia (Cnidaria: Anthozoa): A Literature Survey." Zoological Journal of the Linnean Society 198, no. 2 (May 25, 2023): 677–90.

Lines 111-114 The last sentence of the introduction would have a better place in the discussion of the metabolomic results.

4. Experimental design & Methods

- The Experimental design in terms of sampling strategy has been done rigorously notably with regards to biological replicates. Methods pertaining to sample extraction and metabolomic data acquisitions are adequately described.
- **Line 229** a new subsection should be added with the title ''*Multivariate data analyses*''.
- **Lines 248-250.** Methods and algorithm used to perform HCA should be added.
- **Raw data** pertaining to the LC-MS based metabolomics profiling are freely provided here: https://figshare.com/s/67d62277dc542c0f1d13. The raw data accession is working properly. The Genbank accession codes in table 1 are also adequately assigned.
- **The table 2 compiling sample codes and their description** allow researchers to trace back the available raw data down to their original species. This was done rigorously and is very important in the context of raw data sharing and research transparency.

5. Validity of the Findings

- **Addressed knowledge gap**: The combination of metabolomic and chemometric analyses were shown to contribute to species discrimination and identification of a new species when other analyses such as genomic and morphologic analyses provide unclear results. The benefit of such integrated taxonomy method was illustrated in the manuscript with Muricea species and the *Leptogorgia cf. alba*, respectively. The authors also counterbalanced their findings by acknowledging that metabolomics analyses combined with multivariate analyses did not provide perfect adequacy with Phylogenetic tree obtained though appropriate DNA sequencing.

- **Better addressing the complementarity of results**. The figures assembling and illustrating the collected results should better convey the complementarity of metabolomics to other taxonomic analyses. A panel figure illustrating side by side the results from one type of analyses as opposed

to those obtained by chemometrics would help the reader in understanding such complementarity. This could be done with regards to the Muricea species: a phylogenetic tree could be presented in one panel and on the other the HCA or a heatmap obtained with MS-based metabolomic data. The same color code should be used for the different figures of the manuscript.

- Fig 3 and Fig 4 could be combined in one single figure panel. The PLSDA validation results of fig 3b could be placed in supporting information instead, also the validation results do not show the R2/Q2 values as specified in the material and method section. Table summarizing the cross-validation results would be better than the exported graph from Metaboanalyst web-page.
- The VIP table in fig 4 does not bring enough information to be presented alone, unless some feature annotations are provided. There is no explanation on what the feature names M..T means in the figure legend. The scores are obtained for feature projections on a PC axis whose identity has not been specified.
- On fig 5, the different sample names on the HCA are a too small to be read correctly.

- In general, the interpretation of metabolomics results through multivariate data analyses lacks precisions. There is a **misuse of the term metabolites** or molecules in the article. What is presented throughout the manuscript correspond to chemical features, i.e signals detected by mass spectrometry that are molecular ions, adducts or in-source fragment ions of metabolites. One metabolite can provide different MS features, and since there is no annotation of the MS data, the authors can only use the term chemical feature.

- **Pages 14/15 with PLSDA analysis and VIP scores**.
  - **Line 335 and line 373**: A pairwise post-hoc permutational test would have been more accurate to evaluate the differentiation between each group, that is between species within the Muricea genus and between samples of the Muricea and Leptogorgia genera.
  - **Line 342**: Which PC axis was used for the projection? This was not indicated in the text nor in the figure legend.
  - **Lines 348 and 379**: "*Among the 15 VIPs, XY were highly expressed"*..a feature is not expressed (it is not a gene), a feature is detected with a level of intensity, you can say line 348 "*Among the 15 VIPs, 13 detected features were proportionally the most intense''.*
  - **Lines 350-351, Lines 380-381: an annotation of features** (as these are not compounds at this stage) could have been provided instead. Annotation requires at least to provide a table with the measured and calculated *m/z*, the retention time the types of ion detected (e.g. $[M+H]^+$, $M+NH_4]^+$) and the proposed molecular formula of the detected ion. Such table can already bring some hints as per the possible identity of the detected ion and thus possibly molecule.
  - **Lines 389-390: "t**he VIP analysis showed that both species *M. plantaginea* and *M. squarrosa* produced distinct metabolites from other Muricea species" again here at this stage, the VIP does not show any differences in terms of metabolites presence of absence but only in terms of chemical feature relative intensities…. So the authors can only say '' *The VIP analysis shows that the composition in terms of chemical features is different between M. plantaginea and M. squarrosa''*
  - **Lines 391-392: "***These very specific metabolites could explain the difficulties encountered during the gene amplification of both species.''* The authors can't say this as they don't know what these possible metabolites are. They should explain why they think these metabolites may have affected the PCR process.

6. Other comments

- **The abstract** mentioned "Additionally, this approach allowed a quick selection of putative bioactive compounds for further MS - guided isolation work among the metabolome of species of interest" **This has not been illustrated in the results section of the manuscript**. It seems to be part of an additional objective of the study that has not been performed here in the submitted manuscript.
- **Discussion**: The authors could also discuss their results with regards to the literature pertaining to chemotaxonomically investigations of Octocorals and thus propose some hypothesis as per the identity of discriminant features is concerned. What type of molecules are expected to be possibly discriminant in the analyzed Octocoral samples? The authors could mention previous results with the identification of lipids, steroids and/or terpenoids as discriminant metabolites for other Octocoral species (see also remarks pertaining to the literature).
- **In the discussion, Page 17 line 458** the authors could briefly define what a synapomorphic character is in order to better understand why molecular networking methods would help in identifying such synapormorphic metabolites.
- **In the supporting information:** A series of figures S8-154 were added at the end of the document but are not presented or mentioned in the manuscript, and, thus, do not have their place in the context of this manuscript.
- In general, the article is written in correct English. Nevertheless, a few sentences will require to be reformulated for more clarity as noted below

Page 12 Line 264: "*Therefore, it has tentatively named as Leptogorgia cf.alba''* please correct as: ''Therefore, it has been named tentatively as *Leptogorgia cf. alba''*

Page 15 lines 362: This sentence is strange in my opinion**:** "On the other hand, the dissimilarities between the metabolomic profiles of the three species from the genus Leptogorgia (OC-55, 65, 66, L. alba and OC-67, L. cf. alba and OC-74, 75, 72 L. obscura) were separated, and this result was well supported on both phylogenies (mt-COI and mt-MutS).'' I would suggest the following *"On the other hand, the differences in the metabolomic profiles among the three species from the genus Leptogorgia (OC-55, 65, 66, L. alba, OC-67, L. cf. alba, and OC-74, 75, 72 L. obscura) were evident, and this outcome was well supported by both phylogenies (mt-COI and mt-MutS)*."

Page15 lines 396-398: "*This result highlights a potential of the metabolomic analysis to distinguish, but place closely in the HCA, sister species that the molecular analysis could not reveal so far*."
The sentence is not clear and could be reformulated. For example, ''This result highlights the potential of metabolomic analysis to distinguish closely related sister species, which molecular analysis has not been able to reveal thus far.''

Page 16 lines 413: "….*species delimitation where molecular data could not discriminate*". Please modify for more clarity: "*…species delimitation in cases where molecular data alone could not discriminate*"

Page 17, Lines 468-473: Please reformulate this sentence is too long and something is missing**: "***In Jaramillo et al (2018), using both a untargeted and targeted metabolomic approach with UHPLC-HRMS was demonstrated that this method was very efficient as a complementary tool in zoantharian taxonomy but also for their classification and some specialized metabolites like zoanthamines or 2-aminoimidazole families of natural products identified as key biomarkers for certain species as shown for other groups by Vohsen,Fisher & Baums (2019)."*