

Review of Graco-Roza et al. - Comments to the Authors

The manuscript by Graco-Roza and colleagues, entitled “**Clumpy coexistence in phytoplankton: the role of functional similarity in community assembly**”, investigated both spatial and temporal distributional patterns of a phytoplankton community located in a tropical region of Brazil, the Piabanha river, under the hypothesis of Emergent Neutrality. To do this, they raised two main questions: 1) *differences in species biovolume are linked to their functional differences at the clump level* and 2) *species with higher distinctiveness will show higher relative biovolume within clump*; and use a 1-year time series to confront theoretical expectations with field observations. From their analysis, they identified two stable main clumps in body size with species belonging to three MBFGs, and found that among-species differences in biovolume were driven by functional redundancy at the clump level. In addition, the relative biovolume of species was related to species functional distinctiveness at the clump level. The authors concluded from their study that species with similar traits are selected under given environmental conditions.

This study is topical – especially in the current context of intense forcing on biodiversity - and can be an interesting contribution to improve our knowledge on how species are organised/distributed in nature and whether it is possible to determine species distributions, in different kind of ecosystems, on the basis of a theoretical framework. While this paper has a potential to be accepted, some important shortcomings have to be clarified or fixed to make it suitable for publication. I here summarise my main concerns:

1. Please revise the conclusion in the abstract as no relation with environmental conditions had been clearly investigated; or I have missed that information.
2. It is quite unclear to me what is behind the term “clump”. As defined by the authors, a clump is an aggregation of species. However, it could be useful to provide a more detailed and precise definition of clumpy patterns, to mention whether they are stable or episodic patterns, to discuss the relation between species’ niche and clumps (e.g. Segura et al. 2011), etc...
3. While the Introduction section provide useful information for the readers, more attention should be devoted to the different theory (for example, by defining the main differences between the ecological *sensu* Hutchinson and the niche following the neutral theory); I sincerely trust that it could help to clarify the main message of the paper.

My feeling when reading this section was that more precise information would help to strengthen the paper; when previous works are cited, it could be helpful to be more precise; e.g. “from a wide range of ecosystems”: which kind of ecosystems? Has this theory been tested on different kind of ecosystems (e.g. polar, temperate, etc...) or the same kind of ecosystems in different areas?

In the Introduction section, some sentences need to be redrafted for clarity; e.g. “*These predictions were tested theoretically at both steady (Fort et al. 2010) and fluctuating resource conditions (Sakavara et al. 2018), but beyond robust theoretical results, empirical evidence of EN is still scarce.*” I do not understand the dichotomy made by the authors. The same applies for: “*However, even small differences in organismal traits are linked to species performance in the niche space (McGill et al. 2006; Violle et al. 2007).*”

“*[...] disclose the trigger of stochasticity driving the species coexistence*”: Please, clarify this sentence.

Please clarify the following sentence (e.g. by providing an example from the literature): “*However, although two functionally redundant species most likely show high similarity in trait combination, functional distinctiveness is not directly linked to redundancy*”.

“*Despite the importance of body size, considering only a single trait as a proxy for the niche might obscure clumpy patterns (D’Andrea et al. 2018)*”: I disagree with the word ‘obscure’. To me, using only a single trait, depending on which trait, only allow to discern a small deterministic part of how species are arranged in an ecosystem, but using too many dimensions may obscure the patterns because of numerical bias such as the influence of multicollinearity or the use of non-discriminant dimensions. Note that “*D’Andrea et al. 2018*” has been cited two times in the reference list at the end of the paper.

“[...] *by comparing species' functional differences in high resolution* [...]”: Please, clarify ‘in high resolution’

“[...] *because competition should be stronger within clumps rather than among clumps* [...]”: Please, add a reference to that statement.

4. My major comment (the most critical point to me) is about the clarity/level of detail of the analyses and on how the results are presented in their current form. In essence, the ‘Results’ section is too short without a required level of explanations.

Please consider providing more information on (1) the dataset used in this study (Are environmental parameters - and which - considered in the study?) and (2) how analyses have been performed to help readers interpret the results. A clearer guidance from the authors on how analyses were performed would help readers to better understand the paper and would allow researchers to reproduce analyses. I had some difficulties to understand the different steps of the methodology. This highly influence the understanding of the results and therefore of the ‘Discussion’ section. A map to locate the different sampling sites, and brief explanations on the biological data (only phytoplankton samples) would be useful.

“*Here, we measured species individual* [...]”: It is not clear to me if the measure were performed on all individuals or on only a fraction of the samples.

“*Functional traits were used to classify the species into MBFG*”: a table, in the main text, to resume the number of species per MBFG, site and season would be useful. This applies for the result: the authors did not mention enough the main/typical species that were affected into each (the three main) MBFG in order to make a better link between their results and the functioning of the study area.

“*The niche axis was divided into equally spaced fractions*”: what could be the consequence of this fractioning on the results?

“[...] *we applied linear regressions between species functional distinctiveness (FDist) and relative biovolume (RBiov) standardized to range between zero and one at clump-level*”. My apologies if I am wrong, but I cannot see the results from linear regressions. In the same way, I do not understand how was evaluated “*the relationship between RBiov and FDist at clump-level across river stretches and seasons*”.

The authors should explain better Figures 1 and 2. Figure 1 could be clarified and improved, for example by splitting the spatial representation from the temporal representation by means of two Figures instead of 5 panels; it could help to clarify the reading.

“*There were two significant (randomization test, $p < 0.05$, Figure 2) clumps* [...]”: how was determined the significance?

“*The first aggregation (clump I, 24 species) comprises species ranging from 9 to 10 in log2 volume* [...]”. Please, provide 2-3 species in order to establish a better link with ecology”.

“*The remaining results* [...]”. Please clarify what are the ‘remaining’ results.

“*The mean biovolume of species within clumps differed* [...]”. How was quantified/determined this difference?

“*According to the Mantel tests* [...]”. Results from the Mantel tests were not shown in the text.

Figure 2 is not exploited at all. No explanation was provided for the different components (panels). For example, how were obtained the factorial plans? What is the meaning of the size of the circles on the right panels and how were they calculated? How was calculated the significance of the clumps (results of the tests are not shown in the text)? It is not clear how were obtained the PCoAxes. What is the meaning of looking at “*the relative contribution of the six species with the highest biovolume*” while overlooking the species that contributed most to the multivariate analyses?

To resume, because results are not detailed enough and only partially presented, it is difficult to really understand how the conclusion is supported by the numerical analyses;

5. While interesting points are developed in the ‘Discussion’ section, I cannot see the link between the results and the conclusion, and why, for example, “*results does (do) not support alternative explanations as pure Neutrality*”; the same applies for several points highlighted in the Discussion, and the one I mentioned for the Abstract section (i.e. the relation with environmental conditions). For example:

- “*Present results showed that the clumps in body size are a conspicuous feature of phytoplankton community structure in riverine systems across seasons and river stretches*”

- “[...] *reflecting adaptations to survive in turbulent and mixed conditions*”. A parallel with species ecological traits must be done to better understand that statement.

- “[...] *Our results suggest a strong effect of stabilizing mechanisms at the overall community level as size differences increase*”. Could the authors explain better which analyses/results allowed us to conclude, this is not evident as results were not detailed enough.

- “*Studies evidencing EN using phytoplankton as the model community in coastal, estuarine, and here, riverine ecosystems presented fluctuations in environmental conditions (Segura et al. 2011, 2013), which is in agreement with recent modeling results (Sakavara et al. 2018)*”. Please clarify and document such as statement.

“*Despite larger species showing lower growth rates (Kruk et al. 2010), the elongated shape provides advantage under turbulent conditions (Reynolds et al. 1994) favoring these species under high flow conditions on nutrient rich environments (Irwin et al. 2006)*”. It could be very helpful, for a better understanding of the paper, to establish clearer link among phytoplankton species encountered in the study area, species traits, functional groups and environmental conditions. Note that parameters representing environmental conditions are available (“*A detailed description of the biological and physico-chemical sampling methodology [...]*”), but not clearly described nor related to phytoplankton communities.

Note that the following sentence is incomplete: “[...] *the HDH does not predicts any particular trait distribution and (Vergnon et al. 2009; Ingram et al. 2018).*”