Review Moury et al 2021 PCI EE

In the manuscript « The quasi-universality of nestedness in the structure of quantitative plant-parasite interactions", Moury and coauthors present an analysis of a compilation of plant-parasite interaction matrices and establish that the majority of them have a nested structure.

This is a timely and interesting manuscript which definitely deserves to be published. However, the production of this manuscript has involved a very high volume of work and the density of the information provided makes it a bit difficult to read and to follow. I suggest below ways of reorganising the manuscript and being more synthetic that should help having a better flow. I also ask some scientific and technical questions that should be answered in the manuscript.

1. As explicitly said at the end of the introduction, this paper has two goals: "*(i) to assess the performance of available algorithms to identify nested and modular patterns in matrices of quantitative data and (ii) to determine if these patterns are specific to each pathosystem or show a general trend".* This paper is thus a combination of two papers in a sense, one with a methodological question and the other with a scientific question. The choice to write a single paper is one of the reasons why this manuscript is so information-rich. All the material and methods and results corresponding to the first goal are presented in the Supplementary material but discussed in the main discussion (l392-418), whereas everything related to goal 2 (including a very long material description) is in the main text. I would suggest to balance and synthetize a bit more the presentation of the two goals, to have at least an outline of the Materials and Methods and Results of the methodological comparison of the different algorithm to test for nestedeness and modularity in the main text and to shorten the material section corresponding to the second goal (see next point).

2. The collection of plant-parasite quantitative interaction matrices analysed in this paper is composed of a mix of published and unpublished data. This results in a long Material section because for unpublished studies, technical details about the obtention of the matrix had to be given. I suggest to leave in the main text the information which is given for all matrices (number of plant accession, number of pathogen isolates, information on resistance/virulence QTL if available and pathogenicity variable(s) measured and given in the matrices) and move the technical details on how the matrix were obtained for unpublished studies to the supplemental material.

3. The criterion to build the dataset are not clearly stated: a criterion of minimal size of the matrix is given (l510-511) but there are other published datasets that would satisfy this criterion and are not included. How were the studies to include chosen? Did the authors perform a systematic literature search? Which key words were used?

4. The results of the methodological part (best algorithms and null models in terms of type I and type II error) are taken into account in a blurry way in the analysis of the dataset: the results are presented in the tables for all algorithms and null models (even the ones shown to perform poorly) but then in the written part of the results more weight is given to the results provided by the "good" combinations of algorithm and null model and a unique final number is given. This integration of the results of goal 1 in the analyses for goal 2 needs to be clarified.

5. The main result of the study is that nestedness is very common in plant-parasite interactions and that statistical interaction between plant accession and parasite isolate is usually not significant (result section + l425-427, l445-448). The authors derive this second conclusion from the fact that statistical models with only the two individual factors (without the interaction between them) have a high R^2. The author <u>should test directly</u> whether the

interaction is statistically significant or not; a model with single effects with an R^2 of 0.5 leaves "space" for a significant interaction.

6. For many pathosystems, there is more than one matrix in the dataset. It is not always clear whether the different matrices have been obtained by measuring different traits after the same inoculation events or whether they result from different inoculation experiments and in this second case, whether the plant accessions and the parasite isolates used are identical for all experiments/matrices.

7. In the cases where several matrices have been obtained on the same pathosystem, it would be great to have comments on the similarity of the results between the different matrices (as for example in l330-350).

8. L65-74: this paragraph of definitions does not seem necessary; the two meanings of "interaction" are well-known to readers of this kind of papers.

9. L96-133: the link between models of interaction and the modularity/nestedness has been discussed in other papers (de Vienne et al. 2009, Gallet et al. 2016). And this section of the introduction could be more synthetic.

10. L375-391: this paragraph would be better in the introduction.

11. L562: what do the authors mean by "differential lines"?

12. Supp mat, l39-40: explain why you could not apply certain algorithms to your data.

Minor comments:

- L91: to be rephrased
- L546 and 551: apple <u>tree</u> accessions.
- L758: similar results <u>as</u>
- Supp mat l10: weighted version <u>of</u>
- Supp mat l94: "or" has to be replaced by "of".