by Ines Alvarez, 2020-08-03 06:52
Manuscript: https://www.biorxiv.org/content/10.1101/2020.04.30.069948v3 version V3

last revisions before recommendation

Thanks to the authors.

I'm satisfied with their modifications and just ask them to make the minor modifications of reviewer 3 and to include the link to their dataverse website (https://data.inrae.fr/dataverse/TE-mobility-in-MiV3) in the main manuscript.

We are glad our revised version of the paper has met the criteria for recommendation in PCI Evolutionary Biology.

All the authors would like to thank the three reviewers and the recommender for the ensemble of comments and points raised which collectively allowed a substantial improvement of the manuscript clarity and significance.

We have addressed the minor comments from Reviewer 3 and have added a Data accessibility section with the following information:

“All the raw and filtered data generated in this study as well as details of the experimental procedures, scripts and datasets have been deposited and made publicly available in the institutional INRAE Data Portal at this URL: https://data.inrae.fr/dataverse/TE-mobility-in-MiV3 and cited throughout the text where appropriate, with DOIs available in the references.”

A message will be sent to the authors with the formatting instructions.

Best,

Inez Alvarez

Reviews

Reviewed by anonymous reviewer, 2020-07-13 12:21

I thank the authors for adressing all the raised points. The revised manuscript is ok for me.

We are glad the reviewer is satisfied by the efforts we made to address each point. We thank the reviewer for the constructive comments which helped improve the paper.

Reviewed by anonymous reviewer, 2020-08-02 19:23

The authors have thoroughly revised their manuscript and have done an excellent job of addressing all the reviewer concerns. The authors have also added new data on HGT likelihood

of various TE associated loci, which is very welcome. Both the introduction and discussion sections are substantially enhanced and provide a comprehensive review of the research field. I strongly support that the manuscript be officially recommended on PCI-EvolBio.

We would like to thank the reviewer for the constructive and encouraging comments and are glad the reviewer is satisfied by the efforts we made to address the different points raised.

One minor point: please include the link to their dataverse website (https://data.inrae.fr/dataverse/TE-mobility-in-MiV3) in the main manuscript, currently I could only see it in the response letter from the authors.

We have added a new 'Data availability' section in which we provided the URI of the institutional Dataverse in which we publicly deposited all the data and details of the experimental procedure. Each dataset is also cited in the text wherever appropriate with the corresponding DOI available in the reference list.

Congratulations on a rigorous piece of work.

Many thanks for the supportive and encouraging comments.

Reviewed by Daniel Vitales, 2020-07-17 17:37

I read with great interest this new version of the manuscript prepared by Kozlowski and collaborators. In my opinion, the study looks now clearer and more appealing after the revision. I would like to thank the authors for being so careful addressing my previous comments as well as those form the other reviewers. I think the manuscript could be recommendable at its current state, but I would also like to add some further suggestions that the authors might consider to incorporate in the final paper.

We are glad the reviewer did appreciate the efforts we deployed in addressing carefully each comment of the reviewers which collectively improved the quality of the paper.

In this new version of the manuscript, the authors provide a table summarising the whole (i.e. unfiltered) TE characterization of M. incognita (Table S1). This is a very interesting information allowing the comparison among whole repeatome annotation and the TE landscape represented by canonical (filtered) TEs. To me, the most unexpected result from this comparison is related to the abundance of Maverick elements, being by far the most abundant elements according to the whole repeatome characterization but showing much lower abundances when canonical elements are considered. I missed an explanation to these somewhat contradictory outputs and perhaps a discussion on how this could bias (or not) other results obtained.

We thank the reviewer for his comment. Considering the whole repeatome, Mavericks TE are indeed the most numerous annotations but these numbers dramatically decrease when canonical TEs are isolated from the rest of the repeatome. During the filtering process, most of

the Maverick annotations were removed either because i) they are small while canonical mavericks are between 5 and 20 kb long (Pritham et al., 2007) (median length = 364 bp while our minimum size criterion is 250 bp), ii) because they are too divergent from their consensus (median identity with consensus = 74.8%, which is the lowest value of both DNA and retro Transposons; and is way inferior to our inclusion threshold of 85%), and iii) because they might cover less than one third of their consensus. Other TE orders might present one of these characteristics but Maverick is the only order which combines all those exclusion parameters for most of its annotations. Our hypothesis is that many TEs annotations from the Maverick likely correspond to older copies which could be associated with remnants of a former TE activity. Another explanation could be a bias during the TE identification process (*de novo* step). However, working on other genomes from different phyla (not published yet), we observed different patterns with longer and more conserved Marverick copies. Hence, we think the signal we observe in the *M. incognita* genome is a biological reality.

The Table S5 is another very informative addition (I don't remember having seen it in the previous version) to the manuscript. There, we observe that LTRs, LINEs and Maverick are those elements showing a larger number of substantially expressed putative transposition machinery genes (i.e. suggesting TE activity). Conversely, according to TE-polymorphisms analyses (i.e the number of neoinsertions), MITEs and TIRs are reported as elements that "might have been more active in the genome of M. incognita than elements from other orders". To me, both results seem somewhat contradictory, so perhaps the authors could try to explain this better too.

MITEs are non autonomous elements, and thus, by definition, no genes coding for proteins involved in the transposition machinery are present. As explained in the results, finding no gene for transposition machinery in the non-autonomous TEs yet finding such genes in the autonomous TEs constituted a validation of the approach as well. Concerning TIRs, indeed it is surprising that while being among the most likely recently active TEs, based on the identity graph with their consensus, they are one of the autonomous orders with the less predicted transposition machinery in proportion. Thus, we checked how many TIR elements encompassed a predicted gene, whether it is TE related or not, given the criteria explained in methods (TE covering more than 95% of the gene length). Only 66 TIR annotations out of 3,595 encompass a predicted gene. As a comparison 83 LINE annotations out of 145 encompass a gene. This results is thus not incompatible with a global activity in TIR as almost half of TIR annotations with genes detected (30/66) have a putative transposition machinery. However this might point out some difficulties for gene call software to predict genes in TIR regions.

Finally, the criteria employed to select the 5 HCPTE loci validated by PCR were clearly explained in the point-by-point response letter to the reviewers. In my opinion, this explanation should also be included within the manuscript for a better interpretation of the results obtained. Without this information, the readers could understand that the validated "locus 1" being inserted in an expressed Meloidogyne-specific gene is a signal of adaptation that could be

extrapolated to the rest of HCPTEs. However, this locus 1 was specifically selected among the 22 HCPTEs for having those precise characteristics (e.g. being expressed in Morelos transcriptome and being Meloidogyne-specific). To me, explaining these criteria would make reading easier.

We agree with the reviewer's suggestion and now clearly explained the criteria used to select the 5 genes to be tested in the in methods section.

Other minor points:

L39. As a taxonomic name of a kingdom, I think that "Metazoa" should be written in capital letters.

This has been corrected, thanks for pointing this out.

L100, L105. "Arabidopsis" and "Drosophila" should be written in italics.

This has been corrected, thanks for pointing this out.

L168. The taxonomic name of the phylum should be "Nematoda".

This has been corrected, thanks for pointing this out.

L188. Regarding the differences of TE-content estimations among filtered and unfiltered approaches, here I would specify that "..almost two-thirds of the M. incognita canonical TE content".

We agree this is helpful to remind here that we are presenting filtered canonical TEs and not the raw unfiltered results. We thus added 'canonical'.

L235. I am not sure whether here should be cited Table S2 or Table S4.

Yes, we agree Table S4 would be more appropriate here, we modified the citation accordingly.

L311. The isolate from Morelos is named as "morelos" and Morelos" indistinctly along the manuscript. I think this should be unified.

We agree, and because this is the name of a state in Mexico, we used the version with a capital M everywhere.

L489. I think that "22 out of 33" should be better described as "majority" rather than "vast majority".

We agree, and changed to two-thirds.

Cited References

Pritham EJ, Putliwala T, Feschotte C. 2007. Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. *Gene* **390**:3–17. doi:10.1016/j.gene.2006.08.008