Copy number variations shape genomic structural diversity underpinning ecological adaptation in the wild tomato Solanum chilense

Sent
09-Feb-2024
From
EiC.MBE@gmail.com, EAssist.MBE@gmail.com
To
gasilvaa@unal.edu.co
CC
EiC.MBE@gmail.com, EAssist.MBE@gmail.com
Subject
Editorial Decision to Reject MBE-23-1089
Body
09-Feb-2024

MS: MBE-23-1089
Title: Copy number variations shape genomic structural diversity underpinning ecological adaptation in the wild tomato *Solanum chilense*

Dear Dr. Silva Arias:

Thank you for submitting your manuscript to Molecular Biology and Evolution (MBE). We regret to inform you that it did not receive high enough priority for publication after an in-depth review by the editors and the peer reviewers. Specific comments from the editors and/or external reviewers are included below.

Note that we might consider a substantially revised manuscript that carefully and thoroughly addresses the reviewer concerns.

In general, MBE seeks to publish research, methods, and resources of broad significance in molecular evolutionary biology. Even when the external reviewers find a manuscript to be scientifically and technically sound, the ultimate priority for publication is determined based on the novelty and impact of the work presented. MBE does not publish manuscripts judged by the reviewers to contain mostly descriptive work, confirmatory results, and discoveries with a limited gene and taxonomic scope. All of these factors were considered in deciding the publication priority for your manuscript.

Thank you for considering Molecular Biology and Evolution, and please continue to consider MBE as a venue for the publication of your best work.

Sincerely,

Board of Editors
Molecular Biology and Evolution

Editors' comments to the author:

The paper has been seen by two referees who have extensive comments and raise several issues. Because of the number of issues raised, the paper is not suitable for publication in its current form and will need extensive revisions.

Reviewer(s)' comments to author:
Reviewer: 1

Comments to the Author
This paper examines patterns of genome-wide Copy Number Variations (CNVs) in a collection of 35 wild tomato genotypes and identifies links between CNVs and potential environmental adaptations to more stressful habitats. The authors uncover some particularly intriguing copy number changes in genes involved in photosynthesis and flowering time, which they had not previously identified with SNPs alone.

While I did think there were some potentially interesting findings in this paper that could offer insight into the role of structural variants in adaptive evolution, I did have some questions/concerns about the methods being used to call the CNVs and to test for evidence of selection, which I detail below.

1. It was not clear from the beginning of the paper that the results were all based on short-read sequencing data; in fact, I didn't realize this until I got to the first paragraph of the Materials and Methods section at line 458. Since using short-read data (instead of whole genomes or long reads) can very much make a difference in terms of what CNVs can be identified, I think this needs to be set out clearly from the beginning.
2. While the authors do state that the data were processed in the same way for a previously published study (line 461), I would still like to see some details here related to average read depth statistics, since again this can affect structural variant calling.
3. In the section called 'Identification and genotyping of CNVs' the authors describe using several SV calling tools (LUMPY, Wham, DELLY, and Manta), but I do not see any information describing what parameters were used for each tool. Also, it seems the final CNV set was generated by merging the results of all of the tools, but does this mean that all of the SVs called by each tool were merged into one inclusive data set? Or were only the SVs called by all tools (or more than one tool) kept? Was there any filtering done on the called SVs? The way that this paragraph reads suggests that the raw SV calls from each tool were all saved and put into SVtyper, which, if true, would cause me some concern about false positives.
4. In the same section, 2nd paragraph, the authors describe a simulation pipeline to help validate their SV calls, but I did not see the results of this test anywhere (maybe I missed them).
5. In the section called 'Quantification of gene copy number,' a completely different set of methods is described for identifying CNVs. I don't understand why one pipeline was used to find CNVs in genes, while the previous pipeline was used to find all CNVs. Wouldn't the gene CNV results be contained within the LUMPY/DELLY/Manta results from before? And if the CNV calls do not overlap, isn't that an issue?
6. I am not very familiar with the VST statistic used to find CNVs with signatures of divergence, but if it is very similar to FST, then it could be prone to some of the issues that FST has in low diversity regions. Namely, that FST can have a tendency to become inflated when diversity is low. This could be of particular concern in this system, as the authors do mention that some populations underwent expansion/colonization events which would result in lower diversity in those populations. The usual way to deal with this is to use Dxy along with FST, so I would recommend the authors use that statistic as well.

7. I was a bit confused by exactly how the CAFÉ analysis of gene expansions and contractions was done. In my experience with CAFÉ, it is used to identify gene families that have undergone expansion/contraction, but the way the authors described the results it sounded like they were talking about rates of expansion/contraction of particular gene copies. I think this needs to be clarified in the sections of the manuscript where the expansion/contraction rates are described. Also, when I have used CAFÉ, I thought it required an orthology analysis to group genes into families, so those methods should be described here as well.
8. For the association analyses between CNVs and climate variables, my understanding was that only the highly differentiated CNVs were used. However, I think a more robust analysis would be to look at associations between climate variables and ALL CNVs, and then see how many significant associations overlapped with the highly differentiated CNVs. This would help validate the VST outlier analysis results and could find additional associations between CNVs and environmental variables.

Reviewer: 2

Comments to the Author
Please find my review, as well as annotated PDFs of the manuscript, supplementary text, as well as supplementary figures and tables attached.