PCIEvolBiol #525

Genomic evidence of paternal genome elimination in globular springtails
Kamil S. Jaron, Christina N. Hodson, Jacintha Ellers, Stuart JE Baird, Laura Ross
https://doi.org/10.1101/2021.11.12.468426

Overview

In this study, the authors analysed re-sequencing genomic data in order to investigate an established hypothesis that Paternal Genome Elimination occur in globular springtails. I believe their reasoning is sound, the results interesting, the study convincing and the title accurate. My criticism are only towards the clarity of the different steps taken by the authors to reach their conclusion and the length of the main text. This study is of good quality, however the formating of its manuscript should be improved.

The interest of some analyses is unclear to me, the limits of some interpretation are not explicitely stated and the potentiality of competing interpretation feel sometimes overlooked. I believe this study could be thoroughly stream-lined, by only keeping the necessary and sufficient results in order to build the argumentation towards the likely existence of PGE in globular springtails. This would i) ease the reading, ii) shorten the manuscript, and iii) and create space to discuss the limits and competing hypotheses possibly explaning the results presented. I do believe these limits do not hamper the authors to reach their conclusions, and that their final hypothesis is the most likely to date. But their conclusion will be even stronger after having taken the time to explain how potential issues are unlikely to impact the results or that they do not hamper to reach the author's conlusions. Also it would be the opportunity to clarify precisely what conclusion can be drawn by what analyses. Taken alltogether, I agree with the interpretation of the results, but the process to get to this conclusion felt convoluted, sometimes spending time on trivial aspects and unecessary figures, sometimes not discussing or investigating potential issues.

Note that this reviewer is not an expert in these specific reproductive modes (PGE) and their evolutionary impact on organisms. These aspects of study were thus not « properly » reviewed here, and I would hold with my overall positive review as long as these aspects are considered correctly tackled by other reviewers (e.g. the evolution of PGE in arthropods being correctly depicted).

Mandatory revisions

line 109 : inappropriate figure reference : SM Figure 1 do not really show the elimination of one of the X1 and X2 chromosomes during early embryogenesis. Or it does but I didn't understand how. In both case, the authors need to modify this figure in order for the reader/reviewer to understand how it shows elimination of X1 and X2 chromosomes. Maybe I simply did not find the « Supplementary Figure 1 » ? In any case, this needs to be fixed/clarified.

Line 139-141 : This study does not « demonstrate » uniparental inheritance, it strongly suggests it. This is correctly stated in the abstract, but the sentence here is incorrect. Also, using the wording « co-segregate » is unclear (at least for me) regarding whether the

segregation of X chromosomes and autosome co-occur (i.e. same timing) or is done in two different steps. This should be clarified here.

Line 146 : The figure 2 title should clearly establish that this is a hypothetical model, a working hypothesis. As of now, it might be confused by readers as a conclusion of this study (thus appearing way too soon in the manuscript), or worse, as a state-of-the-art introductory figure (which would render the entire study unnecessary, since PGE would already have been shown). Maybe a title like « Working hypothesis for PGE in globular springtails » ? Another solution (my favorite) would be to place this figure at the end of the manuscript, after PGE is strongly suggested by the results, as a scheme of the new, up-to-date, working hypothesis that the reader will remember.

Line 194-198 : Instead of showing difficult-to-interpretate k-mer spectra due to unevenly spaced ploidy peaks, the authors shoud first (or only) show the distribution of mapping coverage on scaffolds (SM Fig 2 panels B, D and F). This would make understanding the results easier for the readers. Note that If they wish so, the authors they can later show the corresponding k-mer spectra. But these spectra do not add any value to this study, as what makes them more informative (i.e. estimating genome size and heterozygosity) is useless in the context of this manuscript. Consequently, this also diminishes the interest of the Box 1 in the main text. This box is not important to understand the study, and should be either removed or placed within supplementary text if the authors strongly want to keep it in the manuscript.

Line 279: Figure 4 is referred to before Figure 3 (line 342). These two figures should thus be switched in the manuscript. See also my remark on placing Figure 2 at the end of the manuscript.

The estimation of the fraction of sperm cells heavily rely on the accuracy of the peak of the allele frequency distribution shown in SM Figure 5. However, this distribution is very flat (thus, the exact position of the peak is uncertain). While trying to estimate the fraction of sperm cells based on these data is very interesting, the authors should explicitely mention the inherent inaccuracy of these estimate every time they mention it in the manuscript. This does not seem to be a strong results of the study and should not get the quantity of attention it currently gets from the authors. Also, for clarity and for convincing the readers that their method is sound, they should also compute and present the corresponding fraction of sperm cells in Ocin2, as its peak is also not exactly positionned at 0.5. (of course, it is impossible to placed at exactly 0.5). Finally, they should also try their approach on an organism for which this fraction is known. As long as this is not done, I would definitely tone down theses results in the manuscript for two reasons : i) their accuracy is debatable and, more importantly, ii) they do not really help answering whether or not PGE occurs in globular springtails and are somewhat a distraction from the main argumentation line of this manuscript. Additionaly, authors should clearly indicate the respective sequencing depth of the data used for BH3-2 and Ocin2 in the main text and discuss it. Indeed, a much larger coverage in Ocin2 might be sufficient to produce a minor allele frequency distribution closer to 0.5, that is, with more accurate SNP-calling. Alternatively, the order of magnitude higher number of heterozygous variants in Ocin2 might be sufficient to produce a much clearer peak, potentially mechanisticaly closer to 0.5 (i.e. even if every single SNP is called with the same accuracy as for BH3-2). Overall, I believe the estimate of the fraction of sperm cells need to be taken with caution, and that the authors should discuss about the limitation of their otherwise interesting approach.

Authors repeatedly uses PGE as the explanation of a coverage shifts. While this is a possible explanation, other types of GE, that are non-paternal genome elimination (or even simply different ploidy levels within an organism) could also produce such a shift. However, the results presented in figure 4C clearly and convincingly suggest PGE in *Allacma fusca* (or at least in indvidual BH3-2). I thus suggest the authors to clarify their argumentation line, decomposing it in two steps : i) analysis of coverage data + interpretation of the peak being shifted allow to confirm various ploidy levels in the tissue (interpreted as due to genomic elimination, Figure 3) ; ii) coverage distribution of maternal allele fitting the distribution of « major » autosomal allele strongly suggests that the eliminated chromosome are, in fact, of paternal origin. I believe this would clarify the argumentation line of the study, explicitely stating the relative contribution of the different analyses to the main conclusion. This would also help to stream-line the writing.

The authors should explicitely state whether their analyses bring light (or not, which would not be an issue) on the tempo of PGE : with their data, can they re-inforce or contradict the hypothesis that chromosome elimination happens in one step, or in two steps (elimination of X chromosome in early spermatogenesis, and elimination of chromosomes later during the process) ? I believe the answer is no : these data can not validate or invalidate this hypothesis, but this should be discussed in the manuscript, in order to clarify to the reader the state-of-the-art knowledge on reproductive mode in globular springtails.

The last paragraph of the discussion section should be removed, or placed elsewhere (maybe as supplementary text, or at the beginning of the discussion section). I personally do not think that this study « demonstrate the power of a careful bioinformatics analysis » : this is vague and emphatic. I do believe however that discussing the importance of quality checks and data exploration is interesting for the readers. This should however not be placed as the final paragraph of this study about reproductive mode of globular springtails.

English spelling and typos need to thoroughly checked throughout the manuscript.

Suggestions, corrections and « typos »

line 35 : the current sentence finishing the first paragraph is not particularly convincing, it feels a bit « off » the rest of the paragraph. I suggesting re-phrasing it.

Table S1, row 11 : « ~~confromation~~ » => « confirmation »

line 113 : « ~~The X chromosome lacking spermatocytes~~ » => «The spermatocyte lacking the X chromosomes »

line 138 : « in-silico bioinformatics » is redundant, pick your favorite.

Line 138 : « to separate the effect of somatic and germline genomes ». What effect are your referring to ? Please re-phrase.

Line 139 : Please refrain to use wordings such as « innovative ». It is close to meaningless and feels emphatic. Readers will decide in the future wether this study is « innovative » or not.

Line 165 : The authors need to explicitely specify whether DNA amplification step was used in this study during the production of cDNA libraries. Indeed, this would have an importance for later analyses of allelic frequency and the determination of maternal versus paternal chromosomes.

Line 171 : « ~~the bistates~~ » => « biallelic SNPs »

Line 187 : « monoploid ». Either the authors have a reason to not use the word « haploid » and should justify their choice within the text, or they should use « haploid ».

SM Figure 3 title : typo « allelie »

Line 214 : SM Figure 9 caption refers to SM Figure 9.

Line 472 : re-phrase the entire sentence, as the paragraph specifically explain that PGE is NOT the only explanation compatible with biology.

It is my opinion that the overall inclusion of misassigned variants (results and discussion) takes too much space in the current manuscript. While it is good the authors checked it, it did not occur to me when reading the study that it could have had a strong impact on the results. I suggest the authors mention this check in the manuscript, but reduces its importance.