



Pressing NGS data through the mill of Kmer spectra and allelic coverage ratios in order to scan reproductive modes in non-model species

Nicolas Bierne

► To cite this version:

Nicolas Bierne. Pressing NGS data through the mill of Kmer spectra and allelic coverage ratios in order to scan reproductive modes in non-model species. 2021, pp.100142. 10.24072/pci.evolbiol.100142 . hal-03715291

HAL Id: hal-03715291

<https://hal.science/hal-03715291>

Submitted on 6 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Peer Community In Evolutionary Biology

Pressing NGS data through the mill of Kmer spectra and allelic coverage ratios in order to scan reproductive modes in non-model species

Nicolas Bierne based on reviews by Paul Simion and 2 anonymous reviewers
A recommendation of:

Open Access

Genomic evidence of paternal genome elimination in the globular springtail *Allacma fusca*

Kamil S. Jaron, Christina N. Hodson, Jacintha Eilers, Stuart JE Baird, Laura Ross (2022), *bioRxiv*, 2021.11.12.468426, ver. 5 peer-reviewed and recommended by Peer Community in Evolutionary Biology <https://doi.org/10.1101/2021.11.12.468426>

Published: 2022-07-01

Copyright: This work is licensed under the Creative Commons Attribution-NoDerivatives 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nd/4.0/>

Data used for results

<https://www.ncbi.nlm.nih.gov/bioproject/PRJEB44694/>

Scripts used to obtain or analyze results

<http://doi.org/10.5281/zenodo.6645407>

Abstract

Submitted: 18 November 2021, Recommended: 29 June 2022

Cite this recommendation as:

Nicolas Bierne (2022) Pressing NGS data through the mill of Kmer spectra and allelic coverage ratios in order to scan reproductive modes in non-model species. *Peer Community in Evolutionary Biology*, 100142. <https://doi.org/10.24072/pci.evolbiol.100142>

Recommendation

The genomic revolution has given us access to inexpensive genetic data for any species. Simultaneously we have lost the ability to easily identify chimerism in samples or some unusual deviations from standard Mendelian genetics. Methods have been developed to identify sex chromosomes, characterise the ploidy, or understand the exact form of parthenogenesis from genomic data. However, we rarely consider that the tissues we extract DNA from could be a mixture of cells with different genotypes or karyotypes. This can nonetheless happen for a variety of (fascinating) reasons such as somatic chromosome elimination, transmissible cancer, or parental genome elimination. Without a dedicated analysis, it is very easy to miss it.

In this preprint, Jaron et al. (2022) used an ingenious analysis of whole individual NGS data to test the hypothesis of paternal genome elimination in the globular springtail *Allacma fusca*. The authors suspected that a high fraction of the whole body of males is made of sperm in this species and if this species undergoes paternal genome elimination, we would expect that sperm would only contain maternally inherited chromosomes. Given the reference genome was highly fragmented, they developed a two-tissue model to analyse Kmer spectra and obtained confirmation that around one-third of the tissue was sperm in males. This allowed them to test whether coverage patterns were consistent with the species exhibiting paternal genome elimination. They combined their estimation of the fraction of haploid tissue with allele coverages in autosomes and the X chromosome to obtain support for a bias toward one parental allele, suggesting that all sperm carries the same parental haplotype. It could be the maternal or the paternal alleles, but paternal genome elimination is most compatible with the known biology of Arthropods. SNP calling was used to confirm conclusions based on the analysis of the raw pileups.

I found this study to be a good example of how a clever analysis of Kmer spectra and allele coverages can provide information about unusual modes of reproduction in a species, even though it does not have a well-assembled genome yet. As advocated by the authors, routine inspection of Kmer spectra and allelic read-count distributions should be included in the best practice of NGS data analysis. They provide the method to identify paternal genome elimination but also the way to develop similar methods to detect another kind of genetic chimerism in the avalanche of sequence data produced nowadays.

References

Jaron KS, Hodson CN, Ellers J, Baird SJ, Ross L (2022) Genomic evidence of paternal genome elimination in the globular springtail *Allacma fusca*. bioRxiv, 2021.11.12.468426, ver. 5 peer-reviewed and recommended by Peer Community in Evolutionary Biology. <https://doi.org/10.1101/2021.11.12.468426>

Reviews

Toggle reviews

Evaluation round #2

DOI or URL of the preprint: <https://www.biorxiv.org/content/10.1101/2021.11.12.468426v2>

Version of the preprint: 2

Author's Reply, 27 Jun 2022

Thank you for suggestions, we have updated the manuscript so it comply with PCI standards.

Decision by **Nicolas Bierne**, 09 Jun 2022

Dear Dr Jaron,

Sorry for the delay to process this resubmission. I had problem with emails going in junk mails followed by one referee being off line for a long period, and then me. Referees are happy with the revisions you made and I am pleased to accept your preprint for a recommendation in PCI Evol Biol. I'm working on the recommendation text now. Meanwhile, could you correct the typos identified by referee#2 and resubmit again please?

Thank you for your support of the PCI initiative, and sorry again for the delay.

Best regards,

Nicolas Bierne

*Reviewed by **Paul Simion**, 17 May 2022*

PCIEvolBiol #525

Review #2

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>

After reviewing the changes made to the manuscript, i believe the authors answered all the important comments previously made, notably by clarifying several points, the overall argumentation line, and by adding an interesting Supp Figure 10. it is my opinion that the manuscript is thus ready to be recommended by peers.

My own english level is however not good enough to ascertain whether the writing needs further improvements or not.

Reviewed by anonymous reviewer, 12 May 2022

I am happy with the revision of the authors.

Just a few minor corrections to the new SM text 5:

L 149: Add a ":" at the end of the line.

L 153: Fix "catinated".

L 186: "biassed" -> biased

Evaluation round #1

DOI or URL of the preprint: <https://doi.org/10.1101/2021.11.12.468426>

Version of the preprint: 1

Author's Reply, None

[Download author's reply](#)

*Decision by **Nicolas Bierne**, 09 Jun 2022*

Dear Dr Jaron,

We have received three thoughtful reviews of your manuscript entitled "Genomic evidence of paternal genome elimination in globular springtails". The three referees are globally positive, as well as I was when I accepted to handle this preprint for a recommendation. As it is current practice, I'll ask you to revise your ms according to referees' concerns and provide a cover letter where you explain how you modified the ms. Referee 1 and 3 have only a few minor concerns to improve clarity that you should easily account for. Referee 2 who signed his review has a longest list of comments. This referee will likely see the revised version. I am looking forward to reading your revised ms and to work on a nice recommendation to advertise your interesting work.

Best regards,

Nicolas Bierne

Reviewed by anonymous reviewer, 07 Jan 2022

This article presents a new genomic approach for detecting an interesting type of reproduction: paternal genome elimination. In this reproductive mode, males inherit paternal and maternal genomes but only maternal genetic material is present in their spermatozooids. To detect such a reproductive mode, the authors suggest to check whether X-linked alleles show similar coverage than the major autosomal allele, which would be indicative of an excess of maternally inherited genetic material in sperm.

While I do not have the expertise to judge the mathematical aspects of the method, the approach is nicely presented with clear figures explaining the rationale, and the main result showing that *Allacma fusca* males practice paternal genome elimination is convincing. As authors acknowledge it, this approach is only possible in males with large quantity of sperm. While they discuss accordingly the fact that many invertebrates feature high proportion of male germ cells, I would like to know if they can estimate the fraction of sperm necessary to significantly detect paternal genome elimination depending on genome coverage. I think it would be a nice addition to help future scan studies using this kind of approach on several other species. Also, I suggest to edit the title to be more specific about the fact that the main result concerns only one species of globular springtails.

Minor remarks:

L40: fix nested parentheses of citation

L50: I do not think that the verb "culture" can be used for insects

L53: citations should be placed before the comma

L98: add a comma after "In fungus gnats and gall midges"

L272: contains

L467: change "seems to be present already" on "seems to be already present"

L493: fix nested parentheses of citation

L513: fix nested parentheses of citation

Fig1: "heterochmatized" does not seem to be an existing word

Reviewed by [Paul Simion](#), 16 Dec 2021

[Download the review](#)

Reviewed by anonymous reviewer, 13 Dec 2021

In this paper, Jaron et al. test for systematic paternal genome elimination in a species of globular springtails (*Allacma fusca*) by developing computational approaches to disentangle germline and somatic genome contributions to a whole-body sample of gDNA. Specifically, the authors develop a two-tissue mixture model (soma and cells having undergone paternal genome elimination, here, secondary spermatocytes and mature sperms) to estimate the respective contribution of these tissues to the whole-body sample of gDNA and, hence, the sequencing library. This allows them to formally test for systematic paternal genome elimination based on maternal and paternal sequencing coverages. They also validate their approach with a control species (*Orchesella cincta*) that has XO sex determination. Overall, this manuscript is well written, the methods are clear, very well described and most likely reproducible. I only have a few minor comments that mostly involve clarifications.

- I found the wording "aberrant spermatogenesis" (L98) a bit confusing, as paternal genome elimination appears to be systematic. I understand this wording is a legacy from earlier publications, but I wondered if it could be reformulated at this point, given the evidence provided by the authors.

- A minor caveat to the approach developed here is that it requires a reference genome assembly. This should be mentioned somewhere.

ABSTRACT

L21-23: “The genomic approach we developed allows for detection of genotypic differences between germline and soma in all species with sufficiently high fraction of germline in their bodies.”

- This statement is vague (what is a sufficiently high fraction?). Might it be possible to evaluate what is a sufficiently high fraction either by modelling or sequencing read resampling? If not (or beyond the scope of the present study), I suggest rephrasing.

- This statement is also potentially more generalizable (“high fraction of germline in their bodies”) to include a tissue sample (e.g., gonads) containing a mixture of somatic and germline cell (and not only whole-body samples) in larger organisms (e.g., birds).

INTRODUCTION

L59: I would remove the comma.

Figure 1: Heterochmatized --> heterochromatinized? (or heterochromatinised)

DISCUSSION

L464: I wondered if “hybridogenesis” should not be mentioned as well for the Australian carp gudgeons.

L505: How could PGE affect the evolution of reproductive isolation and increase diversification rate? Could it be related to the lack of recombination in males and strict maternal inheritance to reduced effective population size? I would clarify the rationale here.

SUPPLEMENTARY MATERIAL

L129: I think this sentence is incomplete :)

Github repository: some parts could be cleaned up a bit, but everything seems to be there.