Peer Community In
Zoology

Natural history collections continue to inform ground beetle genetics.

Felix Sperling based on peer reviews by Michael Caterino, Julian Dupuis and 1 anonymous reviewer

Marie T. PAULI, Jeremy GAUTHIER, Marjorie LABEDAN, Mickael BLANC, Julia BILAT, Emmanuel F.A. TOUSSAINT (2024) Museomics of *Carabus* giant ground beetles shows an Oligocene origin and *in situ* Alpine diversification. bioRxiv, ver. 5, peer-reviewed and recommended by Peer Community in Zoology.
`https://doi.org/10.1101/2024.03.21.586057`

Submitted: 29 March 2024, Recommended: 19 July 2024

Cite this recommendation as:
Sperling, F. (2024) Natural history collections continue to inform ground beetle genetics.. *Peer Community in Zoology*, 100319. 10.24072/pci.zool.100319

Published: 19 July 2024

Some of the biodiversity of our planet now exists only in museums, due to continuing habitat destruction and climate change. With more than 380 million entomological specimens already preserved in museums (Johnson and Owens 2023), there is much work left to document what we already have. Fortunately, new advances in DNA sequencing have given us the opportunity to get enormous amounts of information from dried specimens on pins. One such advance is HyRAD-X, which uses RAD-derived probes originally developed using RNA extracted from a selection of specimens with high RNA-integrity (Schmid et al. 2017). These exome-limited probes can then be used to capture low-integrity DNA extracted from a single leg from museum specimens, followed by Illumina sequencing of the enriched libraries. Ground beetles allow an excellent demonstration of this approach, as their diversity, large size, and charismatic appearance has led to them being well represented in museums. Using a HyRAD-X probe set previously developed for a higher phylogeny within the subfamily Carabinae (Toussaint et al. 2021), the authors have now applied the same probe set to produce a comprehensive phylogeny for Arcifera, a clade of four subgenera and ten species within the genus Carabus (Pauli et al. 2024). Of the 96 specimens that they started out with, 90% were from natural history collections and 40% dropped out immediately due to poor DNA extraction yield. After filtering the resulting sequence reads for minimum coverage and minimum number of samples per locus, they ended up with 35 museum specimens with an average of 793 loci. Phylogenetic analysis of this data supported the current classification of these beetles. Pauli et al's. (2024) study has effectively shown the power of HyRAD-X methods for applications at the species level. In-house production of probes makes the method accessible, expanding the opportunity to use museum specimens for population genetic research.

References

Johnson KR, Owens, (IFP. 2023) A global approach for natural history museum collections. Science 379,1192-1194(2023). https://doi.org/10.1126/science.adf6434

Pauli MT, Gauthier J, Labédan M, Blanc M, Bilat J, Toussaint EFA (2024) Museomics of Carabus giant ground beetles shows an Oligocene origin and in situ alpine diversification. bioRxiv, ver. 5 peer-reviewed and recommended by Peer Community in Zoology. https://doi.org/10.1101/2024.03.21.586057

Schmid, S., Genevest, R., Gobet, E., Suchan, T., Sperisen, C., Tinner, W. and Alvarez, N. (2017), HyRAD-X, a versatile method combining exome capture and RAD sequencing to extract genomic information from ancient DNA. Methods Ecol Evol, 8: 1374-1388. https://doi.org/10.1111/2041-210X.12785

Toussaint EFA, Gauthier J, Bilat J, Gillett CPDT, Gough HM, Lundkvist H, Blanc M, Muñoz-Ramírez CP, Alvarez N (2021) HyRAD-X Exome Capture Museomics Unravels Giant Ground Beetle Evolution, Genome Biology and Evolution, Volume 13, Issue 7, evab112, https://doi.org/10.1093/gbe/evab112

Reviews

Evaluation round #2

DOI or URL of the preprint: https://doi.org/10.1101/2024.03.21.586057
Version of the preprint: 2

Authors' reply, 26 June 2024

June 26, 2024

Dear recommender, Prof Felix Sperling,

My coauthors and I would like to thank you and the reviewers for your constructive comments and for inviting us to resubmit a revised version of our manuscript. We have found your last comments helpful and have addressed all of them. We include point-by-point responses at the end of this letter with our replies and corresponding line numbers for the revision. We believe that our manuscript is much clearer and hope is now ready for publication in PCI Zoology.

Revision round #2
Decision for round #2 : Revision needed
One more round needed.

This version is better, with good responses to reviewers, but the ms still has editorial issues that necessitate another round of revisions by the authors. I attach the Word version of the main text file, which contains:

1. A few remaining glitches that I have edited in the Word file.

In the new biorxiv version the Figures have been uploaded in jpg format avoiding this issue.

2. Some suggested editorial revisions intended to make some of the phrasing more grammatically smooth or easy to understand (these should all be checked to ensure I haven't altered meaning).

All these suggestions have been integrated.

3. A couple of points in text that refer to other's work/opinions and need citations.

New citations have been included.

4. Noted missing caption for new Suppl. Fig. 5, which means adjusting numbers as well.

Corrected.

5. Noted that Fig. 2. has a number of problems with offset text and symbol blocks.

Corrected.

6. The references list needs some substantial cleaning up to ensure that format and information is consistent. Also please check that all refs are cited in text and vice versa.

Corrected.

Since the remaining items are editorial, I don't think this needs to go our to a second round of review.

Download recommender's annotations

by Felix Sperling, 17 Jun 2024 07:01

Manuscript: `https://doi.org/10.1101/2024.03.21.586057`

version: 2

Decision by **Felix Sperling**, posted 17 June 2024, validated 17 June 2024

One more round needed.

This version is better, with good responses to reviewers, but the ms still has editorial issues that necessitate another round of revisions by the authors. I attach the Word version of the main text file, which contains:

1. A few remaining glitches that I have edited in the Word file.
2. Some suggested editorial revisions intended to make some of the phrasing more grammatically smooth or easy to understand (these should all be checked to ensure I haven't altered meaning).
3. A couple of points in text that refer to other's work/opinions and need citations.
4. Noted missing caption for new Suppl. Fig. 5, which means adjusting numbers as well.
5. Noted that Fig. 2. has a number of problems with offset text and symbol blocks.
6. The references list needs some substantial cleaning up to ensure that format and information is consistent. Also please check that all refs are cited in text and vice versa.

Since the remaining items are editorial, I don't think this needs to go our to a second round of review. **Download recommender's annotations**

Evaluation round #1

DOI or URL of the preprint: `https://doi.org/10.1101/2024.03.21.586057`

Version of the preprint: 1

Authors' reply, 10 June 2024

Download author's reply

Download tracked changes file

Decision by **Felix Sperling**, posted 17 May 2024, validated 21 May 2024

This manuscript has recieved three very positive reviews. All of the reviews raise some editorial points, and two of them request more detailed information and discussion on the lab methods. I agree that it would help to flesh out the methods at least a bit more. Please revise your manuscript accordingly to take advantage of these supportive reviews, responding to (or refuting) each of the points raised.

Reviewed by Michael Caterino, 22 April 2024

Title and abstract

Does the title clearly reflect the content of the article? [**X**] Yes, [ ] No (please explain), [ ] I don't know

Does the abstract present the main findings of the study? [**X**] Yes, [ ] No (please explain), [ ] I don't know

Introduction

Are the research questions/hypotheses/predictions clearly presented? [**X**] Yes, [ ] No (please explain), [ ] I don't know

Does the introduction build on relevant research in the field? [**X**] Yes, [ ] No (please explain), [ ] I don't know

Materials and methods

Are the methods and analyses sufficiently detailed to allow replication by other researchers? [**X – with some commentary on exclusion criteria**] Yes, [ ] No (please explain), [ ] I don't know

Are the methods and statistical analyses appropriate and well described? [**X**] Yes, [ ] No (please explain), [ ] I don't know

Results

In the case of negative results, is there a statistical power analysis (or an adequate Bayesian analysis or equivalence testing)? [ ] Yes, [ ] No (please explain), [ ] I don't know, [**X**] N/A.

Are the results described and interpreted correctly? [**X**] Yes, [ ] No (please explain), [ ] I don't know

Discussion

Have the authors appropriately emphasized the strengths and limitations of their study/theory/methods/argument? [**X**] Yes, [ ] No (please explain), [ ] I don't know

Are the conclusions adequately supported by the results (without overstating the implications of the findings)? [**X**] Yes, [ ] No (please explain), [ ] I don't know

This is an impressive phylogenomic study of a complex lineage of ground beetles, while also serving as a powerful demonstration of the applicability of museomic techniques to such questions. The amount of data acquired from historic specimens was surprising to me (not having explored this area nearly as much as I would like), but the trees support that the vast majority of it is reliable and informative.

With regard to the criteria for excluding certain samples, these aren't too clear early on. Apparently there was nothing detected for the samples initially excluded, rather than simply very low concentrations. (As an aside, it would be interesting to see if any markers could be amplified from those extracts. Given the higher than expected success rates overall, it seems not out of the question that below detectable quantities may still be amplifiable.) Exclusion of some lower concentration samples that produced some data might merit a little more discussion. 12 samples with <150 loci were left out. But it would be interesting to see if these could be placed on data sets limited to a smaller number of shared loci. I'm just interested to see what information can be squeezed out of these old specimens (thinking 100YO types and such) when necessary.

Otherwise, there is a wealth, perhaps an overabundance of discussion of the fine points of Carabus relationships. Only the hardcore fans will find this useful, while the general readers interested more in museomics will begin to wonder whether to continue reading. In many other groups this might be problematic. I suppose in a paper dealing with such a popular genus, the committed readership will be somewhat higher than for many other taxa.

I made a number of other minor corrections and comments in the pdf. But in general this is a very readable and clean manuscript, especially for a pre-print. The figures are attractive, clear, and sufficient to convey the results.

Reviewed by Julian Dupuis, 08 May 2024

Here, the authors use HyRAD-X on a combination of museum and contemporary specimens to generate a large phylogenomic dataset to assess diversification of a genus of ground beetles. This approach allowed them to generate quite a bit of data from a broad assemblage of diverse specimens, which they then used for a whole bunch of phylogenetic and population genetic analyses. From these, they focused discission on both taxonomic considerations within this genus as well as diversification/speciation dynamics. I am no carabid expert, so can't judge that aspect of the study, however from a very superficial review of pertinent Carabus literature, the authors have done a fine job in addressing outstanding taxonomic questions at both the subgeneric and more microevolutionary levels.

Overall, the study is well-written, most analyses are conducted and presented appropriately, and the system-specific discussion is well-presented and as approachable as such system-specific discussions can be for non-experts in that group. From a phylogenomic perspective, bioinformatic pipelines (including code, input files, etc.), analyses, and interpretations are solid. My only bigger-picture critique of the study is that I think more could be done to explore why some samples work and some don't work in this museomics context. As the authors point out, a lot of this is unknowable (specimen collection and preservation conditions), but the authors document that a FA was used on all raw DNA extracts, so I wonder if there are any trends with regard to fragment size distributions that might inform more interpretation of the results presented in Figure 1. Might turn out to just be too variable to make any conclusions there, but I'd say worth a bit of added discussion to the first paragraph of that section of the Discussion. As someone who is exploring these methods in depth, that kind of methodologically important information is quite valuable, even if it's somewhat conjecture or the author's best guess.

Specific comments provided below. I think this study will be a valuable addition to both 1) ground beetle researchers and 2) systematists exploring this new field of museomics. With museomics being such a young field, studies such as these have a great opportunity to add extremely valuable context to the discipline moving forward.

Lines 79-81. Grammar a bit clunky—would rephrase.

Line 92. Not sure what is meant by "inferences made in fine"?

Line 116. Should the abbreviated form of the subgenus name be used here, as it is used down in the results/discussion? i.e., "C. (Chaetocearabus)"? Formatting of subgenera goes back and forth throughout the manuscript it seems. Personally, I like the abbreviated form, since they're all Carabus subgenera.

Lines 119-120. Reword to "where hybrids are found, for example, at the Katara Pass…"

Lines 193-194. Although I think I generally understand what is being said here, I would suggest rephrasing this sentence. "Processed" could be used to mean several diderent things in this context.

Line 205-208. Thank you for this transparency! I completely agree. I would rephrase slightly as "Early sampling erosion and discarded samples are commonly not…".

Section at line 216. Was the same RNAseq-based probe kit used as in Toussaint et al. 2021? It is unclear from this section here, and that seems to be a pivotal step in library prep for this method.

Line 247-249. Commas should be used in the # loci here. Also maybe refer to Table 1 here, so it shows up near this part of the methods?

Line 288. Should be "SNPs"

Line 331. Although barplots and likelihood values are provided in the supplement, I'd also

like to see the deltaK values (especially since that method was used to determine K, as in the methods) and would like to see the likelihood values plotted so they're more interpretable to a reader. Ultimately, following some of the best practices for STRUCTURE (10.1111/mec.14187), seeing the various statistics developed by Puechmaille might help in determining the best value of K, but given the relatively low (and uneven) sample sizes, I don't expect that they'd be terribly informative beyond what is presented here. But, as I note below, providing some more details about those analyses/interpretations would be a good addition here.
Line 348. Rephrase to "enabled us to…"
Lines 350-351. What elution volume was used for these extractions? Without that, the concentrations are rather meaningless. Alternatively, are 260/280 or 260/230 ratios available to contextualize the quality of DNA? Expanding a bit more broadly, I wonder if any of those considerations could help explain the large spread in data return across the various extracts. Or some other consideration from the FA runs that were done on raw extracts? I am not the most familiar with HyRAD/HyRAD-X literature, but trying to figure out what the best practices are with these museomic extraction techniques seems to be the biggest sticking point that may be keeping the broader community from advancing these methods to the next level. Adding a bit more detail/discussion to that logistical topic might be a nice balance to the very detailed taxonomic/biogeographic discussion.
Line 352. I would also argue the same for fresh samples (concentrations not being homogenous)—it looks like almost a bimodal distribution among fresh samples, which might be worth noting in the text (or at least the large spread).
Figure 2. Some mismatch between parts of this figure—I assume the result of bad pdf conversion or something like that. Also, as a color-blind person, I have trouble diderentiating the grey and red asterisks (the boxes are a bit easier), FYI. Also also, great illustration and pictures—cheers Conrad!
Line 434. What value of K was determined to be best, and what did deltaK say? From figure 3, it seems like K=8 was supported, but looking at the supplemental barplots, K=7 shows virtually the same overall pattern, and it's really only fine-scale diderences that are showing up relative to K=6. More detail about the interpretation of these results would be warranted.
Line 445. "i.e." not needed.
Line 668. Personally, I think one or all of these networks should be included in the maintext. The connection here to hybridization is nice, but from a purely species delimitation point of view, the diderences in the diderent mt genes is interesting. Or does a combined 3-gene network show an overall composite of the findings of the three genes individually? That might simplify a main-text figure to one thing, which I think would be a nice addition.
Line 709. BioProject not currently available on NCBI—can't tell if just still embargoed or not.
Figure 3. Symbols are provided to contextualize specimens, as in fig 2, but aren't actually included in the figure (unless they also were subject to some pdf creation/conversion errors).
Note, I'd answer yes to virtually all of the questions presented in the PCI Zool reviewer's list, and think I've adequately addressed my concerns above.

Reviewed by anonymous reviewer 1, 29 April 2024

Lines 105-108 explicitly mention which are the two species
Lines 108-109 structure references and parentheses appropriately
Linea 324 indicate the beta value

Figure 1 homogenize colors in pinned (< 30-year-old)

Does the title clearly reflect the content of the article? [X] Yes, [ ] No (please explain), [ ] I don't know
Does the abstract present the main findings of the study? [X] Yes, [ ] No (please explain), [ ] I don't know
Are the research questions/hypotheses/predictions clearly presented? [X] Yes, [ ] No (please explain), [ ] I don't know
Does the introduction build on relevant research in the field? [X] Yes, [ ] No (please explain), [ ] I don't know
Are the methods and analyses sufficiently detailed to allow replication by other researchers? [X] Yes, [ ] No (please explain), [ ] I don't know
Are the methods and statistical analyses appropriate and well described? [X] Yes, [ ] No (please explain), [ ] I don't know
In the case of negative results, is there a statistical power analysis (or an adequate Bayesian analysis or equivalence testing)? [X] Yes, [ ] No (please explain), [ ] I don't know
Are the results described and interpreted correctly? [X] Yes, [ ] No (please explain), [ ] I don't know
Have the authors appropriately emphasized the strengths and limitations of their study/theory/methods/argument? [X] Yes, [ ] No (please explain), [ ] I don't know
Are the conclusions adequately supported by the results (without overstating the implications of the findings)? [X] Yes, [ ] No (please explain), [ ] I don't know