Revision round #3

Thank you for your thoughtful and detailed feedback on our manuscript. We appreciate the time and effort that went into reviewing our work again, and we understand that some points require further clarification and adjustments.

We would like to acknowledge that this project began almost two years ago. Since then, there may have been new species entries added to NCBI, and we appreciate this being brought to our attention. However, given the circumstances, we face some practical limitations. Two of our co-authors are no longer affiliated with our institute at LMU (WF, SV), and the third (GW) is currently on sabbatical. Additionally, I (Warren Francis) have since moved on to an unrelated position and do not intend to revisit this particular topic in the future. While we are open to making reasonable adjustments, we must acknowledge that it is not feasible to significantly expand or rewrite the manuscript at this stage. Nonetheless, we are certainly willing to revise or remove certain sentences where appropriate to improve clarity and focus.

We would also like to highlight that our study is intended as a focused contribution on a specific topic rather than a comprehensive review. Expanding the discussion to cover all aspects of speciation would be beyond the intended scope of this work. We believe the current focus aligns well with the goals of the paper, and we trust that the available data and code will enable future researchers to build upon our findings if they wish to do so. To this end, all relevant data, along with the Python and R scripts used in our analyses, are publicly available in our repository to facilitate this.

Regarding the remaining points of criticism, we believe there are two valid concerns that we are prepared to address. First, we acknowledge that the original title may have been overly broad. We are open to specifying "of animals" in the title (done so) to clarify that our findings do not extend to plants, fungi, or

bacteria. We believe this adjustment will improve the paper's accuracy without compromising the intended message.

Second, we recognize the oversight regarding *Crassostrea* species nomenclature. As one reviewer pointed out, some species were reassigned to the genus *Magallana* according to Salvi (2017). We updated the relevant figures and text accordingly, as this change does meaningfully alter the interpretation - particularly in terms of how one of the more divergent species pairs should now be classified. We appreciate the reviewer's attention to this important detail.

We are committed to making these revisions in a way that strengthens the manuscript and aligns with the reviewers' constructive feedback. We greatly value the opportunity to improve our work and thank you for your ongoing consideration throughout this process.

In the following, if replies are in first person, it is the first author (Warren Francis) replying.

Decision for round #3 : Revision needed

Review by anonymous reviewer 1, 27 Jan 2025 11:52

I commend the authors for their responses to my comments and for addressing my concerns. I now have only minor comments.

The document with all the supplementary figures greatly facilitated the identification of figures referenced in the manuscript. I thank you for providing it. However, the text within the figures in this document appears slightly blurry. If possible, improving its clarity would enhance readability.

We apologize for the blurry figures in the supplement. We suspect this is due to multiple re-rendering steps introduced by uploading the rendered figures to Google Docs, exporting, and possibly

another re-rendering again at bioRxiv. We will make sure that in a final version of the manuscript, should it be recommended, all figures are of the highest possible quality. In any case all of the supplemental figures were originally PDFs generated in R, and they are available as PDFs on the github repository in the folder "supplements for paper".

Furthermore, while I understand the authors' perspective that the title is already sufficiently long, I would suggest slightly modifying to make it less generic.

Done so, please see above.

Review by anonymous reviewer 2, 20 Jan 2025 07:17

The authors present an updated version of their manuscript studying genome evolution between congeneric species pairs.

As noted previously, I very much enjoyed reading the manuscript and believe it will be a valuable addition to the field. However, the authors have made minimal efforts to address my comments - which is of course their prerogative - but most of my previous concerns therefore remain and I have little more to add.

Thank you for the considerations. We did our best to address the critical points related to methods and results from the previous round. However, beyond the changes made in that round, it is not feasible to add lengthy complex discussion points, for reasons outlined above. Potentially these could be targets of future work for other researchers.

Review by anonymous reviewer 3, 20 Feb 2025 03:30

Warren et al. provide a novel look into the role of protein identity and macro- and microsynteny in speciation processes. In this updated manuscript, they improve their work in which they utilize congeneric species pairs to identify genomic processes that lead to divergence of species and seek to determine whether or not there are universal patterns that drive speciation. In the updated version, little has been changed with the exception of minor text edits. In some cases, reviewers expressed concerns with the methodology (not just minor text details), and the authors dismiss the commentary rather than making substantive edits.

The authors are FRANCIS et al.

A few that clearly still need addressing: 1) Several places were suggested for expansion in the introduction and clarification of the methods and discussion. Only the minor, "low-hanging fruit" edits were changed.

As we mentioned above, we have limited resources to devote to this manuscript anymore. We have tried our best to address the criticisms of the previous round.

Two reviewers highlighted that the title was vague and doesn't accurately capture the extent of the work.
Please see above.

3) Two reviewers expressed concerns about the criteria for species selection and taxonomic resolution. In many cases, the selected species seem somewhat arbitrary (e.g., why is the "mammals" represented predominantly by primates, rather than the diverse array of mammals that are known to have huge chromosomal rearrangements such as Equidae or Cervidae?).

As stated in the text, the additional species, which were not congeneric pairs, were in reference to another study that did the same analysis two decades ago. The insects were mostly from the Zdobnov study (most of their pairs), and that paper did not explain most of their methods, so I had to reinvent the same strategy here. For the primates, I added them simply because we are humans and other colleagues had asked me about the same parameters for humans vs chimps. It seemed like an obvious comparison to make.

Indeed, Cervidae were included in the original set.

4) Reviewer 2 raised a valid point with the correlations between protein identity, macrosynteny, and microsynteny. Without the "insect" comparisons, the already loose correlation would disappear. Rather than adding justifications in text, the author responds to "check back in 5 years" when additional references are available.

The authors' response that there is an insufficient number of references available is also categorically false. As an example, when the authors initially submitted the manuscript to bioRxiv, there were four Daphnia species with NCBI annotated, chromosomelevel assemblies, but choose only to look at two species. It is entirely valid to decide that additional analyses are not possible at this time due to time constraints, etc. It is also entirely valid to not use all the available resources if a research group is simply seeking to explore a phenomenon at a high level. But typically if authors have concerns with portions of the analyses, it requires changes and justifications within the text, rather than snarky responses.

At the time the original analysis was done, spring-summer 2023, there were 3x *Daphnia* species. From the 400 total animal species with chromosome level assemblies on NCBI at that time, 55 were in genera with more than one species, and most were arthropods and mammals. In the interest of focusing on whether there are patterns in animals more broadly, we did not include all possible combinations of arthropods or mammals, which could easily make 90% of the comparisons.

5) The authors did not address reviewer concerns about the divergence time of the taxa in the analyses. The authors respond saying that it is unnecessary due to issues with dating and that divergence times were not available for all species pairs in TimeTree (TimeTree is not the only source of estimates and often does not aligned with peer-reviewed literature). The issue is that

the divergence times may span orders of magnitude different time intervals which is concerning regardless of molecular clock and fossil calibrations. Pairs that may be diverged by hundreds of millions of years(Daphnia) are not apples-to-apples comparisons with pairs diverged by a few million years (Drosophila) even though we as a research community have arbitrarily binned them as congeners. This does not need additional analyses, only additional text.

The reviewer has a good point about the vast variation in divergence times, but this is why I did not want to discuss it in detail in the first place in the paper. This topic should be a paper on its own.

Relevant to this, as you point out later, is that Crassostrea gigas and angulata are now assigned to another genus, Magallana. This is exactly what we were indicating in Figure 3, that there is substantial divergence in protein identity across that genus (now two genera), and that might be hinting that these should not be classified to be the same genus in the first place. I do not work on bivalves, so I could not comment if there are other features that suggest they should be different genera, or whether any of the fossils are accurately defined to the genus or family level. This could mean that if a fossil calibration were based on Crassostrea sensu lato, and used as a calibration for the fossil presence of the genus, this calibration may now include multiple genera, meaning it actually should be a higher level calibration like family and will drag the node much further into the past than if it were used now for only Crassostrea sensu stricto. Potentially the same is the case for other species pairs in our dataset. As you point out, Daphnia has a long divergence time, but potentially also this could be a sign that it should be split into

I do not intend to look into that, but potentially future researchers can.

multiple genera.

Taken together, the work presented here has intellectual merit and is novel, but needs more work.

Thank you for recognizing the intellectual merit of our work. Due to the reasons outlined at the beginning of this document, we are currently unable to undertake extensive additional work. However, we firmly believe that our research establishes a solid foundation for future developments. Notably, we have made all code and data publicly available in our GitHub repository, facilitating further reference and extension.

I also have little new to contribute as the manuscript is nearly identical to the first round, but below are additional comments that have occurred to me since the first time reading the manuscript:

Additional comments: Minor comments: Page 4, Line 2-5: Authors should use full binomial Latin names instead of a mix of genus and common names.

We have added the latin names with common names in parentheses.

General:

1) Assemblies are not perfect and the criteria for selection needs to be described in the main text. The assemblies simply being "chromosome-level" is not enough. Although these are all chromosome-level assemblies, quality within this classification can still vary considerably.

The criterion for "chromosome-level" was based on their designation as such in RefSeq, distinguishing them from "Scaffold" or "Contig." While this is a reasonable consideration, we did not investigate the assembly method. Addressing this aspect would be challenging in this context, as such information is not typically encoded in RefSeq in a way that is easily sortable.

 Several of the assembly hyperlinks in the GitHub direct to "species" hubs on NCBI with multiple assembly versions with different assembly methods.

We apologize that this was not easier to find, but all of the accessions used were in the input data table "species_pair_list.tab" on the repo, in the folder "02-processing_scripts". This table contains the GCF_ numbers that you are looking for.

3) Several assemblies are missing from the list of species in the GitHub. For example, the manuscript Table 1 mentions that two species of Perca were used. In the GitHub "input_datasets" section, only Perca flavescens is listed in the markdown which links to the NCBI species pages. Perca fluviatilis is mentioned in another supplemental document, but as far as I can see there is no link nor text that describes the accession used. All species and accessions need to be listed in one spot not scattered throughout.

Apologies, this was an oversight on my part, since some of the species were downloaded from the web interface and not wget. Please instead refer to the above table, which will have all the GCF_ accession numbers. NCBI has also changed the databases of "Genome" and "Assembly" to "Datasets", so those links go to the legacy pages.

4) The species in the Crassostrea comparisons have been reclassified into two genera, Magallana and Crassostrea. The C. virginica v. C. angulata comparison in the main text is no longer a congeneric comparison. M. angulata v. M. gigas would have to take

its place to remain a congeneric comparison.

This reclassification appears to have been in 2017/2021, though I am not sure why this was not carried through to the NCBI taxonomy until last year.

In any case, I have changed the names for Figure 2, and modified Figure 3 to reflect this, as well as the supplemental figures. The text is slightly modified to explain that.

Review by Jean-Baptiste Ledoux, 29 Jan 2025 13:52

This is the second time I review this paper, entitled "Genomic changes are varied across congeneric species pairs" by Francis and collaborators. In my opinion, the Authors did a good job in considering most of the comments that were made in the previous round. I still believe that the paper would benefit from the addition of some microevolutionary/ popgen inputs in the introduction and discussion. My comment #4 was indeed not a question but a suggestion in this line. The concept of the "grey zone of speciation" fits with the discussion on divergence time among species etc. but with a population genomics perspective. This may broaden a bit the readership targeted by the manuscript while it supports that this species delimitation is a tricky question at the heart of evolutionary biology.

Thank you. As stated above, we have limited resources to continue working on this, so I do not intend to add any further discussion.

Anyway, I would like to acknowledge once again the rephrasing and adjustments made by the Authors that significantly clarify the outputs of the study.

Jean-Baptiste