Francis et al. 2024 BioRxiv – Review Round 1
09.10.2024

Summary

The recent explosion in the availability of chromosome-level genomes permits extensive genomic comparisons beyond the restricted model taxa in which they were previously conducted. The preprint by Francis et al. compares macrosynteny, microsynteny and protein evolution between 22 pairs of congeneric species belonging to five phyla. The authors report correlations between these three variables but do not find a “universal path” to species divergence. The manuscript was a pleasure to read and represents a valuable contribution to the field. In particular, the possibility of using genomic signatures for taxon delimitation is a pressing issue to address. I do not have any major concerns about the manuscript but I do have some intermediate level comments and minor comments, which I list below.

I preface the below comments by saying that the Supplementary Figures appeared not to be available on BioRxiv and though a link to a GitHub repository is provided and files to remake the figures are available, I could not find a document that presented them with their captions. If they are indeed available and I have missed them, I apologise, and suggest making their location more obvious. If not, please make them accessible – it would be most convenient as a PDF on BioRxiv. Please also note that the link to the Github repository is incorrect in the Data Availability section of the BioRxiv preprint (dash between Pal and Muc).

Intermediate-level comments

(1) The logic behind adding the ‘Insect’ and ‘Mammal’ comparisons is currently unclear and needs more explanation. If I understand correctly, these are “to remake the comparisons of previous studies”, at least in the case of the mammals, but the insects are not mentioned in any part of the methods section.

- Which species are used and how many?
- How distantly related are they?
- Is it valid to add these comparisons in to the main dataset or are they too different?

  - Are these comparisons taken into account in the main conclusions of the results section (e.g. “microsynteny broadly correlates with protein identity across animals”)?

The addition or not of these comparisons appears to make a significant impact on the conclusions and therefore should be fully explained and justified. The datasets themselves need something equivalent to table 1 to readers can study their composition.

A significant portion of the first sections of the Results section is devoted to the discussion of these groups. This is not consistent with the relative weight placed on them in the methods section. I wonder if it would be easier for the reader to follow if the main questions of the paper were first addressed without these extra comparisons and then there was a separate section specifically discussing these additions and their importance.

(2) I find the ‘loose correlation’ reported as shown in Figure 1C to be unconvincing to the eyeball test, especially if the ‘insect’ pairs, which as far as I can tell are separate from the main dataset, are removed. Did you do any statistical test to show that there is indeed a relationship between these two variables?

(3) As in point (2), the statement that “microsynteny changes more frequently than macrosynteny” is equivocal (and to me very logical and expected) but seems only very weakly supported by the data in the plot, again with no statistical assessment. Is there any other data supporting this on top of Figure 1D? Can statistical tests be used to test the claim?

(4) “Distribution of protein identity values is similar across species pairs” is again a very clear statement which the evidence immediately available to the reader does not clearly support. On the contrary, the species shown in Figure 3 seem to have highly variable distributions. Indeed, as you say that species “may all be at different stages of divergence”, would it not be tremendously surprising for the distribution of protein identity values to be similar?

Though as said above I could not access the Supplementary Figures so this information may be there, it is quite difficult for the reader to make any assessments of broad trends when the main text figure contains only 2 of the 17 genera studied. I wonder whether there is a way of summarizing this data across all 17 genera to make it easier to assess broad patterns? For instance, could you compare the mean protein identity to pairwise divergence time for all species? This would then allow you to see whether pairs diverge in protein identity at different rates. I know in a later section you discuss the pitfalls of using divergence times in this context and I understand the reluctance to open up that particular can of worms, but it strikes me that even rough estimates of divergence times between each pair of species would add significant context to the analyses.

Minor comments

P1L72 – Implies that synteny-breaking genome rearrangements are not a type of mutation. Surely they are indeed a type of mutation, albeit distinct from point mutations, indels etc.?

P2L10 – This sentence is quite dismissive considering others have argued that macrosynteny changes play a major role in speciation (E.g. Augustijnen et al. 2024 Sci Adv. https://doi.org/10.1126/sciadv.adl0989.) This may be worth briefly discussing or at least rephrasing.

P3L27 (1) – This is not a criticism but simply a question. I am interested in the thinking behind your method of orthology assignment. Why not use a highly-cited and robust program like OrthoFinder which would achieve many of the described steps automatically?

P3L27 (1) – How many orthologues did you have per species pair in your final dataset? If this data is not currently available, it should be reported. Could be added to Table 1 or as Supplementary.

P4L63 – Can you explain in more detail how chromosome fusions and splits are distinguished?

P7L42 – where in Figure 1 is there evidence of ‘a strong correlation between protein identity and divergence time’?

Fig. 1A ‘enes’ -> ‘genes’

General questions from PCI

- Did you read the “guide for reviewers”? (see the Help menu of the thematic PCI or the dedicated blog post) YES
- Is the manuscript well written? YES
- Is the description of the rationale and methods clear and comprehensive? YES
- Are there flaws in the design of the research? NO
- Are there flaws in the analysis? NO
- Are there flaws in the interpretation of results? POSSIBLY – my comments request further explanation of some results.
- Do you have concerns about ethics or scientific misconduct? NO
- Did you detect a spin on the results, discussion or abstract? (a spin is a way of twisting the reporting of results such that the true nature and range of the findings are not faithfully represented, https://doi.org/10.1073/pnas.1710755115) NO
- Is something critical missing? NO

Evaluation of the various components of the article

Title/abstract/introduction

- Does the title clearly reflect the content of the article? YES
- Does the abstract present the supported findings of the study concerned and no other? YES
- Does the introduction clearly explain the motivation for the study? YES
- Is the research question/hypothesis/prediction clearly presented? YES
- Does the introduction build on relevant recent and past research performed in the field? YES

Materials and Methods

- Are the methods and analysis described in sufficient detail to allow replication by other researchers? YES
- Is the experimental plan consistent with the questions? YES
- Are the statistical analyses appropriate? Further analyses required.
- Have you evaluated the statistical scripts and program codes? NO

Results

- Have you checked the raw data and their associated description? NA
- Have you run the data transformations and statistical analyses and checked that you get the same results? NA
- To the best of your ability, can you detect any obvious manipulation of data (e.g. removal)? NO
- Do the statistical results strongly support the conclusion ($p< 10-3$ or $BF>20$)? NA
- In the case of negative results, was a statistical power analysis (or an appropriate Bayesian analysis) performed? NA
- Did the authors conduct many experiments but retain only some of the results? NO

Discussion

- Do the interpretations of the analysis go too far? ? POSSIBLY – my comments request further explanation of some results.
- Are the conclusions adequately supported by the results? ? POSSIBLY – my comments request further explanation of some results.
- Does the discussion take into account relevant recent and past research performed in the field? YES
- Did the authors test many hypotheses but consider only a few in the discussion? NO

References

- Are all the references appropriate? YES
- Are the necessary references present? MOSTLY
- Do the references seem accurate? YES

Tables and figures

- Are the tables and figures clear and comprehensive? YES
- Are all the tables/figures useful? YES
- Are there too many/too few tables and figures? NO
- Do the tables and figures have suitable captions such that they can be understood without having to read the main text? YES