

# Response to reviews

Feb 20, 2023

19-Oct-2022

Dear Dr Sanchez Reyes:

Decision on USYB-2022-152, DateLife: leveraging databases and analytical tools to reveal the dated Tree of Life:

Accept pending minor revisions.

Thank you for your Systematic Biology submission. It has been reviewed by Associate Editor Dr Daniele Silvestro and two reviewers with the relevant expertise. Their comments are listed at the end of this letter. Both reviewers and the AE provide some excellent constructive suggestions that I am sure you will appreciate. The general consensus is that this software is of general interest to and expected to be widely used by the systematics and evolutionary biology community, providing capabilities that will help in the construction of supertrees. I agree with this assessment.

I think that your paper will be a valuable contribution to Systematic Biology once the detailed comments provided below are addressed, especially in relation to data curation and the English style. Thank you very much for your submission.

Sincerely,

Dr Isabel Sanmartín Editor-in-Chief, Systematic Biology [isanmartin@rjb.csic.es](mailto:isanmartin@rjb.csic.es)

Please acknowledge (by email to [sysbio.editorialoffice@oup.com](mailto:sysbio.editorialoffice@oup.com)) receipt of the reviews and give a probable time frame for the return of your revised manuscript. Your revision should be submitted as soon as possible. Any revision not received in a timely manner may have to be considered a new submission.

Your revision must comply with our instructions to authors (in this letter, with additional instructions at [https://academic.oup.com/sysbio/pages/General\\_Instructions](https://academic.oup.com/sysbio/pages/General_Instructions)). This applies even if your original submission deviated from Systematic Biology style and corrections were not included in the notes from the editors or reviewers. Therefore, **READ AND APPLY OUR INSTRUCTIONS BEFORE SUBMITTING A REVISION**. Failure to do so will result in significantly delayed processing of your paper. If anything remains unclear after reading the instructions, use recent issues of the journal to find examples. If necessary, feel free to contact the Managing Editor at [sysbio.editorialoffice@oup.com](mailto:sysbio.editorialoffice@oup.com) for help.

Log into <https://mc.manuscriptcentral.com/systbiol> and enter your Author Center, where you will find your manuscript title listed under “Manuscripts with Decisions.” Under “Actions” click on “Create a Revision.” Your manuscript number will be appended to denote a revision. Your original files are available to you. Delete the old files and replace them with your modified versions. Files that did not require revision can simply be retained

You will be unable to make your revisions in ScholarOne Manuscripts. Instead, revise your paper using a word processing program and save it on your computer. **HIGHLIGHT** the changes to your manuscript within the document using the track changes mode in MS Word or by using bold, strikethrough, or colored text; any method you choose which indicates all changes made. If a large number of changes were made and you feel the document is too cluttered to read easily with all of them shown, feel free to submit a “clean” copy as well, in which changes are not indicated.

If you decide a clean copy would likely be helpful to reviewers and editors, please upload it in the category called Related Files.

When submitting your revised manuscript, address each point made by the Editor, AE and Reviewers IN THE SPACE INDICATED (under “Response to Decision Letter”). Your revision cannot be processed if your responses to reviews are given only in a cover letter. The best way to address each point would be to copy this letter and insert your comments after each point made. Please do not change the order of or delete any of the comments because this makes it difficult to review again and would slow the process. The format of your responses must be compatible with ScholarOne Manuscripts text fields. For example, colored text is not an option, but you could use asterisks (and numbers, spacing, etc.) to clearly distinguish your responses from the text of the reviews.

Feel free to argue your case, with careful consideration and documentation, if you disagree with any of the suggestions. If you feel a reviewer did not understand a point you made, in your response keep in mind that as an author it is your responsibility to make your points clear to the readers.

DO NOT submit your revision in .pdf format. This applies to the main text, tables, and appendices. Individual figures may be in .pdf format.

In the specific case of papers written using LaTeX, in addition to the .tex (and associated style, bib, or etc. files) please include a .pdf generated from those files. Upload the .pdf in the ScholarOne Manuscripts category called Related Files.

Figures must be uploaded separately, with each file name including the figure number. No figures may be imbedded in or tied to the .pdf or .tex files.

After you’ve uploaded your revision, carefully view the ScholarOne Manuscripts version to verify that all figures and other files display correctly, and that you’ve followed all of our author instructions.

Figures: Each figure must be submitted as an individual file. The name of the file should include the figure number. Numbers and letters on figures should be as large and clear as possible, without overlapping any text. No figures may be imbedded in the document. Captions should not be placed on the figures. Figure portions should be referred to in the text and figure captions using lowercase letters. On the figures themselves, portions should be labeled by a lowercase letter followed by a single parenthesis (e.g., “a”), located in the upper left area of the figure portion. Line thickness (including graph axes) should be a minimum stroke weight of 0.5pts, and 1.0pts is recommended for most lines. We prefer vector rather than bitmap figure formats. If desired, see <http://systbio.org/?q=node/138> for an explanation of the difference. If bitmapped figures are necessary, they should be created at a minimum of 300 dpi and should be about 8 inches wide. We accept a wide variety of figure file types, as long as the figures are of sufficiently high resolution.

The cost of printed color figures is \$600 each. Authors are normally expected to cover this cost, but we do have limited funds available for authors who are SSB members, and cannot pay the full amount. If you feel your circumstances create an unusual financial need, please explain in your cover letter. The maximum allowance is one color figure per paper. All color costs can be avoided if you decide to have the figures printed in black and white, but be shown in color in the online version of the paper. If you do this, the captions must be worded such that they are appropriate for both situations (e.g., descriptions should not name colors), because the captions in print and online will be identical even if color is used online only. Do not submit two versions of any figure. Instead, make sure the color figure is also easily legible in grayscale, then submit only the color version. In your cover letter, please make your intentions regarding the use of color clear.

Journal Covers: Please consider submitting a suggested cover image. They can be illustrations of theory, photos of organisms, a combination of the two, or alternatives. These can be uploaded with a revision of the manuscript or may be sent later. Only images for which you have copyright

permission or that are not under copyright may be submitted. The color figure fee does not apply to images chosen by the Editor to be on the cover of the issue. If you have questions about possible cover designs, please email [sysbio.editorialoffice@oup.com](mailto:sysbio.editorialoffice@oup.com). We encourage and greatly appreciate cover image suggestions.

**Data and Online Appendices:** All such files should be deposited in the Dryad data repository. Also, before the manuscript can be published, data accession numbers must be in the text. Nucleotide sequence data and alignments must be submitted to GenBank (<http://www.ncbi.nlm.nih.gov/genbank>) or EMBL (<http://www.ebi.ac.uk>); alignment, input file and tree files must be submitted to TreeBASE (<http://www.treebase.org>); morphological data must be submitted to either Morphbank or MorphoBank. Check over your online appendices or supplemental material (if any) carefully, because they will not be copyedited or proofread, and cannot be changed later. The first time you mention online-only material in the text of your paper, give the doi provided by Dryad as the location where the material can be found. Then, in a separate section after the main text, include a statement such as:

**SUPPLEMENTARY MATERIAL** Supplementary material, including data files and/or online-only appendices, can be found in the Dryad data repository (put your Dryad doi in parentheses here).

We added DOIs for Dryad package with Supplementary Material, L551.

Make sure all section headings conform to Systematic Biology style for first, second and third levels. Use of incorrect styles is potentially confusing and, in any case, is likely to delay processing of the manuscript.

Noted and checked.

Tables should have a single-sentence informative title above the table, with any other descriptive information located below the table in the form of notes and/or specific footnotes.

Noted and checked.

When references are grouped together in parentheses in the text, they should be listed in ascending chronological order. Multiple references in a single year should be alphabetized.

Noted and checked.

All funding used for this work should be listed in a “Funding” section preceding the Acknowledgements. Please give the full official name of each funding body.

Done on L556-558.

Our accepted abbreviation for millions of years ago is Ma. The abbreviation for millions of years duration is myr.

Checked.

Our full instructions are accessible via the link in the upper right area of the ScholarOne Manuscripts page (the link goes to [https://academic.oup.com/sysbio/pages/General\\_Instructions](https://academic.oup.com/sysbio/pages/General_Instructions)). In case of a discrepancy between the information at that site and this letter, follow the instructions in this letter.

If you are the first author and your manuscript is based on work done while you were a student, you would be eligible for the “Publisher’s Award for Excellence in Systematic Research.” When submitting your revised version be sure to indicate student work in the checkbox provided. If two student researchers were heavily involved in the manuscript, please briefly describe the situation in your cover letter.

This work was done while the first author was a Postdoctoral researcher.

Please note that the accepted version of your manuscript will be published on Advance Access prior to typesetting and copyediting; it is therefore important that all files, including tables,

figures, and supplementary files, are properly labelled and with the editorial office at this stage of the process - it is the author's responsibility to ensure that the correct files are submitted. Please also keep in mind that files will be published as submitted (ie, please review table formatting and labelling and position of figure images, etc). For further information please see the Advance Access Publication section at: [https://academic.oup.com/sysbio/pages/General\\_Instructions](https://academic.oup.com/sysbio/pages/General_Instructions).

Noted and checked.

Associate Editor: Dr Daniele Silvestro

Recommendation #1: Accept with minor revisions

Comments to the Author: Dear Dr Luna L Sanchez Reyes and co-authors,

many thanks for submitting your manuscript to Systematic Biology. Your study was reviewed by two highly qualified reviewers who provided an overall positive assessment of the paper while listing a number of things that should be revised and clarified. Based on their assessment and my own reading of the manuscript I invite you to resubmit it after carefully revising to address each and all points raised by the reviewers.

Thanks for your reviews and comments! They were most helpful. We carefully revised the manuscript to address all points raised by the reviewers and yourself and we hope you find the revision suitable.

2. Why are the node ages evenly distributed between calibrations? I would expect an exponential distribution of node ages under a standard birth-death process.

Thanks for pointing this out. We elaborate on this on the "Description" section, at the end of subsection "Dating a tree topology with no branch lengths", L 202-210 (Note that we expanded the subsection "Dating a tree topology" into three sections: "Applying secondary calibrations", "Dating a tree topology" and "Dating a tree with branch lengths").

Briefly, in the absence of genetic data from an alignment, nodes that have no age data from published studies in the chronogram database are assigned an age based solely on other nodes that do have age data from published studies. MrBayes has an option in which you can use a birth-death strict-clock model to sample branch lengths in the absence of genetic data. We have implemented a DateLife function called "make\_mrbayes\_tree", that wraps this MrBayes service. Inherited from MrBayes implementation, the function requires setting parameter values for birth and death rates. If the amount of missing data is large, the chosen parameters can determine the "shape" of the resulting summary chronogram in terms of branch length distribution.

We added a discussion section, "Effects of phylogenetic sampling on downstream analyses" (L445 of manuscript without differences; L527 of manuscript with differences), in which we discuss the effects of chronograms with missing data that have been completed under a certain diversification model, especially when used for analyses that require estimating a diversification rate, which might introduce circularity (Rabosky, 2015. *Evolution* doi.org/10.1111/evo.12817).

MrBayes also allow us to generate a cloud of trees produced with different birth-death rates that could be used for downstream analyses. However, the main goal of the summarizing step in DateLife is to provide a single chronogram that can quickly show in one glance the distribution of node ages based on published data, in the most agnostic way possible.

This is why we chose an algorithm that distributes node ages evenly between calibrations and does not require any assumptions on the underlying model of branch length distribution. The Branch Length Adjuster (BLADJ) algorithm allows this and is the one set as default for the summarizing step. However, users that wish to use MrBayes instead can do so with the "make\_mrbayes\_tree" function.

Please make sure to carefully revise the text to remove typos.

We revised the text throughout to remove all typos we could detect.

I hope you will be willing to revise and resubmit your paper and that you'll find these and the Reviewers' comments useful.

Best regards, Daniele Silvestro

Yes, thanks!

There were many typographical errors in the manuscript which should be corrected prior to publication.

We corrected typos across the text.

A separate but related point that I would also like to see discussed is that synthesizing services such as DL and OTOL seem capable of ending up in loops where bad trees with bad calibration points provide the skeleton for further bad trees based on the former - with their own seemingly well-supported but in fact dodgy secondary calibrations. Is that a risk? What can be done about it?

It is a risk that we are aware of, but it is possible to avoid it. We explain how in the text.

Apart from these general points that might be touched upon a bit more in the Discussion, here now some specifics about the manuscript:

- The Abstract looks like an extreme afterthought. I understand how that works, but please have another look. I see verb disagreement on line 21 and on line 23. Probably needs a comma after databases on line 25. On the same line, 'timeframe' is spelt as one word (fine by me), but elsewhere it's two words. Line 27: 'incetivized' is not a thing. Line 29, 'finding' scans weird, maybe use 'discovery'? Line 36 probably needs 'use' instead of 'using' but the sentence is hard to parse. Line 38, 'awereness' is wrong. In this way, the Abstract is quite different from the rest of the MS, which is otherwise well written.
- On page 7, second paragraph, you state that subspecies are ignored. What do you mean precisely? My guess is that you ignore the subspecific epithet and collapse to species level. Maybe state that more clearly.

Thanks for pointing this out. We actually only ignore subspecies when retrieving data from a more inclusive taxonomic group. When provided by the user, subspecific taxa are processed and searched as regular species. We clarify this in paragraph X, page 7

It would be possible to drop the subspecific epithet and perform a more general search. We added this functionality to our software development plan, so that it can be soon implemented in a future iteration of the software.

- On page 8, fourth paragraph, it's not quite clear whether DL's database syncs automatically with Phylesystem or whether you have volunteered yourself for this task. Which would be noble, but hard to sustain.

It is indeed currently synced "manually". However we are working to have the syncs done automatically. We comment this in the discussion and clarified that it is currently a manual sync on page 8.

- On page 25 you mention the fossilcalibrations.org initiative. Maybe that's a good opportunity to go a bit into what we need as a community. I suspect that, in general, most people in this field think that doing it by themselves is 'better', i.e. do a bunch of sequencing (hybseq right now, I guess?) and then get good primary calibration points. Natural history collections must have many more of those, both as fossils but also from geology (i.e. vicariant events having to do with tectonics, orogeny, etc.). Shouldn't we want *that*?
- On page 26 you discuss some criteria for scoring quality of chronograms. One additional criterion might be where the calibration points are placed. Nodes that have a calibration point between them and the root have less freedom of movement and hence narrower confidence

intervals. Ages ago, I did a bit of simulation work on that (Vos & Mooers, 2004 - definitely no need to cite). Maybe someone else has discussed this a bit better?

- Page 28 line 473: I think it should be either ‘public-funded’ or ‘publicly funded’

Went for “publicly funded”, thanks!

- 
- Page 29, Supplementary Material: it’s probably better to sync the repos with Zenodo and cite the DOI, just so that it’s guaranteed unchanging.

Thanks for the suggestion! We created stable versions for all three repositories on Zenodo, and refer the doi instead of the GitHub address for the repos.