nature portfolio

# Peer Review File

# Rates of passerine body plan evolution in time and space

Corresponding Author: Dr Jacob Berv

Version 0:

Decision Letter:

27th October 2025

*Please ensure you delete the link to your author homepage in this e-mail if you wish to forward it to your co-authors.

Dear Dr Berv,

Your manuscript entitled "Rates of avian body plan evolution in space and time" has now been seen by three reviewers, whose comments are attached. The reviewers have raised a number of concerns which will need to be addressed before we can offer publication in Nature Ecology & Evolution. We will therefore need to see your responses to the criticisms raised and to some editorial concerns, along with a revised manuscript, before we can reach a final decision regarding publication.

We therefore invite you to revise your manuscript taking into account all reviewer and editor comments. Please highlight all changes in the manuscript text file [OPTIONAL: in Microsoft Word format].

We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

When revising your manuscript:

* Include a “Response to reviewers” document detailing, point-by-point, how you addressed each reviewer comment. If no action was taken to address a point, you must provide a compelling argument. This response will be sent back to the reviewers along with the revised manuscript.

* If you have not done so already please begin to revise your manuscript so that it conforms to our Article format instructions at http://www.nature.com/natecolevol/info/final-submission. Refer also to any guidelines provided in this letter.

* Include a revised version of any required reporting checklist. It will be available to referees (and, potentially, statisticians) to aid in their evaluation if the manuscript goes back for peer review. A revised checklist is essential for re-review of the paper.

* Extended Data Figures - please ensure that any supplementary figures and tables that are crucial to the manuscript’s conclusions are converted into Extended Data figures and tables to increase visibility of these data. Extended Data figures and tables are online-only (present in the online PDF and full-text HTML versions of the paper), peer-reviewed display items that provide essential background to the article but are not included in the main article due to space constraints. A maximum of ten Extended Data display items (figures and tables) is permitted.

Please use the link below to submit your revised manuscript and related files:

Link Redacted

**Note:** This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

We hope to receive your revised manuscript within four to eight weeks. If you cannot send it within this time, please let us know. We will be happy to consider your revision so long as nothing similar has been accepted for publication at Nature Ecology & Evolution or published elsewhere.

Nature Ecology & Evolution is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as ‘corresponding author’ on published papers create and link their Open

Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on 'Modify my Springer Nature account'. For more information please visit please visit <a href="http://www.springernature.com/orcid">www.springernature.com/orcid</a>.

Please do not hesitate to contact me if you have any questions or would like to discuss these revisions further.

We look forward to seeing the revised manuscript and thank you for the opportunity to review your work.

[redacted]

Reviewer expertise:

Reviewer #1: phylogenetics, bird body plan evolution

Reviewer #2: behaviour and trait distributions, macroecology

Reviewer #3: passerine skeletal evolution

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

I thank the authors for their interesting work which is ambitious in scope and beautifully visualized.
I have undertaken a careful review, and where I have been motivated to ask the authors questions I have tried to justify my intuition with simulations or example analyses.
I am open to the authors refuting all my questions.
I have attached a detailed pdf review document for the authors, including an example R code framework for a suggested method to interrogate the correlation of time-series of estimated evolutionary parameters with global environmental time-series.
I want to thank the authors for their patience in waiting for my review response, and hope that they find my questions useful.

I have copied my general summary and main questions below:

The manuscript investigates a dataset of limb skeletal length measurements and body masses collected across >2000 passerine bird species. The authors aim to understand patterns of variation in the rates of trait evolution across the passerine family tree. The authors want to know whether major environmental change explains variation in rates of trait evolution. The authors want to know whether weaker covariance between traits is associated with faster rates of phenotypic change.
The authors use a custom code implementation to explore these questions, and arrive at the conclusion that passerine evolution is 'bursty', with multiple sudden increases in trait evolutionary rates followed by gradual declines. They believe that the strongest bursts are associated with episodes of major environmental change and that weaker covariance between traits is associated with faster rates of trait evolution-- perhaps suggesting that traits which evolve independently of one another are more amenable to selective forces.

I am convinced by the authors' claim that there are recurrent bursts in passerine evolution with subsequent slow-downs. I can probably believe that there is a systemically higher rate of speciation in the new world, as opposed to the old world, given that passerines emerged in Gondwanaland and may have already packed niche-space there more tightly. However, it may also be possible there are simply lots of ancient isolated lineages in Gondwanaland and these bias estimates of lineage origination and alter inferred lineage-specific rates of recent evolution.
I am not convinced yet that there is a high-latitude bias for higher lineage origination rates (I would sooner believe this reflects patterns of dispersal and Rapoport's rule). I am not convinced that the equation the authors have employed to estimate recent lineage-specific phenotypic rates is robust, and I propose a way to show this.
I need some more clarification about the way covariance matrices were represented across the family tree and the inference of modularity before I feel capable fully assessing the claims made.

Lineage phenotypic rate estimation:
I believe the authors' chosen equation will cause an over-estimation of phenotypic rate evolution in out-groups relative to diverse niche-packing radiations, even if the current rates of phenotypic evolution are similar in both. I believe this because there are more recent nodes in well-represented radiations, and those will record the deceleration in rates and receive strong weightings. Outgroups with no recent nodes will appear to retain fast rates because the last available and most strongly weighted nodes are basal to bursty events.

Modularity:
Please take the upper triangles of the post-hoc freely estimated sigma-squared matrices, vectorize them, and stack them into a matrix. Please perform PCA on this matrix and review their distribution;
is the clustering of these matrices' properties sympathetic with the temporal distribution of the nodes? Does it emphasize major events? What kinds of changes in covariance are occurring and do they help us better understand the unique

phenotypes that have emerged in different parts of the passerine family tree?

Equal splits:
I would like a stronger consideration of the role of dispersal and milankovic cyclicity mediated extirpations at high and mid-latitudes in setting the apparent lineage division rates in geographic cells. I think the current analysis risks giving the impression that the estimated values in the cells represent endogenous speciation rates and believe they probably do not. I am concerned that many cells with the fastest rates are adjacent to regions with extremely low passerine diversity and therefore may not be well-constrained by the available data in the cells- I point to the eastern Sahara and Canadian Arctic in particular.

Time-series correlations:
I have written an example code which shows that the distribution of evolutionary parameters resolved at the nodes on a phylogeny can be discretized over time and meaningfully correlated with a time series such as d18O.
The procedure is as follows:
For each time step, record all the nodes in the interval.
Compute the difference between the parameters of these nodes and their parents, divided by the elapsed times, to obtain the estimated derivative in the parameter value.
Take the phylogenetic mean of these estimated derivatives.
Normalize this vector by its variance, and correlate it with the variance-normalized vector of all d18O observations for which there were valid observations of the nodal parameter values.
I applied this method to a worked example to demonstrate its utility.

Reviewer #1 (Remarks on code availability):

I inspected the coding functions on the GitHub repository.
There was indeed a well-organized ReadMe file.
The code is comprised of multiple inter-dependent functions which total several thousand lines in length.
It was not within my abilities to manually review all code to my satisfaction in the timescale that is available for review.
The dataset is nominally available, but it is associated with another manuscript.
I manually read through the code to look for mistakes instead of re-running the analysis directly. I did not find any clear mistakes.
My intuition is that future authors who might use the same functions will be unlikely to manually read through the code.
I would have appreciated it if the in-line comments in the coding functions, such as 'TemporalAnalyses-functions.R' included tags where GPT models had been employed to construct code, to increase the transparency surrounding the role of Large Language Model contributions.
I would advocate adding a tag at the beginning and end of sections copied from an LLM output, and the inclusion of the prompt text.
I appreciate the authors are unlikely to have systematically recorded this information, though.

In Sum, I was not presented with any reason to believe the coding implementation makes any mistakes.
However, I believe some of the fundamental equations being implemented in the analyses (such as lineage-specific estimates of recent evolutionary rates) are not well-defined.

As part of my review I wrote new code and conducted simulations, and was able to show that a latitudinal gradient in lineage diversification rates can be produced by biased patterns of dispersal, and would like the authors to consider writing their own simulation to explore their method's performance. My simulation was very simple, and would be improved if species' dispersals were represented by Random-Walks with different variances, and if species' extirpation probabilities were functions of a simulated latitudinal temperature gradient that fluctuates with Milankovic cycles.

Reviewer #2 (Remarks to the Author):

This is an incredibly exciting manuscript, using a recently-published dataset of bird skeletal features, a recently-published avian phylogeny, and brand new phylogenetic comparative methods to demonstrate that (passerine) avian morphology is the result of rare, mostly early, mostly phylogenetically-clustered bursts of evolution. Similar results have been found by other groups using other datasets/methodologies, but this is a really nice demonstration of this phenomenon. This manuscript is written to a very high technical standard, and I found the authors' interpretations to be very insightful. This manuscript is also a rare example of a study that opens up many future avenues for high-quality, important scientific discovery; these results, especially coupled with the twinned data and new methodological approach, will be ones that I'm thinking about for quite a while.

I have two general comments and a small number of line comments, all incredibly minor. Otherwise, though, I *greatly* enjoyed reading this paper, and I salute the authors for all of their hard work!

General comments:

I will apologise to the authors if this is addressed somewhere in their supplement, but I am a little concerned that some of these results are an artefact of sampling bias. Specifically, present-day lineages are undersampled (1/3 of extant species compared with presumably nearly all of the early passerine lineages), which might lower the estimates of recent

morphological diversification rates. If the authors have addressed this in some way, I would recommend highlighting it slightly more (and if not, I would recommend addressing this).

I also would have appreciated – and, again, I apologise if this was somewhere in the supplement and I missed it – a quick summary of what these 12 skeletal traits are. What, morphologically, do they represent? And if possible, how do a reader link these broad, general changes in morphology/'bauplan' to specific ecological functions of individual traits? The argument about increased modularity and functional compartmentalisation is interesting, and likely true, but it would be greatly strengthened by the ability to extract morphological meaning from (some of) these results.

Line comments:

Abstract: As far as I know, I don't think citations in the abstract are a norm for this journal? I would suggest removing these.

References in general: They currently seem to be at the very end of the manuscript, referring to both the main manuscript and the supplement. The authors might want to consider separating these.

L234: The statistical significance of K_{mult} is going to be highly dependent on sample size – how much of this finding is just that avian species richness starts declining around \pm 50?

L234: Also, I would recommend specifying that these K_{mult} values are statistically significant *from zero*, as in theory you could have set up the test to distinguish K_{mult} from some other value. (As a broader point, K_{mult} \neq 0 is a very weak test. Depending on what sort of phylogenetic signal you're interested in, small, non-zero values of K can indicate truly inconsequential signal, particularly across a large sample of species. In other words, I'm not sure that a binary zero-or-not-zero is at all biologically interesting, or at least not as you're currently framing this.)

L430-432: Forgive me if this is in the supplement and I missed it, but what's the spatial sampling like in this dataset? Any notable geographic biases?

L436: parenthesis error

Figure 4: I'm fascinated by the regional variation in this figure! (Not an action point. Just an expression of amazement.) (Also, in general, these figures are done to a remarkably high standard. Well done!)

L498: parenthesis error

L762: To be a little pedantic: "…multivariate trait similarity tracks phylogenetic relatedness UNDER BROWNIAN MOTION, while…" . Specifically, under single-rate Brownian motion! There could be strong phylogenetic signal under some other tempo/mode.

Reviewer #2 (Remarks on code availability):

I glanced at the code, but I did not try to run/reproduce it. It seems well-signposted/well-annotated from what I could tell.

Reviewer #3 (Remarks to the Author):

This manuscript represents an impressive body of analytical work with a very large and comprehensively sampled dataset. It sets out to answer questions relating to the mode, tempo and spatial aspect of passerine body plan evolution, investigated by applying a suite of analyses on a dataset representing the proportional organisation of the passerine skeleton. I think it does a great job of answering these questions and discussing major evolutionary themes, and the paper will be a much-cited contribution to the literature for passerine workers, ornithologists and the wider community of evolutionary biologists. The figures are beautiful and informative. The methods are detailed and the supplement contains a lot of additional background, including illustrated supplementary results. I do however have some issues with how some things are communicated in the main text. My suggestions are pretty much all about wording, so I think this warrants minor revisions and should be accepted once the concerns are addressed.

One problem I have with the way the work is communicated is unnecessary vagueness in the interesting results. There is description of the overarching patterns and processes, but not enough specific details for the reader to truly understand the relevance of the work in terms of passerine evolution. There needs to be more examples of clades, environments, ecological factors, so that we can understand some of the key takeaway points in more context. See my points below for the paragraphs that need specific taxonomic or environmental examples. Doing this would improve the accessibility of the paper.

I think the narrative should be shifted from avian body plan evolution to passerine body plan evolution – I don't agree that passerines are representative of all living birds. On the contrary, they are a peculiar clade – no other group across tetrapods diversified to the same extent in such a short time span, and there has to be something unique (which we currently don't understand well) about passerines that enabled them to diversify like this. Also, passerines are remarkably uniform on a whole skeleton level (see Navalón et al. 2022 which should definitely be cited, is highly relevant for multiple aspects of this paper, but don't think it is cited: see citation below), whereas other bird groups are remarkably varied in their skeletal organisation and proportions. Passerines are incredible and deserve more attention on such a broad scale like this. They

should be highlighted as something extraordinary in evolutionary terms, rather than being used as a model system to explain evolution across all bird groups. It's rare to have a paper exploring passerine morphology so thoroughly with a third of their 6,000 living species represented, and it's great that they are finally being put in the evolutionary and morphological spotlight. I suggest adapting the title to say 'passerine body plan evolution' instead of 'avian', and adjust the wording in the abstract, beginning of the introduction and the conclusion section. If you want the title to sound more enticing, then change it to 'Rates of body plan evolution in space and time across the most diverse avian order' or something similar.

Navalón G, Bjarnason A, Griffiths E, Benson RB. Environmental signal in the evolutionary diversification of bird skeletons. Nature. 2022 Nov 10;611(7935):306-11.

Along a similar vein, I don't think this paper has been cited either and it probably should be as very relevant for passerine skeletal organisation too:

Brinkworth, A., Green, E., Li, Y., Oyston, J., Ruta, M., & Wills, M. A. (2023). Bird clades with less complex appendicular skeletons tend to have higher species richness. Nature Communications, 14(1), 5817.

Line 32: Very pedantic, but I suggest changing 'clade' to 'group' or saying 'order-level clade' because one could argue the most diverse clade of living birds is Neornithes…

Somewhere in the first few paragraphs, you should be more explicit about the dataset. I.e., briefly explain you measured the longest length of 12 ecologically important bones to generate a mass-corrected body plan profile for each species. I had to read the methods and then open the published dataset paper to understand exactly what kind of raw data you analysed, and I think this information should be presented more clearly in the first paragraphs of the main text. Especially considering my next point about body shape.

Line 75: I don't think body shape is the ideal example here, because the data is longest length measurements of individual bones, so saying 'shape' feels misleading and I think should be reserved for specific shape analysis (e.g. landmark morphometrics). Also, several components are missing such as the vertebral column which is an integral part of overall body shape. I would remove the example in parentheses and just keep it at proportional organisation of the skeleton, or perhaps skeletal architecture? But swap 'body shape' for 'body plan' or some alternative (architecture, proportional organisation) throughout the main text. Otherwise the data part feels ambiguous and one might think the data consists of 3D GMM, for example. E.g., body shape is also used in lines 128, 172, 204 and should be changed.

Line 91: Can you use a Latin clade name instead of nine-primaried oscines (I think Emberizoidea), easier when looking at the phylogeny (also change in legend for Fig. 1).

Line 113 and paragraph thereafter: This paragraph is super interesting. Can you list a few of the most important clades here, beyond Passeri and Emberizoidea? One or two taxonomic examples to illustrate your point after this would be good too, about the origins of higher taxa from other studies. Again, when discussing the clades that are 'historically recognised on the basis of their phenotypic distinctiveness' – it would make the text more accessible by including specific examples rather than the reader having to scrutinize the figures. Many passerine workers (myself included) will be thankful for this addition.

Again in paragraph starting line 131, there are a lack of examples – it would be helpful to briefly mention which bird clades have been investigated before to compare with passerines. Also, you mention some studies failed to detect early bursts – could you comment on whether passerines might just have different diversification dynamics to other bird groups? This seems very plausible, considering they are the most diverse group that managed to do it in the last ~45 million years, as opposed to groups of birds that have been around since the Paleocene. The Navalón et al 2022 paper might be appropriate to cite here.

The following paragraph could also benefit from being more specific about which ecosystems were restructured or gave rise to more ecological opportunity; was it the opening of savannahs and grasslands? Etc.

Paragraph starting line 158: It might be something obvious, but it would be helpful to state to what degree passerines specifically reflect the latitudinal species diversity gradient.

Lines 191-192: The reference I mentioned above, Brinkworth et al 2023, might be appropriate to cite here.

Line 254: either change avian for passerine, or delete the word.

Conclusion: see my paragraph near the start about shifting the narrative.

The only issue with figures I have is the colours for the residual map in 3.C are very similar, is it possible to increase the contrast between negative and positive? I know there is not much variability on the map but it's even harder to see this when the colours are pale.

Methods

Overall, the methods are very comprehensive and an extensive analytical suite was applied to the data to investigate evolutionary trends, including some novel methods which will hopefully be applicable for other datasets and taxonomic

groups. The methods and individual steps are also extensively illustrated in the supplementary figures, which is helpful.

Line 423: Would be helpful to list the 12 skeletal elements directly.

Can you be a bit clearer about the body mass data? You say body mass was estimated from voucher metadata (line 429) but why was it estimated? Were multiple masses averaged over a sample from one species? Were any of these estimates cross-checked with data from AvoNet? I also don't think saying 'estimated from voucher metadata' is a very accessible way to write it, I was scratching my head for a while. Perhaps simplify by saying something like 'estimated from associated specimen data'.

In line 449 I would just stick with Eocene because early Eocene is a bit ambiguous, when you look at the chronostratigraphic chart the Eocene is split into unequal sections, and your quoted time frame goes between two sections of the Eocene. Also, you could cite the following preprint that uses the most comprehensive fossil calibration scheme for passerines to date for the crown age:

Luo A, Nguyen JM, Zhou QS, Zhu CD, Ho SY. An Eocene Origin of Passerine Birds Estimated Using Bayesian Tip Dating with Fossil Occurrences. bioRxiv. 2025 May 1:2025-04.

Line 498: I can't find the spelled-out definition of GIC or BIC anywhere in the methods.

Referencing: Some of the reference formatting in the bibliography needs some work.

R Code: The R code is well documented and annotated and is available in GitHub.

Reviewed by Elizabeth Steell

Reviewer #3 (Remarks on code availability):

The code is extremely well annotated and well presented. There is a comprehensive readme file with instructions on how to run it on different systems. All the custom functions are provided in a separate script to the analytical script. Very well organised and ready to be published. I did not install and run the code because I have already spent a lot of time going through this extensive manuscript and writing my fairly lengthy review, and the analyses are the most extensive part of this work.

**********END**********

Version 1:

Decision Letter:

13th January 2026

*Please ensure you delete the link to your author homepage in this e-mail if you wish to forward it to your co-authors.

Dear Dr Berv,

Your manuscript entitled "Rates of avian body plan evolution in space and time" has now been seen by the same three reviewers, whose comments are attached. It's clear that we're very close to being able to accept the paper for publication, but before we do, we would encourage you to undertake the additional recommendations made by reviewer 1, and please note that a condition of publication will be revision of the title and abstract to focus on passerines (which as the reviewer points out, is still no mean feat). We'll send further formatting checks and recommendations after we receive your next version.

We therefore invite you to revise your manuscript taking into account all reviewer and editor comments. Please highlight all changes in the manuscript text file [OPTIONAL: in Microsoft Word format].

We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

When revising your manuscript:

* Include a "Response to reviewers" document detailing, point-by-point, how you addressed each reviewer comment. If no action was taken to address a point, you must provide a compelling argument. This response will be sent back to the reviewers along with the revised manuscript.

* If you have not done so already please begin to revise your manuscript so that it conforms to our Article format instructions

at http://www.nature.com/natecolevol/info/final-submission. Refer also to any guidelines provided in this letter.

* Include a revised version of any required reporting checklist. It will be available to referees (and, potentially, statisticians) to aid in their evaluation if the manuscript goes back for peer review. A revised checklist is essential for re-review of the paper.

* Extended Data Figures - please ensure that any supplementary figures and tables that are crucial to the manuscript's conclusions are converted into Extended Data figures and tables to increase visibility of these data. Extended Data figures and tables are online-only (present in the online PDF and full-text HTML versions of the paper), peer-reviewed display items that provide essential background to the article but are not included in the main article due to space constraints. A maximum of ten Extended Data display items (figures and tables) is permitted.

Please use the link below to submit your revised manuscript and related files:

Link Redacted

**Note:** This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

We hope to receive your revised manuscript within four to eight weeks. If you cannot send it within this time, please let us know. We will be happy to consider your revision so long as nothing similar has been accepted for publication at Nature Ecology & Evolution or published elsewhere.

Nature Ecology & Evolution is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' on published papers create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on 'Modify my Springer Nature account'. For more information please visit please visit <a href="http://www.springernature.com/orcid">www.springernature.com/orcid</a>.

Please do not hesitate to contact me if you have any questions or would like to discuss these revisions further.

We look forward to seeing the revised manuscript and thank you for the opportunity to review your work.

[redacted]

Reviewer expertise:

as before

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

I have attached a detailed PDF response to this review.

The authors have, in response to a previous round of review, introduced new language to their manuscript, redacted a figure illustrating a geographic estimate of speciation rate, and undertaken meaningful supplementary analyses to explore reviewer comments.

There are a couple of methodological points which I am still skeptical about, but I am not going to push for the execution of new code.

I did think about a different way of plotting the geographic association of species' recent evolutionary rates. This would not require new code, it would avoid the need for a range-size correction, be insensitive to short-term changes in species' distributions, and I think it would be able to reveal patterns that the current choice to aggregate results into grid-cells elides. For example, it would be possible to identify regions which have both high and low rates of phenotypic evolution.

I feel the authors should accept previous reviewer suggestions to amend the title of the work to avoid the impression that the findings are applicable to all birds. The achievement of producing a general study of passerines is already considerable and I do not believe that the work's exclusive study of passerines will reduce the motivation that other scientists will have to generalize the authors' greedy heuristic approach to other systems.

Reviewer #1 (Remarks on code availability):

I reviewed the code in the first round of review, and did not re-review it this time, so I have selected 'No' to the question 'did you review the code?'. The authors have added supplementary analyses to the work, but have not changed their core body of analyses.
I wrote a simple version of the code for one of the approaches the authors implemented in their supplementary analyses in the last round of review, so I did not feel it was necessary to review this component of the code. (although I would have if the authors had incorporated any new analyses into the main manuscript).

Reviewer #2 (Remarks to the Author):

I thank the authors for their detailed and professional responses to my review; I think this version of the manuscript is now much clearer. For what it's worth, I also think the authors did a good job of responding to Reviewer 3, though Reviewer 1's comments are beyond my expertise to fully evaluate.

I genuinely have nothing else to add (a rarity for me), and I congratulate the authors on such an interesting, well-designed study.

Reviewer #2 (Remarks on code availability):

From a quick glance, the code looks reproducible and clearly-annotated, but I did not myself try to run it.

Reviewer #3 (Remarks to the Author):

Following revisions, the manuscript is now more accessible and I am happy with the additions and changes to the text, the authors have addressed all my comments satisfactorily. I look forward to seeing the work published soon and will cite it many times in future!

**********END**********

Version 2:

Decision Letter:

10th March 2026

Dear Dr. Berv,

Thank you for submitting your revised manuscript "Rates of passerine body plan evolution in time and space" (NATECOLEVOL-25093100B). It has now been seen again by Reviewer #1 and their comments are below. This reviewer find that the paper has improved in revision, and therefore we'll be happy in principle to publish it in Nature Ecology & Evolution, pending minor revisions to satisfy the reviewers' final requests and to comply with our editorial and formatting guidelines.

If the current version of your manuscript is in a PDF format, please email us a copy of the file in an editable format (Microsoft Word or LaTex)-- we can not proceed with PDFs at this stage.

If you have not done so already, please ensure that you also email us a completed copy of the Reporting summary :

Reporting summary: <a href="https://www.nature.com/documents/nr-reporting-summary.pdf">https://www.nature.com/documents/nr-reporting-summary.pdf</a>

We are now performing detailed checks on your paper and will send you a checklist detailing our editorial and formatting requirements in about a week. Please do not upload the final materials and make any revisions until you receive this additional information from us.

Thank you again for your interest in Nature Ecology & Evolution. Please do not hesitate to contact me if you have any questions.

[redacted]

Reviewer #1 (Remarks to the Author):

I have undertaken a detailed review of the revised work and believe all previous reviewer criticisms and suggestions have been explored.
I have attached a pdf document which details my appraisal of the authors' responses and changes and which in turn synthesizes their comments with the wider literature to explore future research directions.
I thank the authors for their patience and congeniality.

Reviewer #1 (Remarks on code availability):

I have selected 'no' to the question 'did you review the code', because the changes since the last round of review were mostly language and discussion changes. A new supplementary analysis was undertaken but was not mathematically complex and I trust that it has been performed correctly having reviewed the figures and the described supplementary methods.
I had previously undertaken a partial review during the first round of review.

**Below, please find all reviewer comments, followed immediately by our responses (in bold).**

Reviewer #2 (Remarks to the Author):

This is an incredibly exciting manuscript, using a recently-published dataset of bird skeletal features, a recently-published avian phylogeny, and brand new phylogenetic comparative methods to demonstrate that (passerine) avian morphology is the result of rare, mostly early, mostly phylogenetically-clustered bursts of evolution. Similar results have been found by other groups using other datasets/methodologies, but this is a really nice demonstration of this phenomenon. This manuscript is written to a very high technical standard, and I found the authors' interpretations to be very insightful. This manuscript is also a rare example of a study that opens up many future avenues for high-quality, important scientific discovery; these results, especially coupled with the twinned data and new methodological approach, will be ones that I'm thinking about for quite a while.

I have two general comments and a small number of line comments, all incredibly minor. Otherwise, though, I *greatly* enjoyed reading this paper, and I salute the authors for all of their hard work!

**Thank you for your generous comments and constructive feedback!**

General comments:

I will apologise to the authors if this is addressed somewhere in their supplement, but I am a little concerned that some of these results are an artefact of sampling bias. Specifically, present-day lineages are undersampled (1/3 of extant species compared with presumably nearly all of the early passerine lineages), which might lower the estimates of recent morphological diversification rates. If the authors have addressed this in some way, I would recommend highlighting it slightly more (and if not, I would recommend addressing this).

**We agree that this is an important consideration.**

**While our temporal phylogenetic comparative analyses rely on a phylogeny that includes ~2,000 of ~6,000 passerine species, the sample spans ~89% of recognized families, making it generally representative of Passeriformes. However, we agree that missing species, especially among recently diverged lineages, may reduce the number of short terminal branches and thus lead to conservative estimates of phenotypic rates, potentially missing some very recent shifts. Importantly, all the major shifts we detect with strong support subtend larger clades where lineages are well represented, so our core inference of rare, hierarchically nested shifts is unlikely to be an artefact of undersampling tips. Further, we previously described a small sensitivity analysis in the Supplement, in which we excluded 101 taxa that had been manually grafted into the Claramunt et al. (2025) topology. The results were essentially identical to those of our full analysis. We had not previously linked this to the question of the impacts of incomplete sampling, but we do feel that it is relevant, as it suggests that the temporal patterns we recover are somewhat robust to sampling. We have revised the text within the Discussion to better explain why we think incomplete sampling is unlikely to affect our main results (Lines 147-150):**

**"Despite our dense sampling (2,057 species spanning ~89% of passerine families), many recently diverged species remain unsampled, so our inferred pattern of rate variation should be interpreted as a lower bound on the complexity of passerine body-plan diversification (see supplementary sensitivity analysis)."**

**The reviewer's comment is also relevant to our spatial analyses. In that context, we explicitly account for incomplete species sampling by estimating and incorporating a sampling fraction relative to the expected species counts across grid cells as a control variable in all spatial autoregressive lag models. This control variable was not identified as a significant predictor of spatial variation in the rate of phenotype evolution across latitude (e.g., Supplementary Figure 13B). Thus, even after accounting for sampling intensity, the key results of spatial models appear robust to sampling across grid cells (in addition to other possible**

**confounders). We agree that this is an important point and have revised the Discussion to reference these results more prominently (Lines 237-238):**

**"Sampling intensity itself was not a significant predictor of spatial variation in phenotypic rates (e.g., Supplementary Figure 13B)."**

**Broadly, we believe the results we highlight are generally robust to incomplete sampling, though we acknowledge that this is, of course, an important avenue for expanding the dataset in the future. We previously included maps of our sampling regimes in Supplementary Figure 11 and now emphasize that there are exciting opportunities to expand this sampling in the future (Lines 676-678).**

I also would have appreciated – and, again, I apologise if this was somewhere in the supplement and I missed it – a quick summary of what these 12 skeletal traits are. What, morphologically, do they represent? And if possible, how do a reader link these broad, general changes in morphology/'bauplan' to specific ecological functions of individual traits? The argument about increased modularity and functional compartmentalisation is interesting, and likely true, but it would be greatly strengthened by the ability to extract morphological meaning from (some of) these results.

**Thank you for raising this; we completely agree that this information needed to be more prominent. All of the data are linear dimensions of 12 skeletal elements: the lengths of the tibiotarsus, humerus, tarsometatarsus, ulna, radius, keel, carpometacarpus, 2nd digit 1st phalanx, furcula, and femur; the maximum outer diameter of the sclerotic ring, and the length from the back of the skull to the tip of the bill (treating the rhamphotheca as part of the bill when it remains present on the specimen). We have incorporated this information more clearly into the Methods (Lines 452-458). Collectively, these measurements capture the primary axes of avian body-plan variation, including limb proportions, flight and perching apparatus, and cranial/sensory structure, and thus reflect broad morphological organization rather than isolated functional traits. Given their likely influence on nearly every aspect of passerine ecology and physiology, linking individual traits to specific ecological functions is beyond the scope of our analyses. We do agree, however, that more biological interpretation would be a useful addition and have revised the Discussion and supplementary materials to reference new analyses (also in response to requests from reviewer 1) that interrogate how coordinated shifts in multivariate data patterns relate to ideas of shifting modularity and functional compartmentalization (Lines 129-134, 1520-1648).**

Line comments:

Abstract: As far as I know, I don't think citations in the abstract are a norm for this journal? I would suggest removing these.

**Thanks for catching this. We have removed these citations.**

References in general: They currently seem to be at the very end of the manuscript, referring to both the main manuscript and the supplement. The authors might want to consider separating these.

**Thanks for noting this. We have revised the reference formatting in accordance with NEE guidelines.**

L234: The statistical significance of K_{mult} is going to be highly dependent on sample size – how much of this finding is just that avian species richness starts declining around \pm 50?

**Thank you for this helpful comment. In our revision, we now acknowledge that the smaller fraction of significant cells at high northern latitudes may indeed be impacted by smaller local species pools. We have also revised the text to note that the major declines in the proportion of significant cells start closer to ±40°, not ±50°, as we had previously noted; importantly, these are regions where sampling is generally adequate (n ~ 20-30) (Collyer et al. 2022). In our dataset, grid cells with very small sample sizes (n < 10 species) are entirely restricted to the most poleward northern latitudes (> ~68°N). (Lines 271-276):**

**“Further informing these trends, the proportion of grid cells showing significant departures from phylogenetic independence declines around ±40° latitude, consistent with reduced phylogenetic structuring of phenotypic variation at high latitudes (Figure 4D). This apparent weakening of evidence for phylogenetic signal may align with patterns of phenotypic overdispersion and modular organization observed at high latitudes, though part of these declines may reflect limitations of smaller species pools.”**

L234: Also, I would recommend specifying that these K_{mult} values are statistically significant *from zero*, as in theory you could have set up the test to distinguish K_{mult} from some other value. (As a broader point, K_{mult} \neq 0 is a very weak test. Depending on what sort of phylogenetic signal you’re interested in, small, non-zero values of K can indicate truly inconsequential signal, particularly across a large sample of species. In other words, I’m not sure that a binary zero-or-not-zero is at all biologically interesting, or at least not as you’re currently framing this.)

**Thank you for highlighting the need for clarification here. As we understand it, the implementation we use does not actually test $K_{mult}$ against a fixed value (e.g., 0). Rather, we assessed significance within each grid cell using physignal.z (geomorph), which (i) optimizes Pagel’s λ for the observed data and then (ii) evaluates the strength of phylogenetic signal by comparing the observed log-likelihood to a null distribution generated by RRPP permutations of trait values across the tips. This null corresponds to phylogenetic independence (no trait–phylogeny association), so the resulting p-values indicate that the observed trait–phylogeny association is stronger than expected under independence, rather than a direct test of $K_{mult} \neq 0$. The permutation procedure is likelihood-based and is summarized as a Z effect size (Collyer et al. 2022); we show $K_{mult}$ descriptively in Figure 4D. The classification of cells as significant is based on p-values adjusted for multiple testing, under the null hypothesis of independence. We believe this is a useful ‘birds-eye view’ approach because it provides a standardized framework that can be combined with multiple-testing correction.**

**We agree that this needed further clarification and have expanded the Methods section to clarify these details (Lines 802-816). We have also modified Figure 4D and updated the caption to emphasize where species sampling falls below 10 per grid cell, a reasonable caution threshold.**

L430-432: Forgive me if this is in the supplement and I missed it, but what’s the spatial sampling like in this dataset? Any notable geographic biases?

**Thank you for raising this question. We previously included maps of the spatial sampling in the Supplement but did not prominently reference them. We have now revised the text (Lines 147-150, 237-238, 457-458, 676-678) to explicitly direct readers to these figures. Although the dataset is extensive, some geographic gaps remain; however, we account for spatial sampling in all spatial models, and sampling fraction is not a significant predictor of phenotypic rates. We interpret this to suggest that spatial sampling does not strongly influence the patterns we report (please see our response to your previous comment related to sampling for additional detail).**

L436: parenthesis error

**Corrected, thank you.**

Figure 4: I’m fascinated by the regional variation in this figure! (Not an action point. Just an expression of amazement.) (Also, in general, these figures are done to a remarkably high standard. Well done!)

**Thank you! We also agree that these patterns, particularly the biogeographic patterns, are quite remarkable. We plan to explore these patterns in more detail in subsequent work.**

L498: parenthesis error

**Corrected, thank you.**

L762: To be a little pedantic: "…multivariate trait similarity tracks phylogenetic relatedness UNDER BROWNIAN MOTION, while…" . Specifically, under single-rate Brownian motion! There could be strong phylogenetic signal under some other tempo/mode.

**We have changed the language in this section to address your prior comments and now acknowledge this assumption. Line 805.**

Reviewer #2 (Remarks on code availability):

I glanced at the code, but I did not try to run/reproduce it. It seems well-signposted/well-annotated from what I could tell.

**Thank you.**

Reviewer #3 (Remarks to the Author):

This manuscript represents an impressive body of analytical work with a very large and comprehensively sampled dataset. It sets out to answer questions relating to the mode, tempo and spatial aspect of passerine body plan evolution, investigated by applying a suite of analyses on a dataset representing the proportional organisation of the passerine skeleton. I think it does a great job of answering these questions and discussing major evolutionary themes, and the paper will be a much-cited contribution to the literature for passerine workers, ornithologists and the wider community of evolutionary biologists. The figures are beautiful and informative. The methods are detailed and the supplement contains a lot of additional background, including illustrated supplementary results. I do however have some issues with how some things are communicated in the main text. My suggestions are pretty much all about wording, so I think this warrants minor revisions and should be accepted once the concerns are addressed.

**Thank you for your kind words.**

One problem I have with the way the work is communicated is unnecessary vagueness in the interesting results. There is description of the overarching patterns and processes, but not enough specific details for the reader to truly understand the relevance of the work in terms of passerine evolution. There needs to be more examples of clades, environments, ecological factors, so that we can understand some of the key takeaway points in more context. See my points below for the paragraphs that need specific taxonomic or environmental examples. Doing this would improve the accessibility of the paper.

**Thank you for these thoughtful comments. We agree that adding more examples will help orient readers, and we have made numerous changes throughout the manuscript to do this (detailed below). At the same time, we have tried to avoid couching this too specifically in our system, as a goal in our presentation is to appeal to the broadest possible audience and emphasize general macroevolutionary principles: episodic shifts, modularity, and links between climatic variability and phenotypic tempo. Because the framework we develop is also intended to be broadly applicable beyond Passeriformes, we have been careful to present results at a conceptual level accessible to a wide macroevolutionary audience. We believe the targeted additions we have made strike an appropriate balance between specificity and generality while preserving the broader appeal of the work. Thank you again for the suggestion.**

I think the narrative should be shifted from avian body plan evolution to passerine body plan evolution – I don't agree that passerines are representative of all living birds. On the contrary, they are a peculiar clade – no other group across tetrapods diversified to the same extent in such a short time span, and there has to be something unique (which we currently don't understand well) about passerines that enabled them to diversify like this. Also, passerines are remarkably uniform on a whole skeleton level (see Navalón et al. 2022 which should definitely be cited, is highly relevant for multiple aspects of this paper, but don't think it is cited: see citation below), whereas other bird groups are remarkably varied in their skeletal organisation and proportions. Passerines are incredible and deserve more

attention on such a broad scale like this. They should be highlighted as something extraordinary in evolutionary terms, rather than being used as a model system to explain evolution across all bird groups. It's rare to have a paper exploring passerine morphology so thoroughly with a third of their 6,000 living species represented, and it's great that they are finally being put in the evolutionary and morphological spotlight. I suggest adapting the title to say 'passerine body plan evolution' instead of 'avian', and adjust the wording in the abstract, beginning of the introduction and the conclusion section. If you want the title to sound more enticing, then change it to 'Rates of body plan evolution in space and time across the most diverse avian order' or something similar.

Navalón G, Bjarnason A, Griffiths E, Benson RB. Environmental signal in the evolutionary diversification of bird skeletons. Nature. 2022 Nov 10;611(7935):306-11.

Along a similar vein, I don't think this paper has been cited either and it probably should be as very relevant for passerine skeletal organisation too:

Brinkworth, A., Green, E., Li, Y., Oyston, J., Ruta, M., & Wills, M. A. (2023). Bird clades with less complex appendicular skeletons tend to have higher species richness. Nature Communications, 14(1), 5817.

**Thank you for raising these points. We agree that the previous version did not sufficiently situate passerines within the broader context of birds or highlight them as an evolutionarily distinctive group in their own right. In response, we have made a significant revision to the Introduction, adding more discussion of Passeriformes as a group, including their exceptional species richness, temporal dynamics, and distinctive skeletal organization (Lines 69-83). In this context, we now cite Navalón et al. (2022) and Brinkworth et al. (2023), noting that passerines combine relatively conservative skeletal disparity and simple appendicular skeletons with exceptionally high species richness. Further, we clarify in the Abstract and early in the Main text more explicitly and prominently that all empirical analyses focus on Passeriformes (e.g., the first sentence of the Abstract and the opening of the "Empirical dataset and phylogenetic framework" section, Lines 588-590), to avoid any risk of misinterpretation. With these changes and the stronger emphasis on passerines' unique evolutionary characteristics, we hope the revised framing addresses your concerns while also preserving the broader framing of the study.**

**We agree that there is potential value in revising the title to focus on passerines, but we also see some costs to this change and would prefer to retain our original title. First, we agree that passerines are not fully representative of all birds, and we agree that their own history is exciting, important, and deserves more work focused on them as a group. At the same time, our primary goal in this work is to use this densely sampled, globally distributed radiation to address general questions about body plan evolution, climatic variability, and evolvability. To that end, we are concerned that a passerine-specific title might discourage engagement from readers working on other clades who are interested in macroevolutionary principles. We view this more general framing as complementary to other recent studies in this journal (e.g., Cooney et al. 2022, Nat. Ecol. Evol., https://doi.org/10.1038/s41559-022-01714-1).**

Line 32: Very pedantic, but I suggest changing 'clade' to 'group' or saying 'order-level clade' because one could argue the most diverse clade of living birds is Neornithes…

**We have changed this instance to 'order".**

Somewhere in the first few paragraphs, you should be more explicit about the dataset. I.e., briefly explain you measured the longest length of 12 ecologically important bones to generate a mass-corrected body plan profile for each species. I had to read the methods and then open the published dataset paper to understand exactly what kind of raw data you analysed, and I think this information should be presented more clearly in the first paragraphs of the main text. Especially considering my next point about body shape.

**Thank you for raising this point. We agree and have revised the main text to provide a description of the elements (Lines 63-68) and have added the full details at the start of the methods (to avoid breaking the flow of the main body with the details while still providing them). Also, Lines 449-455.**

Line 75: I don't think body shape is the ideal example here, because the data is longest length measurements of individual bones, so saying 'shape' feels misleading and I think should be reserved for specific shape analysis (e.g. landmark morphometrics). Also, several components are missing such as the vertebral column which is an integral part of overall body shape. I would remove the example in parentheses and just keep it at proportional organisation of the skeleton, or perhaps skeletal architecture? But swap 'body shape' for 'body plan' or some alternative (architecture, proportional organisation) throughout the main text. Otherwise the data part feels ambiguous and one might think the data consists of 3D GMM, for example. E.g., body shape is also used in lines 128, 172, 204 and should be changed.

**We agree that "body shape" may be a bit ambiguous; we have therefore replaced this phrasing with "body plan," "skeletal architecture," or "proportional organization" where appropriate.**

Line 91: Can you use a Latin clade name instead of nine-primaried oscines (I think Emberizoidea), easier when looking at the phylogeny (also change in legend for Fig. 1).

**We have revised the text to define this term when it is first used: "nine-primaried oscines (Emberizoidea and Fringillidae)". Now that it is defined in the text and given its broad usage, we have retained it without a second definition in the figure legend.**

Line 113 and paragraph thereafter: This paragraph is super interesting. Can you list a few of the most important clades here, beyond Passeri and Emberizoidea? One or two taxonomic examples to illustrate your point after this would be good too, about the origins of higher taxa from other studies. Again, when discussing the clades that are 'historically recognised on the basis of their phenotypic distinctiveness' – it would make the text more accessible by including specific examples rather than the reader having to scrutinize the figures. Many passerine workers (myself included) will be thankful for this addition.

**Thank you for this suggestion. We provide a full list of all ranks that our method identifies as significant 'rate increases' (Supplementary Table 7) and refer to this table in this section. We attempted to incorporate this full list into the main text but felt its scope did not align with the rest of the manuscript (i.e., we believed it would not be of interest to readers who are not focused on passerines). However, we do agree with the reviewer in general, so we also previously included a more detailed set of examples in the Figure 1 caption, e.g.: "These rate increases align with recognized taxa, for example, Passerida (C), Nine-Primaried Oscines (J), Furnariida (Q, L), Corvinae (G), Cyanocoracinae (I), the 'core' Cotingas (K), Acanthizidae (F), and others (Supplementary Table 7)." We think this is effective because it both provides a list of exemplars and prominently points readers to the complete table.**

Again in paragraph starting line 131, there are a lack of examples – it would be helpful to briefly mention which bird clades have been investigated before to compare with passerines. Also, you mention some studies failed to detect early bursts – could you comment on whether passerines might just have different diversification dynamics to other bird groups? This seems very plausible, considering they are the most diverse group that managed to do it in the last ~45 million years, as opposed to groups of birds that have been around since the Paleocene. The Navalón et al 2022 paper might be appropriate to cite here.

**Thank you for this very helpful suggestion. We agree that adding a few concrete examples here will help provide context, and we have revised the paragraph to explicitly note several well-studied avian radiations (e.g., Darwin's finches, Hawaiian honeycreepers, Madagascan vangas, Himalayan songbirds) to illustrate cases where early pulses of diversification have been inferred (Lines 153-165). We also appreciate the reviewer's point that passerines may differ in their diversification dynamics relative to other avian groups. Because our analyses focus on body-plan evolution within Passeriformes, we are not able to attempt a formal cross-order comparison; we do think this is an exciting avenue for future work.**

**However, we now explicitly acknowledge broader-scale skeletal analyses, such as Navalón et al. (2022), which find that passerines exhibit relatively low skeletal disparity compared to other bird clades. We note that this pattern is complementary to our findings and illustrates how the taxonomic scope of an analysis can influence which aspects of diversification are most apparent: while Navalón et al. portray passerines as morphologically conservative at a pan-avian scale, our densely sampled, within-Passeriformes analysis shows that passerine skeletal diversity has been assembled through rare, hierarchically nested changes in body-plan architecture rather than uniformly gradual divergence. Navalón et al.'s broad but sparse passerine sampling is therefore methodologically and taxonomically complementary to our more focused and densely sampled approach. In response to this and related comments, we have updated the paper's framing, particularly in the Introduction, to make this contrast explicit and to more clearly situate our results within existing work on avian diversification. Lines 69-84, 138-150.**

The following paragraph could also benefit from being more specific about which ecosystems were restructured or gave rise to more ecological opportunity; was it the opening of savannahs and grasslands? Etc.

**Thank you for this thoughtful point. We agree that providing more ecological context for the EOT and MMCO/MMCT will help readers interpret these patterns. We have substantially revised these sections to note the kinds of ecosystem changes documented during these intervals (e.g., contraction and turnover of warm forests and expansion of more open, seasonal habitats), which would have altered resource distributions and opened new ecological opportunities (Lines 165-187). A detailed synthesis of the specific biotic responses across all regions and environments is beyond the scope of our study, but we now make the nature of the ecological restructuring that likely underpinned the opportunities we discuss much more explicit.**

Paragraph starting line 158: It might be something obvious, but it would be helpful to state to what degree passerines specifically reflect the latitudinal species diversity gradient.

**We agree that this point needed to be emphasized more clearly and have revised the text to explicitly state that passerines follow the latitudinal diversity gradient, referencing Figure 1E (Lines 196-198).**

Lines 191-192: The reference I mentioned above, Brinkworth et al 2023, might be appropriate to cite here.

**Thank you for this suggestion. Brinkworth et al. (2023) provide interesting evidence that appendicular skeletal organization is linked to ecological specialization and species richness across bird clades, and highlight passerines as an extreme case of high diversity coupled with relatively simple limb architecture. Because the paragraph in question focuses on spatial variation in phenotypic rates and temperature seasonality, we felt this reference was a better fit for the Introduction, where we now cite it when describing the distinctive combination of skeletal organization and species richness in passerines (Lines 69-83).**

Line 254: either change avian for passerine, or delete the word.

**As outlined above, in our revisions, we tried to walk a fine line between providing more details about passerines while maintaining the broad applicability of the theoretical findings; in the course of those revisions, this ended up being an instance where we kept the framing broad, as this sentence intentionally suggests that our results may be generalizable to other avian clades.**

Conclusion: see my paragraph near the start about shifting the narrative.

**Thank you for this comment. In line with your prior suggestions, as outlined in more detail above, we have adjusted the manuscript's framing to better emphasize passerines and highlight their distinctive evolutionary and skeletal context (e.g., Lines 69-83). We have also adjusted the Conclusion to note explicitly that our**

**empirical findings concern passerine bauplan evolution, while acknowledging that the climatic and macroevolutionary patterns we uncover may have broader relevance for avian body-plan evolution.**

The only issue with figures I have is the colours for the residual map in 3.C are very similar, is it possible to increase the contrast between negative and positive? I know there is not much variability on the map but it's even harder to see this when the colours are pale.

**Thank you for this suggestion. The main conclusion from panel 3C is that there is only modest residual variation after fitting the aggregated SAR-Lag model: most residuals fall between ±0.25, with very few grid cells in the tails. Because we use a symmetric color scale centered on zero and the same palette as the other panels, most cells are mapped to similar light mid-range colors, and increasing the map's contrast tended to exaggerate very small differences. Instead, to better show the rarity and magnitude of large residuals, we now plot the accompanying histogram on a pseudo-logarithmic y-axis ("Counts (pseudo-log)"), which enhances the visibility of tail values while preserving the underlying residual distribution. We have also better clarified the interpretation of residual variation in the text; predicted log-rates span roughly 1.3 units on the natural log scale (about –8.1 to –6.8), corresponding to a ~3–4-fold range in evolutionary rates across grid cells, whereas the 95% highest-density interval for the residuals (±0.25) corresponds to deviations of only about 25% from the predicted rates for 95% of grid cells. This simple check is consistent with the Nagelkerke pseudo-$R^2$ of 0.74 that we previously reported for the aggregated SAR-Lag model.**

Methods

Overall, the methods are very comprehensive and an extensive analytical suite was applied to the data to investigate evolutionary trends, including some novel methods which will hopefully be applicable for other datasets and taxonomic groups. The methods and individual steps are also extensively illustrated in the supplementary figures, which is helpful.

**Thank you for these positive comments about our methods. We have aimed to make the analytical framework fully reproducible and broadly applicable, and to provide all code so that the approach can be readily translated to other datasets and taxonomic groups.**

Line 423: Would be helpful to list the 12 skeletal elements directly.

**Thank you, we now include detailed information on what these elements are.**

Can you be a bit clearer about the body mass data? You say body mass was estimated from voucher metadata (line 429) but why was it estimated? Were multiple masses averaged over a sample from one species? Were any of these estimates cross-checked with data from AvoNet? I also don't think saying 'estimated from voucher metadata' is a very accessible way to write it, I was scratching my head for a while. Perhaps simplify by saying something like 'estimated from associated specimen data'.

**Thank you for highlighting this ambiguity. We agree that our original description was unclear and have revised the Methods to clarify how body mass was obtained and why we refer to it as "estimated". In our dataset, 54.2% of individual skeletons (~70% of species) have a body mass recorded on their museum labels (archived in VertNet); for the remaining individuals, we imputed body mass using the same multivariate phylogenetic imputation framework described in Weeks et al. (2025) and implemented in the R package Rphylopars, which we use to leverage correlations among skeletal traits and the avian phylogeny. Cross-validation against known specimen masses indicates very low error (mean RMSE ≈ 0.16 log-grams; mean p-bias ≈ 0.45%) with this approach. Because our analyses are conducted at the species level, we then use the species means returned by our approach (Weeks et al. 2025) as our body-mass covariate; these are model-based predicted means (i.e., best linear unbiased predictions) that integrate information from all available individual measurements (measured and imputed) and account for both within-species variance and among-trait covariances, rather than simple arithmetic averages. We did not perform an additional cross-check**

**against AVONET, as our goal was to obtain internally consistent mass estimates for the same specimens used in our analyses. In the Methods, we now clarify this procedure (Lines 459-471).**

In line 449 I would just stick with Eocene because early Eocene is a bit ambiguous, when you look at the chronostratigraphic chart the Eocene is split into unequal sections, and your quoted time frame goes between two sections of the Eocene. Also, you could cite the following preprint that uses the most comprehensive fossil calibration scheme for passerines to date for the crown age:

**We deleted 'early' and we will add this citation if it is accepted while our paper is under review.**

Luo A, Nguyen JM, Zhou QS, Zhu CD, Ho SY. An Eocene Origin of Passerine Birds Estimated Using Bayesian Tip Dating with Fossil Occurrences. bioRxiv. 2025 May 1:2025-04.

Line 498: I can't find the spelled-out definition of GIC or BIC anywhere in the methods.

**We now spell these out in line where they are used in the methods.**

Referencing: Some of the reference formatting in the bibliography needs some work.

**We have revised the references to conform to the journal's formatting guidelines and carefully reviewed them for errors.**

R Code: The R code is well documented and annotated and is available in GitHub.

**Thank you.**

Reviewed by Elizabeth Steell

Reviewer #3 (Remarks on code availability):

The code is extremely well annotated and well presented. There is a comprehensive readme file with instructions on how to run it on different systems. All the custom functions are provided in a separate script to the analytical script. Very well organised and ready to be published. I did not install and run the code because I have already spent a lot of time going through this extensive manuscript and writing my fairly lengthy review, and the analyses are the most extensive part of this work.

**Thank you.**

Reviewer #1 (Remarks to the Author):

I thank the authors for their interesting work which is ambitious in scope and beautifully visualized.
I have undertaken a careful review, and where I have been motivated to ask the authors questions I have tried to justify my intuition with simulations or example analyses.
I am open to the authors refuting all my questions.
I have attached a detailed pdf review document for the authors, including an example R code framework for a suggested method to interrogate the correlation of time-series of estimated evolutionary parameters with global environmental time-series.
I want to thank the authors for their patience in waiting for my review response, and hope that they find my questions useful.

**Thank you for your extremely thorough review of our work. We found your comments on our existing analyses to be constructive, and your proposed novel approaches to several issues extremely interesting. We have attempted to address all your questions, and although we were unable to integrate all the analyses you suggested given the scope of the manuscript, we considered them carefully, thoroughly explored their implications for our findings, and attempted to revise our work to fully take advantage of your feedback. Thank you again for your time and effort. Here, we provide integrated responses to your 'summary' comments as well as to comments included in your expanded PDF.**

I have copied my general summary and main questions below:

The manuscript investigates a dataset of limb skeletal length measurements and body masses collected across >2000 passerine bird species. The authors aim to understand patterns of variation in the rates of trait evolution across the passerine family tree. The authors want to know whether major environmental change explains variation in rates of trait evolution. The authors want to know whether weaker covariance between traits is associated with faster rates of phenotypic change.
The authors use a custom code implementation to explore these questions, and arrive at the conclusion that passerine evolution is 'bursty', with multiple sudden increases in trait evolutionary rates followed by gradual declines. They believe that the strongest bursts are associated with episodes of major environmental change and that weaker covariance between traits is associated with faster rates of trait evolution-- perhaps suggesting that traits which evolve independently of one another are more amenable to selective forces.

**Thank you for highlighting the need for clarification about the 170,000 data points versus ~2,000 species (section 3.1 of the PDF). In all comparative analyses, we work at the species level, using species-mean trait values for ~2,057 passerine species. The ~170,000 measurements reflect individual element measurements across ~14,419 voucher specimens in the Skelevision dataset, which we use to estimate species means and quantify measurement uncertainty. We have clarified these details in the Materials and Methods. (Lines 446-458).**

I am convinced by the authors' claim that there are recurrent bursts in passerine evolution with subsequent slow-downs. I can probably believe that there is a systemically higher rate of speciation in the new world, as opposed to the old world, given that passerines emerged in Gondwanaland and may have already packed niche-space there more tightly. However, it may also be possible there are simply lots of ancient isolated lineages in Gondwanaland and these bias estimates of lineage origination and alter inferred lineage-specific rates of recent evolution.
I am not convinced yet that there is a high-latitude bias for higher lineage origination rates (I would sooner believe this reflects patterns of dispersal and Rapoport's rule). I am not convinced that the equation the authors have employed to estimate recent lineage-specific phenotypic rates is robust, and I propose a way to show this.

**Thank you for your encouragement. We appreciate the importance of improving the robustness of our findings. Please find detailed responses to each of your comments outlined below.**

I need some more clarification about the way covariance matrices were represented across the family tree and the inference of modularity before I feel capable fully assessing the claims made.

**To clarify: our spatial modularity analysis is entirely independent of the proportional ("shared-eigenvector") structure assumed in the global shift-detection model. Those analyses are per grid cell, where we estimate a local trait covariance matrix by fitting an mvBM model to the local phylogenetic subtree (accounting for phylogenetic relatedness under BM), convert it to a correlation matrix, and estimate local modularity using Louvain community detection (detailed in the Methods). This approach allows covariance patterns to vary freely across local assemblages and does not inherit constraints from the global shift model. (Lines 789-801).**

Lineage phenotypic rate estimation:
I believe the authors' chosen equation will cause an over-estimation of phenotypic rate evolution in out-groups relative to diverse niche-packing radiations, even if the current rates of phenotypic evolution are similar in both. I

believe this because there are more recent nodes in well-represented radiations, and those will record the deceleration in rates and receive strong weightings. Outgroups with no recent nodes will appear to retain fast rates because the last available and most strongly weighted nodes are basal to bursty events.

**We thank the reviewer for deeply engaging with our proposed phenotypic rate statistic and for describing this scenario, as it highlights an important conceptual point about what this metric is intended to summarize. Our response to these ideas is necessarily detailed, but we hope it is sufficient to more fully explain our thought process and provide additional commentary. This response also directly addresses section 3.2 of the expanded PDF of the reviewer's notes. For clarity, we restate the statistic as implemented. For each tip, we estimate a *whole-lineage mean phenotypic rate* as:**

$$\bar{\sigma}^2_{\mathrm{WPR}} = \frac{1}{Z} \sum_{i=1}^{L} 2^{-a_i/T} \cdot \bar{\sigma}^2_i$$

**Where:**
$\bar{\sigma}^2_{\mathrm{WPR}}$**: Weighted mean phenotypic rate along the root-to-tip path.**
$\bar{\sigma}^2_i$**: Mean variance at node** $i$**, computed across multiple trait dimensions.**
$a_i$**: Age of node** $i$**, measured as time since the present.**
$T$**: Half-life parameter controlling the rate of decay in weighting.**
$Z$**: Normalization constant ensuring weights sum to 1.**
$L$**: Total number of nodes along the root-to-tip path, including both shift and non-shift nodes.**

$$Z = \sum_{j=1}^{L} 2^{-a_j/T}$$

$Z$ **is a normalization constant ensuring that the exponential weights sum to 1, with** $L$ **denoting the total number of internal nodes along the root-to-tip path, including both shift and non-shift nodes. Thus, if all nodes along a lineage share the same regime,** $\bar{\sigma}^2_{\mathrm{WPR}}$ **is exactly that regime's rate, regardless of how many nodes there are. The normalization by** $Z$ **is deliberate: it removes dependence on node count when the regime is constant, and makes the statistic interpretable as a present-biased mean rate along the lineage, rather than as a cumulative amount of evolution that simply increases with the number of nodes.**

**Conceptually, our goal was not to estimate a strict "instantaneous" rate over an identical and sharply bounded time interval (e.g., "the last 5 Myr") for every lineage. Instead, we intended** $\bar{\sigma}^2_{\mathrm{WPR}}$ **to be a present-biased summary of phenotypic evolution along each lineage's entire regime history. Our thinking is that recent rate regimes should contribute more than ancient ones, but deeper history should still influence the summary to some degree. The exponential kernel** $2^{-a_i/T}$**, together with the within-lineage normalization by** $Z$ **yields this kind of present-biased mean of the regime-specific rates along the path.**

**In this context, the behavior the reviewer describes is actually a central, intended feature of the estimator rather than an unintended bias. Consider the specific scenario in the comment: an early high-rate "burst" regime followed by a later low-rate regime. In a densely branching radiation that has accrued many young nodes in the low-rate regime, the majority of the normalized exponential weight inevitably falls on those recent low-rate nodes, so** $\bar{\sigma}^2_{\mathrm{WPR}}$ **lies very close to the low rate. In a very sparsely splitting outgroup that only recently entered the low-rate regime and has few or no very young nodes in that regime, the same old burst node is still down-weighted in absolute terms by the exponential, but it represents a larger fraction of the normalized weight within that lineage, simply because there are fewer young nodes to dilute its influence. As a result, the outgroup's** $\bar{\sigma}^2_{\mathrm{WPR}}$ **will retain more influence from the historical high-rate regime than the radiation's value—even if their current regimes are similar. This is the behavior we intended.**

**Such outgroup lineages are summarized as having a stronger "legacy" of historically high rates than radiations that have been evolving slowly for longer and have many recent nodes in the low regime. In this way, the metric also incorporates information about the lineage-splitting process, another potentially desirable property. The reviewer's comment is effectively evaluating our statistic against a different target quantity (a synchronized "rate over the same recent 5 Myr" for all lineages), whereas our target quantity is a present-biased whole lineage-history summary. We have significantly amended the text to clarify the goals and expected behavior of this summary statistic (Lines 611-651).**

**In principle, we agree that a fixed-window recent rate is definable. Given a piecewise-constant regime model, one could define for each lineage a rate over the last 5 Myr as the time-weighted average of the regime-specific rates over that final 5 Myr segment of the root-to-tip path. However, in practice, such a quantity is extremely unevenly identifiable across lineages: many lineages have no splits (and often no inferred regime changes) within that window, so there is little information to distinguish their "recent" rate from the rate of their terminal regime. Any practical estimator of a strict 5-window rate must therefore either (i) borrow information across lineages through the same kind of shared multiregime model we already employ, or (ii) assume a constant-rate Brownian model as in contrast-based tip-rate estimators. In this empirical context, we suggest our present-biased whole-lineage estimator is a reasonable compromise: it uses the shared multiregime mvBM framework to integrate over each lineage's inferred rate history, with an explicit recency bias, rather than assuming that an identical time-window rate can be estimated with precision for all terminal branches.**

**We are very grateful for the suggested alternative approach that could, in theory, leverage mvMORPH's existing estim() function. The reviewer suggested that we could partly achieve their suggested estimator by truncating each terminal branch 5 Myr before the present, reconstructing the expected trait value at that point under the fitted model, and then defining a recent rate as the Euclidean distance between the observed tip trait and this reconstructed 5-Myr trait, divided by 5 Myr. Conceptually, this is a very elegant way to target a strict 5-Myr instantaneous rate estimate, and we appreciate the reviewer laying out this logic.**

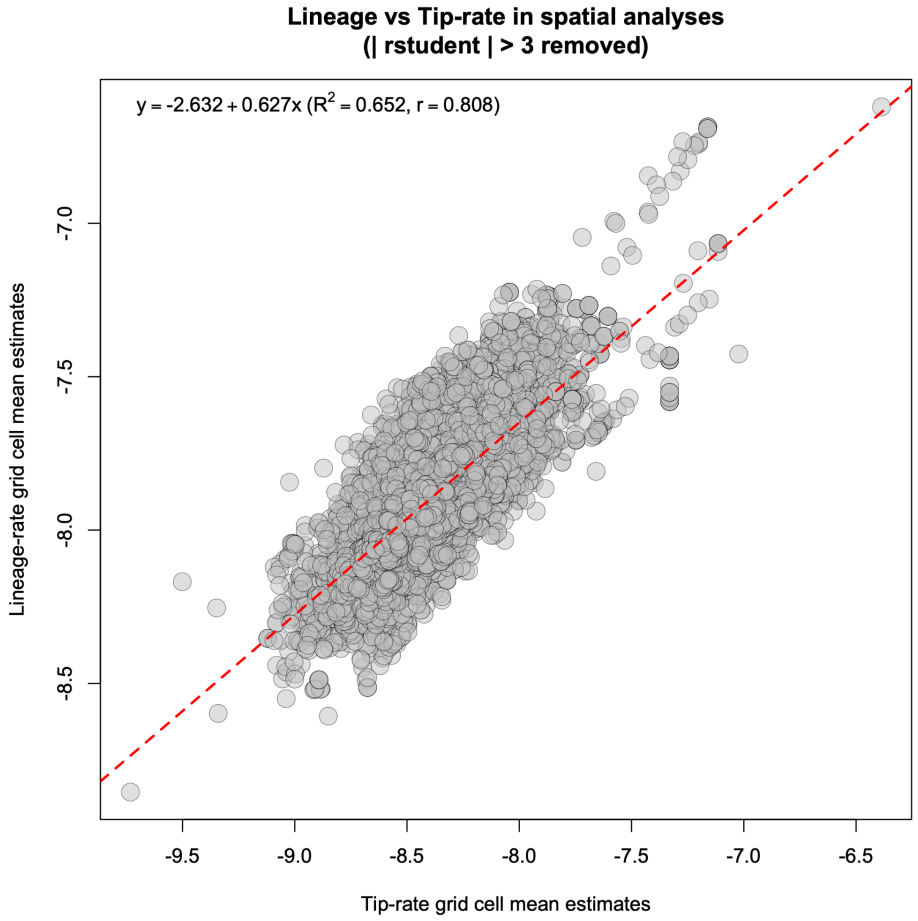
**From a technical standpoint, however, there are practical limitations to implementing this suggestion in our current framework. First, the mvMORPH function estim(), which the reviewer refers to in their derivations, is designed to work with evolutionary model objects of class "mvmorph" (e.g., generated by mvBM() or mvOU()) and stops with an error if the supplied model object is not of that class. Our main multivariate modeling and regime-fitting pipeline for these data is instead based on mvgls() (class "mvgls"), which is not accepted by estim(). Bridging this gap would require either (i) re-fitting an equivalent high-dimensional, multi-regime evolutionary model as an mvBM "mvmorph" object (which, as our early experiments determined, is not feasible with data of this scale and dimensionality), and then constructing a modified tree with pseudo-nodes at 5 Myr along every terminal branch, or (ii) re-implementing the GLS prediction machinery that estim() uses internally, built on the covariance structure from mvgls(), so that we could reconstruct trait values at these 5-Myr horizon points for all terminals. In practice, this would mean writing new code to alter the tree, rebuild the associated covariance matrices, and perform horizon-based GLS reconstructions in a 12-dimensional trait space for >2,000 species. We view this expansion as beyond the scope of this study. We agree that such an approach is appealing and could be a valuable direction for future research, but we do not see how we can adopt it "out of the box" with existing infrastructure.**

**Several additional points help reduce concerns that our proposed lineage-rate summary biases our spatial analyses in significantly problematic ways:**

**<u>1. Model fitting, not the summary statistic, governs whether outgroups "retain" high rates.</u> Our multiregime mvBM model can and does assign low-rate regimes to sparsely sampled clades when the trait data support such shifts. Outgroup lineages are not forced by construction to retain high-rate regimes; if their phenotypic histories indicate true slowdowns, this is reflected in the inferred regime values and therefore in the lineage-rate summary statistic. The extreme case where an outgroup "appears fast" despite currently evolving**

**slowly typically corresponds to situations in which a low-rate regime is not inferred for that lineage at all, which is fundamentally a model-identifiability/data issue rather than an artifact of the exponential weighting.**

**2. Spatial patterns are robust to removing the lineage-history weighting entirely. To ensure that the spatial patterns were not driven by this lineage summary, we repeated our spatial analyses using a much simpler per-tip rate, the rate of the regime currently covering each tip (i.e., a "tip phenotype rate" that ignores ancestral nodes and might behave more like a purely "current" rate). As noted in the initial submission, this tip-level metric is highly correlated with the lineage-rate statistic (r ≈ 0.8) and yields very similar geographic maps (see the plot on the next page). Thus, the latitudinal patterns we report do not depend strongly on the specific way we integrate information along root-to-tip paths, nor on the degree to which dense radiations versus sparse outgroups "record" recent slowdowns (revised text Lines 646-651).**




**Relationship between grid-cell mean lineage-rate and tip-rate estimates used in spatial analyses (log scale). Points show H3 grid cells (|rstudent| > 3 removed); the dashed red line is the ordinary least-squares fit (lineage-rate ~ tip-rate). The two metrics are well correlated (y = −2.632 + 0.627x, $R^2$ = 0.65, r = 0.81), indicating that our spatial analyses are not strongly dependent on whether we summarize phenotypic rates using the lineage-based statistic or a simpler tip-rate measure. We prefer the lineage-based statistic for reasons outlined here.**

**3. Appropriateness for the macroecological question. Our primary goal in applying this approach is to assess how well lineage histories of phenotypic tempo, with a bias toward the present, relate to broad-scale environmental and biogeographic gradients (this is now more clearly articulated, Lines 191-205, 244-249), assuming that present-day trait disparity in a region reflects the accumulated effects of both recent and**

**somewhat deeper rate regimes. In this context, a present-biased lineage summary, such as the one we propose, is arguably more informative than a strict-window estimate of the instantaneous rate, because it partially preserves the influence of older, high-rate episodes while still emphasizing more recent dynamics. The half-life parameter (in our case, T = 5 Myr) sets the temporal scale over which that "memory" decays; lineages that have resided longer in a particular regime are summarized as reflecting that regime more strongly. We acknowledge that the specific choice of the half-life parameter is somewhat arbitrary, but the same type of modeling decision would be required for any fixed time interval chosen to represent 'recent' evolution. Given the robustness we show above, we do not believe that the spatial patterns we infer hinge on this choice. Instead, the lineage-rate statistic should be viewed as a tunable way to emphasize recent evolutionary history in a manner that aligns with the temporal scale of the questions being asked.**

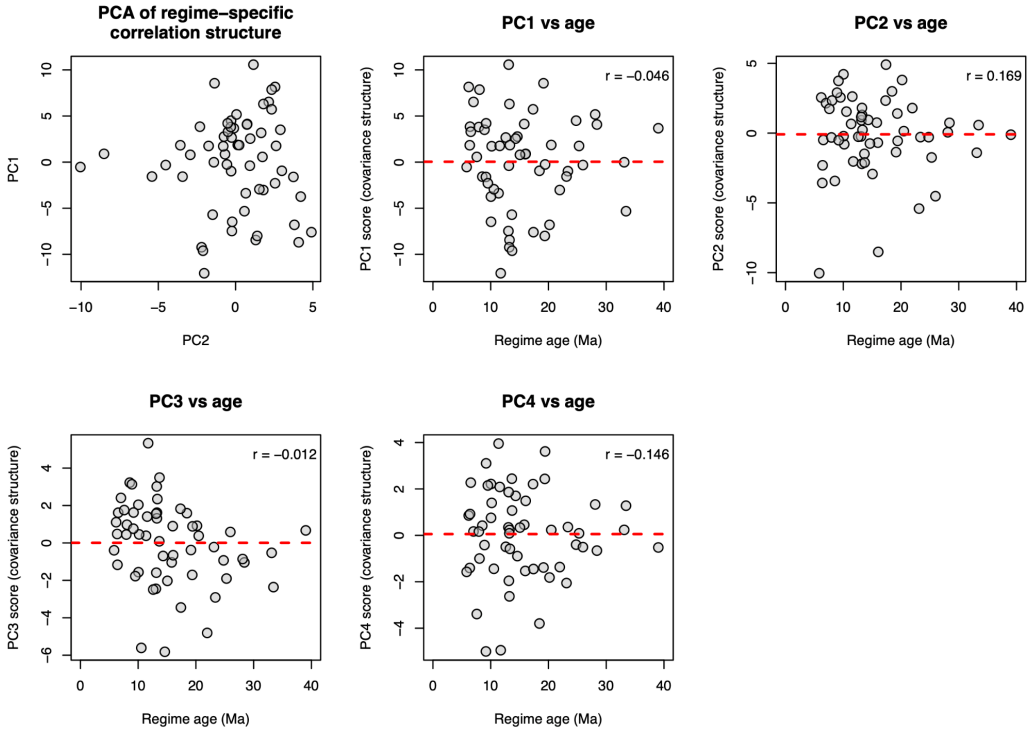
**To avoid misunderstanding, we have clarified the Methods text to state explicitly that our lineage-rate statistic $\bar{\sigma}^2_{\mathrm{WPR}}$ is intended to be a present-biased summary of each lineage's regime history, rather than as an estimate of an identical recent-time-window rate across all lineages, and that its temporal resolution necessarily depends on the density and ages of ancestral nodes along each root-to-tip path (Lines 611-651). We also reiterate that our analyses concern variation in phenotypic rates: $\lambda_{\mathrm{DR}}$ is included only as a control variable in our spatial models, and we do not interpret or comment on spatial patterns in speciation or diversification rates (see additional discussion below in response to your comments on the influence of dispersal).**

Modularity:
Please take the upper triangles of the post-hoc freely estimated sigma-squared matrices, vectorize them, and stack them into a matrix. Please perform PCA on this matrix and review their distribution;
is the clustering of these matrices' properties sympathetic with the temporal distribution of the nodes? Does it emphasize major events? What kinds of changes in covariance are occurring and do they help us better understand the unique phenotypes that have emerged in different parts of the passerine family tree?

**<u>Summary</u>:**
**We agree that it is important to further examine how multivariate covariance structure varies across regimes, in addition to the scalar rate parameters. Though we previously performed a suite of post-hoc analyses on this topic (e.g., Supplementary Figure 4A) to interrogate the meaning of the shift events identified by our search procedure, those earlier analyses summarized only "global" patterns of trait correlation and do not easily reveal the primary axes of morphological integration. Here, we provide extensive additional commentary on these patterns, along with multiple new analyses, in response to the reviewer's summary and to section 3.3 of the expanded PDF notes. While we provide a detailed exploration of these patterns here, our additions to our revised manuscript are necessarily modest. We believe that most of these new results lie outside of our current manuscript's focus and are better explored in detail in future work, though we do add some new discussion of broad patterns in our revision.**

**<u>New analyses</u>:**
**Under our primary search procedure, the covariance structure is shared across the tree for identifiability and computational reasons. As such, the primary inferences in the main text are based on scalar summaries of $\sigma^2$ (e.g., the "BMM" model in mvgls). For the post-hoc analyses presented as supplementary information, however, we refit separate models to a subset of individual regimes (states) identified by our joint search procedure. These post-hoc model fits provide independently estimated variance–covariance matrices for a subset of "terminal" regimes with adequate sample sizes. Following the reviewer's suggestion, we used these post-hoc covariance matrices to further explore variation in the correlation structure (patterns of trait integration) across the regimes resolved by our approach. For these additional analyses, we focus on the primary search results highlighted in the main text (which identified 82 regimes). After filtering out regimes with fewer than 10 tips (to improve signal/noise in the covariance estimates), we obtained 64 regimes for further post-hoc analysis.**

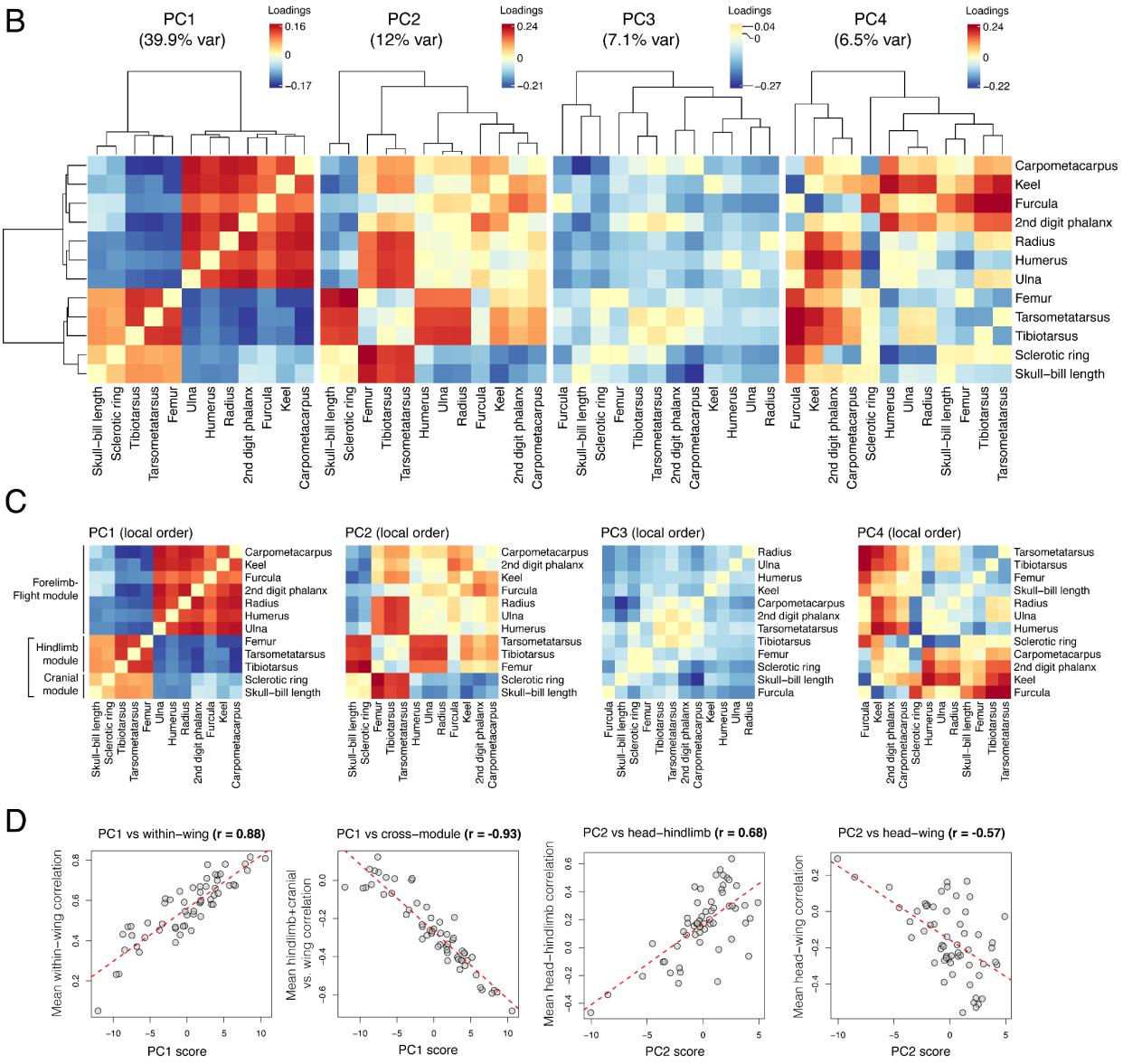
**For each such regime, we extracted the fitted evolutionary covariance matrix, converted it to a correlation matrix (to focus on integration independent of variance scale), vectorized the pairwise trait correlations, stacked these into a regime × correlation matrix, and performed PCA. PC1 explains ~40% of the variance in correlation structure, PC2 ~12%, PC3 ~7%, and PC4 ~6.5%, with remaining variance distributed across higher components. Regimes form a broad, continuous cloud in PC1–PC2 space, and do not show any clustering or relationships with regime age (below). Thus, the distribution of regime-specific covariance structures in this PCA is not obviously related to node ages and does not isolate a set of discrete covariance types associated with temporal events.**




**Diagnostic plots showing principal components of regime-level correlation structure variation and their relationship with regime age. PCA of vectorized correlation matrices estimated for 64 post-hoc regimes; each point is a regime. PC1–PC4 scores are also plotted against regime age (Ma), with least-squares regression lines (red dashed) and Pearson correlation coefficients (r) shown. All correlations with age are negligible, suggesting no temporal trends in the leading axes of correlation structure.**

**Next, at the reviewer's request, we reconstructed each PC's loading vector as a trait × trait matrix to interpret the leading axes in clearer biological terms (new Supplementary Figure 4B-D). PC1 (~40% of variance) clearly separates a hindlimb + cranial set (tibiotarsus, tarsometatarsus, femur, skull–bill length, sclerotic ring) from a distal wing/flight set (carpometacarpus, second digit, radius, ulna, humerus, keel, furcula). Regimes with more positive PC1 scores show stronger correlations within the distal wing/flight apparatus and more negative mean correlations between hindlimb/cranial traits and wing/flight traits, whereas regimes with more negative scores show only moderate within-wing integration and relatively weak cross-module correlations (Supplementary Figure 4B-D). PC1, therefore, captures transitions across a relatively diffuse pattern of integration to one in which the distal wing/flight apparatus behaves as a tightly cohesive module that is negatively coupled to hindlimb and cranial elements. Regimes at the extremes of PC1 are drawn from diverse clades (e.g., ovenbirds and treecreepers, vangas and bushshrikes, New World warblers at the high end; wrens, becards, and honeyeaters at the low end) and do not appear to map clearly onto a single ecological**

**axis such as canopy vs. understory or aerial vs. terrestrial foraging, and thus do not appear to capture a simple ecological syndrome. Instead, they suggest that similar patterns of trait covariation have evolved repeatedly in distantly related passerine clades, reflecting convergence in integration "strategies" rather than in a single ecotype. We visualize these patterns as a series of hierarchically clustered heatmaps generated with the ComplexHeatmap R package (see the next section).**




**Supplementary Figure 4B-C. Trait × trait heatmaps of PC1–PC4 loadings from a PCA of vectorized correlation matrices estimated for 64 post-hoc regimes. The top row (B) shows the PCs in a shared-trait row order (corresponding to PC1 row clustering) with hierarchical clustering dendrograms; the middle row (C) shows the same PCs with rows and columns independently clustered, emphasizing modular structure within each principal axis. Bottom row (D) shows how within and between module correlations vary along PC1 and PC2 (abridged description here). This expanded figure is now integrated into the revised text as an expanded Supplementary Figure 4 (B-C) with an expanded description and caption.**

**PC2 (~12% of variance) reflects correlations involving cranial traits, such that regimes with higher PC2 scores tend to show stronger positive correlations between head morphology (skull–bill length, sclerotic ring) and hindlimb elements (femur, tibiotarsus, tarsometatarsus), and somewhat weaker average correlations between head traits and distal wing/forelimb elements. PC2 thus shifts integration from a head–wing to a head–hindlimb configuration, but the effect is more modest than that of PC1 (e.g., lower r). PC3 (~7%) reflects subtle reweighting of correlations involving the distal autopod ("hand" of the wing: carpometacarpus, second digit); loadings indicate mildly stronger correlations within the distal autopod and hindlimb block, accompanied by more negative loadings on correlations between the distal autopod and hindlimb/cranial traits, consistent with a slight decoupling of the hand from the rest of the skeleton. PC4 (~6.5%) further modulates how axial/pectoral elements—especially the keel and furcula—act as integrative hubs: more positive PC4 values correspond to stronger correlations between keel/furcula and multiple limb elements (hindlimb and distal wing), while other trait pairs contribute less. These axes, therefore, refine how existing hindlimb, distal wing/flight, and cranial/anterior torso groupings are linked, rather than defining entirely new sets of traits that always move together.**

**Collectively, PCs 1–4 show that the correlation structure varies significantly across the regimes our approach identifies, but this variation is multidimensional and continuous. The same broad functional groupings—hindlimb, distal wing/flight apparatus, and cranial/anterior torso—recur across axes, and the main changes involve how strongly these modules are coupled to one another (and, in the case of PC1, whether the distal wing/flight module behaves as a tightly cohesive unit that is negatively coupled to hindlimb/cranial elements).**

**In the Supplementary Text, we previously reported that regimes inferred as "fast" in the joint search tend to have lower overall correlation magnitudes (lower mean absolute off-diagonal correlations, higher effective dimensionality), consistent with a general reduction in global integration in high-rate regimes. The new PCAs of regime-specific correlation matrices are complementary to those results: they confirm that our approach is not only detecting changes in rate (e.g., proportional scaling), but that model shifts also identify a primary axis of shifting integration along which the distal wing/flight module becomes more internally cohesive while trading off with hindlimb and cranial elements (PC1). These correlation differences do not appear to form a few discrete, convergent modular configurations that align with specific events or lineages; rather, they are scattered and relatively subtle, and we do not see obvious clusters in correlation-space that would justify strong claims about modules arising at particular times using these data– their origination likely predates the origin of Passeriformes.**

**We have now added a description of this expanded PCA analysis and our findings to the Supplementary Results (including the lack of strong temporal/clade-based clustering and the interpretation of PC1 as a hindlimb–flight integration axis (Lines 1520-1648). In general, the post-hoc correlation analyses show that high-rate regimes tend to be less globally integrated and that the main integration changes occur along axes modulating hindlimb–forelimb/flight and cranial–wing coupling. These regime-level patterns are also complementary to our existing graph-theoretic analyses of regional modularity (Figure 4A), which show that trait clustering within geographic assemblages varies across space, correlating with estimates of local variation in phenotypic rates. Taken together, we view these independent analyses as informative context for our primary analyses, rather than as a full treatment of modular organization in passerines, which would require additional targeted work and greater skeletal coverage.**

Equal splits:
I would like a stronger consideration of the role of dispersal and milankovic cyclicity mediated extirpations at high and mid-latitudes in setting the apparent lineage division rates in geographic cells. I think the current analysis risks giving the impression that the estimated values in the cells represent endogenous speciation rates and believe they probably do not. I am concerned that many cells with the fastest rates are adjacent to regions with extremely low passerine diversity and therefore may not be well-constrained by the available data in the cells- I point to the eastern Sahara and Canadian Arctic in particular.

**We thank the reviewer for the careful discussion in section 3.4 of their provided PDF and, in particular, for the informative simulation study demonstrating how spatial gradients in Equal Splits–based metrics can arise purely from extirpation and recolonization dynamics even when speciation rate is constant. We fully agree that ES/λDR summaries may be shaped by dispersal, Milankovitch-driven range shifts, and local extirpation, and that when mapped across geographic cells, they should be interpreted carefully. Our aim in using these spatial summaries is partly to test whether latitudinal patterns previously reported for lineage diversification (e.g., Rabosky 2018, Harvey et al 2020) are recapitulated in phenotypic rates, and partly to ground-truth the temporal patterns we recover by asking whether spatial variation in phenotypic rates corresponds to spatial variation in environmental variability.**

**In our manuscript, Equal Splits / λDR is not a focal response variable in any analysis, so we have removed the reference to ES in the description of our proposed lineage-rate metric, as it was too easily conflated with ES-based diversification measures. Our main spatial results and conclusions focus on phenotypic lineage rates obtained from the multivariate shift model. λDR is included only as a control variable in SAR–Lag models, to account for the possibility that our shift procedure (which assigns phenotypic changes at speciation events) might induce some coupling between lineage splitting and phenotypic evolution. We do not interpret the λDR maps as measurements of in-situ speciation probability. In the revised text, we now state this more explicitly (Lines 711-714) and note that refitting the SAR models without λDR produces qualitatively similar latitudinal and seasonal response patterns in phenotypic rates (results not shown).**

**We also agree that there are concerns about species-poor, high-latitude, or desert-margin cells. Our spatial analyses already control for some aspects of assemblage structure (e.g., inverse range-weighted averaging of species-level values (as in Rabosky 2018 https://www.nature.com/articles/s41586-018-0273-1) and explicit richness and sampling covariates in the SAR models). In particular, inverse range-weighting downweights species with very broad geographic ranges—such as those whose distributions extend into desert margins or the far Arctic—so that they contribute less to the cell mean than range-restricted species (note that the range-weighted metrics are more constrained (Supp Figures 9, 10)). Nonetheless, any metric derived from low-diversity cells is necessarily weakly constrained and likely more sensitive to dispersal patterns and extirpation dynamics. More broadly, our intent is to use the spatial maps of phenotypic lineage rates as indices of recent evolutionary dynamics of the lineages currently occupying each cell, i.e., as an approximation of recent evolutionary tempo in those environments. At the same time, as we previously noted in the main text ("Though we do not model the assembly process directly…"), it is true that this framework cannot disentangle where along a species' geographic range phenotypic changes occurred, and dispersal and local extirpation will inevitably blur a strictly local interpretation. In the revised manuscript, we have added a new paragraph describing this caveat (Lines 290-301), and we emphasize that spatially explicit models that jointly incorporate speciation, dispersal, and climate-driven range dynamics will be required to separate recent local evolutionary processes from the effects of range dynamics on the composition and evolutionary histories of assemblages. That said, we believe our results address the primary aims outlined above (e.g., we use these spatial summaries to test whether previously reported latitudinal diversification patterns are recapitulated in phenotypic rates and whether spatial variation in phenotypic rates mirrors spatial variation in environmental variability).**

Time-series correlations:
I have written an example code which shows that the distribution of evolutionary parameters resolved at the nodes on a phylogeny can be discretized over time and meaningfully correlated with a time series such as d18O.
The procedure is as follows:
For each time step, record all the nodes in the interval.
Compute the difference between the parameters of these nodes and their parents, divided by the elapsed times, to obtain the estimated derivative in the parameter value.
Take the phylogenetic mean of these estimated derivatives.
Normalize this vector by its variance, and correlate it with the variance-normalized vector of all d18O observations for which there were valid observations of the nodal parameter values.
I applied this method to a worked example to demonstrate its utility.

**We thank the reviewer for outlining this procedure and for providing example code. The idea of summarizing nodal derivatives in evolutionary parameters over time and correlating those summaries with an environmental time series, such as $\delta^{18}O$, is, to our knowledge, a novel and interesting contribution.**

**In the present manuscript, however, our temporal analyses are framed somewhat differently. Our primary goal is to characterize the occurrence and timing of discrete rate-shift *events* in a piecewise-constant multivariate Brownian model, and to ask whether the *frequency* and *direction* of those shifts align with major episodes of climatic instability. To that end, the main text focuses on three complementary summaries:**

1. **Branchwise rate distribution. We show that inferred rates across branches follow a Gumbel distribution, consistent with a process dominated by rare, high-magnitude shifts (Figure 1A).**

2. **Waiting times between shifts. We show that waiting times along individual tip-to-root paths are Exponential, but across the full tree are best fit by a Weibull distribution, indicating temporal clustering of shift events (Main text, "Historical patterns of body plan evolution").**

3. **Shift frequency vs. climate variability. We summarize the temporal frequency of rate shifts in a sliding window, compare it with a sliding-window measure of $\delta^{18}O$ variability, and ask whether intervals of elevated climatic variability coincide with elevated shift frequencies and whether the direction of shifts differs across those intervals (Figure 2).**

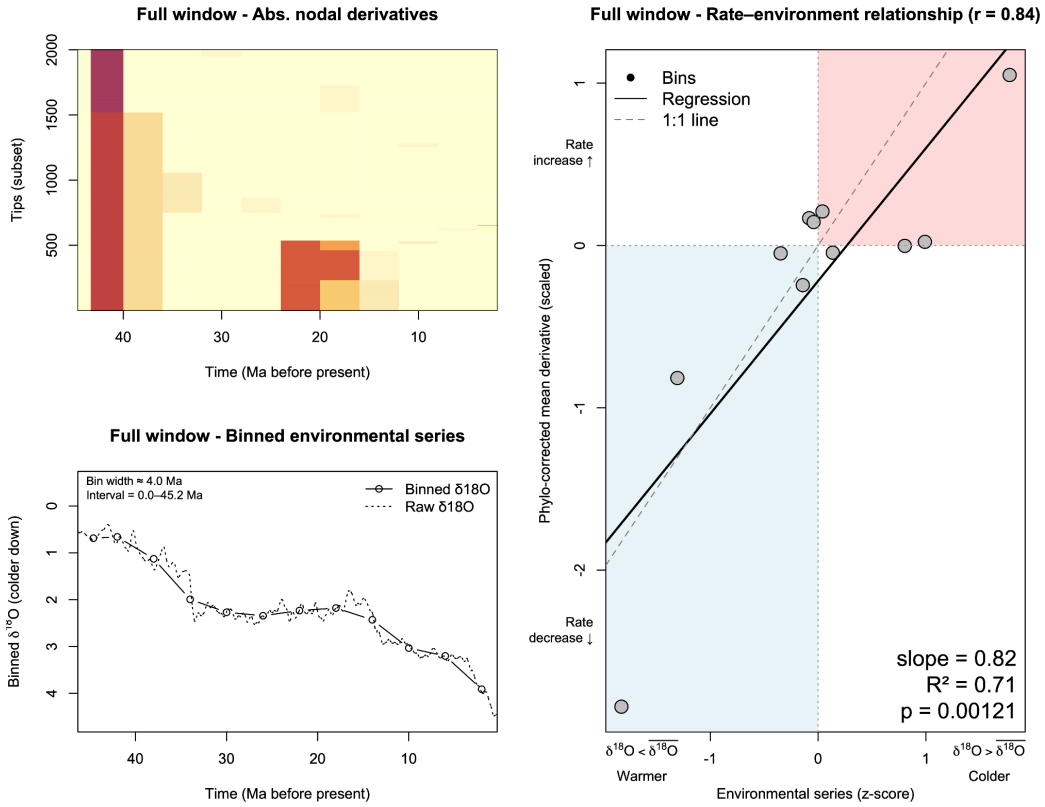
**Together, these analyses support our conclusions that rate variation is dominated by rare bursts, that shift events are temporally clustered, and that peaks in shift frequency align with well-known episodes of climatic upheaval such as the Eocene–Oligocene transition and the Mid-Miocene Climatic Optimum/Transition.**

**By contrast, the procedure suggested by the reviewer focuses on the magnitude of continuous parameter changes between parent and descendant nodes, aggregating node–parent derivatives across the tree at each time slice and then correlating those derivatives with an environmental time series. Applying this approach in our setting requires re-casting our piecewise-constant regime process as a continuous trajectory of nodal rate changes, and it introduces additional modelling choices (e.g., fixed time binning, weighting of shared deep nodes, and treatment of temporal autocorrelation in both the environmental series and the nodal summaries). We agree that this is a promising direction, but we view it as a substantial methodological extension that would be best developed in a dedicated study, ideally with simulations and sensitivity analyses, rather than as a significant addition to an already methods-heavy empirical paper. *We view this type of method as promising for future work and would welcome further development and discussion of these ideas.***

**That said, we did implement a version of the reviewer's idea because we found it interesting! Using our focal model of rate shifts, we (1) discretized time into 2-4-Myr bins (4 Myr shown below) over the 0–45.2 Ma interval, (2) computed node–parent derivatives of the inferred rates, (3) aggregated these derivatives along tip paths within each bin, (4) obtained a phylogenetically corrected mean derivative per bin (following the reviewer's approach), and (5) compared these values to binned Westerhold $\delta^{18}O$ values. This analysis (illustrated on the next page) yields a strong positive association between the mean derivative time series and $\delta^{18}O$ (slope ≈ 0.82, $R^2$ ≈ 0.71, p ≈ 0.0012). In other words, when the climate is cooler than average (higher $\delta^{18}O$ values), mean rate derivatives tend to be positive (net rate increases), whereas warmer-than-average bins tend to show net rate decreases. This node-based analysis, therefore, relates the direction of rate change to the absolute climate state, whereas our sliding-window shift-frequency analysis in our text focuses on the frequency and sign of discrete rate shifts in relation to climatic variability (Figure 2). Taken together, these results are complementary: both indicate that cooling episodes and intervals of heightened climatic variability are associated with accelerations in body-plan evolution, even though they summarize the underlying regime process in different ways. This result is essentially similar to those reported by Clavel (2017; https://www.pnas.org/doi/10.1073/pnas.1606868114) for body-size data.**

**For transparency, we have included our implementation of this derivative-based analysis in an experimental branch of our public GitHub repository https://github.com/jakeberv/passerine-bodyplan-evolution/, branch 'experimental', so that the reviewer can see exactly how we adapted the suggested procedure. Because this analysis is exploratory and rests on additional modelling choices, we have chosen to share it here in our response rather than incorporate it into the main manuscript. Though we considered including it, we are also not sure it would be appropriate to attribute this novel approach to our paper, given that it draws mostly from the reviewer's intellectual contribution.**

**Finally, in light of these comments, we have re-read the manuscript to ensure we do not overstate the strength of the proposed link between $\delta^{18}O$ and our inferred rate shifts. Our main text consistently frames this comparison as a descriptive alignment rather than as a formal statistical test of $\delta^{18}O$ as a predictor of shift magnitude.**




**Time-binned correlation between body-plan rate shifts and the Westerhold $\delta^{18}O$ climate record. Top left: heatmap of absolute nodal rate derivatives across tips and 4-Myr bins (warmer colors = larger absolute change). Bottom left: binned $\delta^{18}O$ series (solid line and points) with the raw curve overlaid (dotted); lower $\delta^{18}O$ indicates warmer conditions. Right: phylogenetically corrected mean rate derivatives (scaled) for each bin plotted against z-scored $\delta^{18}O$. Horizontal and vertical lines mark zero derivative and mean $\delta^{18}O$; blue shading denotes warmer-than-average bins with net rate decreases; red shading denotes colder-than-average bins with net rate increases. The regression (black line) shows a strong positive association (slope = 0.82, $R^2$ = 0.71, p = 0.0012). These patterns are essentially similar to those shown by Clavel (2017; https://www.pnas.org/doi/10.1073/pnas.1606868114) for body-size data.**

Reviewer #1 (Remarks on code availability):

I inspected the coding functions on the GitHub repository.
There was indeed a well-organized ReadMe file.
The code is comprised of multiple inter-dependent functions which total several thousand lines in length.
It was not within my abilities to manually review all code to my satisfaction in the timescale that is available for review.
The dataset is nominally available, but it is associated with another manuscript.
I manually read through the code to look for mistakes instead of re-running the analysis directly. I did not find any clear mistakes.
My intuition is that future authors who might use the same functions will be unlikely to manually read through the code.
I would have appreciated it if the in-line comments in the coding functions, such as 'TemporalAnalyses-functions.R' included tags where GPT models had been employed to construct code, to increase the transparency surrounding the role of Large Language Model contributions.
I would advocate adding a tag at the beginning and end of sections copied from an LLM output, and the inclusion of the prompt text.
I appreciate the authors are unlikely to have systematically recorded this information, though.

**We appreciate the reviewer's careful inspection of our code and their comments about transparency around LLMs. As noted in our Acknowledgements, we did make use of LLMs as programming assistants during this project. In practice, this ranged from small, targeted suggestions (e.g., debugging, syntax) to, in some cases, drafting larger blocks of boilerplate code that we then edited, refactored, and integrated ourselves. All code in the repository has been manually reviewed by the authors, and all modeling choices and algorithms were designed and vetted by humans. We agree in principle that it would be desirable to tag sections of code that originated from LLM outputs and to record the associated prompts. However, our use of these tools was iterative and spread across hundreds of sessions over several years, and we did not systematically track that provenance. As a result, we do not have a way to reconstruct which specific lines or blocks in the current multi-thousand-line codebase stemmed initially from LLM suggestions. Given this reality, our approach to ensuring correctness has focused on manual review, cross-checks, and repeated testing of the code (including extensive unit testing) across multiple analyses. Community standards and tooling for documenting LLM contributions to scientific software are evolving, and we intend to follow best practices.**

In Sum, I was not presented with any reason to believe the coding implementation makes any mistakes.
However, I believe some of the fundamental equations being implemented in the analyses (such as lineage-specific estimates of recent evolutionary rates) are not well-defined.

**Please see our prior detailed responses above.**

As part of my review I wrote new code and conducted simulations, and was able to show that a latitudinal gradient in lineage diversification rates can be produced by biased patterns of dispersal, and would like the authors to consider writing their own simulation to explore their method's performance. My simulation was very simple, and would be improved if species' dispersals were represented by Random-Walks with different variances, and if species' extirpation probabilities were functions of a simulated latitudinal temperature gradient that fluctuates with Milankovic cycles.

**We are very thankful for the time and thought the reviewer invested in our manuscript, including the custom simulations exploring how dispersal and climate-driven turnover might generate spatial gradients in diversification metrics. Those efforts have directly shaped how we now frame our results: in the revision, we more explicitly flag the limitations of interpreting cell-level patterns as purely in situ processes and the potential roles of dispersal and extirpation (in a new paragraph). We fully agree with the reviewer that a process-based simulation framework that jointly models speciation, trait evolution, dispersal, extirpation, and climate (e.g., using random-walk dispersal and Milankovitch-modulated extinction probability) would be**

**an interesting way to assess the performance and interpretability of our approaches. At the same time, we feel that developing, validating, and analyzing such a simulator would constitute a substantial methodological project, which would be difficult to do justice to within the scope of this already extensive empirical study. For that reason, we have chosen to focus our revision on clarifying our empirical claims and approaches in light of the reviewer's insights, and to treat a full simulation-based evaluation of our methods as an important avenue for future work.**

**Author responses in bold.**

The authors have, in response to a previous round of review, introduced new language to their manuscript, redacted a figure illustrating a geographic estimate of speciation rate, and undertaken supplementary analyses to explore reviewer comments. I am, in principle, prepared to see the manuscript published. I remain skeptical about some approaches, but I believe that scientific publishing should accommodate and represent differences of opinions between researchers, so I don't think this should impede progression to publication. I have suggested some possible solutions to problems the authors identified in their response to reviewers. I invite the authors to consider the exploration of these suggested solutions as voluntary. I have also suggested an alternative way of plotting the geographic association of species' recent evolutionary rates- which shouldn't require a correction for range size and which would have an inherently higher dimensionality that isn't accessible when rates are averaged inside grid-cells. I cannot insist on these maps being produced, but would be very interested in knowing what they might look like.

**Author response: We thank the reviewer again for their in-depth engagement with our revised manuscript. We have implemented several of the reviewer's proposed ideas, including the exploration of quantile occupancy maps, and have revised our language in response to other comments. The revisions do not alter our primary narrative, though we have revised some language to be more explicit about possible interpretations of our results (e.g., when referring to cell-level statistics, we now always specify that they are means). While the reviewer noted that the additional suggestions were not necessary for publication, we explored their comments in depth, and these efforts revealed patterns that provide further support and context for our main findings. This is primarily featured in a revised supplementary section (4.2), which includes new text and an additional supplementary figure (Supplementary Figure 17) that addresses the reviewer's final points.**

1 General comments:
I actually think that the authors' most interesting finding relates to George Gaylord Simpson's conception of 'Quantum megaevolution' and 'burstiness', which emerged from his experience as a paleontologist and mammologist– (we can imagine the many faunal turnover pulses in the mammalian fossil record which support this notion). It might therefore seem difficult to investigate this possibility in groups with scant fossil records, such as the passerines, but the authors have used an almost exclusively neontological approach, and have been able to produce beautiful distributions, like their Gumbell and Weibull plots, which show it is possible to perceive this pattern in passerines as a latent feature. This is a very beautiful insight, and I feel that it is potentially overshadowed by the authors' efforts to emphasize the biogeographic elements of the work that are more difficult to defend, or to imply the generality of the work beyond passerines. On this latter point, I accept the authors' intent to encourage the application of their greedy heuristic method beyond passerines, but I don't think this is the same as advertising the manuscript as though its results are applicable to all birds.

The clearest example of this is the authors' preference to retain their original title that appears to advertise an insight general to all birds, when in fact only the passerines have been investigated. I disagree with the authors that using the word 'passerine' will discourage other scientists from attempting to generalize the results- rather it would probably motivate other researchers to perceive the non-inclusion of other types of birds as a knowledge gap that should be addressed. Producing an analysis that generalizes passerine evolution is a massive achievement, so using the word 'passerine' does not undersell what the authors have done.

**Author response: We have revised the title to "Rates of passerine body plan evolution in time and space" to clarify the scope. We also appreciate the reviewer's positive assessment of our findings regarding 'Quantum megaevolution' and 'burstiness' in passerines, and we agree that producing a general analysis of this massive radiation is a significant step forward.**

Similarly, the authors have made some accommodations to reflect the possibility that high-frequency and short duration climate variability have intersected with species dispersal and extinction to determine contemporary bigoegraphic gradients. However, there appears to be an overall willingness to invite the reader to interpret maps of estimated phenotypic evolutionary rates as reliable estimates of phenotypic evolution triggered in or around their projected locations. In my criticism here I am directly using similar language the authors intend to present to their readers: see Line 245:
"We interpret these patterns as evidence that climatic variability has triggered phenotypic innovation both across ≈50 million years of passerine evolution and in more recent history, reflected in strong relationships among seasonal temperature variation, latitude, and recent rates of phenotypic evolution,"

**Author response: We agree that the language in this section was too mechanistic. We have revised the text and now state that the patterns are "consistent with the hypothesis that climatic variability has contributed to phenotypic innovation." We also explicitly reference the new compositional diagnostics (Supplementary Fig. 17) and the caveats regarding dispersal (added to the Discussion) to ensure that readers interpret these maps as approximations of assemblage dynamics rather than precise locations of in situ evolution.**

It is perhaps not possible to distinguish the authors' interpretations from plausible scenarios they highlight, in which elevated extinction and dispersal in unstable boreal climates has resulted in selective samples of species. I am not convinced by efforts to distinguish these scenarios conducted on a grid-cell wise basis, especially in scenarios where grid-cells are populated by a number of species that is too small for me to trust statistical inference of latent parameters.

If possible, I would encourage steering the tilt of the manuscript towards evolutionary 'burstiness' and the value of their greedy heuristic approach, and portraying the potential relationship between biogeography, dispersal and phenotypic evolution as a subject for future study whose value is strongly implicated by nascent observations of correlations that could have multiple mechanistic explanations. Pitching the findings in this way would achieve the authors' stated aims of inviting further work from other research groups.

**Author response: We appreciate this perspective and the connection to Simpsonian 'Quantum megaevolution.' We agree that the recovery of 'burst' dynamics from extant taxa is a critical result. Across both revisions, we have refined the text to emphasize this evolutionary 'burstiness' as a central finding. Regarding the biogeographic patterns, we continue to frame the specific mechanisms (e.g., dispersal vs. in situ evolution) as a subject for future investigation. We have retained the Discussion paragraph explicitly acknowledging that our current spatial framework provides "approximations" of assemblage dynamics and that "spatially explicit mechanistic models... will be valuable for further disentangling these effects." We believe this re-balancing highlights the value of the heuristic approach while remaining appropriately cautious about specific spatial mechanisms.**

2 Methods:

2.1 General comment:

I thank the authors for undertaking new analyses, especially visualizing the patterns of phenotypic covariance found in different subclades. I have no issue with these ancillary analyses, enjoyed reading them and think that they improve the supplementary material. Thank you for making them.

I also reviewed the authors' efforts to relate the derivative of the δ18O record and patterns of evolutionary change. I recognize that their conclusions appear to converge with hypotheses voiced in the wider literature suggesting that cooling events have prompted rapid bird evolution in the Cenozoic. I do note, however, that the salience of this relationship reported in the response to Reviewers looks like it could originate from a small number (2–3) of influential points- a bump in increased rates at the Eocene-Oligocene cooling, and a slump in the Miocene optimum. Presumably these points represent a process that occurs simultaneously across multiple nodes in the phylogeny (as indicated in the main manuscript). I am willing to entertain the relationship.

I want to ask: In Figure 2 of the main manuscript, there is a rapid fall of temperature after the Miocene. This is followed by the time-intervals 'Plei' and 'Plio' in the figure legend. I thought the Pliocene should be before the Pleistocene? Am I missing something? Are these the wrong way around? Please check. If this is a mistake it is worth checking other figures and supplementary axis labels. I am also curious about the fact that there is apparently no contemporary bump in speciation rates in the Pleistocene. Maybe the Earth's climate was already so cold by the Pliocene that further cooling became an obstacle to speciation, rather than a stimulant?

**Author response: Good catch, thank you. This was a typo, and we have corrected it. There are no other impacted figures.**

**With respect to the comment about "no contemporary bump in speciation rates in the Pleistocene," to reiterate, our study does not examine speciation rates as a response variable in any analysis. If the reviewer is referring to the pattern of estimated tree-wide phenotypic shift-frequency (shown in Figure 2), the resolution of this signal is somewhat limited as the sliding window approaches T = 0, because our heuristic search procedure initially draws from a candidate pool of nodes defined by a minimum clade size (10, in our case). Additionally, because our analysis normalizes for lineage sampling within each sliding-window step, we expect that windows near the present will be most heavily down-weighted. It is also unclear whether we should expect an increase in the frequency of such macroevolutionary shift events toward the present, since, as clade size shrinks toward the present, we are inherently less able to detect significant departures from local background data patterns. Either way, we do not present an interpretation of this pattern in the manuscript, which we think reflects a conservative agnosticism about its source.**

2.2 Tip-wise rates:

I recognize that some suggested mathematical methods in previous reviews would require writing code to interface R data objects with different classes and understand the reticence to do this.

The authors consider two possible avenues to do this; either re-fitting their model as a class compatible with mvMORPH's estim function, or re-writing the equations integral to the mvMORPH estim function manually.

My suspicion is that there is a quick work-around using the mvMORPH:::predict.mvgls internal function. However, I have described a more manual solution below: I do not expect the authors to implement this.

***Please note: Additional reviewer comments providing new code and commentary are truncated here for space.***

Anyway, I am fine with all of this provided that the authors present their distribution of tip-wise rates as a data object which possesses 'memory' that is gradually over-written by cladogenesis. I think it is a good idea to more clearly suggest the interpretation to readers that slow inferred tropical tip-rates could reflect a systemically lower long term risk of extinction, leading to tip-rates being depressed by a greater volume of recent evidence of niche-partitioning dynamics.

**Author response: In our revised submission, we have further qualified our estimation of lineage rates as an object that possesses 'memory', as the reviewer suggests. We agree that this is a more appropriate way to portray its expected properties. In our revision, we have also integrated the reviewer's idea of a possible interpretation involving a systematically lower long-term risk of extinction in the tropics, which we agree is an interesting point.**

2.3 On outgroup rate peakiness:
Let's say we wrote a code to simulate the evolution of a trait undergoing a constant exponential decay in its evolution over a family tree. It wouldn't need to be the full passerine family tree- just some convenient family tree which has a similar type of structure. The authors could then apply their greedy heuristic to this simulated dataset, which we know was produced by a single uniform pattern of deceleration and where we have perfect fore-knowledge that all tips have an identical 'tip-rate' of evolution. The only thing that differs between the tips is the distribution of nodes between them and the root. I predict that the exponental kernel assessment of tip-wise rates would highlight high rates in out-groups. Alternatively, I predict the authors would be able to induce 'hot-outgroups' where they currently don't exist by choosing clades which have high basal rates but slow tip rates, and then pruning them of most of their species and re-estimating the rate shifts. I predict the left over species will inherit the basal rate and become hot outgroups.

Do the authors agree with me that this makes intuitive sense and that it would be relatively easy to test? If the greedy-heuristic only takes a short time to run, why not run a simulation like this on the whole passerine tree? Does the profile of 'hot outgroups' look like the hot outgroups in Figure 1? If it does, that's a problem, because it suggests that the tip-rates are being aliased by the node distribution, such that contemporary maps of inferred phenotypic rates are actually telling us more about cladogenesis and extinction than they are about recent evolution.

**Author response: We appreciate the reviewer's intuition regarding potential artifacts in outgroups, but we believe the concern about "rate inheritance" may partially stem from a misunderstanding of the greedy search procedure. The algorithm does not add shifts in any order based on node density (e.g., root-to-tip). Instead, it performs an exhaustive initial scan of all candidate nodes, ranks them by their expected improvement in the Information Criterion (GIC/BIC) relative to the baseline, and then evaluates candidates sequentially from this ranked list. This ensures that the search order is driven by the strength of the statistical signal for each candidate, rather than by tree depth; strong signals in outgroups to large clades are thus isolated early in the process (if the signal in the data identifies them). Furthermore, because all parameters are fully re-estimated conditional on the updated regime structure at every successful step, high rates inferred**

**for outgroups reflect the specific fit of the data to those lineages. They are not a consequence of inheriting parameters from an earlier step in the search. In other words, if an outgroup is placed on a distinct regime due to the order of shift placements, and the model re-estimates that regime as high-rate, then the interpretation is that the empirical data pattern associated with that (potentially depauperate) outgroup lineage requires a high rate, under the model.**

**Unfortunately, the search procedure we use is somewhat impractical to run in a simulation framework for very large trees (e.g., our primary focal analysis required about 7-10 days of HPC time). This guided our prior decision to evaluate model performance on thousands of replicated sub-samples of the passerine phylogeny (e.g., to capture expected variation in tree shape and data patterns). We agree in principle that such a simulation study could be performed. Still, we have elected not to include it in the manuscript (under the reviewer's guidance that it is not necessary).**

Perhaps the authors believe that the node distribution is not a confounding variable because it is itself intrinsically tied to burstiness and cladogenesis, but they explained in their manuscript's introduction that many biologists perceive a 'paradox' in the fact that cladogenesis and phenotypic burstiness might anticorrelate in passerines. (This is, perhaps, not a paradox under George Simpson's Quantum megaevolution description of adaptive radiation, which supposes that phenotypic disparity is produced early, and that taxonomic richness accumulates gradually after phenotypic rates have decelerated). The estimator for recent tip-rates therefore needs to be constructed to be robust to variation in the node distribution caused by the interplay between cladogenesis and extinction etc., if it is meant to inform this debate.

I am bringing this up because if the authors look at Figure 1A of the main manuscript, they will see that there are some outgroups which appear to be constituted by only a few or even a single species. For example, it appears that there is a basal diverging species from Acanthizidae (maybe the Fernwren?) which actually has no nodes in its history more recent than 40 million years ago. This really invites the idea of artifactual signal to me.

If the authors do not wish to investigate any further, I acknowledge that comments such as those at Lines 290-301 Ameliorate some of these problems by identifying their nature to the reader. I feel it would perhaps be best to avoid implying that the authors favor an interpretation (that climate dynamics like annual seasonality are driving 'recent' evolutionary change along contemporary biogeographic gradients), and position the manuscript more as an invitation to others to investigate whether this is true.

The sensation I have is that this is not true at a 'local' scale, and that if there is a pattern being set by glacial-interstadial dynamics, driven by the sculpting effect of extinction, which means that evidence of relatively slow-rate 'niche partitioning' evolution is a lot more likely to be erased at northern latitudes, resulting in the artifactual impression that high latitudes are 'engines' or rapid evolution, when actually the tropics and low-latitude temperate regions could be the real sources of most phenotypic diversity and novel adaptation.

**Author response: We appreciate the reviewer's insights regarding the potential for extinction to sculpt latitudinal patterns. While we maintain that our method effectively identifies signals in the data (as detailed above), we agree that extinction can potentially erode evidence of slow-rate niche partitioning, particularly at northern latitudes. In line with the reviewer's suggestion, we have further revised the text to avoid over-interpreting these patterns as purely 'local' drivers. We now explicitly acknowledge that extinction**

**and dispersal can generate similar spatial signals and frame our results as "approximations of recent evolutionary dynamics" rather than definitive mechanisms.**

2.4 Spatial discretization of evolutionary rates:
Regarding grid-cell wise estimates of phenotypic rate estimation, I remain skeptical of the fundamental precept that rate estimations should be circumscribed by species distributions at fine spatial scales such as these– especially when those grid-cell's communities would have looked very different 10,000 years ago, 20,000 years ago, etc. To give the authors a contemporary example, in the 20th century the collared dove Streptopelia decaocto expanded its distribution across Europe. This would mean all the evolutionary rate estimates in the grid-cells changed as the collared dove arrived, even though nothing 'macro-evolutionary' has actually happened. Even more strangely, because the collared dove's range size would change, the inferred evolutionary rates would change in any cell it was present in as a consequence of the range-size weighting function, even if the communities in those cells were completely the same as before. When the number of species in a grid cell is small, changes like this might be very important.

The choice to plot grid-cell wise estimates on a map, and contrast them with variables like seasonality, inherently invites the readership to draw mechanistic links between them- when the authors explained good reasons on page 6 to expect climatic volatility and spatial autocorrelation to constitute a suite of lurking variables that could perhaps better explain the instantaneous presence or absence of different species across the World.

An alternative approach that avoids grid-cell aliasing, is to plot quantile maps of the geographic distributions of the 'top 25% of most rapidly evolving tips' and the 'bottom 25% of least rapidly evolving tips'. This is robust to the collared dove problem, because it is much more like asking 'what kind of environments do rapidly evolving species prefer', rather than making an explicit spatial link between a lat-lon point and its community. Species could hypothetically change their distribution in a large way between one set of observations and another, but those species would probably still exploit similar climate niches to before, so we wouldn't really mind them doing this. Each new geographic distribution would just be an arbtirary sampling of their potential distribution- and we would believe the potential distribution probably reflects the climate niche that they had around the time of their recent evolution. Maps like this would be robust to community effects. For example, perhaps a huge number of rapidly evolving species live in the tropics, but they are swamped by slow-evolving species that have emerged through fine niche-partitioning. This would reveal that pattern very nicely because the tropics would be highlighted both in a map of 'hot species' and in a map of 'cold species'. The authors' current maps are not capable of showing structure like this, because they treat phenotypic rate as a unidimensional score specific to each cell, rather than a hereditary feature specific to each species.

How can we make a map like this without it being a burdensome task? Maps like this could be drawn using a mosaic of grid-cells, by adding a 'plus one' to the grid cell if a species within the quantile of interest is known to inhabit that cell. This would neatly avoid the problem of small sample sizes in marginal grid-cells- and it would be pretty interesting to contrast these maps with their taxonomic richness etc.

While my previous requests would require the execution of new code that could require a lot of effort, I think it would be very interesting to make a couple of these quantile maps, and that it wouldn't be burdensome. Are the authors open to making them?

**Author response: We appreciate this thoughtful critique and agree with the core point that any grid-cell summary of evolutionary tempo is, by construction, a summary over whichever species occur in that cell at the time of observation, rather than a fixed "intrinsic" property of the location. We also agree that this can invite mechanistic interpretations when mapped against variables such as seasonality. Following the reviewer's suggestion, we implemented the quantile-map approach and added several**

**complementary diagnostics that make within-assemblage mixtures more explicit, while also retaining our existing spatial models.**

**(1) Quantile occupancy maps (addressing grid-cell aliasing and the collared dove example)**

**To reduce sensitivity to interpreting a cell-mean rate as a "rate of place," we generated quantile occupancy maps by ranking species by lineage-rate and defining "fast" and "slow" sets as the top and bottom 25% of the global lineage-rate distribution (with a more stringent 10% tail used as a corroborating sensitivity analysis). We then assign a +1 to each grid cell for each species in the focal quantile known to occur in that cell, producing maps of the geographic distribution of fast- and slow-tailed lineages (Supplementary Fig. 17A–B).**

**This approach directly follows the reviewer's suggestion and mitigates the concern regarding the "collared dove" example in two ways. First, as the reviewer suggested, it provides an interpretive target of "where fast vs. slow lineages occur," which is closer to the environment-preference framing proposed by the reviewer. Second, because these occupancy maps are unweighted, they are not sensitive to the reviewer's concern about range-size weighting.**

**(2) Composition diagnostics using fractional mixtures**

**We also agree with the reviewer's points that raw occupancy counts are confounded by richness gradients and that important structure can be obscured when phenotypic rate is reduced to a single-cell score. To make the structure of community mixtures more explicit, we computed, for each grid cell, the fraction of the local assemblage comprised of globally fast- and slow-tailed lineages (i.e., the proportion of species in the top/bottom tail of the global lineage-rate distribution), and related these fractions to the range-weighted mean lineage rate used in our main spatial analyses (Supplementary Fig. 17C–H).**

**Two key results emerged. First, mean lineage rates covary strongly with tail composition (fast25 fraction: Spearman $\rho = 0.766$; slow25 fraction: $\rho = -0.619$), and the relationships remain strong when excluding species-poor cells (e.g., fast25 fraction: $\rho = 0.796$ for cells with ≥10 species; $\rho = 0.780$ for cells with ≥20 species; slow25 fraction: $\rho = -0.577$ and $-0.605$). Second, cells in the highest decile of mean lineage rates show a pronounced shift in assemblage composition: fast-tailed lineages comprise a much larger share of the assemblage (fast25 = 0.687|0.60 [mean|median] in top-decile mean-rate cells vs 0.262|0.25 elsewhere). In contrast, slow-tailed lineages are strongly depleted (slow25 = 0.094|0.00 vs 0.283|0.282). Patterns for the 10% tails are qualitatively similar (fast10 = 0.241|0.278 vs 0.116|0.089; slow10 = 0.011|0.000 vs 0.097|0.091; not shown).**

**These additional diagnostics clarify the "swamping" scenario highlighted by the reviewer: tropical (and other species-rich) regions can and do indeed include substantial representation from both global tails in occupancy counts. Importantly, however, compositional diagnostics provide a better indication of where fast or slow lineages are over- or underrepresented in global assemblages. Remarkably, our new map of fractional enrichment for fast25 lineages (Supplementary Figure 17C) closely resembles the spatial distribution of mean rates (as corroborated by the correlations reported above).**

**(3) Clarifying interpretation relative to "filtering-only" intuition**

**Finally, to contextualize the magnitude of the observed enrichment in high-mean or low-mean cells (top decile), we compared it to a simple "drop-out only" intuition: if slow-tailed lineages simply dropped out of local assemblages, the expected fast fraction would increase only modestly by denominator shrinkage (e.g., fast10: 0.10 → 0.10/0.90 ≈ 0.11; fast25: 0.25 → 0.25/0.75 ≈ 0.33). The significantly larger enrichment observed in high mean-rate cells (fast25 ~2.4–2.6× higher than elsewhere; fast10 ~2.1–3.1× higher) suggests broader compositional skew than slow-tail loss alone, consistent with a combination of turnover/assembly and ecological processes. We emphasize that these diagnostics are intended to make the within-assemblage structure more transparent and explicit, rather than to serve as standalone mechanistic tests.**

**We have incorporated these analyses and interpretations into a new Supplementary section (Section 4.2) and present the new quantile occupancy maps and composition diagnostics in Supplementary Fig. 17. We have also revised the manuscript's narrative language in a few key places to refer to these patterns and their revised interpretations. These results do not affect our primary analyses, which rely on cell-level means, and they show that spatial variation in cell-wise mean rates is strongly associated with systematic compositional skew (fast-tail enrichment and slow-tail depletion) across assemblages. High mean-rate cells, therefore, reflect broad, systematic shifts in assemblage composition, rather than being attributable to idiosyncratic effects of a small number of taxa.**

3 Minutia

In their response to Reviewer 2, the authors say that the Kmult statistic does not test difference 'with respect to zero'. I feel Reviewer 2 is right, though. The authors' responses about permutations describe a permutation procedure which is contrived to produce a scenario without phylogenetic signal– Kmult = 0. I believe the 'valuable' component of the effect-size calculation is not the expectation that μ will consistently be greatly different from zero, but the need to infer the standard deviation that informs the effect size; $Z = (x-\mu)/\sigma$. Notice that this equation simply resembles a t-statistic difference of mean test and that there is an implicit factor of $\sqrt{n}$ in the numerator. The approach used in geomorph therefore produces a σ that is sensitive to $\sqrt{n}$, which means the test's apparent 'power' increases with sample size when Bienaym´e's formula is not invoked to compare different effect sizes. This is what is being done when the authors plot a latitudinal gradient of effect sizes, so the idea of using the 'different from μ?' statistic as a significance assessor is flawed, but I digress. I think a lot of authors are handling geomorph's Z-scores a bit incorrectly because it is very tempting to interpret them as a common currency

without an exchange rate, and it is not directly intuitive that contrasting scores with each other requires a different type of statistical testing approach, compared to asking whether an individual score exhibits a significant difference from its null distribution. See the documentation of the geomorph ‘compare.pls()’ function as evidence that Dean Adams’ group is aware of this requirement.

**Author response: We agree that the Z-score is mathematically sensitive to sample size, meaning that "significance" in this context reflects statistical power as much as biological signal. While we have retained our existing analysis in our revision, we have added a new statement to the Materials and Methods (Subsection: Testing Global Patterns of Trait Variation) clarifying that the spatial patterns of significance (Figure 4D) reflect the strength of statistical evidence, which is naturally lower in species-poor, high-latitude assemblages.**

**Below, we reproduce Reviewer 1's comments in full and insert responses in bold.**

Reviewer 1

Third Review of 'Rates of passerine body plan evolution in space and time,'

The authors have, in response to previous reviewer comments, undertaken a new supplementary analysis, modified the title of their work, and supplemented their main manuscript text with additional content that will influence the way readers interpret their reported findings. In general I think these changes have improved the manuscript. My overall view of the work is positive.

As part of my review I re-read their updated manuscript and supplement and undertook a broader review of the ornithological literature.

I wish to thank the authors and the editor for their patience in allowing me the time to do this.

The following are my thoughts about the revised manuscript and the author responses, and are mainly aimed at discussing future work they may envision.

The authors believe that their additional exploration has deepened and solidified the leading interpretations they make of their results, and I agree in part. Structure is revealed that presents interesting questions that may perhaps be investigated in future work. For example, I anticipated in Figure S17 that the Amazon basin would host both a high fraction of lineages inferred to possess rapid rates of phenotypic evolution and those inferred to possess low rates of phenotypic evolution. It is interesting that the other major tropical rainforests, such as the central African rainforest, possess neither a large fraction of rapidly or slowly evolving lineages. I am motivated to wonder whether this pattern is real, a result of lower fractional sampling of African species, or whether it could reflect the authors' decision to exclude migratory species, since there are songbirds of Europe and North Asia that overwinter in subsaharan Africa. I am inclined to expect that the pattern is real, and that maybe it is reflective of the multi-millennial scale oscillation subsaharan Africa experiences between humid and dry climate conditions– a pattern that the Amazon is relatively more insulated from, being spared severe aridity [1]. Some authors argued in the past that glacial-cycling caused ephemeral patchiness but not erasure of the Amazon, creating an engine for allopatric speciation. The authors' analysis perhaps breathes some life back into that idea, especially contrasted with subsaharan Africa. Maybe this provides a mechanistic explanation for the old-world new-world differences observed in the study.

I also noticed that the Wallace Line between Borneo and Sulawesi appears quite clearly in Figure S17. This indicates the mechanistic role that species dispersal might play in the data structure. In particular, the dataset the authors interrogate is focused on year-round residents that might have small ranges and be less likely to make ocean crossings. The barrier between the ancient subcontinent of Sundaland and Australasia may have been too difficult for these tropical forest passerine species to cross. I checked the literature and the Wallace line is indeed observed to be a major obstruction to Passerine birds such as Honey eaters [2]. Thinking more deeply on this subject, it is surprising that there are no prominent changes in the frequency of phenotypic rate increases in Figure 2 of the main manuscript sympathetic with major radiations within songbirds that are coincident with the tectonic uplift of the Papua archipelago around 23-15 mya [3], during which time South-East Asia acted as a conduit rather than a barrier to the wider dispersal of Australasian-origin songbird lineages. I checked that the fossil-calibrated node ages, ultimately derived from Oliveros, used in the current study coincide with this uplift- they do. It is pretty interesting that the opening-up of ecological opportunities that this mass-dispersal and radiation must have presented– including the origin of the roots of phenotypically distinctive lineages, did not cause a signature of rate shifts. Maybe these lineages are burning fuses, whose rate increases only emerge later after their dispersal out of Australasia.

I checked the manuscript and was surprised that the Amazon and Wallacean regions are not explicitly mentioned. The Wallacean uplift is mentioned in a reference about the diversification of Corvides though, so

this hypothesis is represented and engaged with. Ultimately, it is hypothesized the mountains pushed up during the Papuan orogen began to weather and cause changes in earth's surface chemistry that drew down vast amounts of CO2 and contributed to Neogene cooling around 8 mya [4], but this shift in climate is underneath the temporal horizon that the current manuscript can investigate. Overall it is interesting that climate change events rather than major dispersal opportunities for passerines are salient in Figure 2 of the main manuscript.

I think the authors' revised title is a good choice. I acknowledge that the authors have added language to their manuscript discussing the potential influence of dispersal and used language such as 'host' to make clear that inferred rates of phenotypic evolution of lineage origination calculated across grid-cells are not necessarily endogenous.

**Response: We thank the reviewer for these thoughtful observations. In response, we added a brief discussion in the Eocene–Oligocene transition section acknowledging that lineage-specific biogeographic expansion may have reinforced ecological opportunity in particular clades, while retaining climatic instability as the broader explanation for the temporal clustering of accelerations. We incorporated this conservatively by referring to clades already identified in our analysis, rather than expanding the manuscript to include a more speculative regional discussion of Amazonian or Wallacean history. We have added a brief note to the supplementary materials (Supplementary Figure 17) to acknowledge those observations.**

The authors are correct that I was curious about recent rate shift frequencies in Figure 2. I agree with the authors' explanation that minimum group sizes in the heuristic search impose a soft temporal horizon beyond which shift frequencies cannot be found. I might be tempted to truncate plot axes around this soft limit to prevent readers from attempting to interpret data under that horizon, but that's an authorial choice.

**Response: We have now clarified this point explicitly in the Figure 2 caption. The caption now states that, because the heuristic search is conditional on a minimum clade size, very recent shift frequencies are subject to a soft temporal horizon near the present.**

The authors trust that high-rates of phenotypic evolution inferred in out-group taxa are not aliased by the distribution of nodes in the tree because the data distribution informs the placement of estimated shifts. I do not disagree with this latter statement, but observe that it is not mutually exclusive with node aliasing effects. e.g. A bird lineage with a very different mean set of traits would probably demand a rate shift to accommodate the displacement of the trait from its ancestors. Perhaps the rate of evolution in this lineage then slows down, and a second (or many) downward rate shift(s) occurs nearer the present– with the overall effect being that the tips would have relatively low rates within the kernel window used to estimate rates of evolution in the recent past. Now imagine that only 1 species from this lineage survives a recent extinction event; the data exists to demonstrate an ancient shift in the mean, but the data to document the deceleration nearer the present are absent. This biases the estimated recent rate of evolution higher than it really is, because the integrated and weighted root-to-tip path does not encounter any recent highly-weighted nodes documenting the rate deceleration; just the ancient node documenting the higher rate. The available tip-sampling therefore does exert an influence on estimated recent rates of phenotypic evolution. Implicitly, the available tip-sampling necessarily determines all rate shifts, but I digress.

Previous rate-shift search approaches used by other authors fit more highly parameterized models that estimate an exponential decay parameter, which reflects the presumption that a lineage's inferred evolutionary rate should cool-off following a bursty event, even if there are no direct daughter nodes to directly inform that estimate (rather it is transferred through the likelihood-maximizing estimate of the variance-covariance tensor that includes adjacent lineages whose daughter nodes do inform the cool-off parameterization). I appreciate that increasing the parameter burden of the models the authors are fitting might not be compatible with their approaches in the current paper, of course, and I am not requesting it.

I just think the authors should be mindful of the fact the representation of recent nodes over which the heuristic search can make rate estimates should be expected to exert an influence in the recent tip-rates when the variance-covariance matrix is represented as a series of nested brownian diffusions. (I would predict that additional sampling at the genus level would tend to depress estimated rates of recent evolution, and that increased sampling at the species level within genera might actually increase tip rates in some groups, because plastic anatomical responses and the effects of hybridization in closely related lineages add noise that could be interpreted as rate increases).

I appreciate the authors' comments that assessing full-phylogeny simulations to confirm node aliasing would be time consuming and will not insist on it. The mvMORPH package and mvgls function draw upon a penalized likelihood optimization procedure, which is computationally expensive and probably explains the long run time. I believe other authors have demonstrated that, when a phylogeny is treated as a directed acyclic graph, that belief propagation can be used to approximate likelihood with substantially reduced run-times– and that the estimates from the two approaches converge for Gaussian trait models. I do not ask for this to be attempted; it is just an avenue for future exploration that might empower the authors to more easily run large simulation suites if they wanted to. I think Cecile Ané has been publishing on this topic recently. She has also thought about reticulated phylogenetic structures, such as those we might expect to result from hybridization.

**Response: We thank the reviewer for these suggestions. The idea of incorporating exponential decay dynamics more directly is a very interesting avenue for future work by the authors.**

**We previously revised the Methods section to clarify the interpretation of our lineage-rate statistic. The manuscript now describes this measure as a present-biased, temporally filtered average of inferred regime history along each root-to-tip path, with the half-life parameter determining how strongly deeper regimes influence recent summaries. While the reviewer's points are exciting ideas, investigating these ideas in a satisfactory way likely requires a fundamentally different approach and exploring them is beyond the scope of this manuscript.**

The authors added language and discussion to their manuscript about the potential roles of extinction and dispersal as determinants of contemporary biogeography and the distribution of estimated grid-cell wise measures like mean rates of phenotypic evolution. I think their changes are reasonable.

**Response: We thank the reviewer for previously highlighting these dynamics and believe the manuscript is improved following the incorporation of this discussion.**

The authors discuss the possibility of filtering (by which I take it they mean the selective extinction of conservative lineages due to glacial cycling in the northern latitudes potentially favoring adaptable, plastic or high-dispersal species) with a thought experiment on denominator shrinkage and compare to observed data. I might be missing part of their point. Grid cells in northern latitudes sometimes contain 10-30 species (Fig S11A) compared to values that exceed 100 in low latitudes. The accumulation of ancient conservative lineages in the tropics, and their elimination at northern latitudes, could be part of the explanation for the compositional difference, couldn't it? I think the denominator shrinkage idea perhaps works if we assume the net number of species remains constant bar extinction, but new species emerge constantly– and most of them probably emerge in the tropics through fine-scale allopatry, because range sizes are small and the tropics have the greatest richness of lineages available to derive new species. Indeed, that is probably part of the reason why the tropics have an absolutely higher richness. My comments on ancient species persistence in the tropics and evidence of high rates of routine purging in the mid-high latitudes is supported by genetic and morphological analyses at the intraspecific level in birds of the americas [5].

In future work, the authors might be motivated to specifically examine whether bursts of Amazonian tropical passerine radiation are precipitated by the onset of intense glacial cycling in the Neogene cool-down, because questions of this nature have been identified as an outstanding problem in the literature. This idea is

disputed by some authors– but Figure 2 in the main manuscript potentially indicates there could be truth in it. The original version of the hypothesis focused on the Quarternary, and the reality might be some version of the hypothesis which is stretched over a much longer arc of geological time. Amazonian passerines present a very interesting case where changes in continental isolation have regulated dispersal events, so it could be possible to tease-apart the potential dual roles of dispersal and climate change on phenotypic evolution and lineage origination.

In any case, I am satisfied that the authors have thought deeply about these potential confounding forces, represented their potential influence in comments in their main manuscript, and explored them in their supplement such that readers may develop these ideas further and respond in future works.

**Response: We thank the reviewer for their thoughtful comments and agree there are many ideas to explore further.**

Minor comments:

An interesting note: Figure 4 subplot B in the main manuscript shows that there is a flat or slightly southern-biased gradient in phenotypic over-dispersion compared to divergence times in the Americas; the old-world/new-world difference isn't just one of intercept but it fundamentally one of slope as well. Perhaps this feature is a reflection of an Amazonian diversity engine.

The global regions the authors use appear to indicate that Asia ends at around 40 degrees Northern latitude/around Manchuria or Mongolia; I assume Siberia has been designated as Europe. It does not make a difference to the analysis, but it occurred to me that the regions are not defined in the manuscript. e.g. the word 'Europe' does not occur outside of figure legends, although the New and Old world continents are disitnguished and I do not doubt that their separation is unambiguous. I checked the supplement and region-specific words like 'Europe', 'Africa' 'America', 'Amazon', 'Arctic' also do not occur.

**Response: We have more clearly defined the regional bins directly in the Figure 4 legend. New World is now specified as North and South America, and Old World as Africa, Europe, Asia, and Australasia.**

On the comment about 'using only extant taxa', I suppose the phylogenies themselves discerned by previous authors were assembled with reference to fossil node calibrations– but I think this is a minor detail.

**Response: We revised this wording. The manuscript now states that these macroevolutionary patterns can be observed "even when phenotypic data are available only for extant taxa."**

References

[1] Mark B Bush and Paulo E de Oliveira. "The rise and fall of the Refugial Hypothesis of Amazonian speciation: a paleoecological perspective". In: Biota Neotropica 6.1 (2006). [2] Hugh A Ford. "Why does the distribution of the Honeyeaters (Meliphagidae) conform so well to Wallace's Line?" In: Faunal and Floral Migration and Evolution in SE AsiaAustralasia. CRC Press, 2001, pp. 147–152. [3] Robert G Moyle et al. "Tectonic collision and uplift of Wallacea triggered the global songbird radiation". In: Nature Communications 7.1 (2016), p. 12709. [4] Peter E Martin et al. "The rise of New Guinea and the fall of Neogene global temperatures". In: Proceedings of the National Academy of Sciences 120.40 (2023), e2306492120. [5] Brian Tilston Smith et al. "A latitudinal phylogeographic diversity gradient in birds". In: PLoS biology 15.4 (2017), e2001073.

**Response: We would like to extend our sincere thanks to the reviewer for their extraordinary efforts to help us improve this manuscript. The manuscript is substantially improved due to their comments, and we hope to interact with them again in the future should they choose to reveal themselves!**