nature portfolio

Peer Review File

Synnovation and confluence explain disparities in species richness across the Andes and Hengduan-Himalaya Mountains

Corresponding Author: Dr Yaowu Xing

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
I have reviewed the responses in the revised manuscript to the comments from the two previous reviewers. In my opinion, most revisions are adequate, although several issues have not been fully addressed. I strongly recommend publication after minor revisions, including:

1. Deleting the “significance” section (which is not a formal section in the high-profile journal Nature Communications and is more appropriate for PNAS), and presenting the results and discussion as separate sections.

2. Adding more discussion of the study’s limitations and future research directions—for example, regarding the genetic basis of deciduous-SDV leaves through transcriptomic analyses, which may arouse greater interest among general readers.

3. Citing both early and recent references related to biogeographic plant studies of the Hengduan Mountains, including Molecular Phylogenetics and Evolution 38 (2006): 31–49; PNAS (2025), https://doi.org/10.1073/pnas.2426017122; and Nature Ecology & Evolution 10 (2026): 794–806. All of these references suggest that ecological opportunities have helped promote species richness in the Hengduan Mountains and the Himalaya, despite the use of different geographic delimitations and terminology.

Reviewer #3

(Remarks to the Author)
Review of manuscript NCOMMS-26-017922-T
Title: “Synnovation and confluence explain disparities in species richness across the Andes and Hengduan-Himalaya Mountains” by Yu et al.

The study is interesting, timely, and provides a significant contribution by comparing diversification dynamics across two major mountain systems (Andes and Himalaya-Hengduan Mountains) within a common and robust phylogenomic framework of the genus Berberis. The integration of large-scale nuclear genomic data with trait-dependent diversification models (SSE-modelling) and (paleo) niche modelling addresses fundamental questions in mountain biogeography.
In this revised version, most concerns raised by previous reviewers have been addressed, particularly through the inclusion of the newly generated nuclear phylogeny and the clarification of taxonomy-related issues in the supplements. The work is methodologically rigorous, and the scope of the taxon sampling is impressive. However, there are a few critical areas regarding the phylogenetic inference and the presentation of the comparative statistics that require clarification or adjustment.

Major points
1. Branch length estimation (Lines 273–279): The manuscript (to me) suggests that only SNP data were used for tree estimation without re-incorporating invariant sites (“dataset 2” in Materials and Methods section, first paragraph). Using only variable sites (SNPs) without an acquisition bias correction (e.g., Lewis MKV model; Lewis 2001) typically leads to a significant overestimation of branch lengths. This, in turn, impacts divergence time estimation and potentially the species

tree topology in ASTRAL-IV, which relies on gene tree quartet frequencies. The authors use the IQ2MC framework, which is state-of-the-art. However, they must clarify if the 'SNP-only' dataset was analysed using an acquisition bias correction (e.g., the +ASC parameter in IQ-TREE or providing constant site patterns). Without this, branch lengths passed to MCMCTree would be systematically overestimated, leading to potentially biased divergence times. The authors must clarify if and how they accounted for this bias to ensure the reliability of the chronogram and subsequent estimations and comparative analyses.
2. Niche modelling rationale: The "Present and past distribution modeling" of deciduous/evergreen Berberis in South America appears somewhat academic. I question the biological significance of modelling HHM species from Asia into the South American paleoclimate, where these taxa never occurred. The authors should more explicitly state what mechanistic insight is gained from this exercise that is not already captured by the trait-diversification and ecological niche divergence analyses.
3. Statistical reporting for PCM: While the use of stepwise ANOVA/LRT for nested models (Table S6) is statistically valid, I strongly recommend using a consistent AICc-based framework for all model comparisons (including non-nested MuHiSSE/CID models). To improve transparency and ease of interpretation, the authors should provide a unified model selection table including k (degrees of freedom/parameters), lnL, AICc, deltAICc, and Akaike weights (wi). It would also be preferable, if the Mu(Hi)SEE models would be specified using standard SSE parameter connotation (e.g., λ1 = λ 2 = λ3, or q12, q13, etc.).

Specific comments
Data availability: Please ensure all data, including the distributional dataset, sampling fractions per clade, and best also the code used for analysis, are available via GenBank/Dryad/GitHub upon publication.

Discussion: The discussion could benefit from a little broader contextualization within recent literature on mountain biodiversity (e.g. https://doi.org/10.1111/geb.70169). While the current manuscript focuses on morphological synnovation and habitat availability, discussing how these factors interact with resource-use strategies, niche partitioning and spatial isolation could provide a more comprehensive view of the mechanisms generating exceptional species richness in mountain hotspots.

Definition of HDM (Line 183): "HDM" is likely a typo for "HHM"?? If not, define upon first mentioning. Furthermore, the exclusion of the Qinghai–Tibet Plateau, mentioned in the response to reviewers, should be explicitly stated in the main text to avoid ambiguity.

Methods (Line 299): The authors likely used TRACER to check ESS and trace plots, rather than FigTree. Please correct this.

Taxon sampling (Line 335): Please provide the specific numbers/percentages used for the "non-random incomplete taxon sampling" correction.

Biogeographic analysis (Line 342ff): Specify the "known paleogeographic and climatic histories" or provide citations. Was the analysis time-stratified, and what was the evidence for the dispersal constraints used?

Lines 344-246: Sentence seems redundant.

Lines 439-441: This sentence appears redundant with the following one.

Line 362: "oth" should be "both".

Trait-dependent diversification (Line 384ff): Where are the results of the MCMC/Bayesian rate comparisons (with 100 randomly sampled phylogenies) shown? Please specify where the posterior distributions and credibility intervals are reported.

Figure 1: The leaf images are aesthetically pleasing, but a clear legend explaining the phylogeny (colours and dots) within the figure itself and not only buried in the caption could be more helpful.

Figure 1 Legend (Line 736): In "non-alpine/ subalpine", a dash (at least) seems to be missing. Should it say "non-alpine/non-subalpine" (meaning neither)?

References: Please check again for consistency (e.g., Ref 48 and 50). Throughout the text, please make sure citations are correct (e.g., line289: ref 64 is IQ2MC, not ref 62).

Version 1:

Reviewer comments:

Reviewer #3

(Remarks to the Author)
Review of revised manuscript NCOMMS-26-017922-T (2nd round): "Synnovation and confluence explain disparities in species richness across the Andes and Hengduan-Himalaya Mountains" by Yu et al.

I thank the authors for their thorough and careful revision. All major and minor concerns raised in the previous round have been addressed: the acquisition bias correction (+ASC) in IQ-TREE has been clarified and its independence from the ASTRAL-IV branch lengths in the IQ2MC framework is well explained in the responses; the niche modelling rationale has been explained with a clearer mechanistic justification; and the comparative model selection has been reformulated into a unified AICc-based framework, also including model descriptions with standard SSE notation. The remaining specific comments regarding typographical errors, redundancies, and data availability have likewise been resolved. I am very looking forward seeing this comprehensive study published.

Best wishes, Nicolai Nürk

Reviewer #1 (Remark to the Author):

I have reviewed the responses in the revised manuscript to the comments from the two previous reviewers. In my opinion, most revisions are adequate, although several issues have not been fully addressed. I strongly recommend publication after minor revisions, including:

Response: We thank the reviewer for the positive evaluation of our revised manuscript. We have addressed the remaining points raised and provide responses below.

1. Deleting the "significance" section (which is not a formal section in the high-profile journal Nature Communications and is more appropriate for PNAS), and presenting the results and discussion as separate sections.

Response: We have removed the "Significance" section and thank the reviewer for the reminder.

2. Adding more discussion of the study's limitations and future research directions—for example, regarding the genetic basis of deciduous-SDV leaves through transcriptomic analyses, which may arouse greater interest among general readers.

Response: We have added a brief statement in the Results and Discussion to acknowledge this limitation and highlight future transcriptomic work on the genetic basis of the deciduous–SDV traits. Please see lines: 233-237.

3. Citing both early and recent references related to biogeographic plant studies of the Hengduan Mountains, including Molecular Phylogenetics and Evolution 38 (2006): 31–49; PNAS (2025), https://doi.org/10.1073/pnas.2426017122; and Nature Ecology & Evolution 10 (2026): 794–806. All of these references suggest that ecological opportunities have helped promote species richness in the Hengduan Mountains and the Himalaya, despite the use of different geographic delimitations and terminology.

Response: We have now incorporated the study published in Molecular Phylogenetics and Evolution 38 (2006): 31–49. The PNAS (2025) reference was already included in the original manuscript. We have carefully evaluated the Nature Ecology & Evolution (2026) study; however, we consider it less directly relevant to the specific questions addressed in our manuscript and have therefore not included it. We nonetheless appreciate the reviewer for drawing our attention to this work.

Reviewer #3 (Remarks to the Author):

Review of manuscript NCOMMS-26-017922-T
Title: "Synnovation and confluence explain disparities in species richness across the Andes and Hengduan-Himalaya Mountains" by Yu et al.

The study is interesting, timely, and provides a significant contribution by comparing diversification dynamics across two major mountain systems (Andes and Himalaya-Hengduan Mountains) within a common and robust phylogenomic framework of the genus Berberis. The integration of large-scale nuclear genomic data with trait-dependent diversification models (SSE-modelling) and (paleo) niche modelling addresses fundamental questions in mountain biogeography.
In this revised version, most concerns raised by previous reviewers have been addressed, particularly through the inclusion of the newly generated nuclear phylogeny and the clarification of taxonomy-related issues in the supplements. The work is methodologically rigorous, and the scope of the taxon sampling is impressive. However, there are a few critical areas regarding the phylogenetic inference and the presentation of the comparative statistics that require clarification or adjustment.

Response: We sincerely thank the reviewer for this thoughtful assessment of our manuscript and the recognition of our works. We also appreciate the reviewer's careful attention to the remaining points concerning phylogenetic inference and the presentation of comparative statistics.

We have carefully considered each comment and have revised the manuscript accordingly. Detailed point-by-point responses are provided below.

Major points
1. Branch length estimation (Lines 273–279): The manuscript (to me) suggests that only SNP data were used for tree estimation without re-incorporating invariant sites ("dataset 2" in Materials and Methods section, first paragraph). Using only variable sites (SNPs) without an acquisition bias correction (e.g., Lewis MKV model; Lewis 2001) typically leads to a significant overestimation of branch lengths. This, in turn, impacts divergence time estimation and potentially the species tree topology in ASTRAL-IV, which relies on gene tree quartet frequencies. The authors use the IQ2MC framework, which is state-of-the-art. However, they must clarify if the 'SNP-only' dataset was analysed using an acquisition bias correction (e.g., the +ASC parameter in IQ-TREE or providing constant site patterns). Without this, branch lengths passed to MCMCTree would be systematically overestimated, leading to potentially biased divergence times. The authors must clarify if and how they accounted for this bias to ensure the reliability of the chronogram and subsequent estimations and comparative analyses.

Response: We thank the reviewer for raising this important methodological point and apologize that this detail was not clearly stated in the original manuscript. We wish to clarify that ascertainment bias correction was implemented from the outset during phylogenetic inference of the SNP-only dataset using the `+ASC` option in IQ-TREE3, following the recommended practice for datasets containing only variable sites (Minh et al., 2020). We have now added the above information to the Materials and Methods section for clarity.

We also appreciate the reviewer's concern regarding the potential propagation of branch-length bias into divergence-time estimation. We would like to clarify that our dating analyses followed the IQ2MC framework, in which the ASTRAL-IV species tree was used only as an unscaled topology (i.e., without branch lengths) to constrain the dating analysis. Thus, branch lengths estimated by ASTRAL-IV were not directly passed to MCMCtree. Instead, the approximate likelihood calculations used for divergence-time inference were obtained through the IQ-TREE3 step, based on the sequence alignment and the specified substitution models via the gradient and Hessian calculations. One advantage of this framework is that time estimation is informed directly by sequence-likelihood information rather than by relying on pre-estimated species-tree branch lengths, while also allowing the use of more flexible and better-fitting substitution models available in IQ-TREE3. Please find the related source data and dating results within the Supplementary Files (Supplementary Data: 2. Divergence time estimation), which are currently hosted in a private figshare repository associated with this manuscript and can be accessed at: https://figshare.com/s/126d2b0469f6ea637ac5

2. Niche modelling rationale: The "Present and past distribution modeling" of deciduous/evergreen Berberis in South America appears somewhat academic. I question the biological significance of modelling HHM species from Asia into the South American paleoclimate, where these taxa never occurred. The authors should more explicitly state what mechanistic insight is gained from this exercise that is not already captured by the trait-diversification and ecological niche divergence analyses.

Response: We thank the reviewer for the insightful comment. Our primary aim was not to reconstruct the historical distributions of HHM taxa in South America, but to directly test a central hypothesis developed in the Introduction: that diversification in *Berberis* is driven by the interaction between functional trait innovation (deciduousness and SDV leaves) and ecological opportunity generated by mountain building.

In this sense, our phylogenetic and trait-dependent diversification analyses indicate

that the deciduous–SDV combination is associated with accelerated diversification in the Hengduan–Himalaya Mountains (HHM), consistent with a key synnovation expanding the fundamental niche into colder, high-elevation environments. This raises a mechanistic question: why did a comparable diversification not occur in the Andes? Is this due to the absence of trait evolution, or to differences in ecological opportunity?

To disentangle these alternatives, we used the climatic niche of extant HHM deciduous–SDV lineages as an empirical proxy for environmental conditions conducive to this diversification strategy. Projecting these niches onto paleoclimate reconstructions of South America allows us to evaluate whether such environmental conditions were historically available in the Andes, independent of whether the taxa themselves were present.

This approach is conceptually analogous to widely used paleoecological niche modeling frameworks, in which present-day species niches are projected onto past climates (e.g., Last Glacial Maximum reconstructions) to infer the temporal availability of suitable habitats, rather than realized species distributions.

Our results indicate that environmental conditions favorable to deciduous *Berberis* likely existed in South America during the Oligocene and Miocene, but subsequently contracted markedly through time. In contrast, comparable habitat contraction was not inferred for the HHM, where suitable environments persisted or expanded alongside mountain uplift.

These findings suggest that the absence of deciduous lineages in the Andes may reflect a reduction in ecological opportunity through time, rather than solely the absence of trait evolution. In this way, the niche projection analysis provides a mechanistic complement to our diversification and trait-evolution analyses, by explicitly linking synnovation to the temporal dynamics of habitat availability across mountain systems.

We have slightly revised the related content in the manuscript, please see line: 265-267.

3. Statistical reporting for PCM: While the use of stepwise ANOVA/LRT for nested models (Table S6) is statistically valid, I strongly recommend using a consistent AICc-based framework for all model comparisons (including non-nested MuHiSSE/CID models). To improve transparency and ease of interpretation, the authors should provide a unified model selection table including k (degrees of freedom/parameters), lnL, AICc, deltAICc, and Akaike weights (wi). It would also be preferable, if the

Mu(Hi)SEE models would be specified using standard SSE parameter connotation (e.g., λ1 = λ 2 = λ3, or q12, q13, etc.).

Response: We sincerely thank the reviewer for the suggestion regarding the presentation of the results of phylogenetic comparative model (PCM) analyses. We agree that a unified information-theoretic framework improves transparency and facilitates interpretation across both nested and non-nested candidate models.

Accordingly, we have revised the supplementary tables to provide a consolidated model-selection table (Supplementary Table 4-10 in the Supplementary Information), in which all candidate MuHiSSE/CID and MuSSE-like models are compared using a consistent AICc-based framework. This new set of tables now reports the number of free parameters (k), maximized log-likelihood (lnL), AICc, ΔAICc, Akaike weights (wi), model rank, and a brief biological interpretation for each model. In addition, following the reviewer's recommendation, we have added a table (Supplementary Table 8, 9 in the Supplementary Information) summarizing the parameterization of each candidate model using standard SSE notation for each scenario. Specifically, turnover/speciation parameters, extinction parameters, and transition-rate parameters are now explicitly described to facilitate comparison among MuHiSSE/CID and MuSSE-like formulations.

Accordingly, we slightly revised the text in the manuscript for clarity, please see lines: 428-430; 455-471.

Specific comments
Data availability: Please ensure all data, including the distributional dataset, sampling fractions per clade, and best also the code used for analysis, are available via GenBank/Dryad/GitHub upon publication.

Response: In this resubmission, we have included comprehensive Supplementary Files containing all datasets used in the analyses, as well as the direct outputs from those analyses. They are currently hosted in a private figshare repository associated with this manuscript and can be accessed at:
https://figshare.com/s/126d2b0469f6ea637ac5

These files are organized following the structure of the Methods section in the manuscript. Detailed descriptions of the contents and file organization are provided in the main README.md file. In addition, to facilitate navigation and interpretation, we have included supplementary README.md files within subfolders corresponding to individual analyses, describing the purpose and naming conventions of files in each directory.

We have also prepared a Source Data Excel file (available in the private figshare repository associated with this manuscript and can be accessed at: https://figshare.com/s/126d2b0469f6ea637ac5) that contains the data used to generate all figures. Where overlap exists between this file and the Supplementary Files, we provide clear references indicating where the original data can be found. This Source Data file covers all figures. Tables are not included in the Source Data file, as they are directly derived from analysis outputs and can be accessed in the corresponding Supplementary Files.

All R scripts used in this study have been made publicly available on GitHub at: https://github.com/odysseyelf/berberis-diversification-scripts.git

The resequencing data generated in this study have been deposited in the GenBank Sequence Read Archive (SRA) under BioProject [PRJNA1457873]. Individual SRA accession numbers (assigned by GenBank) correspond to BioSample accession numbers are provided in Supplementary Files: 0. Species distribution, habitat, and sampling. Sequence probe data and VCF files used for phylogenomic analyses are provided in Supplementary Data: 1. Phylogenomic inference. These files are currently available in a private figshare repository associated with this manuscript: https://figshare.com/s/126d2b0469f6ea637ac5. Please note that the BioProject is currently under private status and not yet publicly accessible. Due to the large volume of data (approximately 3 TB), the sequence data are still being processed by GenBank and are expected to be released in the near future.

Discussion: The discussion could benefit from a little broader contextualization within recent literature on mountain biodiversity (e.g. https://doi.org/10.1111/geb.70169). While the current manuscript focuses on morphological synnovation and habitat availability, discussing how these factors interact with resource-use strategies, niche partitioning and spatial isolation could provide a more comprehensive view of the mechanisms generating exceptional species richness in mountain hotspots.

Response: We thank reviewer for this suggestion. We have revised the Discussion to incorporate recent work on mountain biodiversity, including the suggested study, and added a brief synthesis on how synnovation and habitat availability interact with niche partitioning, resource-use strategies, and spatial isolation. Please see lines: 211-217.

Definition of HDM (Line 183): "HDM" is likely a typo for "HHM"?? If not, define upon first mentioning. Furthermore, the exclusion of the Qinghai–Tibet Plateau, mentioned in the response to reviewers, should be explicitly stated in the main text to avoid ambiguity.

Response: We apologize for the lack of clarity in the original manuscript. “HDM” was

intended to refer specifically to the Hengduan Mountains, whereas “HHM” denotes the broader Hengduan–Himalaya Mountains region used throughout the study. In the Discussion, these terms were used separately when referring to the distinct uplift and geological histories of the Hengduan Mountains and the Himalayan region, but this distinction was not sufficiently explained. We have now clarified the terminology in the revised text. In addition, we have explicitly stated in the Introduction (Line 57) that the HHM region, as defined in this study, excludes the Qinghai–Tibet Plateau as suggested.

Methods (Line 299): The authors likely used TRACER to check ESS and trace plots, rather than FigTree. Please correct this.

Response: We thank the reviewer for catching this error. The reviewer is correct that convergence diagnostics, including trace plots and effective sample size (ESS) values, were assessed using Tracer rather than FigTree. We have now corrected the Methods section accordingly. See lines: 332-334.

Taxon sampling (Line 335): Please provide the specific numbers/percentages used for the "non-random incomplete taxon sampling" correction.

Response: We have revised the Methods section to provide the specific values used for the non-random incomplete taxon sampling correction, including the overall sampling fraction (0.65) and the clade-specific sampling proportions used in the analyses.

Biogeographic analysis (Line 342): Specify the "known paleogeographic and climatic histories" or provide citations. Was the analysis time-stratified, and what was the evidence for the dispersal constraints used?

Response: We thank the reviewer for this important suggestion. We agree that the rationale for the dispersal constraints should be stated more explicitly. We have now revised the Methods section to clarify that the BioGeoBEARS analyses were time-stratified into three intervals (before 30 Ma, 30–5 Ma, and 5–0 Ma), reflecting major changes in continental connectivity and global climate history. We also added the paleogeographic and climatic basis for these constraints, including the role of the Tethys Ocean, the Central American Seaway, weakening Northern Hemisphere land bridges, and late Neogene global cooling, with appropriate citation (DOI: 10.1126/science.1059412). Please see lines: 392-407 and Supplementary Data: 4. Macroevolutionary rates and biogeographic analysis, for the locality and dispersal matrices. These files are currently available through a private figshare repository associated with this manuscript: https://figshare.com/s/126d2b0469f6ea637ac5

Lines 344-246: Sentence seems redundant.

Response: We thank the reviewer for noting this redundancy. The repeated sentence in the biogeographic analysis section has been removed during revision.

Lines 439-441: This sentence appears redundant with the following one.

Response: We thank the reviewer for noting this redundancy. The repeated sentence has been removed.

Line 362: "oth" should be "both".

Response: We thank the reviewer for catching this typographical error. The text has now been corrected from “oth” to “both”.

Trait-dependent diversification (Line 384): Where are the results of the MCMC/Bayesian rate comparisons (with 100 randomly sampled phylogenies) shown? Please specify where the posterior distributions and credibility intervals are reported.

Response: We thank the reviewer for this helpful comment. We apologize that the description of these Bayesian MuSSE analyses, including where the corresponding results were presented, was not sufficiently clear in the original manuscript. We have therefore revised the relevant Methods text for clarity and now explicitly indicate where these outputs are reported. The posterior distributions of state-specific speciation rates across 100 sampled phylogenies for both Scenario 1 and Scenario 2, including mean, median, standard deviation, and 95% credibility intervals was deposited in Supplementary Data: 2. Divergence time estimation and in the Source Data file for Figure 2. In addition, the IQ2MC pipeline does not output a posterior distribution of trees; instead, node age estimates and their associated 95% credibility intervals are stored numerically in the MCMCTree output file (out.txt). The time-calibrated tree used in Figure 1 incorporates these estimates, and the corresponding tree file with credibility intervals is available in Supplementary Data: 2. Divergence time estimation. These files are currently available through a private figshare repository associated with this manuscript: https://figshare.com/s/126d2b0469f6ea637ac5

Figure 1: The leaf images are aesthetically pleasing, but a clear legend explaining the phylogeny (colours and dots) within the figure itself and not only buried in the caption could be more helpful. Figure 1 Legend (Line 736): In "non-alpine/ subalpine", a dash (at least) seems to be missing. Should it say "non-alpine/non-subalpine"

(meaning neither)?

Response: We thank the reviewer for the suggestions. We have revised Figure 1 to include a clearer in-figure legend explaining the color scheme and dot symbols used in the phylogeny, so that key information can be interpreted directly from the figure. We have also corrected the wording in the figure legend to remove ambiguity regarding the intended meaning of “non-alpine/non-subalpine”. See line: 893.

References: Please check again for consistency (e.g., Ref 48 and 50). Throughout the text, please make sure citations are correct (e.g., line289: ref 64 is IQ2MC, not ref 62).

Response: We thank the reviewer for carefully checking the references. We have now thoroughly revised the reference list and in-text citations for consistency.