

Exploring Tropical Montane Phylogenetic Community Structure of Carabidae

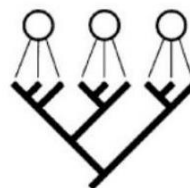
By: Rebekah Hest
BINF*6210 Assignment 4

Github: <https://github.com/rebekahhest/Assignment4/tree/main/assignment4>

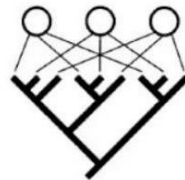
Phylogenetic Community Structure

Studying community assembly through evolutionary relationships provides a framework to infer the ecological processes that drive biodiversity patterns. Phylogenetic diversity quantifies the degree of **phylogenetic clustering** or **overdispersion**. More specifically, the goal is to uncover if local species coexistence is shaped more by environmental filtering, competition, historical biogeography, or stochastic processes.

This research is critical for predicting species responses to climate change, managing ecosystems, and conserving evolutionary diversity (Hoiss et al., 2012). Combining phylogenetic data with distributional data can highlight the underlying drivers of community composition.



Clustering



Overdispersion

Insect Biodiversity in Tropical Mountains

Ecosystems that display rapidly changing conditions along elevational gradients are described as having high environmental turnover and are useful at modelling complex interactions. Tropical mountains contain a wide spectrum of microclimates and are considered global hotspots of biodiversity due to their climatic stability but are underrepresented in ecology and evolution studies (Montano-Centellas et al., 2020).

Hyperdiverse taxonomic groups with highly unresolved phylogenies like insects are ideal candidate study systems due to their environmental sensitivity (Dolson et al., 2021). Despite extensive ecological work, **phylogenetic community structure in tropical montane Carabidae** remains understudied, especially regarding how lineage diversification interacts with elevational gradients.



Study Objective & Competing Hypotheses

Investigate how **phylogenetic community structure** of **Carabidae** varies across a **tropical elevation gradient** and to assess the role of spatial, environmental and biological processes in driving diversity patterns

H1



**Environmental Filtering
(Abiotic Stress)**

At high elevations

- Clustering
- Low α -diversity
- Low β -diversity

H2



**Sky-Island Diversification
(Endemism)**

At high elevations

- Clustering (possible)
- Low-moderate α -diversity
- High β -diversity

H3



**Competition
(Inter-Species)**

At low elevations

- Overdispersion
- High α -diversity
- Moderate β -diversity

Datasets

Carabidae & Elevation Metadata

Public data via BOLD API endpoint (11/26/2025)

- Taxon = **Carabidae**
- Geo = Costa Rica, Ghana, Philippines, South Africa, Peru
- Specimen metadata & raw nucleotides

Pre-filtering (n = **6376**)

Filtering criteria:

- **Tropical** latitudes (-23.5° to 23.5°)
- Marker code: COI-5P
- Sequence length > 500bp
- Elevation & bin_uri fields required

Post-filtering (n = **3433**)

Elevation bands from 0 to 3000m in 100m intervals

BIN Representative Sequences

Barcode Index Number (BIN) used as a proxy for species

- Useful for constructing provisional phylogenies where formal taxonomic identification is incomplete

Unique BINs (n = **398**)

BIN representative criteria:

- Longest read
- Least ambiguous bases ("N")
- Remove contaminated BINs (n = 1)
 - verified by BOLD V5 ID Engine

BIN representatives (n = **397**)

- Sequence length after alignment and trimming = 660bp

Presence/absence **community matrix**



Analysis Pipeline

Data Ingestion & Preprocessing

BOLD Systems API

- Carabidae BIN reps
- Tropical elevation bands
- Presence/absence matrix

Multiple Sequence Alignment

MEGA12

- MUSCLE
- Gap open penalty = 400
- Cluster method = UPGMA
- Substitution model = GTR+G+I

Phylogenetic Tree Reconstruction

IQTREE-2

- Maximum Likelihood
- 1000 bootstrap replicates
- Consensus tree (.contree)

Alpha (α) Diversity

Faith's phylogenetic diversity (PD)

- Sum of branch lengths connecting taxa to root
- Linear models:
 - PD ~ elevation
 - SR ~ elevation
 - PD ~ SR
- Correlation test (PD & SR)

Beta (β) Diversity

Phylosor PCoA

- Fraction of branch-length shared between two communities

Unifrac PCoA

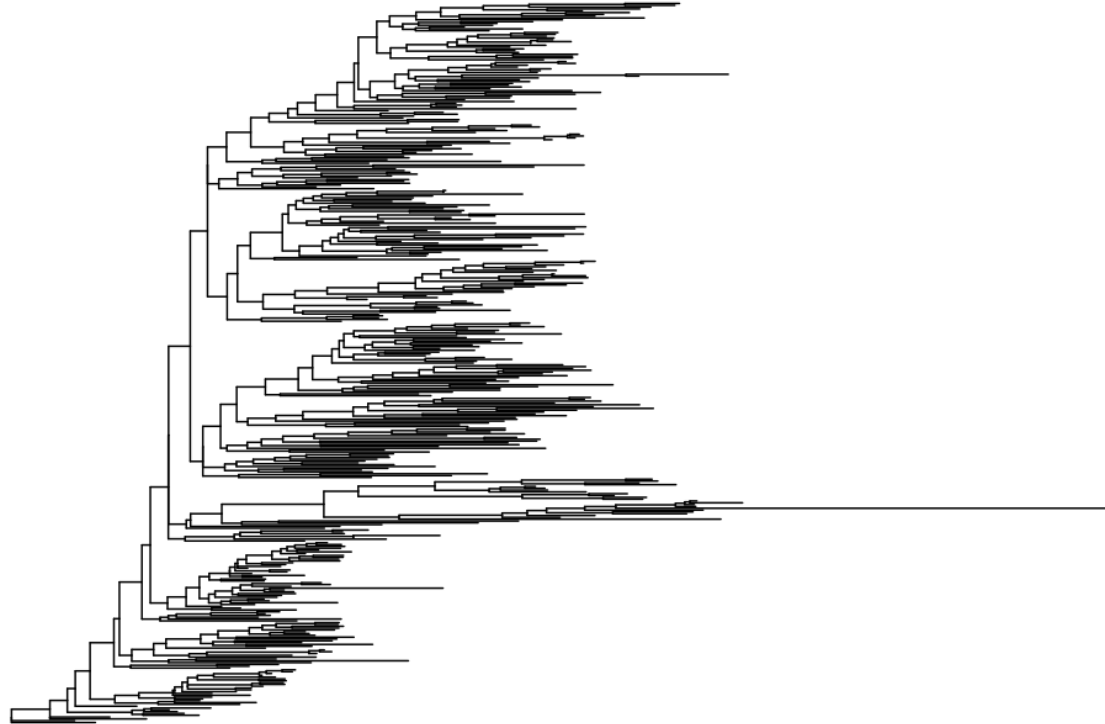
- Unique fraction distance of all sample-pairs
- PERMANOVA

Phylogenetic Community Structure

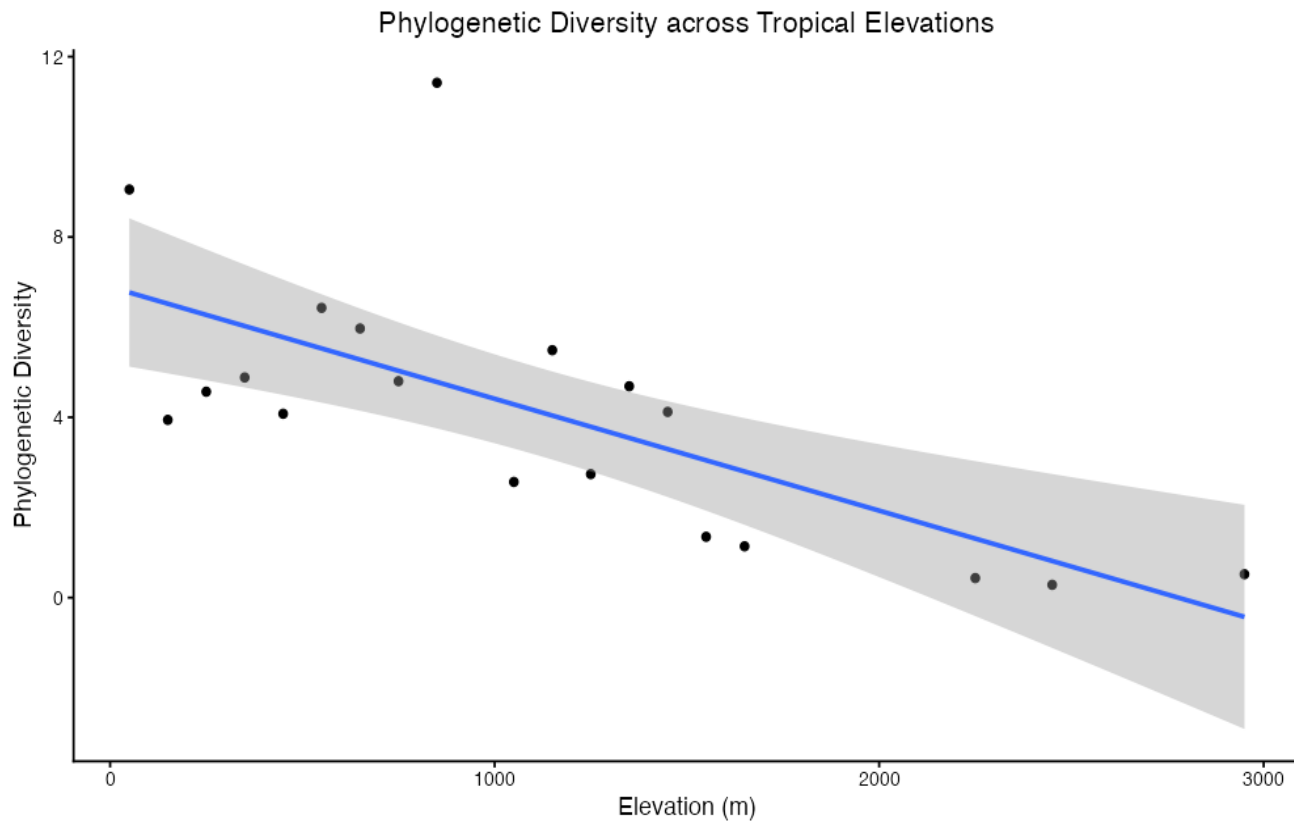
Nearest Taxon Index (NTI)

- Measure of phylogenetic clustering or dispersion
- Standardized effect size (SES) of mean nearest taxon distance (MNTD)

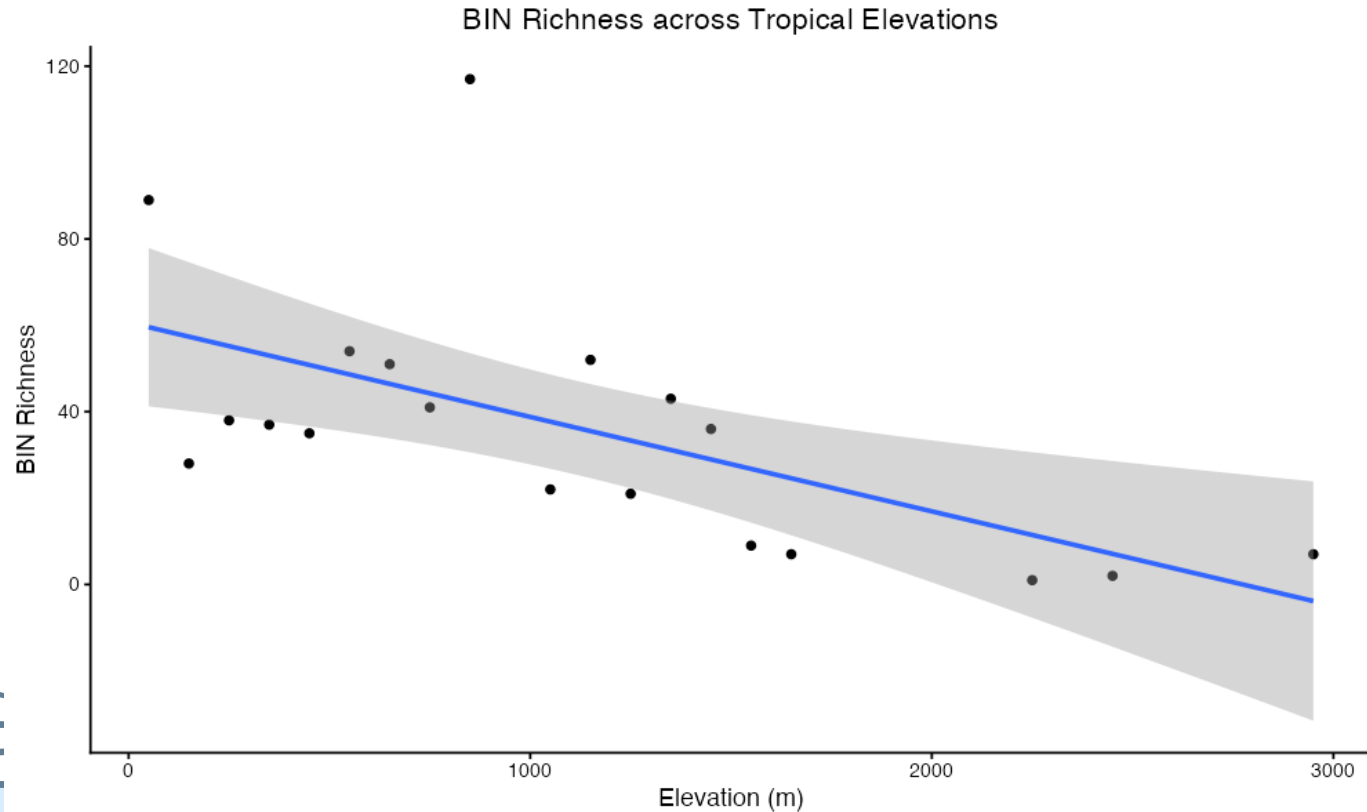
Maximum Likelihood COI-5P Phylogeny of Tropical Carabidae (397 BINs)



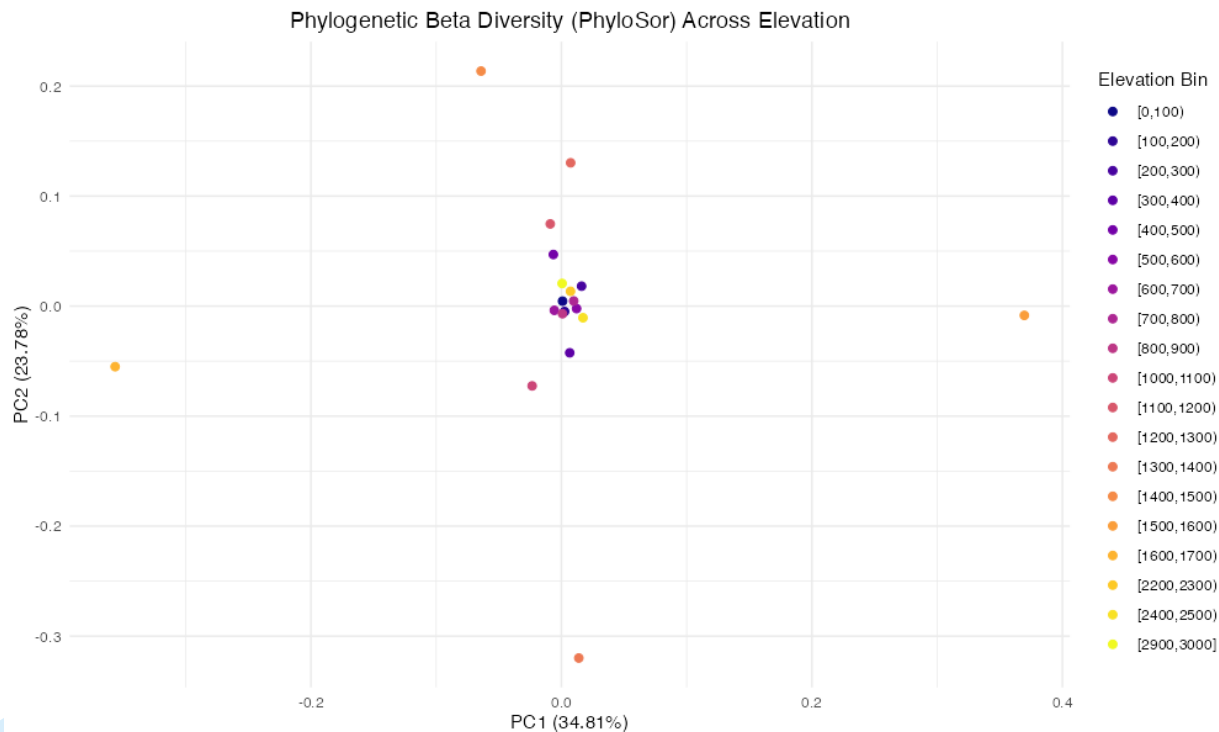
Phylogenetic Diversity Decreases as Elevation Increases



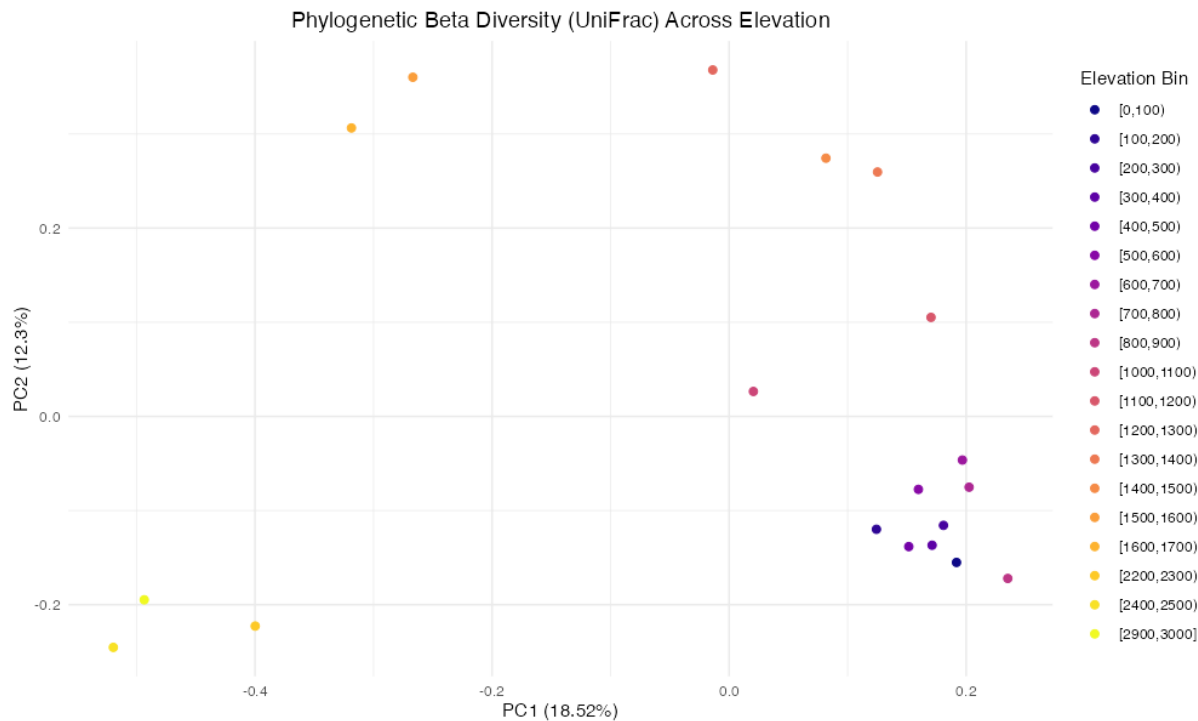
BIN Richness Decreases as Elevation Increases



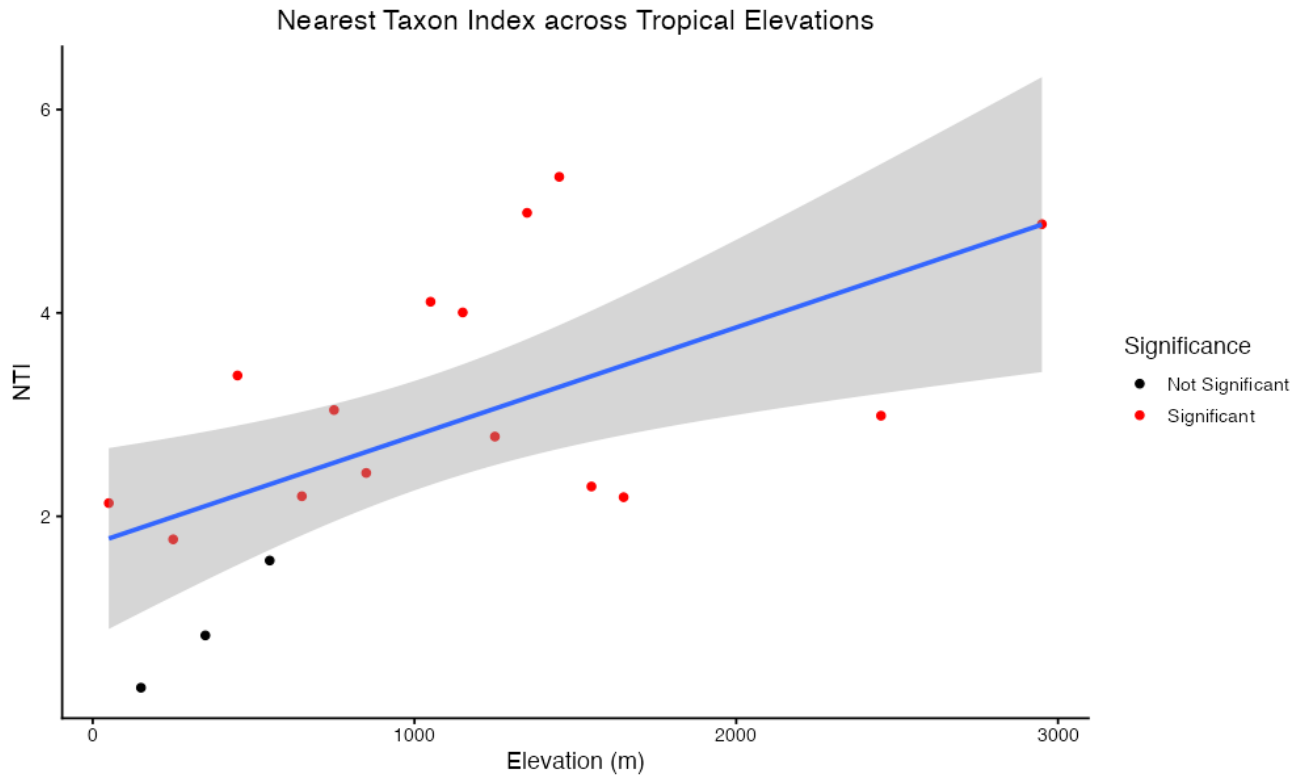
Low/Mid and Peak Elevations are Compositionally Homogenized, whereas Mid-High Transition Zone Reflects Turnover



Low Elevations Have Similar Community Composition, Mid Elevations Exhibit Divergence, and High Elevations Contain Distinct Lineages



NTI Increases with Elevation & Mid/High Elevations Suggest Strong Clustering, unlike Low Elevations



How Does Tropical Carabidae Phylogenetic Diversity Vary Across Elevations?

Alpha Diversity

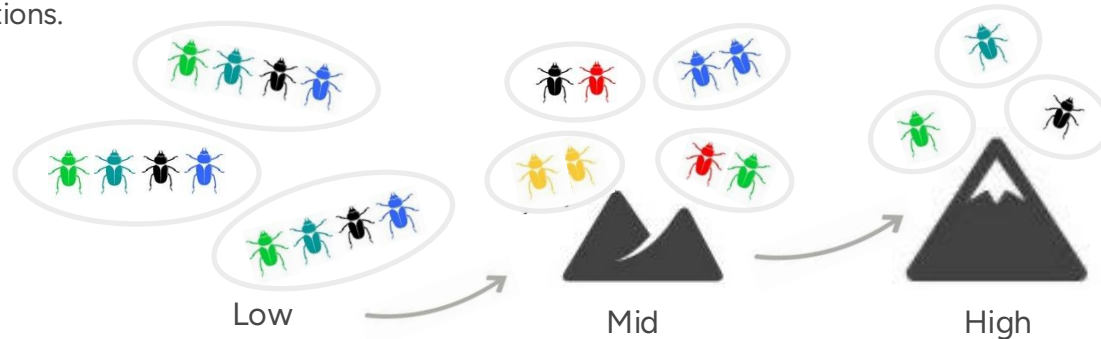
From this analysis, it was found that both phylogenetic diversity and BIN richness were negatively and linearly related to elevation ($p = 0.001$, $p = 0.006$, respectively). This indicates an overall lower degree of within-assemblage (elevation bands) diversity at higher elevations.

Beta Diversity

The Phylosor PCoA analysis revealed clustering at low and mid elevations, while the transition zone from mid to high elevations reflected turnover in community composition. The Unifrac PCoA showed similar findings where low elevation communities clustered together, mid elevation communities are more divergent, and high elevation communities consist of distinct lineages. PERMANOVA confirmed that elevation significantly structures communities with a moderate effect size ($R^2 = 0.25$, $p = 0.001$).

Phylogenetic Community Structure

The Nearest Taxon Index (NTI) values increased with elevation, suggesting stronger phylogenetic clustering at mid and high elevations.



Which Mechanism(s) Shape Phylogenetic Community Structure?

These results provide nuanced support for the three competing hypotheses posited for Carabidae community assembly along tropical elevation gradients.

For **H1 (environmental filtering)**, the low richness and diversity found at high elevations, coupled with the strong phylogenetic clustering suggests that environmental stressors produce fewer communities of more closely related species. This is consistent with the observed increase in NTI with elevation, in that higher elevations favour more-adapted lineages clustered by shared ecological tolerances.

For **H2 (sky-island effects)**, the dissimilar assemblages in the transition zone from mid to high elevations exhibits high turnover and can lead to spatial structure and divergence which supports the idea that somewhat isolated sites ("sky-islands") can lead to distinct lineages.

For **H3 (competition-driven overdispersion)**, the low elevation communities are not strongly overdispersed as indicated by the positive NTI values despite the high alpha diversity and strong phylogenetic clustering suggesting competitive effects at low elevations are weak.

Overall, these findings suggest that **multiple assembly processes act simultaneously along the elevation gradient**: environmental filtering dominates at high elevations, isolation and turnover shape mid- to high-elevation communities, but biotic interactions are not the primary structuring process at low elevations.

Caveats & Limitations

While this study provides insights into phylogenetic community assembly patterns in ground beetles in tropical montane regions, there are several caveats that may have biased the findings.

First, the relatively **small breadth and depth of sampling size** may limit the generalizability of the results and reduce the statistical power for detecting subtle patterns. More specifically, there was uneven or incomplete sampling coverage of across the full elevational gradient in that Carabidae specimens were only collected in 19 of the 30 possible intervals. Additionally, low or uneven number of sequences per elevation band can introduce noise to diversity measures or comparisons.

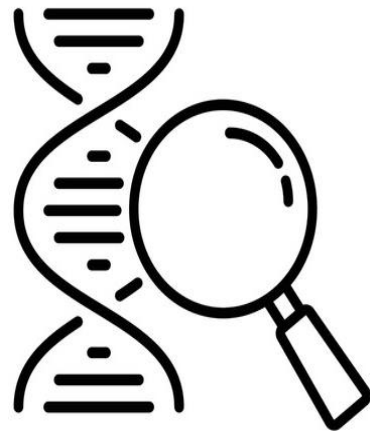
Second, barcode-only phylogenies have **limited reliability in resolving deep evolutionary relationships** or higher-level relationships due to their high degree of saturation (Talavera et al., 2022). Furthermore, reconstructing a tree from a single mitochondrial locus may not capture hybridization, recombination, or complex evolutionary histories (Blanco-Bercial et al., 2014).

Finally, the BOLD database may be **biased toward certain taxa or regions** which could influence which lineages were included in the analysis and affect phylogenetic diversity and structure estimates.

Future Directions

To strengthen and extend this study, follow-up research with specific intentions and goals would be valuable.

1. **Increase sampling across the elevational gradient**, particularly by filling missing bands and improving specimen representation per site to strengthen PD, NTI, and turnover estimates as Carabidae remain under-sampled in tropical montane ecosystems (Loureiro and Smith, 2025).
2. **Build a more resolved phylogeny using multilocus or genomic data**, reducing the limitations of COI-only trees for deeper branching to improve the accuracy of PD or UniFrac analyses (Pentinsaari et al., 2016).
3. **Incorporate environmental variables and functional traits** (e.g., temperature, habitat structure, morphology, thermal tolerance) to directly test mechanisms driving filtering vs. competition (Cadotte & Tucker 2017).
4. **Evaluate geographic isolation (“sky island”) effects** with spatial analyses to explain the pronounced turnover and unique high-elevation lineages detected in this study.



Reflection

Completing this project has afforded me with the opportunity to explore and integrate new phylogenetic and ecological research surrounding my fundamental interests and experience in DNA barcoding and insect biodiversity. Between this project and the entire course, I have significantly strengthened my programmatic skills in R for efficient data preparation, advanced statistical analyses and informative visualizations. I have also gained experience utilizing other tools like MEGA to conduct sophisticated cross-platform research projects.

Moving forward, I aim to further develop my computational and statistical skills to test novel ecological hypotheses and refine my ability to effectively communicate complex results. Furthermore, I will continue to take careful consideration of data organization and quality control to achieve robust results. Lastly, I will not underestimate the importance of time management and adaptability in research planning as challenges will inevitably arise.

Acknowledgements

I would like to extend thanks to Alex Smith and Thanuja Fernando for providing me with critical feedback and insights on resolving deep phylogenetic relationships and the limitations of COI-based tree reconstruction. Their comments highlighted the importance of tree-rooting decisions and subsequent phylogenetic community structure inferences.

I would also like to thank Ramya Manjunath for giving me tips on how to improve R code efficiency through the use of lapply vectorizations.

Finally, I would like to sincerely thank Farah Sadoon for her guidance and support throughout this project. Her insights on analysis tools and methodological choices were invaluable in helping me develop a robust and reliable workflow. Additionally, her expertise in creating visually appealing and creative presentations was tremendously helpful.

Resources:

<https://iqtree.github.io/doc/Tutorial>

<https://www.rdocumentation.org/packages/phylotools/versions/0.2.2>

<https://www.rdocumentation.org/packages/phytools/versions/2.5-2>

<https://pedrohbraga.github.io/CommunityPhylogenetics-Workshop/CommunityPhylogenetics-Workshop.html>

References

- Blanco-Bercial, L., Cornils, A., Copley, N. J., & Bucklin, A. (2014). DNA Barcoding of Marine Copepods: Assessment of Analytical Approaches to Species Identification. *PLOS Currents Tree of Life*, <https://doi.org/10.1371/currents.tol.cdf8b74881f87e3b01d56b43791626d2>
- Cadotte, M. W., & Tucker, C. M. (2017). Should Environmental Filtering be Abandoned?. *Trends in ecology & evolution*, 32(6), 429–437. <https://doi.org/10.1016/j.tree.2017.03.004>
- Dolson, S. J., Loewen, E., Jones, K., Jacobs, S. R., Solis, A., Hallwachs, W., Brunke, A. J., Janzen, D. H., & Smith, A. (2021). Diversity and phylogenetic community structure across elevation during climate change in a family of hyperdiverse neotropical beetles (Staphylinidae). *Ecography*, 44(5), 740–752. <https://doi.org/10.1111/ecog.05427>
- Hoiss, B., Krauss, J., Potts, S. G., Roberts, S., & Steffan-Dewenter, I. (2012). Altitude acts as an environmental filter on phylogenetic composition, traits and diversity in bee communities. *Proceedings. Biological sciences*, 279(1746), 4447–4456. <https://doi.org/10.1098/rspb.2012.1581>
- Loureiro, A. M. M. C., & Smith, A. (2025). No evidence of a common pattern of taxon or phylogenetic diversity across elevation for beetle (Coleoptera) families. *Ecological Entomology*, 50(4), 595–608. <https://doi.org/10.1111/een.13447>
- Montaño-Centellas, F. A., & Loiselle, B. A. (2021). Ecological drivers of avian community assembly along a tropical elevation gradient. *Ecography*, 44(4), 574–588. <https://doi.org/10.1111/ecog.05379>
- Pentinsaari, M., Salmela, H., Mutanen, M., & Roslin, T. (2016). Molecular evolution of a widely-adopted taxonomic marker (COI) across the animal tree of life. *Sci Rep* 6, 35275. <https://doi.org/10.1038/srep35275>
- Talavera, G., Lukhtanov, V., Pierce, N. E., & Vila, R. (2022). DNA Barcodes Combined with Multilocus Data of Representative Taxa Can Generate Reliable Higher-Level Phylogenies. *Systematic biology*, 71(2), 382–395. <https://doi.org/10.1093/sysbio/syab038>