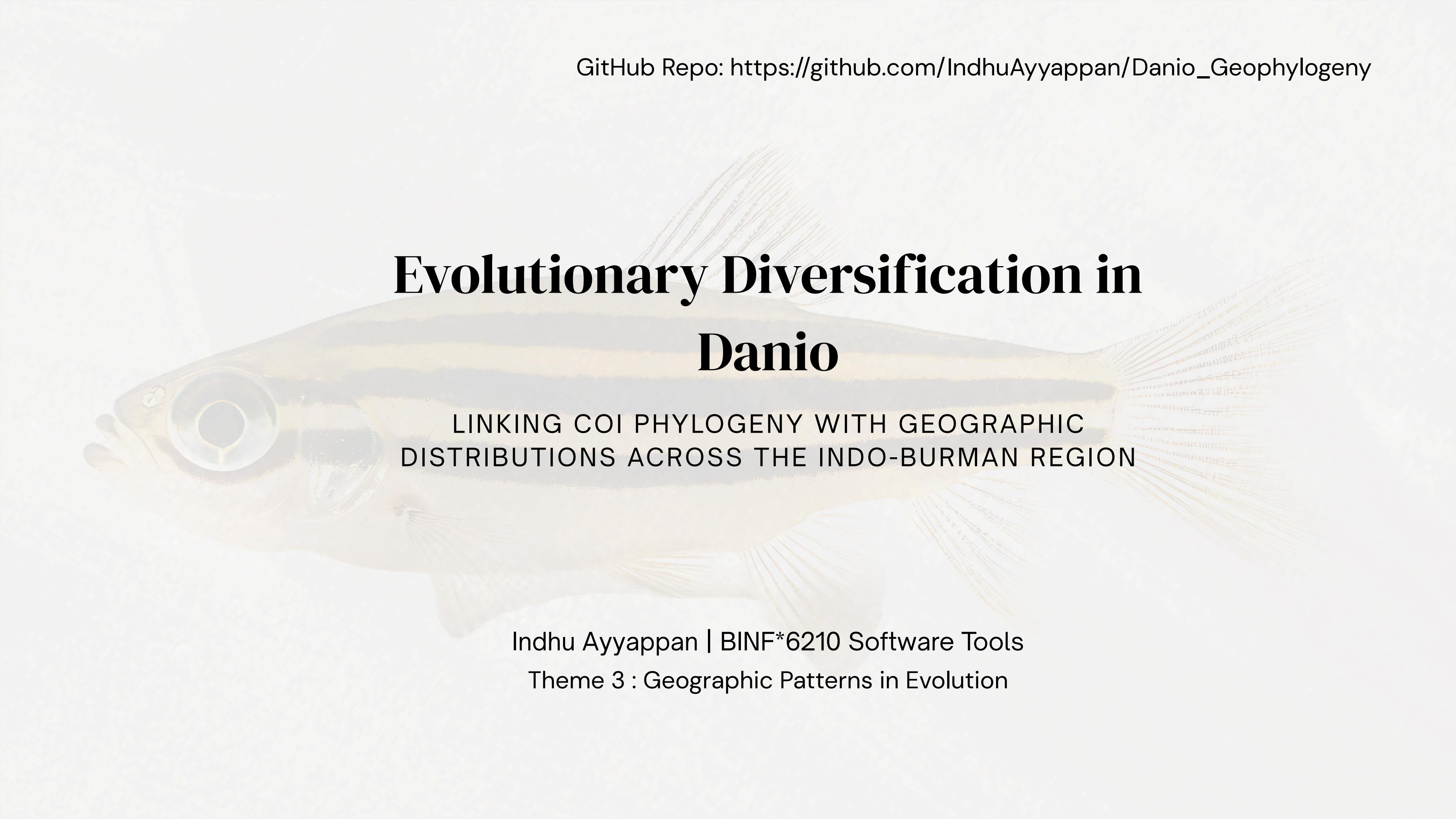


GitHub Repo: https://github.com/IndhuAyyappan/Danio_Geophylogeny



Evolutionary Diversification in Danio

LINKING COI PHYLOGENY WITH GEOGRAPHIC
DISTRIBUTIONS ACROSS THE INDO-BURMAN REGION

Indhu Ayyappan | BINF*6210 Software Tools
Theme 3 : Geographic Patterns in Evolution

Danio species form a diverse freshwater radiation centered in South and Southeast Asia

Why this matters :

- Danio (including the model organism *D. rerio*) are small cyprinid fishes widely used in ecology and evolutionary studies.
- Their greatest species richness occurs in the Indo–Burman biodiversity hotspot (NE India, Bangladesh, Myanmar, Thailand).
- Many species occupy restricted, non-overlapping river basins, making them ideal for studying geographic diversification.
- Previous studies show deep phylogenetic structure, but relationships between clades and geography remain unclear.
- This assignment examines whether related species live in similar or distinct regions, using COI phylogeny + GBIF occurrence data.

This study tests whether closely related Danio species occupy similar or distinct geographic regions

Objective: Determine whether phylogenetically related Danio species live in overlapping or non-overlapping regions

- **Study type:** Exploratory evolutionary analysis combining phylogeny + spatial data
- **Approach:** Build COI tree from NCBI → get GBIF occurrence → link geography + tree
- **Hypothesis:** Sister species will show limited range overlap, supporting local diversification
- **Expected outcome:** Identify geographic hotspots + evaluate diversification patterns

This project integrates publicly available COI sequences from NCBI with GBIF occurrence data

COI sequences (NCBI):

- Source: NCBI nucleotide database
- Data: All *Danio* COI sequences available as of November 2025
- Initial set: ~150 sequences (varies per species)
- Key variable: DNA barcode sequence (cytochrome oxidase I)
- Processing: Filter by length (500–800 bp), keep longest per species



GBIF Occurrence Data:

- Source: GBIF.org API
- Records: Up to 500 cleaned occurrences per species
- Key variables: Latitude, longitude, coordinate uncertainty, species name
- Filters applied: Remove missing coords, large uncertainty, ocean points



Integrated phylogeny + geography reveal evolutionary patterns in *Danio*

COI sequences (NCBI)

- Download COI sequences
- Filter by length
- Keep longest per species

GBIF occurrence data

- Retrieve up to 500 records
- Clean coordinates
- Match species names



Sequence processing

- Align sequences (DECIPHER)
- Build NJ tree (JC69)
- Root with *Microrasbora*



Geographic processing

- Compute centroids
- Compute richness grids
- Compute sister-pair distances



Integrated analyses

- Geophylogeny
- Sampling intensity
- Sister-species distances
- Richness heatmap



Evolutionary questions

- Do clades cluster geographically?
- Do sister species overlap?
- Where is the diversification hotspot?

COI phylogeny resolves major *Danio* clades consistent with published relationships

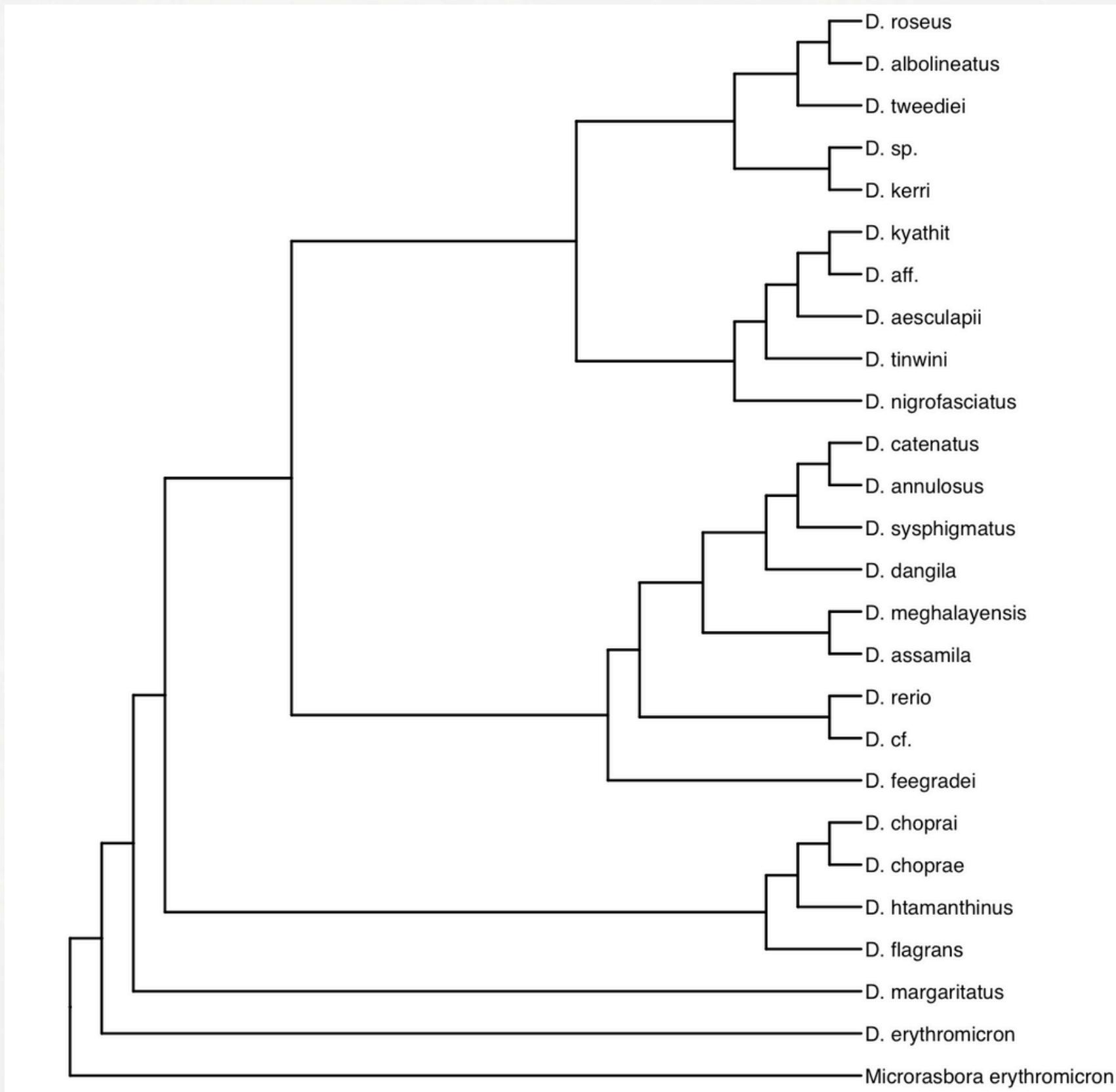


FIGURE 1. COI NEIGHBOR-JOINING PHYLOGENY (JC69 MODEL)

Key steps

- Longest COI sequence per species selected (500–800 bp range)
- Alignment performed using DECIPHER
- NJ phylogeny built under JC69 model
- Tree rooted with *Microrasbora erythromicron*

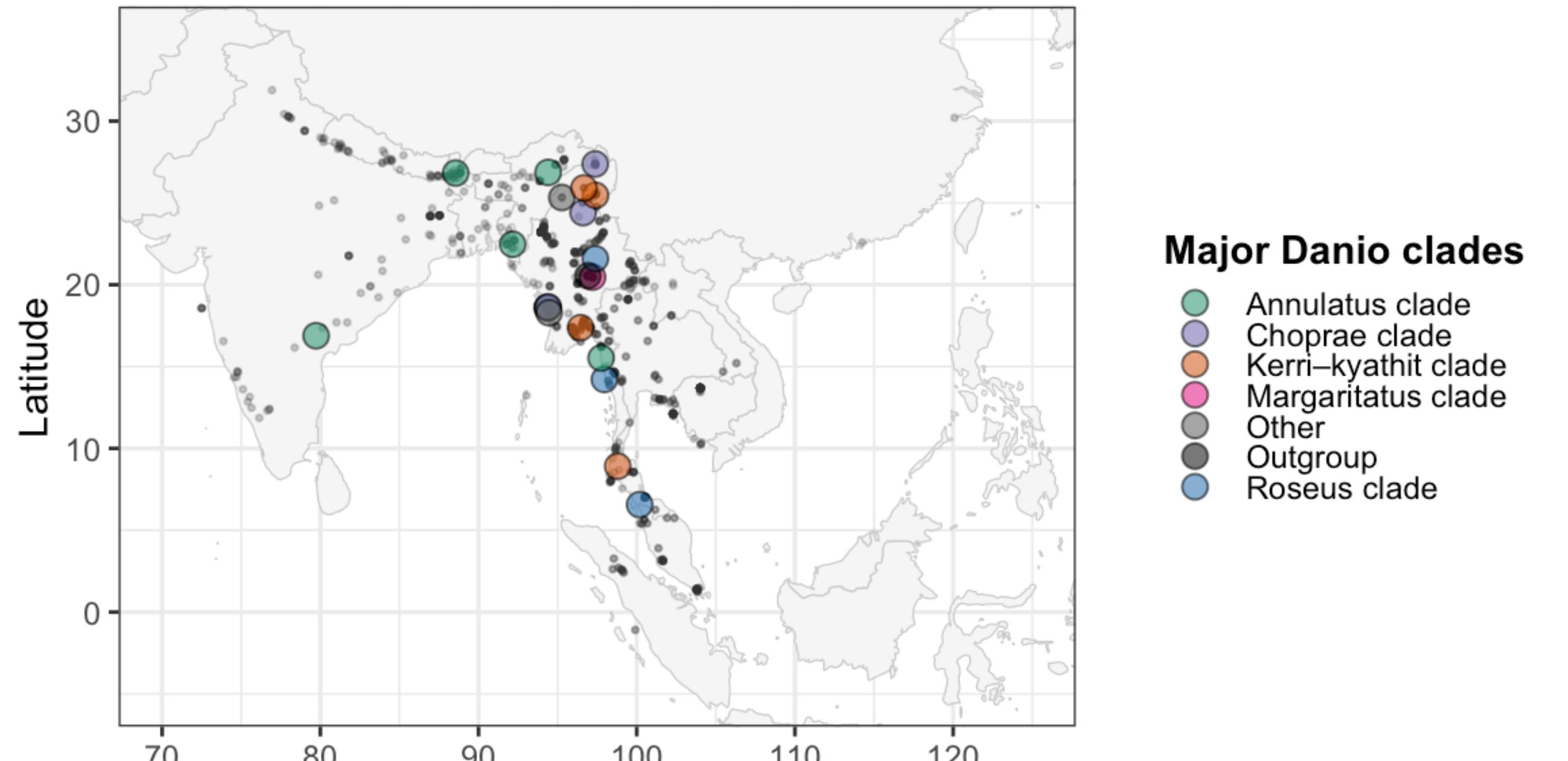
Key findings

- Tree recovers **6 major *Danio* clades**
- Relationships closely match existing cyprinid phylogenies
- Strong backbone supports subsequent geographic comparisons

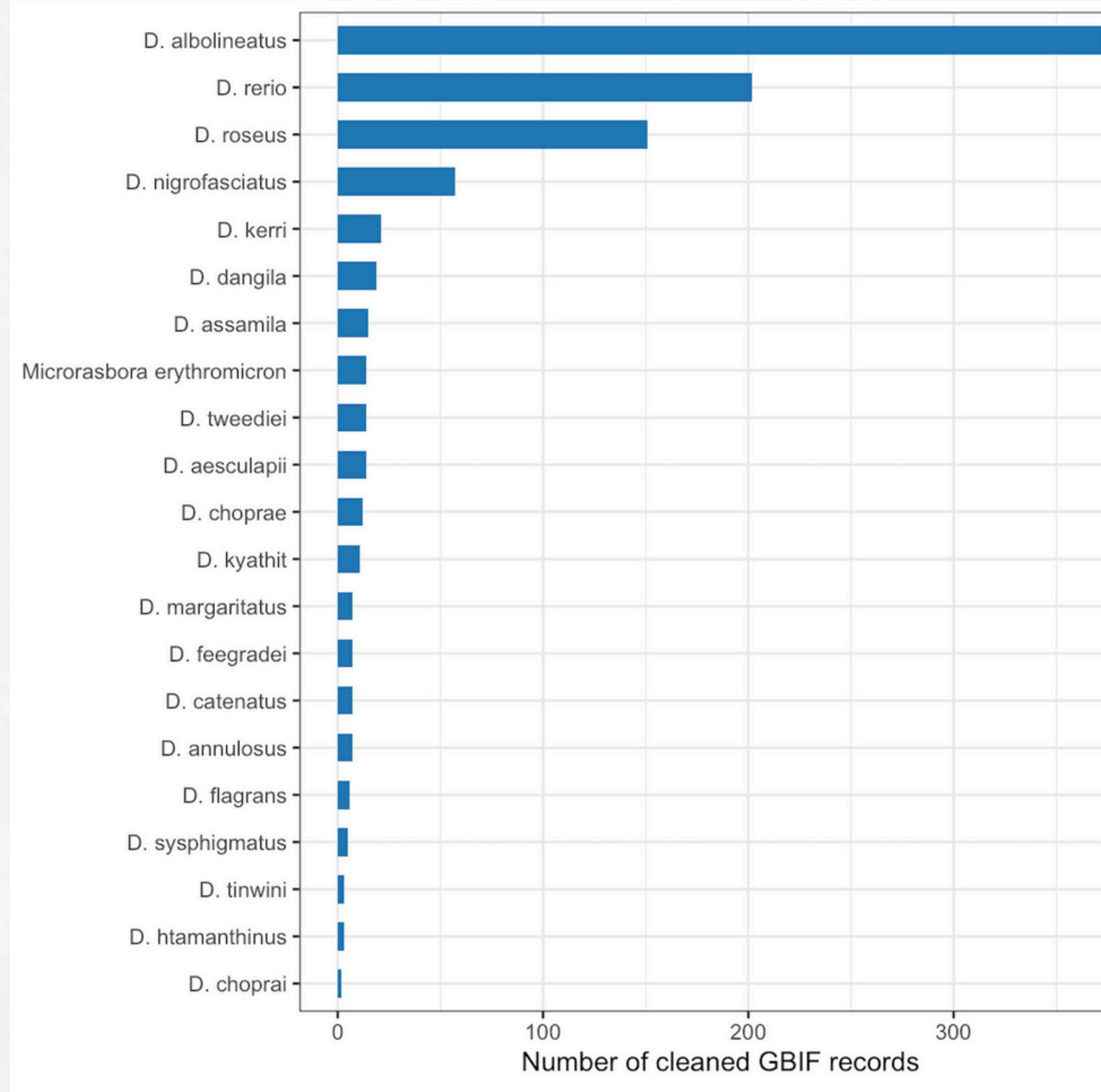
Danio clades show strong geographic clustering in the Indo-Burman hotspot

Key findings

- Most clades cluster in NE India, Bangladesh, and Myanmar
- Geographic clustering strongly matches phylogenetic clade structure
- Indicates a shared regional origin for much of the Danio radiation
- Indicates diversification occurred locally within the hotspot rather than through long-distance dispersal



Sampling intensity is highly uneven across Danio species, with strong bias toward a few taxa

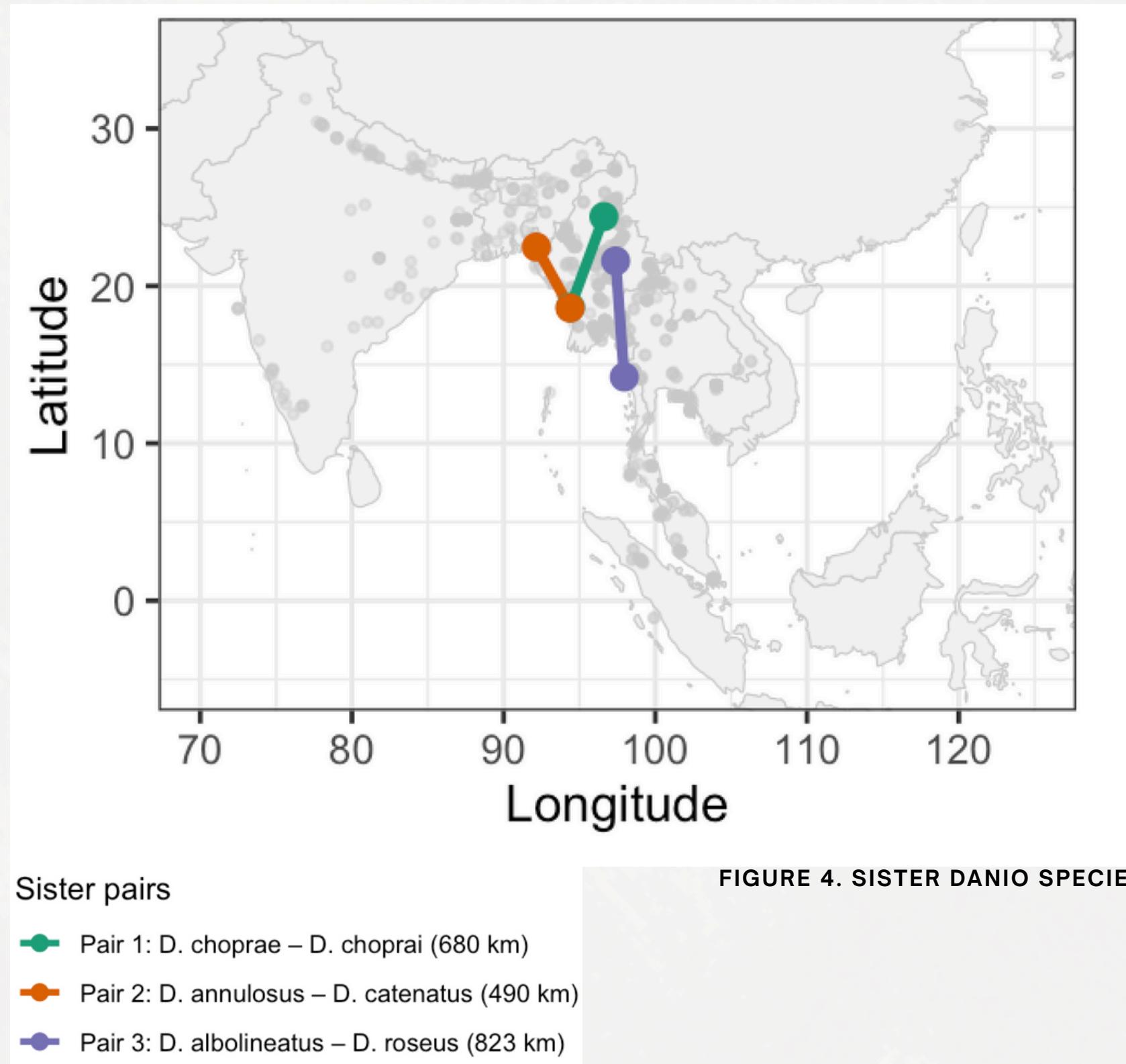


Key findings

- Sampling effort is heavily concentrated in a few widespread species
- Several Danio species have very sparse GBIF representation, limiting geographic inference
- Over-represented species inflate apparent clustering in NE India and Myanmar
- Under-representation in Malaysia, Indonesia, and Laos may contribute to artificially low richness there
- These patterns highlight the need to interpret geographic analyses in the context of sampling bias

FIGURE 3. SAMPLING INTENSITY BARPLOT

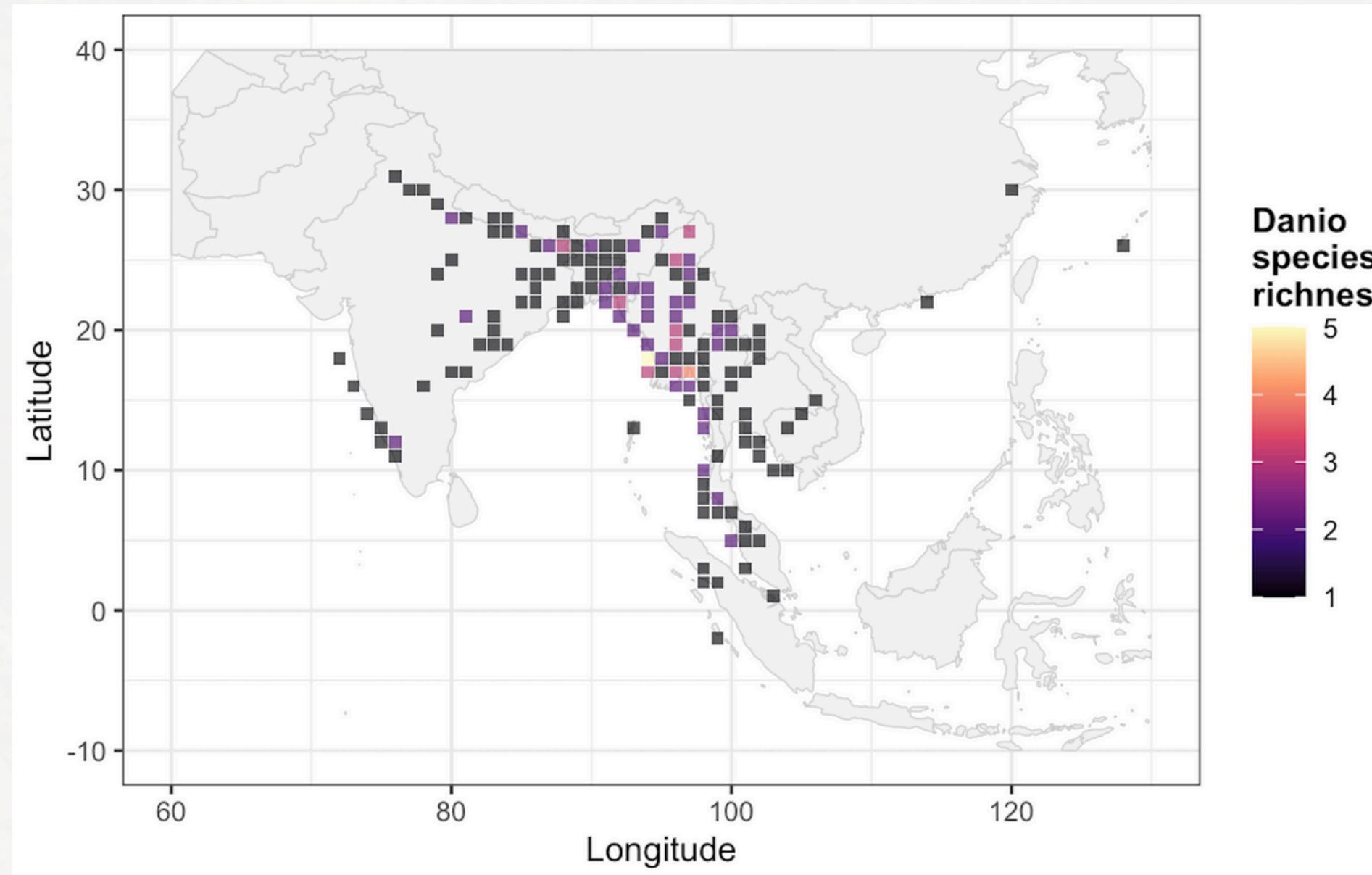
Closely related species live near each other but not together, suggesting local diversification



Key findings

- Sister-species distances range from ~490 km to ~823 km
- No sister pair is fully sympatric — each occupies a distinct but nearby region
- Distances fall within the Indo-Burman region, not across continents
- Pattern indicates local geographic divergence rather than long-distance dispersal
- Supports a model of regional niche partitioning among closely related species

Danio species richness peaks sharply in the Indo-Burman region, forming a clear biodiversity hotspot



Key findings

- Highest richness (up to 5 species per grid cell) occurs in eastern India, Bangladesh, and Myanmar
- Richness declines rapidly moving east toward Thailand and south toward Malaysia
- Hotspot location matches major phylogenetic clades and GBIF centroid clustering
- Pattern supports a single center of diversification with limited spread into surrounding regions

Danio diversification is concentrated in the Indo-Burman hotspot, where geography and phylogeny strongly align

Key conclusions

- COI phylogeny reveals six major Danio clades with deep, well resolved structure
- Geophylogeny shows that related species occur in the same broad region, mainly NE India, Bangladesh, and Myanmar
- Sister species occupy adjacent but non-overlapping ranges, supporting local divergence rather than dispersal
- Species richness peaks in the same hotspot, indicating a shared center of origin
- Geographic patterns and phylogenetic patterns agree, suggesting regional diversification followed by range partitioning

Overall:

Danio evolution appears shaped by local geographic diversification within the Indo-Burman hotspot, rather than long distance dispersal or widespread sympatry.

References

- Fang K., Norén M., Liao T.Y., Källersjö M., Kullander S.O. (2009). Phylogeny and species diversity of the South and Southeast Asian cyprinid genus *Danio* Hamilton (Teleostei, Cyprinidae).
- Mayden R.L., Tang K.L., Conway K.W., Freyhof J., Chamberlain S., Haskins M., Schneider L., Sudkamp M., Wood R.M., Agnew M., Bufalino A., Sulaiman Z., Miya M., Saitoh K., He S. (2007). Phylogenetic relationships of *Danio* within the order Cypriniformes: a framework for comparative and evolutionary studies of a model species. *J. Exp. Zool. (Mol. Dev. Evol.)* 308B:642–654.
- McCluskey B.M., Stout C.C., Conway K.W., Mayden R.L., Chen W.J. (2014). Phylogeny of Zebrafish, a “Model Species,” within *Danio*, a “Model Genus”.
- Conith M.R., Collar D.C., Holcombe S., Wainwright P.C. (2022). Evolution of Feeding Mechanics in the Danioninae, or Why Teleosts Eat with Four Jaws. *Integrative Organismal Biology*.
- Sholihah A., et al. (2020). DNA barcode reference library reveals cryptic diversity in Sundaland Rasborinae. *Sci. Rep.*
- “*Danio*.” ScienceDirect Topics. Elsevier.

The End

THANK YOU FOR VIEWING MY ANALYSIS!

Indhu Ayyappan

BINF 6210 - ASSIGNMENT 4