# An investigation of Geographic Distributions of Tardigrade using R

Bartek Puzio

# Tardigrades are found all over the World

- Tardigrades are highly resilient and tiny animals that are capable of surviving extreme conditions.
- Different species have adapted to different conditions globally.
- Most samples have been collected from the Northern hemisphere.
- Investigating the distribution across the Northern Hemisphere can provide insights into geographic patterns and biodiversity

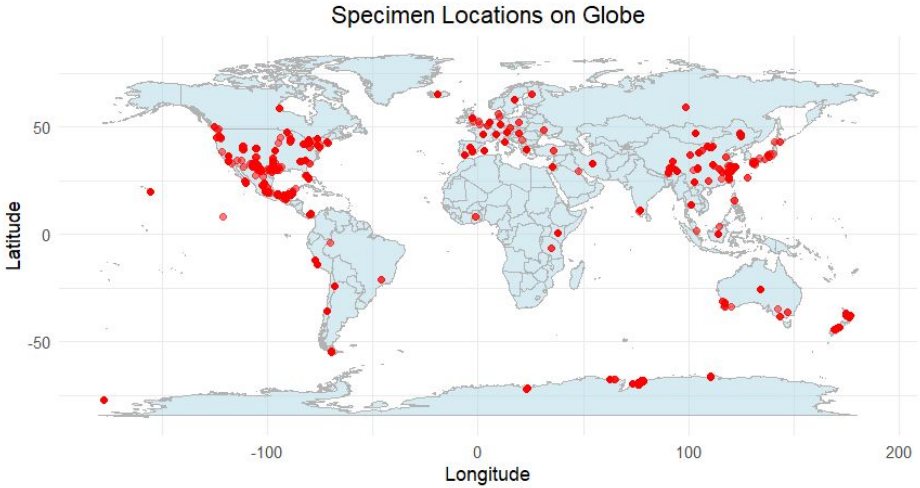
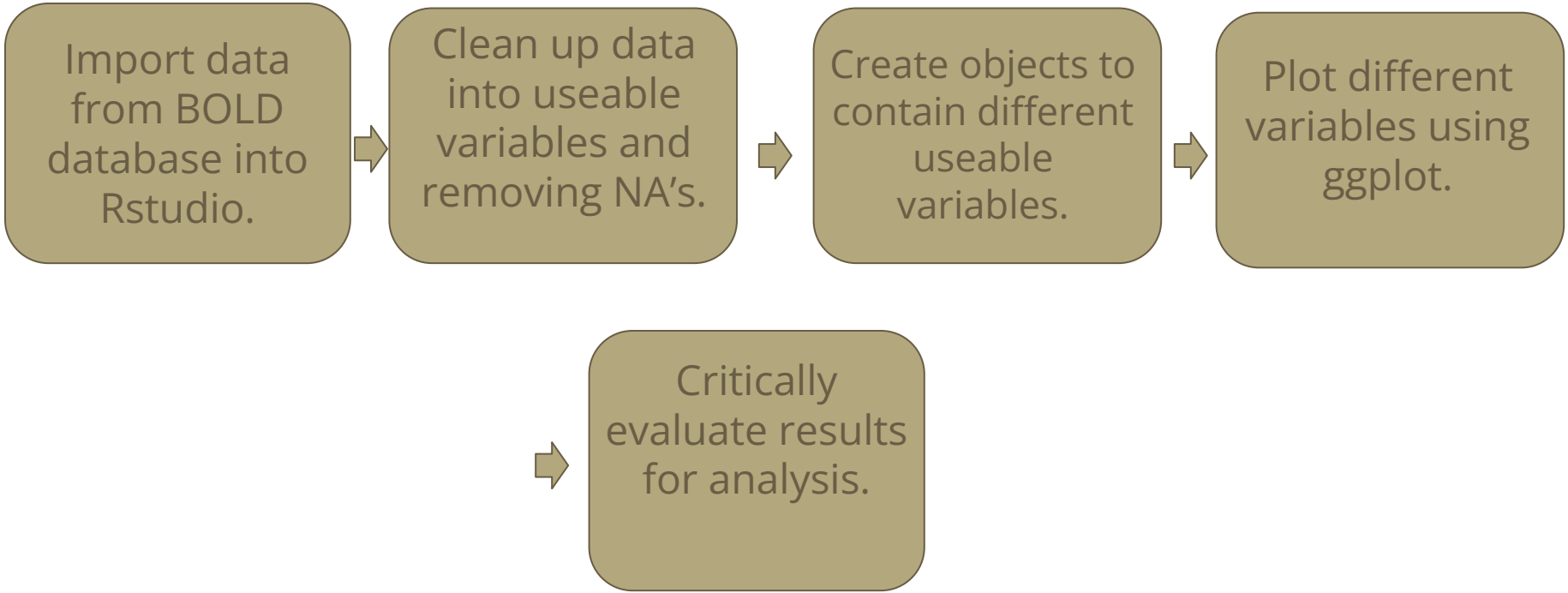



**Figure 1:** All available Tardigrade sampling locations from BOLD database. Where exact coordinates were not available, country centroid coordinates were used in lieu of sampling location. Graph generated from Rstudio using maps package by Puzio 2025

# How does BIN richness vary across latitude bands in the Northern Hemisphere?

# Methods


Import data from BOLD database into Rstudio.
Clean up data into useable variables and removing NA's.
Create objects to contain different useable variables.
Plot different variables using ggplot.
Critically evaluate results for analysis.

# Tardigrade sampling frequency at 40° Latitude

- The latitude of 40° reflects sampling sites across the Northern US, Europe and North Asia.
- The sampling frequency at this latitude does not directly correlate to distribution patterns due to sampling bias in BOLD.

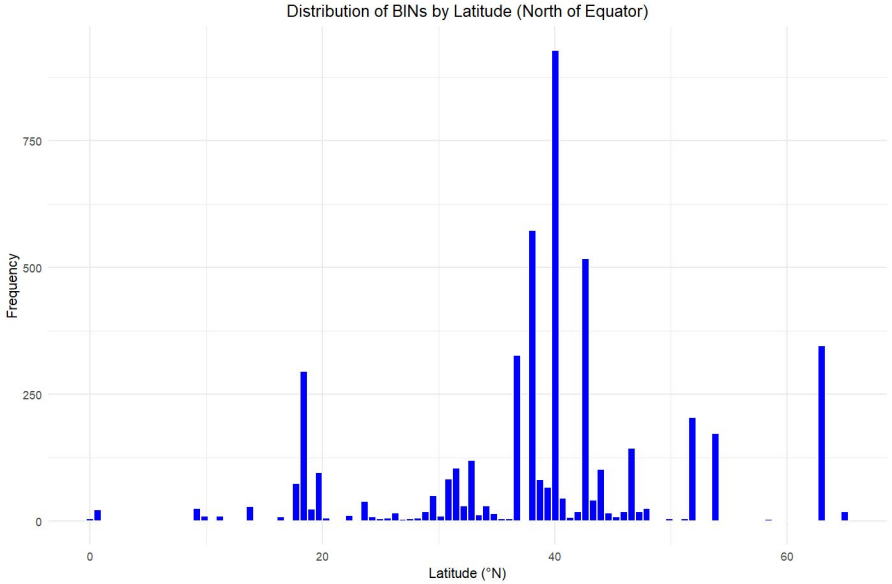



**Figure 2:** Distribution of sampling coordinates arranged by latitude against frequency. The largest frequency of recorded latitudes was found at 40° Latitude. Graph generated using Rstudio by Puzio 2025

# A different approach to distribution by Latitude

- Distribution of BINs vs distribution of unique Bins.
- BIN richness is highest at 20° rather than 40°, which figure 2 would suggest.
- Unique bins at these latitudes reflects genetically differentiated populations.

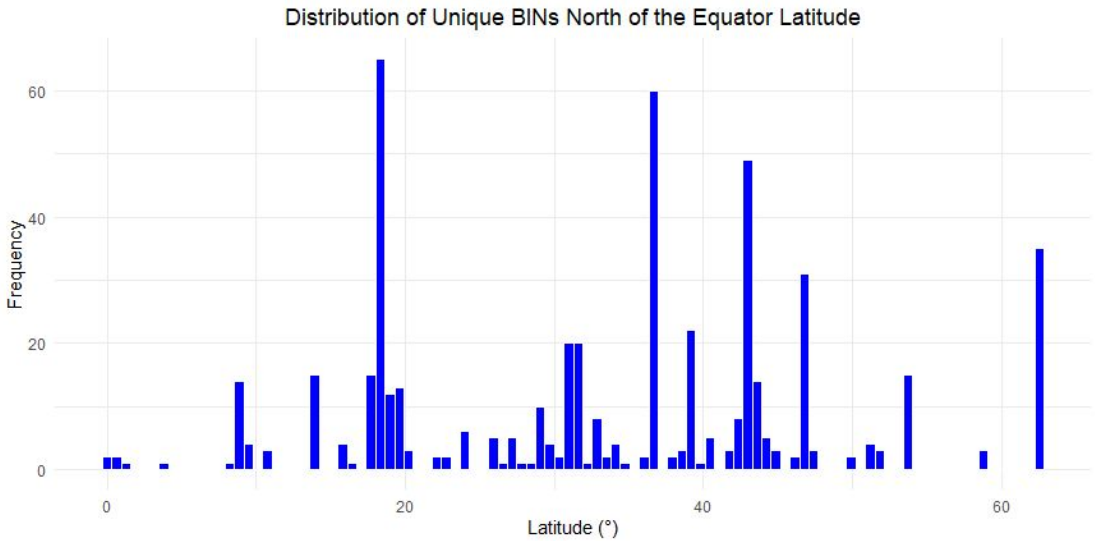



**Figure 3:** Distribution of unique BINs of Tardigrade organized by Latitude north of the equator. This graph provides deeper insights into the regions of concentrated biodiversity in the Northern hemisphere. Graph generated using Rstudio by Puzio 2025

# How does BIN richness vary across latitude bands in the Northern Hemisphere?

- Mid-latitude regions (Europe, North America, East Asia) are more frequently studied and better represented in the BOLD database.
- Temperate zones have diverse ecosystems that support a wide range of tardigrade habitats.
- Extreme heat or cold could reduce metabolic activity and population growth in Tardigrades.

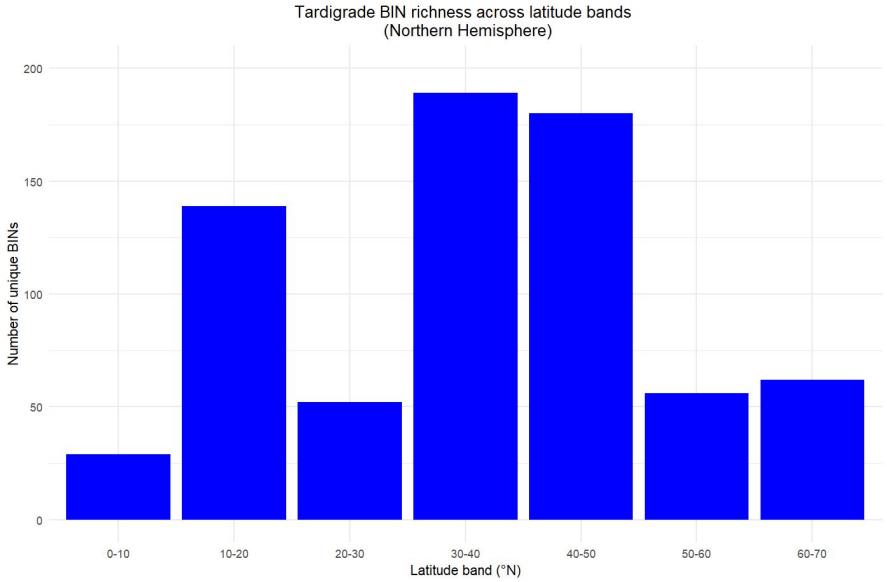



**Figure 4:** Tardigrade BIN richness across latitude bands in the Northern Hemisphere. Peaks between 30–50°N suggest greater genetic diversity or sampling intensity in these regions. Graph generated using Rstudio by Puzio 2025

# Cryptic species are prevalent in the Northern Hemisphere

- In the top 15 countries sampled from the dataset in the northern hemisphere, each country had more BINS's than named species.
- This result may be explained by the occurrence of **Cryptic species** that are yet to be classified.
- Different populations of the same species can become genetically divergent due to geographic isolation over time.

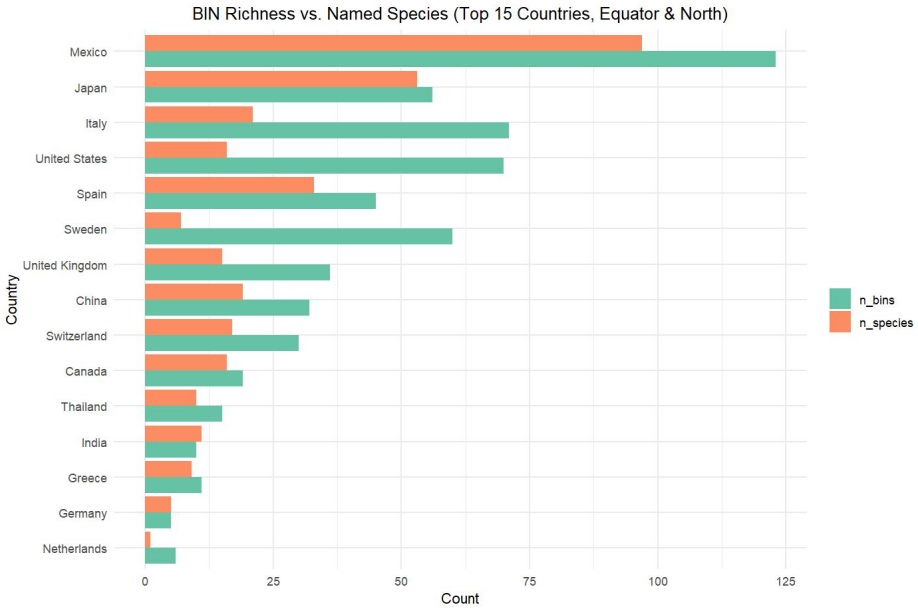



**Figure 5:** BIN richness vs named species from the top 15 countries that lie in Northern hemisphere. More bins are found comparatively against named species. Graph was generated in Rstudio by Puzio 2025

# Dataset limitations and Findings

- The dataset was limited in geographic coverage, particularly in tropical and Arctic regions, which likely underestimates true global diversity.
- Sampling bias was encountered at 40° Latitude, misleading initial findings.
- More BINs have been identified than species consistently across the dataset, implying the presence of cryptic species in this phyla
- Environmental metadata could be used to determine stronger conclusions in habitat selection

```r
data.b.ratio <- data.b %>% filter(!is.na(lat), lat > 0)

n_bins <- n_distinct(data.b.ratio$bin_uri, na.rm = TRUE)
n_species <- n_distinct(data.b.ratio$species, na.rm = TRUE)
ratio <- n_bins / n_species

c(n_bins = n_bins, n_species = n_species, ratio = ratio)

# Result: 1.81. This suggests that there are more distinctive clusters, or mitochondrial lineages, than named species in this dataset. This can be an indicator of cryptic species diversity or populations that have been geographically isolated for some time.
```

**Figure 6:** Code from Rstudio calculating the ratio between unique BINs and unique species from the dataset. A ratio of 1.81 implies almost twice as many unique BINs to unique species. Code was generated by Puzio 2025

# Dataset limitations and Findings (Continued)

- These results were partly consistent with my expectations. I anticipated that tardigrades, being tolerant of extreme conditions, would show some diversity across all latitudes.
- I expected a normal distribution around the equator rather than a sharp mid-latitude peak at 20°.
- This hypothesis stems from the tropical climate and favourable conditions near the equator for Tardigrade proliferation.
- This project deepened my understanding of how latitude, climate, and sampling bias work together to shape observed biodiversity patterns from BOLD.

# Next steps

- I would increase the number of sampling sites outside of the most frequently sampled latitudes to produce a wider dataset that includes more habitats in different regions and climates.
- I would also aim to improve the quality of entry data in the BOLD database to include more metadata that can be used for research instead of using NA.
- Record sampling effort of different regions and normalize data before beginning analysis to improve any statistical tests that could be applied in a larger,deeper project.
- Finally, I would include more metadata in my own investigation in hopes of answering a deeper question using more variables.

# References


https://bookdown.org/rdpeng/RProgDA/basic-plotting-with-ggplot2.html

https://chatgpt.com

https://ggplot2.tidyverse.org/reference/geom_histogram.html

https://r4ds.hadley.nz/data-visualize

https://stackoverflow.com/questions/23130604/plot-coordinates-on-map


- I reached out to peers (Stephanie, Yaz, Maddy and Sodiq) for guidance in setting up my studio (Working directory) and applying relevant packages such as styler for formatting my code and appropriately commenting my code. I used their advice to guide my research and start my project on a good foundation.
- Yaz and I shared our incomplete projects with each other to practice formatting the zip file for our final submission to make sure the pathways used to call the data file were correct and working. This taught us about setting working directories from our studios to make sure the code would run on different devices without issue.
- The help from my peers on this project made me confident in my submission and taught me important lessons in making sure my code can run on different devices.