

Exploring the Geographic Patterns of Chromadorea in Relation to Cricetidae and the Latitudinal Diversity Gradient

1. Introduction

The pattern of higher species richness near the equator which is known as the latitudinal diversity gradient (LDG), has been extensively studied in a myriad of free-living organisms. It is widely acknowledged as one of the earth's predominant patterns of biodiversity, and its existence has been proven in numerous taxa. Yet, there are some exceptions to this theory; for instance, parasites present a particularly intriguing case that challenges the LDG (1). Kamiya et al. (2014) examined parasites of plants, fungi and animals through a meta-analysis of numerous publications, however they were unable to identify any significant correlations between the latitude and parasite species richness (2). It is known that hosts traits can significantly affect parasites species diversity and abundance (3).

Interestingly, nematodes are known to parasitize Cricetidae, one of largest families in the order Rodentia. Notably, there is a lack of significant research investigating the factors affecting the geographical distribution of these parasites. Therefore, this study aimed to analyze the geographical distribution of the class Chromadorea (phylum Nematoda) and assess whether their geographical distribution follows the LDG theory or aligns more closely with the geographical distribution of their host. Data for this analysis was obtained from the Barcode of Life Data (BOLD) system.

2. Code – Part 1

```
## 1 - Packages used ----
```

```
#Uncomment and run if you need to install the package or load the library
```

```
#installed packages for bold data acquisition
```

```
#install.packages("BiocManager")
```

```
#BiocManager::install("sangerseqR")
```

```
#install.packages("remotes")

#remotes::install_github("ropensci/bold")

#library(bold)

#installed package for world map

#install.packages("rworldmap")

library(tidyverse)

library(rworldmap)

library(viridis)

conflicted::conflict_prefer("filter", "dplyr")

## 2 - BOLD data acquiring and saving ----

#df_bold = bold_specimens(taxon = c("Cricetidae", "Chromadorea"))

#write_tsv(df_bold, "Bold.tsv")

## 3 - READING DATA ----

#Remember to set your working directory

#Reading in data and Selecting relevant columns for the study

df_bold = read_tsv("../data/Bold.tsv") %>%

  select(class_name, species_name, lat, lon)

## 4 - Exploring Data and Checking Formatting ----

# Check class

class(df_bold)

# See variables

names(df_bold)

# Summary and check for NAs and overall data structure

summary(df_bold)

# Check for rows with empty strings or space
```

```
df_bold %>%
  filter(if_any(where(is.character), ~ . == "" | . == " "))
```

```
# Filtering out rows with NA value
```

```
df_bold = na.omit(df_bold)
```

3. Code – Part 2

```
## 5 - Map representing worldwide distribution of species of the class Chromadorea ----
```

```
# Tibble for Chromadorea species count grouped based on latitude and longitude
```

```
df_chro_ll = df_bold %>%
```

```
  filter(class_name == "Chromadorea") %>%
```

```
  mutate(lat = round(lat), lon = round(lon)) %>%
```

```
  group_by(lat, lon) %>%
```

```
  summarise(species_count = n_distinct(species_name),
```

```
    .groups = "drop")
```

```
# Get world map data and converting it to data frame for ggplot2
```

```
df_world_map = getMap(resolution = "high") %>%
```

```
  fortify()
```

```
# Plot Chromadorea species geographical distribution on world map
```

```
ggplot() +
```

```
  geom_polygon(data = df_world_map, aes(x = long, y = lat, group = group),
```

```
    fill = "lightgray", color = "gray") +
```

```
  geom_point(data = df_chro_ll, aes(x = lon, y = lat, color = species_count), size = 0.7, alpha
= 0.6) +
```

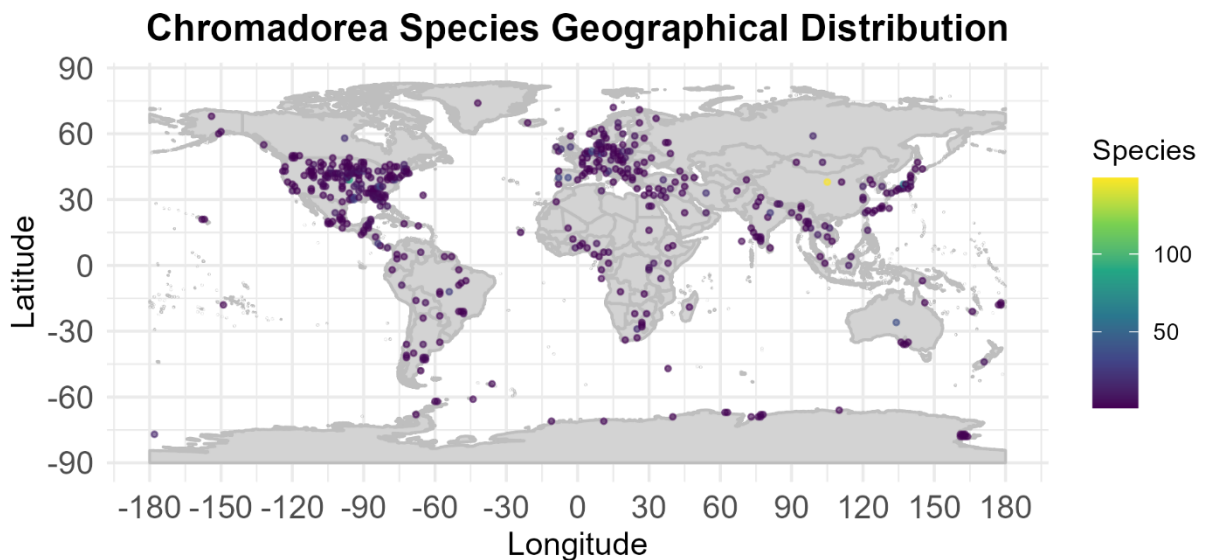
```
  scale_color_viridis_c(option = "viridis", name = "Species") +
```

```
  scale_x_continuous(breaks = seq(-180, 180, by = 30)) +
```

```
  scale_y_continuous(breaks = seq(-90, 90, by = 30)) +
```

```
labs(title = "Chromadorea Species Geographical Distribution",
      x = "Longitude",
      y = "Latitude") +
theme_minimal() +
theme(plot.title = element_text(hjust = 0.5, size = 14, face = "bold"),
      axis.title = element_text(size = 12),
      axis.text = element_text(size = 12))

# Exporting the Plot with size writable for A4 paper
ggsave("Chromadorea Species Geographical Distribution.PNG", width = 6.4, height = 3)
```



```
## 6 - Correlation between distance from equator (degree) and species richness of
Chromadorea and Cricetidae ----

# Tibble with the species count of each targeted taxa grouped based on distance from equator
(degree)

df_dist_eq = df_bold %>%
  mutate(dist_eq_deg = round(abs(lat))) %>% #Converting latitude into distance from equator
(degrees)

group_by(dist_eq_deg) %>%
```

```

summarise(

  Cricetidae = n_distinct(species_name[class_name == "Mammalia"]),

  Chromadorea = n_distinct(species_name[class_name == "Chromadorea"])

)

# Shapiro-Wilk test of normality to determine if the Cricetidae data follows normal
distribution

shapiro_test_cric = shapiro.test(df_dist_eq$Cricetidae[df_dist_eq$Cricetidae != 0]) %>%

  print()

# ==> Result Shapiro-Wilk test

# p-value = 0.000281 ==> data does not follow normal distribution.

# Spearman's Rank Correlation -- To determine if there is a correlation between latitude and
Cricetidae species richness

spearman_cric = cor.test(df_dist_eq$dist_eq_deg[df_dist_eq$Cricetidae != 0],
df_dist_eq$Cricetidae[df_dist_eq$Cricetidae != 0], method = "spearman") %>%

  print()

# ==> Result Spearman's Rank Correlation

# Rho ( $\rho$ ) value = -0.46 ==> moderate negative correlation

# p-value = 2.567e-05 ==> correlation is statistically significant

# There is a statistically significant moderate negative correlation between the distance from
the equator (degrees) and species richness

# Shapiro-Wilk test of normality to determine if the Chromadorea data follows normal
distribution

shapiro_test_chro = shapiro.test(df_dist_eq$Chromadorea[df_dist_eq$Chromadorea != 0])

%>%

  print()

```

```

# ==> Result Shapiro-Wilk test

# p-value = 2.071e-10 ==> data does not follow normal distribution.

# Spearman's Rank Correlation -- To determine if there is a correlation between latitude and
Chromadorea species richness

spearman_chro = cor.test(df_dist_eq$dist_eq_deg[df_dist_eq$Chromadorea != 0],
df_dist_eq$Chromadorea[df_dist_eq$Chromadorea != 0], method = "spearman") %>%
  print()

# ==> Result Spearman's Rank Correlation

# Rho ( $\rho$ ) value = --0.09134011 ==> very weak negative correlation

# p-value = 0.4422 ==> correlation is not statistically significant

# There is no significant relationship between the distance from the equator (degrees) and
species richness

# Combining Cricetidae and Chromadorea species count into a single column

df_dist_eq_long = df_dist_eq %>%
  pivot_longer(cols = c(Cricetidae, Chromadorea),
    names_to = "group",
    values_to = "species_richness")

# Scatter plot represents correlation between species richness (Chromadorea and Cricetidae)
and distance from the equator

ggplot(df_dist_eq_long, aes(x = dist_eq_deg, y = species_richness)) +
  geom_point() +
  geom_smooth(method = "loess", se = TRUE, size = 1) +
  labs(title = "Correlation between species richness vs distance from equator",
    x = "Distance from Equator (degrees)",
    y = "Species Richness") +

```

```

facet_wrap(~ group, scales = "free_y") +

scale_x_continuous(breaks = seq(0, 90, by = 10)) +

theme(

  plot.title = element_text(hjust = 0.5, size = 14, face = "bold"),

  axis.text = element_text(size = 12),

  axis.title = element_text(size = 12),

  panel.grid.major = element_line(color = "gray80", size = 0.5),

  panel.grid.minor = element_line(color = "gray100", size = 0.25),

  strip.text = element_text(size = 12, face = "bold")

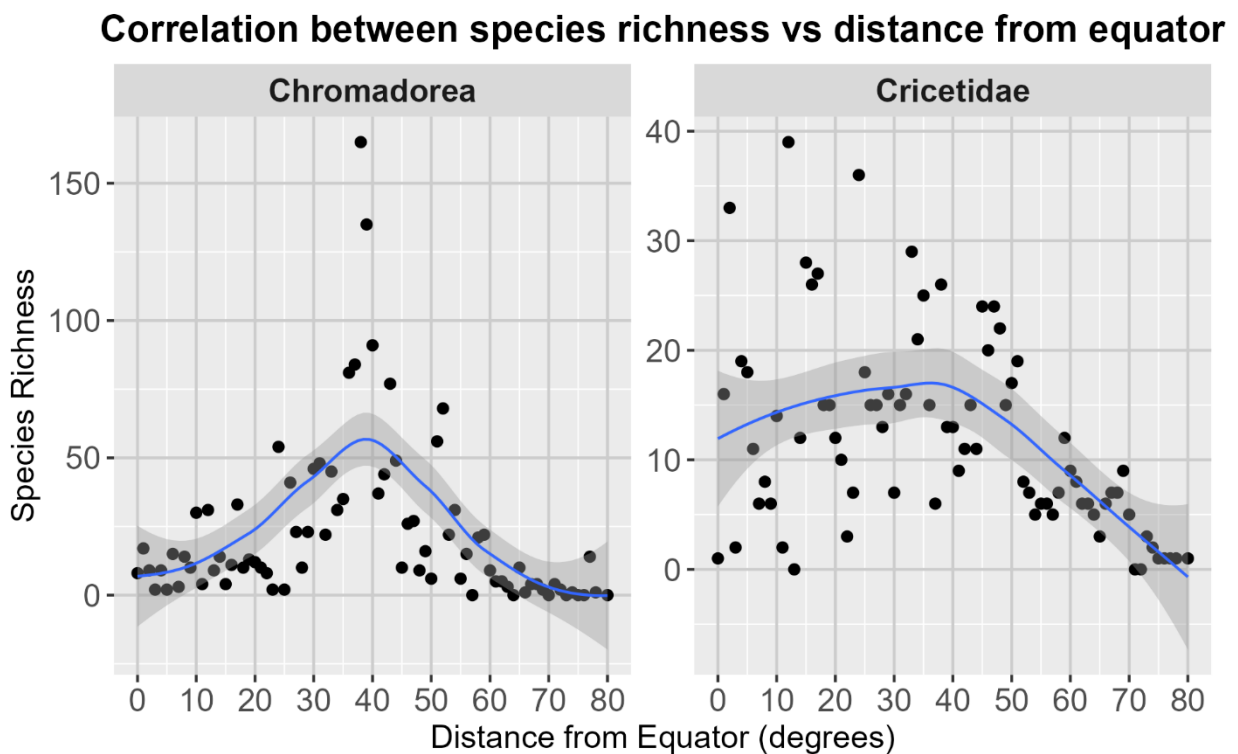
)

# Exporting the Plot with size suitable for A4 paper

ggsave("Correlation between species richness vs distance from equator.PNG", width = 6.4,

height = 4)

```



```

## 7 - Correlation between Cricetidae and Chromadorea species distribution ----

# Separate tibbles for Chromadorea and Cricetidae species count grouped based on latitude
with new source column for combining the data

df_chro_lat_sp = df_bold %>%
  filter(class_name == "Chromadorea") %>%
  mutate(lat = round(lat)) %>%
  group_by(lat) %>%
  summarise(species_count = n_distinct(species_name)) %>%
  mutate(source = "Chromadorea")

df_cric_lat_sp = df_bold %>%
  filter(class_name == "Mammalia") %>%
  mutate(lat = round(lat)) %>%
  group_by(lat) %>%
  summarise(species_count = n_distinct(species_name)) %>%
  mutate(source = "Cricetidae")

# Shapiro-Wilk test

shapiro_test_chro_lat_sp = shapiro.test(df_chro_lat_sp$species_count) %>%
  print()

# ==> Result Shapiro-Wilk test

# p-value = 2.929e-05 ==> data does not follows normal distribution.

# Shapiro-Wilk test

shapiro_test_cric_lat_sp = shapiro.test(df_cric_lat_sp$species_count) %>%
  print()

# ==> Result Shapiro-Wilk test

# p-value = 0.9222 ==> data does not follows normal distribution.

```



```
# Formatting a new data frame with the species count of both Chromadorea and Cricetidae to study their correlation
```

```
df_comb_lat_sp = rbind(df_cric_lat_sp, df_chro_lat_sp)
```

```
df_comb_lat_sp_wid = df_comb_lat_sp %>%
```

```
  pivot_wider(names_from = source, values_from = species_count)
```

```
# Spearman's Rank Correlation -- To determine if there is a correlation between Chromadorea and Cricetidae species richness
```

```
correlation_result = cor.test(df_comb_lat_sp_wid$Cricetidae,
```

```
df_comb_lat_sp_wid$Chromadorea, method = "spearman") %>%
```

```
  print()
```

```
# ==> Result Spearman's Rank Correlation
```

```
# Rho ( $\rho$ ) value = 0.304 ==> moderate positive correlation
```

```
# p-value = 0.005362 ==> correlation is statistically significant
```

```
# There is statistically significant moderate positive correlation between the geographical distributions of Cricetidae and Chromadorea
```

```
# Scatter Plot of Cricetidae vs. Chromadorea Abundance
```

```
ggplot(df_comb_lat_sp_wid, aes(x = Cricetidae, y = Chromadorea)) +
```

```
  geom_point(color = "darkgreen") +
```

```
  geom_smooth(method = "loess", se = TRUE, color = "navy", size = 1) +
```

```
  labs(title = "Correlation between Cricetidae and Chromadorea Abundance",
```

```
        x = "Cricetidae Abundance",
```

```
        y = "Chromadorea Abundance") +
```

```
  theme_minimal() +
```

```
  theme(
```

```
    plot.title = element_text(hjust = 0.5, size = 14, face = "bold"),
```

```

axis.text = element_text(size = 12, colour = "black"),

axis.title = element_text(size = 12),

panel.grid.major = element_line(color = "gray70", size = 0.5),

panel.grid.minor = element_line(color = "gray85", size = 0.25)

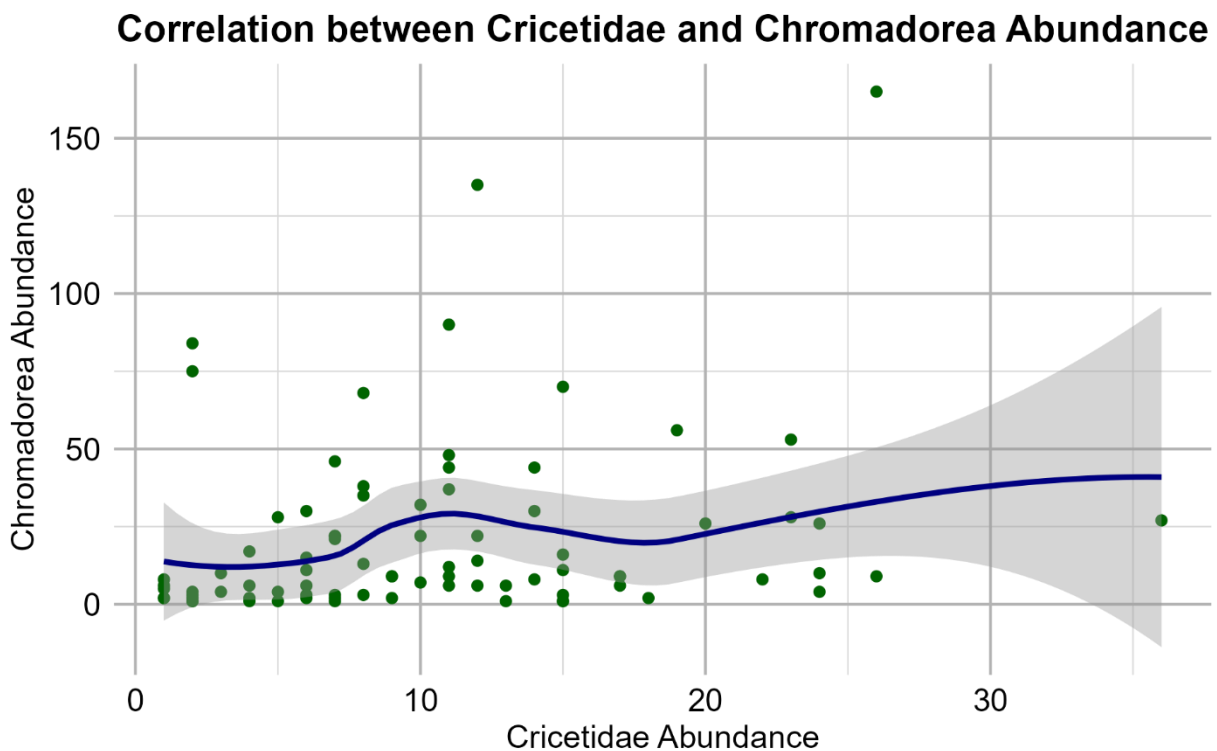
)

# Exporting the Plot with size suitable for A4 paper

ggsave("Correlation between Cricetidae and Chromadorea Abundance.PNG", width = 6.4,

height = 4)

```



4. Discussion and conclusion

The main goal of this project was to analyze the geographical distribution of the class Chromadorea and the factors influencing it. Initially, global species richness of Chromadorea was visualized on a map, revealing that species richness near the equator was not as high as in other areas. To assess whether the distribution of family Cricetidae or class Chromadorea

follows the classic LDG, Spearman's Rank correlation was applied, and the results were visualized using scatter plots.

A moderate negative correlation was observed between the distance from the equator (in degrees) and Cricetidae species richness. However, no significant correlation was found between Chromadorea species richness and the distance from the equator. When combining and plotting the data for both taxa, a statistically significant medium positive correlation was found between the geographical distributions of Cricetidae and Chromadorea. In conclusion, parasites, such as those within Chromadorea, are more likely to follow the geographical distribution patterns of their hosts, supporting the hypothesis that parasites species richness is more affected by their hosts' distributions rather than the classical LDG (4).

On major limitation of this study was the availability of latitude data, about 70 percent of the data initially obtained from BOLD system were missing latitude data. Also, it is recommended to study this correlation on a larger scale, focusing on samples from rodents infected with these parasites.

5. Acknowledgement

I would like to express my sincere gratitude to Dylan Harding, Moiz Syed and Thomas Tekle for their assistance throughout the project though offering advice for new packages or codes to explore, in addition to running my script to ensure that it was error-free.

6. References

1. Zhang Y, Song YG, Zhang CY, Wang TR, Su TH, Huang PH, et al. Latitudinal Diversity Gradient in the Changing World: Retrospectives and Perspectives. *Diversity* (Basel) [Internet]. 2022 Apr 25;14(5):334. Available from: <https://www.mdpi.com/1424-2818/14/5/334>
2. Kamiya T, O'Dwyer K, Nakagawa S, Poulin R. What determines species richness of parasitic organisms? A meta-analysis across animal, plant and fungal hosts. *Biological Reviews*. 2014 Feb;89(1):123–34.
3. Watve MG, Sukumar R. Parasite abundance and diversity in mammals: Correlates with host ecology [Internet]. Vol. 92. 1995. Available from: <https://www.pnas.org>

4. Preisser W. Latitudinal gradients of parasite richness: a review and new insights from helminths of cricetid rodents. *Ecography*. 2019 Jul 1;42(7):1315–30.
5. Data import with the tidyverse : :CHEATSHEET. Available from:
<https://rstudio.github.io/cheatsheets/data-import.pdf>
6. Data tidying with tidyr : : CHEATSHEET. Available from:
<https://rstudio.github.io/cheatsheets/tidyr.pdf>
7. Data visualization with ggplot2 : : CHEATSHEET. Available from:
<https://rstudio.github.io/cheatsheets/data-visualization.pdf>
8. <https://cran.r-project.org/web/packages/rworldmap/rworldmap.pdf>