

The Coevolution of Birds and Mammals

Madison Downing

March 18, 2017

Summary

1. Many studies have been done in light of new information on climate change and habitat destruction to better understand how we as humans are influencing extinction and species diversity. In order to collect more information on the subject it can be important to look at the bigger picture by exploring several groups over all of geological history.
2. Understanding the location of fossil records provides information for future studies and helps refine areas of interest.
3. Exploring the timeline of fossil records allows visualization of past diversity trends over time.
4. The PyRate program allows for exploration of predicted speciation and extinction rates using collected fossil records.

Introduction

When attempting to learn more about current species and how the current environment influences said species it can be important to look to the past first. For example, in recent studies of current extinction rates, these rates must be compared to past extinction rates, for the findings to hold any weight (Ceballos *et al.* 2015). Here we investigate the fossil records, from the paleobiology database, of Aves and Laurasiatheria (a representative for the group mammalia). We more specifically analyze the location of the fossils, the timeline of the species in each group, and their projected extinction and speciation rates.

We chose to study two groups rather than one because we are interested in learning how the groups evolved in relation to each other and how they might have influenced each other. When studying the genome size of each group, it was found that both groups had a similar amount of DNA despite variations in the addition and removal of genome sequences during evolution (Kapusta *et al.* 2017). It was also found that despite having similar genome sizes, there is less genetic variance in birds. This is explained by the avian constraint hypothesis which names constraints on proteins as the reason for this lack of genetic divergence (Stanley & Harrison 1999). With this information we expect that speciation and extinction rates and species timelines for the two groups will be relatively similar, but that mammals will progress quicker than birds.

Methods

The Data set

The data used in this study comes from the paleobiological database website. All data files are csv formatted files and consist of a header describing the data set, a line to describe the information in each column of data, and several thousand fossil occurrence entries. Since the main focus of this paper is to compare bird and mammal fossil records, data for the groups Aves and Mammalia were downloaded. However, the number of occurrences recorded for each group was drastically different. The Aves data had only around 6,000 entries, where as the Mammalia data had over 100,000. In order to make the data sets more even and therefore more easily comparable, a sample group was chosen to represent Mammalia. Laurasiatheria was decided upon because it included a large proportion of all extant mammals and it's timeline matched that of the Aves group. The group still had more entries than Aves (around 16,000), but it was a much more manageable number for comparisons.

Data Preparation

All working python code described in this section can be further investigated in the file called 'python_code.ipynb'

Removing the header

To remove the header we can use the following function named 'removing_header.sh', which takes in three arguments: the number of the line at which to start the new file, the input file, and the output file. The line number to start at is included as a variable because certain information may need to be included or excluded depending on the function that is going to be run. For example, any files to be run through python should not have any header information, but the PyRate program requires files to have the header line that describes each column of information.

```
#!/bin/bash/

tail -n +$num_lines $input_file > $output_file
```

Extracting Data

For comparing locations:

The following function takes in two arguments, a file to read from and a file to write to. The purpose of this code is to take a file and extract the genus name, latitude, and longitude of each record and write this to each line of a new csv file.

```
def locations(openfile, writefile):
    newfile = openfile
    output = open(writefile, "w")
    with open(newfile, 'r') as ff:
        readfile = ff.readlines() # open and read file
    for line in readfile: #go through each line
        taxon = line.split(",")[9]
        if taxon == "species": #only continue if the line contains a species
            species = line.split(",")[8].replace('"','') #extract species names
            genus = species.split(" ")[0]
            lat = line.split(",")[18].replace('"','') #extract the latitude
            lng = line.split(",")[17].replace('"','') #extract the latitude
            outline = "{},{},{}\n".format(genus,lat,lng) #write the format
            output.write(outline) #write to the file
    output.close()
```

new output files:

```
##          genus latitude  longitude
## 1 Ornimegalonyx 22.58333 -83.666664
## 2      Anhinga 29.70000 -82.599998
## 3    Promilio 29.70000 -82.599998
## 4       Buteo 40.90000 -104.866669
## 5   Hadrogyps 35.30000 -118.500000
## 6    Primapus 50.78333  -0.683333

##          genus latitude longitude
## 1   Cynodictis 50.67720  -1.0856
## 2   Saturninia 50.66667  -1.5500
## 3   Stehlinia 50.66667  -1.5500
## 4  Enaliarctos 35.49278 -118.8483
## 5 Pinnarctidion 35.49278 -118.8483
## 6   Indarctos 25.01667  102.0667
```

For exploring timelines:

The following functions are used to create a file containing the genus name, the species name, and the minimum and maximum age for that species. The first function takes the file to read from as an argument. It creates a dictionary with the species name as the key and a list of the average estimated ages for the fossils of that species. Each fossil record has some uncertainty of its exact age, so its age is listed at the maximum age and the minimum age. These are averaged to find the mean age for that record. The second function takes a file to write to and a dictionary name as arguments. This function takes the genus name, species name, and highest and lowest age for that species and writes it to a new csv file. In order to better visualize and compare the two timelines, some entries were removed from the Aves data file so that we are only looking at the age ranges that overlap. This was done in unix.

```

def ages_dict(filename):
    newfile = filename
    with open(newfile, 'r') as ff:
        canid_recs = ff.readlines() #open file
    from collections import defaultdict #download default dict

    species_ages = defaultdict(list) #make an empty dictionary

    for line in canid_recs:
        taxon = line.split(",")[6]
        if taxon == '"species"': #only continue if the line contains a species
            species = line.split(",")[5] #extract species name
            minage = line.split('"')[10] #extract upper age estimate
            maxage = line.split('"')[11] #extract lower age estimate
            mean_age = (float(minage) + float(maxage)) / 2 #find the average of the age
            species_ages[species].append(mean_age) #if the species is not already in the
    return(species_ages)

def range_file(filename, dictname):
    output = open(filename, "w")

    for key in dictname.keys():
        ages = dictname[key]
        minage = min(ages) #extract lowest age
        maxage = max(ages) #extract highest age
        genus = key.replace('"','').split(" ")[0] #extract genus
        species = key.replace('"','') #extract species
        outline = "{},{},{},{}\n".format(genus,species,minage,maxage) #write format
        output.write(outline) #write to new file
    output.close()

#the following is unix code
sort --field-separator=',' --key=3 -g aves_ranges.csv > sorted_aves_ranges.csv #sort by
head -n 1027 sorted_aves_ranges.csv > sorted_aves_ranges1.csv #remove entries past 70 Ma

new output files:

##          genus          species minage maxage
## 1    Accipiter    Accipiter fasciatus 0.00585 0.00585
## 2    Aegotheles  Aegotheles novaezealandiae 0.00585 0.06300
## 3         Alca          Alca torda 0.00585 9.43300
## 4         Alle          Alle alle 0.00585 1.29985
## 5         Anas          Anas platyrhynchos 0.00585 4.46650
## 6 Anomalopteryx  Anomalopteryx didiformis 0.00585 1.29400
##          genus          species minage maxage

```

```
## 1      Acinonyx      Acinonyx jubatus 0.00585 3.96050
## 2      Alopex       Alopex lagopus 0.00585 1.05000
## 3      Amblonyx     Amblonyx cinerea 0.00585 0.00585
## 4      Antrozous    Antrozous pallidus 0.00585 7.60000
## 5      Aonyx        Aonyx capensis 0.00585 3.96050
## 6 Arctocephalus Arctocephalus forsteri 0.00585 0.06885
```

For exploring Speciation and Extinction Rates:

First a program will be run in r to format our data for PyRate, but this program requires a list of all extant species for the group. To find the names of the extant species a python function, that identifies all unique occurrences of species with a minimum age of 0 ma, is used.

```
def extant_species_list(filename):
    import numpy
    with open(filename, 'r') as ff:
        readfile = ff.readlines() #read the file
    extant_species = [] #create our primary list
    cleaned_extant_list = [] #create a cleaned list
    for line in readfile:
        taxon = line.split(",")[6]
        min_ma = line.split(",")[11]
        if taxon == '"species"' and min_ma == '"0"': #only continue if the occurrence is
            species_name = line.split(",")[5].replace('"', '') #identify the species name
            extant_species.append(species_name) #add the species name to the primary list
    extant_list = numpy.unique(extant_species) #once the list is complete remove any repeats
    for item in extant_list:
        cleaned_extant_list.append(item) #turn this list of unique names into a new list
    print(cleaned_extant_list) #print the list
```

In an r script called process_laura_data.r, another r script is called to format the original data file using the newly created list of extant species. The data set for this portion should include a header labeling each column.

```
#tell the code where to find the formatting r script
source("~/PyRate/pyrate_utilities.r")
```

```
#give the list of extant species
extant_mammals = c('Acinonyx jubatus', 'Alopex lagopus', 'Amblonyx cinerea', 'Antrozous

# use the extract.ages.pbdb() function in pyrate_utilities to reformat the data set
extract.ages.pbdb(file= "pyrate_laura.csv", extant_species=extant_mammals)
```

The same code is repeated in process_aves_data.r using the list of extant aves species.

```
#tell the code where to find the formatting r script
source("~/PyRate/pyrate_utilities.r")
```

```
#give the list of extant species
extant_birds = c('Accipiter fasciatus', 'Accipiter gentilis', 'Accipiter striatus', 'Aeg
# use the extract.ages.pbdb() function in pyrate_utilities to reformat the dataset
extract.ages.pbdb(file= "pyrate_aves.csv",extant_species=extant_birds)
```

In order to run pyrate, the 2.7 version of Python must be used. To switch the working terminal to this version use the following shell command.

```
source activate py27
```

To run PyRate and create the files that will eventually be used to visualize speciation and extinction rates, use the following shell code.

```
python ~/PyRate/PyRate.py inputfile_PyRate.py -n 1000000
```

The input file name will change depending on the name of the data file used in the r script that formatted the data for pyrate. The number of repetitions can also be changed, here 1,000,000 repetitions are used.

Results

Comparing Locations

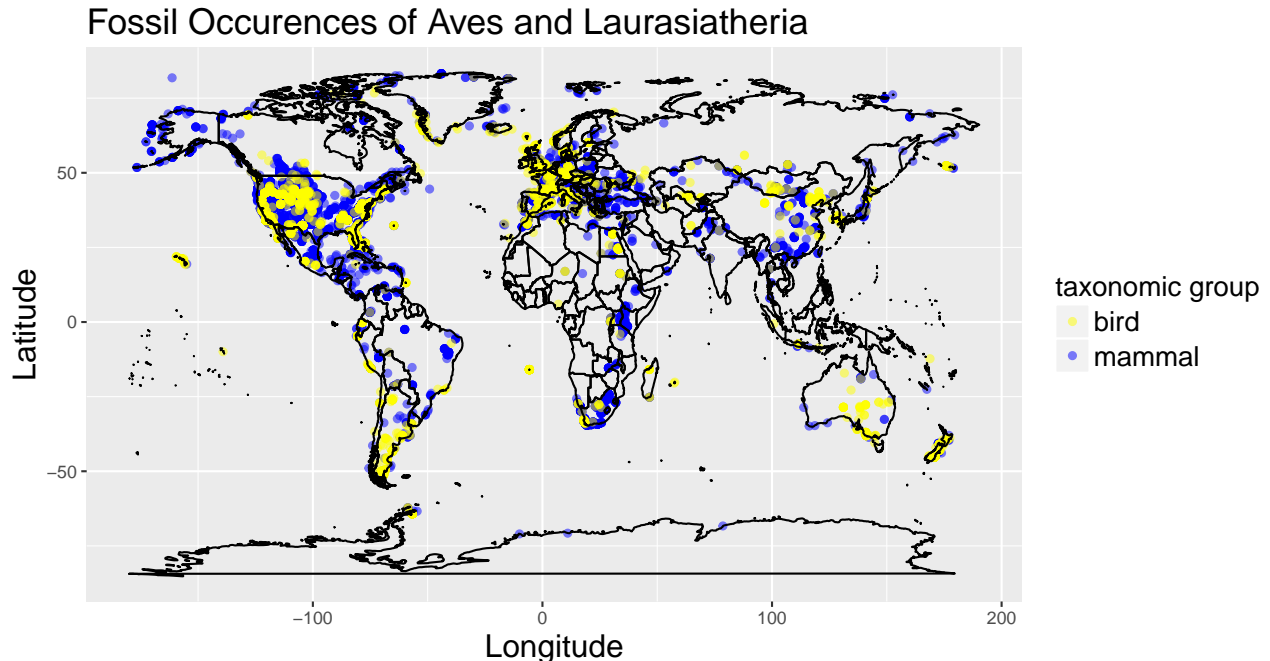


Figure 1: Fossil Locations

This map does little to help draw any conclusions about the way that birds and mammals actually interact, but it does help refine areas of interest. We see that the most overlap in fossil records occurs in the United States, Eastern Europe, and East Asia.

Exploring Fossil Timelines

In figure 2 we are able to visualize the amount of species and which species were present at a given time period. The color groups represent a common genus. From this graph we see that many new species appeared around 20 Ma ago and that the extant species present appeared around 5 Ma ago.

Figure 3 follows the same structure as the preceeding graph, but because the Laurasiatheria group contains more species it looks slightly different. However, it seems to follow a similar trend of regular loss and replenishment of species. This group appeared around 70 Ma and the extant species appeared around 7 Ma ago.

```
##  
## Attaching package: 'dplyr'  
  
## The following objects are masked from 'package:stats':
```

Aves Fossil Occurrences

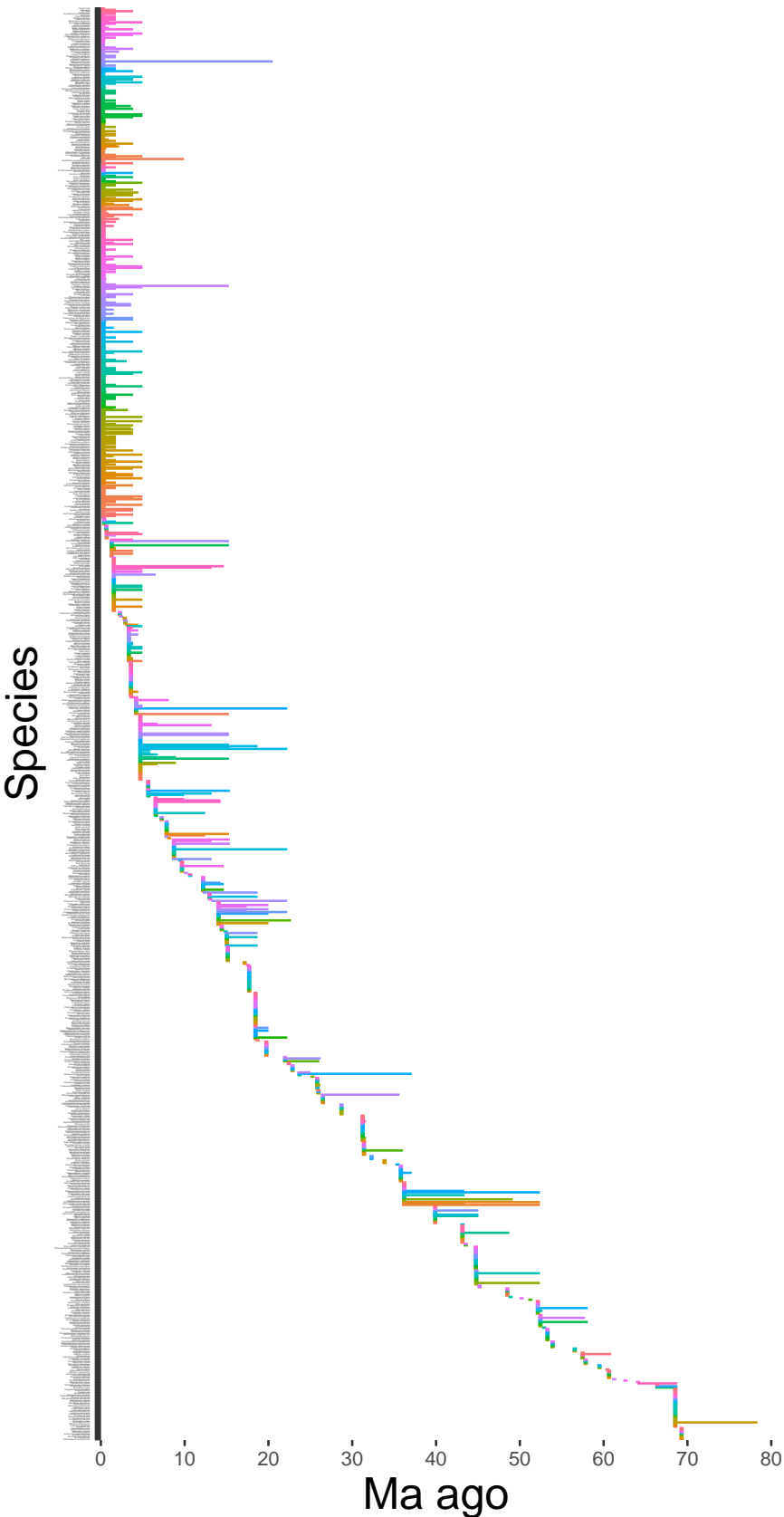


Figure 2: Aves Timeline

Laurasiatheria Fossil Occurrences

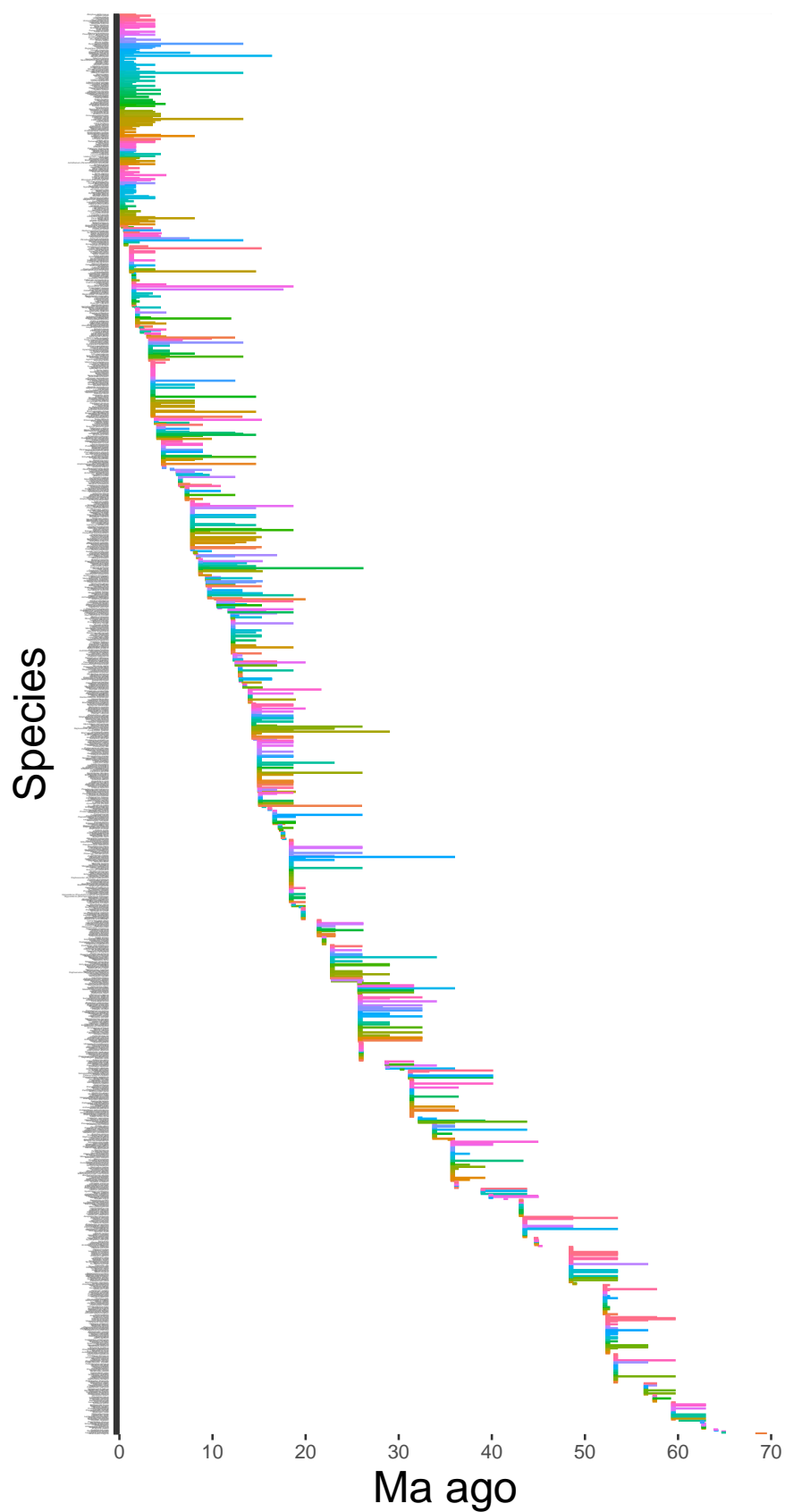


Figure 3: Laurasiatheria Timeline

```
##
##      filter, lag
## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union
```

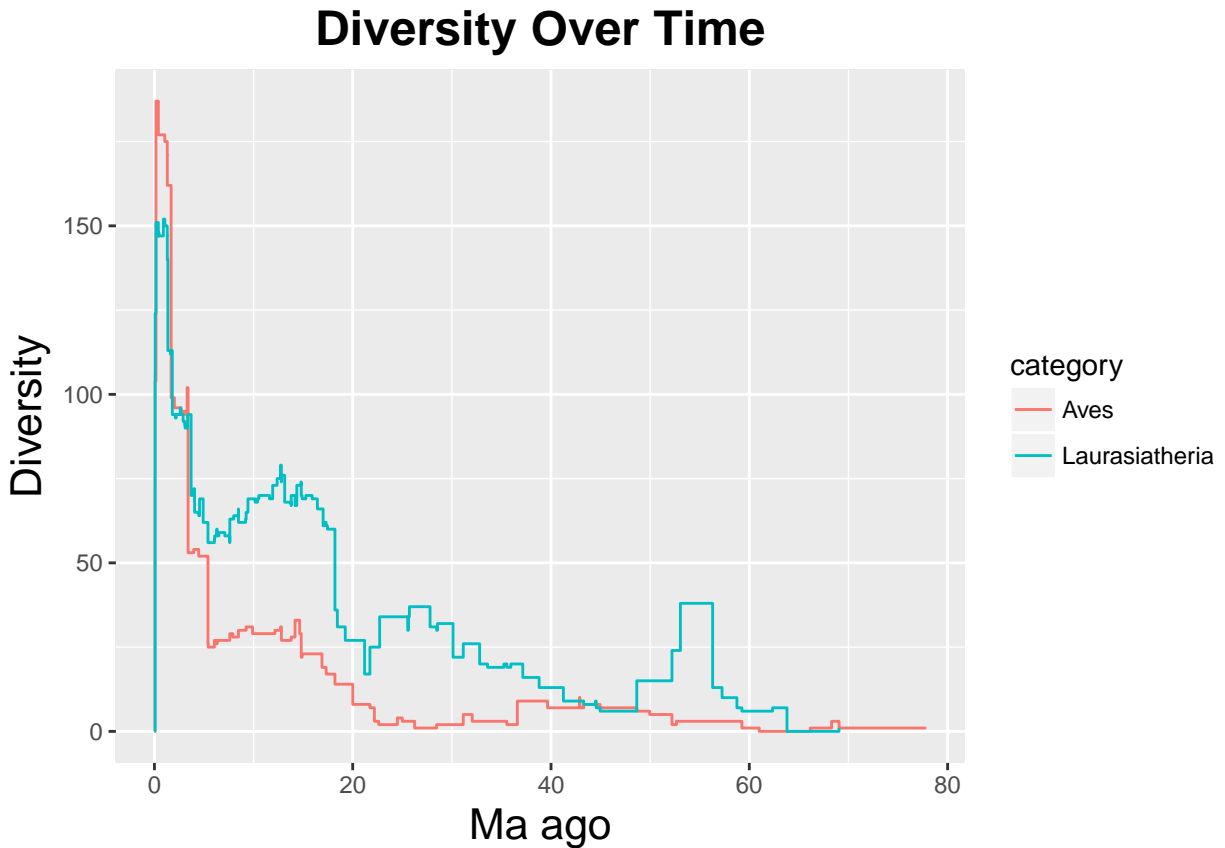


Figure 4: diversity over time

This plot shows numerically, how many species are present at a given time and how that amount changes over time. Although the two graphs do not match up completely, trends common to both groups can still be seen. Both groups experience increases in species around 30 Ma and 17 Ma ago. They also experience decreases in species around 22 Ma and 5 Ma ago, before increasing greatly closer to present day.

Exploring Extinction and Speciation Rates

In order to visualize the PyRate projections, first navigate into the `pyrate_mcmc_logs` directory, then run the following unix code.

```
python ~/PyRate/PyRate.py -plot pyrate_aves_1_marginal_rates.log
```

The file can be changed depending on the group of interest, but the file will always be a `marginal_rates.log` file.

An important thing to note is that the projections for the Aves data covers a larger expanse of time than the Laurasiatheria data. When comparing the two sets only the common time periods should be considered. In addition, all graphs display a sharp change towards present day, but this can be accounted for by the lack of fossil records to provide information for estimations. The first figure estimates the change in speciation rate over time. Both data sets follow a similar pattern of a steady speciation rate around 0.2. The second figure estimates the change in extinction rate over time. Again, both data sets remain relatively constant around 0.1, but the Laurasiatheria data does gradually increase over time. The third figure projects the net diversification rate over time. Here, the aves data begins below 0 and gradually increases until it experiences a sharper increase around 10 Ma ago. The Laurasiatheria data remains above zero, but gradually decreases overall. The fourth figure displays the change in longevity over time. The Aves data gradually increases in longevity, but the Laurasiatheria data drastically decreases.

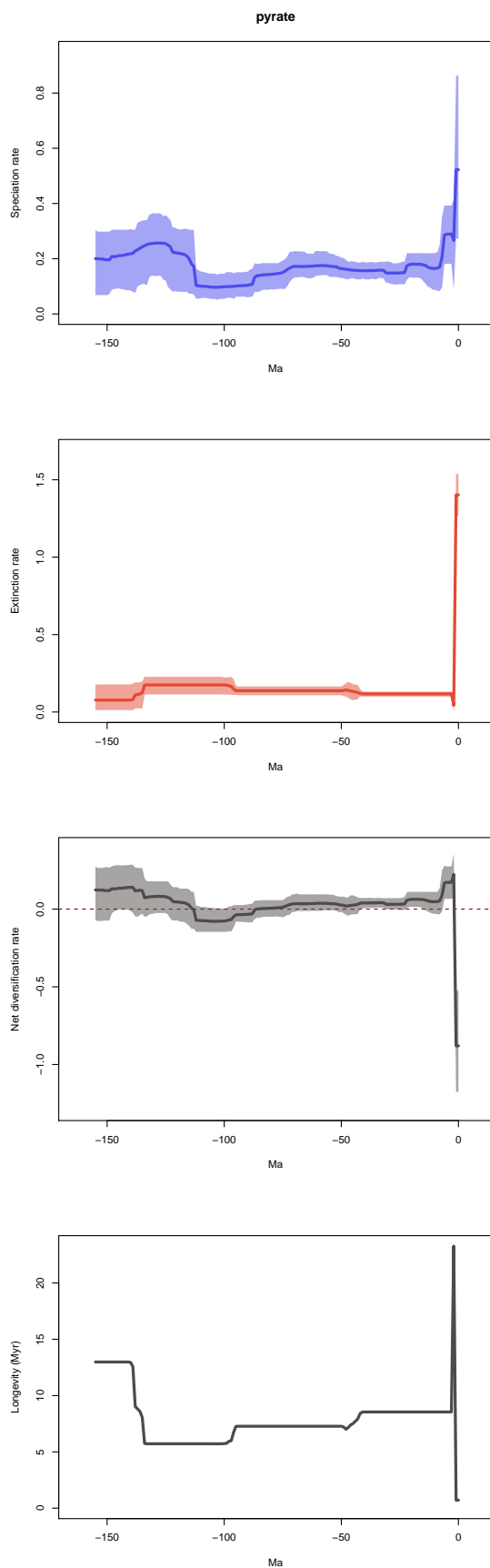


Figure 5: PyRate Aves Data Output

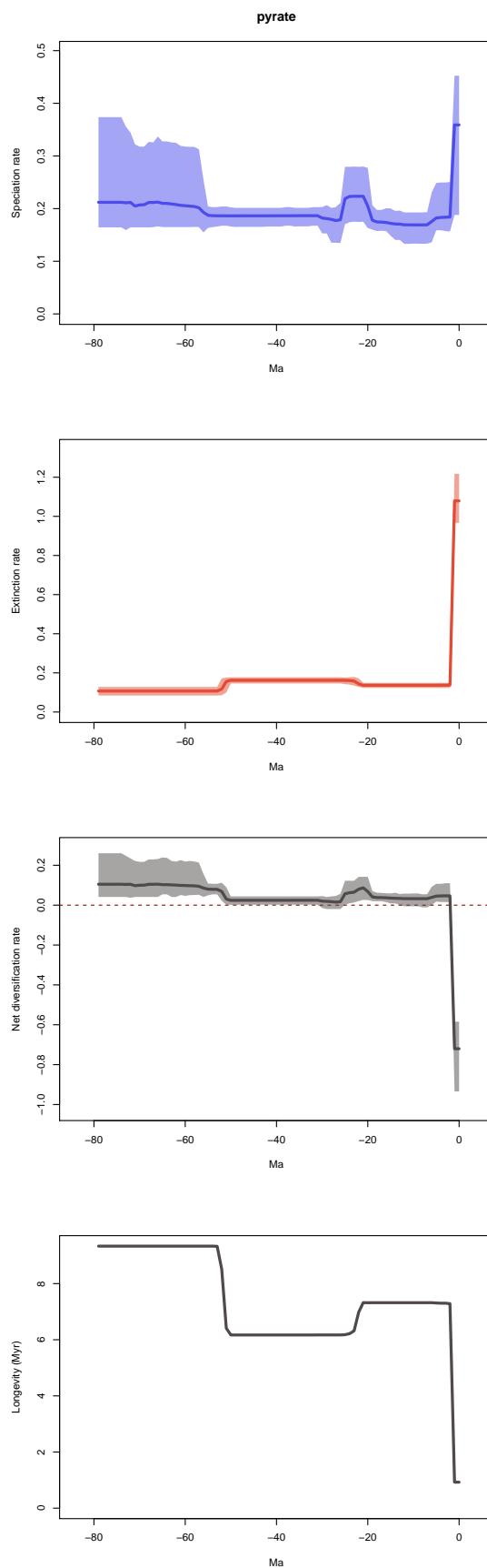


Figure 6: PyRate Laurasiatheria Data Output

Conclusion

Through the analysis of the paleobiology fossil records for Aves and Laurasiatheria we see many similarities in the way that the groups have grown and changed over time. This is especially apparent in the predictions made by the PyRate program where larger trends can be visualized. Figure 4 indicates that although the two groups did not go through identical changes, they did go through the same general changes. This could indicate environmental influences that would be able to change both groups. Similarly, the PyRate projections were not identical, but did have similarities in their general trends. These observations are useful in looking at trends as a whole because birds and mammals make up a large portion of the vertebrate population. With information about genetic trends throughout paleological history, we can use the information and observations gathered from this study to better understand changes in phenotypes through time. This information can also be useful when looking at important extinction and climate events throughout earth's history. Lastly, the map in figure 1 can narrow down the areas to observe when attempting to compare the history of birds and mammals. Although this study does not allow for any conclusions about the causes of these trends to be made, it does provide information necessary for new studies

Git Hub Link

<https://github.com/Madison2296/finalproject>

References

- 1.Ceballos, G., Ehrlich, P.R., Barnosky, A.D., Garcia, A., Pringle, R.M. & Palmer, T.M. (2015). Accelerated modern human induced species losses: Entering the sixth mass extinction. *Science Advances*, 1.
- 2.Kapusta, A., Suh, A. & Feschotte, C. (2017). Dynamics of genome size evolution in birds and mammals. *Proceedings of the National Academy of Sciences*, 114, E1460–E1469.
- 3.Stanley, S.E. & Harrison, R.G. (1999). Cytochrome b evolution in birds and mammals: An evaluation of the avian constraint hypothesis. *Molecular Biology and Evolution*, 16, 1575.