Effects of Extinction and Physiological Adaptations on Squamata Species Richness

Yee Chau

March 24, 2017

1   Summary

1. Paleobiological data also offers location data with longitudinal and latitudinal values for each location where fossils of squamata were collected. These locations can give more information on how concentration of collection locations can help us infer how environment and climate may have played roles in species richness.

2. Fossil occurrence data can shed light on changing species richness throughout time. Due to this interest in the effects of how events in time, where time is represented by a given mya value, affected species richness before and after a value, we attempt to show how, and if, certain events made significant changes, whether to increase or decrease richness.

3. Finally, we investigate the beginning and significance of the development of caudal autotomy (CA) in the order squamata and show how species richness was affected by the development of the behavior. Pyrate plots and histograms can possibly be used to evaluate if CA increased survival of certain species, led to diversification/speciation, and resulting in overall higher richness in years following the behavioral development.

2   Introduction

Squamata are the largest recent order of reptiles including all lizards and snakes, as well as most modern reptiles [1] Modern squamates originated

around the mid jurassic period and are a monophyletic sister group to the tuatara, where together they are a sister group to crocodiles and birds [3]. In addition, with over 10,000 species under the order, it holds the place of the second largest order of extant vertebrates. However, in the past, squamates suffered a heavy extinction near the Cretaceous-Tertiary boundary extinction and a smaller loss around the Eocene-Oligocene extinction.

3   Description

The functions illustrated in the following section will allow further exploration into the how different aspects of squamates have allowed them to exist for so long, especially since they are the second largest order of extant vertebrates. Through these analyses, perhaps it will lead to more explanations on why things have ended up the way they are, how and why diversification and speciation rates have been affected.

4   Data Preparation

Before analyzing the species fossil occurrence data based on location concentrations, it was necessary to first clean the data downloaded from the Paleobiology data base through the terminal. The code below was inputted to the terminal to ensure that the csv did not need further cleaning so that later usage would be clear and without problem.

```
tail -n + 2 lizard_pbdb.csv | cut -d "," -f 0,3,4,8,11,12 |
grep -w "species" | sort -n
```

The output from the shell script outputs the data shown below, which is the data set from which the Lizard_coords dataset that is used for function 4.1.

4.1 Coordinate Data					
Collection No.	Longitude	Latitude	No. of Occ	max_ma	min_ma
3133	-74.683334	39.971943	2	59.2	56
3534	-92.147224	32.234165	1	41.3	38
4894	135.331390	34.350555	1	70.6	66
5241	-77.916664	18.333332	1	47.8	41.3
5607	-107.92583	38.901669	1	83.5	70.6
6802	-90.286110	32.474167	1	38	33.9
7515	-88.606392	35.233612	3	83.5	70.6

The python functions below produce the file that contains only the necessary data for the figure 4.2. The second function requires the first function to be able to get the proper data, and carries out the purpose of writing a new file that will only contain the four variables that will be used to produce the output. The functions are shown below:

Listing 1: 4.2 Data Preparation

```python
from collections import defaultdict
species_ranges = defaultdict(list)
for line in lizard_reader:
    species = line.split('","')[0] # specifies which column is
        needed for species and splits
    min_age = line.split('","')[3] # specifies which column is
        needed for minage and splits
    max_age = line.split('","')[2] # specifies which column is
        needed for maxage and splits
    mean_age = (float(min_age) + float(max_age))/2 # gets the mean
    print (species, mean_age)
    species_ranges[species].append(mean_age)

output = open("/Users/yeechau/Downloads/eeb-174-final-project/
Lizard_Data/Results_Directory/squamata_ranges.csv", "w")

for key in species_ranges.keys():
    ages = species_ranges[key]
    min_age = min(ages)
    max_age = max(ages)
```

3

```
genus = key.split(" ")[0].strip('"')
species = key.strip('"')
outline = "{},{},{},{}\n".format(genus, species, min_age,
    max_age)
output.write(outline)
print(outline)
```

The function then returns the data set shown, listing genus, species, minimum age, and maximum age:

4.2 Range Data			
Genus	Species	min_age	max_age
Aciprion	Aciprion formosum	33.599999	33.5999999
Aciprion	Aciprion majus	33.5999999	33.59999999
Acrochordus	Acrochordus dehmi	8.47050000	13.7890000
Acrodontopsis	Acrodontopsis robustus	42.90000006	42.90000000
Acteosaurus	Acteosaurus tommasinii	96.55	96.55
Adamisaurus	Adamisaurusmagnidentatus	74.8	77.85

5 Methods

5.1 Estimation of Species Richness as a Result of a Given Time

This first function is an aimed at estimating the general significance of effects of disaster events, such as mass extinctions or geographical disasters, on species richness by showing the number of species present before and after the event. From this analyzation we hope to gain a greater understanding of the areas where squamates were concentrated most greatly at different eras, and if we can infer the types of environments they occupied. From those assumptions, perhaps it will also be possible to figure out which species prospered the most and what led to their success. It utilizes the longitudinal and latitudinal values from the file produced as a result of the 4.1 Data Preparation code in the previous section.

The R function takes an input that represents a point in time, such as the Cretaceous-Tertiary boundary which occurred approximately 66 million years ago, and uses the value using mya. The input is represented by the "years" in the function. So in the case of the K-T boundary, the function would take 66 as the argument. Three libraries are utilized, rworldmap, ggplot2, and

dplyr to perform the necessary objective of laying the points over a world map and placing them based on their geographical locations after the data set has been split into two new sets, where pre_era set represents all species fossil occurrences before the input and post_era represents all species fossil occurrences after the input. The function is illustrated below, followed by the image that is produced from the function:

Listing 2: 4.1 R example

```r
lizards <-
    read.csv("/Users/yeechau/Downloads/eeb-174-final-project/
    Lizard_Data/Results_Directory/lizards_coords.csv", header =
    FALSE, stringsAsFactors = FALSE)
lizard_species <-
    read.csv("/Users/yeechau/Downloads/eeb-174-final-project/
    Lizard_Data/Results_Directory/lizard_pbdb.csv", header = FALSE,
    stringsAsFactors = FALSE)
colnames(lizards) <-
    c("collection_no","record_type","formation","lng","lat",
    "collection_name","collection_subset","collection_aka","n_occs",
    "early_interval","late_interval","max_ma","min_ma","reference_no")
head(lizards)

library(rworldmap)
library(ggplot2)
library(dplyr)
lizard_life <- function(years) {
    pre_era <- lizards%>%filter(min_ma < years)
    post_era <- lizards%>%filter(min_ma > years)
    print(pre_era) # prints the data in a table format, with
        location name
    print(post_era) # everything after input value will be printed
        in a separate dataset
    newmap <- getMap(resolution = "low")
    plot(newmap)
    points(pre_era$lng, pre_era$lat, pch = 20, col = "red", cex = .9)
    points(post_era$lng, post_era$lat, pch = 18, col = "black", cex
        = .9)}
```
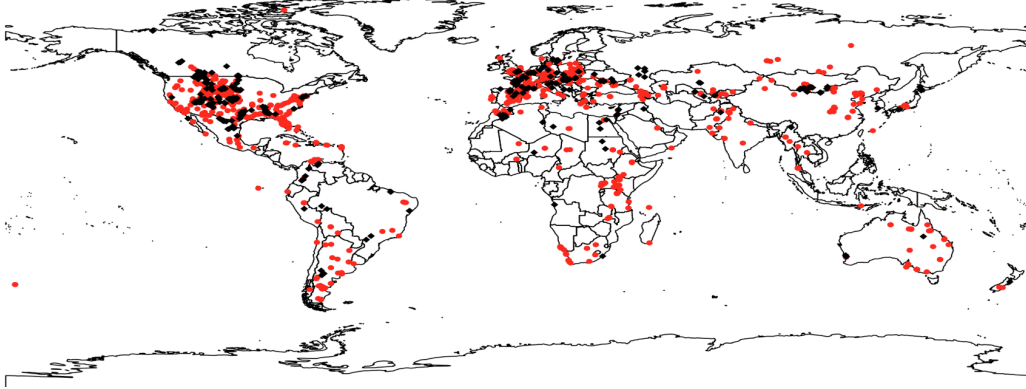
Figure 1: World map showing the points for every location where fossil collection of species under the order squamata occurred. The map is a render of the code shown above, with an overlay of the the pre-era points (red) over the post-era points (black). Pre-era shows points before the time period input in millions of years (e.g. 66 mya), while post-era shows the points after the time period input.

5.2 Utilization of PBDB Data to Visualize Squamate Fossil Occurrence Through Time

Expanding upon the idea of inferring from location data based upon fossil occurrences, as shown the previous section, the aim of this function is to organize all occurrences of each species by age under the order Squamata. By showing all the species from approximately 100 million years ago to the present (or when the most recent samples from the order squamata were collected), this allows us to have a different and more visual view of the data and see which species are extant or became extinct recently. However, some species are excluded from the plot, as they did not have enough data to be placed on the plot.

Libraries ggplot2 and forcats were used to produce the image and reorder the species by their minimum age, as forcats will do for categorical variables. After declaring the columns and having the command head() show a portion of the data, we move on to clarifying each parameter that we want the plot to adhere to and customize the title and labels. ggsave allows us to have

a pdf version of the output, which has much more detail than the output shown in the RStudio window.

Listing 3: 4.2 R example

```r
library(ggplot2)
library(forcats)
setwd("/Users/yeechau/Downloads/eeb-174-final-project/
Lizard_Data/Results_Directory")
LIZARDSS <-
    read.csv("/Users/yeechau/Downloads/eeb-174-final-project/
Lizard_Data/Results_Directory/squamata_ranges.csv", header = F,
    as.is = T)
names(LIZARDSS) <- c ("genus", "species", "minage", "maxage")
head(LIZARDSS)

LIZARDSS_occ <- ggplot(LIZARDSS, aes( x = fct_reorder(species,
    minage, .desc = T), maxage, colour = genus))
LIZARDSS_occ <- LIZARDSS_occ + geom_linerange(aes(ymin = minage,
    ymax = maxage + 0.5)) + theme(legend.position="none") +
    coord_flip() + theme(axis.text.y = element_text(size=.5)) +
    scale_y_continuous(limits=c(0, 100), expand = c(0, 0),
    breaks=c(0, 25, 50, 75, 100)) + labs(title = "Squamata Fossil
    Occurrences", x = "Species", y = "Ma ago") + theme(plot.title =
    element_text(hjust = 0.5, size=22, face = "bold"), axis.title
    =element_text(size=20))
LIZARDSS_occ

ggsave("faceted-ranges.pdf", scale = 6)
library(tidyr)
library(dplyr)
diversity <- LIZARDSS %>% gather(key = type, value = age, minage,
    maxage) %>% mutate(count = ifelse(type == "maxage", 1, -1)) %>%
    group_by(age) %>% summarise(count = sum(count)) %>%
    arrange(-age, -count) %>% mutate(diversity = cumsum(count))
ggplot(diversity, aes(x = age, y = diversity)) + geom_step()
```
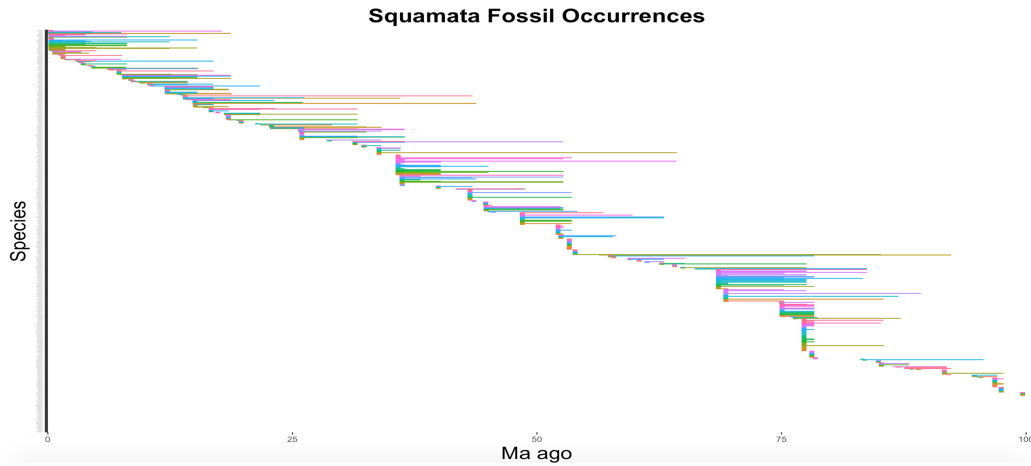
Squamata Fossil Occurrences

Figure 2: Squamata fossil occurrences organized by minimum age, where the youngest species are located at the top, and the oldest at the bottom. From the occurrences you can see which species are still extant based on recent fossil collections. A much clearer version of the plot is accessible under my github linked below under the filename faceted-ranges.pdf.

5.3 Trait Correlation to Species Survival

The purpose was to investigate the beginning and significance of the development of caudal autotomy in the order squamata and show how species richness was affected by the development of the behavior. With this idea in mind, I was unable to create a phylogeny in time, but I made a function that creates a dictionary for the duration of survival for each species in the fossil record, but there needs to be further improvements on showing which are still currently living and that their "duration" is still continuing. From the data of the duration, I was allowed the freedom to choose specific families where the CA behavior had appeared and families where it did not appear, and after extracting the species from those families, I wanted to see if it was possible to see if species that developed the CA behavior were more likely to have longer durations. Although I know that duration can be affected by many other factors and one behavior may not have an effect, I wanted to see if there was even a slight evidence that those with CA existed for a longer period of time. The function is listed below, with each step commented:

8

Listing 4: 4.3 Python example

```python
#making a dictionary for species and ages, meaning duration of
    their existence
def life_duration(filename):
    liz_dict = {} #creates an empty dictionary
    liz_recs = open(filename, "r", encoding = "ISO-8859-15")
    for line in liz_recs:
        line = line.replace('\"', '')
        record_elements = line.split(",")
        species = record_elements[0] # species in column 0
        max_age = record_elements[2] # max age to column 2
        min_age = record_elements[3] # min age to column 3
        age = float(max_age) - float(min_age) # calculates the age
            of the fossil
        liz_dict[species] = age # assigns age as the values in the
            dictionary
    return liz_dict

mydict =
    life_duration("/Users/yeechau/Downloads/eeb-174-final-project/Lizard_Data/Results_Di
mydict
```

6 Results

From the functions that I have made to clean and organize my data into more visual plots or datasets that will answer my questions about the order squamata and the species within it, I have reached certain conclusions because of them. When I was making the first function, I was focused on the effects of extinction events, and in particular, the K-T extinction event. I had made the assumption from a few papers that I had read that extinction events increased speciation and subsequent diversification. However, from the datasets that were produced from my function 4.1, it seems that either extinction events do not drive speciation, or squamates in particular are not very resilient when it comes to extinction.

However, in my squamata fossil occurrences I was able to see the extant species and how it correlated to the ones living today

But at the same time, because some of the results I have obtained are

9

very different from my previous assumptions and extrapolations, I now have more questions. In order to narrow down the analysis for the purpose of this paper, two families were chosen under the order squamata based upon their development or absence of the caudal autotomy behavior. Given their development or absence of the behavior, there was further analysis done to help infer whether or not CA conferred a higher survival rate, and therefore a higher speciation and diversification rate to lead to higher overall species richness.

7   Discussion

Many of the functions gave a more generalized analyzation of the questions outlined in this paper, as the fossil record is not the most accurate, but with further exploration, more data sets, and improvement of my code, many of my conclusions and questions can be answered with more accuracy and possibly take into account errors that were not accounted for previously. For example, the world map and fossil occurrence data that I plotted to estimate how environment may have played a role in species survival and richness, was unable to account for the fact that some environments do not fossilize organisms as well as others do, which may have caused a skew in the data, where more species occurred in areas that were better at fossilization rather than more species fossil occurrences where more species actually survived. Therefore, I would like to further improve my insight on how to make my functions more accurate and, though it may not be able to entirely account for all of the error, make a more logical argument.

Unfortunately, my PyRate run encountered an error, so I was unable to use the million iteration marginal log, but I want to try with more iterations next and see if the data will be more concise and support my initial ideas. PyRate output can be seen in the appendix. The diversification and speciation rates that result from PyRate will be more useful with more iterations. Most importantly, the analyzation with caudal autotomy was of large interest to me, so I would like to explore more on creating phylogenic trees and observing how other traits that have developed in squamates have either aided or hindered their survival and richness. Also, as I only focused on the single behavior of caudal autotomy, I would like to see how other behaviors or traits may or may not confer higher survival that may contribute to overall richness.

8   Data Accessibility

The functions mentioned in this paper, along with further scripts, outputs, and data sets used, can be accessed at: https://github.com/ciaoyee/eeb-174-final-project.

Listings

References

[1] "Diversity of Vertebrates." *Diversity of Vertebrates.* Sam Houston State University, n.d. Web. 13 Mar. 2017.

[2] Dominguez-Lopez, M.E., Ortega-leon, A.M. & Zamora-abrego, G.J. Rev. Chil. de Hist. Nat. *Tail autotomy effects on the escape behavior of the lizard Gonatodes albogularis (Squamata: Sphaerodactylidae), from Crdoba, Colombia.* (2015) 88: 1. doi:10.1186/s40693-014-0010-6

[3] Jones ME, Anderson CL, Hipsley CA, Mller J, Evans SE, Schoch RR. *Integration of molecules and new fossils supports a Triassic origin for Lepidosauria (lizards, snakes, and tuatara).* BMC Evolutionary Biology. 2013;13:208. doi:10.1186/1471-2148-13-208.

[4] Pyron, R., Frank T. Burbrink, and John J. Wiens. *"A Phylogeny and Revised Classification of Squamata, including 4161 Species of Lizards and Snakes."* BMC Evolutionary Biology 13.1 (2013): 93.

[5] Zani, Peter A. *"Patterns of Caudal-autotomy Evolution in Lizards."* Journal of Zoology 240.2 (1996): 201-20.

9   Appendix

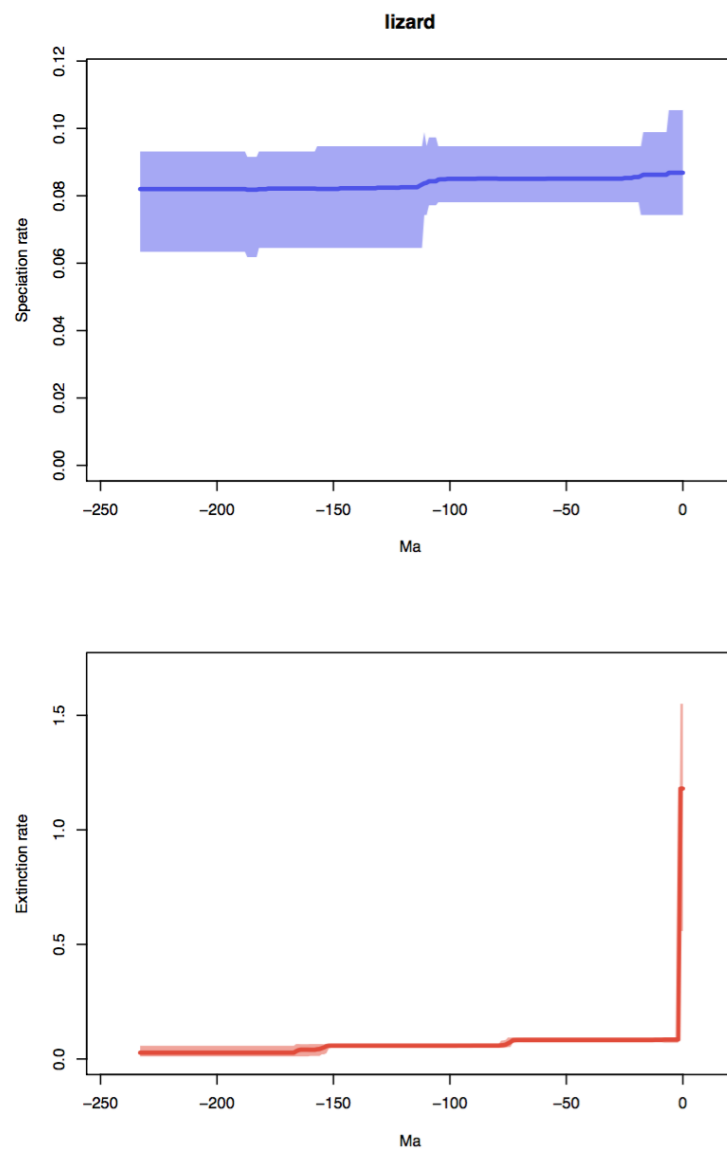Additional images produced with a shorter PyRate run that expands on questions and ideas posed in the paper.
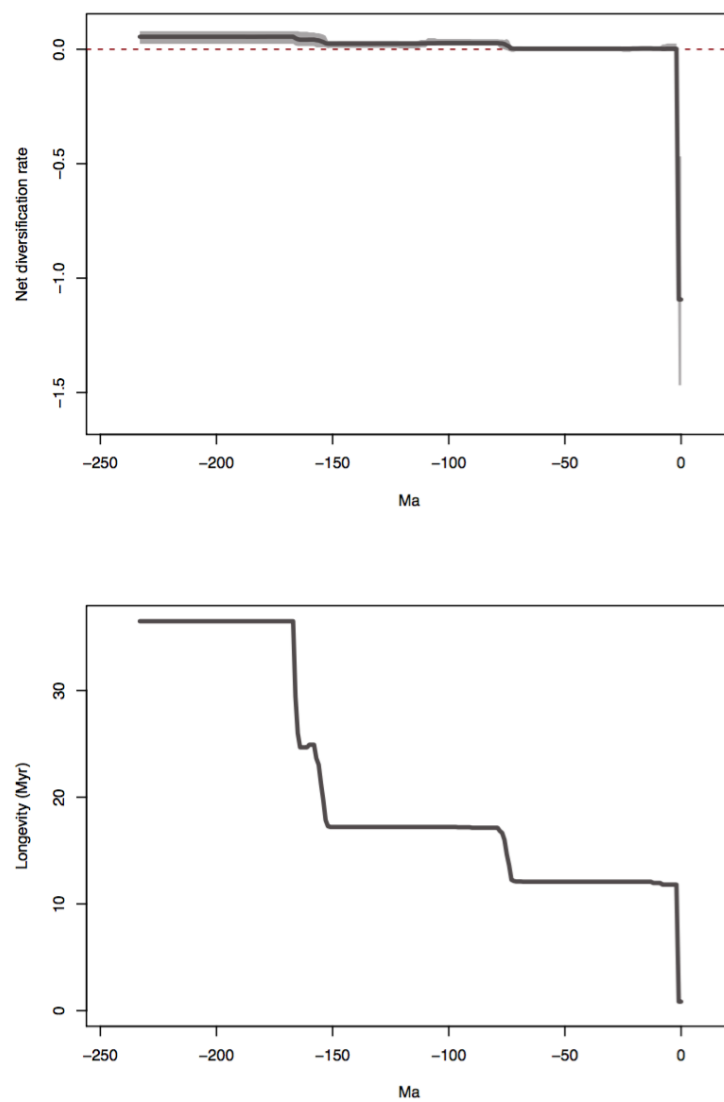
Figure 3: wowowowow

Figure 4: wowowowow