

Examining Mustelidae Fossil Records, Occurrences, Speciation, and Diversity over Time

Tiffany Dang

Abstract

The Mustelidae family, comprising of carnivorous mammals such as otters and weasels, is the largest family within the Carnivora order. Evidently, these species have survived and outcompeted many other families within this order for millions of years. However, there are much fewer extant species compared to the species diversity within this family millions of years ago. Thus, this paper seeks to examine the fossil occurrences, the speciation rates, and the extinction rates throughout history. In this project, I will be fetching the mustelidae data from PaleoBiology Database in the shell, organizing relevant information such as the maximum and minimum ages of the fossil ranges of each species in python, and visualizing the data in graphs created in Rstudio. Additionally, I will be using PyRate on the mustelidae file to visualize the diversification and extinction rates. I found that while there were many species lost throughout history, there has also been enough speciation and growth to account for the loss, leading to constant rates of extinction and species diversification throughout history.

Introduction

The Mustelidae family consists of carnivorous mammals, such as weasels, polecats, otters, and wolverines. This family is the largest family of the Carnivora order. With multiple subfamilies within Mustelidae, it is believed that the lineages went through adaptive radiation, leading to rapid speciation dating back to the early Eocene period (Marmi *et al.* 2004). Fossils of carnivores are often hard to find considering carnivores are more rare than their prey. Thus, researchers rely on dentitures to track carnivores (Martin 1989). Additionally, researchers have found that delayed implantation, a form of diapause, which displays itself through a pause in early developmental stages, is highly prominent in mustelids. While delayed development seems to be detrimental to the survival of predators, the mustelids appear to persist through this. Thus, by examining the time periods of mustelids and the environment during these times, scientists hope that findings will shed light on the adaptive value of delayed implantation and factors that possibly contributed to the survival of mustelids despite delayed implantation (Thom *et al.* 2004). In addition to the rapid speciation that occurred in the early Eocene period, it appears that the mustelids have successfully taken advantage of dietary and environmental niches, allowing them to coexist. For example, the European polecat and American mink, both carnivorous mammals that are similar in size, would be

expected to compete against one another for survival. However, research has found that in addition to rodents, polecats rely on amphibians to supplement their diet while minks prey on fish (Lodé 1993). In general, habitat and resource partitioning have allowed many Mustelidae species to exist in sympatry, a possible reason for existence as the largest family within the Carnivora order. This paper seeks to explore the mustelid occurrences, diversification rates, and extinction rates of mustelids through history and examine times of possible species loss.

Materials and Methods

Retrieving and Cleaning Data (Shell):

In the shell, I downloaded a csv file containing a list of fossil occurrences of species within the mustelidae family from the Paleobiology Database.

```
wget https://paleobiodb.org/data1.2/occs/list.csv?datainfo&rowcount&base_name=mustelidae
```

This file contains various information, including time period the species lived, the maximum millions of years ago the species lived, and the minimum million years ago the species lived. I renamed the file to one that could be easily identified and called upon in later code.

```
mv pbdb_data\14.csv mustelidaepbdb.csv
```

In order to obtain accurate results, I checked for a sufficient sample size in the csv file. The following code sets the delimiter as a comma, searching through the accepted names(column 10) and counted the number of lines containing unique names starting from the 19th line, disregarding the heading.

```
tail -n +19 mustelidaepbdb.csv | cut -d "," -f 10 | sort | uniq | wc -l
```

Organizing and Compiling Relevant Data (Python):

Using the information contained in the csv file, which included the species names and the maximum and minimum age of the fossil range for each species in millions of years ago, I created dictionaries to extract and organize the relevant information. First, I created a function that would return a dictionary containing the species' accepted names as the keys and the time the species fossil was first seen as recorded in the maximum million years ago column(column 15) of the csv file. This function looped through each line of the csv file, extracting the species name and appending the time of the species' first arrival as the value.

```
## gives dictionary of species and its first arrival
def first_occ(filename):
    firstdict = {}
    #initiate dictionary that will hold the species aname and the first arrival of the
    mustfile = open(filename, "r", encoding = "ISO-8859-15") #opens and reads the file
    mustrecord = mustfile.readlines()[19:] #make sure to start at numerical data
```

```

#for loop that will fill the dictionary with the species names and their arrival d
for line in mustrecord:
    mustelements = line.split(",") #separates each element by a comma
    species = mustelements[9][1:-1] #remove quotation marks around the species name
    first_arrival = float(mustelements[14][1:-1]) #remove quotation marks around the
    firstdict[species] = first_arrival #assigns the key-- species a value --first a
return firstdict

```

Similarly, I created a function that returned a separate dictionary containing the species' accepted names as the keys and the time the species was last seen based on the minimum million years ago column (column 16) in the csv file. Like the previous function, the code block loops through each line of the csv file, recording the species name as the key and appending the time the species was last seen in the fossil record as the value.

```

def last_occ(filename):
    lastdict = {}
    #initiate a dictionary that will hold the species name and the time it was last se
    mustfile = open(filename, "r", encoding = "ISO-8859-15") # opens and reads the file
    mustrecord = mustfile.readlines()[19:] #starts reading at the numerical data
    #for loop that will fill the dictionary with the species names and the time it was
    for line in mustrecord:
        mustelements = line.split(",") #tells the computer that each element in the fil
        species = mustelements[9][1:-1] #gets the species name without the quotation ma
        last_seen = float(mustelements[15][1:-1]) # gets the time the species was last
        lastdict[species] = last_seen #assigns the key--species name-- a value-- the ti
    return lastdict

```

Based upon the previous two dictionaries, which organized the species and its associated time of arrival and extinction, I wanted to examine the duration of each species. In order to do so, I created a loop that would read each line of the csv file, recording the species name as the key, calculating the age by subtracting the time the species was last seen from the time it was first seen, and appending this age as the value for the species in a dictionary. The time a species was first seen and the time a species was last seen were originally read as strings in the csv file, preventing conversions to floats and thus calculations. In order to solve this issue, the values stored in the max_age and min_age variables only started at the second character and ended before the last character, excluding the quotation marks around the values. From here, the values could be converted into floats and the age could be calculated and appended.

```

## gives the age of each species based on the max mya and min mya
def age_dictionary(filename):
    agedict = {}
    #initiate the dictionary
    mustfile = open(filename, "r", encoding = "ISO-8859-15") #opens and reads the file
    mustrecord = mustfile.readlines()[19:] #starts at the data, ignores the headers etc
    #for loop to fill the dictionary
    for line in mustrecord:

```

```

record_elements = line.split(",")
species = record_elements[5] #assigns the elements in column 6 as the species
max_age = record_elements[14][1:-1] #removed the quotation marks around the num
min_age = record_elements[15][1:-1] # removed the quotation marks around the nu
age = float(max_age) - float(min_age) #converting the strings in columns 15 and
agedict[species] = age #assigns the key--species a value--age
return agedict

```

In order to declutter the workspace with multiple dictionaries, the information in the previous dictionaries were condensed into a single dictionary. In order to better organize and compile the fossil time range of each species, I created another function using defaultdict from collections that returned a dictionary holding a list as the value for each species name. The list includes the time the species was first seen in million years ago as well as the time the species was last seen in million years ago based on fossil records. This dictionary encompasses the information necessary to visualize the data later in Rstudio.

```

## large dictionary that has both the species' first and last ages
def create_gendict(filename):
    # uses the default dictionary we learned in class on thursday
    from collections import defaultdict
    #uses the function above that created a dictionary filled with the species' name a
    arrivaldict = first_occ(filename)
    #uses the function above that created a dictionary filled with the species' name a
    enddict = last_occ(filename)

    generaldictionary = defaultdict(list)

    #creates a dictionary. if the key already exists, the loop will append the value t
    for index in (arrivaldict, enddict):
        for key, value in index.items():
            generaldictionary[key].append(value)
    return generaldictionary
    #in the end, each species name should have a time of first occurrence and a time o

```

After organizing the necessary information into various dictionaries, I wanted to create a clean csv file that contained only the data required for my graphs. This csv file holds the genus name, the species name, the minimum age, and the maximum age for each species in the fossil record. The keys and values were extracted from the dictionary created using the previous function and recorded into a new csv file.

```

output = open("must_summary.csv", "w")
generaldictionary = create_gendict("mustelidaepbdb.csv")
for key in generaldictionary.keys():
    ages = generaldictionary[key]
    minage = min(ages)

```

```

maxage = max(ages)
genus = key.split(" ")[0]
species = key
#print(genus, key, minage, maxage)
outline = "{},{},{},{}\n".format(genus, species, minage, maxage)
output.write(outline)
output.close()

```

Visualizing Fossil Occurrences (RStudio):

Using the cleaned file, the species durations were plotted in a graph using ggplot in Rstudio. In order to ensure the correct data was being inputted into Rstudio, I opened the file and examined the data using the following code.

```

library(ggplot2)
mustelidae <- read.csv("//home/eeb177-student/Desktop/eeb-177/finalproj/mustelidae_summa
names(mustelidae) <- c("genus", "species", "minage", "maxage")
head(mustelidae)

```

After confirming the information, the graph was formatted with the time (measured in millions of years ago) on the x-axis and the species names on the y axis. The graph was then adjusted for optimal viewing of the information depicted. Various adjustments include flipping the coordinates, removing the tick marks, and placing and adjusting the font size of the graph and axes titles.

```

mustelidae_occ <- ggplot(mustelidae, aes( species, ymin = maxage, ymax=minage, colour = 
mustelidae_occ <- mustelidae_occ + geom_linerange()
mustelidae_occ <- mustelidae_occ + theme(legend.position="none") #gets rid of legend
mustelidae_occ <- mustelidae_occ + coord_flip() #flips the coordinates horizontally
mustelidae_occ <- mustelidae_occ + theme(axis.text.y = element_text(size=2)) #resizes
mustelidae_occ <- mustelidae_occ + theme(axis.ticks.y=element_blank()) # gets rid of ti
mustelidae_occ <- mustelidae_occ + scale_y_continuous(limits=c(0, 25), expand = c(0, 0)
mustelidae_occ <- mustelidae_occ + labs(title = "Mustelidae Fossil Occurrences", x = "Sp
mustelidae_occ

```

Diversification and Extinction Rates (PyRate):

In addition to the fossil occurrences, I wanted to examine the species diversification and extinction rates of the mustelids throughout the time periods. In order to run PyRate, I had to switch to Python 2.7 in the shell. PyRate ran through the data set 250,000 times in order to generate species diversification and extinction rate graphs. In order to run PyRate, I downloaded a new mustelidae file from Paleobiology Database.

```
wget -O "https://paleobiodb.org/data1.2/occs/list.csv?datainfo&rowcount&base_name=Mustelidae"
```

I then created an R script to translate the data, with the extant species inputted manually, to a format that could be read in PyRate.

```
source("~/PyRate/pyrate_utilities.r")

# we need to give the utilities a list of extant species
extant_mustelidae = c("Taxidea taxus", "Sardolutra ichnusae", "Pteronura brasiliensis",

# use the extract.ages.pbdb() function in pyrate_utilities to reformat our dataset...
extract.ages.pbdb(file= "mustelidae_occ.csv", extant_species=extant_mustelidae)

# you should see the following output:

# "This function is currently being tested - caution with the results!"
# replicate 1
# PyRate input file was saved in: ./mustelidae_occ_PyRate.py
```

I then ran PyRate using the newly formatted data in the shell using the following code.

```
python ~/PyRate/PyRate.py canid_occ_PyRate.py -n 250000
```

After running the code, a new directory was created from which I was able to plot and examine the species diversification and extinction rates with the following code.

```
python ~/PyRate/PyRate.py -plot mustelidae_occ_1_marginal_rates.log #plotting the graph
see canid_occ_1_marginal_rates_RTT.pdf # views the graphs just created
```

Results and Discussion

Mustelidae Fossil Occurrences

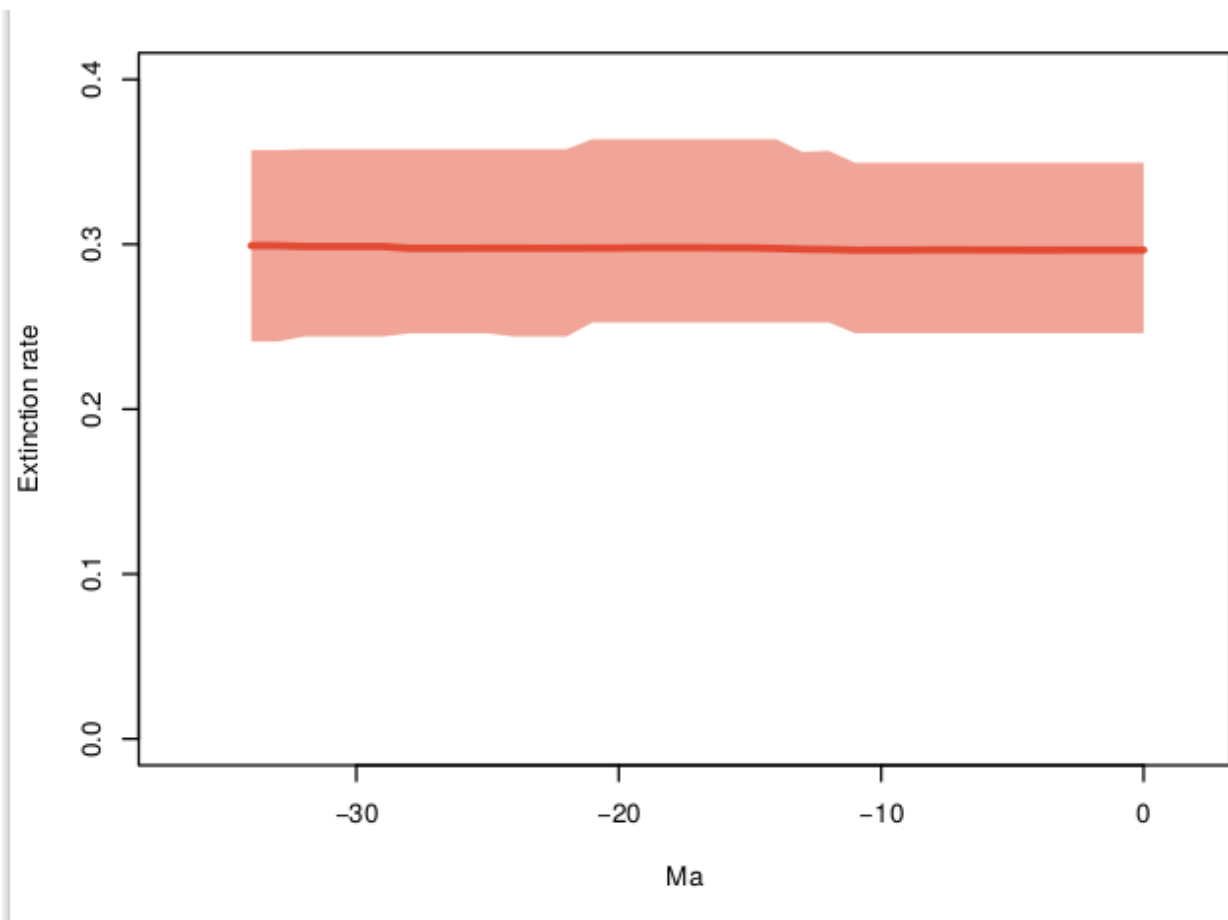
```
## Warning: Removed 8 rows containing missing values (geom_linerange).
```

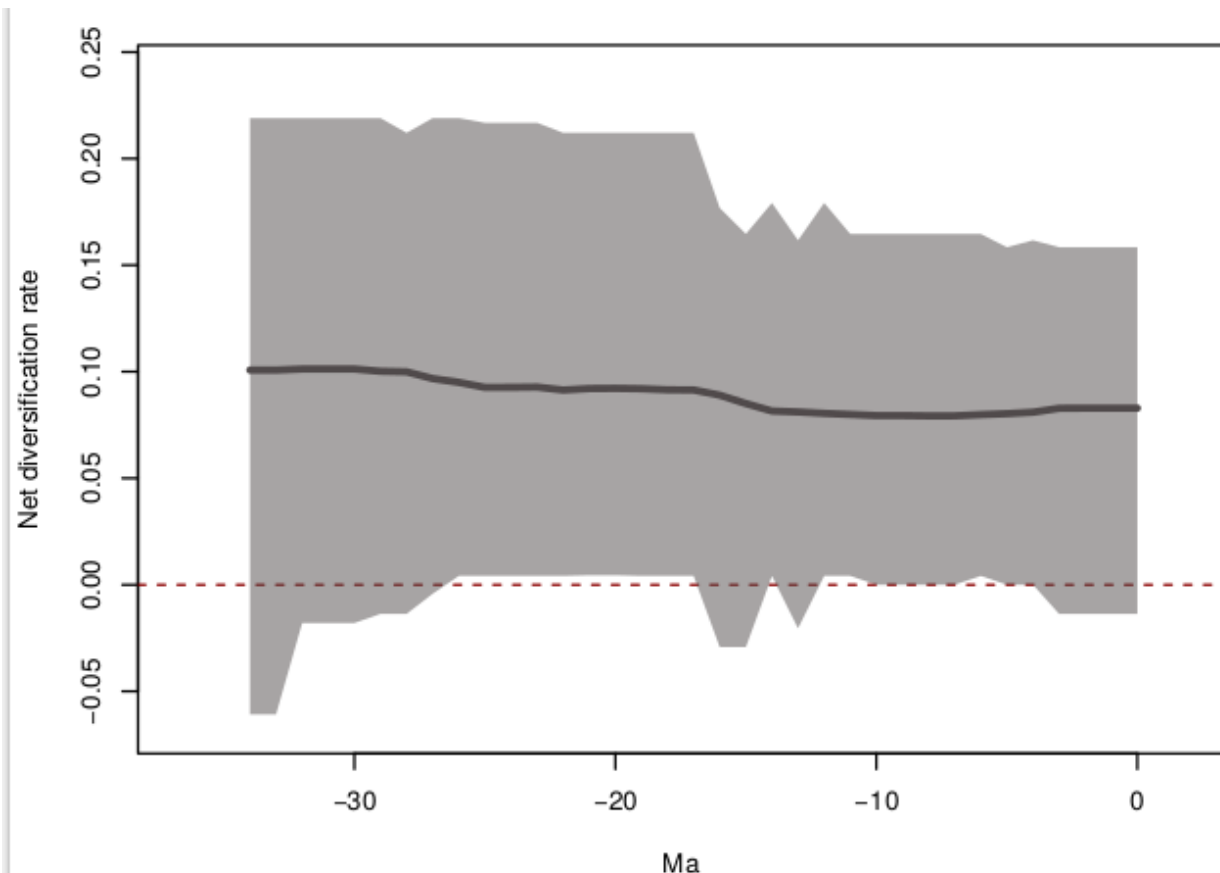
Species

Time (Millions of Years Ago)

0 5 10 15 20 25

7





Despite the many species that appeared to die off throughout history, it appears that the species diversification rates and the extinction rates remain constant. This shows that even with the changing environment throughout these millions of years, mustelids have been able to adapt to their conditions and coexist with other species while outcompeting other predators. Mustelids have demonstrated sympatric speciation, with multiple species existing in the same geographic area. While mustelids are carnivorous mammals, and therefore subsist on the same general diet, it appears that they also partition these resources, with some species, such as otters, preying on fish, while others, such as minks, consume birds and mammals. Additionally, physical features, such as vision, differ between different species, contributing to additional resource partitioning. For example, minks' vision has been adapted for hunting on terrain while otters have the visual acuity and lung capacity to hunt for fish (Powell & Zielinski 1983). These characteristics allow various species to further partition resources and continue to coexist despite their similar features and diet. This ability to exploit niches may have contributed to the success of the Mustelidae family through time.

Conclusion

Unlike many other species presented in class, the graphs generated from Python as well as the ones from generated in Rstudio, it appears that mustelids do not demonstrate the same sharp decline in species diversification and sharp increase in extinction rates in the recent

millions of years. Rather, the family appears to have constant species diversification and extinction rates through time. While they have been able to efficiently partition resources in the local environment, it is important to stress the need for conservation in order to maintain these various niches. For example, while minks normally utilize streams, whereas otters tend toward larger lakes, during the winter seasons, minks and otters are limited to the same bodies of water, increasing resource overlap, and thus competition. Similarly, if we continue to see lakes and other bodies of water continue to dry out, we may see a decline in fish, which would limit the food available, forcing the animals to fall back toward the same rodents. With limited resources, we could see an increase in competition once again and thus a possible decline in the number of species. Therefore, we must ensure that we continue to conserve our environment in order for these mustelids to continue thriving

GitHub link

<https://github.com/tdaang/finalproj>

References

- Lodé, T. (1993). Diet composition and habitat use of sympatric polecat and american mink in western france. *Acta Theriologica*, **38**, 161–166.
- Marmi, J., Lopez-Giraldez, J.F. & Domingo-Roura, X. (2004). Phylogeny, evolutionary history and taxonomy of the mustelidae based on sequences of the cytochrome b gene and a complex repetitive flanking region. *Zoologica Scripta*, **33**, 481–499.
- Martin, L.D. (1989). Fossil history of the terrestrial carnivora. *Carnivore behavior, ecology, and evolution*, pp. 536–568. Springer Nature.
- Powell, R.A. & Zielinski, W.J. (1983). Competition and coexistence in mustelid communities.
- Thom, M.D., Johnson, D.D.P. & Macdonald, D.W. (2004). The evolution and maintenance of delayed implantation in the mustelidae (mammalia: Carnivora). *Evolution*, **58**, 175–183.