DISSERTATION RESEARCH:
Cenozoic mammals and the biology of extinction

PI: Kenneth D. Angielczyk, Co-PI: Peter D. Smits

Introduction

Background

Why species go extinct at different rates remains one of the most fundamental questions in paleobiology. It is expected that for the majority of geological time, extinction is biologically non-random [1, 28, 32, 39, 40, 51, 52]. Determining which and how biological traits influence extinction risk is vital for understanding the differential diversification of life during the Phanerozoic and making predictions about species' vulnerability to anthropogenic impacts. Yet investigations into the factors that affect extinction rate have proved inconclusive and often contradictory results.

For example, species with larger geographic ranges tend to have lower extinction rates than species with smaller geographic ranges [29, 32, 34, 51, 64]. This pattern may be random with respect to differences in organismal traits [59]. However, this may not always be the case and organismal traits directly related to a species environmental interactions may play an important role in extinction risk. The inclusion of organismal trait data are necessary to determine if and how both species and organismal traits contribute to differences in extinction risk or not. By modeling extinction via traits related to environmental preference, the relative importance of species and organismal level properties can be elucidated.

The Law of Constant Extinction [78], which states that a taxon's extinction risk within a given adaptive zone is age-independent (memoryless), is a classic generalization about extinction. This law is the foundation for the Red Queen hypothesis as well as most approaches for quantifying extinction. However there is an increasing amount of evidence contrary to this law [16, 17, 57, 65]. By analyzing survival patterns within adaptive zones during periods of background extinction, it should be possible to both estimate the effects of various ecological strategies on survival and determine if extinction is age-independent or dependent. Periods of background extinction provide a great opportunity to study how traits are related to survival because they represent the majority of geologic time, remain relatively predictable and change slowly [32, 60].

Finally, the availability and stability of a species' biotic and abiotic environment are crucial for the establishment and persistence of the species. The quantification of all interactions in a community is difficult at best (though it has been accomplished for small temporal windows and limited spatial scales [7, 50, 61]), obscuring how such interactions contribute to the spatial heterogeneity of selection

pressures that species experiences. An alternative approach is to focus on on spatially co-occurring taxa, which are a more general subset of the environment and can be studied at a greater temporal duration. This approach is similar to the demarcation of biomes as opposed to food webs [80]. By tracking the number, membership, and similarity of communities over time and how this relates to taxon ecology, it should be possible to estimate spatial heterogeneity of selection pressure.

It is within this framework that I propose to study how ecological traits associated with environmental preference have affected both Cenozoic mammal spatio-temporal dynamics and extinction risk. I am testing the following hypotheses: does dietary category, locomotor category, and body size have predictable effect on the spatial distribution of taxa through time, is extinction non-random with respect to ecology during times of background extinction, and is extinction taxon age independent among Cenozoic mammals? By leveraging the mammal fossil record and integrating over these hypotheses, I am able ask the question "why do species go extinct at different rates?" Cenozoic mammals represent an ideal group and time period because their fossil record is well sampled, well resolved both temporally and spatially, and the ecology and phylogeny of individual species are generally understood [4, 6, 37, 42, 71].

Large spatial and temporal scale analyses and syntheses of the mammal fossil record have been restricted to North American and Europe, lacking information from the Southern Hemisphere [2, 3, 6, 24, 36–38, 42, 56, 77]. The funds from this grant will allow for significant enhancement of these projects through the inclusion of a third and rarely integrated continent: South America [48, 76]. By including the South American mammal fossil recod along with the North American and European records, I will be able to better understand both spatio-temporal dynamics and extinction risk at both regional and global scales.

Study system

Mammals are motile organisms which can track their preferred environment over time and space. Three important traits that describe the relationship between mammals and their environmental context are dietary category, locomotor category, and body size [13, 14, 38, 43, 44, 71, 72]. Each of these traits describe different aspects of a taxon's adaptive zone such as energetic cost, population density, expected home range size, set of potential prey items, and dispersal ability. I will be studying how these traits affect both spatial distribution of taxa and extinction risk in Cenozoic mammals from North America, Europe, and South America.

Dietary category roughly describes the trophic relations of a taxon, a central component of its biotic environment. The categories used here are coarse groupings of similar ecologies: carnivore, herbivore, omnivore, and insectivore. The first three of these represent commonly used groupings of mammals in paleontological studies [38, 54], while the fourth is a biologically important grouping. Price et al. [54] found that mammalian herbivores and carnivores have a greater diversification rates than omnivores which may indicate that these traits are better for survival. An increase in diversification can be due to either an increase in speciation relative to extinction or a decrease in extinction relative to speciation. Which scenario occurred, however, is impossible to determine from a phylogeny of only extant organisms [55]. By analyzing the fossil record of extinct organisms, the results of Price et al. [54] can be better understood from a mechanistic perspective.

Locomotor categories describe the motility of a taxon, plausibility of occurrence in a particular habitat, and dispersal ability. Dispersal ability is important for determining both the extent of a taxon's geographic range and ability to track changing environments [10, 25, 33] which then affects both extinction risk and community similarity. Here, mammals are categorized as either arboreal, ground dwelling, or scansorial. With the transition from primarily closed to open environments during the Cenozoic [11, 35, 75, 76], it is expected that arboreal taxa during the Paleogene will have a greater expected duration than Neogene taxa while the opposite will be true for ground dwelling taxa. In comparison, taxon duration of scansorial taxa is expected to remain relatively similar between the two time periods because it represents a mixed environmental preference that may be viable in either closed or open environments.

Body size, here defined as mass, has an associated energetic cost in order to maintain homeostasis which in turn necessitates a supply of prey items. Many life history traits are associated with body size such as reproductive rate, metabolic rate, and home range size [12, 13, 53, 71]. Body size may affect extinction risk because as body size increases, home range size increases [13]. If individual home range size scales up to reflect a species geographic range, this would mean that extinction risk would decrease. Alternatively, it could be argued that as body size increases, reproductive rate decreases [39], populations get smaller [83], and generations get longer [49], all of which increase extinction risk. A negative relationship between mammal body size and taxon duration has been observed [15, 42] though this is inconsistent between continents [42, 77].

By expanding to include a third continent, South America, and analyzing species level data I hope to elucidate how differences in taxonomic diversity at a continental level might affect both spatial distribution and extinction risk. Through collaboration with Dr. Richard Madden, I have access to a near complete geochronology of South American fossil mammals. However, I am missing the relevant biological trait information for many species. Travel to institutions in South America (i.e. Argentina) is vital for fully incorporating this information.

Preliminary results and proposed work

Community connectedness

How ecology affects the spatial distribution of taxa, and how this affects the relation between provincial and cosmopolitan taxa over time, is important for understanding the spatial heterogeneity of selection pressures. If ecologically similar taxa are, on average, present in many places then it is expected that selection pressures are spatially homogeneous. In contrast, if ecologically similar taxa are spatially restricted, it would be expected that selection pressures are spatially heterogeneous. To measure taxon spatial distributions over time I use a network theoretic approach where taxa and localities are nodes connected in a bipartite network [66, 80, 81]. An occurrence of a species at a particular area creates a link between that species node and that area node. Links are defined only as taxon-locality occurrences, where species cannot be connected to species and spatial nodes cannot be connected to spatial nodes.

This strategy allows for taxon occurrences to be analyzed wholesale without use of a possibly arbitrary similarity metric to summarize the pair-wise relationships of all communities [66]. An

additional advantage of this approach is the application of alternative measures of community structure such as the map equation [62, 63] which estimates the number of modules in a network based on its topology. A module in this context is a set of taxa and localities which are more connected to each other than to other modules, effectively a biome or province. This approach has been used in the study of communities at the Permo-Triassic boundary [66] and bivalve provinces at the Cretaceous-Tertiary boundary [81]. Here, I propose to apply this approach to the mammalian record for the entire Cenozoic to estimate changes in the spatial heterogeneity of selection with regards to ecology.

The four measures of community structure used here are relative number of taxa unique to a locality, relative taxon provincial occupancy, biogeographic connectedness (which is similar to biome evenness) [66], and network code length. Code length is an information theoretic measure of how most efficiently to divide a network into modules and acts here as a measure of total spatial similarity [62, 63]. In the context of this study, taxa are defined as species and spatial units are from a uniform latitude-longitude grid. Because of the high spatio-temporal resolution of the mammalian fossil record for North America and Europe [4, 24, 38, 47], temporal units are 2 million year (My) bins and spatial units are from a 2x2 lat/long grid.

In addition to these four measures of community structure, I will measure phylogenetic similarity of the taxa at a locality. Phylogeny may play an important role in community structuring where closely related taxa may be "repulsed" due to competitive exclusion or "clumped" because of environmental filtering [82]. While it is infeasible to infer an explicit phylogenetic hypothesis for all observed mammal species on all continents, all taxa have some degree of hierarchical information. By combining multiple large scale extinct mammal phylogenies with taxonomic information, an informal phylogeny can be constructed. From there, measures of phylogenetic similarity such as the phylogenetic species variability [30] can be estimated for every locality.

Preliminary results are restricted to North America and Europe using fossil occurrence information downloaded from both the Paleobiological Database (PBDB, `http://paleodb.org/`) and the Neogene Old World Database (NOW; `http://www.helsink.fi/science/now/`). Dietary and locomotor assignments for each species were taken from the PBDB and the NOW. Comparisons of locality phylogenetic similarly have yet to be analyzed. These results are primarily exploratory. See Figure 1 for species-level dietary category summary.

Preliminary results show that all measures biome similarity are generally volatile through out the Cenozoic (Fig. 1). Relative taxon provincial occupancy (Occ) is relatively constant for most of the Cenozoic with a few spikes during periods of poor spatial sampling. Code length, a measure of the number and uniqueness of biomes, shows increasing separation between dietary categories during the Cenozoic. Both biogeographic connectedness (BC) and the relative number of taxa unique to a locality (E) fluctuate the most, though it appears to be mostly random movement around a constant mean.

Future analysis of these network measures will be done using a hierarchical Bayesian time series model where, for each continent, the different ecological combinations are exchangeable samples from a common underlying continent-specific mammalian distributional pattern. Because only two continents are currently sampled, stable estimates of an additional hierarchical level where continents are drawn from a common underlying global mammal distributional pattern is not possible [26].
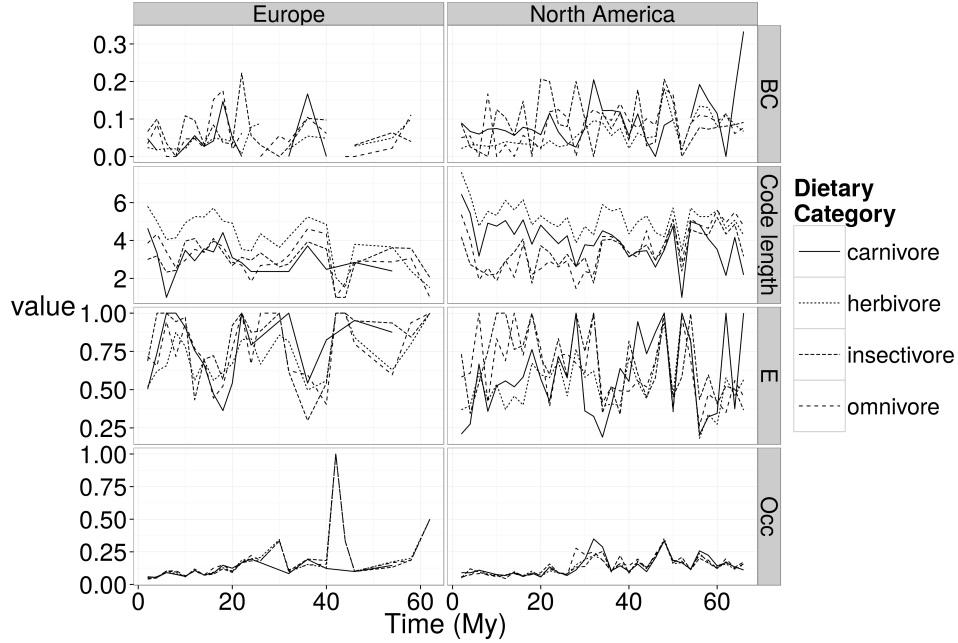
Figure 1: Biogeographic network summary statistics for North American (right) and European (left) species-level occurrence information. From top to bottom, these summaries are biogeographic connectedness (BC), code length in bits, relative number of taxa unique to a locality (E), and the relative taxon provincial occupancy (Occ). Measures are separated based on dietary category. Missing values are due to poor sampling in that bin.

The inclusion of a third continent, South America, would allow for potentially stable estimates of a underlying common global mammal distributional pattern.

Extinction risk

How ecology affects extinction risk is extremely important for having a more mechanistic understanding of the differential proliferation of taxa. Related to this is the Law of Constant Extinction which states that extinction is taxon age independent [78]. I use the mammalian fossil record to measure both the effect of ecology on taxon duration as well as taxon age upon extinction.

Species durations will be modeled in a survival analysis framework where the effect of the above traits can be quantified as the effect upon a taxon's duration. This approach has a long history in paleobiology [9, 19, 68, 69, 78, 79] that has matured greatly as this analytical framework is commonly used in epidemiology, engineering, and demography studies [41]. What makes survival analysis different from various linear modeling strategies such as logistic regression is that both duration and event occurrence are incorporated explicitly. In logistic regression, for example, only event occurrence is modeled, while in linear regression only duration is modeled. Importantly, because some taxa may have originated before the Cenozoic or not have gone extinct yet, this information can be explicitly modeled as a censored duration which increases the total information included in the model.

Many paleontological analyses of extinction selectivity and survivorship are performed at the genus level [17, 22, 29, 42, 77] though there are potential biases in accurately modeling a specific level process using genus level data [57, 58, 65, 67, 79]. Differences in species and genus level extinction risk can be attributed to speciation. Namely, if a trait has no effect on extinction risk at the specific level but decreases genus extinction risk this difference is most likely caused by differences in speciation rate. Because of this potentially elucidating property, I will be analyzing both species and genus level extinction risk. Importantly, this moves beyond simple analysis of diversification rates by partially decomposing the two aspects of diversification: speciation and extinction.

My current analyses are restricted to North America and Europe fossil occurrence information downloaded from both the PBDB and the NOW. Duration was measured as the difference between first appearance (FAD) and last appearance (LAD), in millions of years. This value is most certainly a truncated version of the true species duration and so some amount of correction must be made for sighting failure [5, 73, 74]. As with the study of community connectedness described above, dietary and locomotor assignments for each species were taken from the PBDB and the NOW, supplemented with data from the primary literature. Body size estimates are based on a combination of measurement data from the PBDB, NOW, and a species by species literature search. These estimates are based on the common practice of using regression equations for estimating mass from measures such as tooth area or skull length [3, 4, 38, 70, 77].

Preliminary analyses were performed solely within a maximum likelihood and multi-model inference framework. Duration information was fit to either an exponential or Weibull distribution, both common distributions for wait times. The exponential distribution corresponds to age-independent extinction or the Law of Constant Extinction [78] while the Wiebull distribution allows for age-dependence in extinction risk. Dietary and locomotor categories were broken up into $r-1$ binary predictor variables, where r is the number of unique states for both traits, for a total of five binary predictors. A constant term was also estimated which corresponds to the remaining dietary and locomotor states. For North America, the final biological predictor was natural log of body mass. Body size has not yet been included in the analysis of the European record; that data collection effort is not yet been completed. As is common practice, the k parameter of the Weibull distribution was estimated as a single constant [41]. Because of the volume of preliminary results, presented here are survival functions based on dietary and locomotor category for both North American and European species (Fig. 2).

There are pronounced differences between the survival patterns of these two continents. North American species show little variation in extinction risk based on diet, with carnivore taxa being on average the longest lived. In Europe the differences are more striking with omnivorous taxa followed by insectivorous being the longest lived on average (Fig. 2a). For both North America and Europe, locomotor category appears to have more effect upon extinction risk, with North American ground dwelling taxa and European arboreal taxa being the longest lived on average for their respective continents (Fig. 2b).

In future analyses, measures of climate and geographic range will also be included as continuous predictors of duration. Global climate, though not a taxon's local climate, will be proxied via mean δO^{18} isotope value as calculated from Zachos et al. [84] over the duration of the species. Mean provincial occupancy over the duration of the species will be calculated using the network theoretic approach described above.
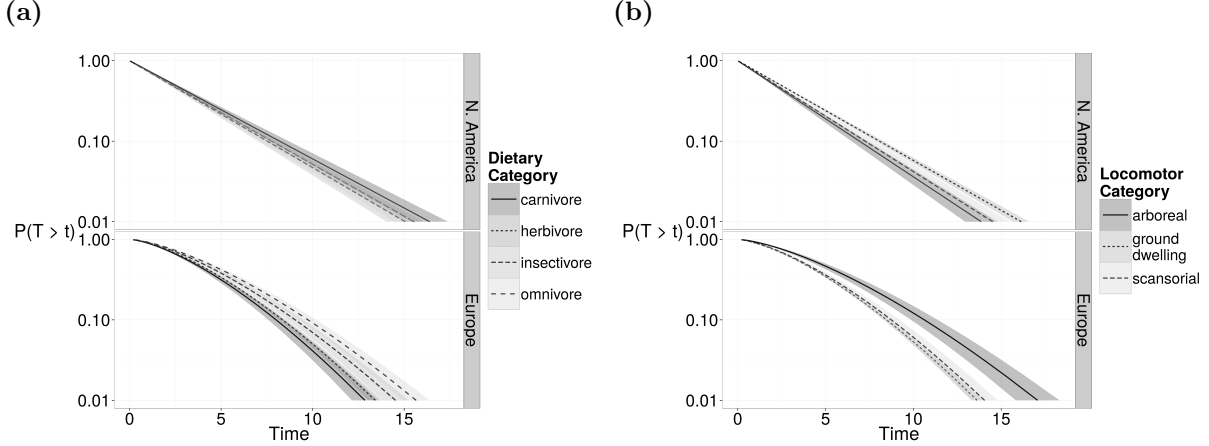
Figure 2: Survival curves for mammal species based on dietary (a) and locomotor category (b) for North America and Europe. North American curves are from an exponential model while European survival curves are from a Weibull model. Distribution choice is based on model selection results. This curve is an estimate that a species will survive till time T having survived until time t.

Future modeling work will be done in a purely Bayesian framework in order to 1) better quantify the uncertainty with which estimates are made, 2) combine the analysis of the different continents in a hierarchical framework in order to estimate a common underlying distribution of survival times, 3) incorporate phylogenetic distance into a prior on a frailty coefficient [8, 31] to account for potential phylogenetic autocorrelation in species durations, and 4) to move on a continuous model expansion framework as opposed to one of discrete model choice [26]. All of these changes will dramatically improve the interpretability and meaning of the results. Importantly, estimates under goal 2 would not be possible without the inclusion of a third continent, South America.

A unique aspect of the South American record is the presence of multiple ungulate-like orders that are rare or absent from other continents (e.g. Notoungulata) [18, 45, 46, 48]. These convergent orders combined with the other ungulate-like orders present in North American and Europe make it possible to estimate, in a hierarchical Bayesian framework, the order-specific distributions of taxon durations as well as the underlying distribution common to all ungulate-like taxa [26].

Intellectual merit

I propose to analyze how ecology affects both spatial distribution and temporal duration of mammalian species. The mammal fossil record allows me to model complex relationships that are both more spatially and temporally explicit than normally possible.

The Law of Constant Extinction, that extinction within an adaptive zone is taxon age-independent [78], is a foundational statement underlying both the Red Queen hypothesis and many approaches to quantifying extinction intensity [5, 20, 21, 23, 57, 65]. With an increasing amount of evidence contrary to this law [16, 17, 57, 65], it is important to explicitly model and test this claim. The survival analytical approach outlined above is an explicit statistical framework for estimating the

effect of taxon age on survival.

Changes in provinciality are associated with changes in community stability and the heterogenity of selective pressures [66, 81]. By estimating changes in the spatial heterogeneity of selection along with complementary ecologically explicit survival analyses, it may be possible to actually address the major question of "why do species go extinct at different rates?" The spatial and temporal resolution of the mammalian fossil record provides a unique opportunity to study changes in spatial relations over long periods of time. I take an alternative and relatively new approach to identifying and quantifying biotic provinces [66, 80, 81] to analyse the changes in the degree of endemism and relative spatial similarity of whole continents.

I have the unique opportunity to collaborate with Dr. Richard Madden and include his dataset of South American fossil occurrence information in the synthesis, which has been historically restricted to North America and Europe. The inclusion of South America into the analysis of mammalian diversification will provide a much needed Southern Hemisphere perspective and will greatly improve the generality of these analyses and conclusions. One of the major limiting factors which has prevented the inclusion of South America in large-scale studies of Cenozoic mammal evolution is that the majority of this literature is in Spanish. This lack of exposure is similar to the lament that, in the mid 20th Century, the work of continental authors writing in German were under represented in paleontological syntheses [27]. My hope is that through this study much of the information seemingly locked away will be made available to the paleobiological community.

Broader impacts

My past and current research are based out of natural history museum collections I have been involved in several wonderful outreach programs. I have participated as an expert on both marsupial mammals and fossil mammals with the "Meet the Mammals" and "Dino Day" programs at the Burke Museum of Natural History and Culture during my undergraduate in Seattle. These programs were designed specifically to reach out to children and gave them the opportunity to interact directly with the museums collections. I plan to continue to participate in similar outreach initiatives at the Field Museum of Natural History in Chicago such as "Meet the Scientist" program where I can directly present my research and specimens to the public in the middle of the museum.

Most of the literature on South American fossil mammals is in Spanish and is effectively unavailable to many researchers. One of my goals with this study is to open up this literature. Related to this, I plan to learn conversational Spanish so that I can engage with students and other researchers in South America about my project. I am committed to providing whatever data I gather will be made available through the Dryad data archiving service (`http://datadryad.org/`) and will be entered into the Paleobiological Database (PBDB, `http://paleodb.org/`) so that South America will continue to be included in future paleobiological studies.

Mental health of students and researchers is both important and an interest of mine. I am involved with GSMHAB, a graduate student group involved with outreach and communication of mental health services at the University of Chicago campus. The group specifically targets graduate students, which make up two-thirds of the campus' student population. Additionally, I will be involved with mentoring entering first-year graduate students through the SAIL program at the university.