ECTOTHERM DIVERSITY THROUGH TIME: THE INTERPLAY OF SAMPLING BIASES, ENVIRONMENTAL DRIVERS AND MACROEVOLUTION

by

Terri Jade Cleary

A thesis submitted for the degree of

Doctor of Philosophy

University College London

September 2018

[This page intentionally left blank]

I, Terri J. Cleary, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

[This page intentionally left blank]

Abstract

The fossil record is inherently biased, but few studies have examined the effects of these biases on our understanding of non-marine vertebrate ectotherms through time. Here, I investigate the taxon richness of non-marine lepidosaurs and turtles from the Triassic–Paleogene (252–23 Ma) using subsampling methods and completeness metrics to correct for/highlight sampling and preservation biases. Additionally, I used generalized least-squares regressions to study the relationships between taxon richness and other potential explanatory variables (e.g. temperature, sampling proxies).

The lepidosaur record is more poorly sampled than the turtle record, but for both clades the Southern Hemisphere is clearly undersampled compared to the Northern Hemisphere. Consequently, the majority of sequential richness signal is observable in Northern Hemisphere continents only.

Lepidosaur and turtle richness patterns differed prior to the Eocene, but were very similar afterwards. Across the K-Pg boundary, North American lepidosaurs decreased in richness while turtles increased. Both clades exhibited high richness in the Late Cretaceous and in the early and late Eocene, often corresponding to times of very warm global climates. During these times richness was greater at higher latitudes than at lower latitudes, which is the opposite of the modern latitudinal biodiversity gradient. Richness decreased at the Eocene–Oligocene boundary in the Northern Hemisphere, coincident with a steep drop in global temperatures, but remained relatively high compared to the Mesozoic. It is unclear whether taxa with lower cold tolerance migrated south at these times or went extinct, due to poor sampling in the tropics.

Ectotherm taxon richness through time is strongly influenced by sampling biases, including the number of collections available and the prevalence of Lagerstätten (localities of exceptional preservation). Temperature might have influenced lepidosaur richness, but other factors not tested for might have also been important for ectotherm richness, such as precipitation (particularly for freshwater turtles).

Impact Statement

The work done in this thesis adds to the wealth of literature on the effects of biases on the fossil record, and how we can account for and ameliorate these biases. Prior to this thesis there had been few studies on the fossil records of turtles, and almost none on lepidosaurs, that accounted for these biases when discussing richness through time. My results provide a useful comparison to uncorrected face-value observations of richness in the literature, and a cautionary tale against interpreting richness changes based on poorly sampled areas or time periods. I have also added non-marine lepidosaur and turtle richness studies to the pool of comparative results for future researchers who wish to try the same methods on other clades; I edited R code included in the appendices that could be very useful to others who want to do this. One chapter of this thesis is already published (Chapter 2), and I aim to publish the rest in the future to disseminate my results more widely. I will also speak (and have spoken previously) at conferences about this work, creating opportunities for future collaborations based on the work I have done here.

Outside of my field, my results may provide a useful deep-time comparison for those examining the predicted responses of ectotherms to present anthropogenic climate change. Ectotherms are expected to respond strongly to changes in temperature, and researchers often examine past occurrences of rapid climate change in order to learn about potential future responses of organisms. Having a record of long-term richness that has been corrected for the majority of sampling biases may provide a very useful resource for these researchers. Alongside journal articles and conference talks, I have participated in public engagement and outreach activities where I have discussed my work in this light, so this aspect of my research is very easy to disseminate outside of academia. In addition, the knowledge gained with regards to how past ectotherms responded to climatic changes, and how this relates to the current situation, could be applied to reports which are used to influence government policy.

Acknowledgements

There are many people, both inside academia and outside without whom I would have been unable to reach the end of this journey. First and foremost, I have been fortunate to have three incredible supervisors: Paul Barrett, Susan Evans and Roger Benson, who supported me both in my research and in my development as an independent researcher. You three are the kind of people I aspire to be in science, and I hope I can support students of my own even half as well someday. Elsewhere in academia, I would like to thank Phil Mannion and Pat Holroyd for their very helpful discussions of aspects of this thesis. David Nicholson's work on turtles prior to mine inspired some of this thesis, and he was happy to help me understand how all the methods worked when I was first starting out. Paul Valdes provided Cenozoic land area data that was invaluable to one chapter of this thesis. Richard Butler was always on hand in the early days of my entering data into the Paleobiology Database, and I am very thankful for every answer he readily gave to my (many) questions. Everyone I interacted with while editing the PBDB was incredibly gracious, and I am thankful to all of them.

Outside of academia, there are too many people to name. Most of them know how much I appreciate them, so I will keep this short. Special thanks go to my ever-supportive family, my flatmate (who put up with my nerdy pursuits and took pity on my lowly student status by treating me to dinner more than a few times), and my office-mates who shared in my frustrations and my triumphs. I made it.

[This page intentionally left blank]

Table of Contents

[This page intentionally left blank]

List of Tables

[This page intentionally left blank]

List of Figures

Table of Contents for Electronic Appendices

Appendix 1: Miscellaneous files for R scripts used in multiple chapters:

- **Appendix 1.1:** List of marine tetrapod names (“Marine terms”).
- **Appendix 1.2:** List of ootaxon names (“Egg terms”).
- **Appendix 1.3:** List of ichnotaxon names (“Trace terms”).
- **Appendix 1.4:** Time bins used to make PBDB data consistent (“Time bins – stages”).

Appendix 2: Data, scripts and files used in Chapter 2:

- **Appendix 2.1:** Data used in Chapter 2, original PBDB download and references, major contributors; GLS data and results.
- **Appendix 2.2:** Script for filtering PBDB download of lepidosaurs, and processing before and after SQS.

Appendix 3: Script and files used in SQS analyses for all chapters that use it:

- **Appendix 3.1:** Perl script used for running SQS (“quorumSubsample4-3.pl”).
- **Appendix 3.2:** Time bins used in SQS analyses in Chapters 2, 3 and 5 (“Time bins”).
- **Appendix 3.3:** Time bins used in SQS analyses in Chapter 4 (“Time bins (lats)”).

Appendix 4: Non-marine tetrapod data used in all chapters:

Appendix 4.1: Non-marine tetrapod original PBDB download, references and major contributors.

Appendix 4.2: Script for filtering PBDB download of tetrapods, and compiling tetrapod-bearing formations and collections for all chapters except Chapter 4.

Appendix 5: Data and script for GLS analyses using $\delta^{18}O$ data from Prokoph (2008):

Appendix 5.1: R script for interpolating data from Prokoph.

Appendix 5.2: Data from Prokoph before and after interpolation.

Appendix 6: R script for running GLS analyses in Chapters 2 and 3.

Appendix 7: Data, scripts and files used in Chapter 3:

Appendix 7.1: Data used in Chapter 3, original PBDB download and references, major contributors; GLS data and results.

Appendix 7.2: Script for filtering PBDB download of turtles, and processing before and after SQS.

Appendix 7.3: File for excluding turtle occurrences from tetrapod group in plot of latitudinal distribution of tetrapods versus turtles (Fig. 3.9): ("turt exclude").

Appendix 7.4: File for excluding turtle occurrences from tetrapod group in plot of latitudinal distribution of tetrapods versus turtles (Fig. 3.9): ("turt exclude2").

Appendix 8: Data and scripts used in Chapter 4:

Appendix 8.1: Data and results from GLS analyses of lepidosaurs and turtles.

Appendix 8.2: Land area calculations for GLS analyses.

Appendix 8.3: Scripts for processing lepidosaur data from Ch. 2 into latitudinal bins:

Appendix 8.3.1: R script for global lepidosaur latitudes.

Appendix 8.3.2: R script for European lepidosaur latitudes.

Appendix 8.3.3: R script for North American lepidosaur latitudes.

Appendix 8.4: Scripts for processing turtle data from Ch. 2 into latitudinal bins:

Appendix 8.4.1: R script for global turtle latitudes.

Appendix 8.4.2: R script for European turtle latitudes.

Appendix 8.4.3: R script for North American turtle latitudes.

Appendix 8.5: Script for processing tetrapod data into latitudinal bins of collections and formations.

Appendix 8.6: Script for running GLS analyses on latitudinal data.

Appendix 9: Data, script and file used in Chapter 5:

Appendix 9.1: Data on completeness including metric, explanation of abbreviations used in dataset, references; GLS data and results.

Appendix 9.2: Script for running GLS on completeness data and species-level SQS analyses.

Appendix 9.3: File for excluding snake and amphisbaenian taxa from lepidosaur data (“snakesamphis”).

Appendix 9.4: Script for running z-tests on Mesozoic VS Paleogene completeness data

1. Introduction

1.1 Accounting for geological and anthropogenic biases in the fossil record

1.1.1 Background

Scientists have long been interested in attempting to reconstruct patterns of taxon richness through deep time, and how they fluctuated and responded to past climatic, biotic and environmental events, as well as the influence of extinctions and evolutionary radiations. Investigations into fossil diversity began with curves constructed from ‘raw’ data, i.e. face-value counts of taxa in designated time bins, and attempts to highlight the inherent biases in the fossil record (Raup, 1972, 1976a, 1976b; Raup and Sepkoski, 1982; Sepkoski, Jr., 1993; Smith, 2001).

It is unlikely that all taxa that have ever existed will have become fossilised, and of these fossils not all will be discovered. A number of filters exist between an organism that lived in the past and our finding their remains; these filters can bias and distort the record that we recover and thereby shift and modify observed past richness away from ‘true’ richness. Filters can include:

1) Inherent preservability: some taxa will be more favourably preserved than others due to physical traits (e.g. large body size, possession of a calcified skeleton) that aid their survival through taphonomic processes (e.g. Raup, 1972; Forey et al., 2004; Brown et al., 2013). This is linked to:

2) Lagerstätten: groups of organisms with lower inherent preservability are found more frequently in areas of exceptional preservation (“konservat

[conservation] Lagerstätten"; Seilacher, Reif and Westphal, 1985; Allison, 1988). These areas, often created by special local environmental circumstances during deposition (e.g. the anoxic shallow lagoonal localities hypothesised to be prevalent during the deposition of the Late Jurassic lithographic limestones in the Solnhofen region; e.g. Munnecke, Westphal and Kölbl-ebert, 2008) are non-randomly distributed and can preserve organisms that are difficult to find elsewhere. As a result they can overinflate comparative richness estimates and increase sampling unevenness. This can influence:

3) Collection biases: substantial collection focus on certain geographic areas (particularly if those areas contain Lagerstätten) or on certain time periods affects sampling evenness. This can overinflate long-term richness curves, particularly around notable events such as mass extinctions, where scientists might sample more intensely. With many tetrapod groups, for example, there has been a substantial focus on North America and Western Europe, the areas with the highest numbers of palaeontologists. By contrast, the Southern Hemisphere has been more poorly studied due to a lower density of workers. Additionally, some organisms are more difficult to collect than others, and so are sometimes overlooked; microvertebrates, for example, often require screen-washing of sediments to separate out small bones (Blain et al., 2010).

4) Geological biases: the rock record is unevenly distributed and this controls the number of opportunities that palaeontologists have to sample their fossil content. Many fossiliferous rocks are eroded away before we can sample them,

and accessibility to those remaining is determined by weathering and uplifting processes. Rocks of some time periods have limited exposure at the surface, and are distributed unevenly, which leads to geographically uneven richness estimates within time periods. In broad terms, the availability of more rock should increase the chances of finding new taxa, and so the rock and fossil richness records should track each other to a certain extent (e.g. Peters and Heim, 2011; Smith and Benson, 2013; Dunhill, Hannisdal and Benton, 2014).

5) The Signor-Lipps Effect: the first and last occurrences of a taxon are unlikely to be recorded (Signor and Lipps, 1982). When estimating richness levels through time we must take this into account as it affects the timings of originations and extinctions, and may confuse the speed and effects of mass extinction events.

6) The Pull of the Recent: younger parts of the fossil record are likely to be over-sampled due to the greater availability of rock (as there has been less time for erosion to remove them). Additionally, extant taxa can "pull" forward the ranges of related fossil taxa (e.g. families or genera), even if known occurrences of those taxa are temporally far apart (Raup, 1972).

7) Literature biases: novel taxa are more likely to be published than new records of established taxa. This creates a disproportionate view of richness and creates situations where only one occurrence of a taxon is known to the general scientific community (and to aggregation databases such as the Paleobiology Database) despite many specimens existing in museum

collections (so-called “dark data”; Marshall et al., 2018). There is also the added error of many new species being named from fragmentary remains, which can overinflate diversity (Alroy, 2010a).

As all of these biases have the potential to distort our reading of the fossil record, a number of methods have been developed to over the past few decades in order to identify or alleviate some of these issues. They can be separated into a number of categories, which are described below.

1.1.2 Measuring preservation biases

Some methods examine the preservation of individual specimens or taxa, in order to assess the preservational completeness of the fossil record, and are commonly referred to as “completeness metrics”. This information can be used to identify areas where low preservation potential (or high incidence of incomplete preservation) and low richness could be correlated. Incomplete preservation obscures information in fossil specimens, which can lead to an inability to identify them to lower taxonomic levels and thereby excludes these specimens from inclusion in richness estimates, even if they represent undocumented species. A number of different completeness metrics have been employed in the past, including simple metrics that assess overall completeness of specimens on a short numeric scale (Benton, 2008b), to those that generate a more robust measure of completeness by splitting the body into sections and assessing the percentage or categorical completeness of each section (e.g. Mannion and Upchurch, 2010; Beardmore et al., 2012). The Character Completeness Metric, alternatively, examines the number of phylogenetic characters that can be coded from each region of a skeleton (assuming that poorly preserved skeletons have a more limited amount of information that can be gleaned from them), and assign completeness percentages this way (Mannion and Upchurch, 2010). Completeness metrics have been used to explore the fossil records of various taxa, including dinosaurs, marine reptiles and echinoderms (Smith, 2007; Mannion and Upchurch, 2010; Cleary et al., 2015; Driscoll et al., 2018). The practicality of these metrics vary by taxa, with some easier to categorise in this way than others (e.g. taxa with conservative body plans are easier to score consistently than highly disparate clades). For a more detailed review of these metrics, see Chapter 5.

1.1.3 Modelling approaches to biases

Models are created to assess the fit of sampling proxies (such as the area of available rock outcrop) to taxon richness data. The most common method using this approach takes raw taxonomic richness and a proxy for either a geological or anthropogenic sampling bias to create a model where raw richness is perfectly predicted by the sampling proxy and then deviations from this model are examined to extract 'true' diversity signals. This method was first proposed by Raup (e.g. Raup, 1976b) and was developed further into what has become known as 'the residual diversity method' by Smith and McGowan (2007) and Lloyd (2012). Regression analyses are often used with multiple proxies to detect which one explains the richness data best. Commonly used proxies are the number of taxon-bearing formations and taxon-bearing collections, as representations of rock record bias and human sampling bias, respectively (e.g. Barrett, McGowan and Page, 2009). A linear model is fitted to ordered richness and proxy data, and used to calculate modelled richness. These values are subtracted from original raw richness to leave the 'residuals', the underlying richness that is not explained by the sampling proxy (Smith and McGowan, 2007). Residuals can be compared with a raw richness time series to outline where the two differ, and also with other factors (e.g. sea level) to detect evolutionarily meaningful patterns. This method has been employed by numerous authors (e.g. Barrett, McGowan and Page, 2009; Benson et al., 2010; Mannion et al., 2011; Smith, Lloyd and McGowan, 2012; Pearson et al., 2013; Newham et al., 2014; Grossnickle and Newham, 2016).

In an extension of this method, Lloyd (2012) introduced several improvements that address some of the issues that arose from a simple linear model approach. First, a number of different models are fitted to the data, including linear, exponential, sigmoidal, and polynomial models. The best one is chosen through calculation of the Akaike Information Criterion (AIC), which assesses goodness of fit against the complexity of a model. Standard deviations are stored for the chosen model. Residuals are produced as usual, but 95% confidence intervals are generated using the stored deviation data, in order to provide a measure of significant excursions from the model, which the original method (Smith and McGowan, 2007) could not provide. Model-based approaches are less demanding of large sample sizes, and tend to be flexible in the number and types of biases that can be considered (Lloyd, 2012). The use of various model types can help to emphasise a shared result, or help to reveal flaws when they are in conflict, and provides a level of comparative analysis that cannot be found with subsampling.

While models can ‘correct’ richness data to a certain degree, residuals may not represent ‘true’ richness fluctuations. There may be other bias signals affecting the observed richness that has not been accounted for by the model (Smith and McGowan, 2007), and so the degree to which any remaining signal is biological is debated (Lloyd, 2012). The Common Cause explanation (Peters, 2005; Peters and Heim, 2011), for example, has some sampling biases and observed richness driven by a third biologically-related factor (a ‘common cause’), and therefore ‘correcting’ the bias removes part of the biological signal as well. It is sometimes difficult to discern whether Common Cause or bias is impacting the pattern observed in richness data, as these two different mechanisms produce very similar effects. However, some studies have developed methods to address this (e.g. Butler et al., 2011; Hannisdal and Peters, 2011; Smith and Benson, 2013), such as comparing richness and rock records to marine carbonate isotopes and estimates of sea level. Common cause appears to influence shallow marine assemblages during parts of the sea level cycle (Benson and Butler, 2011), but may be a less important factor on land (Butler et al., 2011).

Residuals can be affected by the measure of proxy used; for example, methods of measuring rock outcrop area can differ greatly (e.g. mapped area versus actual rock exposure area; Benton et al., 2011), and so the degree to which one is 'correcting' richness can vary. They can also be influenced by some factors discussed above, such as the presence of Lagerstätten (e.g. Benson et al., 2010). Steps can be taken to account for the latter's spurious inflations of richness, such as the removal of formations that produce >50% of diversity in a time bin (Lloyd, 2012), but this also removes a substantial portion of the data (that may contain a real signal) rather than fixing the issue. It is also possible to include the presence/absence of Lagerstätten (and their frequency per time bin) as an additional variable in models (Benson and Butler, 2011; Butler, Benson and Barrett, 2013).

Recently, palaeontologists have been moving away from the use of the residuals method. Several papers have suggested that the method produces wholly unreliable results (Sakamoto, Venditti and Benton, 2017; Dunhill et al., 2018) and should not be used with taxon richness estimates. They highlight problems with independent sorting and that creating residual corrections based on a false linear relationship will lead to spurious predictions.

As an alternative, the use of multivariate models is a more recent innovation that examines multiple proxies simultaneously, and so may offer a more robust method of accounting for biases in raw richness than residuals. One version of these methods, generalised least-squares (GLS), can address issues of autocorrelation and quantitatively assess the effects of adding additional variables to a model. As with the residuals method, AIC values are used to identify the best fitting models, with the best combination of proxies, but also give a lower weighting to models with a higher number of variables. GLS has been applied in numerous studies (e.g. Hunt, Cronin and Roy, 2005; Marx and Uhen, 2010; Benson and Butler, 2011; Benson and Mannion, 2012; Smith and Benson, 2013; Dean, Mannion and Butler, 2016; Nicholson et al., 2016) and provide a great comparison to subsampling techniques when assessing palaeodiversity. These are discussed in more detail in later chapters.

1.1.4 Subsampling approaches to biases

Sampling standardisation, or subsampling, generates sampling-corrected richness counts via the random drawing of data from time or spatial intervals (palaeodiversity studies tend to use the former, so time is what I will refer to from here), to retrieve an even or fair sample (Alroy, 2010b). They tend to handle richness data alone, without examination of potential driving factors (e.g. rock outcrop availability). There are many different subsampling methods; here I will discuss two frequently-used techniques that differ in their method of "fair" sampling: rarefaction (AKA "classical rarefaction") and shareholder quorum subsampling.

Rarefaction arose first, and was developed to address the problem of how to compare samples of different sizes directly (Sanders, 1968). A dataset is divided into time intervals and then a uniform sampling quota is set, which is a benchmark number of items that must be drawn in each interval. Items are sampled until the quota is met for each interval, and this is done repeatedly so an average can be taken (Alroy, 2010a). “Items” can be either individual specimen/taxon occurrences (e.g. Miller and Foote, 1996; Fastovsky et al., 2004) or whole collections of taxa (e.g. Alroy, 1996; Alroy et al., 2001). A rarefaction curve is plotted, showing the number of taxa against the sample size, and is used to extrapolate the number of taxa one would discover in smaller samples with an identical species composition (Sanders, 1968; Miller and Foote, 1996). It is not possible to extrapolate for larger samples, however, so there are constraints on minimum sample size (Mannion et al., 2011).

Rarefaction curves are weak with respect to subdivision of datasets, and also tend to produce increasingly flatter curves as the sampling quota decreases (Alroy, 2010b). This method is easily influenced by dominant taxa; if a taxon is particularly abundant within an assemblage, it will likely be oversampled compared to rarer taxa and risk “dampening” the resultant taxic-richness curve (Alroy, 2010b). Rarefaction makes the assumption that uniform sampling quotas equate to fair sampling (Alroy, 2010a), which would require organism abundances to remain constant through time when they obviously do not (Alroy, 2010b; Benton et al., 2011). If a species pool increases the chances of discovering any one taxon from a single item draw will decrease (Alroy, 2010b), and vice versa. Because rarefaction does not account for the varying abundance of taxa through time, it has limited use as a method of correcting diversity counts, although it was widely applied prior to the exposition of these criticisms.

Shareholder quorum subsampling (SQS) was introduced to try and account for varying species pool size and in an attempt to represent rarer taxa (Alroy, 2010c). The technique is the same as rarefaction initially, but differs in its method of drawing samples to avoid the drawbacks of uniform sampling (Alroy, 2010b). As Alroy explains, it is incorrect to assume that uniform sampling equates to fair sampling. If you have a uniform quota for each time bin, you over- or undersample when the size of a species pool changes via immigration or extinction; SQS tries to account for this by increasing sampling intensity when species richness increases.

Furthermore, instead of using fixed counts of data items as a sampling quota, SQS uses the "coverage" of taxon frequencies sampled in a time interval (Alroy, 2010b). Frequencies are obtained by working out the proportion of occurrences that belong to each taxon. To demonstrate, imagine a time interval with various taxa. Species A has a frequency of 0.2, Species B= 0.1, and C= 0.05. A collection containing all three would have a coverage of 35% for that time interval. Instead of drawing taxa until a counting quota is met, sampling is carried out until a defined coverage is met. This coverage target is a "shareholder quorum", with taxa as the "shareholders". SQS aims to correct for unknown taxa by oversampling where coverage is low, since uniform sampling is not fair sampling. As the degree of undersampling varies depending on the richness of each interval, we modify the 'observed' coverage with an estimator, Good's u (Good, 1953). This compares the number of taxa that occur only once to the total occurrences in a bin, using $1 - n1/O$ ($n1$= no. of single-occurrence taxa; O= total no. of occurrences). Time bins with a high number of single-occurrence taxa and thus a low u value are subsampled more intensely to better capture underlying richness. If u is lower than the chosen quorum for an analysis, however, then that time bin will be excluded from analysis as it is too poorly sampled. Unlike rarefaction, SQS can produce the same relative diversity estimate regardless of the quota level or subdivision of dataset. SQS has been used in a large number of studies since its creation (e.g. Mannion et al., 2012, 2015; Nicholson et al., 2015; Benson et al., 2016; Grossnickle and Newham, 2016; Tennant, Mannion and Upchurch, 2016; Davies et al., 2017; Dunne et al., 2018), and its use is still widespread in palaeontology.

SQS is not an unbiased estimator of diversity, however, since coverage is still an estimate of the available fossil record; it is relative, rather than absolute, richness (Starrfelt and Liow, 2016). Furthermore, SQS assumes species additions and subtractions are random, which is not always the case when ecologically dominant taxa are involved (Alroy, 2010a). To counteract this, the most common taxon can be ignored in frequency calculations, but included in diversity counts. When additions/subtractions are of rare species, however, coverage may be underestimated (Alroy, 2010b). Rarer taxa are more fairly represented than when rarefaction is employed, but SQS still retrieves "most of the common taxa and a stochastic assortment of the rare ones" (Alroy, 2010c), meaning that it still probably fails to capture the entire diversity signal in a time interval. It may also struggle with very small sample sizes (Benson and Mannion, 2012).

An additional sampling standardisation method, recently proposed as a possible alternative to SQS, is the "True Richness estimated using a Poisson Sampling" model (TRiPS; Starrfelt and Liow, 2016). This is a parametric approach to sampling standardisation (opposed to SQS's non-parametric approach) that models sampling as a Poisson process to calculate maximum-likelihood estimates of 'true' underlying richness (Close et al., 2017). TRiPS infers a single sampling rate for all taxa in a time bin, then uses that sampling rate to estimate a per-lineage binomial probability per time bin (i.e. the probability a taxon would be sampled in that time bin). This binomial probability is used to derive the most likely 'true' richness curve (Starrfelt and Liow, 2016). Due to its recent invention, it has yet to be applied in many studies, but has been tested alongside SQS in some (e.g. Brocklehurst et al., 2017; Close et al., 2017, 2018).

Unfortunately TRiPS is prone to influence from the prevalence of very common taxa, as it fits a single sampling rate for all taxa in a bin (Starrfelt and Liow, 2016). A common taxon can overinflate the binomial sampling probability for a time bin, causing it to saturate prematurely. This results in TRiPS ceasing to extrapolate richness while many taxa remain undiscovered, producing an underestimate of past richness for that time bin. Additionally, if used with samples of unequal sizes (as occurs frequently with data derived from the fossil record), richness estimated with TRiPS may be as misleading as uncorrected taxon richness curves (Close et al., 2018).

SQS and other coverage-based estimators of richness appear to perform more consistently (producing more stable estimates) than extrapolator-type methods like TRiPS when face-value richness counts fluctuate greatly. All methods of estimating richness, however, are affected by differences in sampling evenness (i.e. geographical sampling biases). As evenness decreases, it becomes harder to detect rare taxa and richness estimates become downwardly biased (Close et al., 2018).

There will never be a perfect method to estimate past richness, as unknown data that cannot be accounted for even with the most robust statistical methods will always exist, but coverage-based estimators appear to be one of the best methods that we have currently for examining richness in deep time. It is important to use them to explore the effects of biases on our observations of past richness to obtain the best estimates of richness possible, even if we know that ‘corrected’ richness is still likely not perfect ‘true’ richness.

1.2 Fossil record biases and ectotherms

Many studies have examined biases in the fossil record using the above methods, but few have examined the record of non-marine ectothermic reptiles in detail. Understanding patterns of ectothermic reptile taxic richness through time is relevant to understanding the origins of their exceptional extant richness. This is especially true for lepidosaurs, consisting of Squamata (lizards, snakes, amphisbaenians) and Rhynchocephalia (the extant tuatara plus many extinct forms), which are the most speciose extant reptile group (>10,000 species; Uetz, Freed and Hošek, 2018). Ectotherms have important roles today as indicators of global change, so it is also important to understand how past richness responded to climatic changes given current environmental change, as many studies indicate that ectothermic taxa in particular will be heavily affected by ongoing climatic warming (e.g. Araújo, Thuiller and Pearson, 2006; Bickford et al., 2010; Foufopoulos, Kilpatrick and Ives, 2011; Moreno-Rueda et al., 2012; Urban, Richardson and Freidenfelds, 2014).

While non-marine crocodylians have been examined extensively (e.g. Hutchison, 1982; Brochu, 2003; Toljagic and Butler, 2013; Bronzati, Montefeltro and Langer, 2015), and their fossil record has been examined using some of the methods mentioned above (Mannion et al., 2015; Jonathan P Tennant, Mannion and Upchurch, 2016), non-marine lepidosaurs and turtles (Testudinata) have received less attention.

Numerous studies have examined changes in species richness in terrestrial lepidosaur communities around key events in Earth history such as mass extinctions, but most of these have concentrated on the underlying causes of taxic richness change, rather than analysing the pattern of change itself (e.g. Böhme, 2007; Rage, 2012; Apesteguía, Gómez and Rougier, 2014). While lepidosaur taxon richness has been examined in past studies, for either small regions (e.g. Alifanov, 2000; Jones, 2006; Albino, 2011), whole continents (e.g. Augé, 2005; Nydam, 2013; Rage, 2013) or globally (Evans, 1998, 2003), few have used methods that account for the inherent biases in the fossil record.

Similarly, many studies have examined aspects of turtle evolutionary history, with some focusing on important time periods such as mass extinction events or times of dramatic climatic change (e.g. Hutchison and Archibald, 1986; Hutchison, 1992; Hirayama, Brinkman and Danilov, 2000; Holroyd, Hutchison and Strait, 2001; Corsini, Smith and Leite, 2011; Holroyd, Wilson and Hutchison, 2014; Nicholson et al., 2015, 2016). Most of these studies have focused on North America, where there is a long-term accessible rock record, particularly in the western and central USA and western Canada. The vast majority of these are based on face-value interpretations of the fossil record and, though they acknowledge that sampling biases affect these observations, few employ methods to correct for these biases. Although some wider studies have included non-marine turtles in their estimates of changing richness through time, and used methods that alleviate various biases in the fossil record, these have focused on non-marine tetrapods as a whole, including turtles but not examining their richness in isolation or detail (Kalmar and Currie, 2010a; Benson et al., 2013, 2016). The exception to this are two papers by Nicholson et al. (2015, 2016), who used subsampling to investigate the richness of marine and non-marine turtle genera through the Mesozoic Era.

Using some of the methods described above, I will investigate biases in the ectotherm fossil record from their first appearances in the Triassic until the end of the Paleogene, focusing on non-marine lepidosaurs and turtles. This will partially be expanding upon work done by Nicholson et al. (2015, 2016) on Mesozoic turtle richness. Analyses using sampling-correction methods have been carried out on Mesozoic–Recent crocodylians (Mannion et al., 2015) which will provide a third ectotherm clade for comparison to my data. For the purposes of this thesis, the term "ectotherm" excludes non-avian dinosaurs, as they represent a more complicated clade whose thermoregulatory methods are heavily debated (e.g. Bakker, 1972; Seebacher, 2003; Köhler et al., 2012; Werner and Griebeler, 2014).

The specific aims and objectives of this thesis are as follows:

- To construct sampling-corrected richness curves for non-marine lepidosaurs and turtles from the Triassic–Paleogene and to compare these curves to uncorrected face-value observations of richness and to each other. This will address two fundamental questions: 1) what are the major patterns in the taxon richness of these clades through time?; and 2) are there common patterns in ectothermic vertebrate taxon richness through time?

- I will investigate the latitudinal biodiversity gradients (LBGs) of extinct lepidosaurs and turtles to test whether their modern day LBGs (characterised by high taxon richness in the tropics, with decreasing richness towards the poles) also existed in the past. This will address whether it is possible to measure reliably past LBGs despite biases in the record and, if possible, will form the basis for examining this major macroecological pattern over extended temporal scales.

- The quality of the lepidosaur fossil record will be evaluated to determine if preservation biases have had an effect on the number of identifiable taxa (and thus lower-level taxonomic richness). How applicable are existing completeness metrics to lepidosaurs, and how informative are they when being used to assess lepidosaur fossil record biases? How does the lepidosaur fossil record compare with those of other extinct tetrapods?

- Finally, I will assess whether Triassic–Paleogene non-marine ectothermic tetrapods had similar patterns of richness through time that could indicate potential shared responses to climatic, environmental or biotic change or, if not, propose alternative reasons that might have generated the observed patterns.

2. Non-marine Lepidosaur Richness from the Triassic–Paleogene

2.1 Introduction

Lepidosaurs, consisting of Squamata (lizards, snakes, amphisbaenians) and Rhynchocephalia (the extant tuatara plus many extinct forms), are a highly successful modern clade with over 10,000 extant species (Uetz, Freed and Hošek, 2018) and a long evolutionary history spanning back to the Middle Triassic (Jones et al., 2013; Simões et al., 2018). Lepidosaurs throughout time vary greatly in traits such as body size, from tiny extant geckos such as *Sphaerodactylus* (<20 mm; Meiri, 2008) to the extinct mosasaurs (up to approx. 17 m in body length, although this is likely overestimated; Grigoriev, 2014). Members of the clade have evolved specialisations for marine, fossorial and arboreal life, and exhibit wide variation in diet, behaviour and habitat. Historically, squamates acquired several features that have been regarded as key adaptations thought to contribute to their success, including limblessness (Gans, 1975; Wiens, Brandley and Reeder, 2006), cranial kinesis (Frazzetta, 1962) and venom-delivery apparatus (Fry et al., 2006). They display varying methods of reproduction (Shine, 1983), including viviparity, oviparity, ovoviviparity and even parthenogenesis in some taxa (Kearney, Fujita and Ridenour, 2009).

There is continued debate over the timing of lepidosaur origins, but the confirmed presence of closely related groups (archosauromorphs), as well as evidence from molecular phylogenies suggests that it must have occurred by the Late Permian (e.g. Gorr, Mable and Kleinschmidt, 1998; Wiens, Brandley and Reeder, 2006; Hugall, Foster and Lee, 2007; Albert et al., 2009; Pyron, 2010; Mulcahy et al., 2012). Despite this, the fossil record currently indicates that the major radiation of the clade occurred much later. The earliest known rhynchocephalian fossils are from the Ladinian of the Northern Hemisphere (Fraser and Benton, 1989; Sues and Olsen, 1990; Sues and Hopson, 2010; Jones et al., 2013), but the absence of any record prior to the Middle Triassic obscures their geographic origins.

A large temporal gap previously separated the earliest rhynchocephalian fossils (Jones et al., 2013) from those of the earliest known squamates, limiting understanding of their early diversification, but this has potentially been resolved in a recent paper (Simões et al., 2018). Prior to this the earliest squamate fossils were from the Early or Middle Jurassic of India (Evans, Prasad and Manhas, 2002; though there is some uncertainty in the dating of these remains, see Prasad and Manhas, 2007), and the Bathonian (Middle Jurassic) of the UK (Evans, 1994) and Central Asia (Nessov, 1988). A revision of the Middle Triassic lepidosauromorph *Megachirella wachtleri* by Simões et al. (2018) reidentified it as a stem squamate, shifting the earliest known squamate back 75 million years to the Anisian of Italy. Whether this identification is verified by others remains to be seen, but the occurrence of both rhynchocephalians and squamates in the Middle Triassic is consistent with their status as sister taxa that should have originated around the same time.

Unequivocal snake fossils first appear in the Early Cretaceous (Cuny et al., 1990; Gardner and Cifelli, 1999; Martill, Tischlinger and Longrich, 2015) and might have had a Gondwanan origin given their wider distribution and higher species-richness in this region during their early evolution (Evans, 2003; Evans and Jones, 2010), although there is disagreement on the identification of older specimens from the Middle Jurassic as stem-group snakes (Caldwell et al., 2015). Debate is ongoing about whether their origins were terrestrial/fossorial (Vidal and Hedges, 2004;

Longrich, Bhullar and Gauthier, 2012a; Hsiang et al., 2015; Martill, Tischlinger and Longrich, 2015; Da Silva et al., 2018) or marine (Caldwell and Lee, 1997; Caprette et al., 2004; Lee, 2005; Lee et al., 2016). Amphisbaenians, which are entirely fossorial, first appear in the early Paleocene of Belgium (Folie, Smith and Smith, 2013) and the USA (Sullivan, 1985) and the late Paleocene of Africa (Augé and Rage, 2006), though there are disputed Late Cretaceous finds (Blain et al., 2010). Their fossil record is extremely sparse, and little is known about how they diversified, with modern species-richness concentrated in Africa and South America (Gans, 1967; Augé, 2012).

Although rhynchocephalians were abundant during the early history of lepidosaurs, their abundance and geographic distribution began to contract after the Early Jurassic. They made their final appearance in Asia during the Sinemurian despite the presence of younger microvertebrate localities that should be suitable for their preservation (Wu, 1994; Evans, 1998; Evans and Jones, 2010; Benson et al., 2013). They subsequently continued to decline in the Northern Hemisphere, culminating in their last appearances in North America and Europe during the Early Cretaceous (Evans, 1998; Evans, Prasad and Manhas, 2001; Nydam and Cifelli, 2002). Although they are present in some Gondwanan localities in the Late Cretaceous, the Cenozoic fossil record of rhynchocephalians is sparse and, except for one occurrence in the Paleocene of Argentina (Apesteguía, Gómez and Rougier, 2014), confined to the area in which they are found currently, New Zealand (Jones et al., 2009). Squamates, by contrast, expanded their ranges from their first appearances in the Triassic and Jurassic of Europe and Asia to become near-cosmopolitan in the modern day, except in the coldest regions of the world, and are key organisms within the terrestrial ecosystems they inhabit (Pianka and Vitt, 2003).

As mentioned in Chapter 1, numerous studies have examined changes in species richness in terrestrial lepidosaur communities around key events in Earth history such as mass extinctions, but most of these have concentrated on the underlying causes of taxic richness change, rather than analysing the pattern of change itself (e.g. Böhme, 2007; Rage, 2012; Apesteguía, Gómez and Rougier, 2014). Some have reviewed large portions of the lepidosaur fossil record through time, for either small regions (e.g. Alifanov, 2000b; Jones, 2006; Albino, 2011), whole continents (e.g. Augé, 2005; Nydam, 2013; Rage, 2013) or globally (Evans, 1998, 2003), but these have been based primarily on qualitative observations and direct readings of the fossil record. Few of the above-mentioned studies on the lepidosaur fossil record mention biases that affect the face-value fossil record, and even fewer employ methods to account for these biases, with the exception of two recent studies on marine tetrapod diversity that include mosasaurs (Benson et al., 2010; Benson and Butler, 2011).

Beyond lepidosaur-focused research, there are now many studies that have examined the richness of the tetrapod (e.g. Kalmar and Currie, 2010; Benson et al., 2013, 2016; Close et al., 2017, 2018) and the marine invertebrate (e.g. Sepkoski, Jr., 1993; Foote, 2000; Alroy et al., 2001) fossil records with regards to sampling biases in the rock record. Whilst results differ, the majority of these larger-scale studies recover some significant correlations between various proxies representing sampling and/or geological biases and palaeobiodiversity, particularly for proxies representing rock availability or worker effort.

Many methods have been developed to untangle geological and sampling biases from underlying ‘true’ richness, but few have been used on the non-marine lepidosaur record. In this chapter I examine the taxon richness of non-marine lepidosaurs from the Triassic–Paleogene (252–23 Ma), constructing sampling-corrected taxon-richness curves from an expanded dataset of their occurrences through time. I discuss how solid the basis is for unravelling lepidosaur taxic richness through time and discuss some of the driving factors that might generate these patterns, and note the problems still associated with the lepidosaur fossil record and the fossil record as a whole.

2.2 Methods

2.2.1 Data

Triassic–Paleogene occurrences of lepidosaurs were downloaded from the Paleobiology Database (PBDB; www.paleobiodb.org), accessed via Fossilworks (www.fossilworks.org) on 27 September 2016, following extensive updating of the taxonomy and occurrences of those taxa already entered (resulting in more than 500 changes to the previously published data) and the addition of more than 1000 new occurrences after an extensive review of the literature. The relative contributions of different researchers to this dataset are provided in Appendix 2. I filtered the data in R (see Appendix 2 for script) following download to remove trace fossils, marine taxa, and taxonomically indeterminate records, and to correct inconsistencies in the dating of time bins; the remaining dataset consists of 1971 occurrences representing 449 genera. I used genera in my analyses because of the problems associated with using species counts in the fossil record (e.g. preservation difficulties preventing identification to species level, uneven taxonomic treatments: Robeck, Maley and Donoghue, 2000; Krug et al., 2009). Of course there are still potential problems with wastebasket taxa even at the generic level, but this is the highest one can go and still study macroevolutionary patterns in detail, as the signal is too coarse at higher taxonomic levels (Krug et al., 2009). Using genera allowed for the inclusion of occurrences that were indeterminate at species level, as these comprise a geographically widespread wealth of data (>650 occurrences) that would otherwise be excluded.

2.2.2 Subsampling

In order to account for potential sampling biases in the fossil record, I used Shareholder Quorum Subsampling (SQS; Alroy, 2010a, 2010b) to generate a subsampled richness estimate curve for lepidosaurs, which I compared to the curve of uncorrected face-value genus counts. As explained in Chapter 1, SQS uses taxon frequencies calculated from the proportion of occurrences that are represented by taxa in each time bin: for example, a taxon that has 15% of occurrences will have a frequency of 0.15 for that bin (its “share”). Specimens are drawn and their frequencies are summed to generate a coverage value (the proportion of the total distribution of taxa represented by specimens drawn so far). Sampling stops when a desired level of coverage (the “quorum”) is reached. Each species can only have its share counted towards the total once, so a quorum cannot be reached by sampling only multiple specimens of one species. An estimator (Good’s u) is used to assess the sampling quality of time bins by subtracting the number of single-occurrence taxa from 1. Therefore, an increased number of singletons lowers u, and is indicative of poor sampling. Good’s u is used to adjust the quorum of each time bin based on these singleton proportions (to estimate underlying ‘true’ richness) and therefore time bins cannot be sampled above their u value, and are excluded from analyses where the quorum level is higher than this.

SQS was carried out using a Perl script (version 4.3, see Appendix 3) on global and continental scale data at a quorum level of 0.4, which produces a sufficient measure of relative standing diversity (Alroy, 2010b; Benson et al., 2016). For my analyses, some standard stratigraphic stages were merged to reduce variance in durations between bins (Table 2.1), resulting in interval durations of approximately 9 Ma. Occurrences were only included in the analyses if their stratigraphic ages were known with sufficient certainty to be assigned to a single time bin. With the global data, this excluded 5% of total occurrences, with continental data exclusions ranging from 1–3% (Europe, North America, Indo-Madagascar, Oceania), to 14% (Africa, Asia) and 34% (South America). I also included the amalgamated Jehol Group as a separate point in the figures, which crosses several time bins and was thus excluded from the analyses otherwise, in order to show the generic richness in this Lagerstätte. Countries included in each continent are shown in Table 2.2.

Table 2.1: Time bins of roughly 9Ma each used in SQS and GLS analyses, and the standard stratigraphic intervals they correspond to. Previous Triassic bins have no occurrences named at genus level and so are excluded from this table.

Bin name	Epoch/stage equivalent	Base (Ma)	Midpoint (Ma)
Pg5	Rupelian-Chattian	33.9	28.465
Pg4	Bartonian-Priabonian	41.2	37.55
Pg3	Lutetian	47.8	44.5
Pg2	Ypresian	56.0	51.9
Pg1	Selandian-Thanetian	61.6	58.8
Pg0	Danian	66.0	63.8
K8	Maastrichtian	72.1	69.05
K7	Campanian	83.6	77.85
K6	Turonian-Coniacian-Santonian	93.9	88.75
K5	Cenomanian	100.5	97.2
K4	Albian	113.0	106.75
K3	Aptian	125.0	119.0
K2	Hauterivian-Barremian	132.9	128.95
K1	Berriasian-Valanginian	145.0	138.95
J6	Kimmeridgian-Tithonian	157.3	151.15
J5	Callovian-Oxfordian	166.1	161.7
J4	Bajocian-Bathonian	170.3	168.2
J3	Toarcian-Aalenian	182.7	176.5
J2	Pliensbachian	190.8	186.75
J1	Hettangian-Sinemurian	201.3	196.05
Tr5	Rhaetian	208.5	204.9
Tr4	Norian	227.0	217.75

Table 2.2: List of continent assignments for the countries in my dataset. Russia is included twice due to its large geographic spread, and the data was edited to reflect this, in order to separate out the localities closer to Europe or Asia.

Continent	Included countries
Africa	Algeria, Egypt, Libya, Morocco, Niger, South Africa, Sudan, Tanzania
Asia	China, Japan, Kazakhstan, Kyrgyzstan, Lebanon, Mongolia, Myanmar, Pakistan, Russian Federation, South Korea, Uzbekistan
Europe	Austria, Belgium, France, Germany, Germany, Hungary, Italy, Luxembourg, Poland, Portugal, Romania, Russian Federation EU, Spain, Switzerland, United Kingdom
Indo-Madagascar	India, Madagascar
North America	United States, Canada, Mexico
Oceania	Australia, New Zealand
South America	Argentina, Brazil, Bolivia, Colombia, Peru, Venezuela

To highlight underlying data quality, I marked the number of collections that contributed to each data point in the subsampled curves. In order to avoid problems with Good's *u* for very small sample sizes (it can be overestimated stochastically for very small samples), I excluded estimates of richness that were based on data from fewer than five collections from most of the figures (a figure with these added back in is also included in my results for reference).

As a comparison for the SQS analyses, I also used classical rarefaction (CR), which employs uniform sampling rather than coverage to construct estimated richness curves. Uniform sampling has been criticised for its tendency to dampen genuine richness signals (Alroy, 2010b), since the quota is restricted to the level of the bin with the poorest observed taxonomic richness and thus often flattens the 'corrected' richness curve. Nevertheless, I compare results obtained from the two subsampling techniques in order to look for common patterns of lepidosaur richness changes through time, as no currently available subsampling method can account for all possible problems with the data,

To demonstrate visually the effects of geographic biases on my results, I also plotted the palaeolatitudinal distribution of lepidosaur occurrences through time versus those for other terrestrial tetrapods. Palaeolatitudes are calculated by Fossilworks based on the present-day coordinates of collections using tectonic plate rotation data from the GPlates model (Wright et al., 2013).

2.2.3 Comparing richness to proxies for sampling and environmental factors

I used generalized least-squares regression (GLS) to examine the relationship between uncorrected generic richness, proxies for sampling (non-marine tetrapod-bearing collections (TBCs) and tetrapod-bearing formations (TBFs)), sea level, non-marine area and palaeotemperature. Time bins remain the same as in the SQS analyses above (Table 2.1). GLS allows autoregressive (AR) models to be fitted to the data, which benefits the analysis by reducing the chance of overestimating statistical significance of regression lines due to serial correlation (e.g. Chatfield, 2003). Sampling proxy data was obtained from the PBDB, from a dataset of non-marine tetrapods (61316 occurrences downloaded on 10 March 2017; see Appendix 4). Non-marine area was obtained from the palaeocoastline reconstructions of Smith, Smith and Funnell (2004). For palaeotemperature I used sea surface $\delta^{18}O$ records from Prokoph, Shields and Veizer (2008) compiled by Mannion et al. (2015), which I interpolated into 0.1 Ma intervals for ease of subsequent aggregation (see Appendix 5 for script and data); more negative values indicate warmer temperatures, and vice versa. All proxy data was aggregated into within-bin means to enable comparisons with richness data. Sea level (Miller et al., 2005) and palaeotemperature data are only available from the Jurassic onward, and so analyses range from bins J4—Pg5. Duration of time bins was included as an additional non-optional explanatory variable to ensure this was not influencing results, as in Marx and Uhen (2010).

I fitted AR models of orders one or two to combinations of the above explanatory variables, and then compared the results of the GLS analyses using Akaike's Information Criterion for small sample sizes (AICc). Akaike weights were calculated from this information to identify the best combination of explanatory variables from those tested. AIC methods reward models for goodness of fit, but penalise those with higher numbers of explanatory variables (Burnham and Anderson, 2002). I also manually calculated R^2 values from the GLS output using the generalised R^2 of Nagelkerke (Nagelkerke, 1991). R^2 provides an estimate of the amount of variance in generic richness explained by the variables in each model. To check normality and homoskedasticity of each model's residuals I used the Jarque-

Bera (Jarque and Bera, 1980) and Breusch-Pagan (Breusch and Pagan, 1979) tests. The sampling proxies and generic richness were ln-transformed prior to analysis to ensure normality and homoskedasticity of residuals. All analyses were carried out in R version 3.2.2, using the packages lmtest version 0.9-35 (Hothorn et al., 2017), nlme version 3.1-131 (Pinheiro et al., 2017), qpcR version 1.4-0 (Spiess and Ritz, 2014) and tseries version 0.10-40 (Trapletti and Hornik, 2017), following the approach of Benson and Mannion (2012). See Appendix 6 for GLS script.

2.3 Results

2.3.1 Global face-value and subsampled curves

Patterns of relative change in subsampled richness estimates for pooled global occurrences are highly similar to face-value genus counts (Figure 2.1). Many time bins do not provide subsampled richness estimates, as they are too poorly sampled to meet the lowest quorum level (indicated by low values of Good's u). Even at lower quorum levels (Figure 2.2) the situation is not much improved, and a signal could not be recovered for many Jurassic and mid-Cretaceous time bins.

There was a higher generic richness before the Triassic–Jurassic (Tr–J) boundary than afterwards in both the face-value and corrected data. At a quorum level of 0.4 there are few well-supported data points prior to the Late Cretaceous. The Late Cretaceous had the highest 'global' standing richness estimate, in bin K7 (Campanian), followed by lower estimated richness in K8 (Maastrichtian) and across the K–Pg boundary. There is a subsequent gap in coverage until Pg4 (Bartonian–Priabonian) in the 0.4 quorum data, but at a quorum level of 0.3 (Figure 2.2) the data suggest that the trend is more like that seen in the face-value data, with sequential peaks and troughs culminating in a lower richness estimate in the Oligocene than in the preceding time bin (Pg4–Pg5), which is still relatively high compared to the overall curve. This low a quorum level is unreliable, however. Very similar trends are recovered using CR (Figure 2.3), though there are some issues with missing data and flattened trends at lower quotas.

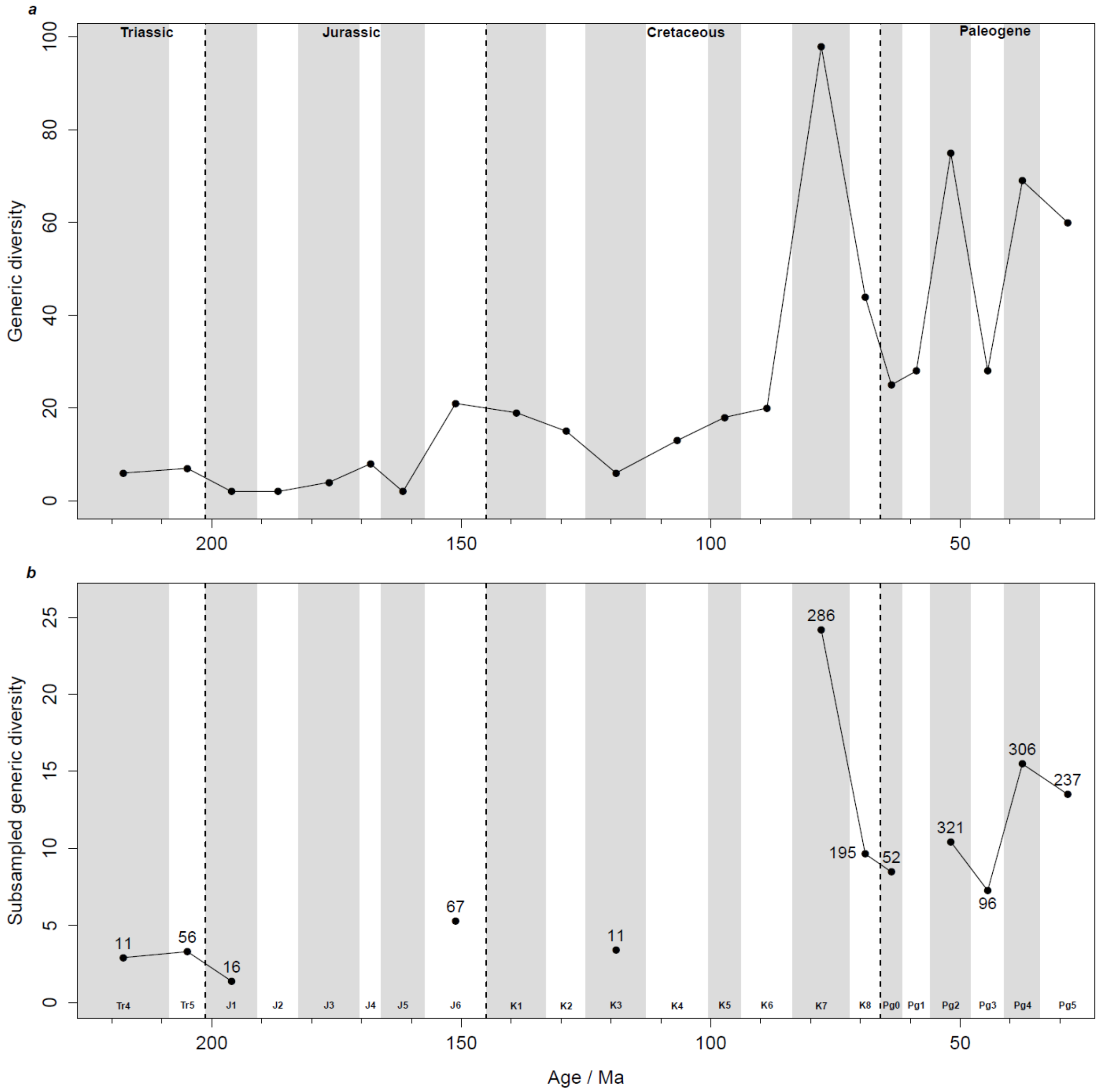

Figure 2.1: Non-marine lepidosaur global generic richness from Triassic–Paleogene using *(a)* uncorrected (face-value) richness and *(b)* subsampled (SQS) richness at quorum 0.4. Numbers on figure 1*b* indicate number of collections drawn for each time bin subsampled, as an indicator of underlying data quality. T, Triassic; J, Jurassic; K, Cretaceous; Pg, Paleogene.

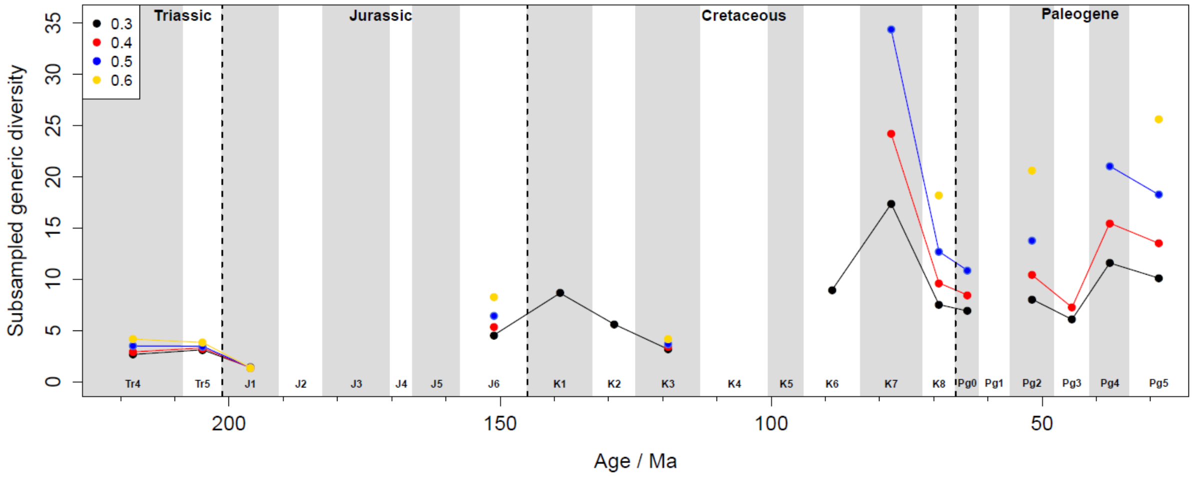

Figure 2.2: Subsampled global non-marine lepidosaur generic richness from Triassic–Paleogene at quorum levels 0.3–0.6. T, Triassic; J, Jurassic; K, Cretaceous; Pg, Paleogene.

Figure 2.3: Rarefied global non-marine lepidosaur generic richness from Triassic–Paleogene at a quota of 15, 30 and 45. Numbers indicate number of collections drawn for each time bin subsampled, as an indicator of underlying data quality. T, Triassic; J, Jurassic; K, Cretaceous; Pg, Paleogene.

2.3.2 Continental-level subsampling

Splitting the ‘global’ data into continental regions reveals a major problem of uneven sampling between continents (Figure 2.4) and demonstrates that literal readings of the apparently ‘global’ curve can lead to the recovery of erroneous macroevolutionary signals. For example, Indo-Madagascar has exceptionally poor coverage, and robust SQS results (i.e. those underpinned by >5 collections per time bin) are only recovered in one time bin. Africa and Oceania have so few collections providing data for each time bin that no robust results were obtained and these areas were excluded from figure 2.4 (see Figure 2.5 for comparison). Therefore, little can be said about richness trends on these continents except that in the Maastrichtian (K8), richness in Indo-Madagascar was much lower than in North America. There are further substantial differences in the numbers of collections attributed to each data point between continents, which are particularly marked between the relatively well-sampled areas of North America and Europe versus the rest of the world.

A drop in richness was recovered across the Tr–J boundary in Europe, but other regions do not reach the minimum quorum level in the final Triassic bin, so it is unclear what is occurring outside Europe at this time. This means that the ‘global’ record from this time is almost entirely driven by the European signal. Comparatively high richness was found in the Late Jurassic (K6; Kimmeridgian–Tithonian) in Europe and becomes lower in the Early Cretaceous (K2), but after this time a robust generic richness signal is absent until the Paleogene, and it is not possible to track sequential trends in richness before this time. Asia and South America had a relatively intermediate to low estimated richness during the Cretaceous, but the data available are sparse and so these estimates are less reliable. For the amalgamated Jehol Group, the richness estimate was relatively low, but is more robustly supported than the Asian K3 point.

Figure 2.4: Subsampled non-marine lepidosaur generic richness from Triassic–Paleogene at quorum 0.4, for individual continents: AS, Asia; EU, Europe; IM, Indo-Madagascar; NAm, North America; SA, South America. Also included is the combined Jehol Group. Numbers indicate collections drawn for each time bin subsampled, as an indicator of underlying data quality. T, Triassic; J, Jurassic; K, Cretaceous; Pg, Paleogene. Arrows indicate approximate timings of major finds in the lepidosaur record: 1, formerly earliest squamate fossils; 2, earliest snake fossils; 3, earliest amphisbaenian fossils. The recently proposed earliest potential squamate is outside the temporal scope of this figure.

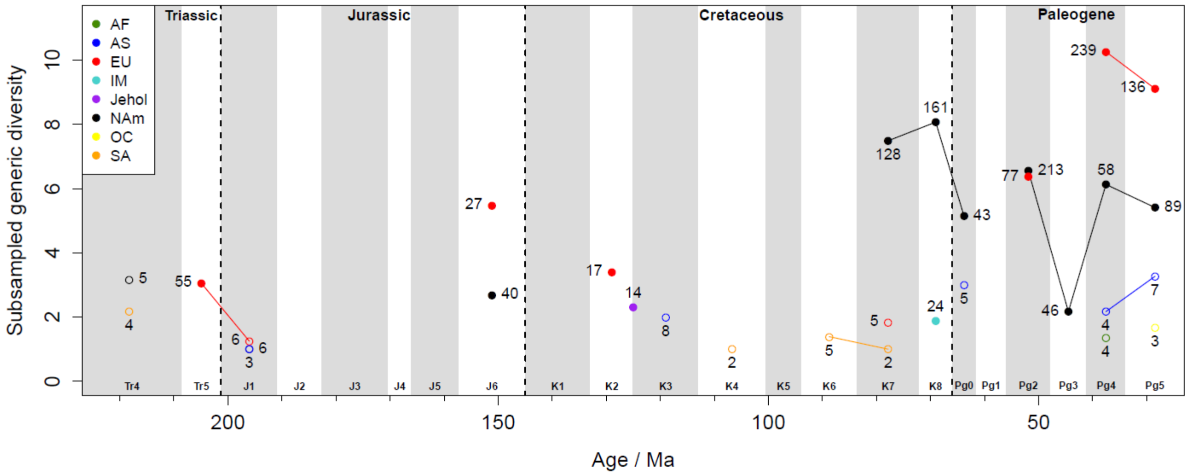

Figure 2.5: Subsampled non-marine lepidosaur generic richness from Triassic–Paleogene at quorum 0.4, for individual continents, including those points with fewer than 5 collections: AF, Africa; AS, Asia; EU, Europe; IM, Indo-Madagascar; NAm, North America; OC, Oceania; SA, South America. Also included is the combined Jehol Group. Numbers indicate number of collections drawn for each time bin subsampled, as an indicator of underlying data quality. T, Triassic; J, Jurassic; K, Cretaceous; Pg, Paleogene.

In the latest Late Cretaceous, North America shows the same proportionately high levels of generic richness as the 'global' curves, but with one notable difference. In North America, K8 had a slightly higher richness than K7 (probably within stochastic error of each other) and the drop across the K–Pg boundary was steeper than in the 'global' curve. In addition, this latest Cretaceous peak no longer represents the highest richness in the time series, and it is proportionately much lower than the peak seen in the 'global' curve at this time (which is due to the fact that the 'global' curve is an amalgamation). No other continents exhibit any sequential trends through this time period, but seem to have had a low to intermediate richness when compared to North America. After a break, there was the same peak and trough pattern as in Figure 2.1 for Pg2–5 in North America, and partially in the European data (the missing trough is poorly sampled but present at quorum 0.3; see Figure 2.6), although it is more pronounced than in the 'global' data. In Pg4 there was extremely high relative richness in Europe, and high richness in North America, followed by decreases in both regions over the Eocene–Oligocene boundary, as in the 'global' data.

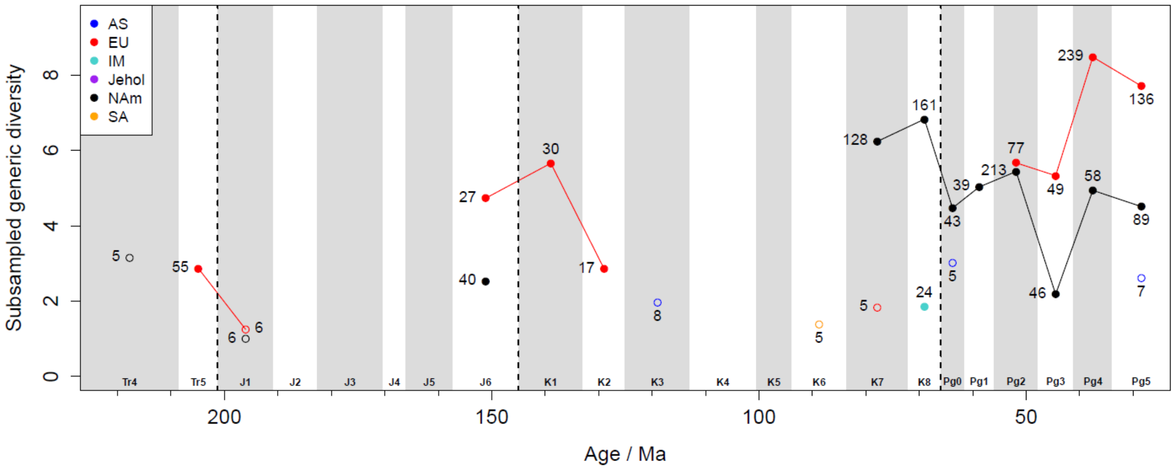

Figure 2.6: Subsampled non-marine lepidosaur generic richness from Triassic–Paleogene at quorum 0.3, for individual continents: AS, Asia; EU, Europe; IM, Indo-Madagascar; NAm, North America; SA, South America. Also included is the combined Jehol Group. Numbers indicate number of collections drawn for each time bin subsampled, as an indicator of underlying data quality. T, Triassic; J, Jurassic; K, Cretaceous; Pg, Paleogene.

The same trends can be seen at quorum 0.5 (Figure 2.7a), with a few points dropping out due to poorer sampling. Europe appears to be badly sampled in a number of intervals including bin K2 (Hauterivian–Barremian), Pg2 (Ypresian) and Pg4 (Bartonian–Priabonian), and so these points are not recovered at quorum 0.5. Pg3 North America is too poorly sampled and also drops out at this level. The very high richness in Europe during Pg5 (Oligocene) remains. At quorum 0.6 (Figure 2.7b) only North American points are recovered, with everything except Pg4 still recovered. Late Cretaceous richness remains markedly high, with a notable drop across the K–Pg boundary; richness had almost reached these levels again by Pg2.

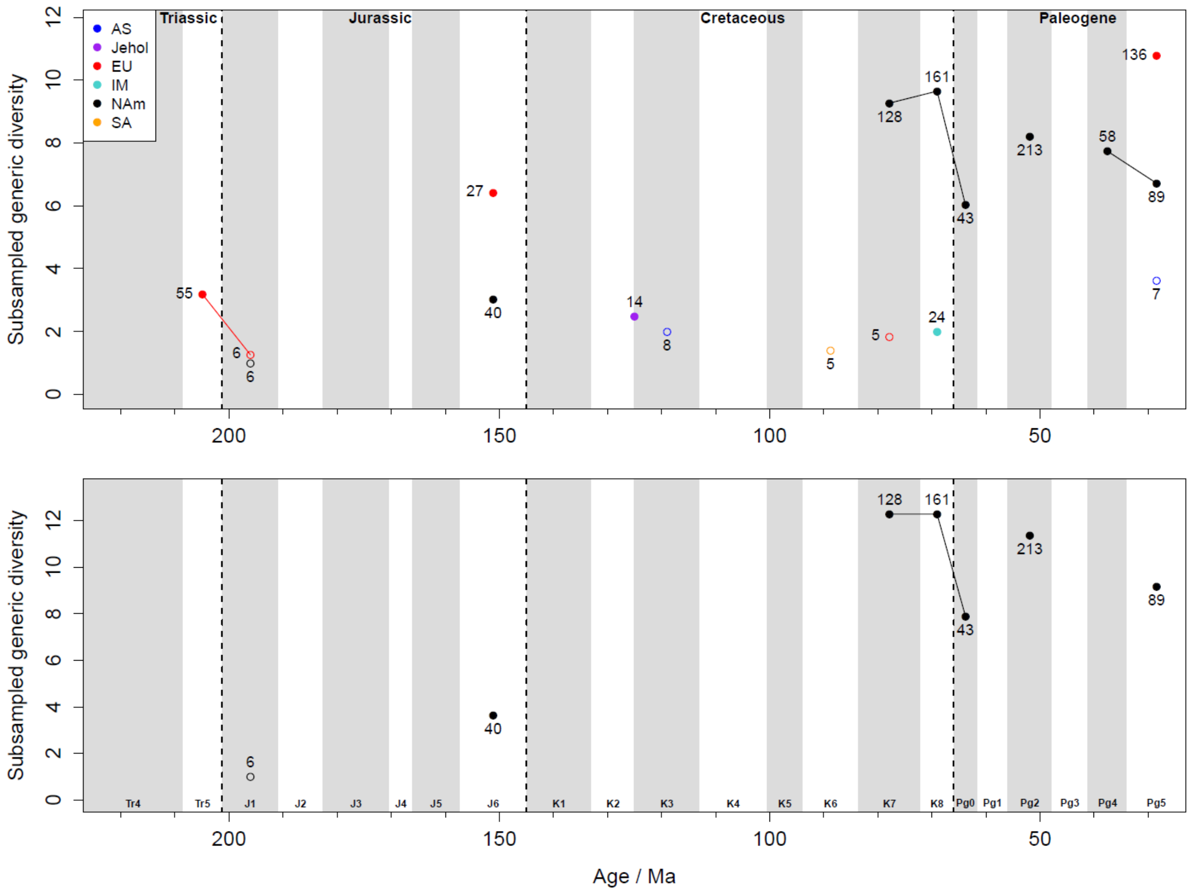

Figure 2.7: Subsampled non-marine lepidosaur generic richness from Triassic–Paleogene at quorum (*a*) 0.5, and (*b*) 0.6 for individual continents: AS, Asia; EU, Europe; IM, Indo-Madagascar, NAm, North America; SA, South America. Also included is the combined Jehol Group. Numbers indicate number of collections drawn for each time bin subsampled, as an indicator of underlying data quality. T, Triassic; J, Jurassic; K, Cretaceous; Pg, Paleogene.

In the CR continental results (Figure 2.8), I recovered the same trends but very dampened compared to SQS. At a quota of 15, which is unreliably low, everything except those points from very few collections are the same, with three differences. Points at K7 for Asia, Pg1 (Selandian–Thanetian) for North America and Pg3 (Lutetian) for Europe are recovered; the Asian K7 point had the highest richness of all bins. At a quota of 30, still very low, everything pre-K7 drops out, along with Pg1 North America and Pg2–3 Europe, which is similar to the losses from SQS as I moved up quorum levels. The Asian K7 point remains and is much higher than any other in the time period covered.

I highlighted the disparity between sampling in different continents by plotting the palaeolatitudes of every terrestrial lepidosaur occurrence through time against those for all other terrestrial tetrapods (Figure 2.9). It is clear that the vast majority of lepidosaur occurrences are from the Northern Hemisphere, and this appears to be a trend for tetrapods as a whole. The absence of lepidosaur specimens at very high latitudes is notable, even when other tetrapods are present.

To test sensitivity to taxonomic sampling, I separated snake occurrences from the dataset and ran new analyses on these 'snake-only' and 'other lepidosaur' datasets. While snakes are too poorly sampled to produce meaningful results, with the exception of a slight drop across the Eocene–Oligocene boundary, their exclusion from the lepidosaur dataset causes a few points that had previously dropped out of analyses (Pg1 of North America and Pg3 of Europe) to be recovered (see Figure 2.10). This is most likely due to a decrease in the number of singleton taxa that would have otherwise been included in these bins, thus raising the maximum quorum level that they could be sampled at, allowing us to recover a signal at quorum 0.4.

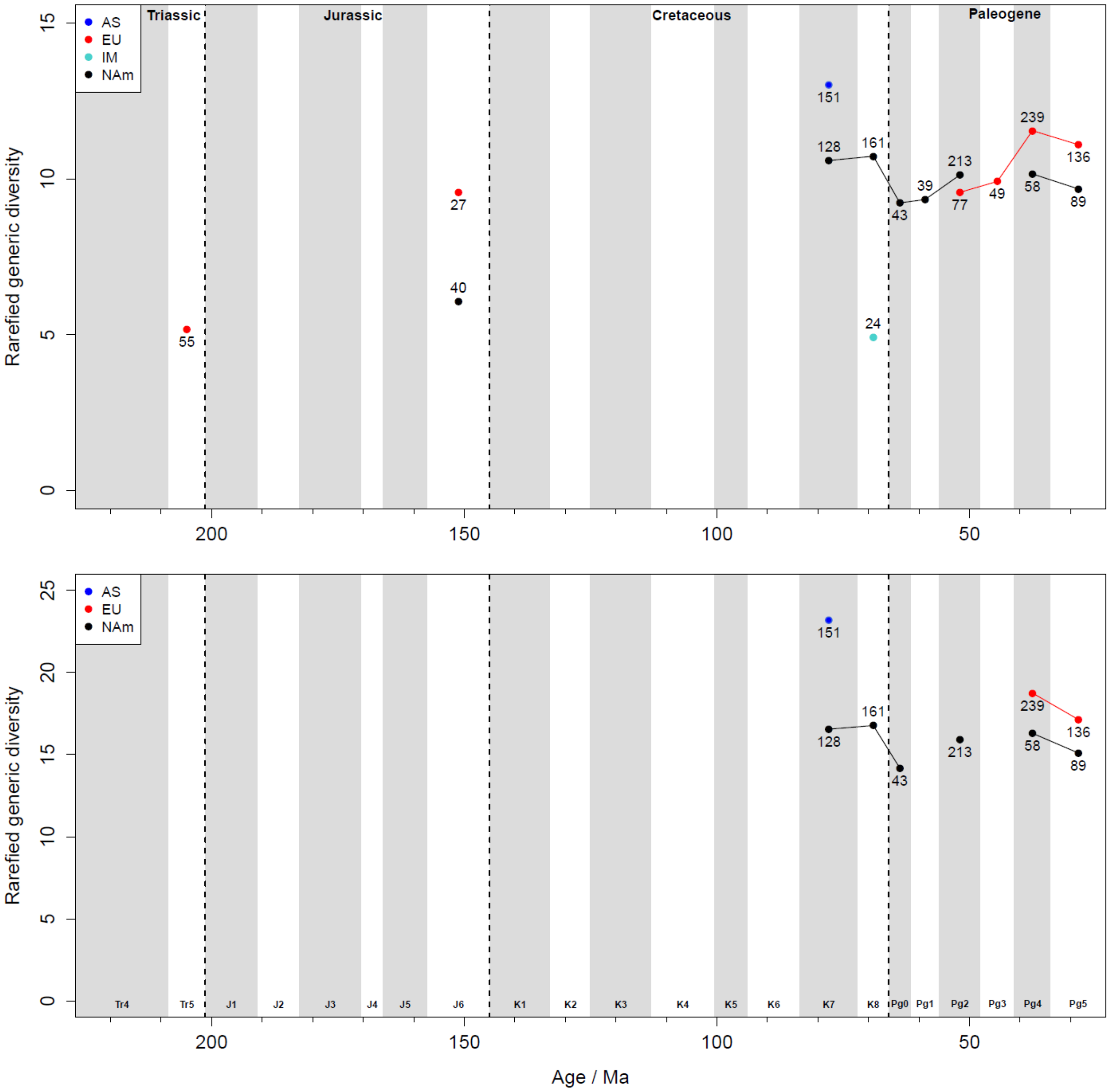

Figure 2.8: Rarefied non-marine lepidosaur global generic richness from Triassic–Paleogene using a quota of *(a)* 15 and *(b)* 30 for individual continents. AS, Asia; EU, Europe; IM, Indo-Madagascar; NAm, North America. Numbers indicate number of collections drawn for each time bin subsampled, as an indicator of underlying data quality. T, Triassic; J, Jurassic; K, Cretaceous; Pg, Paleogene.

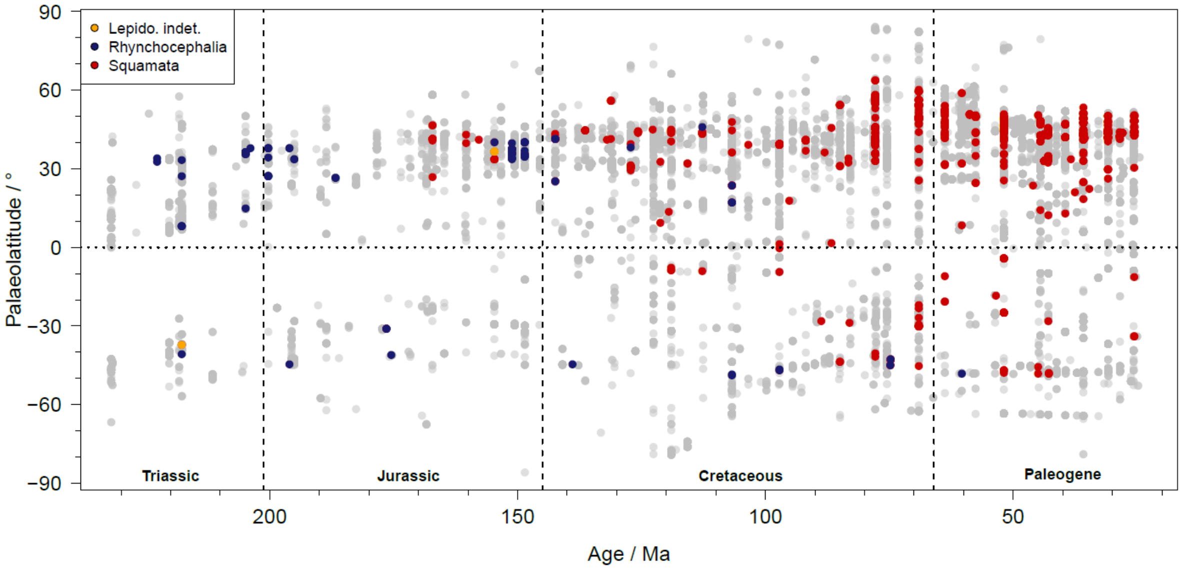

Figure 2.9: Palaeolatitudinal distribution of non-marine lepidosaur occurrences from Triassic–Paleogene; light blue circles indicate non-lepidosaur tetrapod occurrences. T, Triassic; J, Jurassic; K, Cretaceous; Pg, Paleogene.

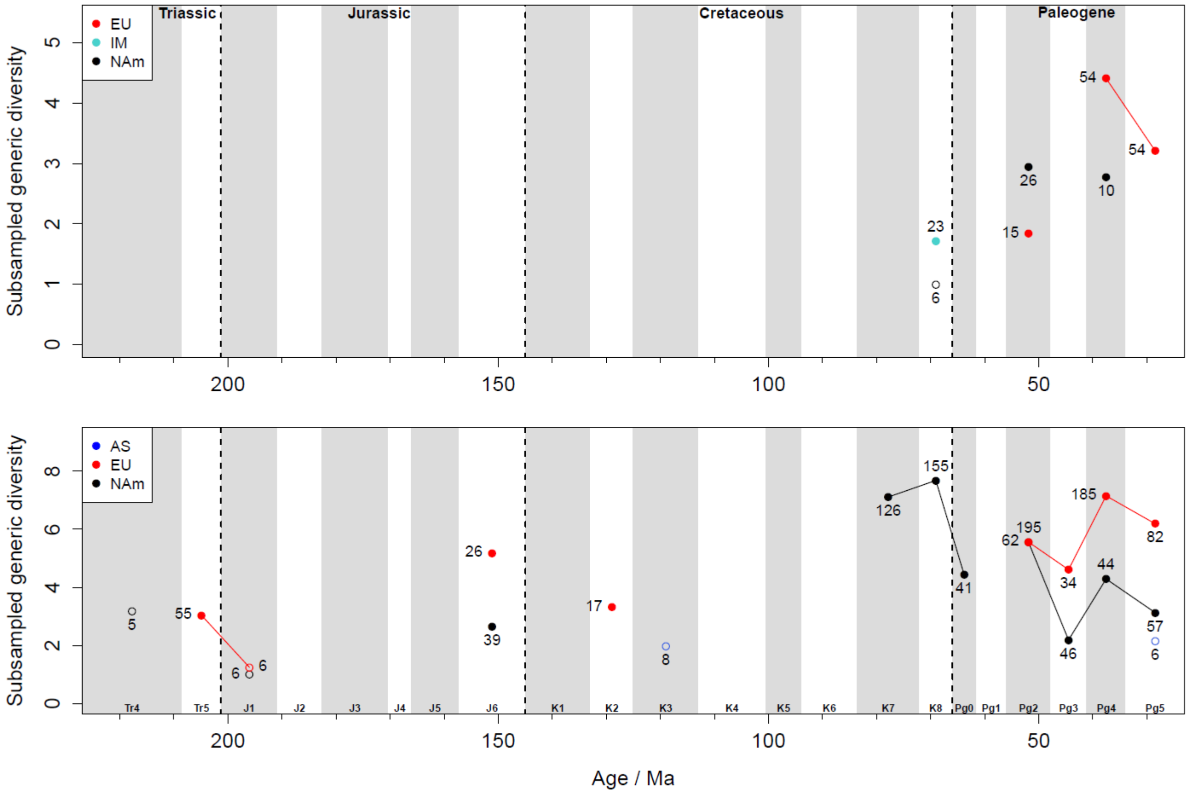

Figure 2.10: Subsampled generic richness at quorum 0.4 for (*a*) non-marine snakes and (*b*) non-marine lepidosaurs excluding snakes for individual continents: AS, Asia; EU, Europe; IM, Indo-Madagascar; NAm, North America. Numbers indicate number of collections drawn for each time bin subsampled, as an indicator of underlying data quality. T, Triassic; J, Jurassic; K, Cretaceous; Pg, Paleogene.

2.3.3 Generalized least-squares analyses

I examined various explanatory variables that may be driving face-value generic richness using GLS (Table 2.3). Duration was not found to be significant for any of the models tested and did not change the order of the best models, so these results are presented in Appendix 2. Seven models had higher AICc weights than the null model. The best models based on weights all featured tetrapod-bearing collections (TBCs), except for the tetrapod-bearing formations (TBFs) model which was selected as the fifth best. The TBCs + $\delta^{18}O$ palaeotemperature model emerged as the best model (AICc= 34.6), followed by TBCs only (AICc= 37.5), and then TBCs + non-marine area (AICc= 39.2). While TBCs + palaeotemperature has been identified as the best model, with a moderately good R^2 value, within this model TBCs is highly significant (p<0.0001) but the palaeotemperature proxy is not (p= 0.76). In all of the best model combinations, TBCs is the only variable with significant p values. Furthermore, the AICc values of this model and the TBCs-only model are similar, so the AICc weight of the TBCs only model is low but non-negligible. This makes it more difficult to choose between these two models, but implies strongly that sampling (TBCs) is the most important explanatory variable.

Table 2.3: Summary of various model fits to observed (face-value) generic richness of non-marine lepidosaurs, ordered by AICc score. N= 17; "TBCs", tetrapod-bearing collection; "TBFs", tetrapod-bearing formations; "$\delta^{18}O$" is the palaeotemperature proxy. Column abbreviations are AR, order of AR model chosen; LL, log-likelihood.

Model name	**AR**	**Sampling (TBCs or TBFs)**			**Sea level (SL)**			**Non-marine area (NMA)**			**$\delta^{18}O$**			**R^2**	**LL**	**AICc**	**AICc weight**
		slope	*t*-value	*p*	slope	*t*-value	*p*	slope	*t*-value	*p*	slope	*t*-value	*p*				
TBCs + $\delta^{18}O$	**2**	**1.183**	**14.133**	**0.000**	**-**	**-**	**-**	**-**	**-**	**-**	**0.022**	**0.315**	**0.758**	**0.724**	**-10.4**	**34.6**	**0.716**
TBCs	**1**	**1.018**	**5.856**	**0.000**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**0.559**	**-14.3**	**37.5**	**0.163**
TBCs + NMA	**2**	**1.176**	**14.107**	**0.000**	**-**	**-**	**-**	**0.004**	**0.713**	**0.487**	**-**	**-**	**-**	**0.638**	**-12.7**	**39.2**	**0.072**
TBCs + SL	2	1.171	6.993	0.000	0.000	0.082	0.936	-	-	-	-	-	-	0.611	-13.3	40.4	0.039
TBCs + NMA + $\delta^{18}O$	2	1.172	13.234	0.000	-	-	-	0.007	0.722	0.483	-0.044	-0.376	0.713	0.584	-13.8	45.0	0.004
TBFs	1	1.216	2.963	0.010	-	-	-	-	-	-	-	-	-	0.310	-18.2	45.2	0.004
TBCs + SL + $\delta^{18}O$	2	1.142	6.055	0.000	0.001	0.244	0.811	-	-	-	0.031	0.367	0.720	0.534	-14.8	46.9	0.002
Null model	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-21.3	48.9	<0.001
TBFs + $\delta^{18}O$	2	1.365	3.377	0.005	-	-	-	-	-	-	0.060	0.269	0.792	0.279	-18.5	50.9	<0.001
TBCs + SL + NMA	2	1.152	6.609	0.000	0.001	0.158	0.877	0.004	0.699	0.497	-	-	-	0.380	-17.2	51.8	<0.001
$\delta^{18}O$	2	-	-	-	-	-	-	-	-	-	-0.029	-0.096	0.925	0.017	-21.2	53.2	<0.001
TBFs + SL	2	0.960	2.372	0.033	0.015	2.256	0.041	-	-	-	-	-	-	0.150	-19.9	53.7	<0.001
TBFs + NMA	2	1.528	4.064	0.001	-	-	-	0.022	1.233	0.238	-	-	-	0.090	-20.5	54.8	<0.001
NMA	2	-	-	-	-	-	-	-0.041	-1.968	0.068	-	-	-	-0.110	-22.2	55.2	<0.001
SL	2	-	-	-	0.025	4.275	0.001	-	-	-	-	-	-	-0.137	-22.4	55.6	<0.001
TBFs + SL + $\delta^{18}O$	2	0.916	2.354	0.035	0.017	2.733	0.017	-	-	-	0.225	1.414	0.181	0.151	-19.9	57.2	<0.001
TBCs + SL + NMA + δ18O	2	1.181	5.747	0.000	0.000	-0.045	0.965	0.007	0.643	0.532	-0.049	-0.326	0.750	0.308	-18.2	57.8	<0.001
$\delta^{18}O$ + SL	2	-	-	-	0.027	4.683	0.000	-	-	-	0.244	1.309	0.212	-0.126	-22.3	58.5	<0.001
TBFs + NMA + $\delta^{18}O$	2	1.533	3.931	0.002	-	-	-	0.026	1.130	0.279	-0.085	-0.330	0.747	0.046	-20.9	59.1	<0.001
$\delta^{18}O$ +NMA	1	-	-	-	-	-	-	-0.019	-0.655	0.523	0.043	0.133	0.896	-0.376	-24.0	59.9	<0.001
TBFs + SL + NMA	2	1.030	2.558	0.024	0.015	2.217	0.045	0.020	1.380	0.191	-	-	-	-0.144	-22.4	62.2	<0.001
SL + NMA	2	-	-	-	0.025	4.077	0.001	0.013	0.778	0.450	-	-	-	-0.599	-25.3	64.4	<0.001
TBFs + SL + NMA + $\delta^{18}O$	2	0.967	2.258	0.043	0.016	2.235	0.045	0.009	0.442	0.666	0.140	0.561	0.585	-0.197	-22.8	67.1	<0.001

2.4 Discussion

2.4.1 Lepidosaur richness curves and the influence of geographic spread

SQS and CR recovered the same broad richness trends through time, but the CR quota levels that recover sufficient data are unreliably small and generally drop more points than the SQS analyses. Where points that SQS does not recover are found (e.g. K7 Asia) there are issues with many single-occurrence taxa that point to an overall poorly sampled interval which does not compare reliably with other intervals when studying the relative richness that subsamples produce. Unfortunately, CR curves suffer from the problem of dampening at lower quota levels (see Alroy, 2010b for a full discussion of this) and this means that some quotas with useful information exhibit very shallow fluctuations in richness. As a result of their general similarity, I choose to focus on the SQS results herein.

The lepidosaur fossil record is variably sampled, but sampling is often poor across multiple time bins and regions, as seen in Figures 2.1b and 2.4 versus Figure 2.1a, where many time bins dropped out of the SQS analysis as the data could not be subsampled at even very low quorum levels. This issue obscures overall richness trends through time, and prevents us from presenting robust scenarios for a large portion of early lepidosaur history. Similarly, sampling affects what we observe in 'global' trends; many portions of the sampling-corrected 'global' curve result from a small number of continents producing the majority of the signal (Figure 2.4). Additionally, 'global' curves tend to be cumulative based on the number of continents contributing to a time bin (i.e. double the continents, double the estimated richness) in absence of any underlying changes in richness. For example, Benson et al. (2016) found that the palaeogeographic spread of global fossil localities explained 72% of the variance in subsampled 'global' species richness for terrestrial tetrapods (and see

Alroy, 2010b; Close et al., 2017). This means that apparently 'global' richness curves predominantly tell us about the number of continents that have been sampled of fossils in each interval. The presence or absence of data from different regions in adjacent time intervals have the potential to create false peaks and troughs that could lead to misinterpretations of true 'global' richness, so 'global' curves should always be interpreted with caution. As such, I will focus more on what my continental-level results show than the 'global' curve.

2.4.2 Mesozoic lepidosaur richness and the diversification of major clades

North America and Europe are well sampled compared with the rest of the world, which affects what can be concluded about overall trends in taxic richness. This Northern Hemisphere bias is shown clearly in the palaeolatitudinal distribution of occurrences through time (Figure 2.9), which is similar to that recorded for other tetrapod groups (Mannion et al., 2015; Nicholson et al., 2015). This bias appears to be much more acute for lepidosaurs, but it is not currently possible to determine if it reflects sampling issues that might be specific to lepidosaurs (such as their small size and low recovery potential), a genuinely different geographic distribution from that of other tetrapods (due to ecological or physiological dissimilarities) or a combination of these factors.

Despite this, the signal that I recovered is still informative for first order patterns of lepidosaur richness. At a quorum level of 0.4 (remaining at 0.5; see Figure 2.7), there was a stable and low level of rhynchocephalian richness in the latest Triassic that declined across the Tr–J boundary. This interval is only resolved for Europe (as all other continents are too poorly sampled; Figure 2.4), and studies of taxon changes in the fissure-fill faunas of the southwest of England indicate a shift from a Late Triassic fauna dominated by reptiles, with a diverse array of rhynchocephalians, to a mammaliaform-dominated Early Jurassic fauna (there is no evidence to discern whether this is a local environmental signal or more widespread; Whiteside et al., 2016). Early Jurassic rhynchocephalians were reduced to two genera (*Gephyrosaurus* and *Clevosaurus*), though they remained dominant in abundance despite low generic richness (Whiteside et al., 2016). By contrast, the Tr–J boundary effect on mammal generic richness is apparently negligible (Newham et al., 2014). This faunal shift might be related to the end-Triassic mass extinction, which was likely caused by environmental change associated with long-term volcanism in the Central Atlantic Magmatic Province (CAMP; Blackburn et al., 2013; Lucas and Tanner, 2015; Percival et al., 2017).

There is a sampling gap (caused by poor data), that prevents us from estimating richness for many regions and intervals, through most of the Jurassic and Cretaceous until the latest Cretaceous. It is punctuated by a few individual points in various time bins where the presence of various Lagerstätten (areas of exceptional quality or quantity of preservation) have ensured that sampling is good enough to be detected at this quorum level. These include the Late Jurassic (J6) localities of the Morrison Formation (western North America), Solnhofen (Germany), Cerin (France) and Guimarota (Portugal), and the Early Cretaceous sites of Las Hoyas (K2; Spain) and the Jehol Group (partially in K3; China). This is unfortunate, as I lack information on how richness patterns were affected by some of the first recorded occurrences of squamates in the Middle Jurassic (Nessov, 1988; Evans, 1994; Evans, Prasad and Manhas, 2002) and by the diversifications of the many extant terrestrial clades that are thought to have occurred during the Early Cretaceous (e.g. scincids, xenosaurid anguimorphs, acrodont iguanians; Evans, Wang and Li, 2005).

Bin K1 (Berriasian–Valanginian) is too poorly sampled in both the 'global' and continental curves at quorum 0.4 to observe anything across the J–K boundary, but at quorum 0.3 (Figure 2.6) there was a rise across the boundary in Europe. This is contrary to results obtained for the majority of large-bodied contemporaneous reptiles (Mannion et al., 2015; Jonathan P. Tennant, Mannion and Upchurch, 2016), but is consistent with the findings of other studies on mammals (Newham et al., 2014), non-marine turtles (Nicholson et al., 2015) and small-bodied tetrapods in general (Jonathan P. Tennant, Mannion and Upchurch, 2016). Only Europe is recoverable here, however, and therefore this should be interpreted as a regional phenomenon and as unreliable due to the low quorum level.

Southern hemisphere (Gondwanan) continents are represented in time bin K6 (Turonian–Santonian), but there are too few localities to be interpretable. Rhynchocephalians still occupied South America at this time, despite their disappearance from Laurasian continents during the Jurassic and Early Cretaceous (Evans and Jones, 2010; Apesteguía and Jones, 2012), and were found alongside numerous snakes (which were uncommon in Laurasia at this time; Evans and Jones, 2010).

In the 'global' data I recovered the highest richness for the entire dataset in the Campanian (K7). However, in the continental-level curves this large Late Cretaceous peak disappears (since it is an agglomeration of varying regions) and in the North American curve the difference between these bins and the Paleogene is not as large. In Europe it appears as if the Eocene had the highest richness (Pg4), but there is only one poorly supported Late Cretaceous point (and less points overall) so comparisons are more difficult.

Only North America is well sampled enough to show sequential richness during the end-Cretaceous, with a wealth of squamate taxa from Canada and the USA that increased slightly in richness from the Campanian to the Maastrichtian (however, these estimates are probably within stochastic error for each other; at higher quorum levels they are nearly or exactly the same- see Figure 2.7). This result is contrary to that of Nydam (2013), who found higher taxonomic richness in the Campanian, but noted that this result could have been affected by sampling biases, as he did not use any sampling-correction methods. Lower relative richness was found in Indo-Madagascar at this time (and also in K7 Europe but this data point is not as robustly supported, with only 5 collections). 'Global' Campanian high richness is heavily influenced by finds from Mongolia, which has an incredibly rich fauna (with >50 genera), in conjunction with North America. Unfortunately, this Asian signal is lost when richness is examined at the continent-level, as there are numerous single-occurrence taxa in these Mongolian faunas and so the SQS algorithm deems the continent too poorly sampled to be counted. Moreover, it is likely that revision of the Mongolian faunas is needed, as there are many named taxa (e.g. Alifanov, 2000) that have not been examined in detail subsequently and the generic richness of these sites might change substantially in the future. This observation highlights a major issue in diversity studies in palaeontology as a whole, which is reliance on an existing body of literature that is heavily biased towards the description of new taxa or localities, rather than documenting in detail every occurrence of common or widespread taxa. For examples, museum collections that house large numbers of duplicate specimens from places like the Campanian of Mongolia (e.g. IVPP; S.E. Evans, personal observation) that are not individually listed in the literature, do not contribute towards publications-based analyses such as these. This was found to be the case for the Cenozoic marine invertebrate literature compared to collections in the West

Coast of the USA (Marshall et al., 2018). However, these collections data can be made available and will be potentially important for future work, through increasing the known occurrences of taxa currently published as singletons, for example.

2.4.3 Paleogene lepidosaur richness and the diversification of major clades

The K–Pg boundary is heavily sampled due to its macroevolutionary importance, and many tetrapod lineages went extinct at this time (MacLeod et al., 1997). Lepidosaurs were no exception to this, as the large-bodied polyglyphanodontians (also known as borioteiioids) died out and caused a drop in 'global' richness across the boundary (Longrich, Bhullar and Gauthier, 2012b); Wilson (2013) found the same drop for larger-bodied mammals. This extinction is observable in the North American data, as polyglyphanodontians dominated assemblages for the entire Late Cretaceous, and there is a good Danian record in the region to observe faunal changes across the boundary. Other higher taxa remain relatively unscathed, however, and it has been proposed that polyglyphanodontians may have already been in decline prior to the extinction event (MacLeod et al., 1997). Faunal recovery is rapid, and by the end of the Danian a more modern squamate fauna was beginning to diversify. This can be seen in the continent-level curves at a quorum level of 0.3 (Figure 2.6), but North America is not well sampled enough in the later Paleocene for a sequential trend to be detected at a quorum of 0.4. The same is true of the other continents even at the lower quorum level, and so I cannot draw many conclusions regarding the effects of the K–Pg extinction on lepidosaur richness and recovery elsewhere.

After the sampling gap in the early Paleocene, there is a period of excellent sampling in both North America and Europe that reveals some interesting patterns. The Paleocene–Eocene Thermal Maximum (PETM) heralded a much warmer climate and opportunities for taxa to diversify into niches and areas that may have been previously thermally limiting to smaller ectothermic taxa (Head et al., 2009, 2013; Smith, 2011). This was followed quickly by the Early Eocene Climatic Optimum (EECO; e.g. Zachos et al., 2001). At this time richness in Europe was much higher than in the Campanian, with a multitude of localities yielding rich faunas of snakes and lizards. Iguanians appeared for the first time in Europe (Augé, 2005) and amphisbaenians were numerous by this time, having diversified at least as early as the Paleocene (Folie, Smith and Smith, 2013; Rage, 2013), and were found in most European localities despite their fossorial nature. Richness was at approximately the same level in North America; many Paleocene lineages persisted and richness was boosted by probable immigration from lineages that were previously restricted to more southerly environments (e.g. stem-polychrotid iguanians; Smith, 2009).

A large drop in richness occurred in North America during Pg3 (Lutetian). A similar drop is observed for Europe at quorum 0.3 (although not to the same extent; see Figure 2.6) but obviously this is unreliable. This may be linked to the end of the EECO, where temperatures began to decline (Zachos et al., 2001). In Europe, much of the middle Eocene is poorly sampled due to a lack of localities compared to the early Eocene. Some Lagerstätten occur at this time (Messel and Geiseltal, Germany), but it is difficult to compare fossils from these localities to those in other time bins due to extreme differences in preservation (Rage, 2013). Interestingly, amphisbaenians are not found even in localities with the special taphonomic conditions of Lagerstätten, suggesting a genuine absence (or very low richness) at this time (Augé, 2012) or perhaps a bias against fossorial animals. This differs from the findings of Mannion et al. (2015) for crocodylians, where European generic richness was relatively high in the middle Eocene, with a much lower richness in the late Eocene. This is also true of their North American data points, which are the inverse of what was observed for bins Pg2–4, with high richness in Pg3 and lower richness on either side of this bin.

There is a lack of squamate localities in central North America from the middle Eocene (Smith, 2011), which makes it difficult to track sequential faunal changes in this area. Woodburne et al. (Woodburne, Gunnell and Stucky, 2009) found a loss in richness for mammals after the EECO, although these data are uncorrected for sampling, and an associated loss of floral richness that was connected with increasing aridity during this period. The same may have been true for lizard genera, particularly for the more thermophilic taxa that supposedly immigrated into the area during the EECO (Smith, 2009), but studies of early and late Eocene localities in Wyoming show that many of these taxa seem to have persisted throughout the entire period (Smith, 2011). Squamates, therefore, do not appear to have been as affected as mammals (at least in central North America), but the middle Eocene sampling gap makes it difficult to tell if richness was consistent, or if taxa migrated away and then returned during the climatic amelioration in the latest Eocene (Zachos et al., 2001). The large drop in richness from Pg2–3 is almost certainly an artefact of sampling entirely different areas of North America during these two time bins, which SQS cannot account for (see Figure 2.11). The early Eocene record is comprised of localities mostly from Wyoming, whereas the middle Eocene squamate localities are Californian. It may be that richness was always proportionately lower in California than in Wyoming, but without good coeval data from both regions it is currently impossible to tell.

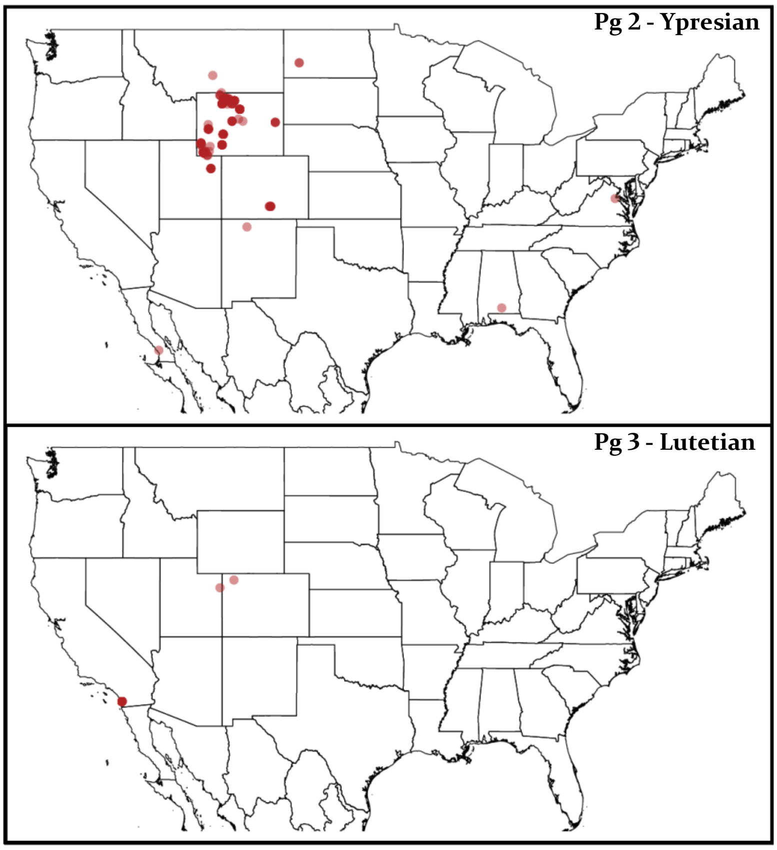

Figure 2.11: Distribution of North American lepidosaur localities in bins Pg2 and Pg3. Darker shades indicate layered localities.

Both Europe and North America had high generic richness in the late Eocene, in contrast to the lows in the previous interval. Richness in Europe was at its maximum during this period, with an incredible array of snakes, lizards and amphisbaenians from a wide range of localities. The rich fauna was composed of taxa that were already established in the region in combination with the first European appearance of cordyliforms, which are indicative of an environment that was still warm enough to support thermophilic taxa despite the overall cooling during the Eocene (Bolet and Evans, 2013). Amphisbaenians continued to diversify, and the first blanid (an extant taxon endemic to Europe) appeared in France at this time (Augé, 2012). A decline in small-bodied mammalian richness at this time might have allowed niche diversification for squamate taxa (Rage, 2012) and been a contributing factor underpinning this high richness. Late Eocene faunal richness in central North America (Wyoming) was very similar to that of the early Eocene, meaning that (central) North American squamates also continued to flourish.

The Eocene–Oligocene (E–O) boundary was marked by a sudden substantial drop in temperature, which caused increased seasonality and the onset of glaciation in Antarctica (Ivany, Patterson and Lohmann, 2000; Zachos et al., 2001; Liu et al., 2009), and is associated with a significant turnover event known as the Grande Coupure. The event was first recognised in the terrestrial realm for mammals (Stehlin, 1909; Hooker, 1989; Meng and McKenna, 1998; Costa et al., 2011; Eronen et al., 2015), but has been recognised subsequently for squamates (e.g. Rage and Augé, 1993; Augé and Smith, 2009; Rage, 2013) and pseudosuchians (Mannion et al., 2015). The event is believed to have been much more severe in Europe, as (combined with climate change) a major fall in eustatic sea level (Miller et al., 2005, 2008) connected a long-isolated Europe with Asia. This possibly allowed Asian taxa to immigrate into Europe and outcompete established taxa (Hooker, Collinson and Sille, 2004).

Squamate turnover may also have been exacerbated by the diversification of mammalian carnivorans (Augé and Smith, 2009). My data show a fall in richness across the boundary (Pg4–5) in both Europe and North America, which is slightly steeper in Europe, but richness remained relatively high afterwards when compared with that across the entire time span of the study.

In Europe, it is difficult to track any effects of the Grande Coupure in the literature at finer scales outside of the Quercy Phosphorites localities of southern France (although some unpublished work also exists for the Hampshire Basin of the UK; Green, 1998), which limits our understanding of when certain taxa disappeared from the continent as a whole. What is certain is that Oligocene faunas were markedly different from late Eocene faunas, with the exception of a small handful of survivors that did not persist for long after the boundary (except *Plesiolacerta*; Rage, 2013). Glyptosaurine lizards disappeared from Europe entirely and the modern snake family Colubridae appeared for the first time. Iguanians, once thought to have become extinct in Europe at the E–O boundary, appear in one locality in the late Oligocene of France (*Geiseltaliellus*; Augé and Pouit, 2012). Intense sampling of other Oligocene French localities has yet to produce further specimens, suggesting a genuine absence during the early Oligocene, but it is impossible to tell if taxa retreated to refugia in other parts of Europe due to the scarcity of localities in other regions. It is difficult to track the progress of Asian immigration also, since the number of eastern European and Asian localities from this time bin are limited. Richness remained relatively high in the continental curves after the E–O boundary, which could represent the rapid re-diversification of taxa (and the possible immigration of Asian taxa, as occurred for mammals) by the end Oligocene, filling vacant niches (Augé and Smith, 2009; Rage, 2012).

In North America the Grande Coupure event was previously thought to be negligible for lepidosaurs, but new finds in the late Eocene suggest that North American squamates also experienced a similar but less severe extinction at the end of the Eocene (Smith, 2006). This is consistent with what I found in the sampling-corrected curves, though the decline is not marked. The available data points may be within stochastic error of each other, as between the K7–K8 bins, but the quorum 0.5 curve still shows a similar decrease (unlike with K7–8 where the trend becomes shallower; see Figure 2.7). Many taxa with tropical affinities disappeared, but it is impossible to tell whether they migrated southwards or went extinct due to a lack of coeval localities from tropical North America (Smith, 2011). North American mammal taxa at this time apparently passed through the boundary relatively unscathed (Prothero, 1994), possibly due to the fact that mountain uplifting in the middle–late Eocene pre-adapted taxa to cooler climatic conditions (Eronen et al., 2015).

I found a rise in richness across the Eocene–Oligocene boundary for Asian taxa but, due to a very low number of localities in Pg4, data are not adequate to support robust patterns and so were excluded (see Figure 2.5). Previous studies of central Asian mammals suggest that they experienced a similar turnover to those in Europe and North America, with the transition from a warm temperate to an arid or semi-arid climate causing a change in the dominant faunas from large-bodied perissodactyls to smaller rodents and lagomorphs (Meng and McKenna, 1998; Sun et al., 2014). Less is known about squamate faunas over the boundary in this region, although it has been suggested that early Oligocene herpetofaunas featured many taxa analogous to modern arid-dwelling species, in comparison to late Eocene faunas that were more indicative of humid climatic conditions and mild winter temperatures (Böhme, 2007).

It is not possible to examine the effects of the Grande Coupure on squamate taxa from the Southern Hemisphere due to a lack of occurrences. Low richness in Africa and Oceania was recovered in the subsampled curves before and after the E–O boundary, respectively (Figure 2.5), but these data are too incomplete to reveal sequential patterns in taxon richness. It is important to note that many first occurrences of major lepidosaur taxa discussed above conflict greatly with the divergence dates derived from molecular phylogenies (e.g. Wiens and Lambert, 2014; Zheng and Wiens, 2016; Pyron, 2017). These phylogenies, while disagreeing on smaller-scale details, propose that crown group squamates began to diversify as early as the Middle Jurassic (Wiens and Lambert, 2014), with amphisbaenians diversifying approximately 100 Mya (mid-Cretaceous; Wiens and Lambert, 2014) or as early as 146 Mya (latest Jurassic; Zheng and Wiens, 2016). Snake phylogenies appear to provide better divergence estimates, with some studies placing their divergence date at approximately 125–130 Myr (Early Cretaceous; Zheng and Wiens, 2016), which is almost coincident with the earliest unequivocal snake fossil finds. Total evidence dating (Pyron, 2017), which attempts to reconcile molecular and morphological characters, comes closer to the divergence dates provided by the fossil record (e.g. diversification of Amphisbaenia at ~60 Myr), but still proposes large ghost lineages for other groups due to the lack of pre-Late Cretaceous material to calibrate these models. It is difficult to test these and other molecular-based estimates of diversification times and rates when the fossil record is so poor. So, at least for now, the origination times of major taxa remain unresolved.

2.4.4 Relationships between lepidosaur richness and potential driving variables

As mentioned above, it is difficult to select for the best model from my results, as AICc values for the best models are very similar. With regards to AICc weights, the TBCs + $\delta^{18}O$ palaeotemperature model is four times better than the TBCs only model. This is not a large difference compared to, for example, the difference between these two models and the null model (where palaeodiversity is constant and variation is explained by error), which has a weight of <0.001. While the R^2 value is improved by the addition of these extra explanatory variables (but still remains moderate, at 0.63 in the best model), the non-significant *p* values of every explanatory variable besides in all of the best models suggests that the sampling proxy (number of collections) is the important factor in our results. This finding is consistent with research on pterosaurs (Butler, Benson and Barrett, 2013), marine reptiles (Benson and Butler, 2011), sauropodomorph dinosaurs (Benson and Mannion, 2012), Late Cretaceous turtles (Nicholson et al., 2016) and dinosaurs as a whole (Mannion et al., 2012), which also found the number of collections as either the best explanatory variable, or the best in conjunction with non-marine area and/or the presence of Lagerstätten (which is another proxy for sampling biases). The low to intermediate R^2 values of all the best models (0.63 or lower; see Table 2.3), however, demonstrate that there is a good proportion of variance in the face-value generic data that remains unexplained by sampling alone. This suggests that other factors were also driving lepidosaur richness through time, perhaps biotic interactions that are difficult to test for in the fossil record, such as competitive exclusion or ecological responses to local environmental change, or other abiotic variables or evolutionary innovations that were not tested in this study.

2.4.5 Conclusions

Overall, I recovered some interesting richness trends from the Mesozoic–Paleogene lepidosaur fossil record, which have served to test the observations made by others while correcting for previously overlooked unequal sampling issues. Lepidosaur richness was low from the Triassic until the Late Cretaceous, with a few exceptions (particularly the Late Jurassic of Europe, which features numerous localities of exceptional preservation e.g. Cerin, Solnhofen and Guimarota). Richness became higher in the Late Cretaceous (Campanian–Maastrichtian), in North America, but was low in Indo-Madagascar. North America is well sampled enough that a decrease in richness across the K–Pg boundary is observable (when larger-bodied taxa were lost during the mass extinction) that remained at a moderately high level compared to earlier intervals.

In the Paleogene, both North America and Europe maintained a high level of richness that fluctuated with climate, except during the Lutetian (likely due to poor geographic spread of localities in North America; Europe is too poorly sampled at quorum 0.4). In Europe, richness was much higher at the end of the Paleogene than every other interval in the analysis. There was a fall in richness on both continents following the Grande Coupure at the Eocene–Oligocene boundary, though richness remained moderately high in North America and very high in Europe moving towards the Neogene. This implies a relatively late acquisition of the incredible lepidosaur richness that we observe today, though we are lacking much of the data that would help us understand how they became so diverse, particularly in the tropics.

Unfortunately, the lepidosaur record suffers from many issues related to sampling biases, particularly geographic biases, which obscure sequential patterns in diversity. Rhynchocephalians disappeared from the Northern Hemisphere in the

Early Cretaceous, and their record (and the Southern Hemisphere record) is so poor post-Cretaceous that we know little about the processes that restricted them to the limited distribution they occupy today. It is also impossible to pinpoint the geographic origins of lizards or snakes, aside from inferences based on their relative early abundances in Laurasia and Gondwana, respectively (Evans, 2003), as Triassic to Early Cretaceous records are currently too poor for most continents. Even in time bins with relatively good sampling there are periods when locally excellent records exist but which cover only a small proportion of the total land area available (e.g. Western Europe, Western Interior of North America). Although it is tempting to extrapolate the signals generated from these unevenly distributed data to large-scale macroevolutionary patterns, these results demonstrate that it is crucial to examine these data at regional and local levels to avoid generating false hypotheses. However, until more data become available it behoves us to refine the signals we can extract from the data we have available in order to try and reconstruct rigorously tested patterns of past diversity, as these form useful benchmarks for understanding the origin and evolution of the modern biota. In particular, we must continue to look for new fossiliferous localities worldwide, particularly in the Southern Hemisphere, in order to improve our data and the robustness of our conclusions.

This chapter was published as: Cleary, T.J., Benson, R.B., Evans, S.E. and Barrett, P.M. (2018) 'Lepidosaurian diversity in the Mesozoic–Palaeogene: the potential roles of sampling biases and environmental drivers', *Royal Society open science*, 5(3), pp. 171830.

[This page intentionally left blank]

3. Non-marine Turtle Richness from the Triassic–Paleogene

3.1 Introduction

Testudinata, consisting of terrapins, tortoises and turtles (collectively referred to as 'turtles' herein), is a unique group of reptiles spanning from the Late Triassic (Joyce, 2017) to the modern day. Today, it is a relatively small clade compared to the extant Lepidosauria, with approximately 90 genera containing 350 species (Uetz, Freed and Hošek, 2018) and is divided into two major lineages, Cryptodira (hidden neck turtles) and Pleurodira (side neck turtles). Extant pleurodires are found only in the Southern Hemisphere, and are restricted to aquatic or semi-aquatic lifestyles. Cryptodires, on the other hand, are common in both hemispheres and occupy a range of aquatic and terrestrial niches, including the marine realm (Pough, Janis and Heiser, 2013). Both extant lineages extend back into the Jurassic, alongside a variety of stem-taxa that diverged prior to the cryptodire/pleurodire split (Joyce, 2017), with a rich and widespread fossil record. Turtle origins are heavily debated, with multiple theories on origins and which group turtles split from (see Joyce (2015) for a good summary).

Turtles have remained an important component in non-marine vertebrate faunas from their origins until today and are a key clade for studying the effects of current climate change on extant vertebrates (e.g. Araújo, Thuiller and Pearson, 2006; Foufopoulos, Kilpatrick and Ives, 2011; Urban, Richardson and Freidenfelds, 2014). In addition, many modern reptiles (including turtles) have gender ratios determined by temperature, and these are expected to be heavily affected by warming (Bickford et al., 2010). The ability of individual taxa to cope with rapid climate change varies, but ectotherms appear to have relatively poor plasticity with regards to thermal tolerances (Gunderson and Stillman, 2015).

As mentioned in Chapter 1, there have been a number of studies that have focused on face-value observations of richness of turtles at key moments in time, with a focus on major extinction events or climatic changes (e.g. Hutchison and Archibald, 1986; Hutchison, 1992; Hirayama, Brinkman and Danilov, 2000; Holroyd, Hutchison and Strait, 2001; Corsini, Smith and Leite, 2011; Holroyd, Wilson and Hutchison, 2014). Besides inclusion in whole-tetrapod analyses, only a few studies have examined the turtle fossil record using methods that account for or alleviate sampling biases, both for Mesozoic non-marine turtles (Nicholson et al., 2015, 2016).

Here, I investigate non-marine turtle richness through time using an expanded, updated dataset for non-marine turtle distribution and richness that extends from the Late Triassic–Paleogene (237–23 Ma). This expanded interval allows more detailed comparisons with the other long-term datasets that have been generated for other non-marine tetrapod clades (e.g. Newham et al., 2014; Mannion et al., 2015; Cleary et al., 2018; Chapter 2) in order to detect similarities in patterns, which might have common causes, or notable differences, which might indicate clade-specific responses to long term environmental change. I generate sampling-corrected taxon-richness curves and discuss some of the driving factors that may influence these richness patterns, assessing problems with the turtle fossil record and the fossil record as a whole.

3.2 Methods

3.2.1 Data

Testudinate occurrences from the Triassic–Paleogene were downloaded from the Paleobiology Database (PBDB; www.paleobiodb.org), accessed via Fossilworks (www.fossilworks.org) on 30 October 2017. Details of individual contributions to the database are listed in Appendix 7. After download, data were filtered to remove marine and ichnofossil occurrences as well as those not identifiable to the generic level, and time bins were standardised (see Appendix 7 for script). The resultant dataset consists of 2666 occurrences that represent 295 genera.

I recognise that there are problems with using genera rather than species, particularly with wastebasket taxa that have been used widely in the turtle fossil record (e.g. "*Trionyx*"). However, species taxonomy is also muddled in many clades, but genera can be assigned with confidence in many assemblages (e.g. Holroyd and Hutchison, 2002). For the latter reason, I include genus-level occurrences that are specifically indeterminate (over 1000 occurrences).

3.2.2 Subsampling

I used Shareholder Quorum Subsampling (SQS; Alroy, 2010a, 2010b) to produce subsampled richness estimate curves for non-marine turtles at global and continental levels. SQS uses taxon occurrence frequencies rather than absolute counts of occurrences, to determine when the subsampling quota (= quorum) has been met. These frequencies are the proportion of total occurrences that each taxon accounts for in a time bin. Occurrences are sampled until a desired summed frequency quorum is met, with each taxon that is drawn contributing its frequency only once per bin. This would result in fair and comparable sampling of the abundance-frequency distribution only if all intervals/regions were equally-sampled to being with. However, the observed abundance-frequency distribution represents only a subset of the true distribution. Therefore, Good's u, the proportion of single-occurrence taxa ('singletons') is used to estimate the coverage of a time bin's dataset (i.e., how much of the 'true' underlying richness has been recovered). By dividing a chosen quorum by each time bin's value of u, one can estimate the 'true' coverage of each bin and subsample more intensively when coverage is low (for further information on this, see Alroy, 2010a). Consequently, if a time bin's estimated coverage is lower than the chosen quorum level, then that bin will be excluded from the analysis as the quorum cannot be met.

SQS was performed using the v. 4.3 Perl script (see Appendix 3), on global and continent-level turtle occurrences at quorum levels of 0.4–0.7. Quorum level 0.4 is considered to be the lowest level at which a sufficient level of standing diversity can be recovered (Alroy, 2010b; Benson et al., 2016). Time bins represent the same merged stages as in Chapter 2 (Table 2.1). Occurrences were only included in the analyses if their stratigraphic ages were known with sufficient certainty to be assigned to a single time bin. Continental exclusions ranged from less than 10% (Indo-Madagascar, North America, Europe) to very high exclusions due to high stratigraphic uncertainty (Asia, 41%; Oceania, 67%, but only had 12 collections total). To account for the broad or poorly known dating of some Asian localities, I also included individual points for the Jehol (plotted between bins K2 and K3) and Nemegt (plotted between K7 and K8) biotas. The countries included in my continental analyses are listed in Table 3.1.

Table 3.1: List of continent assignments for the countries in my dataset. Russia is included twice due to its large geographic spread, and the data was edited to reflect this, in order to separate out the localities closer to Europe or Asia.

Continent	Included countries
Africa	Algeria, Angola, Cameroon, Egypt, Eritrea, Ethiopia, Kenya, Lesotho, Libya, Malawi, Mali, Morocco, Namibia, Niger, Senegal, Somalia, South Africa, Sudan, Tanzania, Tunisia, Zambia, Zimbabwe
Asia	China, Georgia, Japan, Kazakhstan, Kyrgyzstan, Laos, Lebanon, Mongolia, Myanmar, North Korea, Oman, Pakistan, Palestinian Territory, Russian Federation, South Korea, Syria, Tajikistan, Thailand, Uzbekistan
Europe	Austria, Belgium, Croatia, Czech Republic, Denmark, France, Germany, Hungary, Italy, Luxembourg, Macedonia, Netherlands, Norway, Poland, Portugal, Romania, Russian Federation EU, Slovenia, Spain, Sweden, Switzerland, Ukraine, United Kingdom
Indo-Madagascar	India, Madagascar
North America	United States, Canada, Mexico
Oceania	Australia, New Zealand
South America	Argentina, Brazil, Bolivia, Chile, Colombia, Peru, Uruguay

In order to avoid problems with small sample sizes when using Good's *u*, points generated from fewer than five collections were excluded from plots, except for one figure with these points added back in for reference. All SQS curves are annotated with the numbers of collections that contribute to each point to display underlying data quality.

I also used classical rarefaction (CR), which uses absolute counts of occurrences rather than coverage when subsampling. Using absolute counts with a set quota ('uniform sampling') has been criticised as this can dampen genuine diversity signals (Alroy, 2010b). Nevertheless, I use this method as a comparison to SQS as in Chapter 2 in order to see if results vary due to the application of different subsampling techniques.

As a visual aid to my discussion of geographic biases in particular, I plotted the palaeolatitudinal distributions of every occurrence used in my analyses based on major clade (Cryptodira and Pleurodira, Paracryptodira, and stem taxa). Palaeolatitudes were calculated by the PBDB and Fossilworks using tectonic plate rotation data from the GPlates model (Wright et al., 2013).

3.2.3 Comparing richness to proxies for sampling and environmental factors

To examine relationships between face-value occurrence data and proxies for sampling and environmental variables I used generalized least-squares regressions (GLS). Sampling proxies considered herein include the number of non-marine tetrapod-bearing collections (TBCs) and tetrapod-bearing formations (TBFs), which were calculated from the same PBDB tetrapod data as in Chapter 2 (61316 occurrences downloaded on 10 March 2017; see Appendix 4). Environmental proxies consist of sea-level (Miller et al., 2005), non-marine area (Smith, Smith and Funnell, 2004) and palaeotemperature ($\delta^{18}O$ record; Prokoph, Shields and Veizer, 2008 via Mannion et al., 2015). I interpolated palaeotemperature data into 0.1 Ma intervals to facilitate subsequent aggregation into time bin averages (see Appendix 5). Negative $\delta^{18}O$ values indicate warmer temperatures and vice versa. As in Chapter 2 these data were only available from the Jurassic onwards, so my GLS analyses include bins J4 to Pg5 (Table 2.1). I added time bin duration as a non-optional variable to determine if it was influencing my model outcomes, following Marx and Uhen (2010).

For GLS I fitted autoregressive (AR) models of order one to all possible combinations of the above proxies and compared the models using Akaike's information criterion for small sample sizes (AICc). I calculate Akaike weights to identify the best model from those tested; AIC weights reward goodness of fit but penalise a higher number of variables in a model (Burnham and Anderson, 2002). I manually calculated R2 values using the method of Nagelkerke (1991), which provides an estimate of how much variance in raw richness is explained by each model. Sampling proxies and raw richness were ln-transformed prior to analyses to ensure homoskedasticity and normality of the models' residuals; Jarque-Bera (Jarque and Bera, 1980) and Breusch-Pagan (Breusch and Pagan, 1979) tests were used to check the residuals after results were generated. All analyses were carried out in R v. 3.4.3, using the packages lmtest v. 0.9–35 (Hothorn et al., 2017), nlme v. 3.1–131 (Pinheiro et al., 2017), qpcR v. 1.4–0 (Spiess and Ritz, 2014) and tseries v. 0.10–40 (Trapletti and Hornik, 2017), following the approach of Benson & Mannion (2012). GLS script can be found in Appendix 6.

3.3 Results

3.3.1 Global face-value and subsampled curves

Figure 3.1 shows uncorrected richness compared against subsampled richness using SQS at quorum 0.4. In the uncorrected curve richness increased steadily from a very low level in the Late Triassic until the mid-Cretaceous, where it sharply rose until the end of the period. There was a sharp decline over the Cretaceous-Paleogene (K-Pg) boundary followed by a sequential pattern of peaks and troughs in the Paleogene. Richness remained high compared to pre-Late Cretaceous levels. The subsampled curve has a number of marked differences, including a peak in richness in the Early Cretaceous and in bin K6 (Turonian–Santonian). There was an increase in richness across the K-Pg boundary, and while the peaks and troughs remain in the Paleogene their scale is uneven compared to the uncorrected curve. The time bins J2, J3 and K3 are too poorly sampled to be recovered at quorum 0.4 and so do not appear on the figure. At higher quorum levels the latest Cretaceous (K7–K8) and late Eocene (Pg4) had the highest peaks in richness (Figure 3.2). Classical rarefaction results are very similar to SQS results but the signal is dampened (Figure 3.3).

3.3.2 Continental-level subsampling

Subsampled richness for individual continents at quora of 0.4-0.6 (Figure 3.4–3.6) indicates substantial heterogeneity among regions. Due to poor sampling between continents, robust estimates of regional richness cannot be recovered prior to the Late Jurassic even at a quorum of 0.4. The only exception to this is the recovery of low richness in the Late Triassic of Europe. Additionally, of the Southern Hemisphere continents, subsampled richness estimates can only be recovered for South America, and there only for a few time intervals.

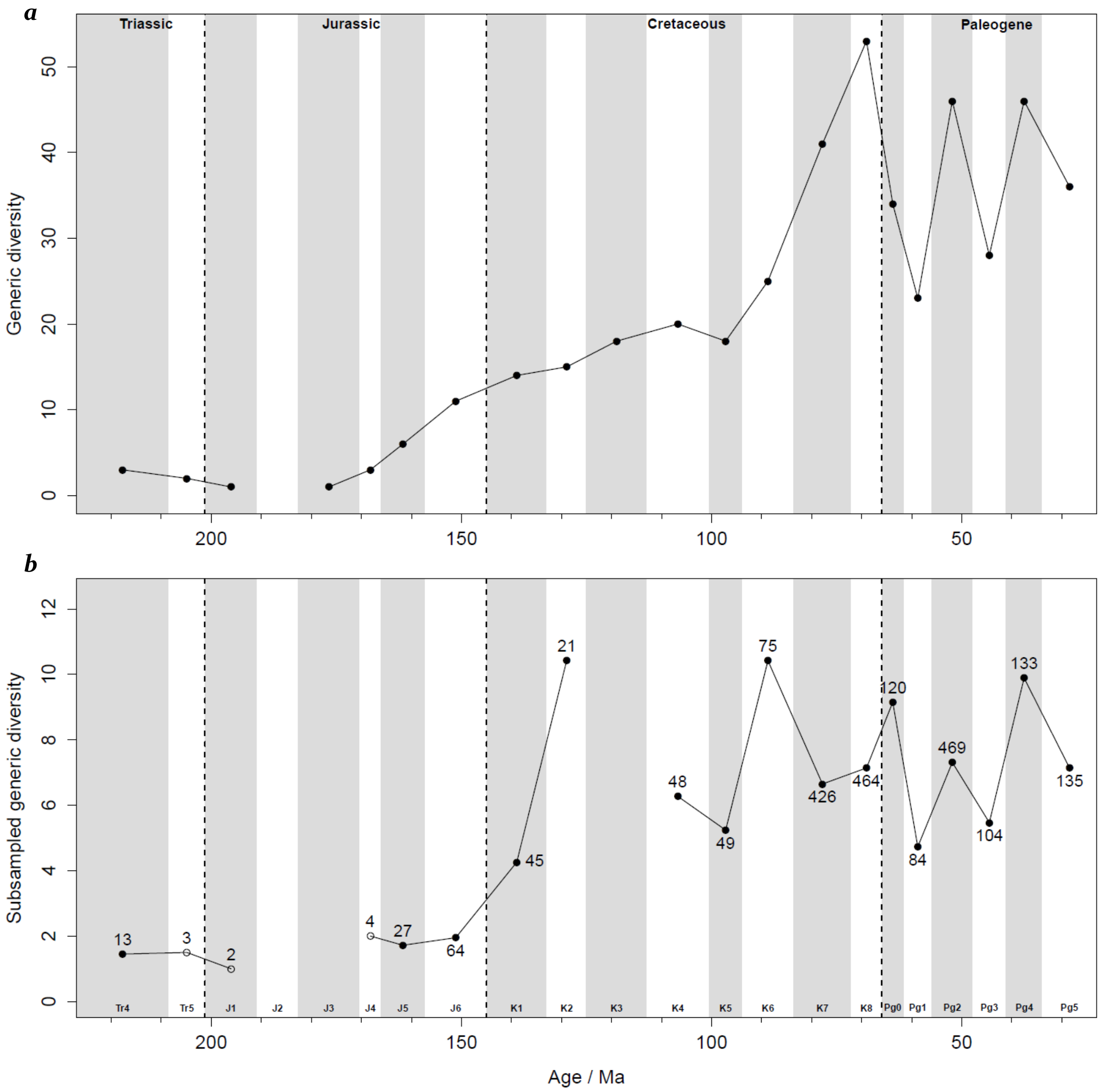

Figure 3.1: Non-marine turtle global generic richness from Triassic–Paleogene using *(a)* uncorrected (face-value) richness and *(b)* subsampled (SQS) richness at quorum 0.4. Numbers on figure 1*b* indicate number of collections drawn for each time bin subsampled, as an indicator of underlying data quality. T, Triassic; J, Jurassic; K, Cretaceous; Pg, Paleogene.

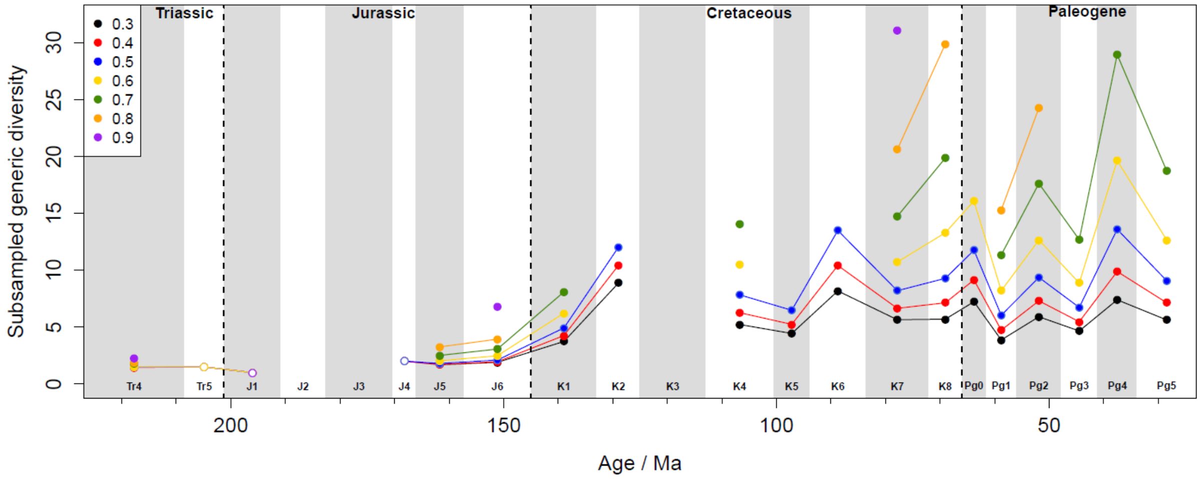

Figure 3.2: Subsampled global non-marine turtle generic richness from Triassic–Paleogene at quorum levels 0.3–0.9. T, Triassic; J, Jurassic; K, Cretaceous; Pg, Paleogene.

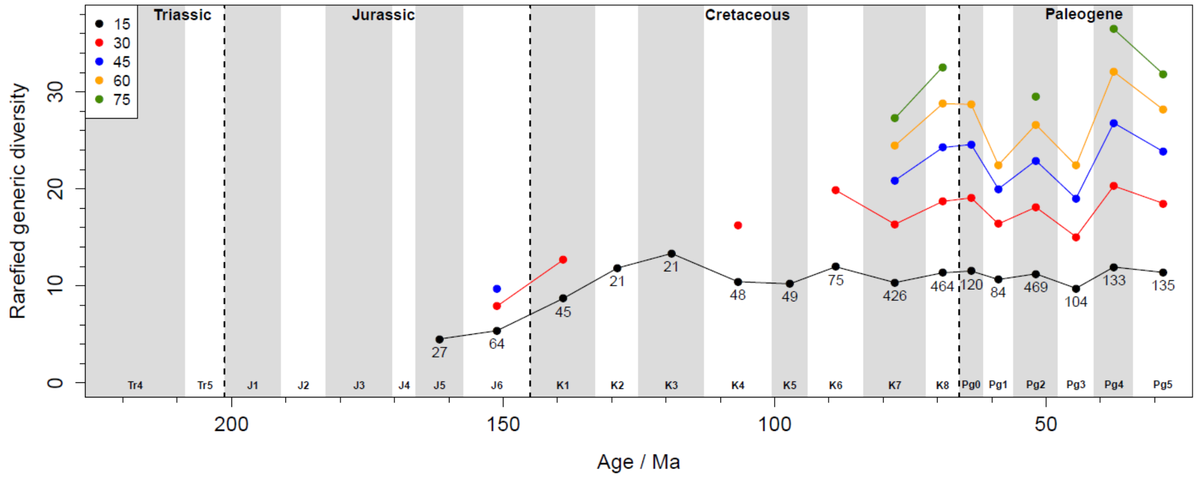

Figure 3.3: Rarefied global non-marine turtle generic richness from Triassic–Paleogene at quotas of 15–75 in intervals of 15. Numbers indicate number of collections drawn for each time bin subsampled, as an indicator of underlying data quality. T, Triassic; J, Jurassic; K, Cretaceous; Pg, Paleogene.

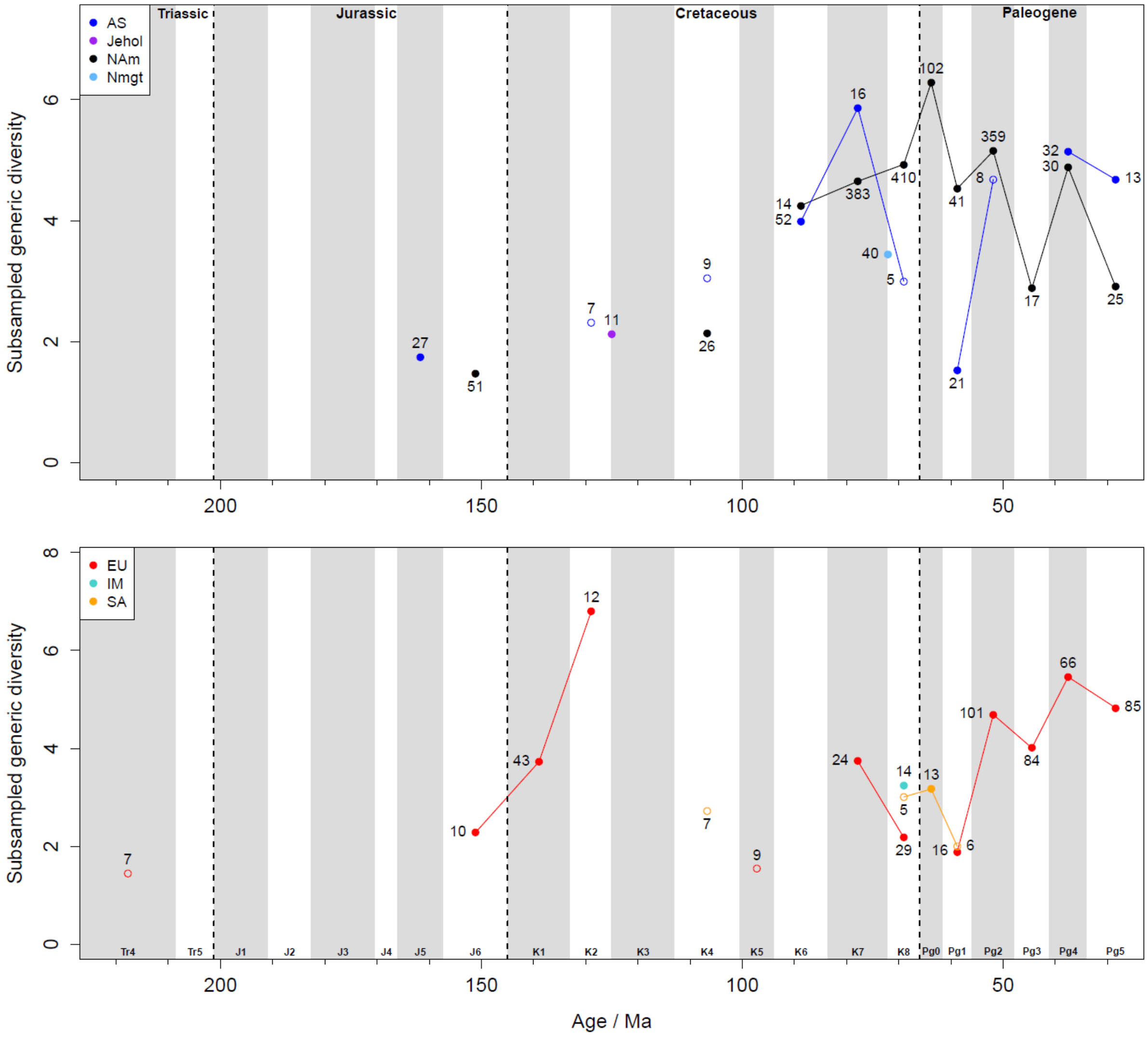

Figure 3.4: Subsampled non-marine turtle generic richness from Triassic–Paleogene at quorum 0.4, for individual continents: AS, Asia; EU, Europe; IM, Indo-Madagascar; NAm, North America; SA, South America. Also included are single points representing the Jehol Group and Nemegt (Nmgt) Formation. Numbers indicate number of collections drawn for each time bin subsampled, as an indicator of underlying data quality. T, Triassic; J, Jurassic; K, Cretaceous; Pg, Paleogene.

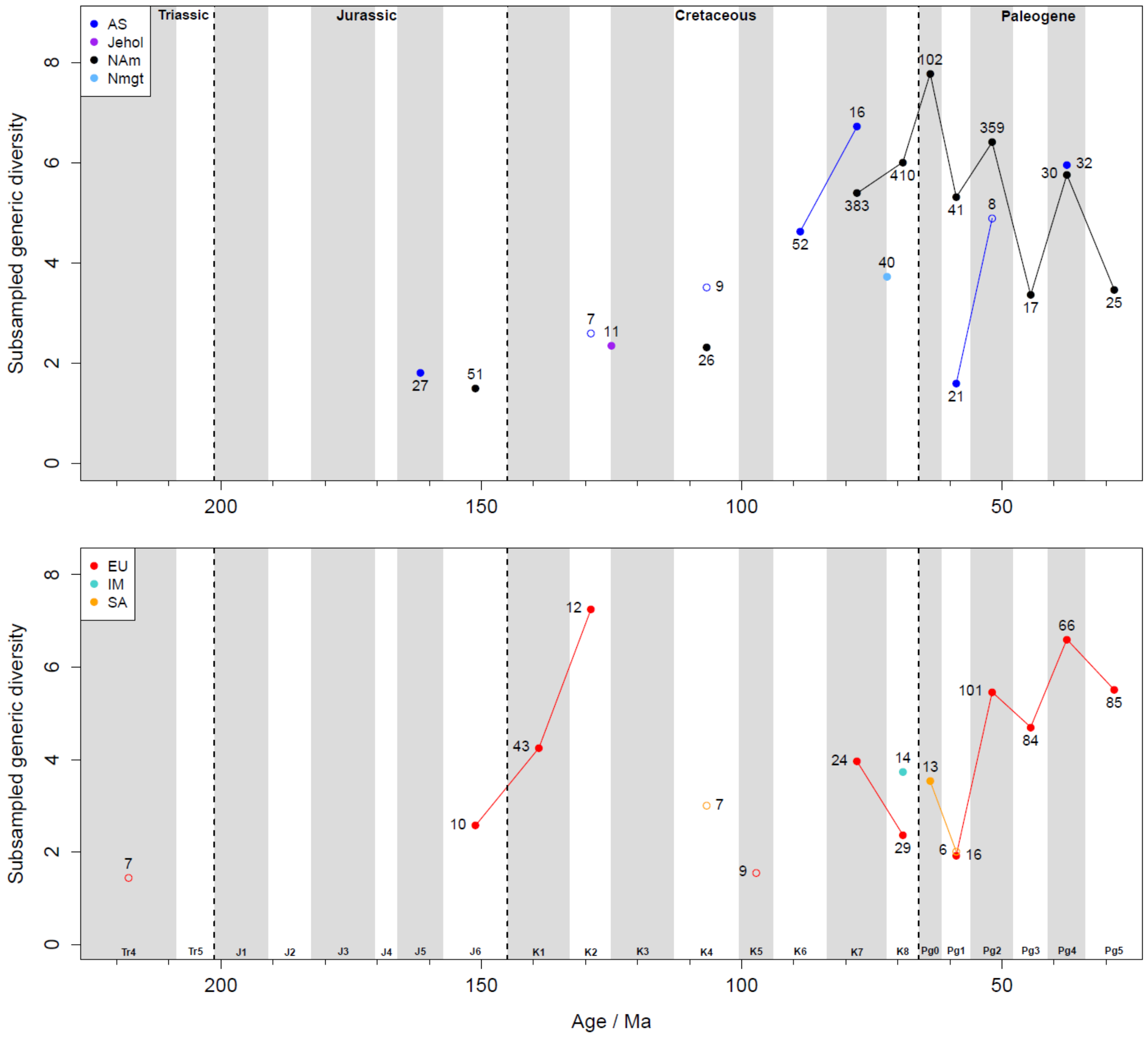

Figure 3.5: Subsampled non-marine turtle generic richness from Triassic–Paleogene at quorum 0.5 for individual continents: AS, Asia; EU, Europe; IM, Indo-Madagascar; NAm, North America; SA, South America. Also included are single points representing the Jehol Group and Nemegt (Nmgt) Formation. Numbers indicate number of collections drawn for each time bin subsampled, as an indicator of underlying data quality. T, Triassic; J, Jurassic; K, Cretaceous; Pg, Paleogene.

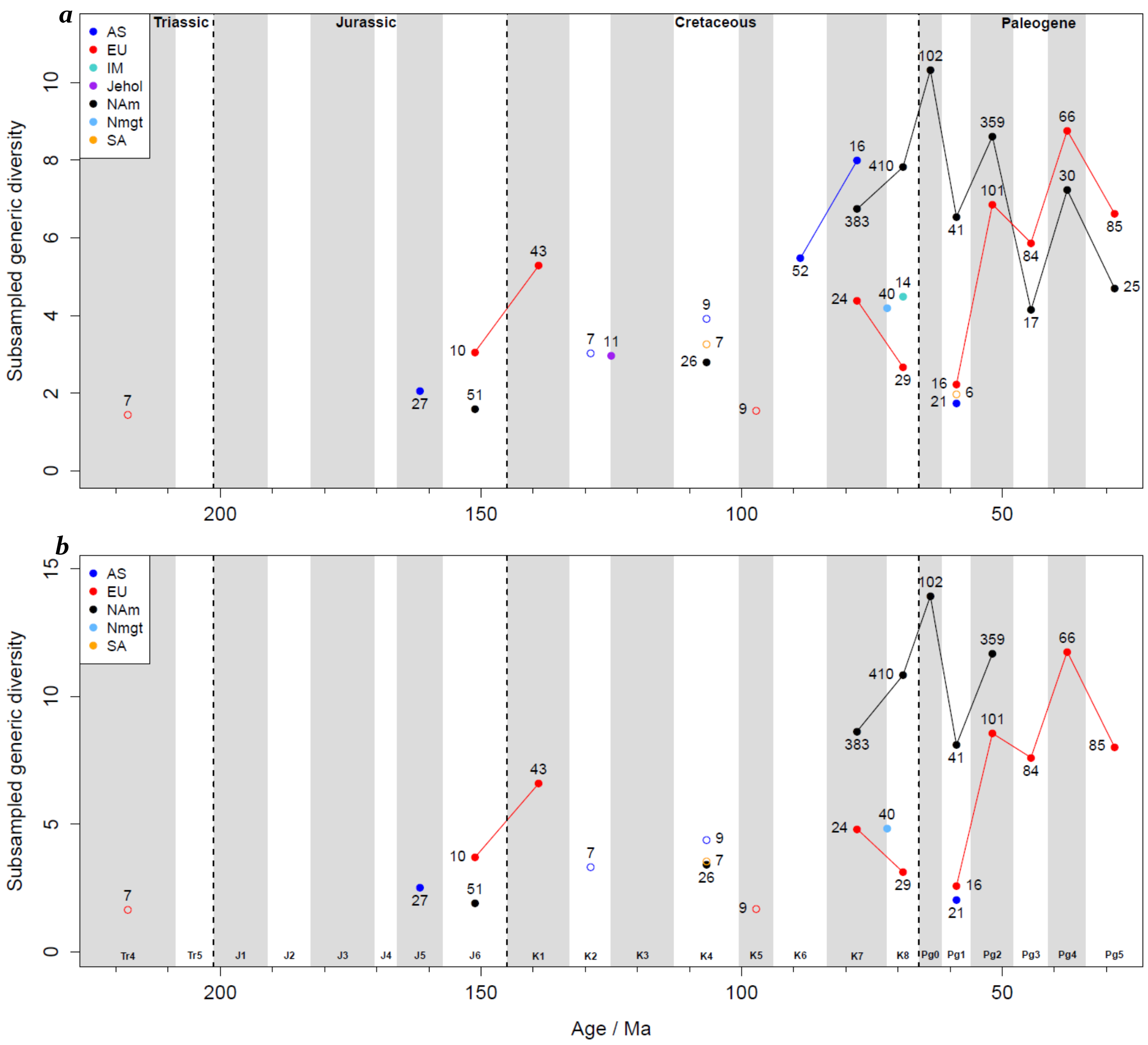

Figure 3.6: Subsampled non-marine turtle generic richness from Triassic–Paleogene at quorum (*a*) 0.6 and (*b*) 0.7 for individual continents: AS, Asia; EU, Europe; IM, Indo-Madagascar; NAm, North America; SA, South America. Also included are single points representing the Jehol Group and Nemegt (Nmgt) Formation. Numbers indicate number of collections drawn for each time bin subsampled, as an indicator of underlying data quality. T, Triassic; J, Jurassic; K, Cretaceous; Pg, Paleogene.

In the Late Jurassic, richness was still low in Europe, Asia and North America (Figure 3.4). Over the J–K boundary and into the Early Cretaceous, however, richness increased sharply in Europe to its highest level in the entire study interval (in bin K2). After this point, all continental level curves are punctuated by a large gap, which is populated by isolated and mostly uninformative points until the Late Cretaceous. The occurrences from the Jehol Group that fit into a single time bin contribute a large portion of the richness of Asia in the late Early Cretaceous, and subsampling the biota alone suggests that it was characterised by relatively low richness compared to K2 Europe. Richness estimates can also be made for North America and South America in bin K4 (Albian; although the estimate for South America is based on data from only seven localities), and also indicate relatively low richness. Overall, these findings suggest substantial regional variation in the genus richness of turtles in the Early Cretaceous.

In the Late Cretaceous there are a variety of richness patterns among data points representing North America, Asia, Europe, South America and Indo-Madagascar. In Asia, richness increased from the Turonian–Santonian (K6) to a peak in the Campanian (K7), and then decreased in the final bin of the Cretaceous (K8, Maastrichtian; although this estimate is based on data from only five localities). The well-sampled Nemegt Formation had a similar level richness to that of the Maastrichtian (K8) of Asia, which was medium compared to North America. On this continent there was a steady, high and increasing richness across the late Late Cretaceous (K6–K8). In Europe, sampling is too poor to recover K6, but in K7–K8 it had a low and decreasing richness. In Indo-Madagascar and South America there was also a low Maastrichtian (K8) richness similar to Europe.

Across the K-Pg boundary subsampled richness increases among North American turtles. Richness appears to have neither increased nor decreased in South America but cannot be estimated at a quorum greater than 0.4 (Figure 3.5, 3.6). Subsampled North American turtle richness was at an all-time high in Pg0 (Danian). The richness of other regions cannot be estimated during this time. Richness declined from the early to late Paleocene (Pg0 to Pg1, Danian–Thanetian) in both North America and South America. Turtle richness in Pg1 of Asia and Europe was low, comparable to its lowest levels of the Late Jurassic–Cretaceous, and in striking contrast to the high richness in Pg0 of North America.

A return to high richness levels occurred in Pg2 (Ypresian) in North America, Europe and Asia. These continents had roughly comparable richness at this time, although Europe's was slightly lower. Richness cannot be estimated for other continents. North America, Asia and Europe exhibited similar richness patterns of alternating increase and decrease during the Paleogene, with peaks in richness during Pg2 and Pg4 (Ypresian and Bartonian–Priabonian). The absolute values of richness attained in Europe and Asia were especially similar, whereas North America showed particularly deep declines in richness during Pg3 (Lutetian) and Pg5 (Oligocene). At the Eocene–Oligocene (E-O) boundary there was a fall in richness for all three Northern Hemisphere continents, with the drop in North America appearing to be steeper than the others. The approximate concordance among patterns in North America, Asia and Europe from Pg2 (Ypresian) until at least the Pg5 (Oligocene) contrasts with their idiosyncratic regional patterns of diversity during the Cretaceous.

At a quorum level of 0.5 (Figure 3.5) the results are similar to those at 0.4, but with a few key differences. First, the higher quorum means those intervals that still yield signal change in scale as I am sampling more of what is present. In this case this means that the peak in North American richness in the Paleocene (Pg0) was now the highest richness overall, just surpassing that in the K2 bin of Europe. Second, some bins that are too poorly sampled to be recovered at quorum 0.5 drop out. These bins are K6 of North America, K8 and Pg5 of Asia and K8 of South America. At quorum 0.6 (Figure 3.6) many more bins drop out: K2 of Europe (eliminating the large Early Cretaceous peak), Pg2 and Pg4 of Asia (removing its record after the Paleocene), the Jehol Group, and Pg0 of South America (meaning that I recover few meaningful results from the Southern Hemisphere at this quorum level). At quorum 0.7 (Figure 3.6) the Late Cretaceous of Asia drops out, as does K8 Indo-Madagascar and North America after Pg2.

Rarefaction recovered similar trends (Figure 3.7), but even at very low quota levels many of the bins I recovered with SQS cannot be calculated; all Southern Hemisphere points are missing. In addition, the trends that are identified are much flatter than those found in my SQS analyses.

I also looked at the major taxonomic split in turtles (Cryptodira and Pleurodira) to examine differences in the richness of these clades through time (Figure 3.8, 3.9). These clades were assigned to taxa manually by Pat Holroyd (pers. comm.; see Appendix 7 for code and original data assignment file). Pleurodira, which today is less species-rich and confined to the Southern Hemisphere, is predictably more poorly sampled, while cryptodires contribute most of the richness record for non-marine turtles. The Paleogene part of the North American genus richness curve and the entire Asian curve appear to have been driven almost entirely by cryptodires, while the South American and Indo-Madagascan data were dominated by pleurodires. The European record was a mix of both clades, which shared the same trends in varying magnitudes except at the E-O boundary, where cryptodiran richness increased in comparison to the fall in richness for all turtles.

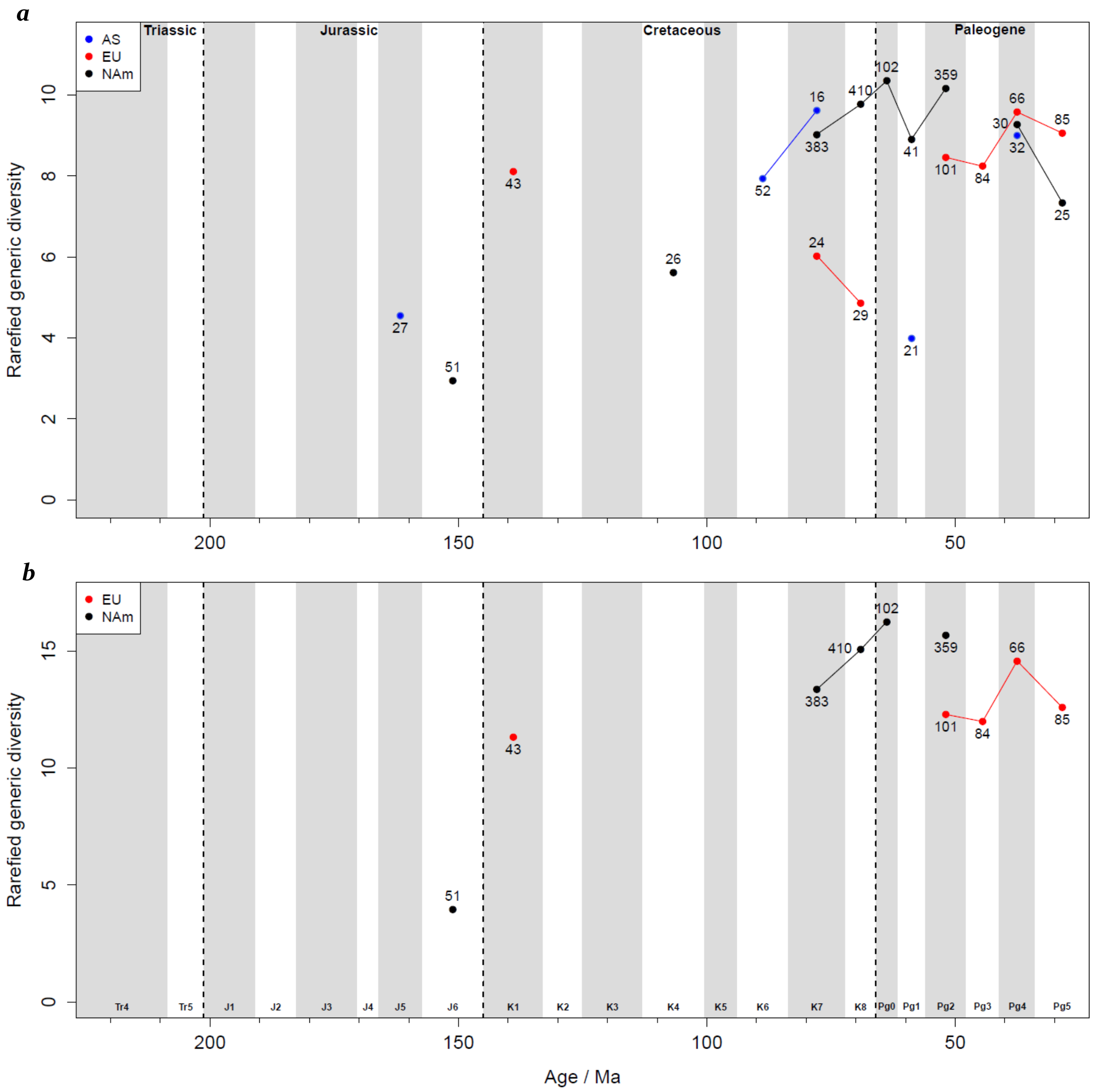

Figure 3.7: Rarefied non-marine turtle global generic richness from Triassic–Paleogene using a quota of *(a)* 15 and *(b)* 30 for individual continents. AS, Asia; EU, Europe; NAm, North America. Numbers indicate number of collections drawn for each time bin subsampled, as an indicator of underlying data quality. T, Triassic; J, Jurassic; K, Cretaceous; Pg, Paleogene.

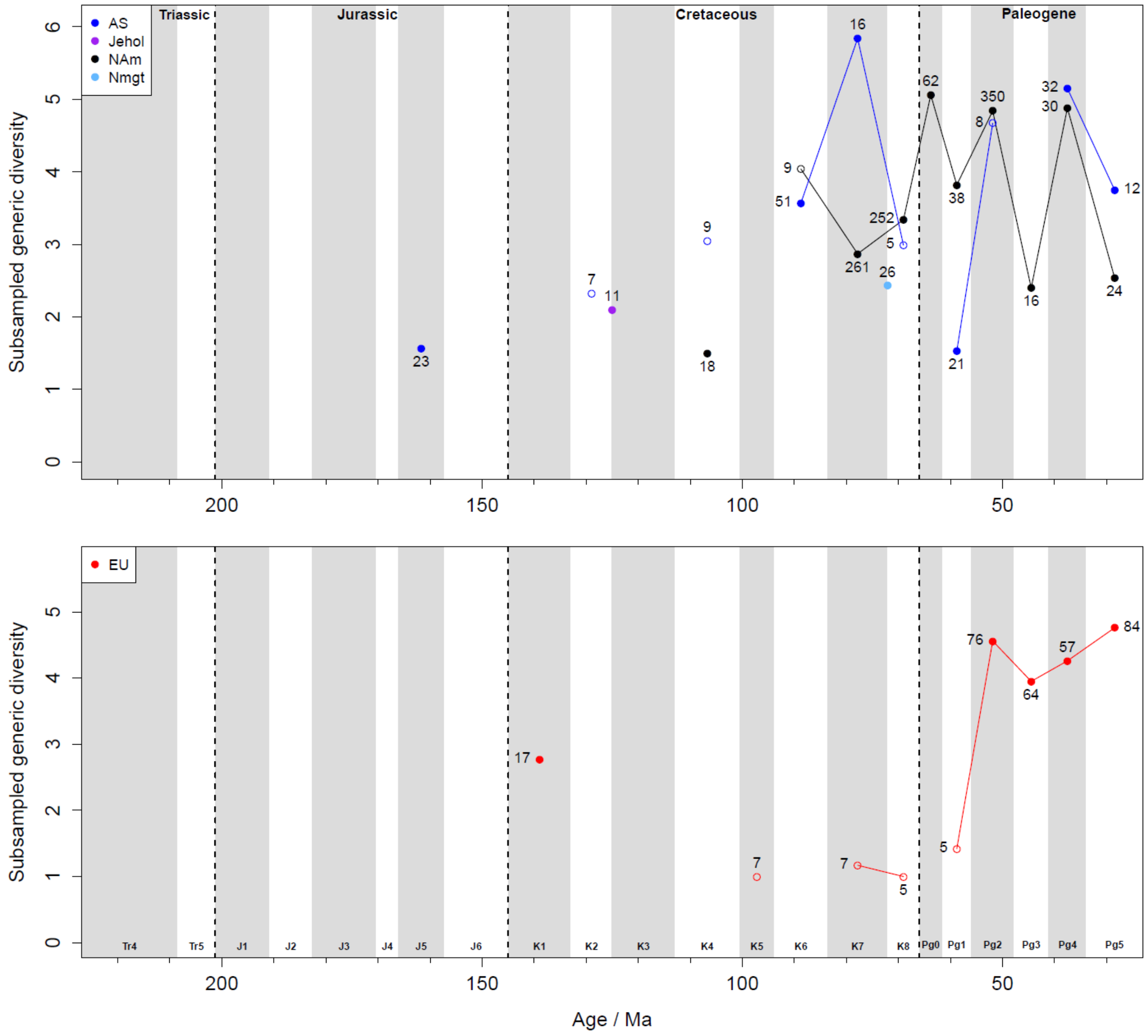

Figure 3.8: Subsampled generic richness at quorum 0.4 for Cryptodira. AS, Asia; EU, Europe; NAm, North America. Also included are single points representing the Jehol Group and Nemegt (Nmgt) Formation. Numbers indicate number of collections drawn for each time bin subsampled, as an indicator of underlying data quality. T, Triassic; J, Jurassic; K, Cretaceous; Pg, Paleogene.

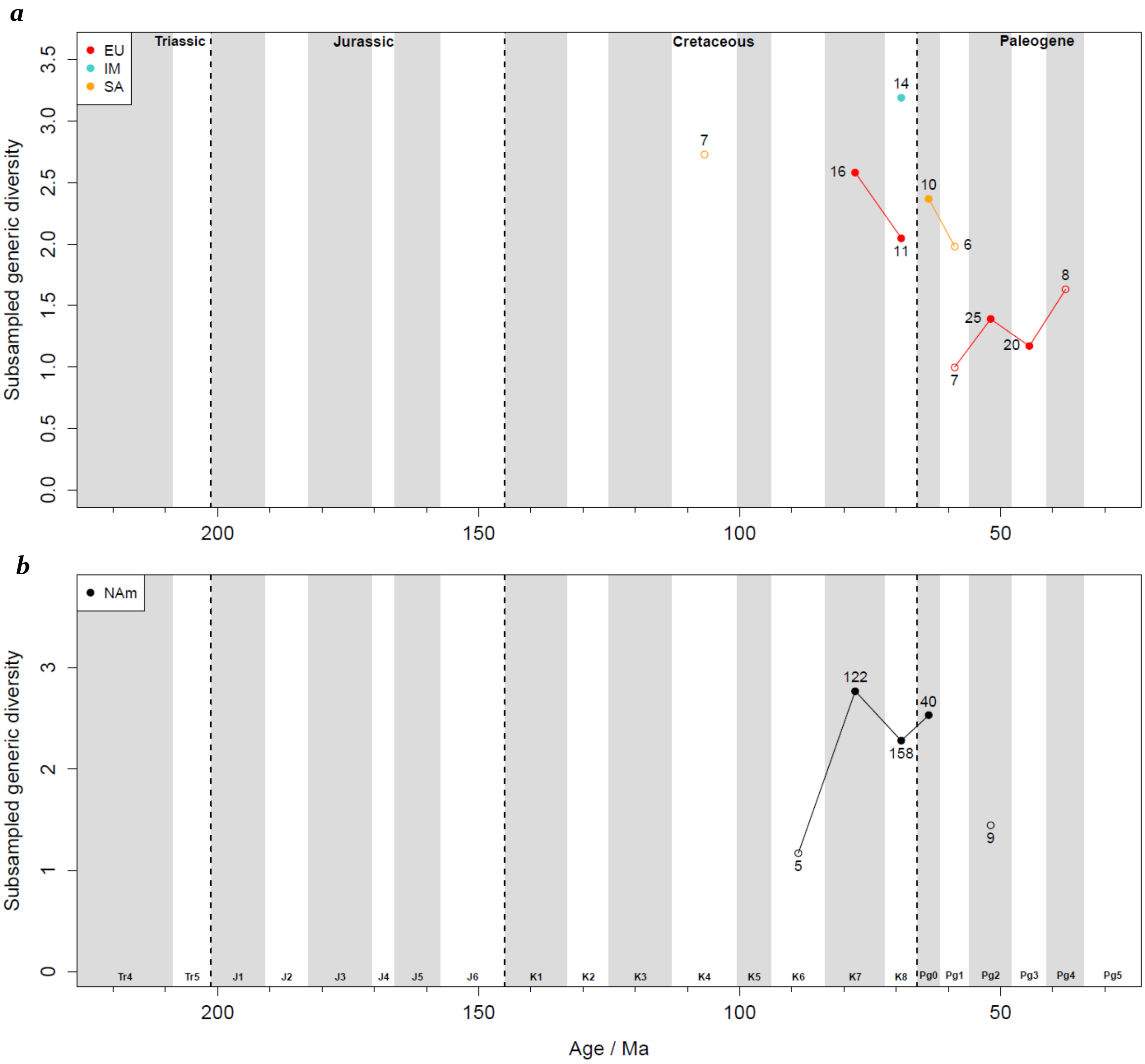

Figure 3.9: Subsampled generic richness at quorum 0.4 for (*a*) Pleurodira and (*b*) Paracryptodira. EU, Europe; IM, Indo-Madagascar; NAm, North America; SA, South America. Numbers indicate number of collections drawn for each time bin subsampled, as an indicator of underlying data quality. T, Triassic; J, Jurassic; K, Cretaceous; Pg, Paleogene.

3.3.3 Generalized least-squares analyses

Of the various combinations of explanatory variables that I tested with generalized least-squares regression (GLS; Table 3.2), only tetrapod-bearing collections (TBCs) were found to be a better fit than the null model (AICc; 18.94 versus 23.60). The tetrapod-bearing formations (TBFs) model trailed just behind the null (AICc = 23.84). The other models including non-marine area, sea level, and temperature had much lower AIC weights and so do not appear to have any clear influence on richness. Bin duration was not significant in any of the models and did not change the outcome of the model rankings, and so the analyses using this as an extra compulsory variable are presented in Appendix 7. While the TBCs-only model is recovered as the best model, with a significant p value ($p < 0.01$), the model's R2 value is low (0.35), and so there are likely other variables that were not tested in these analyses that contribute towards driving turtle generic diversity.

Table 3.2: Summary of various model fits to observed (face-value) generic richness of non-marine turtles, ordered by AICc score. N= 17; “TBCs”, tetrapod-bearing collection; “TBFs”, tetrapod-bearing formations; “δ^{18}O” is the palaeotemperature proxy. LL, log-likelihood.

Model name	**Sampling (TBCs or TBFs)**			**Sea level (SL)**			**Non-marine area (NMA)**			**δ^{18}O**			**R^{2}**	**LL**	**AICc**	**AICc weight**
	slope	*t*-value	*p*	slope	*t*-value	*p*	slope	*t*-value	*p*	slope	*t*-value	*p*				
TBCs	**0.355**	**3.633**	**0.003**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**0.347**	**-5.04**	**18.94**	**0.769**
1 (null)	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-8.67**	**23.60**	**0.075**
TBFs	**0.385**	**2.024**	**0.061**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**0.129**	**-7.49**	**23.84**	**0.066**
TBCs + δ18O	0.355	3.605	0.003	-	-	-	-	-	-	-0.087	-0.850	0.410	0.266	-6.04	23.93	0.063
δ18O	-	-	-	-	-	-	-	-	-	-0.085	-0.615	0.548	-0.109	-9.55	27.96	0.008
TBCs + NMA	0.319	2.969	0.010	-	-	-	-0.008	-0.824	0.424	-	-	-	0.866	8.41	28.66	0.006
TBFs + δ18O	0.378	1.930	0.074	-	-	-	-	-	-	-0.066	-0.524	0.609	0.018	-8.51	28.87	0.005
NMA	-	-	-	-	-	-	-0.020	-1.773	0.097	-	-	-	-0.276	-10.74	30.34	0.003
TBCs + SL	0.356	3.420	0.004	0.000	-0.038	0.970	-	-	-	-	-	-	-0.106	-9.52	30.89	0.002
TBFs + NMA	0.278	1.190	0.254	-	-	-	-0.011	-0.803	0.435	-	-	-	-0.251	-10.57	32.98	0.001
TBCs + NMA + δ18O	0.320	2.942	0.011	-	-	-	-0.008	-0.819	0.428	-0.087	-0.843	0.415	-0.09	-9.40	34.13	<0.001
SL	-	-	-	0.004	0.602	0.556	-	-	-	-	-	-	-0.61	-12.72	34.29	<0.001
δ18O +NMA	-	-	-	-	-	-	-0.020	-1.744	0.103	-0.086	-0.667	0.516	-0.421	-11.65	35.15	<0.001
TBFs + SL	0.374	1.890	0.080	0.002	0.367	0.719	-	-	-	-	-	-	-0.434	-11.73	35.31	<0.001
TBCs + SL + δ18O	0.363	3.449	0.004	-0.001	-0.294	0.774	-	-	-	-0.097	-0.872	0.399	-0.23	-10.43	36.18	<0.001
TBFs + NMA + δ18O	0.266	1.102	0.291	-	-	-	-0.011	-0.815	0.430	-0.073	-0.565	0.582	-0.404	-11.55	38.43	<0.001
δ18O + SL	-	-	-	0.003	0.433	0.672	-	-	-	-0.066	-0.450	0.660	-0.791	-13.62	39.09	<0.001
TBFs + SL + δ18O	0.372	1.818	0.092	0.001	0.224	0.826	-	-	-	-0.058	-0.423	0.679	-0.612	-12.72	40.78	<0.001
TBCs + SL + NMA	0.323	2.864	0.013	-0.001	-0.199	0.845	-0.009	-0.819	0.427	-	-	-	-0.633	-12.84	41.01	<0.001
SL + NMA	-	-	-	0.001	0.176	0.862	-0.019	-1.603	0.131	-	-	-	-1.101	-14.98	41.80	<0.001
TBFs + SL + NMA	0.279	1.150	0.271	0.001	0.188	0.854	-0.010	-0.711	0.490	-	-	-	-1.061	-14.82	44.97	<0.001
TBCs + SL + NMA + δ18O	0.328	2.890	0.014	-0.002	-0.467	0.649	-0.009	-0.874	0.399	-0.103	-0.920	0.376	-0.804	-13.68	46.82	<0.001
δ18O + SL + NMA	-	-	-	0.000	-0.020	0.984	-0.020	-1.617	0.130	-0.087	-0.619	0.547	-1.325	-15.84	47.01	<0.001
TBFs + SL + NMA + δ18O	0.266	1.059	0.311	0.000	0.020	0.984	-0.011	-0.752	0.467	-0.072	-0.511	0.618	-1.298	-15.74	50.94	<0.001

3.4 Discussion

3.4.1 Turtle richness and the effects of subsampling and regional analysis

Uncorrected global face-value genus counts show a steadily increasing richness from the first appearance of turtles up to the early Late Cretaceous, then a more abrupt increase to the Maastrichtian (K8). This was followed by a large drop over the K-Pg boundary and a series of fluctuations throughout the Paleogene, with peaks during Pg2 (Ypresian) and Pg4 (Bartonian-Priabonian) (Figure 3.1). Paleogene fluctuations resulted from apparently synchronised, continent-scale fluctuations in subsampled regional data for North America, Asia and Europe (Figure 3.4). These continents contributed the majority of Paleogene turtle fossils, and patterns elsewhere are largely uncharacterised. In contrast to the Paleogene fluctuations, other features of the global face-value curve changed drastically when subsampling was applied, and especially when it was applied to regional datasets. Instead, there was low richness in the Late Jurassic that became much higher in the Early Cretaceous. This global trend is due to the combination of a well-sampled European record (see Figure 3.4) plus some underlying input from North America and Asia that isn't recovered fully in the individual continent curves. These findings illustrate the pitfalls of analysis of 'global' diversity patterns in the fossil record. Many of the nominally 'global' trends are in fact caused by richness patterns present in one or two dominant continents plus a combination of others that are not individually well-sampled enough to provide subsampled richness estimates at quorum 0.4.

I was not able to characterise patterns of richness in many (or any, in some cases) regions during the Early–Middle Jurassic and mid-Cretaceous. The turtle fossil record is very poorly sampled during these intervals (Nicholson et al., 2015), and further field exploration is required before this will be possible. I recovered a large peak in the Turonian–Santonian (K6) from well-sampled Asia and North America, and lower richness in the final two Cretaceous bins. In the continental SQS results the Northern Hemisphere continents experienced different changes in richness at different times, and so this cumulative ‘global’ rise is not as simple as it appears (see below). The rise across the K-Pg boundary is a net rise driven almost entirely by North America, and the subsequent fluctuations in the Paleogene are primarily due to changes in North America and Europe, with some input from Asia and South America. All of the continents for which I recover Paleogene data at quorum 0.4 show the same fluctuating peaks and troughs in richness (albeit at differing scales) when they appear, which makes the ‘global’ curve very consistent at this time (at least for the Northern Hemisphere). Whether these ‘global’ patterns are driven by climate, sampling biases not accounted for by SQS, or something else remains to be seen (but see the results of the multivariate analyses).

When using classical rarefaction (CR) I recovered the same trends as SQS, but with fewer recovered points overall, particularly for the more poorly sampled Southern Hemisphere continents. Sample sizes had to be extremely low in order to generate even these reduced points, making interpretations unreliable, and trends were much flatter than those identified with SQS. This is a problem with CR that has been pointed out by Alroy (2010b), and as such I will be focusing on SQS results only herein.

3.4.2 Mesozoic richness patterns and major clade diversifications

Turtle diversity of the Triassic was low compared to later intervals in all the analyses. The first unambiguous stem-group turtles appeared during Tr4 (Norian). The fossil record for this time is sparse for turtles and consists of stem members of the clade only (Joyce, 2017). I recovered a Rhaetian (Tr5) point at quorum 0.4 at the same level as Tr4 but this is excluded from Figure 3.4 due to the very low number of collections that contribute to it (see Figure 3.10 for inclusion of points supported by <5 collections). Similarly low levels of richness not reaching the collection threshold were recovered for Asia and North America in Tr4 and J1 respectively. The low level of richness in Tr4 persists up to quorum 0.6 (see Figure 3.6), and although the European record is the only one robust enough to be recovered in my analyses, by the Norian (Tr4) turtles had already achieved a cosmopolitan distribution (Joyce, 2017). This supports the earlier observation that poor sampling is obscuring much about the origins and early history of turtles (Anquetin et al., 2009; Joyce, 2017).

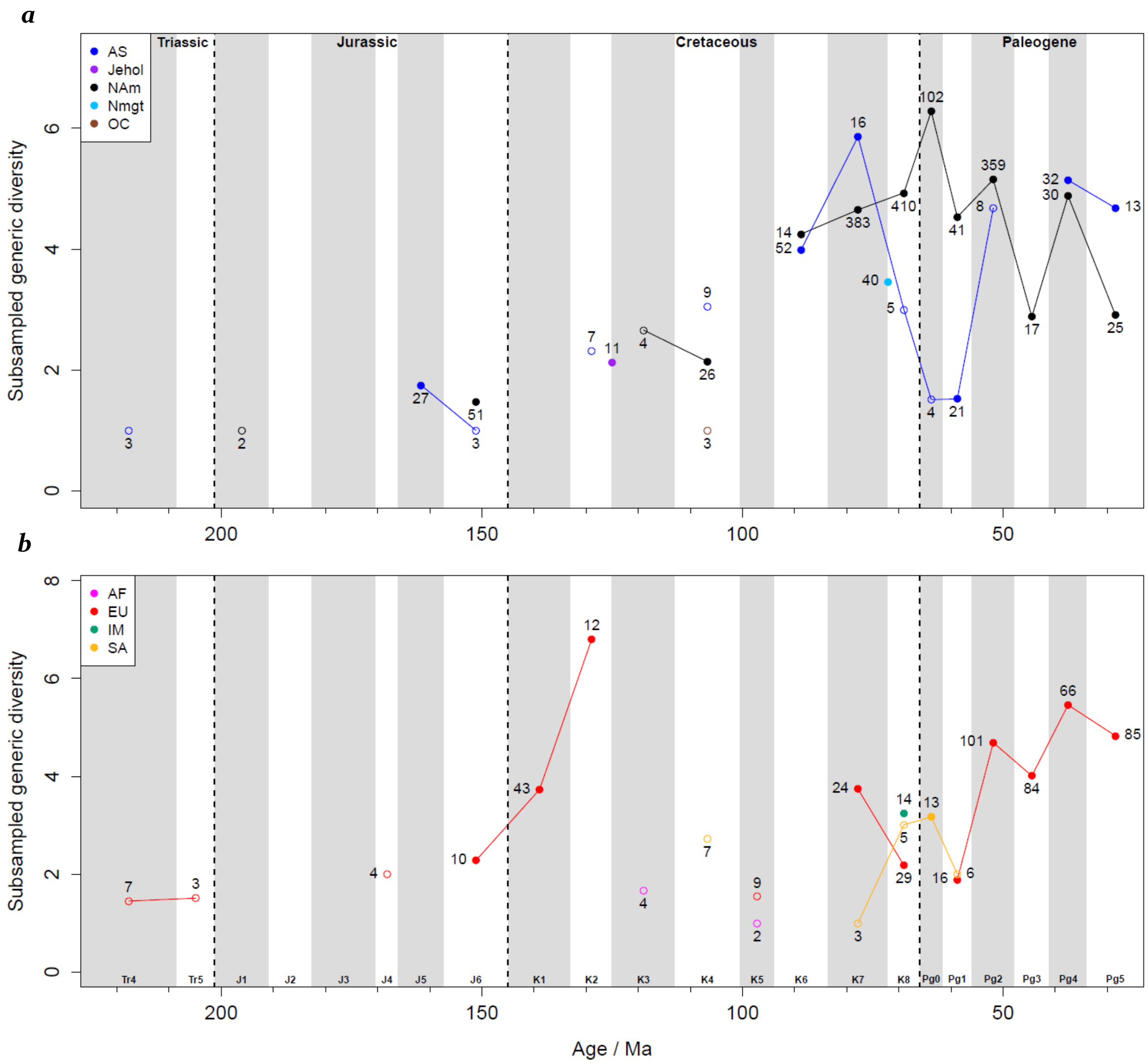

Figure 3.10: Subsampled non-marine turtle generic richness from Triassic–Paleogene at quorum 0.4, for individual continents, including those points with fewer than 5 collections: AF, Africa; AS, Asia; EU, Europe; IM, Indo-Madagascar; NAm, North America; OC, Oceania; SA, South America. Also included is the combined Jehol Group. Numbers indicate number of collections drawn for each time bin subsampled, as an indicator of underlying data quality. T, Triassic; J, Jurassic; K, Cretaceous; Pg, Paleogene.

The Early to Middle Jurassic is very poorly sampled and nothing is recovered even at quorum 0.4 (with the exception of one unreliable North American point mentioned above). This limits what can be understood about the origin of the turtle crown group, which is thought to occur in the Middle to Late Jurassic, based on the first appearance of crown group fossils in the Late Jurassic that were geographically widespread shortly afterwards (Sterli, 2008; Anquetin et al., 2009). These observations suggest that turtles could have been widespread and abundant during this lengthy undersampled interval. Other non-marine vertebrate fossil records also have major sampling gaps at the same time, including crocodylians (Mannion et al., 2015), lepidosaurs (Chapter 2; Cleary et al., 2018) and mammals (Newham et al., 2014).

The late Middle and Late Jurassic had a number of lineage-splitting events that gave rise to six distinct groups of crown and stem turtles. These lineages developed, perhaps vicariously, in three different geographic areas, with one crown and stem lineage in each. In Euramerica, these crown and stem lineages are Paracryptodira and Helochelydridae (=Solemydidae), respectively. Cryptodira and Sichuanchelyidae diversify in Asia, and Pleurodira and Meiolaniiformes are predominantly Gondwanan in distribution (Joyce et al., 2016; Joyce, 2017). It is possible, however, that Paracryptodira may belong to the cryptodire stem (Joyce, 2017), and so this splitting event may be more complex than thought currently. In my analyses, no clear patterns are recovered until the Late Jurassic (J5–J6), where I still find a relatively low richness, with a slightly higher richness in Europe. Notably, I recover no meaningful diversity trends from the Southern Hemisphere, which means I cannot comment on the origins or early diversification of Pleurodira, one of the two key lineages of extant turtles.

The Jurassic–Cretaceous boundary is poorly sampled for all continents except Europe, but here richness rose between the Kimmeridgian–Tithonian (J6) and the Berriasian–Valanginian (K1), and then sharply rose during the Hauterivian–Barremian (K2). High sampling in Europe at this time is due in part to the very productive Wealden and Purbeck localities in the United Kingdom. Turtles increased substantially in richness during the Early Cretaceous, particularly between K1 and K2, although the K2 signal is lost at quorum 0.6 (Figure 3.6). More derived members of Eucryptodira and Paracryptodira flourished (Pérez-García, 2012, 2014) and the Xinjiangchelyidae made their first appearance in Europe (likely immigrating from Asia; Pérez-García et al., 2015). Gondwanan taxa (pleurodires and meiolaniids) remained absent from Laurasia at this time (Perea et al., 2014). This increase in richness from the Jurassic to the Cretaceous is consistent with findings on other terrestrial vertebrates, including lepidosaurs (Cleary et al., 2018 and Chapter 2, although sampling is very poor), mammals (Newham et al., 2014) and small-bodied tetrapods in general (Jonathan P. Tennant, Mannion and Upchurch, 2016), but contrary to pseudosuchian richness patterns (Mannion et al., 2015).

In the later Early Cretaceous there was low to medium richness in Asia, corresponding to the origin and diversification of stem Trionychoidea (softshell turtles, e.g. *Perochelys*) and a diverse array of Xinjiangchelyidae (Brinkman, Rabi and Zhao, 2017). The separately plotted Jehol Group (between K2 and K3) had comparable richness to that of the cumulative Asian record at this time. This is consistent with the fact that this Lagerstätte makes a large contribution to what we know of the turtle record from this time, and highlights potential issues with sampling discrepancies between areas with vastly different preservation potentials. South American richness was at a similar level in the Albian (K4); richness was contributed to by pleurodire turtles (Pelomedusoides) collected entirely from the

Santana Formation in the Araripe Basin of Brazil, as South America is otherwise very poorly sampled in the Early Cretaceous. At this time in North America, new lineages, such as the endemic Baenidae, were establishing themselves (Hirayama, Brinkman and Danilov, 2000; Joyce and Lyson, 2015) alongside taxa persisting from the Late Jurassic (e.g. pleurosternids, solemydids), but the continent is too poorly sampled to reveal how this influenced richness in sequential time bins. All of the points recovered from the Early Cretaceous (minus that for the Jehol Group and K2 Europe) are robust up to quorum 0.6, and show areas of spectacular preservation or study that highlight the continued diversification and persistence of non-marine turtles despite the overall low level of sampling that my results indicate.

In the Late Cretaceous I recovered more robust trends from the Northern Hemisphere, particularly Asia and North America. The Cenomanian (K5) still suffers from poor sampling apart from a very low richness recovered for Europe, but between the Turonian and Santonian (K6) there was a higher richness in both North America and Asia, which coincides with the occurrence of the Cretaceous Thermal Maximum (CTM; Forster et al., 2007). The occurrence of turtles at high latitudes in the Turonian (71°N; see Figure 3.11) suggests that the higher temperatures allowed for a range expansion, as turtles tend to have a conservative latitudinal range compared to other terrestrial tetrapods at other times (Nicholson et al., 2015).

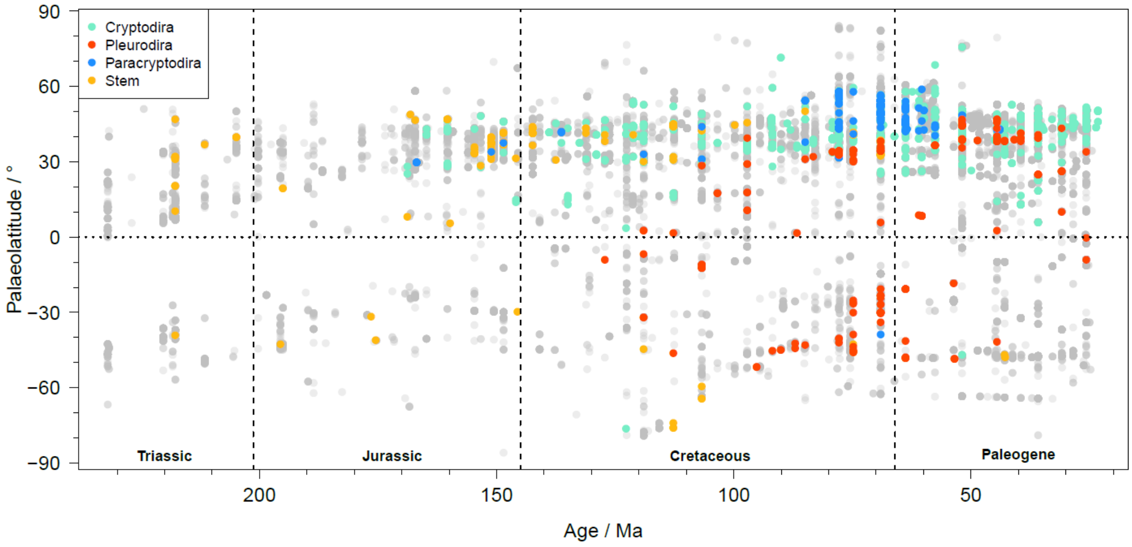

Figure 3.11: Palaeolatitudinal distribution of non-marine turtle occurrences from Triassic–Paleogene; light blue circles indicate non-turtle tetrapod occurrences. T, Triassic; J, Jurassic; K, Cretaceous; Pg, Paleogene.

During K7 (Campanian) a difference between Asia and North America emerged. In Asia, there was a peak in richness (the highest recovered for the continent in the studied timespan), with high turtle diversity in Mongolia and China. In the Asian Late Cretaceous, soft-shelled trionychids remained diverse and important components of the fauna (Brinkman, Rabi and Zhao, 2017), alongside terrestrial nanhsiungchelyids (Danilov and Syromyatnikova, 2008) and stem testudinoids (the precursors of extant land turtles, e.g. *Lindholmemys*; Hirayama, Brinkman and Danilov, 2000; Tong et al., 2017). Poor dating restricted the inclusion of some formations and collections in my analyses, which may contribute to the decrease in richness in the Maastrichtian (K8). It should also be noted that lower numbers of collections contribute to the final two Asian Cretaceous bins, meaning that these results are not as robust as those from North America; the K8 Asian bin disappears above quorum 0.4 indicating poorer sampling compared to K6, K7 and the lone Nemegt Formation, which persist until at least quorum 0.6. The Nemegt Formation is much more robustly sampled, with 40 collections, but despite this richness remains comparable to K8 Asia. This suggests that richness was genuinely lower in Asia at this time compared to previously in the Late Cretaceous and to North America. This is likely due to a difference in environmental conditions: the beds in Mongolia and China were deposited in arid or semiarid conditions, whereas the Western Interior beds of North America were mostly deposited in wet coastal or near-coastal environments (Hirayama, Brinkman and Danilov, 2000). The terrestrial nanhsiungchelyids would likely have coped better with this aridity, but the environment may have excluded other aquatic-based clades.

European subsampled richness decreased between the Campanian and Maastrichtian (K7–K8). Campanian richness was comparable to that of North America at this time, as non-marine pleurodires (bothremydids) appeared in the Northern Hemisphere for the first time during the Santonian and quickly became the dominant turtle clade in the Late Cretaceous of Europe. This is theorised to be a direct result of the CTM, as modern pleurodires have stricter environmental requirements than cryptodires and so a period of warming may have facilitated their dispersal to the other continents (Pérez-García, 2017). Richness decreased in the Maastrichtian, which could be attributed to poor sampling, but the countries sampled are very similar so it might be that a genuine decrease occurred at this time. In Romania, for example, the turtle fauna recovered so far is monogeneric, consisting solely of the meiolaniiform *Kallokibotion*.

South America and Indo-Madagascar are sufficiently sampled in the Maastrichtian (up to at least quorum 0.6 for the latter) to estimate subsampled genus richness. In both continents richness was comparable to the Maastrichtian of Asia, with podocnemids dominating all assemblages. Other Gondwanan continents are too poorly sampled for this time period.

In North America there was a steady rise in richness from K6–K8, following the CTM. Trionychids first appeared in the record here during the Santonian (Brinkman, Rabi and Zhao, 2017), as pleurodires did in Europe, perhaps again due to the warmer conditions facilitating immigration from Asia. The terrestrial Nanhsiungchelyidae also appeared here in the Late Cretaceous (*Basilemys*; Holroyd and Hutchison, 2002). Regional differences are observed in the uncorrected record between the Campanian and Maastrichtian (Brinkman, 2003; Holroyd, Wilson and Hutchison, 2014) but the overall pattern is of a continuing high level of richness up to the end of the Cretaceous contributed to by high immigration from other continents. Richness increased in K7–8 in North American lepidosaurs (Chapter 2; Cleary et al., 2018), but mammal richness decreased (Newham et al., 2014). In non-marine crocodylians richness was much higher in Europe than in North America (Mannion et al., 2015). When turtles are split into major clades (Figure 3.8, 3.9) it appears that the richness of cryptodire genera decreased between K6 and K7 before increasing again in K8, while paracryptodires did the opposite, resulting in a net shallow increase over the three time bins. Many new lineages emerged in North America at this time, including basal chelydrids (snapping turtles; Santonian) and kinosternoids (mud sliders; Campanian), and after a hiatus (due to the fluctuating coverage of the Western Interior Seaway) baenids are recovered in the record after the Santonian and were very diverse (Holroyd and Hutchison, 2002; Joyce and Lyson, 2015; Joyce, 2016; Joyce and Bourque, 2016), leading to a diverse, rich assemblage of non-marine turtles throughout the Late Cretaceous.

3.4.3 Results compared to Nicholson et al. (2015)

The results from the Mesozoic portion of my study corroborate those recovered by Nicholson et al. (2015). Some differences can be ascribed to my exclusion of points recovered from very low collection numbers, and the addition of data points representing the Early Cretaceous Jehol Biota of China, and the Late Cretaceous Nemegt Formation of Mongolia. When points contributed to by <5 collections are added back in (Figure 3.10), three differences remain between my analyses and Nicholson et al.'s despite matching time binning schemes. These are the Late Jurassic (J5–6) and Early Cretaceous (K2–4) trends in Asian richness that Nicholson et al. find and I did not, and the presence of bin K6 for North America that I recovered and they did not. The first two differences can be explained primarily by the reclassification of several taxa since the first study was conducted. The peak in Asian richness the authors recovered in the mid-Early Cretaceous disappeared from my data due to both this and changes in the stratigraphic ages of certain rock units and localities (e.g. Khodzhakul Formation, Uzbekistan). This resulted in Asian turtles of bin K3 (Aptian) becoming too poorly sampled to provide an estimate of diversity at a quorum of 0.4 and the reduction of richness of bin K4 (Albian). Finally, the ability to make a subsampled diversity estimate in North American bin K6 is due to the increased number of specimens from the Canadian Milk River Formation recorded in my dataset.

3.4.4 Paleogene richness patterns and major clade diversifications

The K-Pg mass extinction has been intensely studied for its effects on global richness, most famously for the demise of the non-avian dinosaurs (e.g. Alvarez et al., 1980; Clemens, 1982; Hutchison and Archibald, 1986). While there are few instances worldwide where we can follow the fossil record through the boundary without hiatuses, previous literature has suggested that turtles had very high rates of survival, at least in North America and possibly elsewhere (Hutchison, 1982; Hutchison and Archibald, 1986; Sheehan and Hansen, 1986; Hutchison and Holroyd, 2003; Holroyd, Wilson and Hutchison, 2014).

In North America, which has the best continuous rock record across the K-Pg boundary, there was a definitive rise in richness between the Maastrichtian (K8) and Danian (Pg0). The earliest richness in the Puercan (= Danian) Tullock Member of the Fort Union Formation of Montana was observed at face-value as being just as high as in the preceding Maastrichtian Hell Creek Formation (Holroyd, Wilson and Hutchison, 2014). Minimum survivorship was 75% for the northern Great Plains region, with 18 of 24 genus-level lineages confidently known to have survived over the K-Pg boundary. Holroyd et al. (2014) highlight that stratigraphic uncertainty means that we cannot tell if certain turtle genera made it to the boundary or were extinct long prior (perhaps even hundreds of thousands of years prior) to the K-Pg event and so true survivorship could be much higher. Nevertheless, a large proportion of Cretaceous genera are found in the earliest Paleocene, including many baenid and trionychid genera as well as members of smaller lineages such as kinosternoids, pleurosternids and adocids. My subsampled curves are consistent with high turtle survivorship, and a diversification of taxa between the end of the Cretaceous and the earliest Paleocene in North America.

In South America there was essentially no difference between pre- and post-boundary richness, and the same podocnemoid genera appeared in bins on either side of the boundary in my data. However, the number of collections contributing to K8 is low, and the geographic spread of the data is very different between the two bins. This is exacerbated by uncertainties in the dating of key formations, which are regarded as ‘Campanian–Maastrichtian’, meaning that my analyses exclude them. Once the ages of these formations are better understood it may be possible to observe richness trends more precisely for South America, but based on the evidence currently available it appears that turtle richness changed little over the boundary here.

The common face-value trend observed was high survivorship through the K-Pg boundary, which my subsampled results are consistent with, at least for North and South America. Prominent losses over the boundary were the Nanhsiungchelyidae and Sichuanchelyidae, both consisting of land turtles (Danilov and Syromyatnikova, 2008; Joyce et al., 2016). It is likely that differences in ecologies played a large part in turtle extinction selectivity at the boundary, and that aquatic turtles had a much higher survivorship; this has been noted by a number of authors studying both turtles and (similarly aquatic) non-marine crocodylians (e.g. Hutchison, 1982; Hutchison and Archibald, 1986; Brochu, 2003; Bronzati, Montefeltro and Langer, 2015; Mannion et al., 2015). A few papers have suggested that the increased survivorship of aquatic clades could be due to differences in food webs compared to fully terrestrial clades. After the bolide impact the atmosphere could have filled with dust and debris, obscuring sunlight and causing the collapse of primary productivity (Alvarez et al., 1980). This collapse would affect every organism in food webs associated with primary productivity. Modern aquatic habitats, however, rely more on detrital-based food webs rather than in-situ primary productivity, and so aquatic environments may have been buffered against the long-term environmental changes following the K-Pg event (Sheehan and Hansen, 1986; Robertson et al., 2013). These hypotheses are remain speculation, however, as they are impossible to test.

Later in the Paleocene, richness remained relatively low for everywhere except North America, with late Paleocene richness levels in Asia, Europe and South America reaching down to nearly their lowest points in the analysis. Clearly there is either a taphonomic or sampling bias occurring here that is not accounted for by SQS, or there is something genuinely different about North America compared to the other continents at this time. Nevertheless, many modern lineages began to appear during this interval, including testudinoids in both North America (platysternids; Vlachos, 2018) and Asia (testudinids, the extant tortoise clade; Claude and Tong, 2004).

By the early Eocene turtles had richness levels close to, if not higher than, those in the Late Cretaceous, although only the Northern Hemisphere continents are well-sampled enough to track from this point forwards in my subsampled curves. Higher temperatures after the Paleocene-Eocene Thermal Maximum (PETM) opened high latitude dispersal routes (e.g. Beringia) for interchange between Eurasia and North America (Godinot and de Broin, 2003). In Asia testudinids continued to diversify, and based on current fossil evidence dispersed to other continents from there, though whether tortoises truly originated in this region is debated (Hofmeyr et al., 2017). In Europe, the higher temperatures appear to also have facilitated more dispersals of pleurodires from Gondwana via Africa (Pérez-García, 2017). Carettochelyids (a clade of soft-shelled turtles), which were previously endemic to Asia (and are restricted to Australasia today), appeared in Europe and North America in the Ypresian (Joyce, 2014). North American assemblages continued to contain a high richness of turtles, with a turnover in the mid-Wasatchian Mammal Age (Ypresian, Pg2) that resulted in a rapid increase in large-bodied herbivorous testudinid and dermatemydid turtles. This corresponded to a similar mammalian turnover event and an increase in thermophilic flora (Holroyd, Hutchison and Strait, 2001). The warm climate enabled turtles to colonise high latitudes again for the first time since the Late Cretaceous, and in North America kinosternid and carettochelyid turtles reached at least 75°N at this time alongside squamates, amphibians and crocodylians (Estes and Hutchison, 1980; see also Figure 3.11). For comparison, the modern maximum latitude of extant turtles in North America is approximately 58°N (*Chrysemys picta*; Angielczyk, Burroughs and Feldman, 2015). It is unfortunate that the Southern Hemisphere is so poorly sampled at this time, as other terrestrial vertebrates appear to have had similar, southerly range expansions due to the warmer temperatures during the PETM (e.g. lepidosaurs; Head et al., 2009; Albino, 2011) in South America. Given modern pleurodire environmental requirements (Pérez-García, 2017) it seems likely that they might have responded favourably to warmer conditions in this region, though further evidence is required to test this hypothesis.

Turtle richness was lower in the early middle Eocene (Lutetian, Pg3) in both North America and Europe, in a pattern very similar to that found for non-marine lepidosaurs (Chapter 2; Cleary et al., 2018). For the European data the countries being sampled appear to be approximately the same, and so the reason for this decline in richness is intriguing. As with lepidosaurs, this decline could be linked to the temperature decline that occurred at the end of the Early Eocene Climatic Optimum (Woodburne, Gunnell and Stucky, 2009) or taphonomic differences between Pg2 and Pg3. Mannion et al. (2015) found the inverse pattern for non-marine crocodylians, with Pg3 containing a higher richness of taxa than the two surrounding bins. In North America, turtle decline in the Lutetian (containing most of the Uintan North American Land Mammal Age, or NALMA) may have been due to a decrease in aquatic environments and increasing aridity, which is suggested by Hutchison (1992) based on the reduction in carapace size of aquatic turtles in the Uintan, while tortoises (testudinids) remained mostly unaffected. The decrease in Pg3 is also partially caused by the difficulties in correlating NALMAs with standard European stratigraphic stages, as, for example, the Uintan overlaps the Lutetian and Bartonian stages, which means that if collectors do not specify the section of the Uintan that the fossils are derived from they cannot be included in my analyses. This may be one of the reasons why the North American dip in richness I recovered was more severe than that seen in Europe.

High richness is recovered in the late Eocene (Pg4) on both continents, with European richness at its highest for the Triassic–Paleogene range. Asia is well-sampled enough to be recovered again, with richness levels comparable to North America. There are fewer monogeneric localities in Europe in Pg4 compared to Pg3, and numerous collections are known from the United Kingdom, but the geographic spread of collections does not drastically change. In North America, Baenidae made their last appearances in the fossil record in the late Uintan NALMA (early in Pg4), and are assumed to have become extinct by the end of the Eocene (Joyce and Lyson, 2015), while testudinids became more common. This coincided with increasing aridity throughout the later Eocene (Hutchison, 1992). Precipitation (and the associated availability of aquatic habitats) is the most important factor that determines species richness in extant aquatic turtles (Iverson, 1992; Waterson et al., 2016), and therefore increasing aridity could have had an effect on the distribution and richness of aquatic versus terrestrial turtles. The interaction of various environmental variables is complicated, however, and aridity alone may not necessarily cause aquatic taxa to decline in richness.

The Eocene–Oligocene boundary (Pg4–5) has been the subject of intense interest due to the Grande Coupure ('great break'), an extinction or turnover event recognised for many groups of terrestrial vertebrates, particularly mammals (Stehlin, 1909; Hooker, 1989; Meng and McKenna, 1998; Costa et al., 2011; Eronen et al., 2015) and squamates (Rage and Augé, 1993; Augé and Smith, 2009; Rage, 2013; Cleary et al., 2018; Chapter 2). It coincides with a sudden decrease in temperature, increased seasonality and the onset of Antarctic glaciation (Ivany, Patterson and Lohmann, 2000; Zachos et al., 2001; Liu et al., 2009). In my analyses, the turtle richness of all three Northern Hemisphere continents decreased across this boundary. The decrease appeared more severe in North America until we analyse results at quorum 0.6 (Figure 3.6), where it was very similar to Europe but at a lower starting richness level.

In Asia relatively few Oligocene deposits have been sampled, and so examining the effects of this climatic shift is more difficult; for example, the Asian Pg5 bin is not recovered above quorum 0.4. It appears, however, that large parts of the continent experienced the same transition from warmer temperate conditions to an arid or semi-arid cooler climate with a short rainy season and fewer permanent bodies of water (Meng and McKenna, 1998; Böhme, 2007; Sun et al., 2014). This likely affected vertebrate assemblages, and Böhme (2007) noted that late Eocene Mongolian faunas contained many aquatic forms that do not appear in the Oligocene. Nevertheless, it will be interesting to investigate the effects of the Grande Coupure further when more data are available from Asian Oligocene localities.

In Europe, where the Grande Coupure was first recognised (Stehlin, 1909; Hooker, 1989), there was a reduction in pleurodire richness (Figure 3.9a), which corresponds to the known stricter environmental requirements of the extant members of that clade. Conversely, cryptodires increased in richness (Figure 3.8). Little attention has been paid to European turtle responses to the event, but richness remained relatively high during the Oligocene. Geoemydids (European pond turtles) and trionychids were still relatively abundant even in the early Oligocene, which suggests that European aquatic turtles were not affected much by supposedly global environmental changes.

North American turtle faunas underwent changes at this time, with emydids and trionychids disappearing after very rare occurrences in the early Oligocene, leaving testudinid-dominated assemblages in the continental interior. Testudinid body size decreased across the Eocene–Oligocene boundary (Hutchison, 1992), corresponding with global cooling and drying (Hutchison, 1982), but temperatures were still relatively warm and testudinid geographic ranges remained unchanged. Evidence for a brief warmer period in the late Oligocene is suggested by the presence of pleurodires in South Carolina and Florida at this time (Weems and Knight, 2012; Bourque, 2016). There are also instances of taxa that were previously thought to have gone extinct at the boundary in North America (e.g. carettochelyids), but were more recently found as fragmentary remains in the late Oligocene (Joyce, 2014). The differences in responses between continents could indicate a number of possibilities, including that sampling for the Oligocene is much poorer (particularly in North America) and with effort and time the effects of the Grande Coupure may be found to have been less severe than currently thought. There is a lack of localities for the late Oligocene of North America (Arikareean NALMA; Joyce and Bourque, 2016) that could exacerbate the steep drop in richness found here compared to the rest of the Northern Hemisphere.

Conversely, or perhaps additionally, the sudden global temperature drop could have caused migration of aquatic taxa out of the sampled areas to more southerly latitudes. In North America for example, kinosternids, chelydrids, emydids and trionychids are found widely in extant faunas, and it is hypothesised that these turtles persisted at lower latitudes e.g. in the Gulf Coast region until conditions were more favourable further north (Hutchison, 1992). This was also theorised to be the case for North American squamates (Smith, 2011), but the lack of more southerly localities limit the testing of these hypotheses.

Nevertheless, the climatic changes of the Eocene–Oligocene boundary likely limited the geographic ranges of turtles with very strict temperature tolerances (e.g. pleurodires), and allowed those with wider temperature tolerances and semi-terrestrial to terrestrial habitats to spread (testudinids). By the end of the Oligocene, turtle taxonomic composition closely resembled that of modern-day assemblages.

3.4.5 Correlations of turtle richness with potential driving variables

I compared models containing various combinations of explanatory variables (sampling, temperature, land area, sea-level) to face-value turtle richness using generalized least-squares regressions. Of these, only the model with the number of tetrapod-bearing collections (TBCs) per bin was found to be a better fit than the null model based on AICc values (Table 3.2). It appears that turtle richness dynamics are complex and do not correspond to common drivers of richness such as palaeotemperature and non-marine area. Even though it was chosen as the best model, the TBCs-only model had a very low R2 value (0.35), and so does not explain the variance in richness well; it is just the best of a group of bad models. The number of fossil-bearing collections is recovered as the best explanatory variable or as a component of a model alongside non-marine area, temperature and/or Lagerstätten for a number of vertebrate groups including sauropodomorph dinosaurs (Benson and Mannion, 2012), dinosaurs in general (Mannion et al., 2012), pterosaurs (Butler, Benson and Barrett, 2013), and marine reptiles (Benson and Butler, 2011).

Lepidosaur richness was found to correlate best with TBCs and a palaeotemperature proxy (δ18O; Chapter 2; Cleary et al., 2018), and palaeotemperature correlated closely with variation in the richness of cetaceans (along with diatom richness; Marx and Uhen, 2010) and European crocodylians (Mannion et al., 2015). It is surprising that, as ectotherms, turtles do not respond as strongly to temperature as other extant groups, though North American crocodylian richness also does not correspond closely to palaeotemperature (Mannion et al., 2015). It is likely, as discussed above, that precipitation and the availability of aquatic environments were more important to turtle richness alongside temperature influences (as in modern faunas; Iverson, 1992), since a large portion of Triassic–Paleogene turtles were aquatic, but these factors are hard to measure directly in the fossil record. Ecological niche modelling of the Maastrichtian versus the modern day found that precipitation drove niche expansion of aquatic turtles, while thermal limits provided important constraints on niche changes (Waterson et al., 2016), and should be repeated on other time periods to see if these major influences change.

3.4.6 Conclusions

Temperature, rainfall, and habitat heterogeneity are thought to drive patterns of richness in turtles both now and in deep time (Holroyd and Hutchison, 2002; Waterson et al., 2016). However, our analyses demonstrate that the variance in observed turtle richness through geologic time is also influenced by a significant sampling component that must be accounted for in interpreting the timing and pattern of diversification of turtles through time. Through the analyses presented here, I have done that, showing where sampling is not yet sufficient and discovering previously-unrecognized richness peaks and shared long-term patterns of diversity change.

Turtle richness was relatively low and poorly sampled until the mid-Late Cretaceous, with the exception of a few instances such as the Late Jurassic–Early Cretaceous of Europe. Richness was high following the Cretaceous Thermal Maximum, but in all recovered continents except North America richness declined to a lower level by the Maastrichtian (K8). Non-marine turtles appear to have been relatively unaffected by the K-Pg mass extinction, besides the extinction of most land turtles, with high survivorship in North America and a stable richness across the boundary in South America. I corroborated the previous findings of Nicholson et al. (2015) for Mesozoic turtles, with some minor updates resulting from taxonomic or stratigraphic changes since that publication. In the Paleogene, richness in all Northern hemisphere continents (where recovered) fluctuated in a series of peaks and troughs, with European richness at a consistently higher level than Asia and North America from the middle Eocene (Pg3; Lutetian) onwards. All three continents had a fall in richness at the Eocene–Oligocene boundary. This pattern of peaks and troughs corresponds to the one recovered for lepidosaurs, which suggests similar driving factors for both groups whether that be climate, sampling factors not corrected for by SQS, or a combination of both. GLS analyses suggest that sampling (particularly the number of collections bearing tetrapod fossils) somewhat influences recovered face-value richness through time, but much variation remains unexplained by sampling alone and further investigation is required. Ecological niche models may provide an alternative method of examining potential driving factors like precipitation that are difficult to directly observe in the fossil record.

With regards to current climatic change, it appears that responses to warming will differ between major turtle clades if past richness curves can be relied upon. Pleurodires, for example, will likely benefit from a warmer climate as it facilitated their dispersal on several occasions and they appear to have a preference for warmer temperatures. However, as aquatic turtles they are also reliant on rainfall-driven aquatic habitat availability, and so clade responses to current climatic warming will depend on the interaction of numerous factors. Ultimately turtle survivability will depend on the rapidity of their adaptation to changing conditions, and their ability to migrate to locales within their temperature tolerances. The data we have gathered on past richness provides a useful framework with which to predict future responses.

Despite the wealth of information recovered, the turtle record still suffers heavily from various collection biases, especially those related to uneven geographic sampling. This is particularly true for the Southern Hemisphere, where poor sampling obstructs the understanding of key events in the evolutionary history of turtles. The Early to Middle Jurassic is also very poorly sampled for non-marine vertebrates, making it difficult to pinpoint the temporal and geographic origins of crown group turtles. Even in well-sampled continents there are problems with fossil localities only representing a portion of the total land area for that time period (e.g. just the Western Interior of North America), and evenness can really affect richness estimates. It is therefore crucial to examine richness at more local scales, to avoid generalising what could be local patterns of richness to continental and global scales. Unfortunately, this will be difficult until more robust information is available, so we must rigorously assess the data we do have, make new field collections and increase assessments of neglected museum collections in order to improve our datasets for the future.

[This page intentionally left blank]

4. Latitudinal Biodiversity Gradients in Ectotherms

4.1 Introduction

The latitudinal biodiversity gradient (LBG) is the modern pattern of high species richness at the tropics (30°N–30°S) with decreasing richness moving into lower latitudes towards the poles (Willig, Kaufman and Stevens, 2003). It applies to most animal taxa, except the majority of amphibian families (Wiens, 2007) and some marine bivalves, birds and mammals (Chown, Gaston and Williams, 1998; Procheş, 2001; Krug, Jablonski and Valentine, 2007; Kiel and Nielsen, 2010). LBGs are thought to be driven by climate and seasonality, but it is difficult to isolate individual factors for study (e.g. Erwin, 2009; Archibald et al., 2010; Krug et al., 2010; Archibald, Greenwood and Mathewes, 2013; Mannion et al., 2014).

Modern lepidosaurs adhere to the general LBG (Roll et al., 2017), with some taxa deviating due to stricter environmental preferences (e.g. Amphisbaenia; Lewin et al., 2016) or the presence of large areas of unsuitable habitat within certain latitudinal belts (e.g. the Sahara). Extant non-marine turtles have their highest richness in hotspot areas that are mostly situated at low or mid- latitudes (Roll et al., 2017), such as those located in the south-eastern United States (where high richness extends into latitudes below 40°N), south-eastern Asia and Amazonian South America. Large areas of these hotspots coincide with the presence of major river drainage basins (e.g. Mississippi, Amazon, Ganges), as the richness of many non-marine turtle clades is dependent on average rainfall and its associated effects on the availability of suitable aquatic habitats (Iverson, 1992).

Many studies have discussed if the modern LBG was present in deep time and, if not, when it was established. Some theorise that the LBG was established as early as the Cambrian (500 Ma), and continued throughout time with some periods of shallower gradients when latitudinal differences in richness were less pronounced (e.g. Crame et al., 2001; Mittelbach et al., 2007). Most researchers agree that the present day strong LBG was established around 30 Ma (Thomas and Gooday, 1996; Crame et al., 2001; Archibald et al., 2010). A number of these studies used palaeolatitudinal occurrence data to determine if modern LBGs applied to plants (Rees et al., 2004; Peralta-Medina and Falcon-Lang, 2012), marine invertebrates (Powell, 2007), and non-marine tetrapods (Rose et al., 2011; Mannion et al., 2012, 2015) between the Mesozoic and now. Of particular interest are the palaeolatitudinal distributions of ectothermic non-marine tetrapods which, given their lowered tolerance with regards to temperature changes versus endotherms today (Gunderson and Stillman, 2015), should provide interesting data on how past climates potentially affected LBG compositions in the past. These data could also lend themselves to predicting possible outcomes of future climatic warming.

Unfortunately, many of the above mentioned studies failed to account for issues with sampling biases, which unfairly weight richness in areas where sampling intensity is high. This means that their results regarding whether a modern-type LBG was present or absent might be inaccurate. A small number of studies have taken steps to alleviate these biases, using subsampling methods (Rose et al., 2011; Mannion et al., 2012, 2015; Nicholson et al., 2016; Brocklehurst et al., 2017). Of these, only two examined ectothermic non-marine tetrapods in detail. Deep time LBGs have been investigated for Mesozoic turtles (Nicholson et al., 2016), and for non-marine crocodylians (Mannion et al., 2015) from their origins to present day, with varying results.

Here, I investigate LBGs for non-marine turtles and lepidosaurs in deep time, from the Triassic–Paleogene (252–23 Ma), building on the previous work of Nicholson et al. (2016). I test whether past LBGs existed and whether they correspond to the modern LBG. It might be expected that past LBGs might be shallower than the modern LBG, or even absent, due to the greenhouse climatic conditions that existed, particularly in the Mesozoic. I compare my findings with previous studies on palaeolatitudinal richness, in order to identify any common patterns, focussing particularly on other ectothermic taxa.

4.2 Methods

4.2.1 Data

I used the same Triassic–Paleogene datasets for non-marine lepidosaurs and turtles as described previously in Chapters 2 and 3 (see Appendices 2 and 7). Genera were used for the analyses in this chapter rather than species to maximise data volume by allowing the inclusion of specifically-indeterminate specimens. The datasets have 1971 occurrences of lepidosaurs (449 genera) and 2666 occurrences of turtles (295 genera), respectively. For regression analyses I once again used tetrapod-bearing formations and collections as proxies of sampling opportunity and effort; the data are from the same download of non-marine tetrapod occurrences used in Chapters 2 and 3, with 65,866 occurrences of 7173 genera (Appendix 4). The latitudinal spread of these occurrences can be seen in Figure 2.9 and 3.9, which plotted the latitudinal spread of tetrapod occurrences against non-marine lepidosaurs and turtles, respectively.

4.2.2 Latitudinal richness curves

The methods for this chapter are adapted from those used by Nicholson et al. (2016), who analysed the latitudinal distribution of Mesozoic non-marine turtles. I repeated their analyses for Mesozoic turtles and extended the dataset through the end of the Oligocene to provide an appropriate comparison to my lepidosaur data. In order to examine spatial distributions through time, I separated the occurrences of both clades into 15 degree latitudinal bands using the palaeolatitude data included in the Paleobiology Database downloads. These are calculated using tectonic plate rotation data from the GPlates model (Wright et al., 2013). Bands of 15 degrees were the smallest divisions of the data possible in order to show ecologically important results but still retain good sample sizes in each. After separating each clade into latitudinal bands I plotted raw uncorrected generic richness and subsampled richness using shareholder quorum subsampling (SQS) at quorum 0.4 (see previous chapters for SQS methodology). The time slices used for the Mesozoic were the Triassic (Tri), Jurassic (Jur), Campanian (Camp), Maastrichtian (Maas) and an amalgamated rest of the Cretaceous (Berriasian–Santonian; RCret), to generate results comparable to the analyses performed by Nicholson et al. (2016). The Paleogene was separated into the Paleocene (Pal), Ypresian (Ypr), rest of the Eocene (Lutetian–Priabonian; REoc) and the Oligocene (Olig). A summary of time slices and durations can be found in Table 4.1.

Table 4.1: Time slices used for latitudinal plots and GLS analyses, and the standard stratigraphic intervals they correspond to.

Bin name	Abbrev.	Stage equivalent (if applicable)	Base (Ma)	Top (Ma)	Range (Ma)
Latitudinal plots					
Triassic	**Tri**		252	201.3	50.7
Jurassic	**Jur**		201.3	145	56.3
Rest of Cretaceous	**RCret**	Berriasian–Santonian	145	83.6	61.4
Campanian	**Camp**		83.6	72.1	11.5
Maastrichtian	**Maas**		72.1	66	6.1
Paleocene	**Pal**		66	56	10
Ypresian	**Ypr**		56	47.8	8.2
Rest of Eocene	**REoc**	Lutetian–Priabonian	47.8	33.9	13.9
Oligocene	**Olig**		33.9	23.03	10.87
Generalized least-squares					
Jurassic	**Jur**		201.3	145	56.3
Early Cretaceous	**ECret**	Berriasian–Albian	145	100.5	44.5
Late Cretaceous	**LCret**	Cenomanian–Maastrichtian	100.5	66	34.5
Early Paleogene	**EPalg**	Danian– Lutetian	66	41.2	24.8
Late Paleogene	**LPalg**	Bartonian–Chattian	41.2	23.03	18.17

To examine further the most heavily sampled continents and their influence on supposedly 'global' richness patterns, I isolated occurrence data from North America and Europe and plotted both uncorrected and subsampled (SQS quorum 0.4) latitudinal richness curves in the same way as above for lepidosaurs and turtles.

4.2.3 Comparing richness to proxies for sampling and environmental factors

I used generalized least-squares regressions (GLS) to examine the relationship between uncorrected generic richness in different latitudinal bands and factors representing sampling or the availability of emergent land area in each band for five time slices. The time slices used were the Jurassic, Early Cretaceous, Late Cretaceous, Early Paleogene (Danian–Lutetian) and Late Paleogene (Bartonian–Chattian); these slices are longer than the time slices used in the previous chapters' general richness curves to allow for sufficient sample sizes in these analyses (see Table 4.1). The

Triassic record was too poorly sampled to be included for either clade. Instead of extracting the number of tetrapod-bearing collections and formations per time bin as in previous chapters, I separated the data into the above mentioned time slices and then into latitudinal bands within each slice (see Appendix 8). Nicholson et al. (2016) chose to use the number of turtle-bearing formations in their Mesozoic analyses, but I chose to use tetrapod-bearing formations instead as I believe that, despite the inclusion of some non-suitable habitats for turtles, the number of tetrapod-bearing formations better represents the potential opportunities to recover turtle fossils and overall sampling effort.

The third variable I used was emergent land area, which is named as such to separate it from the “non-marine area” variable used in previous chapters. Instead of using the non-marine area data from my other GLS analyses, which cannot be easily separated into latitudinal bins, I used the Mesozoic land area data from Nicholson et al. (2016), and obtained data for the Paleogene from Paul Valdes (pers. comm.). Land area fractions for the Paleogene were provided in 5 degree bands for each stage, which I multiplied by the total area in each band provided by Nicholson et al. (2016). The resultant areas were summed into 15 degree bins (see Appendix 8 for all land area data).

After assembling the data for each time slice for the two clades, I removed any latitudinal bands that contained zero genera. This was to prevent the artificial inflation of statistical power and the clustering of data points at the boundary of the data series. Data were log transformed prior to regressions. The combinations of variables I tested were all three individually, plus each sampling variable combined with land area, for a total of six models including an intercept-only null model. Variables were fitted with autoregressive (AR) models of order one, and were evaluated using Akaike's information criterion for small sample sizes (AICc) in R version 3.5.1. As in previous chapters, I generated R^2 values using the method of Nagelkerke (1991) and checked the normality and homoskedasticity of residuals using Jarque-Bera and Breusch-Pagan tests (Breusch and Pagan, 1979; Jarque and Bera, 1980). Packages used were lmtest v. 0.9-36 (Hothorn et al., 2017), nlme v. 3.1-137 (Pinheiro et al., 2017), qpcR v. 1.4-1 (Spiess and Ritz, 2014) and tseries v. 0.10-45 (Trapletti and Hornik, 2017). The script for all analyses from this chapter can be found in Appendix 8.

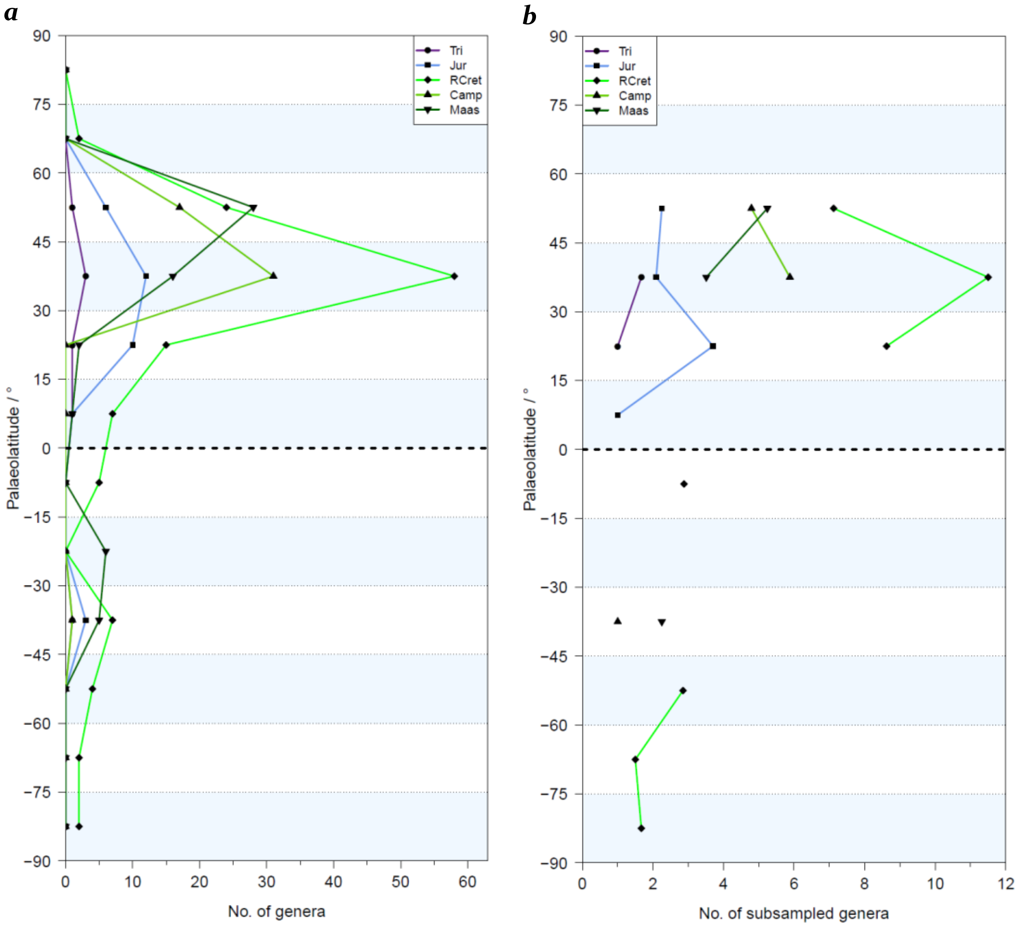

Figure 4.1: Latitudinal generic richness for global Mesozoic non-marine turtles using *(a)* uncorrected (face-value) richness and *(b)* subsampled (SQS) richness at quorum 0.4. Time slice abbreviations are Tri, Triassic; Jur, Jurassic; RCret, rest of Cretaceous (Berriasian–Santonian); Camp, Campanian; Maas, Maastrichtian.

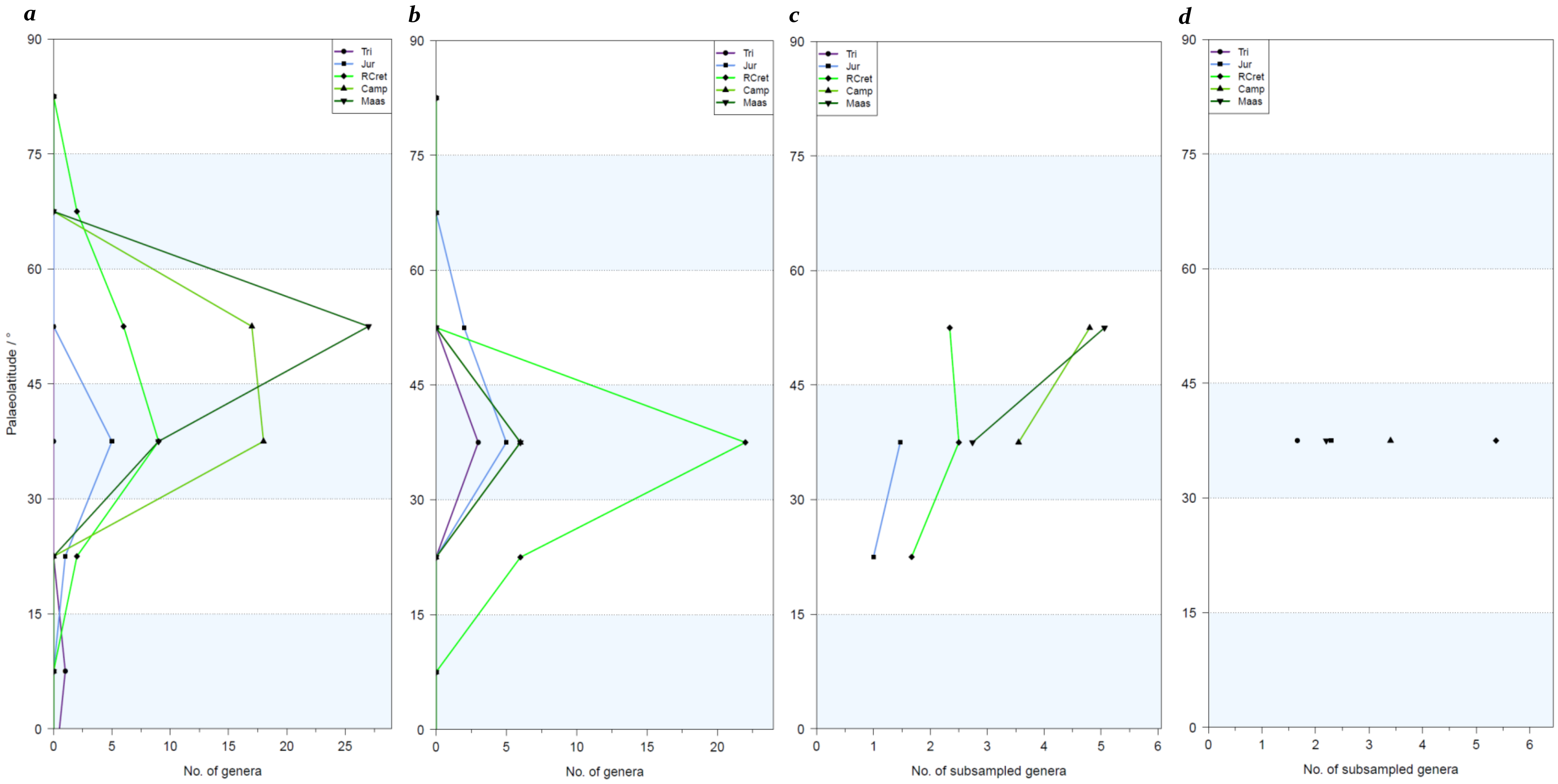

Figure 4.2: Latitudinal generic richness for Mesozoic non-marine turtles in North America and Europe. Figures show uncorrected (face-value) richness for (*a*) North America and *(b)* Europe, and subsampled (SQS) richness at quorum 0.4 for (*c*) North America and (*d*) Europe. Time slice abbreviations are Tri, Triassic; Jur, Jurassic; RCret, rest of Cretaceous (Berriasian–Santonian); Camp, Campanian; Maas, Maastrichtian.

4.3 Results

4.3.1 Latitudinal richness plots

Mesozoic turtles

Figures 4.1 and 4.2 show latitudinal richness curves plotted for Mesozoic non-marine turtles at the global and continental scales, respectively. In Figure 4.1a, which shows the uncorrected face-value data, there is a clear Northern Hemisphere peak between 30 and 60°N, which is clearly distinct from the generally low levels of richness in the Southern Hemisphere between 15 and 60°S, which remain low regardless of the time slice. The highest richness levels were in the RCret slice (representing most of the Cretaceous) between 30 and 45°N. The Campanian had very high richness at this same latitude, and in the following Maastrichtian stage this richness remained approximately the same but shifted north to 45–60°N. Triassic and Jurassic richness peaked at 30–45°N but both had much lower richness levels both overall and in these two bands than the other time slices.

After subsampling (Figure 4.1b), many latitudinal bins were not recovered because of poor sampling. Most of the remaining signal is in the Northern Hemisphere, except for some isolated instances between 30 and 45°S and a trend of relatively low to medium RCret richness at very high southern latitudes. The Triassic, as expected from the raw data, had relatively low turtle richness between 15 and 45°N. The Jurassic had a low richness near to the equator, and medium richness at higher latitudes (30–60°N), while its highest richness levels were found between 15 and 30°N. The RCret bin still had the highest richness by far out of all time slices, with its highest richness levels in the 30–45°N band. The Campanian and Maastrichtian retain their opposite trends from the uncorrected richness curves, although richness levels in the 45–60°N band are much closer after subsampling. The Campanian had the highest turtle richness of the two stages when considered at the global scale.

When I separated out North America and Europe (Figure 4.2a-c and b-d, respectively), the origins of certain trends in the global curves emerge. In the uncorrected North American data (Figure 4.2a) the Campanian had very similar high richness levels between 30 and 60°N while the Maastrichtian had a large peak in richness at 45–60°N but relatively low richness in the band below this. These trends were not present in the face-value European data (Figure 4.2b), which had relatively low richness for every time slice except the amalgamated rest of the Cretaceous (RCret). European Campanian and Maastrichtian richness is exactly the same.

After subsampling, many latitudinal bands drop out for being too poorly sampled, as with the global data. North America (Figure 4.2c) had relatively low turtle richness in the Jurassic between 15 and 45°N, and in the RCret slice between 15 and 60°N. The Campanian and Maastrichtian, unlike in the global results, appear to have had more congruent trends in North America, with lower richness at 30–45°N and higher richness at 45–60°N. The Maastrichtian had a greater difference between the two bands than the Campanian, which is congruent with the raw data for this continent. In Europe (Figure 4.2d) the Mesozoic record is more poorly sampled, and the only band recovered for all time slices is 30-45°N, with the scale of recovered subsampled richness much lower than North America. The RCret slice had the highest richness, with the Campanian second. The Jurassic and Maastrichtian had roughly the same turtle richness, and the Triassic had the lowest.

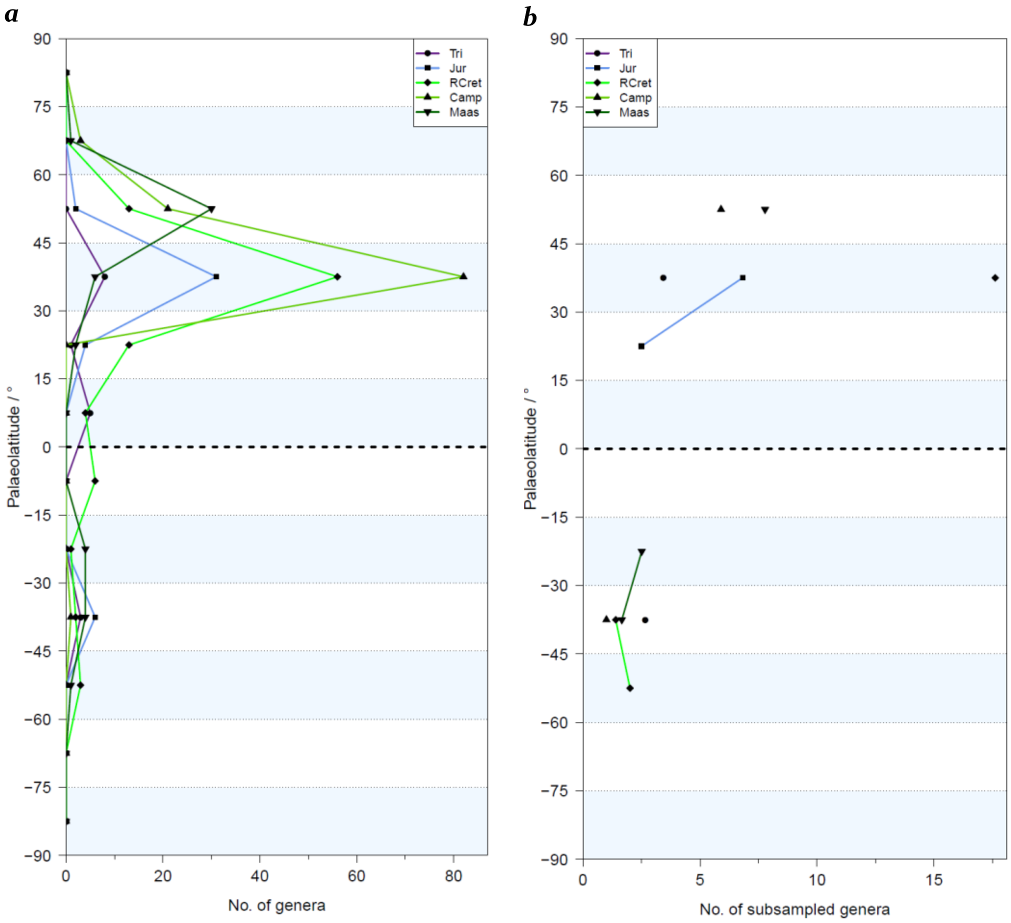

Figure 4.3: Latitudinal generic richness for global Mesozoic non-marine lepidosaurs using *(a)* uncorrected (face-value) richness and *(b)* subsampled (SQS) richness at quorum 0.4. Time slice abbreviations are Tri, Triassic; Jur, Jurassic; RCret, rest of Cretaceous (Berriasian–Santonian); Camp, Campanian; Maas, Maastrichtian.

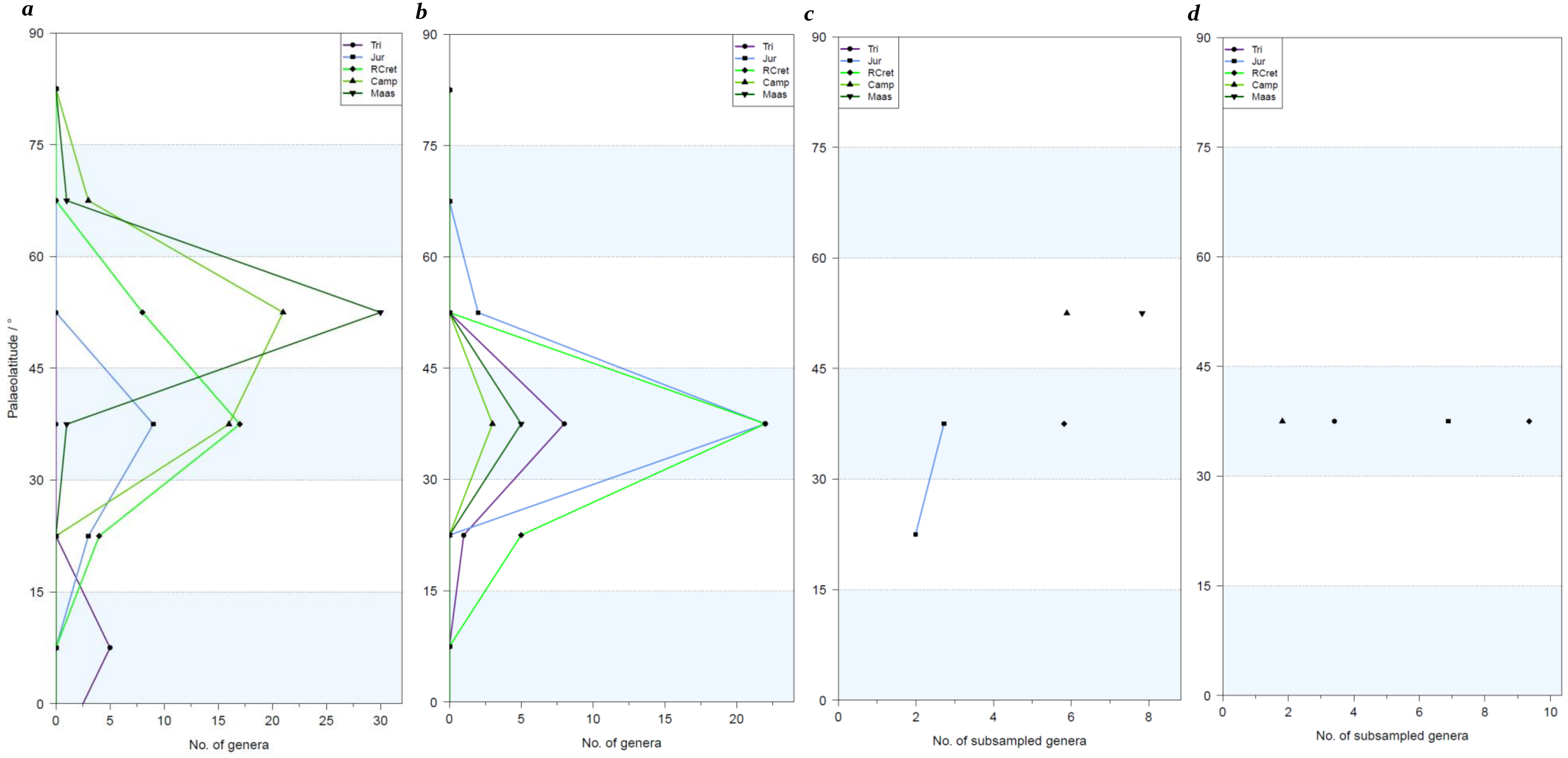

Figure 4.4: Latitudinal generic richness for Mesozoic non-marine lepidosaurs in North America and Europe. Figures show uncorrected (face-value) richness for (*a*) North America and *(b)* Europe, and subsampled (SQS) richness at quorum 0.4 for (*c*) North America and (*d*) Europe. Time slice abbreviations are Tri, Triassic; Jur, Jurassic; RCret, rest of Cretaceous (Berriasian–Santonian); Camp, Campanian; Maas, Maastrichtian.

Mesozoic lepidosaurs

Global uncorrected generic richness for lepidosaurs (Figure 4.3a) is very similar to that of turtles, but the Triassic and Jurassic had higher richness at 30–45°N. Additionally, the positions of the RCret and Campanian slices are reversed, with Campanian richness higher than RCret in the same latitudinal band. Uncorrected Campanian lepidosaur richness far exceeded turtle richness. The Maastrichtian again had its highest richness levels in the 45–60°N band, and its richness levels were about the same as those for turtles. Southern Hemisphere richness levels were low, especially compared to those in the north.

Once subsampled, most of these trends in the data disappear (Figure 4.3b). As a result, these data are not easily comparable with those of turtles, but the trends that are recovered appear markedly different. In the south, a relatively low level of richness was found for every bin (except the Jurassic, which is missing) in band 30–45°S, and there were low levels of richness in the RCret time slice between 30 and 60°S and in the Maastrichtian between 15 and 45°S. In the Northern Hemisphere, most individual points recovered were between 30 and 60°N, with richness in the RCret very high in the 30–45°N band. I recovered a low richness for the Triassic in band 30–45°N. The Jurassic had a low richness between 15 and 30°N and richness comparable with 45–60°N Campanian and Maastrichtian between 30 and 45°N. The Campanian and Maastrichtian had medium levels of lepidosaur richness at this high latitude; the richness levels of the latter were slightly higher. These results, excepting perhaps the RCret slice, are very different to the global turtle results.

Uncorrected North American latitudinal richness differed from the global average (Figure 4.4a). Most notably, Maastrichtian lepidosaur richness was the highest overall, with Campanian richness levels second, both within the 45–60°N band, as with the turtle record. The RCret and Jurassic slices peaked in band 30–45°N as in the global curves, and Triassic data are only present near the equator (0–15°N). The European lepidosaur curves were very similar to those obtained for turtles, except that the Jurassic had a much higher richness comparable to the RCret slice. In the uncorrected curve (Figure 4.4b) the Campanian and Maastrichtian have the lowest richness levels, and most occurrences lie within the 30–45°N band.

Unfortunately, the Mesozoic is very poorly sampled for both continents, so that even when subsampling at the low quorum of 0.4 few clear signals are recovered. In North America (Figure 4.4c), there was a low to medium level of lepidosaur richness in the Jurassic between 15 and 45°N. The RCret slice is recovered with a higher richness than the Jurassic at 30–45°N; the Campanian had the same level of richness as the RCret time bin, but at 45–60°N, and the Maastrichtian had the highest richness in the North American Mesozoic in this same band. Europe is very poorly sampled, and as with turtles the only richness result recovered lies in band 30–45°N (Figure 4.4d). The highest richness in this band was found in the RCret time slice, but unlike turtles this is followed by the richness recovered for the Jurassic, which is due largely to the high levels of uncorrected European lepidosaur data. The Triassic and Campanian both had low levels of richness, and the Maastrichtian is too poorly sampled to be recovered.

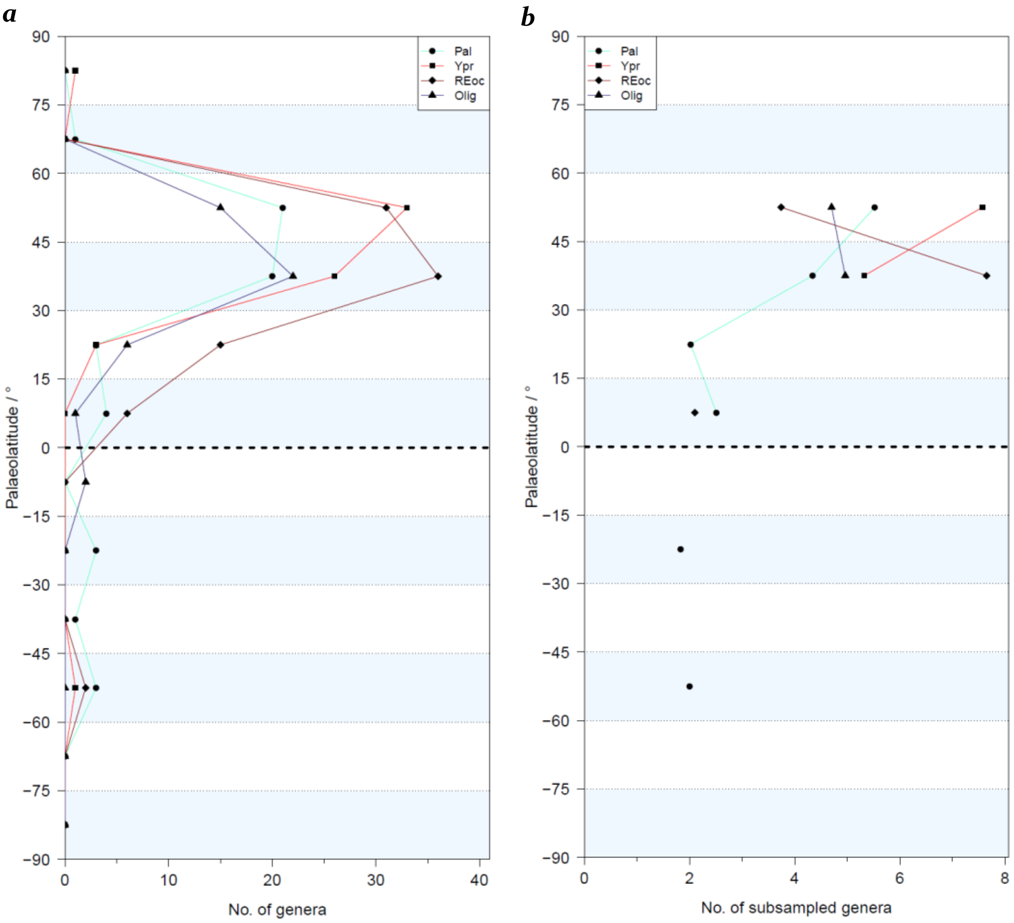

Figure 4.5: Latitudinal generic richness for global Paleogene non-marine turtles using *(a)* uncorrected (face-value) richness and *(b)* subsampled (SQS) richness at quorum 0.4. Time slice abbreviations are Tri, Triassic; Jur, Jurassic; RCret, rest of Cretaceous (Berriasian–Santonian); Camp, Campanian; Maas, Maastrichtian.

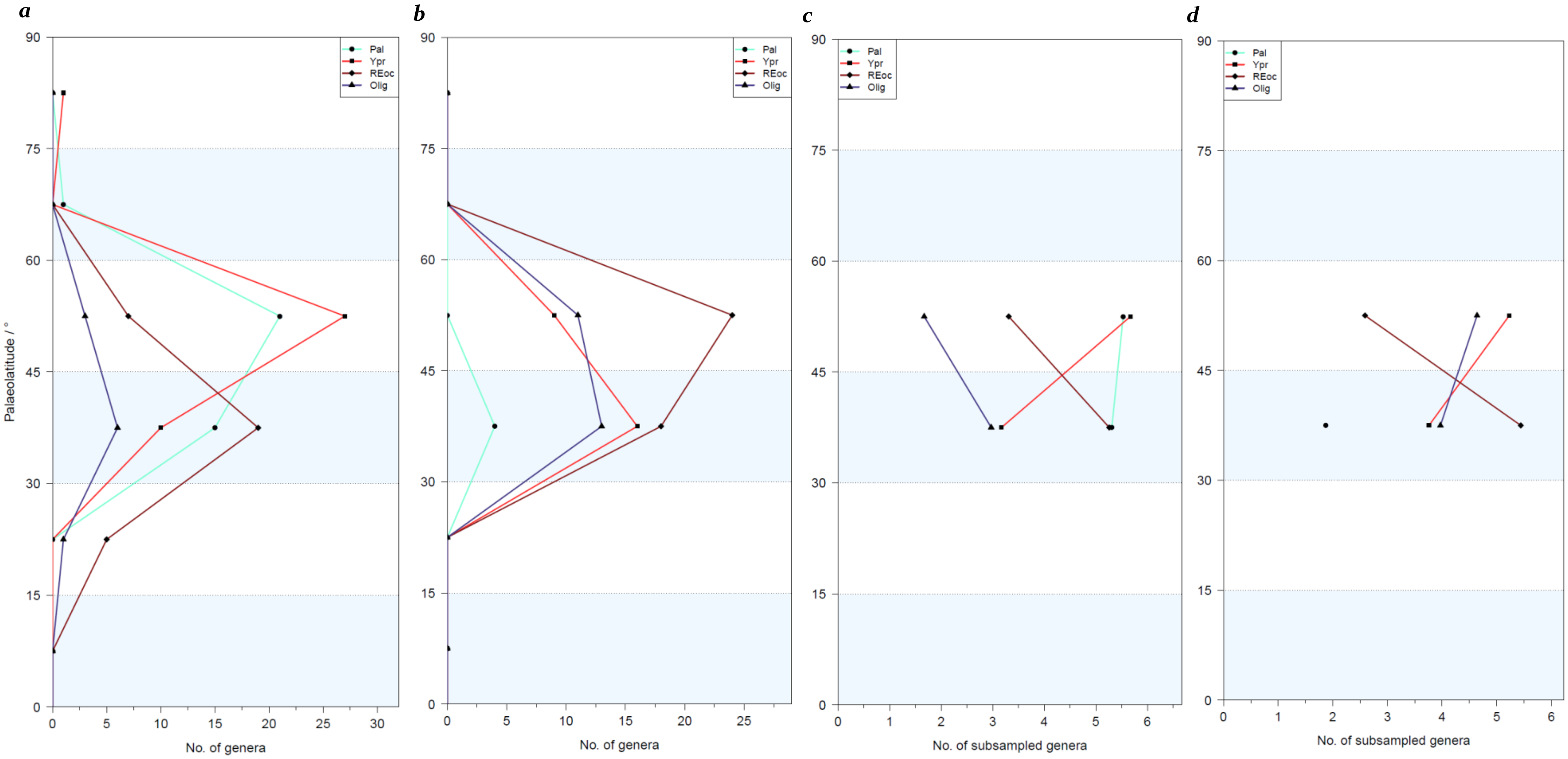

Figure 4.6: Latitudinal generic richness for Paleogene non-marine turtles in North America and Europe. Figures show uncorrected (face-value) richness for (*a*) North America and *(b)* Europe, and subsampled (SQS) richness at quorum 0.4 for (*c*) North America and (*d*) Europe. Time slice abbreviations are Tri, Triassic; Jur, Jurassic; RCret, rest of Cretaceous (Berriasian–Santonian); Camp, Campanian; Maas, Maastrichtian.

Paleogene turtles

As with the Mesozoic, there is a clear Northern Hemisphere peak in the uncorrected global generic richness (Figure 4.5a). The Paleocene and Ypresian show peaks at 45–60°N, though Paleocene richness levels were almost as high in the 30–45°N band just below. The collective "rest of the Eocene" time slice (REoc) and the Oligocene exhibited their peak generic richness in the 30–45°N band, with Oligocene turtle richness slightly higher than in the Paleocene. The REoc slice had a medium level of richness in the 15 to 30°N band. There were very small peaks in the Southern Hemisphere, particularly for the Paleocene at 15–30°S and for most time slices at 45–60°S.

Post-subsampling (Figure 4.5b), generic richness signals are not recovered in the majority of latitudinal bands due to poor sampling, and only a Paleocene richness result is recovered for the Southern Hemisphere. The Paleocene time slice had relatively low turtle generic richness at 15–30°S and 45–60°S. This low level of richness was also present between 0 and 30°N. Between 30 and 60°N there was a much higher level of richness in the Paleocene, with the highest richness levels for the time found at 45–60°N. This high richness was overshadowed by the Ypresian, which had the highest richness at this latitude, and a higher richness than the Paleocene at 30–45°N. The collective REoc slice had the highest richness overall, in the 30–45°N band, but a much lower richness in the 45–60°N band and low richness between 0 and 15°N. The Oligocene had a relatively consistent high richness of turtles between 30 and 60°N, comparable to the average Paleocene levels of richness in these two latitudinal bands.

In the raw continental data, the distributions of generic richness in North America and Europe are very different (Figure 4.6a, b). In North America, the placement of the peaks roughly corresponds with the global curve, though at different magnitudes. The Ypresian and Paleocene richness curves peak at 45–60°N and the former had the highest overall richness. There are some occurrences of turtles at 75–90°N in the Ypresian. The REoc and Oligocene slices had their highest peaks in the 30–45°N band but were much reduced compared to the global curves. In Europe the REoc slice had the highest overall richness, but in a higher band (45–60°N) than in the global curve. The Paleocene had a very low richness in Europe and only had occurrences between 30–45°N. The Ypresian and Oligocene have roughly the same richness levels between 30 and 60°N, but the Oligocene richness is slightly higher at 45–60°N and the Ypresian is slightly higher at 30–45°N.

Subsampled North American data (Figure 4.6c) changes little from the uncorrected data except during the Paleocene time bin. The Paleocene had a consistently high level of richness between 30 and 60°N, matching the highest richness of both the Ypresian and REoc slices. The Ypresian and REoc slices had opposite trends of roughly the same high and medium richness levels between 30 and 60°N, with the highest richness of the Ypresian at 45–60°N and the highest of the REoc at 30–45°N. The Oligocene had a low to medium turtle richness between these two latitudinal bands, with a comparable richness to the Ypresian in band 30–45°N. In European subsampled richness (Figure 4.6d) the Paleocene is only recovered at 30–45°N with a low richness, as this is the only latitude to have occurrence data. The REoc bin is reduced compared to the uncorrected figure, and the pattern between it and the Ypresian was similar to that found in North America, with opposites of richness between 30 and 60°N. The lowest REoc richness was much lower than the lowest Ypresian richness, however. The Oligocene and Ypresian had very similar richness levels, as in the uncorrected data, but the trend of one being higher than the other is different here. Instead, the Ypresian had a slightly higher richness at 45–60°N and the Oligocene had a slightly higher richness at 30–45°N. The difference between the Oligocene and Ypresian in the lower latitudinal band was smaller than in the higher band.

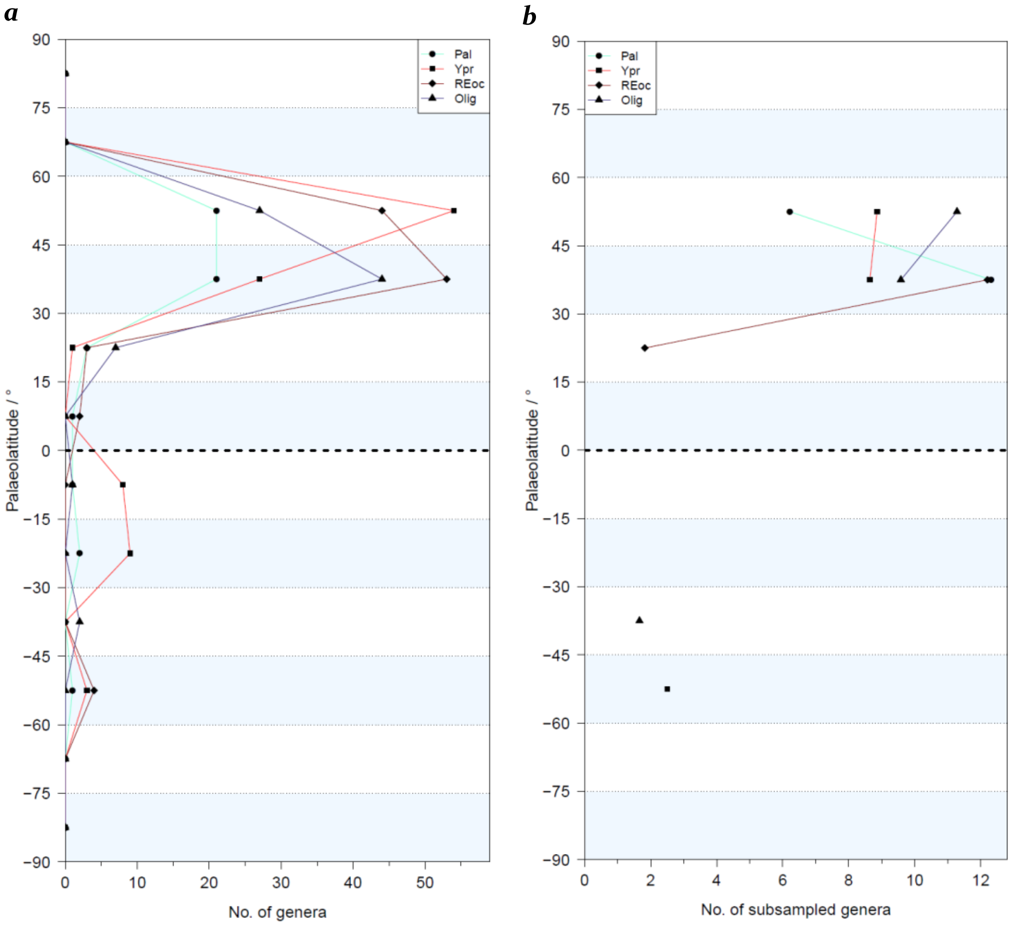

Figure 4.7: Latitudinal generic richness for global Paleogene non-marine lepidosaurs using *(a)* uncorrected (face-value) richness and *(b)* subsampled (SQS) richness at quorum 0.4. Time slice abbreviations are Tri, Triassic; Jur, Jurassic; RCret, rest of Cretaceous (Berriasian–Santonian); Camp, Campanian; Maas, Maastrichtian.

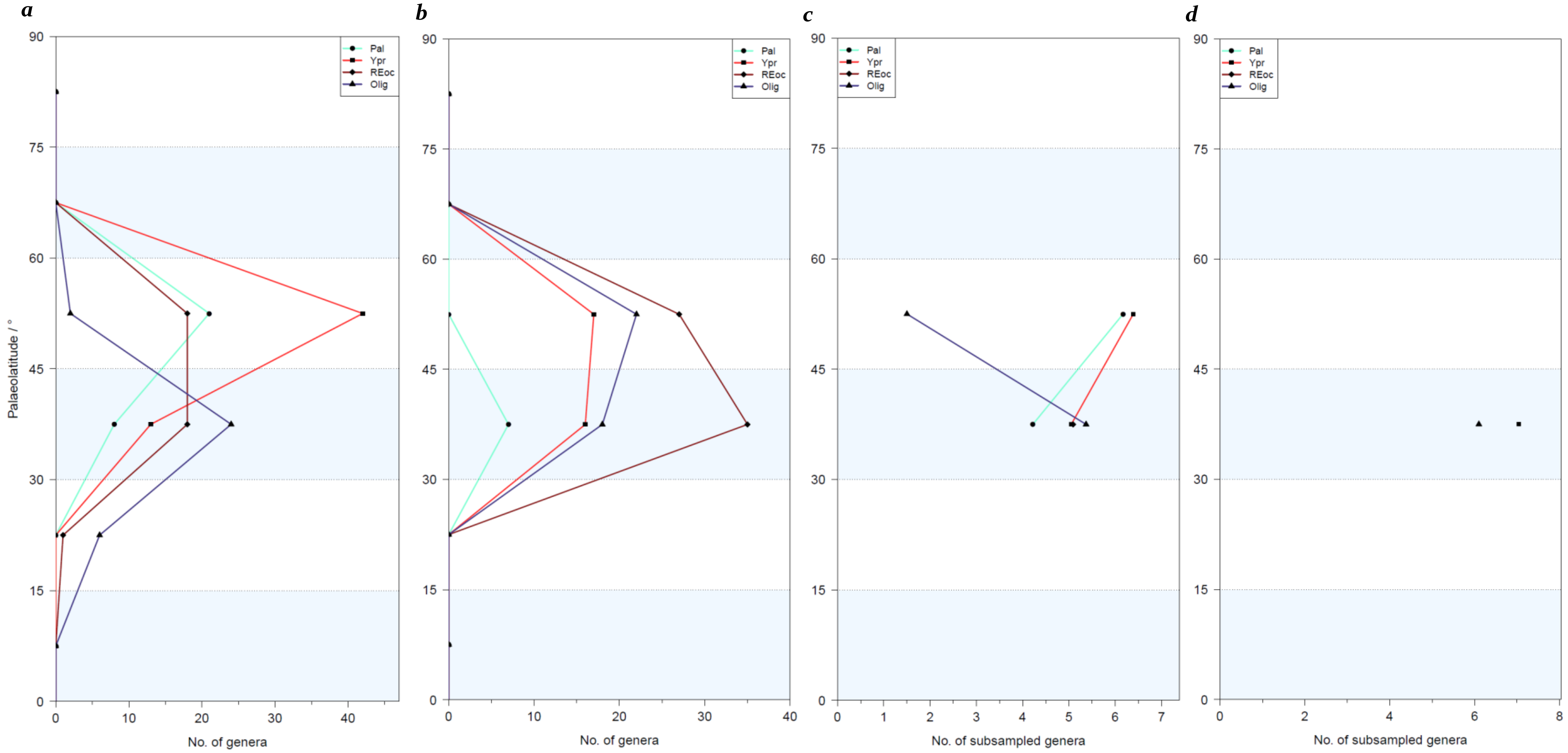

Figure 4.8: Latitudinal generic richness for Paleogene non-marine lepidosaurs in North America and Europe. Figures show uncorrected (face-value) richness for (*a*) North America and *(b)* Europe, and subsampled (SQS) richness at quorum 0.4 for (*c*) North America and (*d*) Europe. Time slice abbreviations are Tri, Triassic; Jur, Jurassic; RCret, rest of Cretaceous (Berriasian–Santonian); Camp, Campanian; Maas, Maastrichtian.

Paleogene lepidosaurs

Uncorrected Paleogene lepidosaur richness curves (Figure 4.7a) are very similar to those of turtles except for a higher Oligocene richness at 30–45°N. There was a low richness between 0 and 30°S for Ypresian lepidosaurs. The number of genera present overall is much higher than for turtles.

After SQS (Figure 4.7b), as with turtles, data within most of the bands outside 30–60°N drop out of the analysis due to poor sampling. There was low richness during the Ypresian between 45 and 60°S, and between 30 and 45°S during the Oligocene. In the Northern Hemisphere there was a high richness of lepidosaurs in the Paleocene between 30 and 45°N, with medium levels at 45–60°N. The REoc slice had the same high richness as the Paleocene at 30–45°N, and very low richness between 15 and 30°N. The Ypresian had a consistent relatively high level of richness at 30–60°N. Oligocene richness was higher than in the Ypresian in these two bands, and had a very high richness in the higher of the two (45–60°N).

North American uncorrected lepidosaur generic richness (Figure 4.8a) is somewhat similar to that for turtles during the Paleocene and Ypresian but different in the REoc and Oligocene. It is dissimilar to global uncorrected lepidosaur richness. Ypresian richness remains very high at 45–60°N, but the Paleocene and REoc slices had roughly the same, medium, richness in this band. In band 30–45°N richness was low to medium for everything, with the Paleocene having the lowest and Oligocene the highest richness levels. Europe is very different from the global data (Figure 4.8b), and somewhat similar to the pattern seen in turtles. The REoc peak was the opposite of that recovered for turtles, with higher richness at 30–45°N. The Paleocene had occurrences only at 30–45°N and very low richness. The Ypresian and Oligocene were very similar with higher richness in the Oligocene for both bands between 30 and 60°N.

Subsampling the North American data recovers signals between 30–60°N only (Figure 4.8c). The REoc slice is only well enough sampled at 30–45°N and had the same richness as most of the other slices in this band except the Paleocene, which had a slightly lower richness. At 45–60°N the Paleocene and Ypresian had a very similar high richness, while the Oligocene had a very low richness. This band exhibits very similar richness to that of turtles. The European Paleogene lepidosaur record is very poorly sampled outside of band 30–45°N (Figure 4.8d), and only the Ypresian and Oligocene are recovered at quorum 0.4, the former with marginally higher richness.

4.3.2 Generalized least-squares analyses of latitudinal bands

The results of the GLS analyses, including coefficients and individual variable values, can be found in Tables 4.2 and 4.3 for turtles and lepidosaurs, respectively. For Jurassic turtles, I found that the null model was the best model for explaining variation in patterns of richness (AICc= 19.99, R^2= 0.54). For lepidosaurs, it was the number of tetrapod-bearing formations (TBFs; AICc= 17.47, R^2= 0.51) very closely followed by the null model (AICc= 17.68, R^2= 0.46). Both turtles and lepidosaurs had a very low N for this time period (4 and 5, respectively).

The model best fitting latitudinal richness patterns in the Early Cretaceous was TBFs for both turtles (AICc= 20.97, R^2= 0.41) and lepidosaurs (AICc= 21.17, R^2= 0.49). For turtles the number of tetrapod-bearing collections (TBCs) was a close second (AICc= 21.05, R^2= 0.41). In the Late Cretaceous the best model picked by AICc was TBCs for turtles (AICc= 15.69, R^2= 0.68) and TBFs for lepidosaurs (AICc= 24.72, R^2= 0.55).

For Early Paleogene turtles, AICc selected TBCs as the best model (AICc= 17.38, R^2= 0.75) with TBFs close behind (AICc= 17.48, R^2= 0.75), while for lepidosaurs land area was chosen (AICc= 26.65, R^2= 0.26). In the Late Paleogene data TBFs were selected as the best model for both turtles (AICc= 16.24, R^2= 0.39) and lepidosaurs (AICc= 22.88, R^2= 0.33). As with the Jurassic both clades had a low N for this time period, with N= 5 for turtles and N= 6 for lepidosaurs.

Table 4.2: Summary of model fits using GLS with autoregressive model of order one to per-latitudinal band richness of non-marine turtles in several time periods, ordered by best to worst model as chosen by AICc. Model abbreviations are C, tetrapod-bearing collections; F, tetrapod-bearing collections; L, land area. Column abbreviations are JB, Jarque-Bera results and BP, Breusch-Pagan results.

	Model	N	Sampling (C or F)			Land area (L)			Log-likelihood	R^2	Phi	J-B	B-P	AICc	AICc weight
			Value	*t*-value	*p*-value	Value	*t*-value	*p*-value							
Jurassic	**Null**	**5**	**-**	**-**	**-**	**-**	**-**	**-**	**-6.333**	**0.000**	**0.326**	-	-	**19.999**	**5.410E-01**
	F	5	0.918	7.006	0.006	-	-	-	-3.676	0.268	-0.680	0.782	0.735	21.352	2.750E-01
	C	5	0.590	2.598	0.081	-	-	-	-4.757	0.169	0.050	0.717	0.339	23.513	9.334E-02
	L	5	-	-	-	0.073	0.036	0.973	-4.786	0.166	0.374	0.767	0.145	23.572	9.062E-02
	FL	5	0.815	3.062	0.092	0.056	0.056	0.961	-3.129	0.314	-0.150	0.672	0.960	40.257	2.159E-05
	CL	5	0.547	2.030	0.179	-0.129	-0.092	0.935	-3.624	0.273	0.274	0.736	0.317	41.248	1.315E-05
E Cretaceous	**F**	**9**	**0.767**	**4.988**	**0.002**	**-**	**-**	**-**	**-5.482**	**0.408**	**0.490**	**0.737**	**0.058**	**20.964**	**3.587E-01**
	C	9	0.650	4.807	0.002	-	-	-	-5.523	0.406	0.596	0.709	0.515	21.045	3.445E-01
	CL	9	0.547	4.706	0.003	0.709	2.189	0.071	-3.723	0.519	0.666	0.589	0.730	22.247	1.889E-01
	FL	9	0.692	3.456	0.014	0.322	0.629	0.552	-5.050	0.438	0.499	0.699	0.106	24.900	5.014E-02
	L	9	-	-	-	1.180	1.993	0.087	-7.815	0.222	0.762	0.683	0.690	25.630	3.481E-02
	Null	9	-	-	-	-	-	-	-9.945	0.000	0.840	-	-	26.461	2.297E-02
L Cretaceous	**C**	**8**	**0.817**	**12.634**	**0.000**	**-**	**-**	**-**	**-2.644**	**0.681**	**0.345**	**0.602**	**0.290**	**15.687**	**8.552E-01**
	CL	8	0.859	14.177	0.000	-0.532	-1.650	0.160	-1.749	0.713	0.614	0.704	0.244	19.498	1.272E-01
	F	8	1.133	5.622	0.001	-	-	-	-6.689	0.487	0.108	0.674	0.922	23.778	1.497E-02
	FL	8	1.188	5.003	0.004	-0.477	-0.574	0.591	-5.874	0.534	0.257	0.804	0.756	27.749	2.055E-03
	Null	8	-	-	-	-	-	-	-12.360	0.000	0.219	-	-	31.387	3.333E-04
	L	8	-	-	-	1.171	0.688	0.517	-10.683	0.179	0.215	0.809	0.067	31.766	2.757E-04
E Paleogene	**C**	**9**	**0.714**	**16.388**	**0.000**	**-**	**-**	**-**	**-3.691**	**0.746**	**-0.045**	**0.386**	**0.899**	**17.381**	**4.528E-01**
	F	9	1.161	14.615	0.000	-	-	-	-3.740	0.745	0.036	0.637	0.115	17.479	4.312E-01
	CL	9	0.773	15.540	0.000	-0.383	-1.722	0.136	-3.064	0.764	-0.372	0.698	0.304	20.927	7.690E-02
	FL	9	1.066	12.770	0.000	0.290	1.276	0.249	-3.741	0.745	-0.287	0.646	0.252	22.282	3.905E-02
	L	9	-	-	-	1.590	1.745	0.124	-13.574	0.188	0.067	0.784	0.105	37.149	2.309E-05
	Null	9	-	-	-	-	-	-	-15.341	0.000	0.431	-	-	37.254	2.191E-05
L Paleogene	**F**	**5**	**0.948**	**10.926**	**0.002**	**-**	**-**	**-**	**-1.117**	**0.389**	**0.153**	**0.503**	**0.889**	**16.235**	**5.630E-01**
	Null	5	-	-	-	-	-	-	-5.305	0.000	1.000	-	-	17.943	2.397E-01
	C	5	0.760	11.443	0.001	-	-	-	-2.686	0.265	-0.420	0.778	0.810	19.372	1.173E-01
	L	5	-	-	-	2.299	0.985	0.397	-3.070	0.231	0.982	0.769	0.462	20.141	7.986E-02
	FL	5	0.984	5.119	0.036	-0.220	-0.199	0.861	-0.181	0.453	0.068	0.649	0.363	34.362	6.520E-05
	CL	5	0.929	4.790	0.041	-1.431	-0.879	0.472	-0.949	0.401	-0.562	0.834	0.850	35.898	3.025E-05

Table 4.3: Summary of model fits using GLS with autoregressive model of order one to per-latitudinal band richness of non-marine lepidosaurs in several time periods, ordered by best to worst model as chosen by AICc. Model abbreviations are C, tetrapod-bearing collections; F, tetrapod-bearing collections; L, land area. Column abbreviations are JB, Jarque-Bera results and BP, Breusch-Pagan results.

	Model	N	Sampling (C or F)			Land area (L)			Log-likelihood	R^2	Phi	J-B	B-P	AICc	AICc weight
			Value	*t*-value	*p*-value	Value	*t*-value	*p*-value							
Jurassic	**F**	**4**	**0.883**	**148.75**	**0.000**	**-**	**-**	**-**	**1.266**	**0.512**	**-1.000**	**0.717**	**0.282**	**17.468**	**5.090E-01**
	Null	4	-	-	-	-	-	-	-4.840	0.000	-0.661	-	-	17.681	4.576E-01
	C	4	0.936	5.515	0.031	-	-	-	-1.629	0.315	-0.627	0.777	0.911	23.258	2.815E-02
	L	4	-	-	-	-0.751	-0.392	0.733	-3.313	0.164	-0.530	0.761	0.846	26.626	5.225E-03
	CL	4	1.131	12.652	0.050	0.916	2.957	0.208	-0.221	0.419	0.000	0.757	0.429	∞	0.00000
	FL	4	1.251	2.728	0.224	0.844	0.623	0.645	-1.535	0.322	0.000	0.862	0.203	∞	0.00000
E Cretaceous	**F**	**7**	**1.261**	**10.422**	**0.000**	**-**	**-**	**-**	**-5.087**	**0.494**	**-0.456**	**0.623**	**0.496**	**21.174**	**6.857E-01**
	C	7	1.062	9.349	0.000	-	-	-	-6.173	0.425	-0.553	0.709	0.216	23.346	2.314E-01
	FL	7	1.283	6.283	0.003	-0.082	-0.122	0.909	-4.647	0.520	-0.346	0.620	0.533	27.295	3.214E-02
	L	7	-	-	-	1.956	1.369	0.229	-8.776	0.219	0.363	0.576	0.276	28.551	1.715E-02
	Null	7	-	-	-	-	-	-	-10.879	0.000	0.539	-	-	28.559	1.708E-02
	CL	7	0.977	6.668	0.003	0.570	0.959	0.392	-5.318	0.480	-0.527	0.663	0.640	28.635	1.644E-02
L Cretaceous	**F**	**9**	**1.103**	**6.405**	**0.000**	**-**	**-**	**-**	**-7.358**	**0.549**	**0.525**	**0.519**	**0.399**	**24.716**	**5.474E-01**
	C	9	0.852	7.432	0.000	-	-	-	-8.182	0.503	-0.224	0.720	0.658	26.363	2.402E-01
	FL	9	1.047	5.766	0.001	0.703	1.010	0.352	-6.292	0.602	0.504	0.427	0.811	27.385	1.441E-01
	CL	9	0.795	6.295	0.001	0.676	0.980	0.365	-7.173	0.559	-0.082	0.608	0.959	29.145	5.976E-02
	L	9	-	-	-	1.923	1.221	0.262	-12.033	0.219	0.460	0.860	0.174	34.065	5.106E-03
	Null	9	-	-	-	-	-	-	-14.132	0.000	0.404	-	-	34.835	3.475E-03
E Paleogene	**L**	**7**	**-**	**-**	**-**	**2.196**	**1.892**	**0.117**	**-7.823**	**0.261**	**0.577**	**0.598**	**0.455**	**26.646**	**4.528E-01**
	Null	7	-	-	-	-	-	-	-10.398	0.000	0.784	-	-	27.596	2.815E-01
	F	7	0.912	1.751	0.140	-	-	-	-8.871	0.164	0.499	0.686	0.303	28.743	1.587E-01
	C	7	0.453	1.606	0.169	-	-	-	-9.829	0.065	0.366	0.642	0.547	30.659	6.088E-02
	FL	7	0.578	0.980	0.382	1.608	1.226	0.287	-6.952	0.333	0.575	0.553	0.548	31.905	3.265E-02
	CL	7	0.152	0.396	0.712	1.840	1.173	0.306	-7.828	0.261	0.549	0.399	0.799	33.657	1.360E-02
L Paleogene	**F**	**6**	**1.471**	**5.287**	**0.006**	**-**	**-**	**-**	**-5.442**	**0.328**	**0.313**	**0.493**	**0.522**	**22.883**	**5.322E-01**
	Null	6	-	-	-	-	-	-	-8.825	0.000	1.000	-	-	24.650	2.201E-01
	L	6	-	-	-	0.156	0.048	0.964	-6.784	0.213	1.000	0.694	0.793	25.568	1.391E-01
	C	6	0.927	2.829	0.047	-	-	-	-7.245	0.170	0.460	0.398	0.792	26.491	8.765E-02
	FL	6	1.697	4.154	0.025	-1.000	-0.640	0.568	-3.917	0.439	0.155	0.884	0.115	29.833	1.648E-02
	CL	6	1.665	4.335	0.023	-3.339	-1.511	0.228	-5.211	0.346	-0.468	0.791	0.261	32.422	4.517E-03

4.4 Discussion

4.4.1 Latitudinal biodiversity gradients through the Mesozoic

There is a clear peak in Northern Hemisphere richness for Mesozoic non-marine turtles and lepidosaurs (Figs. 4.1 and 4.3), centred around 30–60°N, with earlier Mesozoic periods having higher richness levels at slightly lower latitudes, particularly for turtles (e.g. Jurassic peak at 15–30°N). Global richness levels are unreliable (see previous chapters for further discussion) but, due to sampling issues eliminating a lot of information, I will compare differences between the two clades at a global scale before moving on to discuss the continental results. I will focus primarily on the subsampled richness plots, as these are more reliable indicators of generic richness than uncorrected occurrence data plots.

The Triassic is very poorly sampled for both clades. Both clades were much less diverse, having potentially originated in the Triassic or just prior (e.g. Li et al., 2008; Pyron, 2010; Mulcahy et al., 2012; Jones et al., 2013; Joyce, 2015), and this low richness shows in both the uncorrected and subsampled data (when recovered). In the North American and European plots (Figs. 4.2 and 4.4) the Triassic either is not recovered (as in North America where the Triassic record is scarce) or had very low richness (Europe). This contrasts somewhat with the Late Triassic dinosaur record, which had a medium richness between 30–60°N and low richness at 0–30°N (Mannion et al., 2012), and contrasts greatly with that of crocodiles, which had a strong palaeotropical-centred richness peak at this time (Mannion et al., 2015) that was similar to modern day tropical richness peaks (Willig, Kaufman and Stevens, 2003).

Jurassic intervals are better sampled and turtle richness, while still relatively low, was much higher than other Jurassic bins in band 15–30°N. This is due to the large numbers of localities from southern Asia (particularly China), and the large-scale originations in the Late Jurassic of crown-group turtle clades (Anquetin et al., 2009). Lepidosaurs had a low level of richness in the Jurassic in this band, with higher richness at 30–45°N due to good European and North American records. Richness is much lower (when recovered) after splitting the data into continents. Jurassic richness appears to have been relatively consistent for North American turtles and lepidosaurs between 30 and 45°N, but was relatively higher in Europe for lepidosaurs, likely partially aided by the presence of major Lagerstätten in the Late Jurassic (Cerin, Solnhofen) and the Early Jurassic fissure fill localities of the United Kingdom. To a certain extent, the lower latitudinal richness of the Jurassic compared to later intervals represents the past positions of the most heavily-sampled areas of the Northern Hemisphere continents as they drifted northwards after the breakup of Pangaea. My global turtle richness results appear to correspond with the palaeotropical peak found for dinosaurs when examining SQS-corrected richness, with highest Jurassic richness at 0–30°N (Mannion et al., 2012).

The collective "rest of the Cretaceous" (RCret) time slice had a high richness of turtles and lepidosaurs. The highest Mesozoic richness levels of both clades are found at 30–45°N in this slice (though the slice is better sampled over multiple latitudes for turtles). In North America the RCret slice had a lower turtle and lepidosaur richness compared to the combined global and European data, partially due to large gaps in sampling during the Early Cretaceous in North America. There was a low richness of lepidosaurs in the Southern Hemisphere between 30 and 60°S comprised of mostly Argentinian sphenodontians and snakes, and a low to medium richness of primarily Australian turtles between 30 and 90°S, a wider range and higher richness than today (Roll et al., 2017). Unfortunately I lack data for either clade in lower latitudes, so cannot compare results to those for non-marine crocodylians, which in the early Late Cretaceous of North Africa had richness levels comparable to palaeotemperate regions in other Cretaceous time slices (Mannion et al., 2015).

The Campanian and Maastrichtian are difficult to compare with the preceding RCret slice, as a much higher RCret richness is likely, at least in part, to be a product of its much longer slice duration, so I will not comment on how richness levels compare directly. These two latest Cretaceous stages are better sampled for turtles than lepidosaurs, but at 45–60°N both clades had a slightly higher global richness in the Maastrichtian than the Campanian. For turtles there was a much higher richness in the Campanian than the Maastrichtian in the 30–45°N band (Figure 4.1b), meaning that turtles were richest at a higher latitude in the Maastrichtian than in the Campanian. However, the Campanian was richer overall, as the lowest richness in the Campanian was only slightly lower than the highest Maastrichtian richness. This trend changes somewhat when examining the North American turtle data (Figure 4.2b); the Campanian and Maastrichtian were both slightly richer at 45–60°N than 30–45°N, with the Campanian still richer than the Maastrichtian in the latter band albeit with a smaller difference between the two stages. In Europe, which is not well sampled in the late Mesozoic, the Campanian was also richer than the Maastrichtian at 30–45°N for turtles, explaining the larger difference in the global Northern Hemisphere curves compared to North America. European Campanian lepidosaur richness was very low, lower than any previous Mesozoic time slice, and the Maastrichtian was too poorly sampled to be recovered at quorum 0.4.

In the Southern Hemisphere there was relatively low latest Cretaceous turtle richness, which was contributed to by localities in Argentina and Brazil at 30–45°S, and a low richness of snakes in India and Madagascar during the Maastrichtian at 15–45°S. Both lepidosaurs and turtles had relatively high richness in the latest Cretaceous, potentially owing to favourable environmental conditions after the Cretaceous Thermal Maximum (Forster et al., 2007) that allowed for wide geographic ranges, even for thermophilic taxa. Unfortunately a lot of the world is poorly sampled for this period outside North America, making it difficult to investigate if the latitudinal richness of other continents followed the same trends or not. In the case of lepidosaurs, even North America is poorly sampled, and so there is less that can be said about their richness for the latest Cretaceous and, indeed, for much of the Mesozoic.

Very high crocodylomorph richness in the Late Cretaceous Adamantina Formation at palaeotropical latitudes in the Southern Hemisphere could suggest that crocodylomorphs did not experience a palaeotemperate peak in richness at this time (Mannion et al., 2015), in contrast to dinosaurs (Mannion et al., 2012) and the results for lepidosaurs and turtles presented herein, but data for all of these groups are currently limited due to poor sampling. Other studies of the plant and brachiopod fossil records suggest a palaeotemperate peak also (Rees et al., 2004; Powell, 2007; Coiffard and Gomez, 2012; Peralta-Medina and Falcon-Lang, 2012), but these studies do not utilise sampling correction and so their findings may be less robust.

My Mesozoic turtle richness results correspond almost exactly to the analyses carried out by Nicholson et al. (2016), with the exception of the RCret and Maastrichtian richness curves being shallower in their work. This is probably due to revisions of various turtle taxa that took place between our separate data downloads from the Paleobiology Database, as a number of papers reviewing Mesozoic turtle taxonomy have been published since 2016 (e.g. Anquetin, Püntener and Joyce, 2017; Georgalis and Joyce, 2017; Joyce, 2017). Regardless, my reassessment of turtles from this time period reaffirm their analyses and I have demonstrated the heavy influence of North America on many of the 'global' trends in Mesozoic turtle latitudinal richness.

4.4.2 Latitudinal biodiversity gradients through the Paleogene

As with the Mesozoic, there is a strong peak in generic richness for both clades in the Northern Hemisphere between 30 and 60°N for the Paleogene. For turtles, only the Paleocene is well-sampled enough for any diversity signal to be recovered outside the Northern Hemisphere, with a relatively low richness of turtles in Bolivia and Argentina. Low richness was also found between the equator and 30°N, while a medium to relatively high richness was present at 30–60°N, with the highest richness in the northernmost of the two bands. For lepidosaurs, which are more poorly sampled, a Paleocene signal is only recovered between 30–60°N after subsampling, and with the opposite trend to turtles, with a higher richness at 30–45°N. When looking at continental richness, the Paleocene is poorly sampled in Europe for both clades, except for low turtle richness between 30 and 45°N with no occurrences outside this latitude. In North America, Paleocene turtle richness was consistently very high between 30 and 60°N, as turtles passed through the K-Pg mass extinction event relatively unscathed and diversified shortly afterwards (Holroyd et al. 2014; see also Chapter 3). Lepidosaurs also had a high richness in the North American Paleocene, with a greater richness at 45–60°N than 30–45°N. Rose et al. (2011) found a very flat richness gradient for non-volant mammals in the Western Interior of North America (compared to the steep gradient of modern mammals) after using some measure of subsampling (not well explained in their paper), suggesting that the modern LBG did not exist in the Paleocene. However, the authors did not sample any contemporaneous palaeotropical-latitude localities, and so their conclusions cannot be applied outside this area.

The Ypresian was a time of continued high richness for both clades. There was very high turtle and lepidosaur richness at 30–60°N, with relatively high turtle richness at 30–45°N (comparable with the high Paleocene richness of 45–60°N). This high richness was contributed to by localities in both North America and Europe, which both had similar patterns of high turtle richness at 45–60°N and lower richness at 30–45°N (with lower richness in North America than Europe in the lower latitudinal band). Lepidosaurs also had a relatively high consistent Ypresian richness between 30–60°N. Richness in the 45–60° band was not as high as for turtles due to the poorly sampled European record at this latitude. Ypresian richness was very high at 30–45°N in Europe. North America, however, had the same pattern of higher richness in the 45–60°N band as turtles and that closely matches the pattern observed in Paleocene lepidosaurs on this continent. The continued high richness of both clades at higher latitudes may be attributed to the Paleocene-Eocene Thermal Maximum (PETM) and Early Eocene Climatic Optimum (EECO) events between the end of the Paleocene and the early Eocene (mid-Ypresian: Zachos et al. 2008; Hyland et al. 2017). Latitudinal temperature gradients were low, and the polar regions lacked ice, which allowed the spread of organisms, particularly ectotherms, to much higher latitudes than the present day (Zachos, Dickens and Zeebe, 2008; Bijl et al., 2009; Payros et al., 2015; Roll et al., 2017). Iguanians first appeared in Europe at this time (Augé, 2005), increasing richness, and higher temperatures would have enabled many ectotherms to move into niches that were previously thermally limiting both for climatic tolerances and body size (Head et al., 2009, 2013; Smith, 2009). For example, the warmer climate is thought to have facilitated the evolution and dispersal of large herbivorous turtles in North America, some of which reached very high northern latitudes (above 75°N, see Figure 4.6a; Estes and Hutchison, 1980), and the increased presence of more thermally-limited pleurodire turtles in Europe (Pérez-García, 2017).

Crocodylians were also present at very high latitudes at this time, and were still diverse between 30–60°N (Mannion et al., 2015). Polar Eocene plant richness was apparently comparable to modern temperate latitudes (Harrington et al., 2012). Fossil insect richness was analysed by Archibald et al. (2010), who found that palaeotemperate (>50°N) richness was higher than modern temperate richness and comparable to extant tropical insect richness. They suggested that a low "tropics-like" seasonality enabled this high richness, but did not sample any contemporaneous Eocene palaeotropical localities. It is possible therefore, that palaeotropical early Eocene richness was even higher than the tropics today, and that the modern LBG still existed in the Paleogene.

After the Ypresian, the richness of both turtles and lepidosaurs decreased both globally and in North America and Europe (see Chapters 2 and 3). In the global turtle latitudinal data, there was a continued high richness in the "rest of the Eocene" (REoc) time slice (comparable to peak global Ypresian richness) at 30–45°N, but lower richness in the 45–60°N band. In both North America and Europe there was lower turtle richness in the REoc at 45–60°N, the opposite trend to the Ypresian. In North America this richness level was comparable to the lowest Ypresian richness in band 30–45°N, and in Europe (at least for turtles) it was close to the low richness of the poorly sampled Paleocene. In the lepidosaur record 45–60°N is too poorly sampled to be recovered at both global and the continental scales, but richness at 30–

45°N was still very high, matching peak Paleocene richness. There was also a very low richness recovered at 15–30°N in the global lepidosaur curve that is not recovered in the two continental curves, as the REoc slice is poorly sampled and only appears at 30–45°N in North America, with a similar richness to the Ypresian. It is possible to attribute this change in latitudinal richness patterns to the gradual cooling of the climate and increasing aridity in the later Eocene (Zachos et al., 2001; Woodburne, Gunnell and Stucky, 2009), with higher latitude ectotherms retreating to warmer southerly latitudes as temperature gradients increased. This is proposed to be the case for lepidosaurs (Smith, 2011) and turtles (Hutchison, 1992) in the late Eocene and early Oligocene of North America, but the issue is likely exacerbated by poor sampling in the middle Eocene and Oligocene (see Chapters 2 and 3).

Turtle and lepidosaur richness patterns differ considerably from each other during the Oligocene. In the former, there was a consistent medium-to-high level of richness between 30 and 60°N, at approximately the average of the Paleocene in these two bands. Lepidosaurs, in comparison, had a very high Oligocene richness, higher than the Ypresian, with peak richness at 45–60°N. In the North American curves, both clades had a higher richness at 30–45°N than 45–60°N, with turtles having a low to medium overall richness, and lepidosaurs having a low richness at 45–60°N but a relatively high richness at 30–45°N. Both clades had a peak Oligocene richness roughly comparable to the lowest Ypresian richness levels. Notably, the North American lepidosaur trend of higher richness in the 30–45°N band was contrary to the overall global lepidosaur trend. European lepidosaurs are too poorly sampled at 40–60°N but presumably are well sampled enough when combined with the other two Northern Hemisphere continents to produce the higher richness observed in the global curve. European Oligocene richness at 30–45°N is high, but lower than corresponding Ypresian richness. North American and European turtle richness were opposites, with slightly higher richness at 45–60°N than at 30–45°N in Europe; European richness was much higher overall, comparable to the Ypresian though slightly lower in the more northerly latitudinal band.

North American lepidosaurs appear to have maintained a high richness at 30–45°N while turtles had a lower richness than previous stages. This could be because of the effects of increasing aridity and decreasing temperatures at the time, exacerbated by the final breakup of Antarctica from the other Southern Hemisphere continents, which caused a relatively sudden temperature drop and an increase in seasonality (Zachos, Dickens and Zeebe, 2008; Bijl et al., 2009; Archibald et al., 2010). The Grande Coupure, an event where several groups of terrestrial vertebrates (particularly mammals, and possibly lepidosaurs) underwent turnover due to rapid climatic change (e.g. Stehlin 1909; Hooker 1989; Rage & Augé 1993; Hooker et al. 2004; Augé & Smith 2009; Costa et al. 2011; Eronen et al. 2015), is possibly supported by the higher richness of North American ectotherms at 30–45°N compared to 45–60°N. Faunas in the continental interior shifted to terrestrial testudinid-dominated assemblages, and taxa such as the pond turtles (Emydidae) disappeared from the record. As the latter are found in modern North America, it is likely these taxa migrated southwards and returned when conditions were more favourable (Hutchison, 1992). Lepidosaurs also supposedly experienced some turnover in North America, though not to the same extent as turtles, with taxa of “tropical affinities” (e.g. stem corytophanine and polychrotine lizards) going extinct or moving southwards where there are no coeval localities to examine richness (Smith, 2011).

The event does not appear to effect European turtles in the same way; although richness still decreased for the continent, it remained relatively high (see Chapter 3). Pleurodires, which are more temperature-dependent (Pérez-García, 2017), decreased in richness, while cryptodires increased. Unfortunately, lepidosaurs are too poorly sampled to study any trends in latitudinal richness changes for the Oligocene. Perhaps the differences between Europe and North America exist because turtles are more heavily affected by the availability of aquatic environments (and the rainfall that partially drives their availability) than temperature (Iverson, 1992), and Europe did not experience as marked a shift towards aridity as North America at this time? It could also be the case that Europe is better sampled over a wider geographic area at this time than North America. There is a heavy bias in the lack of available North American localities in the late Oligocene (Joyce and Bourque, 2016), which may have caused the very low richness of ectotherms in the 45–60°N band. European lepidosaurs are, unfortunately, too poorly sampled to reveal much about Oligocene latitudinal richness here. Non-marine crocodylians decline over the Eocene–Oligocene boundary with low richness at palaeotemperate latitudes, suggesting a common response in ectotherms, but poor sampling elsewhere renders it difficult to determine additional trends in richness gradients (Mannion et al., 2015).

The modern LBG supposedly developed around the Eocene–Oligocene boundary (Thomas and Gooday, 1996; Crame et al., 2001; Archibald et al., 2010; Fenton et al., 2016). From a taxonomic perspective, the Oligocene is when modern lineage compositions began to appear for both turtles and lepidosaurs. In terms of latitudinal distribution, by the Oligocene North American ectotherm latitudinal gradients were approaching those of today, with richness highest below 40°N (modern turtle richness tracks the drainage basins of the southeast United States and is very high, while lepidosaurs are more equally distributed: Buhlmann et al., 2009; Roll et al., 2017). In Europe, Oligocene ectotherm latitudinal distribution is still very different from the modern day, as turtle and lepidosaur richness in Europe today is very low compared to other continents (Roll et al., 2017). This different distribution and high richness could have partially been contributed to by the land connection between Asia and Europe, which allowed for taxon migration and interchange (Hooker, Collinson and Sille, 2004), and created a very different situation to North America. There are also differences in the amount of land area between Europe and North America that could have contributed to these differences.

Overall, there appears to be a consistent peak in richness in palaeotemperate latitudes (at least in the Northern Hemisphere) from the Triassic until at least the Paleogene for both turtles and lepidosaurs. While individual fluctuations in time slices sometimes differ, these similar richness peaks (corresponding in part to those found for dinosaurs; Mannion et al., 2012) suggests that there may be similar environmental driving factors affecting the distributions of the two clades through time. Surprisingly, despite similar life histories to turtles, non-marine crocodile latitudinal richness (Mannion et al., 2015) does not appear to correspond to my results here. This could be due to the decreased tolerance for cool temperatures in the majority of crocodylians. In North America, thermal tolerances of crocodylians appear to be lower, with cool-tolerant alligatorines persisting (Hutchison, 1992) while other taxa either migrate south or go extinct (e.g. Crocodylidae; Markwick, 1994). Alligators remain dominant on this continent today in richness and geographic abundance, with the highest latitude distribution of all extant crocodylians (Markwick, 1994).

Of course, it is difficult to properly assess LBGs of turtles and lepidosaurs compared to the tropical peaks (30°N–30°S) we find today due to the overwhelmingly poor sampling of the record outside of 30–60°N, minus a few exceptions. I can comment on how richness fluctuates within this area, but the missing data from the surrounding latitudes obscures any major conclusions I could draw about whether changes in latitudinal richness can be attributed to migrations, extinction events, or both. This is particularly true in the Southern Hemisphere, which tends to produce isolated richness points at latitudes that contain better-sampled countries (e.g. Brazil, Argentina, Indo-Madagascar) but little elsewhere. Sampling, or lack thereof, appears to have a very strong effect on what we observe in the fossil record and the patterns inferred therefrom.

4.4.3 Potential drivers of latitudinal richness through time

I used generalized least-squares (GLS) regressions in order to better understand the potential effects of sampling and land area on the uncorrected latitudinal richness of non-marine turtles and lepidosaurs. The results showed that sampling biases played the largest role in explaining the recovered patterns of richness in most intervals. This is consistent with the finding that sampling biases played a large role in shaping our knowledge of the global richness of both clades through time from the Triassic–Paleogene (see Chapters 2 and 3).

For lepidosaurs (Table 4.3), the largest influence of the three factors tested was the number of tetrapod-bearing formations (TBFs) in every time slice except the Early Paleogene. Values of R^2 vary between time slices, from low (Late Paleogene, 0.33) to medium (Late Cretaceous, 0.55), suggesting that while TBFs may affect latitudinal richness they do not explain a majority of the variation in richness and other factors must have been at play. AICc chose land area as the best model for explaining Early Paleogene richness, but the R^2 value is very low (0.26); as the null model is chosen as the second best model in this time slice it appears that none of the three factors tested here explain variation in Early Paleogene lepidosaur latitudinal richness satisfactorily. It is plausible that, as the interval containing the PETM and EECO, temperature had a large impact on lepidosaur distribution at this time, and possibly in other time slices also. The null model is recovered as the second best model in three of the five time slices (Jurassic and both Paleogene slices), which could be due to poor sampling, as *N* is low (particularly in the Jurassic), or the influence of other factors not tested in my analyses.

Turtles (Table 4.2) have either TBFs or tetrapod-bearing collections (TBCs) as the best explanatory model in every time slice except the Jurassic, where the null model was chosen as best. TBFs were chosen as the best model for the Early Cretaceous and Late Paleogene. AICc chose TBCs as the best model for the Late Cretaceous and Early Paleogene. R^2 values vary, from low (Late Paleogene, 0.39) to high (Early Paleogene, 0.75). The Jurassic has a very low *N* (5) and so the null being chosen is understandable; this could also explain the low R^2 for TBFs in the Late Paleogene. In the Early Cretaceous TBFs and TBCs are very close, but both have relatively low R^2 values (0.41), suggesting some other untested factor is influencing latitudinal richness at this time, perhaps temperature as suggested above, or rainfall/aquatic environment availability. This is likely as many modern turtles have distribution patterns driven by rainfall (Iverson, 1992), and prior to the Paleogene non-marine turtles were largely tied to water (Claude and Tong, 2004). Ecological niche modelling of Maastrichtian and modern turtle faunas found that temperature was important to turtles in general, but that precipitation was the most important factor to aquatic clades (Waterson et al., 2016), which likely applies to other time periods. Late Cretaceous TBCs have a high R^2 (0.68), and the second best model (collections plus land area) has an R^2 of 0.71. TBCs and TBFs have very close AICc values for the Early Paleogene, and high R^2 values (0.75), suggesting that sampling has a large influence on our understanding of the latitudinal richness of Late Cretaceous and Early Paleogene turtles.

My results agree with those of Nicholson et al. (2016) for the most part, although I replaced the number of turtle-bearing formations with tetrapod-bearing formations. In the Jurassic, both of our analyses recovered the null model as best. In the Early Cretaceous, despite the change in scope, I recovered formations as the best model. Naturally the R^2 value for TBFs is not as large as with turtle-bearing formations, given that this value is naturally more closely linked with richness. Both I and Nicholson et al. (2016) found that AICc chose TBCs as the best model for Late Cretaceous richness, though the high value of R^2 I found was not as high as in their study (0.95). This likely is a product of slightly different data, since taxonomic revision has occurred between when we both downloaded our datasets, and a consequence of different analytical procedures. I cannot find Nicholson et al.'s (2016) exact GLS procedure in their paper or supplementary information, and after adding back in zero-occurrence latitudinal bands and adding a start of 1 to all variables with zeroes (to account for using logs of data) I still cannot achieve an R^2 value as high as that reported in their results. Their study found that sampling accounted for the vast majority of the variation in both Cretaceous bins, which differs a little from my results here, where the picture is more complex.

Mannion et al. (2012) found that the number of dinosaur-bearing collections (DBCs) plus non-marine area explained a high amount of dinosaur latitudinal richness (in 10° bands rather than 15°; R^2= 0.7) in the Cretaceous. They used the whole Cretaceous rather than splitting it into two as I did here: consequently, it is more difficult to directly compare their results to mine. Nevertheless, they found that sampling played a relatively large part in the variation of latitudinal richness, consistent with my results for turtles and lepidosaurs. They also found a strong influence of DBCs on Jurassic richness (R^2= 0.85), but as my data are much more poorly sampled in the Jurassic it is again difficult to draw direct comparisons.

Land area was found to be more important to the dinosaur record than turtles or lepidosaurs, particularly for the Cretaceous. Land area plus TBCs was the second best model chosen to explain variance in Late Cretaceous turtle richness, and the best model for Early Paleogene lepidosaurs, but sample size was very small in the latter. It is likely that land area was genuinely important in the Cretaceous, as there was much less land area in the tropics then, particularly in the Northern Hemisphere tropics (e.g. Smith, Smith and Funnell, 2004). It is probable that land area is more important to dinosaurs because they are simply better sampled in the palaeotropics than lepidosaurs and turtles, so patterns of richness related to land area can be more fully explored.

Besides issues with low *N* values, both non-marine turtle and lepidosaur latitudinal richness appears to be heavily influenced by sampling biases, represented here by the number of tetrapod-bearing collections or formations. This influence varies between time slices, with sampling explaining only a small part of richness variation in some time slices (e.g. Early Cretaceous), and a large portion in others (Early Paleogene turtles). It is unlikely that the same factors are responsible for influencing richness in all time slices, since they encompass times of varying geographic configurations and resulting climate.

4.4.4 Conclusions

The latitudinal biodiversity gradient of non-marine turtles and lepidosaurs in the past was largely palaeotemperate in nature, at least in the Northern Hemisphere. The findings of high Northern Hemisphere palaeotemperate richness for the majority of the Triassic–Paleogene corresponds with the general finding of a palaeotemperate peak in richness for Mesozoic dinosaurs (and uncorrected brachiopod and plant richness). A palaeotropical peak in Northern Hemisphere Jurassic dinosaur richness was also found for non-marine turtles but not lepidosaurs. Despite supposedly similar ecologies as ectotherms, richness patterns here do not correspond to those of non-marine crocodylians, which appear to have had a palaeotropical richness peak in the Late Triassic, and a relatively equal richness between the palaeotemperate and palaeotropical latitudes in the Late Cretaceous. Unfortunately, both the turtle and lepidosaur fossil records suffer from a lack of coeval palaeotropical and Southern Hemisphere localities that render the tracking of richness between time slices and comparisons to other terrestrial vertebrates difficult. It is possible, therefore, that genuine palaeolatitudinal distributions of both clades were very different to what we observe in the fossil record with the data we currently have.

Areas in the modern day that are heavily sampled for fossils (USA and southern Canada, Western Europe, northern China) correspond to palaeotemperate areas in the studied time period, so sampling biases are at least partially linked to latitudinal variation in both clades from the Jurassic to the Late Paleogene. Lepidosaur latitudinal richness is explained best by the number of tetrapod-bearing formations, except in the Early Paleogene, but R^2 values are not high enough for TBFs to be the sole factor affecting richness. Latitudinal richness of turtles is best explained by TBFs or the number of tetrapod-bearing collections, except in the Jurassic where the null model was selected as best. R^2 values are higher for the Late Cretaceous and Early Paleogene, suggesting sampling proxies explain a large portion of richness in these two time slices. There is still a large portion of variation in richness that is not explained by sampling biases or land area, which could be attributed to palaeolatitudinal temperature gradients or, in the case of turtles, the availability of aquatic environments (linked to rainfall), but acquiring latitudinal data for these factors in the past is problematic.

It is difficult to comment on whether the LBG of past turtles and lepidosaurs differs substantially from the modern LBG, due to poor sampling at low latitudes. The finding that land area is not well-correlated with palaeolatitudinal richness could support the hypothesis that modern richness is not a product of increased land area in the tropics, but we need to gather more information to be sure, given that sampling issues are prevalent in our current data.

5. Preservation Biases and the Fossil Record of Lizards

5.1 Introduction

Our observations of past life are affected by a number of factors, including the differences in the inherent preservability of taxa. Some taxa have a higher chance of being recovered by researchers due to the presence of parts that are more likely to survive the fossilisation process (e.g. a calcified skeleton or large body size; Raup, 1972; Forey et al., 2004; Brown et al., 2013). This can affect estimates of richness through time, as taxa that do not preserve as well may be less likely to be identified down to lower taxonomic levels. This means they are not counted towards richness estimates, and a very diverse fauna might not be recognised due to scrappy preservation.

Among the many methods used to assess the quality of the fossil record are completeness metrics, which assess the preservation quality of individual specimens or taxa. Completeness metrics originated with observations of preservation biases in the fossil record (Fountaine et al., 2005; Smith, 2007; Benton, 2008a, 2008b), and spawned simple metrics that measured overall quality of specimens. For example, Benton (2008b) assessed the quality of dinosaur specimens based on a scale of 1 to 5, with 1 representing the most fragmentary isolated remains and 5 complete skeletons. He found that, until the mid-1970s, the majority of dinosaur type specimens consisted of very incomplete fossils (isolated teeth or bones). Most of these specimens were subsequently designated as *nomina dubia*.

Mannion and Upchurch (2010) improved upon these simple metrics, and created two more robust methods of measuring preservation quality using sauropodomorph dinosaurs as exemplars. Both of their metrics divide the skeleton into a number of areas, and then weight them with percentages based on how much information each individual portion contributes. For their Skeletal Completeness Metric, the percentages are drawn from how much of the physical skeleton (i.e. numbers and relative sizes of skeletal elements) is represented in each subdivision. By contrast, the Character Completeness Metric takes into account the average number of phylogenetic characters that can be scored from each skeletal region using alternative phylogenetic datasets and assigns percentages using these. This study provided metrics that were more objective than previous attempts to quantify preservation differences, and was the first to compare these more quantitative data with potential driving factors of completeness such as sea level and rock outcrop area.

Another study took a similar approach of dividing skeletons into regions and assessing preservation quality for extinct marine crocodiles, but used a non-weighted approach (Beardmore et al., 2012). The method is an intermediate between Benton (2008b) and Mannion and Upchurch's (2010) methodologies. Beardmore et al. (2012) rated each skeletal region on a scale of 0 to 4 based on the number of elements (bones) present, with scoring criteria specific to each region. For example, if 25–50% of dorsal vertebrae are present the score would be 2 for that region, whereas limb scores are based on the total number of elements present, e.g. a humerus plus radius is scored as 2. The authors then calculated percentage completeness from the total of all regions divided by the total possible score. The authors also integrated a measure of articulation similar to their skeletal completeness metric in order to examine the relationship between completeness, disarticulation and taphonomic processes.

Since then, many studies have used or adapted these three major approaches and applied them to other tetrapod clades (Brocklehurst et al., 2012; Brocklehurst and Fröbisch, 2014; Cleary et al., 2015; Dean, Mannion and Butler, 2016; Verrière, Brocklehurst and Fröbisch, 2016; Davies et al., 2017; Tutin and Butler, 2017; Driscoll et al., 2018), often comparing them to richness and variables representing various sampling issues or environmental factors through time (e.g. number of rock formations available per time period, sea level, temperature proxies) with varying results between taxa.

Here I will assess the completeness of non-marine lepidosaurs from the Triassic–Paleogene using an adapted completeness metric. Due to the difficulty of applying completeness metrics to legless taxa with unclear trunk and tail regions and variable limb numbers, I excluded snakes and amphisbaenians from this study. Hereafter this paraphyletic group will be referred to as "lizards" in order to distinguish it from the whole lepidosaur clade used in previous chapters. Within the remaining lizard group, there are still some issues with disparate shapes and proportions of skeletal elements between taxa, and so I decided to adapt the non-weighted skeletal completeness metric used by Beardmore et al. (2012) to examine lizard specimens. The aim of this study is to investigate the preservation quality of lizards and compare the results to uncorrected and subsampled taxonomic richness obtained in Chapter 2, as well as to environmental factors and sampling proxies, to uncover any significant relationships that may exist.

5.2 Methods

5.2.1 Data and exclusions

I used the skeletal completeness metric created by Beardmore et al. (2012) for marine crocodilians (referred to hereafter as the BSCM) and adapted it to assess the completeness of lizards from the Triassic—Paleogene (252-23 Ma). I compiled a matrix of 414 species of non-marine lizards, representing the majority of all currently valid lepidosaur species from this time interval excluding *nomina dubia*, limbless taxa (snakes and amphisbaenians) and five species for which I could not find sufficient information (see Appendix 9). Taxa preserved in amber and those preserved only as osteoderms were also excluded for difficulty of assessment. Completeness data was collected for the most complete specimen of each species included in the dataset, following the approach of Mannion and Upchurch's (2010) Skeletal Completeness Metric 1. If no specimen was more complete than the holotype in each species, then the holotype was chosen by default. Information on completeness was taken from the literature, using both detailed written descriptions and figures in tandem when available. I collected additional data on each specimen including stratigraphic age, locality information and geological setting, using both the literature (references are included with completeness data in Appendix 9) and the Paleobiology Database (www.paleobiodb.org).

5.2.2 Completeness metric

I divided the lepidosaur skeleton into eight regions: head, dorsal vertebrae, ribs, left and right forelimb, left and right hindlimb, and caudal vertebrae. Limb girdles were excluded as they are often obscured in compressed specimens and difficult to assess. The "dorsal vertebrae" category includes the cervical and sacral vertebrae. Specimens are scored from 0 (absent) to 4 (complete) in each category based on the number of elements preserved (see Figure 5.1 for regions diagram and Table 5.1 for metric criteria). In the case of limb bone regions, elements consist of four subunits comprising the humerus, radius-ulna, metacarpals and phalanges or femur, tibia-fibula, metatarsals and phalanges. Elements are generally marked as complete if present, unless half or more is missing; then they are given a score of 0.5. A limb region can therefore have a score of 2 but be comprised of partial elements of all four component elements. The head is also further sub-weighted to account for the prevalence of taxa represented by a single skull or mandible bone. Single cranial bone taxa are scored at 0.5, while specimens comprising only a single tooth are scored as 0.1. The scores for all eight regions are summed and divided by the total possible score (32) to get the BSCM score, which I then converted into a percentage. Individual scores were compiled into the same time bins used in previous chapters (Table 2.1) to examine average completeness through time.

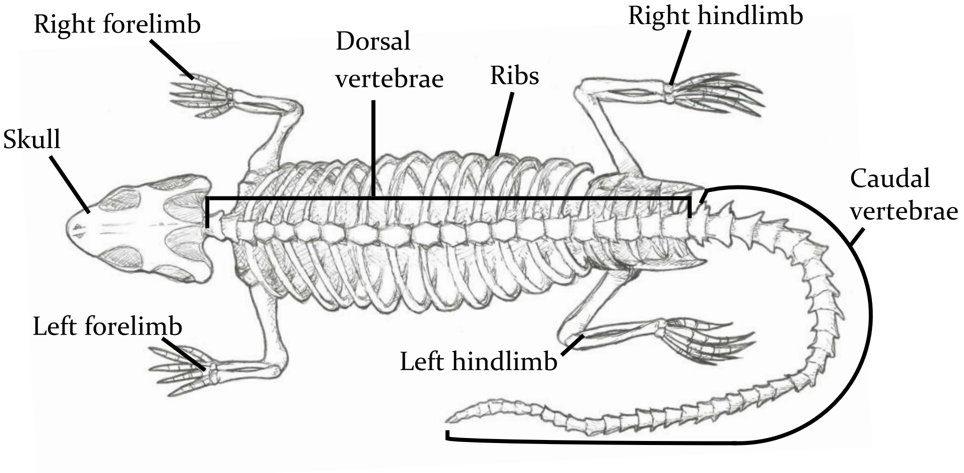

Figure 5.1: Division of regions for the skeletal completeness metric. Regions are scored up to a maximum of 4 based on the proportion of elements present.

Table 5.1: Completeness metric used for lepidosaurs in this chapter, adapted from Beardmore et al (2012). Percentage completeness is obtained from the sum of all categories divided by 32, multiplied by 100. "Elements" = cranial bones, or subunits of limbs. Limb subunits are humerus, radius-ulna, metacarpals and phalanges OR femur, tibia-fibula, metatarsals and phalanges. Limb elements are generally marked as complete if present, unless half or more is missing; then, 0.5 scores are used.

BSCM	**0**	**0.1**	**0.5**	**1**	**2**	**3**	**4**
Skull	Absent	Single tooth	1 element or “partial mandible”	2-3 elements remain; limited pres.	1/2 elements	1-3 elements missing; skull shape recognisable	All elements observed
Dorsal verts	0%	NA	NA	1-25%	25-50%	50-75%	75-100%
Ribs	0%	NA	NA	1-25%	25-50%	50-75%	75-100%
Left forelimb	Absent	NA	Partial element	1/4 elements present	2/4 elements	3/4 elements	All present
Right forelimb	Absent	NA	Partial element	1/4 elements present	2/4 elements	3/4 elements	All present
Left hindlimb	Absent	NA	Partial element	1/4 elements remain	2/4 elements	3/4 elements	All present
Right hindlimb	Absent	NA	Partial element	1/4 elements remain	2/4 elements	3/4 elements	All present
Caudal verts	0%	NA	NA	1-25%	25-50%	50-75%	75-100%

Additionally, I compiled summaries of the number of specimens comprised of only certain elements, e.g. the number of specimens consisting of only mandibular bones versus other combinations, to highlight those elements that are most commonly found (or that are considered to be the most diagnostic). I tested whether the proportion of these mandible-only specimens was significantly different in the Mesozoic compared to the Paleogene using a 2-sample z-test. I also examined the proportion of single-bone specimens and the proportion of specimens comprising both cranial and postcranial elements (partial or full skeletons) in the two time periods with the same test.

Using the same data and methods as presented in Chapter 2, I ran Shareholder Quorum Subsampling (SQS) on a lizard dataset (i.e. original occurrence data minus snakes and amphisbaenians; see R code in Appendix 9) at species level in order to obtain a comparable sampling-corrected curve to compare to completeness through time.

5.2.3 Comparisons to sampling biases and environmental variables

I used generalized least-squares regression (GLS) to examine relationships between average species completeness per bin and a number of variables. These included the number of lizard species per bin (non-subsampled richness), the number of terrestrial tetrapod-bearing collections (TBCs) and formations (TBFs), sea level, non-marine area, and palaeotemperature. Sea level and palaeotemperature data are only available from the mid-Jurassic onwards, so I only analysed bins J4 to Pg5 (Table 2.1). The sampling proxies (TBCs and TBFs) were obtained from the same PBDB file of terrestrial tetrapods as in previous chapters. Sea level data is from Miller et al. (2005) and non-marine area is taken from the palaeocoastline reconstructions of Smith et al. (2004). Palaeotemperature is represented by a $\delta^{18}O$ proxy from Prokoph et al. (2008), compiled by Mannion et al. (2015) which I interpolated as described in previous chapters in order to aggregate the data into within-bin means. The means of all proxies were calculated for each bin. Completeness, the number of species, and the sampling proxies were ln-transformed prior to analysis.

I fitted autoregressive (AR) models of order one to all combinations of the above variables (switching out TBFs and TBCs as they both represent a sampling proxy). AR models check for serial correlation and reduce the chances of overestimating the significance of regression lines (Chatfield, 2003). Akaike's information criterion for small sample sizes (AICc) was used to identify the best model out of all the models tested. AICc preferentially chooses models for goodness of fit but penalises those models with higher numbers of variables (Burnham and Anderson, 2002). I manually calculated R^2 values from GLS outputs (Nagelkerke, 1991), and checked normality and homoskedasticity of residuals using Jarque-Bera and Breusch-Pagan tests (Breusch and Pagan, 1979; Jarque and Bera, 1980). All analyses were carried out in R v. 3.5.1, using the packages lmtest v. 0.9-36 (Hothorn et al., 2017), nlme v. 3.1-137 (Pinheiro et al., 2017), qpcR v. 1.4-1 (Spiess and Ritz, 2014) and tseries v. 0.10-45 (Trapletti and Hornik, 2017) following the approach of Benson and Mannion (2012).

5.3 Results

5.3.1 Completeness through time

Figure 5.2 shows the individual completeness of all specimens, the average completeness per time bin (of those specimens that fit into a single time bin) and the subsampled richness of lizard species using SQS at quorum 0.4. Figure 5.3 shows the same but subsampled richness is replaced with the uncorrected species diversity curve. At face-value they reveal the incongruence between completeness levels and both raw and subsampled richness through time.

Average completeness remains relatively low throughout the entire studied period, but with some instances of exceptional preservation for a small number of taxa. These instances of exceptional preservation can pull the average completeness of a bin that has mostly poorly complete specimens to an unrepresentative medium level of completeness. These instances are particularly prominent in the Late Jurassic (J5-J6: Callovian-Oxfordian, Kimmeridgian-Tithonian), Early Cretaceous (K2-4: Hauterivian-Barremian, Aptian, Albian) and middle Eocene (Pg3: Lutetian). These times of higher average completeness do not visually correspond to peaks in recovered subsampled or uncorrected richness of species, otherwise we might expect that completeness should be very high in the Campanian (K7) and in bins Pg2 (Ypresian) and Pg4–5 (Priabonian–Bartonian, Oligocene). In bins Pg2-4 of the uncorrected species richness curve (Fig. 5.3) the two curves show opposite trends, with one falling when the other is rising and vice versa. There are many bins missing in the subsampled richness curve due to poor sampling of the lepidosaur record, however, so comparisons between this curve and completeness are difficult outside of the Late Cretaceous. The lack of data also prevents any quantitative comparisons of subsampled richness to completeness.

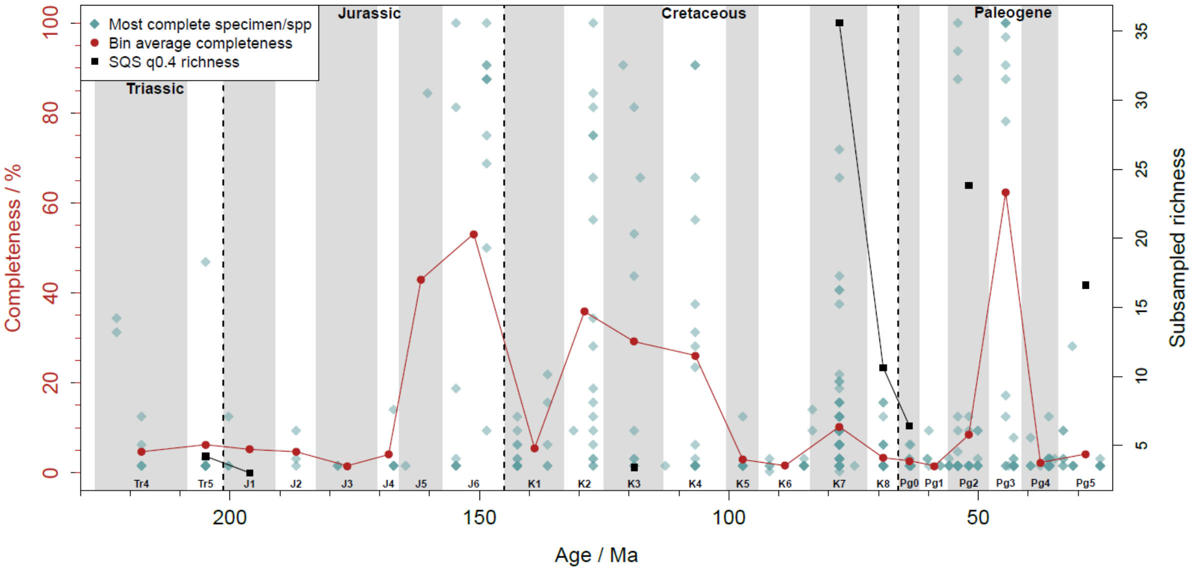

Figure 5.2: Skeletal completeness from Triassic–Paleogene of the most complete specimen of all lizard species (light blue) plotted against the average completeness of lizards in each time bin (red) and subsampled (SQS) lizard species richness at quorum 0.4. T, Triassic; J, Jurassic; K, Cretaceous; Pg, Paleogene.

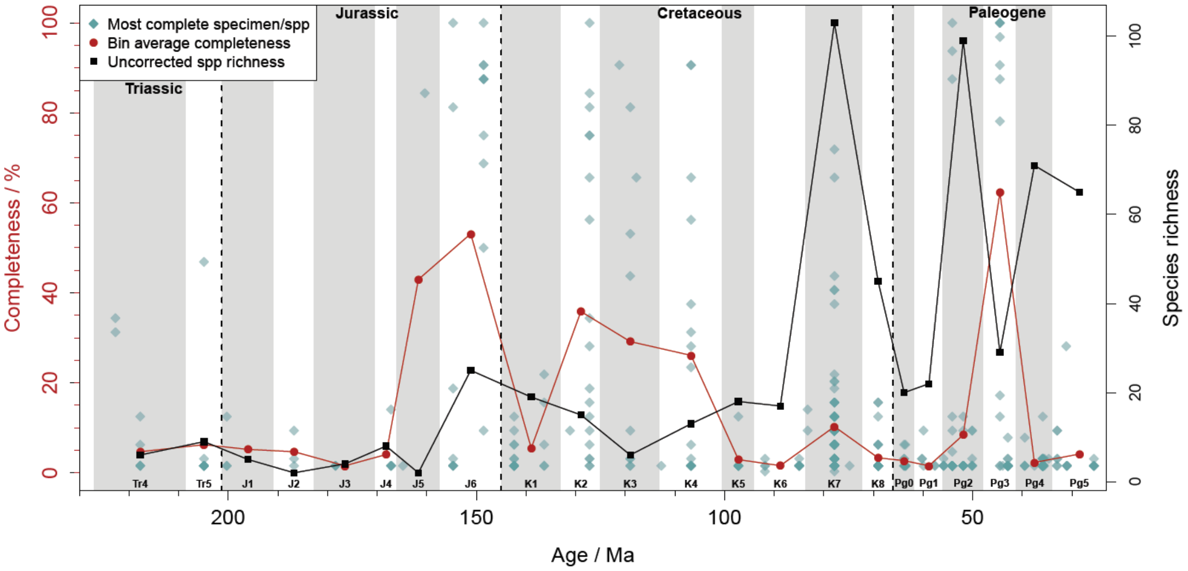

Figure 5.3: Skeletal completeness from Triassic–Paleogene of the most complete specimen of all lizard species (light blue) plotted against the average completeness of lizards in each time bin (red) and uncorrected (face-value) lizard species richness. T, Triassic; J, Jurassic; K, Cretaceous; Pg, Paleogene.

5.3.2 Dataset composition

The number of species where the best preserved specimen is comprised of a single bone versus multiple bones is roughly equal (49.8% versus 50.2%, respectively). Of the total species in my dataset, 74.6% of the best preserved specimens are represented by only cranial bones. Figure 5.4 shows the breakdown of four categories of preservation, where the best preserved specimen of each species is represented by only mandibular bones, only cranial bones (including combinations of mandibular and other cranial elements), only postcranial bones (e.g. vertebrae or limbs), or cranial and postcranial elements (e.g. partial or full skeletons). Of the 74.6% of specimens composed of only cranial elements, more than half have only mandibular elements (41.6% versus 33% for other cranial element combinations). Many of the 49.8% of specimens represented by single bones are known solely from dentaries.

If I partition the data into the Mesozoic and Paleogene, there are significantly more instances of single-bone taxa in the latter (65.8% of 152 Paleogene species versus 40.4% of 260 in the Mesozoic; χ-squared = 23.758, $p < 0.0001$). More than half of all Paleogene taxa have a best preserved specimen consisting of mandibular elements only (significantly less than the 36.2% in the Mesozoic: Fig. 5.5; χ-squared = 7.725, $p = 0.003$), and there are also fewer partial or full skeletons than in the Mesozoic ("cranial and postcranial elements": Paleogene, 14.5%; Mesozoic, 30%; χ-squared = 11.75, $p = 0.0003$).

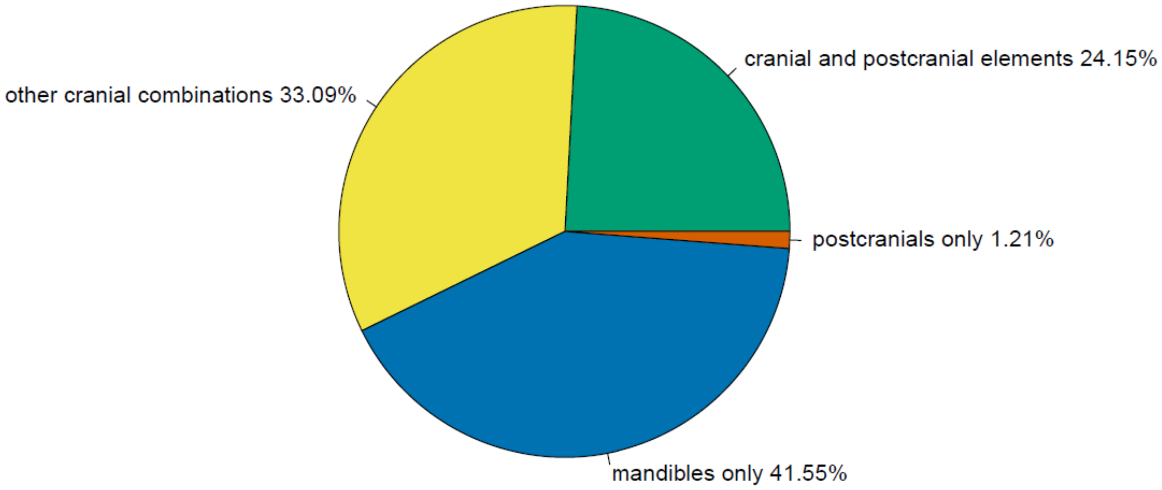

Figure 5.4: Pie chart showing the division of categories for the most complete specimen of all lizard species. "Other cranial combinations" denotes specimens comprised of non-mandibular cranial bones including or excluding mandibular elements.

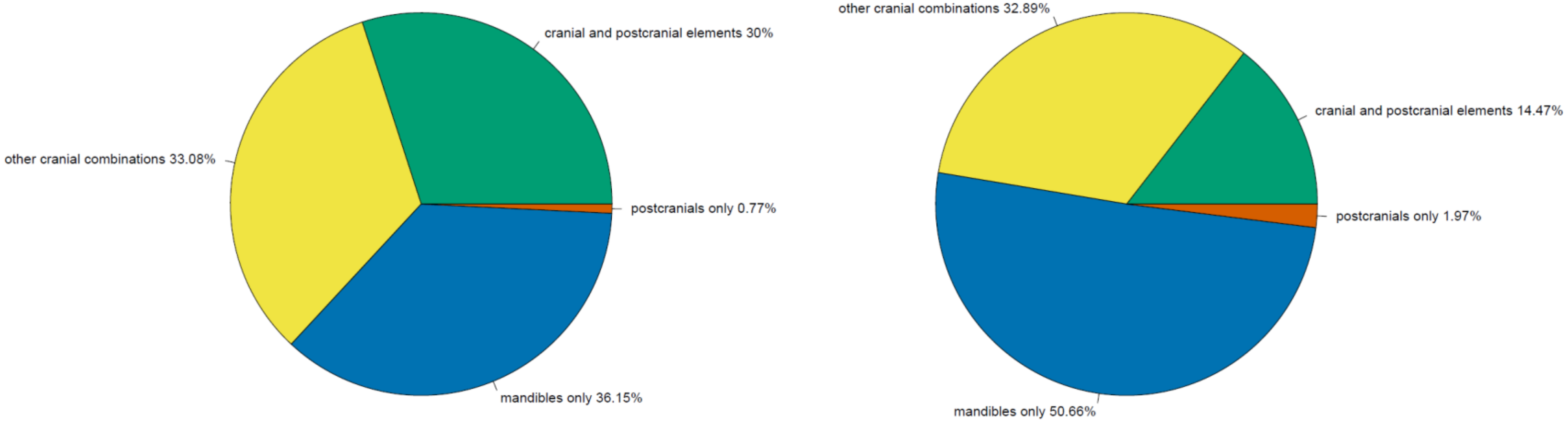

Figure 5.5: Pie chart showing the division of categories for the most complete specimen of all lizard species in the (*a*) Mesozoic and (*b*) Paleogene. “Other cranial combinations” denotes specimens comprised of non-mandibular cranial bones including or excluding mandibular elements.

5.3.3 Generalized least-squares analyses

I tested many combinations of variables including the number of species, sampling proxies (tetrapod-bearing collections and formations), sea level, non-marine area and a palaeotemperature proxy ($\delta^{18}O$) against average completeness. AICc chose the null model as the best model (Table 5.2; AICc = 61.674, weight = 0.274), meaning that no single variable or combination of variables tested is significantly related to completeness. The null model was also chosen as best when data was split into the Mesozoic and Paleogene (see Appendix 9: Mesozoic, AICc= 41.054, weight= 0.242; Paleogene, AICc= 25.847, weight= 0.451). In terms of explaining completeness, none of the variables tested appear to be important.

Table 5.2: Summary of the best 20 model fits using GLS to per-bin completeness of lizards from the Triassic–Paleogene, ordered by best to worst model as chosen by AICc. An AR model of order one was applied to each combination of variables; N= 17. Model abbreviations are R, species richness; C, tetrapod-bearing collections; F, tetrapod-bearing collections; S, sea level; N, non-marine area; O, δ^{18}O temperature proxy. Full results including all models and values for coefficients of all components within each model (which did not fit on one page) and R script can be found in Appendix 9.

Model name	**R^2**	**LL**	**AICc**	**AICc weight**
Null	**0.000**	**-27.75**	**61.76**	**2.781E-01**
F	**0.084**	**-27.00**	**62.86**	**1.605E-01**
R	**0.032**	**-27.47**	**63.80**	**1.004E-01**
RC	0.183	-26.03	63.91	9.494E-02
O	0.012	-27.65	64.15	8.443E-02
C	0.011	-27.66	64.17	8.355E-02
RF	0.125	-26.61	65.07	5.330E-02
FO	0.094	-26.91	65.66	3.968E-02
RO	0.042	-27.39	66.62	2.450E-02
CO	0.027	-27.51	66.87	2.160E-02
RCO	0.192	-25.93	67.20	1.835E-02
RFO	0.138	-26.49	68.31	1.051E-02
N	-0.330	-30.18	69.21	6.724E-03
FN	-0.213	-29.39	70.62	3.317E-03
S	-0.446	-30.88	70.62	3.315E-03
NO	-0.280	-29.84	71.53	2.101E-03
RN	-0.295	-29.95	71.74	1.897E-03
CN	-0.305	-30.01	71.87	1.775E-03
FS	-0.324	-30.13	72.11	1.574E-03
RCN	-0.084	-28.43	72.19	1.511E-03

5.4 Discussion

5.4.1 Explanations for patterns of lizard completeness through time

As discussed above, Figures 5.2 and 5.3 show relatively low completeness through time for lizards, punctuated by bins of mid-level completeness. These times are characterised by a number of very highly or totally complete specimens that raise the bin average due to an increased range in completeness levels. The primary reason for these times of higher completeness are that the bins in question are occupied by localities of exceptional preservation: Lagerstätten.

In bin J6 (Kimmeridgian–Tithonian), the localities of Solnhofen and Painten in Germany contain lithographic limestones that preserve fully articulated and exceptionally preserved skeletons, often with associated soft tissues (e.g. Bennett, 2002; Klug et al., 2015). Included in bin K2 (Hauterivian–Barremian) are the localities of the fine-grained volcanic Yixian Formation in China (Zhou, 2014), and the lithographic limestones of La Huérguina Formation in Las Hoyas and Uña, Spain (Bailleul et al., 2011). In K3 (Aptian) there is more of the Yixian in China and the Plattenkalk of the Crato Formation in Brazil (Fielding, Martill and Naish, 2005), and in K4 (Albian) there are the fine-grained limestones of the La Cavere Formation in Pietraroja, Italy (Dal Sasso and Signore, 1998), and of Tlayua Quarry in Mexico (Alvarado-Ortega et al., 2007). Finally, in Pg3 (Lutetian) there are more localities of exceptional preservation in Germany, with the finely laminated shales of Messel and the lignites of Geiseltal (Smith and Wuttke, 2012). Bin J5 (Callovian-Oxfordian), whilst also featuring a higher level of completeness, contains only two specimens with the average sitting between them, and so I have chosen to disregard this bin when discussing the results.

K7 (Campanian) and Pg2 (Ypresian) also have specimens with higher levels of completeness. In the Campanian there are several localities in the Gobi Desert of China and Mongolia that are highly productive for finding lizard fossils (e.g. Borsuk-Bialynicka, 1984; Alifanov, 2000; Gao and Norell, 2000). The facies here are less likely to produce complete specimens, however, so there are only a few exceptional specimens amongst the many found. In the Ypresian there is a single Lagerstätte in my data, Tynsky Quarry in Wyoming, USA, which exposes the Green River Formation, again a limestone which produces exceptionally preserved specimens (Hellawell and Orr, 2012). There are many more specimens of lower completeness in both K7 and Pg2, however, which keep the bin averages low. Other studies that examined Lagerstätten found their effects influenced the completeness of ichthyosaurs (Cleary et al., 2015), mosasaurs (Driscoll et al., 2018) and pterosaurs (Dean, Mannion and Butler, 2016). By contrast, Verrière, Brocklehurst and Fröbisch (2016) did not find any effect of Lagerstätten on the parareptile record.

One hypothesis regarding specimen preservation quality is that a greater average completeness (and thus higher quality preservation) in a bin should drive the identification and naming of more taxa, since the remains found might be more diagnostic (Mannion and Upchurch, 2010). These results demonstrate that this is not the case with non-marine lizards, however. From both Fig. 5.2 and 5.3 and the generalized least-squares regression results (see below) it is clear that there is no relationship between the completeness of specimens and the number of species found in any particular time bin. Although it is hard to examine completeness and sampling-corrected richness because of very poor sampling in the lepidosaur record, the bins that are recovered by SQS are clearly incongruent with average completeness.

Other studies using completeness metrics found no correlation between these and taxon richness through time for parareptiles (Verrière, Brocklehurst and Fröbisch, 2016), plesiosaurs (Tutin and Butler, 2017), ichthyosaurs (Cleary et al., 2015), mosasaurs (Driscoll et al., 2018), eutherian mammals (Davies et al., 2017), and early tetrapods and amphibians (Benton et al., 2013). Mannion and Upchurch (2010) found a correlation only for the Early and Late Cretaceous (negative and positive correlations respectively) in sauropodomorph dinosaurs. A negative correlation was found for synapsids (Brocklehurst and Fröbisch, 2014), and a positive correlation for early amniotes (Benton et al., 2013), Mesozoic birds (Brocklehurst et al., 2012) and pterosaurs (Dean, Mannion and Butler, 2016). This mixture of results indicates that higher completeness is not a simple proxy for more reliable taxonomic identification per se.

In the case of lizards, the lack of correlation between completeness and both subsampled and uncorrected species richness through time can be explained when examining the makeup of the dataset. As stated in the results, out of the best-preserved specimens of every species represented in these data, 75% feature only cranial elements (Fig. 5.4), and within this more than half are only comprised of mandibular bones. From the total of 414 specimens, roughly half were represented by single bones. It is clear, however, that the prevalence of low completeness does not adversely affect the recognition of new taxa, as the parts commonly preserved appear to be highly diagnostic. Indeed, many of the specimens contained in the dataset are holotypes that consist only of single dentaries or maxillae. This is particularly true in the Paleogene (Fig. 5.5), which has a much higher instance of single-bone and mandible-only taxa than the Mesozoic, and yet richness remains very high in some Paleogene bins. Additionally, completeness is sometimes lower in species because of the loss of articulation suffered either taphonomically or during collection. Lizards, which are usually very small in size, are often found via screen-washing of sediments (e.g. Blain et al., 2010). This results in a scattering of small disarticulated bones that may or may not belong to one individual but which are all assigned different specimen numbers due to uncertainty. This means that there are a large number of different diagnostic elements associated with a species but that individual specimen completeness is low.

Essentially, even though the skeletal completeness of the lepidosaur record is poor it does not appear to have had a deleterious effect on the naming of new genera and species. Only the most diagnostic elements are needed, e.g. dentaries and maxillae, and these elements are very recognisable and so are found even in places with poorly preserved and/or highly disarticulated skeletons. An example of this can be found in the work of Augé (2005), who examined the lizard fossil collections of the European Paleogene, a good portion of which consist of highly disarticulated individual bones yet produce a rich array of taxa. The intensive sampling here could also explain why the Paleogene has a significantly higher proportion of taxa comprising single bones or mandibular elements only compared to the Mesozoic.

Alternatively, poor lizard completeness could be due to scientists being overeager and naming many species based on fossil material despite low preservation quality, as occurred with dinosaurs in the past (Benton, 2008a, 2008b). Regardless of which explanation is true, this is why completeness does not correlate with richness either visually (Fig. 5.2, 5.3) or in the GLS analyses (Table 5.2; below).

5.4.2 Relationships between completeness and other variables

As well as the non-relationship between species richness and completeness, the results of my GLS analyses reveal a lack of relationships between completeness and other variables commonly considered to represent major sampling and environmental factors. This lack of relationship persists even when the dataset is split into Mesozoic and Paleogene segments (although the number of Paleogene bins is too small to run the analyses fully; see Appendix 9).

Other studies have also used GLS or similar techniques to assess a variety of relationships between completeness and other variables. In several instances, as in this study, no significant correlations were found between completeness and sea level, collections, formations, or environmental variables (Brocklehurst et al., 2012; Verrière, Brocklehurst and Fröbisch, 2016; Davies et al., 2017; Tutin and Butler, 2017). However, completeness was found to be negatively related to sea level for sauropodomorphs (Mannion and Upchurch, 2010), significantly related to the number of collections (positive trend), formations and sea level (negative trends) for ichthyosaurs (Cleary et al., 2015), and AIC scores favoured strontium isotope ratios (used as a sea level proxy) as the best explanation for variance in eutherian mammal completeness (Davies et al., 2017). Dinosaur-bearing formations plus Lagerstätten were recovered as the best model explaining pterosaur completeness (Dean, Mannion and Butler, 2016).

The differences between taxa with regards to how skeletal completeness correlates (or does not) with potential explanatory variables are probably due to differences in the preservational probability of each taxon, the time periods (and thus facies) that each taxon ranges through, and the environments that each exists in (e.g. ocean versus fully terrestrial habitats). For lizards, it appears that none of the explored factors were as important for explaining their completeness as the presence of Lagerstätten, which are more likely to preserve microvertebrates in good condition. It would be pertinent therefore in the future to try to add some method of accounting for the presence and absence of Lagerstätten in GLS analyses to see if this affects the fit of models.

5.4.3 Completeness metrics and their applicability to the lepidosaur record

Lepidosaur body plans are disparate, which presents complications when applying a standard completeness metric to every species. Attempting to code for completeness of features that are present in some taxa but absent in others is difficult, which is why I chose to remove legless taxa and examine only "lizards". After the removal of these taxa, applying some completeness metrics still presents issues, particularly with those metrics that require more specific assessments of specimens (e.g. the Skeletal Completeness Metric of Mannion and Upchurch (2010)). For example, there is a lot of variation in body elongation, and so if only the anterior of a specimen is known it is difficult to assess the percentage of the vertebral column that is present. The application of categorical completeness alleviates some of these problems, which is why I applied the completeness metric created by Beardmore et al. (2012) to lizards.

Other completeness metrics should be applied in the future to investigate how results compare to my findings here. Given more time, I would like to apply other methods created by Mannion and Upchurch (2010) to the lizard fossil record. A composite metric comprising the total known completeness of each species based on multiple specimens would be useful to compare against my "best-preserved specimen" scores. In particular I would like to use the "Character Completeness Metric" created by Mannion and Upchurch (2010), which used phylogenetic characters and assessed how many of these characters could be coded from each region to measure preservation. The analyses might be hampered by a lack of comprehensive phylogenies of extinct lepidosaurs, or by the large lepidosaur phylogenetic tree and thus abundance of disparate characters, but it could provide an interesting comparison to the physical preservation metric I used here.

5.4.4 Conclusions and future work

The skeletal completeness of lizard fossils through time does not adhere to a predictable pattern based on the variables I tested here. Certain time bins are affected by the existence of localities with exceptional preservation, which serve to raise the average completeness levels, but this inflation of completeness does not increase the identification of new species. This is partially due to the fact that a large portion of new taxa are named from fragmentary parts; whether these parts are extremely diagnostic or whether there is an overabundance of naming taxa based on insufficient information is unclear.

Despite the usefulness of completeness metrics with other taxa, here I found no relationship between the completeness of individual specimens and potential explanatory variables. The completeness of individual lepidosaurs does not appear to have an effect on their long-term trends of richness through time, such as those demonstrated in Chapter 2.

Future work using this method with lizards should compile a composite metric to include more specimens and increase data quality, and investigate phylogenetic completeness as demonstrated by Mannion and Upchurch (2010). This may alleviate the problems with many lizard taxa being based on disarticulated specimens, and could change the current lack of significant relationships between completeness and the variables tested in my regression analyses. It would also be interesting to further examine preservation and completeness differences between

Mesozoic and Paleogene lizards, to better assess the differences I found in this study and discover if they persist after the addition of more data.

6. Discussion

6.1 Richness of ectotherms through time

I constructed richness curves for raw, face-value genus-level observation counts and sampling-corrected counts using subsampling methods for non-marine turtles and lepidosaurs from the Triassic–Paleogene. After constructing these richness curves, the first obvious difference between the uncorrected and corrected curves is in the number of bins that are missing results due to poor sampling. This is more prevalent in the lepidosaur record than in the turtle record. The large differences that were present between richness peaks and troughs in the uncorrected data are reduced in the sampling-corrected data, but it is still clear that even when the data are interpreted as conservatively as possible there are times of fluctuating high and low richness between the Triassic and Paleogene. Unfortunately, because of the poorly sampled nature of both records, it is difficult to compare the subsampled curves of the two clades before the Late Cretaceous. Both clades appear to have had relatively low levels of richness before this time, punctuated by a few instances of times with exceptional quality or quantity of preservation (Lagerstätten) that raised richness levels. Examples of this include the K6 (Kimmeridgian–Tithonian) bin in the Jurassic, which includes deposits like the Morrison Formation (USA), Solnhofen (Germany), Cerin (France) and Guimarota (Portugal), the K2 (Hauterivian–Barremian) bin for Europe (which includes Las Hoyas and Uña in Spain and the Wealden Group in the UK), and the Jehol Group in China.

During the latest Cretaceous (Campanian–Maastrichtian, K7–8) there is high generic richness for both clades in North America compared to that seen in other continents. At this time the climate was much warmer than today, which likely allowed for increased richness and geographic ranges of ectothermic taxa, particularly those that are thought to exhibit stricter environmental requirements (e.g. pleurodire turtles). Richness is high for both clades between 30 and 60°N during the Campanian and Maastrichtian, and highest at 45–60°N in North America (Europe is much more poorly sampled at this time and contributes little useful information, particularly for lepidosaurs).

Over the Cretaceous–Paleogene (K-Pg) boundary, there is a marked difference in how non-marine ectotherm clades respond to this mass extinction event. Non-marine lepidosaur richness decreased over the K-Pg boundary, following the loss of the herbivorous polyglyphanodontid lizards, but otherwise the majority of higher taxa remained unscathed. It has been proposed, however, that polyglyphanodontids were already in decline before the extinction event (MacLeod et al., 1997). Non-marine turtles lost a number of clades, a large portion of which were either semi- or fully terrestrial (e.g. Nanhsiungchelyidae), but richness was much higher in the lowest Paleocene, suggesting that turtles were relatively unaffected by the K-Pg extinction. Mannion et al. (2015) also found a rise or plateau in richness for subsampled non-marine crocodylian data across this boundary. These results support the hypothesis that taxa with primarily aquatic ecologies had much higher survival rates across the K–Pg boundary than purely terrestrial taxa (Sheehan and Hansen, 1986). It was proposed that, after the bolide impact, dust would have accumulated in the atmosphere, reducing sunlight; this would have had a detrimental effect on primary productivity and all trophic levels reliant on it (e.g. Alvarez et al., 1980). This could explain the loss of the herbivorous polyglyphanodontids and of terrestrial turtles. Sheehan and Hansen (1986) hypothesised that many organisms in freshwater ecosystems were less reliant on primary productivity, as modern freshwater analogues have detritus-based food chains, and that this acted as a buffer during this mass extinction event.

During the Paleogene, non-marine lepidosaurs and turtles appear to have had broadly similar patterns of richness, with high richness in the Ypresian (Pg2) coincident with two short-lived hyperthermals, the Paleocene-Eocene Thermal Maximum and Early Eocene Climatic Optimum. At this time high global temperatures likely allowed the evolution of large body sizes (e.g. giant boid snakes, Head et al., 2009; large turtles; Holroyd, Hutchison and Strait, 2001) and geographic range expansions (e.g. turtles, crocodylians and squamates were present above 75°N; Estes and Hutchison, 1980). Turtles and lepidosaurs had very high richness through the Paleocene and Ypresian at 45–60°N, corresponding to richness levels in the Campanian and Maastrichtian. Both clades also had lower (but still high) richness at 30–45°N in North America and for European Ypresian turtles; the Paleogene of Europe (and the Ypresian for European lepidosaurs) is too poorly sampled to examine richness in adjacent latitudinal bins. Crocodylians were still very diverse in the palaeotemperate latitudes at this time compared to their modern day distributions, which are restricted to the tropics (Mannion et al., 2015).

Both clades underwent a decrease in richness during the Lutetian (Pg3) and then a return to the richness levels of the Ypresian or higher during the end of the Eocene (Pg4), particularly in Europe. Both clades decreased in richness over the Eocene–Oligocene boundary but richness remained high compared with the lows recovered in previous time bins (except for North American turtles, which matched Pg3 richness levels). Temperatures through the later Paleogene were declining (Zachos et al., 2001) so it would be expected that ectotherms decreased in range and abundance through the rest of the Eocene, but this is not the case for either of these clades. It is possible that taxa were able to adapt to the gradual cooler temperatures and increasing aridity in the late Eocene and so richness recovered, but that some taxa could not acclimatise for the sharp temperature drop across the Eocene–Oligocene boundary or were outcompeted by immigrating taxa, as has been proposed to occur in Europe during the Grande Coupure event (but see section 6.2 below for further discussion). It is also possible that both non-marine turtles and lepidosaurs have shared sampling issues that create similar patterns in richness during the Paleogene.

Through the latter half of the Eocene and the Oligocene, turtles had higher richness at 30–45°N than at 45–60°N, the opposite of the pattern in the Early Paleogene. In the Oligocene turtles appear to have had a lower richness compared to the Ypresian, with a lower richness in North America than Europe. Lepidosaurs are too poorly sampled in the late Eocene to reveal much, but maintained relatively high richness at 30–45°N. By the Oligocene North American turtle richness was approaching the pattern seen during the Holocene, with higher (but medium compared to other Paleogene bins) richness between 30–45°N than at 45–60°N. Europe, however, maintained high turtle richness between 30–60°N with slightly higher richness at 45–60°, contrary to the comparatively low turtle richness in this region during the Holocene (Roll et al., 2017). Lepidosaurs followed a similar pattern to turtles in the North American Oligocene, except that lepidosaur richness remained high at 30–45°N; sampling is too poor at 45–60°N in Europe to recover a pattern, but richness is high at 30–45°N. However, global lepidosaur richness at 45–60°N is still very high. This difference in latitudinal richness distribution in North America compared to the earlier Paleogene coincides with declining temperatures towards the Neogene, and could show the migration of both clades to more southerly latitudes as a response, as has been hypothesised for both turtles and lepidosaurs previously

(Hutchison, 1992; Smith, 2011). The European turtle and global lepidosaur pattern of higher richness at 45–60°N than 30–45°N could correspond to the fact that the more heavily sampled localities of the time (mid- to northern France, southern UK, Germany) lie within the former, but more evidence is required. There could have also been climatic differences between the large continental landmass of North America and the smaller more oceanic-influenced landmasses of Europe, such as those that exist today (Kottek et al., 2006).

Crocodylians, by comparison, declined in richness and eventually disappeared from palaeotemperate latitudes altogether (Mannion et al., 2015). One might expect, given the similarities in the richness patterns of aquatic turtles and crocodylians over the K–Pg boundary, that these two clades would follow similar richness patterns during the later Paleogene, but this is not the case. In the late Eocene, crocodylians had a higher richness in Pg3 at a time when lepidosaurs and turtles had low richness; conversely, they exhibited a large drop in richness towards Pg4 when the other two clades increased in richness. The time bins used in this study are the same as those employed in Mannion et al.'s (2015) analyses, which eliminates bin duration differences as a possible explanation for the observed patterns. One possibility is that crocodylians might have been less able to adapt to changing climate than the other two ectotherm groups, though additional work is needed to test this hypothesis.

The majority of the analyses that I ran on both the overall and latitudinal generic richness of lepidosaurs and turtles from the Triassic–Paleogene identified proxies representing sampling (numbers of tetrapod-bearing collections (TBCs) or formations (TBFs); mostly the former) as the best models explaining face-value richness patterns. However, except for medium-to-high R^2 values of the best models for overall lepidosaur richness (TBCs plus $\delta^{18}O$; 0.72), and for latitudinal distribution in the Late Cretaceous (TBCs; 0.68) and Early Paleogene (TBCs; 0.75) for turtles, the explanatory power of sampling proxies was relatively low. This suggests that other factors not tested by my analyses likely contributed to the richness of lepidosaurs and turtles through time. Modern richness patterns are driven by the interaction of numerous variables (e.g. temperature, rainfall, altitude), so it should be expected that past richness is reliant on the interaction of similar complicated models, some of the components of which may be hard to test for in deep time. This appears to be the case in models created for Maastrichtian turtles by Waterson et al. (2016), who found that turtle niche occupation was determined primarily by temperature and rainfall variables.

Land area does not appear to have contributed to the recovered richness patterns of turtles or lepidosaurs through the studied time period. AICc values chose land area as a better explanation for variance in Early Paleogene lepidosaur richness over sampling proxies but the R^2 was very low (0.26), which suggests that it is a poor contributor to richness patterns and that other untested factors are more influential. This contrasts with the results obtained for non-avian dinosaurs, for which Cretaceous richness was influenced in part by land area (Mannion et al., 2012), perhaps due to the frequent occurrence of large-bodied forms, that required large land areas to support viable populations (Burness, Diamond and Flannery, 2001), compared to the much smaller-bodied turtles and lepidosaurs. It is also likely that land area appeared to be more important to Cretaceous dinosaurs because they are better sampled in the palaeotropics than turtles and lepidosaurs, and so a better comparison can be made between richness and land area for the former than the latter clades.

It is also likely that other factors, such as the availability of aquatic environments for non-marine aquatic turtle richness, are more important, as in extant taxa (Iverson, 1992; Angielczyk, Burroughs and Feldman, 2015).

Alongside issues of varying sampling effort, I examined the effects of preservational biases on face-value richness counts of non-marine lizards (squamates excluding snakes and amphisbaenians), using non-weighted completeness metrics that measure individual specimen preservation. More specific (e.g. percentage-based) completeness metrics were difficult to apply to lepidosaurs as their body plans vary widely, even when excluding limbless taxa, so these methods are not easily applicable to this clade. I found no correlation between the average completeness of the best-preserved specimens of taxa in a time bin and average richness for lepidosaurs. Many lepidosaur specimens comprise single elements, which are still diagnostic to lower taxonomic levels, meaning that high preservation is not needed for taxa to be counted towards generic or species richness. While specimen preservational biases do not affect the identification of taxa (and thus do not negatively impact richness levels in any particular time bin), the presence of Lagerstätten does influence lepidosaur richness somewhat due to the increased chance and ease of finding microvertebrates. Elsewhere, special sampling techniques such as screenwashing sediments are often required to recover microvertebrates, but in Lagerstätten, particularly those with fine-grained rocks (e.g. lithographic limestones) that preserve fully articulated specimens, lepidosaurs are more likely to be found and recorded in richness counts. It is facies type, therefore, that is more important to lepidosaur richness than individual specimen completeness, whereas the latter just coincides with the presence of Lagerstätten.

6.2 The limitations of richness estimations

With regards to tracking taxon richness alongside major changes in temperature and other climatic factors, it should be noted that we are hampered by sampling biases even after correcting face-value observations with methods such as subsampling. After combining continental-scale richness and the same richness divided into latitudinal bands per time slice, it becomes clear that there are many major large-scale geographic biases in the record, and that these prevent us from fully understanding what occurred around some major extinction and turnover events. For example, richness decreased for all of the ectothermic taxa discussed above at the Eocene–Oligocene boundary (in every Northern Hemisphere continent that is well-sampled enough at quorum 0.4 to be recovered). Taking North America as an example, it appears that richness decreased at higher latitudes (45–60°N) but remained high or stable at lower palaeotemperate latitudes (30–5°N), meaning a net decrease was observed. Unfortunately sampling is too poor at palaeotropical latitudes (0–30°N) and in the majority of the Southern Hemisphere to discover if taxa became extinct or simply migrated out of the area to more southerly refugia, besides finding their records in later palaeotemperate localities. This makes it difficult to draw rigorous conclusions about how ectotherms responded to events such as sudden climatic change (e.g. the Grande Coupure) on a global scale when only a part of the evidence we need to fully explore these events exists currently.

This major drawback with respect to poor geographic sampling also affects the methods used to correct raw face-value observation counts for sampling biases. Essentially, the areas that are sampled through time are not equal. In Chapter 2 I used an example to explain that richness estimates between bins Pg2 and Pg3 of North America are biased by the fact that the majority of localities sampled in each bin are in completely separate parts of the USA and this caveat also applies to many other instances in the fossil record. Close et al. (2017, 2018) investigated the effects of evenness on the results obtained using simulated and real-world richness data with various richness estimators (e.g. SQS, CR, TRiPS). They found that every method was affected by differences in evenness, particularly when sampling was relatively limited, as it becomes more difficult to detect rarer taxa when geographic spread is uneven. This means that richness through time obtained by subsampling methods at large scales may still be far from 'true' richness, and that it is better to examine richness at smaller, regional scales. There are also problems created by the use of the palaeontological literature to inform face-value (and subsequent subsampled) richness curves, as the literature is inherently biased towards novel taxa. This can inflate singleton counts, as subsequent finds of the same taxon may not be reported, which in turn confounds the calculations of richness estimators (Alroy, 2010a). This problem is investigated by Marshall et al. (2018), who surveyed the differences between the literature and museum collections of Cenozoic marine invertebrates and found that there was a lot of unused data in the museum collections.

In terms of the methods that are currently available, coverage-based subsampling methods such as SQS are still the best techniques we have available to examine richness in deep time, and we should continue to use coverage-based sampling until better methods are developed. Close et al. (2018) recommend using these methods with sampling-standardised discovery curves, which show the chronological addition of occurrences to represent sampling effort through time. This would help researchers to understand more clearly how literature biases through time interact with differences in sampling effort in different time bins, and the sampling-corrected curves produced from variably biased data.

6.3 Future work

Accounting for geographic spread over large geographic scales is difficult, but examining richness at smaller regional levels, or using methods that standardise geographic spread, should help to ameliorate the shortcomings of richness estimators such as SQS. Close et al. (2017) did this for a Triassic–Ypresian dataset of non-volant terrestrial tetrapods, using minimum-spanning trees to standardise geographic spread, and found that it helped to produce richness curves that, while they may not represent 'true' richness, reduced the effect of uneven spread of localities and created approximately comparable points of richness through time. It would be interesting to apply these methods to the lepidosaur and turtle records to see how richness varies compared to curves generated using SQS. Additionally, richness estimators should be applied to ectotherm clades that have yet to be studied in detail (especially those that have extant descendants such as fish and lissamphibians), to provide further comparisons to the reptile clades examined herein.

Although skeletal completeness metrics did not appear to affect lepidosaur richness through time, or correlate with any major environmental factors, it would be interesting to determine if the addition of a proxy for the presence or absence of Lagerstätten influences the fit of models in GLS analyses against completeness. If possible other completeness metrics, such as those that use the presence of phylogenetic characters to assess preservation, should be tested with the lepidosaur fossil record, to see if the same lack of correlation occurs. Completeness metrics should also be applied to the non-marine turtle record as a comparison to those groups already tested.

In order to examine latitudinal richness of ectotherms in more depth, I would like to apply methods such as ecological niche modelling to lepidosaurs and turtles in deep time. This enables us to simulate where suitable habitats for taxa existed and compared them to the fossil record. This has already been done for Maastrichtian turtles (Waterson et al., 2016), but should be expanded to examine more time periods. This would provide an extra dimension for studying the effects of changing climate on latitudinal richness of both clades, and possibly assist in projecting what richness should look like in areas that are more poorly sampled (e.g. the palaeotropics).

Museum collections represent a cache of 'dark data' that is not often represented in the literature, and so I would like to examine the differences between the literature and museum collections for certain time periods, as done by Marshall et al. (2018), and create subsampled richness curves based on combined museum and literature data. This could decrease the number of singletons and increase coverage in the studied time bin and improve subsampled richness signals.

Finally, if given more time, I would extend the studied time period of both clades to encompass the Neogene and possibly extend towards the modern day, as a great amount of climatic upheaval continues to occur in this time. It would also help to explain how the richness of turtles and lepidosaurs developed towards their extant levels compared to those in the past.

6.4 Conclusions

The fossil record of terrestrial lepidosaurs and turtles, as well as for non-marine tetrapods as a whole, is poorly sampled, particularly outside of the Northern Hemisphere, and this colours our view of their taxon richness levels through time when using face-value observations. Sampling estimators can help to alleviate the biases caused by variance in sampling effort, but richness estimates are affected when geographic unevenness is high, as occurs frequently in the fossil records of both clades. Regardless, these estimators are currently the best methods we have for examining taxon richness through time and they do appear to account for some, if not all, of the sampling biases affecting the fossil record. In order to gain the best overview of diversity through time, it is better to combine multiple methods of examining these biases, as I have done here, because investigating single clades from multiple angles helps to highlight common patterns in richness, biases or both and thus provide tests of their robustness. It is also better to use these methods for sampling correction, while acknowledging that they may not be a panacea for biases, than to rely solely on face-value richness curves as so many other studies have done in the past.

Understanding patterns of taxon richness through time for ectothermic vertebrates is important for understanding their extant richness patterns. This is particularly important in the face of the rapidly changing climates of today, as ectotherms are expected to react significantly to sudden temperature changes. Non-marine lepidosaurs and turtles had fluctuating taxon richness from the Triassic to the Paleogene, with peaks in richness coincident with times of very warm climates (e.g. Late Cretaceous, Paleocene–Eocene boundary). It is possible that both clades were buffered from rapidly decreasing temperatures over the Eocene–Oligocene boundary by gradual adaptation to slowly decreasing temperatures throughout the preceding Eocene, or by opportunities to migrate to warmer southerly latitudes. Richness at the Eocene–Oligocene boundary falls, but remains relatively high compared to richness levels pre-Late Cretaceous, with a few exceptions that could represent poor adaptation to associated increasing aridity, other local climatic conditions or problems with sampling (e.g. Oligocene North American turtles). My findings suggest that both clades flourished at higher temperatures, which could bode well for current reptile richness under most current scenarios of future anthropogenic warming. However, a number of factors that are prevalent today did not exist in the past and these could alter many of the potential responses to current climatic warming: for example, latitudinal migration might be difficult due to anthropogenic settlement. Besides providing baseline data and potential analogues for understanding the impact of current climatic change, further work could help elucidate how the taxon richness of lepidosaurs and turtles reached their current levels, especially with respect to the highly diverse lepidosaurs, particularly if these data were extended into the Neogene and beyond.

[This page intentionally left blank]

References

Albert, E. M., San Mauro, D., García-París, M., Rüber, L. and Zardoya, R. (2009) Effect of taxon sampling on recovering the phylogeny of squamate reptiles based on complete mitochondrial genome and nuclear gene sequence data. *Gene*, 441, pp. 12–21. doi: 10.1016/j.gene.2008.05.014.

Albino, A. M. (2011) Evolution of Squamata reptiles in Patagonia based on the fossil record. *Biological Journal of the Linnean Society*, 103, pp. 441–457. doi: 10.1111/j.1095-8312.2011.01691.x.

Alifanov, V. R. (2000a) Macrocephalosaurs and the early evolution of lizards of Central Asia. *Transactions of the Paleontological Institute of the Russian Academy of Sciences*, 272, p. 126.

Alifanov, V. R. (2000b) The fossil record of Cretaceous lizards from Mongolia. in Benton, M. J., Shishkin, M. A., Unwin, D. M., and Kurochkin, E. N. (eds) *The Age of Dinosaurs in Russia and Mongolia*. Cambridge: Cambridge University Press, pp. 368–389.

Allison, P. A. (1988) Konservat-lagerstatten: cause and classification. *Paleobiology*, 14(4), pp. 331–344.

Alroy, J. (1996) Constant extinction, constrained diversification, and uncoordinated stasis in North American mammals. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 127, pp. 285–311.

Alroy, J. et al. (2001) Effects of sampling standardization on estimates of Phanerozoic marine diversification. *Proceedings of the National Academy of Sciences of the United States of America*, 98(11), pp. 6261–6266. doi: 10.1073/pnas.111144698.

Alroy, J. (2010a) Fair sampling of taxonomic richness and unbiased estimation of origination and extinction rates. *Paleontological Society Papers*, 16, pp. 55–80.

Alroy, J. (2010b) Geographical, environmental and intrinsic biotic controls on Phanerozoic marine diversification. *Palaeontology*, 53(6), pp. 1211–1235. doi: 10.1111/j.1475-4983.2010.01011.x.

Alroy, J. (2010c) The shifting balance of diversity among major marine animal groups. *Science*, 329(5996), pp. 1191–1194. doi: 10.1126/science.1189910.

Alvarado-Ortega, J., Espinosa-Arrubarrena, L., Blanco, A., Vega, F. J., Benammi, M. and Briggs, D. E. G. (2007) Exceptional preservation of soft tissues in Cretaceous fishes from the Tlayua Quarry, Central Mexico. *Palaios*, 22(6), pp. 682–685. doi: 10.2110/palo.2006.p06-059r.

Alvarez, L. W., Alvarez, W., Asasro, F. and Michel, H. V (1980) Extraterrestrial cause for the Creaceous--Tertiary extinction, experimental results and theoretical interpretation. *Science*, 208(4448), pp. 1095–1108.

Angielczyk, K. D., Burroughs, R. W. and Feldman, C. R. (2015) Do turtles follow the rules? Latitudinal gradients in species richness, body size, and geographic range area of the world's turtles. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 324(3), pp. 270–294. doi: 10.1002/jez.b.22602.

Anquetin, J., Barrett, P. M., Jones, M. E. ., Moore-Fay, S. and Evans, S. E. (2009) A new stem turtle from the Middle Jurassic of Scotland: new insights into the evolution and palaeoecology of basal turtles. *Proceedings of the Royal Society B: Biological Sciences*, 276(1658), pp. 879–886. doi: 10.1098/rspb.2008.1429.

Anquetin, J., Püntener, C. and Joyce, W. G. (2017) A review of the fossil record of turtles of the clade Thalassochelydia. *Bulletin of the Peabody Museum of Natural History*, 58(2), pp. 317–369. doi: 10.3374/014.058.0205.

Apesteguía, S., Gómez, R. O. and Rougier, G. W. (2014) The youngest South American rhynchocephalian, a survivor of the K/Pg extinction. *Proceedings of the Royal Society B: Biological Sciences*, 281(20140811), pp. 1–6. doi: 10.1098/rspb.2014.0811.

Apesteguía, S. and Jones, M. E. H. (2012) A Late Cretaceous tuatara (Lepidosauria: Sphenodontinae) from South America. *Cretaceous Research*. Elsevier Ltd, 34, pp. 154–160. doi: 10.1016/j.cretres.2011.10.014.

Araújo, M. B., Thuiller, W. and Pearson, R. G. (2006) Climate warming and the decline of amphibians and reptiles in Europe. *Journal of Biogeography*, 33, pp. 1712–1728. doi: 10.1111/j.1365-2699.2006.01482.x.

Archibald, S. B., Bossert, W. H., Greenwood, D. R. and Farrell, B. D. (2010) Seasonality, the latitudinal gradient of diversity, and Eocene insects. *Paleobiology*, 36(3), pp. 374–398.

Archibald, S. B., Greenwood, D. R. and Mathewes, R. W. (2013) Seasonality, montane beta diversity, and Eocene insects: testing Janzen's dispersal hypothesis in an equable world. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 371, pp. 1–8. doi: 10.1016/j.palaeo.2012.10.043.

Augé, M. L. (2005) Évolution des lézards du Paléogène en Europe. *Mémoires du Muséum national d'Histoire naturelle*. Muséum national d'Histoire naturelle, 192, p. 369.

Augé, M. L. (2012) Amphisbaenians from the European Eocene: a biogeographical review. *Palaeobiodiversity and Palaeoenvironments*, 92(4), pp. 425–443. doi:

10.1007/s12549-012-0104-6.

Augé, M. and Pouit, D. (2012) Presence of iguanid lizards in the European Oligocene Lazarus taxa and fossil abundance. *Bulletin de la Societe Geologique de France*, 183(6), pp. 653–660. doi: 10.2113/gssgfbull.183.6.653.

Augé, M. and Rage, J. C. (2006) Herpetofaunas from the upper Paleocene and lower Eocene of Morocco. *Annales de Paleontologie*, 92, pp. 235–253. doi: 10.1016/j.annpal.2005.09.001.

Augé, M. and Smith, R. (2009) An assemblage of early Oligocene lizards (Squamata) from the locality of Boutersem (Belgium), with comments on the Eocene-Oligocene transition. *Zoological Journal of the Linnean Society*, 155, pp. 148–170. doi: 10.1111/j.1096-3642.2008.00435.x.

Bailleul, A., Ségalen, L., Buscalioni, A. D., Cambra-Moo, O. and Cubo, J. (2011) Palaeohistology and preservation of tetrapods from Las Hoyas (Lower Cretaceous, Spain). *Comptes Rendus Palevol*. Academie des sciences, 10(5–6), pp. 367–380. doi: 10.1016/j.crpv.2011.05.002.

Bakker, R. T. (1972) Anatomical and ecological evidence of endothermy in dinosaurs. *Nature*, 238, pp. 81–85. doi: 10.1038/239137a0.

Barrett, P. M., McGowan, A. J. and Page, V. (2009) Dinosaur diversity and the rock record. *Proceedings of the Royal Society B: Biological Sciences*, 276, pp. 2667–2674. doi: 10.1098/rspb.2009.0352.

Beardmore, S. R., Orr, P. J., Manzocchi, T. and Furrer, H. (2012) Float or sink: modelling the taphonomic pathway of marine crocodiles (Mesoeucrocodylia, Thalattosuchia) during the death–burial interval. *Palaeobiodiversity and*

Palaeoenvironments, 92(1), pp. 83–98. doi: 10.1007/s12549-011-0066-0.

Bennett, S. C. (2002) Soft tissue preservation of the cranial crest of the pterosaur Germanodactylus from Solnhofen. *Journal of Vertebrate Paleontology*, 22(1), pp. 43–48.

Benson, R. B. J. and Butler, R. J. (2011) Uncovering the diversification history of marine tetrapods: ecology influences the effect of geological sampling biases. *Geological Society, London, Special Publications*, 358(1), pp. 191–208. doi: 10.1144/SP358.13.

Benson, R. B. J., Butler, R. J., Alroy, J., Mannion, P. D., Carrano, M. T. and Lloyd, G. T. (2016) Near-stasis in the long-term diversification of Mesozoic tetrapods. *PLoS Biology*, 14(1), p. e1002359. doi: 10.1371/journal.pbio.1002359.

Benson, R. B. J., Butler, R. J., Lindgren, J. and Smith, A. S. (2010) Mesozoic marine tetrapod diversity: mass extinctions and temporal heterogeneity in geological megabiases affecting vertebrates. *Proceedings of the Royal Society B: Biological Sciences*, 277, pp. 829–34. doi: 10.1098/rspb.2009.1845.

Benson, R. B. J. and Mannion, P. D. (2012) Multi-variate models are essential for understanding vertebrate diversification in deep time. *Biology Letters*, 8(1), pp. 127–130. doi: 10.1098/rsbl.2011.0460.

Benson, R. B. J., Mannion, P. D., Butler, R. J., Upchurch, P., Goswami, A. and Evans, S. E. (2013) Cretaceous tetrapod fossil record sampling and faunal turnover: Implications for biogeography and the rise of modern clades. *Palaeogeography, Palaeoclimatology, Palaeoecology*. Elsevier B.V., 372, pp. 88–107. doi: 10.1016/j.palaeo.2012.10.028.

Benton, M. J. (2008a) Fossil quality and naming dinosaurs. *Biology Letters*, 4, pp. 729–32. doi: 10.1098/rsbl.2008.0402.

Benton, M. J. (2008b) How to find a dinosaur, and the role of synonymy in biodiversity studies. *Paleobiology*, 34(4), pp. 516–533.

Benton, M. J., Dunhill, A. M., Lloyd, G. T. and Marx, F. G. (2011) Assessing the quality of the fossil record: insights from vertebrates. *Geological Society, London, Special Publications*, 358(1), pp. 63–94. doi: 10.1144/SP358.6.

Benton, M. J., Ruta, M., Dunhill, A. M. and Sakamoto, M. (2013) The first half of tetrapod evolution, sampling proxies, and fossil record quality. *Palaeogeography, Palaeoclimatology, Palaeoecology*. Elsevier B.V., 372, pp. 18–41. doi: 10.1016/j.palaeo.2012.09.005.

Bickford, D., Howard, S. D., Ng, D. J. J. and Sheridan, J. a. (2010) Impacts of climate change on the amphibians and reptiles of Southeast Asia. *Biodiversity and Conservation*, 19, pp. 1043–1062. doi: 10.1007/s10531-010-9782-4.

Bijl, P. K., Schouten, S., Sluijs, A., Reichart, G. J., Zachos, J. C. and Brinkhuis, H. (2009) Early palaeogene temperature evolution of the southwest Pacific Ocean. *Nature*. Nature Publishing Group, 461(7265), pp. 776–779. doi: 10.1038/nature08399.

Blackburn, T. J. et al. (2013) Zircon U-Pb geochronology links the End-Triassic extinction with the Central Atlantic Magmatic Province. *Science*, 340(6135), pp. 941–945. doi: 10.1126/science.1234204.

Blain, H.-A., Canudo, J.-I., Cuenca-Bescós, G. and López-Martínez, N. (2010) Amphibians and squamate reptiles from the latest Maastrichtian (Upper Cretaceous) of Blasi 2 (Huesca, Spain). *Cretaceous Research*. Elsevier Ltd, 31(4), pp. 433–446. doi:

10.1016/j.cretres.2010.06.001.

Böhme, M. (2007) Herpetofauna (Anura, Squamata) and palaeoclimatic implications: preliminary results. *Annalen des Naturhistorischen Museums in Wien. Serie A fur Mineralogie und Petrographie, Geologie und Paläontologie, Anthropologie und Prähistorie*, 108A, pp. 43–52.

Bolet, A. and Evans, S. E. (2013) Lizards and amphisbaenians (Reptilia, Squamata) from the late Eocene of Sossís (Catalonia, Spain). *Palaeontologia Electronica*, 16(1), pp. 1–23.

Borsuk-Bialynicka, M. (1984) Anguimorphans and related lizards from the Late Cretaceous of the Gobi Desert. *Palaeontologia Polonica*, 46, pp. 5–105.

Bourque, J. R. (2016) Side-necked turtles (Testudines, Pleurodira) from the ancient Gulf Coastal Plain of Florida during middle Cenozoic megathermals. *Chelonian Conservation and Biology*, 15(1), pp. 23–35. doi: 10.2744/CCB-1159.1.

Breusch, T. S. and Pagan, A. R. (1979) A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, 47(5), pp. 1287–1294.

Brinkman, D. B. (2003) A review of nonmarine turtles from the Late Cretaceous of Alberta. *Canadian Journal of Earth Sciences*, 40(4), pp. 557–571. doi: 10.1139/e02-080.

Brinkman, D. B., Rabi, M. and Zhao, L. (2017) Lower Cretaceous fossils from China shed light on the ancestral body plan of crown softshell turtles. *Scientific Reports*, 7(1), pp. 1–11. doi: 10.1038/s41598-017-04101-0.

Brochu, C. A. (2003) Phylogenetic approaches towards crocdylian history. *Annual Review of Earth and Planetary Sciences*, 31(1), pp. 357–397. doi: 10.1146/annurev.earth.31.100901.141308.

Brocklehurst, N., Day, M. O., Rubidge, B. S. and Fröbisch, J. (2017) Olson's Extinction and the latitudinal biodiversity gradient of tetrapods in the Permian. *Proceedings of the Royal Society B: Biological Sciences*, 284(1852). doi: 10.1098/rspb.2017.0231.

Brocklehurst, N. and Fröbisch, J. (2014) Current and historical perspectives on the completeness of the fossil record of pelycosaurian-grade synapsids. *Palaeogeography, Palaeoclimatology, Palaeoecology*. Elsevier B.V., 399, pp. 114–126. doi: 10.1016/j.palaeo.2014.02.006.

Brocklehurst, N., Upchurch, P., Mannion, P. D. and O'Connor, J. (2012) The completeness of the fossil record of Mesozoic birds: implications for early avian evolution. *PloS One*, 7(6), p. e39056. doi: 10.1371/journal.pone.0039056.

Bronzati, M., Montefeltro, F. C. and Langer, M. C. (2015) Diversification events and the effects of mass extinctions on Crocodyliformes evolutionary history. *Royal Society Open Science*, 2, p. 140385. doi: 10.1098/rsos.140385.

Brown, C. M., Evans, D. C., Campione, N. E., O'Brien, L. J. and Eberth, D. a. (2013) Evidence for taphonomic size bias in the Dinosaur Park Formation (Campanian, Alberta), a model Mesozoic terrestrial alluvial-paralic system. *Palaeogeography, Palaeoclimatology, Palaeoecology*. Elsevier B.V., 372, pp. 108–122. doi: 10.1016/j.palaeo.2012.06.027.

Buhlmann, K. A. et al. (2009) A global analysis of tortoise and freshwater turtle distributions with identification of priority conservation areas. *Chelonian Conservation and Biology*, 8(2), pp. 116–149. doi: 10.2744/CCB-0774.1.

Burness, G. P., Diamond, J. and Flannery, T. (2001) Dinosaurs, dragons, and dwarfs: the evolution of maximal body size. *Proceedings of the National Academy of Sciences*, 98(25), pp. 14518–14523. doi: 10.1073/pnas.251548698.

Burnham, K. P. and Anderson, D. R. (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. New York: Springer-Verlag. doi: 10.1007/b97636.

Butler, R. J., Benson, R. B. J. and Barrett, P. M. (2013) Pterosaur diversity: untangling the influence of sampling biases, Lagerstatten, and genuine biodiversity signals. *Palaeogeography, Palaeoclimatology, Palaeoecology*. Elsevier B.V., 372, pp. 78–87. doi: 10.1016/j.palaeo.2012.08.012.

Butler, R. J., Benson, R. B. J., Carrano, M. T., Mannion, P. D. and Upchurch, P. (2011) Sea level, dinosaur diversity and sampling biases: investigating the 'common cause' hypothesis in the terrestrial realm. *Proceedings of the Royal Society B: Biological Sciences*, 278(1709), pp. 1165–70. doi: 10.1098/rspb.2010.1754.

Caldwell, M. W. and Lee, M. S. Y. (1997) A snake with legs from the marine Cretaceous of the Middle East. *Nature*, pp. 705–709. doi: 10.1038/386705a0.

Caldwell, M. W., Nydam, R. L., Palci, A. and Apesteguía, S. (2015) The oldest known snakes from the Middle Jurassic-Lower Cretaceous provide insights on snake evolution. *Nature Communications*, 6(5996), pp. 1–11. doi: 10.1038/ncomms6996.

Caprette, C. L., Lee, M. S. Y., Shine, R., Mokany, A. and Downhower, J. F. (2004) The origin of snakes (Serpentes) as seen through eye anatomy. *Biological Journal of the Linnean Society*, 81(4), pp. 469–482. doi: 10.1111/j.1095-8312.2003.00305.x.

Chatfield, C. (2003) *The Analysis of Time Series: An Introduction*. Florida: Chapman and Hall.

Chown, S. L., Gaston, K. J. and Williams, P. H. (1998) Global patterns in species richness of pelagic seabirds: the Procellariiformes. *Ecography*, 21(4), pp. 342–350.

Claude, J. and Tong, H. (2004) Early Eocene testudinoid turtles from Saint Papoul, France, with comments on the early evolution of Testudinoidea. *Oryctos*, 5(January), pp. 3–45.

Cleary, T. J., Benson, R. B. J., Evans, S. E. and Barrett, P. M. (2018) Lepidosaurian diversity in the Mesozoic–Palaeogene: The potential roles of sampling biases and environmental drivers. *Royal Society Open Science*, 5(3). doi: 10.1098/rsos.171830.

Cleary, T. J., Moon, B. C., Dunhill, A. M. and Benton, M. J. (2015) The fossil record of ichthyosaurs, completeness metrics and sampling biases. *Palaeontology*, pp. 1–16. doi: 10.1111/pala.12158.

Clemens, W. A. (1982) Patterns of extinction and survival of the terrestrial biota during the Cretaceous/Tertiary transition. *Geological Society of America Special Papers*, 190, pp. 407–413.

Close, R. A., Benson, R. B. J., Upchurch, P. and Butler, R. J. (2017) Controlling for the species-area effect supports constrained long-term Mesozoic terrestrial vertebrate diversification. *Nature Communications*. Nature Publishing Group, 8(May), p. 15381. doi: 10.1038/ncomms15381.

Close, R. A., Evers, S. W., Alroy, J. and Butler, R. J. (2018) How should we estimate diversity in the fossil record? Testing richness estimators using sampling-standardised discovery curves. *Methods in Ecology and Evolution*, 9(6), pp. 1386–1400. doi: 10.1111/2041-210X.12987.

Coiffard, C. and Gomez, B. (2012) Influence of latitude and climate on spread, radiation and rise to dominance of early angiosperms during the Cretaceous in the Northern Hemisphere. *Geologica Acta*, 10(2), pp. 181–188. doi: 10.1344/105.000001701.

Corsini, J. A., Smith, T. R. and Leite, M. B. (2011) Turtle decline in the early Oligocene of Western Nebraska. *Journal of Herpetology*, 45(2), pp. 238–243.

Costa, E., Garcés, M., Sáez, A., Cabrera, L. and López-Blanco, M. (2011) The age of the 'Grande Coupure' mammal turnover: new constraints from the Eocene-Oligocene record of the eastern Ebro Basin (NE Spain). *Palaeogeography, Palaeoclimatology, Palaeoecology*. Elsevier B.V., 301(1–4), pp. 97–107. doi: 10.1016/j.palaeo.2011.01.005.

Crame, J. A., Antarctic, B., Cross, H. and Road, M. (2001) Taxonomic diversity gradients through geological time. *Diversity and Distributions*, 7(4), pp. 175–189.

Cuny, G., Jaeger, J.-J., Mahboubi, M. and Rage, J.-C. (1990) Les plus anciens serpents (Reptilia, Squamata) connus. Mise au point sue l'âge géologique des serpents de la partie moyenne du Crétacé. *Comptes rendus de l'Académie des sciences. Série II, Mécanique, physique, chimie, sciences de l'univers, sciences de la terre*, 311(10), pp. 1267–1272.

Dal Sasso, C. and Signore, M. (1998) Exceptional soft-tissue preservation in a theropod dinosaur from Italy. *Nature*, 392(6674), pp. 383–387. doi: 10.1038/32884.

Danilov, I. G. and Syromyatnikova, E. V. (2008) New materials on turtles of the family Nanhsiungchelyidae from the Cretaceous of Uzbekistan and Mongolia, with a review of the Nanhsiungchelyid record in Asia. *Proceedings of the Zoological Institute RAS*, 312(1/2), pp. 3–25.

Davies, T. W., Bell, M. A., Goswami, A. and Halliday, T. J. D. (2017) Completeness of the eutherian mammal fossil record and implications for reconstructing mammal evolution through the Cretaceous/Paleogene mass extinction. *Paleobiology*, 43(4), pp. 521–536. doi: 10.1017/pab.2017.20.

Dean, C. D., Mannion, P. D. and Butler, R. J. (2016) Preservation bias controls the fossil record of pterosaurs. *Palaeontology*, 59(2), pp. 225–247. doi: 10.1111/pala.12225.

Driscoll, D. A., Dunhill, A. M., Stubbs, T. L. and Benton, M. J. (2018) The mosasaur fossil record through the lens of fossil completeness. *Palaeontology*, pp. 1–25. doi: 10.1111/pala.12381.

Dunhill, A. M., Hannisdal, B. and Benton, M. J. (2014) Disentangling rock record bias and common-cause from redundancy in the British fossil record. *Nature Communications*. Nature Publishing Group, 5, p. 4818. doi: 10.1038/ncomms5818.

Dunhill, A. M., Hannisdal, B., Brocklehurst, N. and Benton, M. J. (2018) On formation-based sampling proxies and why they should not be used to correct the fossil record. *Palaeontology*, 61(1), pp. 119–132. doi: 10.1111/pala.12331.

Dunne, E. M., Close, R. A., Button, D. J., Brocklehurst, N., Cashmore, D. D., Lloyd, G. T. and Butler, R. J. (2018) Diversity change during the rise of tetrapods and the impact of the 'Carboniferous rainforest collapse'. *Proceedings of the Royal Society B: Biological Sciences*, 285(1872), p. 20172730. doi: 10.1098/rspb.2017.2730.

Eronen, J. T., Janis, C. M., Chamberlain, C. P. and Mulch, A. (2015) Mountain uplift explains differences in Palaeogene patterns of mammalian evolution and extinction between North America and Europe. *Proceedings of the Royal Society B: Biological Sciences*, 282(1809), p. 20150136. doi: 10.1098/rspb.2015.0136.

Erwin, D. H. (2009) Climate as a Driver of Evolutionary Change. *Current Biology*. Elsevier Ltd, 19(14), pp. R575–R583. doi: 10.1016/j.cub.2009.05.047.

Estes, R. and Hutchison, J. H. (1980) Eocene lower vertebrates from Ellesmere Island, Canadian Arctic Archipelago. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 30(C), pp. 325–347. doi: 10.1016/0031-0182(80)90064-4.

Evans, S. E. (1994) A new anguimorph lizard from the Jurassic and Lower Cretaceous of England. *Palaeontology*, pp. 33–49.

Evans, S. E. (1998) Lepidosaurian faunas from the Early Cretaceous: a clade in transition. *New Mexico Museum of Natural History and Science Bulletin*, 14, pp. 195–200.

Evans, S. E. (2003) At the feet of the dinosaurs: the early history and radiation of lizards. *Biological Reviews of the Cambridge Philosophical Society*, 78, pp. 513–551. doi: 10.1014/S1464793103006134.

Evans, S. E. and Jones, M. E. H. (2010) The origins, early history and diversification of lepidosauromorph reptiles. in Bandyopadhyay, S. (ed.) *New Aspects of Mesozoic Biodiversity*. Berlin: Springer-Verlag, pp. 27–44. doi: 10.1007/978-3-642-10311-7.

Evans, S. E., Prasad, G. V. R. and Manhas, B. K. (2001) Rhynchocephalians (Diapsida: Lepidosauria) from the Jurassic Kota Formation of India. *Zoological Journal of the Linnean Society*, 133, pp. 309–334. doi: 10.1006/zjls.2000.0266.

Evans, S. E., Prasad, G. V. R. and Manhas, B. K. (2002) Fossil lizards from the Jurassic Kota Formation of India. *Journal of Vertebrate Paleontology*, 22(2), pp. 299–312. doi: 10.1671/0272-4634(2002)022.

Evans, S. E., Wang, Y. and Li, C. (2005) The Early Cretaceous lizard genus Yabeinosaurus from China: resolving an enigma. *Journal of Systematic Palaeontology*, 3(February 2015), pp. 319–335. doi: 10.1017/S1477201905001641.

Fastovsky, D. E., Huang, Y., Hsu, J., Martin-McNaughton, J., Sheehan, P. M. and Weishampel, D. B. (2004) Shape of Mesozoic dinosaur richness. *Geology*, 32(10), pp.

877–880. doi: 10.1130/G20695.1.

Fenton, I. S., Pearson, P. N., Jones, T. D., Farnsworth, A., Lunt, D. J., Markwick, P. and Purvis, A. (2016) The impact of Cenozoic cooling on assemblage diversity in planktonic foraminifera. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1691), p. 20150224. doi: 10.1098/rstb.2015.0224.

Fielding, S., Martill, D. M. and Naish, D. (2005) Solnhofen-style soft-tissue preservation in a new species of turtle from the Crato Formation (Early Cretaceous, Aptian) of north-east Brazil. *Palaeontology*, 48(6), pp. 1301–1310. doi: 10.1111/j.1475-4983.2005.00508.x.

Folie, A. F., Smith, R. and Smith, T. (2013) New amphisbaenian lizards from the Early Paleogene of Europe and their implications for the early evolution of modern amphisbaenians. *Geologica Belgica*, 16(4), pp. 227–235.

Foote, M. (2000) Origination and extinction components of taxonomic diversity: Paleozoic and post-Paleozoic dynamics. *Paleobiology*, 26(4), pp. 578–605. doi: 10.1666/0094-8373(2000)026<0578:OAECOT>2.0.CO;2.

Forey, P. L., Fortey, R. A., Kenrick, P. and Smith, A. B. (2004) Taxonomy and fossils: a critical appraisal. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 359(1444), pp. 639–653. doi: 10.1098/rstb.2003.1453.

Forster, A., Schouten, S., Baas, M. and Sinninghe Damsté, J. S. (2007) Mid-Cretaceous (Albian-Santonian) sea surface temperature record of the tropical Atlantic Ocean. *Geology*, 35(10), pp. 919–922. doi: 10.1130/G23874A.1.

Foufopoulos, J., Kilpatrick, A. M. and Ives, A. R. (2011) Climate change and elevated extinction rates of reptiles from Mediterranean islands. *The American Naturalist*, 177(1), pp. 119–129. doi: 10.1086/657624.Climate.

Fountaine, T. M. R., Benton, M. J., Dyke, G. J. and Nudds, R. L. (2005) The quality of the fossil record of Mesozoic birds. *Proceedings of the Royal Society B: Biological Sciences*, 272(1560), pp. 289–294. doi: 10.1098/rspb.2004.2923.

Fraser, N. C. and Benton, M. J. (1989) The Triassic reptiles Brachyrhinodon and Polysphenodon and the relationships of the sphenodontids. *Zoological Journal of the Linnean Society*, 96(4), pp. 413–445. doi: 10.1111/j.1096-3642.1989.tb02521.x.

Frazzetta, T. H. (1962) A functional consideration of cranial kinesis in lizards. *Journal of Morphology*, 111(3), pp. 287–319.

Fry, B. G. et al. (2006) Early evolution of the venom system in lizards and snakes. *Nature*, 439(7076), pp. 584–588. doi: 10.1038/nature04328.

Gans, C. (1967) A check list of Recent amphisbaenians (Amphisbaenia, Reptilia). *Bulletin of the American Museum of Natural History*, 135, pp. 61–106.

Gans, C. (1975) Tetrapod limblessness: evolution and functional corollaries. *American Zoologist*, 15(2), pp. 455–467.

Gao, K. and Norell, M. A. (2000) Taxonomic composition and systematics of Late Cretaceous lizard assemblages from Ukhaa Tolgod and adjacent localities, Mongolian Gobi Desert. *Bulletin of the American Museum of Natural History*, 249(249), pp. 1–118. doi: 10.1206/0003-0090(2000)249<0001:TCASOL>2.0.CO;2.

Gardner, J. D. and Cifelli, R. L. (1999) A primitive snake from the Cretaceous of Utah. *Special Papers in Palaeontology*, 60(60), pp. 87–100.

Georgalis, G. L. and Joyce, W. G. (2017) A review of the fossil record of Old World turtles of the clade Pan-Trionychidae. *Bulletin of the Peabody Museum of Natural*

History, 58(1), pp. 115–208.

Godinot, M. and de Broin, F. de L. (2003) Arguments for a mammalian and reptilian dispersal from Asia to Europe during the Paleocene-Eocene boundary interval. *Deinsea*, 10, pp. 255–275.

Good, I. J. (1953) The population frequencies of species and the estimation of population parameters. *Biometrika*, 40(3/4), pp. 237–264.

Gorr, T. A., Mable, B. K. and Kleinschmidt, T. (1998) Phylogenetic analysis of reptilian hemoglobins: trees, rates, and divergences. *Journal of Molecular Evolution*, 47, pp. 471–485. doi: 10.1007/PL00006404.

Green, B. P. (1998) *Fossil lizard assemblages from the Hampshire Basin (Palaeogene) of southern England*. University College London.

Grigoriev, D. V (2014) Giant Mosasaurus hoffmannii (Squamata, Mosauridae) from the Late Cretaceous (Maastrichtian) of Panza, Russia. *Proceedings of the Zoological Institute RAS*, 318(2), pp. 148–167.

Grossnickle, D. M. and Newham, E. (2016) Therian mammals experience an ecomorphological radiation during the Late Cretaceous and selective extinction at the K–Pg boundary. *Proceedings of the Royal Society B: Biological Sciences*, 283(1832), pp. 1–8. doi: 10.1098/rspb.2016.0256.

Gunderson, A. R. and Stillman, J. H. (2015) Plasticity in thermal tolerance has limited potential to buffer ectotherms from global warming. *Proceedings of the Royal Society B: Biological Sciences*, 282(1808), pp. 20150401–20150401. doi: 10.1098/rspb.2015.0401.

Hannisdal, B. and Peters, S. E. (2011) Phanerozoic Earth system evolution and marine biodiversity. *Science*, 334(6059), pp. 1121–1124. doi: 10.1126/science.1210695.

Harrington, G. J., Eberle, J., Le-Page, B. A., Dawson, M. and Howard Hutchison, J. (2012) Arctic plant diversity in the early Eocene greenhouse. *Proceedings of the Royal Society B: Biological Sciences*, 279(1733), pp. 1515–1521. doi: 10.1098/rspb.2011.1704.

Head, J. J. et al. (2009) Giant boid snake from the Palaeocene neotropics reveals hotter past equatorial temperatures. *Nature*, 457(7230), pp. 715–717. doi: 10.1038/nature10271.

Head, J. J., Gunnell, G. F., Holroyd, P. A., Hutchison, J. H. and Ciochon, R. L. (2013) Giant lizards occupied herbivorous mammalian ecospace during the Paleogene greenhouse in Southeast Asia. *Proceedings of the Royal Society B: Biological Sciences*, 280, p. 20130665. doi: 10.1098/rspb.2013.0665.

Hellawell, J. and Orr, P. J. (2012) Deciphering taphonomic processes in the Eocene Green River Formation of Wyoming. *Palaeobiodiversity and Palaeoenvironments*, 92(3), pp. 353–365. doi: 10.1007/s12549-012-0092-6.

Hirayama, R., Brinkman, D. B. and Danilov, I. G. (2000) Distribution and biogeography of non-marine Cretaceous turtles. *Russian Journal of Herpetology*, 7(3), pp. 181–198.

Hofmeyr, M. D., Vamberger, M., Branch, W., Schleicher, A. and Daniels, S. R. (2017) Tortoise (Reptilia, Testudinidae) radiations in Southern Africa from the Eocene to the present. *Zoologica Scripta*, 46(4), pp. 389–400. doi: 10.1111/zsc.12223.

Holroyd, P. A. and Hutchison, J. H. (2002) Patterns of geographic variation in latest Cretaceous vertebrates: evidence from the turtle component. *Geological Society of America Special Papers*, 361, pp. 177–190.

Holroyd, P. A., Hutchison, J. H. and Strait, S. G. (2001) Turtle diversity and abundance through the lower Eocene Willwood Formation of the southern Bighorn Basin. *University of Michigan Papers on Paleontology*, 33(33), pp. 97–107.

Holroyd, P. A., Wilson, G. P. and Hutchison, J. H. (2014) Temporal changes within the latest Cretaceous and early Paleogene turtle faunas of northeastern Montana. *Geological Society of America Special Papers*, 503, pp. 299–312. doi: 10.1130/2014.2503(11).

Hooker, J. J. (1989) British mammals in the Tertiary period. *Biological Journal of the Linnean Society*, 38(1), pp. 9–21. doi: 10.1111/j.1095-8312.1989.tb01558.x.

Hooker, J. J., Collinson, M. E. and Sille, N. P. (2004) Eocene-Oligocene mammalian faunal turnover in the Hampshire Basin, UK: calibration to the global time scale and the major cooling event. *Journal of the Geological Society*, 161(2), pp. 161–172. doi: 10.1144/0016-764903-091.

Hothorn, T., Zeileis, A., Farebrother, R. W., Cummins, C., Millo, G. and Mitchell, D. (2017) lmtest: testing linear regression models. Available at: https://cran.r-project.org/package=lmtest.

Hsiang, A. Y., Field, D. J., Webster, T. H., Behlke, A. D., Davis, M. B., Racicot, R. A. and Gauthier, J. A. (2015) The origin of snakes: revealing the ecology, behavior, and evolutionary history of early snakes using genomics, phenomics, and the fossil record. *BMC Evolutionary Biology*, 15, pp. 1–22. doi: 10.1186/s12862-015-0358-5.

Hugall, A. F., Foster, R. and Lee, M. S. Y. (2007) Calibration choice, rate smoothing, and the pattern of tetrapod diversification according to the long nuclear gene RAG-1. *Systematic Biology*, 56(04), pp. 543–563. doi: 10.1080/10635150701477825.

Hunt, G., Cronin, T. M. and Roy, K. (2005) Species-energy relationship in the deep sea: a test using the Quaternary fossil record. *Ecology Letters*, 8(7), pp. 739–747. doi: 10.1111/j.1461-0248.2005.00778.x.

Hutchison, J. H. (1982) Turtle, crocodilian, and champsosaur diversity changes in the Cenozoic of the north-central region of western United States. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 37(2–4), pp. 149–164. doi: 10.1016/0031-0182(82)90037-2.

Hutchison, J. H. (1992) Western North American reptile and amphibian record across the Eocene/Oligocene boundary and its climatic implications. in Prothero, D. R. and Berggren, W. A. (eds) *Eocene-Oligocene Climatic and Biotic Evolution*. Princeton: Princeton University Press, pp. 451–463.

Hutchison, J. H. and Archibald, J. D. (1986) Diversity of turtles across the Cretaceous/Tertiary boundary in Northeastern Montana. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 55(1), pp. 1–22. doi: 10.1016/0031-0182(86)90133-1.

Hutchison, J. H. and Holroyd, P. A. (2003) Late Cretaceous and early Paleocene turtles of the Denver Basin, Colorado. *Rocky Mountain Geology*, 38(1), pp. 121–142. doi: 10.2113/gsrocky.38.1.121.

Hyland, E. G., Sheldon, N. D. and Cotton, J. M. (2017) Constraining the early Eocene climatic optimum: a terrestrial interhemispheric comparison. *Bulletin of the Geological Society of America*, 129(1–2), pp. 244–252. doi: 10.1130/B31493.1.

Ivany, L. C., Patterson, W. P. and Lohmann, K. C. (2000) Cooler winters as a possible cause of mass extinctions at the Eocene/Oligocene boundary. *Nature*, 407(6806), pp. 887–890. doi: 10.1038/35038044.

Iverson, J. B. (1992) Global correlates of species richness in turtles. *Herpetological Journal*, 2(3), pp. 77–81.

Jarque, C. M. and Bera, A. K. (1980) Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6, pp. 255–259. doi: 10.1016/0165-1765(80)90024-5.

Jones, M. E. H. (2006) The Early Jurassic clevosaurs from China (Diapsida: Lepidosauria). *New Mexico Museum of Natural History and Science Bulletin*, 37, pp. 548–562.

Jones, M. E. H., Anderson, C. L., Hipsley, C. A., Müller, J., Evans, S. E. and Schoch, R. R. (2013) Integration of molecules and new fossils supports a Triassic origin for Lepidosauria (lizards, snakes, and tuatara). *BMC Evolutionary Biology*, 13, p. 208. doi: 10.1186/1471-2148-13-208.

Jones, M. E. H., Tennyson, A. J. D., Worthy, J. P., Evans, S. E. and Worthy, T. H. (2009) A sphenodontine (Rhynchocephalia) from the Miocene of New Zealand and palaeobiogeography of the tuatara (Sphenodon). *Proceedings of the Royal Society B: Biological Sciences*, 276, pp. 1385–1390. doi: 10.1098/rspb.2008.1785.

Joyce, W. G. (2014) A review of the fossil record of turtles of the clade Pan-Carettochelys. *Bulletin of the Peabody Museum of Natural History*, 55(1), pp. 3–33. doi: 10.3374/014.055.0102.

Joyce, W. G. (2015) The origin of turtles: a paleontological perspective. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 324B, pp. 181–193. doi: 10.1002/jez.b.22609.

Joyce, W. G. (2016) A review of the fossil record of turtles of the clade Pan-

Chelydridae. *Bulletin of the Peabody Museum of Natural History*, 57(1), pp. 21–56. doi: 10.3374/014.057.0103.

Joyce, W. G. (2017) A Review of the Fossil Record of Basal Mesozoic Turtles. *Bulletin of the Peabody Museum of Natural History*, 58(1), pp. 65–113. doi: 10.3374/014.058.0105.

Joyce, W. G. and Bourque, J. R. (2016) A review of the fossil record of turtles of the clade Pan-Kinosternoidea. *Bulletin of the Peabody Museum of Natural History*, 57(1), pp. 57–95. doi: 10.3374/014.057.0104.

Joyce, W. G. and Lyson, T. R. (2015) A review of the fossil record of turtles of the clade Baenidae. *Bulletin of the Peabody Museum of Natural History*, 56(2), pp. 147–183.

Joyce, W. G., Rabi, M., Clark, J. M. and Xu, X. (2016) A toothed turtle from the Late Jurassic of China and the global biogeographic history of turtles. *BMC Evolutionary Biology*. BMC Evolutionary Biology, 16(1), pp. 1–29. doi: 10.1186/s12862-016-0762-5.

Kalmar, A. and Currie, D. J. (2010a) The completeness of the continental fossil record and its impact on patterns of diversification. *Paleobiology*, 36(1), pp. 51–60. doi: 10.1666/0094-8373-36.1.51.

Kalmar, A. and Currie, D. J. (2010b) The completeness of the continental fossil record and its impact on patterns of diversification. *Paleobiology*, 36(1), pp. 51–60. doi: 10.1666/0094-8373-36.1.51.

Kearney, M., Fujita, M. K. and Ridenour, J. (2009) Lost sex in the reptiles: constraints and correlations. in Schon, I., Martens, K., and van Dijk, P. (eds) *Lost Sex: The Evolutionary Biology of Parthenogenesis*. Baarn: Springer, pp. 447–474. doi: 10.1007/978-90-481-2770-2.

Kiel, S. and Nielsen, S. N. (2010) Quaternary origin of the inverse latitudinal diversity gradient among southern Chilean mollusks. *Geology*, 38(10), pp. 955–958. doi: 10.1130/G31282.1.

Klug, C., Fuchs, D., Schweigert, G., Röper, M. and Tischlinger, H. (2015) New anatomical information on arms and fins from exceptionally preserved Plesioteuthis (Coleoidea) from the Late Jurassic of Germany. *Swiss Journal of Palaeontology*, 134(2), pp. 245–255. doi: 10.1007/s13358-015-0093-y.

Köhler, M., Marín-Moratalla, N., Jordana, X. and Aanes, R. (2012) Seasonal bone growth and physiology in endotherms shed light on dinosaur physiology. *Nature*, 487(7407), pp. 358–361. doi: 10.1038/nature11264.

Kottek, M., Grieser, J., Beck, C., Rudolf, B. and Rubel, F. (2006) World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, 15(3), pp. 259–263. doi: 10.1127/0941-2948/2006/0130.

Krug, A. Z., Jablonski, D., Roy, K. and Beu, A. G. (2010) Differential extinction and the contrasting structure of polar marine faunas. *PLoS ONE*, 5(12). doi: 10.1371/journal.pone.0015362.

Krug, A. Z., Jablonski, D. and Valentine, J. W. (2007) Contrarian clade confirms the ubiquity of spatial origination patterns in the production of latitudinal diversity gradients. *Proceedings of the National Academy of Sciences of the United States of America*, 104(46), pp. 18129–34. doi: 10.1073/pnas.0709202104.

Krug, A. Z., Jablonski, D., Valentine, J. W. and Roy, K. (2009) Generation of Earth's first-order biodiversity pattern. *Astrobiology*, 9(1), pp. 113–124. doi: 10.1089/ast.2008.0253.

Lee, M. S. Y. (2005) Molecular evidence and marine snake origins. *Biology Letters*, 1(2), pp. 227–230. doi: 10.1098/rsbl.2004.0282.

Lee, M. S. Y., Palci, A., Jones, M. E. H., Caldwell, M. W., Holmes, J. D. and Reisz, R. R. (2016) Aquatic adaptations in the four limbs of the snake-like reptile Tetrapodophis from the Lower Cretaceous of Brazil. *Cretaceous Research*. Elsevier Ltd, 66, pp. 194–199. doi: 10.1016/j.cretres.2016.06.004.

Lewin, A. et al. (2016) Patterns of species richness, endemism and environmental gradients of African reptiles. *Journal of Biogeography*, 43(12), pp. 2380–2390. doi: 10.1111/jbi.12848.

Li, C., Wu, X. C., Rieppel, O., Wang, L. T. and Zhao, L. J. (2008) An ancestral turtle from the Late Triassic of southwestern China. *Nature*, 456(7221), pp. 497–501. doi: 10.1038/nature07533.

Liu, Z. et al. (2009) Global cooling during the Eocene-Oligocene climate transition. *Science*, 323(5918), pp. 1187–1190. doi: 10.1126/science.1166368.

Lloyd, G. T. (2012) A refined modelling approach to assess the influence of sampling on palaeobiodiversity curves: new support for declining Cretaceous dinosaur richness. *Biology Letters*, 8(1), pp. 123–126. doi: 10.1098/rsbl.2011.0210.

Longrich, N. R., Bhullar, B.-A. S. and Gauthier, J. A. (2012a) A transitional snake from the Late Cretaceous period of North America. *Nature*, 488(7410), pp. 205–208. doi: 10.1038/nature11227.

Longrich, N. R., Bhullar, B.-A. S. and Gauthier, J. A. (2012b) Mass extinction of lizards and snakes at the Cretaceous-Paleogene boundary. *Proceedings of the National Academy of Sciences of the United States of America*, 109(52), pp. 21396–21401. doi:

10.1073/pnas.1304029110.

Lucas, S. G. and Tanner, L. H. (2015) End-Triassic nonmarine biotic events. *Journal of Palaeogeography*. Elsevier Ltd, 4, pp. 331–348. doi: 10.1016/j.jop.2015.08.010.

MacLeod, N. et al. (1997) The Cretaceous- Tertiary biotic transition. *Journal of the Geological Society*, 154, pp. 265–292. doi: 10.1144/gsjgs.154.2.0265.

Mannion, P. D., Benson, R. B. J., Carrano, M. T., Tennant, J. P., Judd, J. and Butler, R. J. (2015) Climate constrains the evolutionary history and biodiversity of crocodylians. *Nature Communications*. Nature Publishing Group, 6, p. 8438. doi: 10.1038/ncomms9438.

Mannion, P. D., Benson, R. B. J., Upchurch, P., Butler, R. J., Carrano, M. T. and Barrett, P. M. (2012) A temperate palaeodiversity peak in Mesozoic dinosaurs and evidence for Late Cretaceous geographical partitioning. *Global Ecology and Biogeography*, 21(9), pp. 898–908. doi: 10.1111/j.1466-8238.2011.00735.x.

Mannion, P. D. and Upchurch, P. (2010) Completeness metrics and the quality of the sauropodomorph fossil record through geological and historical time. *Paleobiology*, 36(2), pp. 283–302. doi: 10.1666/09008.1.

Mannion, P. D., Upchurch, P., Benson, R. B. J. and Goswami, A. (2014) The latitudinal biodiversity gradient through deep time. *Trends in Ecology and Evolution*. Elsevier Ltd, 29(1), pp. 42–50. doi: 10.1016/j.tree.2013.09.012.

Mannion, P. D., Upchurch, P., Carrano, M. T. and Barrett, P. M. (2011) Testing the effect of the rock record on diversity: a multidisciplinary approach to elucidating the generic richness of sauropodomorph dinosaurs through time. *Biological Reviews of the Cambridge Philosophical Society*, 86(1), pp. 157–81. doi: 10.1111/j.1469-

185X.2010.00139.x.

Markwick, P. J. (1994) ‘Equability”, continentality, and Tertiary “climate”: the crocodilian perspective”’. *Geology*, 22(7), pp. 613–616. doi: 10.1130/0091-7613(1994)022<0613:ECATCT>2.3.CO;2.

Marshall, C. R. et al. (2018) Quantifying the dark data in museum fossil collections as palaeontology undergoes a second digital revolution. *Biology Letters*, 14(20180431), pp. 2–5. doi: 10.1098/rsbl.2018.0431.

Martill, D. M., Tischlinger, H. and Longrich, N. R. (2015) A four-legged snake from the Early Cretaceous of Gondwana. *Nature Communications*, 349(6246), pp. 416–419. doi: 10.1126/science.aaa9208.

Marx, F. G. and Uhen, M. D. (2010) Climate, critters, and cetaceans: Cenozoic drivers of the evolution of modern whales. *Science*, 327, pp. 993–996. doi: 10.1126/science.1185581.

Meiri, S. (2008) Evolution and ecology of lizard body sizes. *Global Ecology and Biogeography*, 17(6), pp. 724–734. doi: 10.1111/j.1466-8238.2008.00414.x.

Meng, J. and McKenna, M. (1998) Faunal turnovers of Palaeogene mammals from the Mongolian Plateau. *Nature*, 394(July), pp. 364–367. doi: 10.1038/28603.

Miller, A. I. and Foote, M. (1996) Calibrating the Ordovician radiation of marine life. *Paleobiology*, 22(2), pp. 304–309.

Miller, K. G. et al. (2005) The Phanerozoic record of global sea-level change. *Science*, 310(5752), pp. 1293–8. doi: 10.1126/science.1116412.

Miller, K. G., Browning, J. V., Aubry, M.-P., Wade, B. S., Katz, M. E., Kulpecz, A. A. and Wright, J. D. (2008) Eocene-Oligocene global climate and sea-level changes: St.

Stephens Quarry, Alabama. *Bulletin of the Geological Society of America*, 120(1–2), pp. 34–53. doi: 10.1130/B26105.1.

Mittelbach, G. G. et al. (2007) Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecology Letters*, 10(4), pp. 315–331. doi: 10.1111/j.1461-0248.2007.01020.x.

Moreno-Rueda, G., Pleguezuelos, J. M., Pizarro, M. and Montori, A. (2012) Northward Shifts of the Distributions of Spanish Reptiles in Association with Climate Change. *Conservation Biology*, 26(2), pp. 278–283. doi: 10.1111/j.1523-1739.2011.01793.x.

Mulcahy, D. G., Noonan, B. P., Moss, T., Townsend, T. M., Reeder, T. W., Sites, J. W. and Wiens, J. J. (2012) Estimating divergence dates and evaluating dating methods using phylogenomic and mitochondrial data in squamate reptiles. *Molecular Phylogenetics and Evolution*, 65, pp. 974–991. doi: 10.1016/j.ympev.2012.08.018.

Munnecke, A., Westphal, H. and Kölbl-ebert, M. (2008) Diagenesis of plattenkalk: Examples from the Solnhofen area (Upper Jurassic, southern Germany). *Sedimentology*, 55(6), pp. 1931–1946. doi: 10.1111/j.1365-3091.2008.00975.x.

Nagelkerke, N. J. D. (1991) A note on a general definition of the coefficient of determination. *Biometrika*, 78(3), pp. 691–692. doi: 10.1093/biomet/78.3.691.

Nessov, L. A. (1988) Late Mesozoic amphibians and lizards of Soviet Middle Asia. *Acta Zoologica Cracoviensia*, 31(14), pp. 475–486.

Newham, E., Benson, R., Upchurch, P. and Goswami, A. (2014) Mesozoic mammaliaform diversity: the effect of sampling corrections on reconstructions of evolutionary dynamics. *Palaeogeography, Palaeoclimatology, Palaeoecology*. Elsevier B.V., 412, pp. 32–44. doi: 10.1016/j.palaeo.2014.07.017.

Nicholson, D. B., Holroyd, P. A., Benson, R. B. J. and Barrett, P. M. (2015) Climate-mediated diversification of turtles in the Cretaceous. *Nature Communications*. Nature Publishing Group, 6, p. 7848. doi: 10.1038/ncomms8848.

Nicholson, D. B., Holroyd, P. A., Valdes, P. and Barrett, P. M. (2016) Latitudinal diversity gradients in Mesozoic non-marine turtles. *Royal Society Open Science*, 3, p. 160581. doi: 10.1098/rsos.160581.

Nydam, R. L. (2013) Squamates from the Jurassic and Cretaceous of North America. *Palaeobiodiversity and Palaeoenvironments*, 93, pp. 535–565. doi: 10.1007/s12549-013-0129-5.

Nydam, R. L. and Cifelli, R. L. (2002) A new teiid lizard from the Cedar Mountain Formation (Albian–Cenomanian boundary) of Utah. *Journal of Vertebrate Paleontology*, 22(2), pp. 276–285. doi: 10.1671/0272-4634(2002)022.

Payros, A., Ortiz, S., Millán, I., Arostegi, J., Orue-Etxebarria, X. and Apellaniz, E. (2015) Early Eocene climatic optimum: environmental impact on the North Iberian continental margin. *Bulletin of the Geological Society of America*, 127(11–12), pp. 1632–1644. doi: 10.1130/B31278.1.

Pearson, M. R., Benson, R. B. J., Upchurch, P., Fröbisch, J. and Kammerer, C. F. (2013) Reconstructing the diversity of early terrestrial herbivorous tetrapods. *Palaeogeography, Palaeoclimatology, Palaeoecology*. Elsevier B.V., 372, pp. 42–49. doi: 10.1016/j.palaeo.2012.11.008.

Peralta-Medina, E. and Falcon-Lang, H. J. (2012) Cretaceous forest composition and productivity inferred from a global fossil wood database. *Geology*, 40(3), pp. 219–222. doi: 10.1130/G32733.1.

Percival, L. M. E., Ruhl, M., Hesselbo, S. P., Jenkyns, H. C., Mather, T. A. and Whiteside, J. H. (2017) Mercury evidence for pulsed volcanism during the end-Triassic mass extinction. *Proceedings of the National Academy of Sciences of the United States of America*, 114(30), pp. 7929–7934. doi: 10.1073/pnas.1705378114.

Perea, D., Soto, M., Sterli, J., Mesa, V., Toriño, P., Roland, G. and Da Silva, J. (2014) Tacuarembemys kusterae, gen. et sp. nov., a new Late Jurassic-? Earliest Cretaceous continental turtle from western Gondwana. *Journal of Vertebrate Paleontology*, 34(6), pp. 1329–1341. doi: 10.1080/02724634.2014.859620.

Pérez-García, A. (2012) High diversity of pancryptodiran turtles in the Lower Cretaceous of Europe. *Cretaceous Research*. Elsevier Ltd, 36, pp. 67–82. doi: 10.1016/j.cretres.2012.02.004.

Pérez-García, A. (2014) A new species of Brodiechelys (Testudines, Pan-Cryptodira) from the Early Cretaceous of Spain: systematic and palaeobiogeographic implications. *Acta Palaeontologica Polonica*, 59(2), pp. 333–342. doi: 10.4202/app.2012.0059.

Pérez-García, A. (2017) A new turtle taxon (Podocnemidoidea, Bothremydidae) reveals the oldest known dispersal event of the crown Pleurodira from Gondwana to Laurasia. *Journal of Systematic Palaeontology*. Taylor & Francis, 15(9), pp. 709–731. doi: 10.1080/14772019.2016.1228549.

Pérez-García, A., Espílez, E., Mampel, L. and Alcalá, L. (2015) A new European Albian turtle that extends the known stratigraphic range of the Pleurosternidae (Paracryptodira). *Cretaceous Research*, 55, pp. 74–83. doi: 10.1016/j.cretres.2015.02.007.

Peters, S. E. (2005) Geologic constraints on the macroevolutionary history of marine animals. *Proceedings of the National Academy of Sciences of the United States of*

America, 102(35), pp. 12326–12331. doi: 10.1073/pnas.0502616102.

Peters, S. E. and Heim, N. A. (2011) Macrostratigraphy and macroevolution in marine environments: testing the common-cause hypothesis. *Geological Society, London, Special Publications*, 358(1), pp. 95–104. doi: 10.1144/SP358.7.

Pianka, E. R. and Vitt, L. J. (2003) *Lizards: Windows to the Evolution of Diversity*. Berkeley and Los Angeles: University of California Press.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. and R Development Core Team (2017) nlme: linear and nonlinear mixed effects models. Available at: https://cran.r-project.org/package=nlme.

Pough, F. H., Janis, C. M. and Heiser, J. B. (2013) Turtles. in *Vertebrate Life*. 9th edn. New York: Pearson, pp. 287–308.

Powell, M. G. (2007) Latitudinal diversity gradients for brachiopod genera during late Palaeozoic time: links between climate, biogeography and evolutionary rates. *Global Ecology and Biogeography*, 16(4), pp. 519–528. doi: 10.1111/j.1466-8238.2007.00300.x.

Prasad, G. V. R. and Manhas, B. K. (2007) A new docodont mammal from the Jurassic Kota Formation of India. *Palaeontologia Electronica*, 10(2), p. 7A:11p.

Procheș, Ș. (2001) Back to the sea: secondary marine organisms from a biogeographical perspective. *Biological Journal of the Linnean Society*, 74(2), pp. 197–203. doi: 10.1006/bijl.2001.0565.

Prokoph, A., Shields, G. A. and Veizer, J. (2008) Compilation and time-series analysis of a marine carbonate δ18O, δ13C, 87Sr/86Sr and δ34S database through Earth history. *Earth-Science Reviews*, 87(3–4), pp. 113–133. doi: 10.1016/j.earscirev.2007.12.003.

Prothero, D. R. (1994) The late Eocene-Oligocene extinctions. *Annual Review of Earth and Planetary Sciences*, 22(1), pp. 145–165. doi: 10.1146/annurev.ea.22.050194.001045.

Pyron, R. A. (2010) A likelihood method for assessing molecular divergence time estimates and the placement of fossil calibrations. *Systematic Biology*, 59(2), pp. 185–194. doi: 10.1093/sysbio/syp090.

Pyron, R. A. (2017) Novel approaches for phylogenetic inference from morphological data and total-evidence dating in squamate reptiles (Lizards, Snakes, and Amphisbaenians). *Systematic Biology*, 66(1), pp. 38–56. doi: 10.1093/sysbio/syw068.

Rage, J.-C. and Augé, M. L. (1993) Squamates from the Cainozoic of the Western part of Europe. A review. *Revue de Paleobiologie*, 7, pp. 199–216.

Rage, J. C. (2012) Amphibians and squamates in the Eocene of Europe: what do they tell us? *Palaeobiodiversity and Palaeoenvironments*, 92, pp. 445–457. doi: 10.1007/s12549-012-0087-3.

Rage, J. C. (2013) Mesozoic and Cenozoic squamates of Europe. *Palaeobiodiversity and Palaeoenvironments*, 93, pp. 517–534. doi: 10.1007/s12549-013-0124-x.

Raup, D. M. (1972) Taxonomic diversity during the Phanerozoic. *Science*, 177(4054), pp. 1065–1071. doi: 10.1126/science.177.4054.1065.

Raup, D. M. (1976a) Species diversity in the Phanerozoic: a tabulation. *Paleobiology*, 2(4), pp. 279–288.

Raup, D. M. (1976b) Species diversity in the Phanerozoic: an interpretation. *Paleobiology*, 2(4), pp. 289–297.

Raup, D. M. and Sepkoski, J. J. (1982) Mass extinctions in the marine fossil record. *Science*, 215(4539), pp. 1501–1503. doi: 10.1126/science.215.4539.1501.

Rees, P. M., Noto, C. R., Parrish, J. M. and Parrish, J. T. (2004) Late Jurassic climates, vegetation, and dinosaur distributions. *The Journal of Geology*, 112(6), pp. 643–653.

Robeck, H., Maley, C. and Donoghue, M. (2000) Taxonomy and temperal diversity patterns. *Paleobiology*, 26(2), pp. 171–187. doi: 10.1666/0094-8373(2000)026<0171:TATDP>2.0.CO;2.

Robertson, D. S., Lewis, W. M., Sheehan, P. M. and Toon, O. B. (2013) K-Pg extinction patterns in marine and freshwater environments: the impact winter model. *Journal of Geophysical Research: Biogeosciences*, 118(3), pp. 1006–1014. doi: 10.1002/jgrg.20086.

Roll, U. et al. (2017) The global distribution of tetrapods reveals a need for targeted reptile conservation. *Nature Ecology and Evolution*. Springer US, 1(11), pp. 1677–1682. doi: 10.1038/s41559-017-0332-2.

Rose, P. J., Fox, D. L., Marcot, J. and Badgley, C. (2011) Flat latitudinal gradient in Paleocene mammal richness suggests decoupling of climate and biodiversity. *Geology*, 39(2), pp. 163–166. doi: 10.1130/G31099.1.

Sakamoto, M., Venditti, C. and Benton, M. J. (2017) 'Residual diversity estimates' do not correct for sampling bias in palaeodiversity data. *Methods in Ecology and Evolution*, 8(4), pp. 453–459. doi: 10.1111/2041-210X.12666.

Sanders, H. L. (1968) Marine benthic diversity: a comparative study. *The American Naturalist*, 102(925), pp. 243–282.

Seebacher, F. (2003) Dinosaur body temperatures: the occurrence of endothermy and ectothermy. *Paleobiology*, 29(1), pp. 105–122. doi: 10.1666/0094-8373(2003)029<0105:DBTTOO>2.0.CO;2.

Seilacher, A., Reif, W.-E. and Westphal, F. (1985) Sedimentological, ecological and temporal patterns of fossil Lagerstatten. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 311(1148), pp. 5–24. doi: 10.1098/rstb.1985.0134.

Sepkoski, Jr., J. J. (1993) Ten years in the library: new data confirm paleontological patterns. *Paleobiology*, 19(1), pp. 43–51.

Sheehan, P. M. and Hansen, T. A. (1986) Detritus feeding as a buffer to extinction at the end of the Cretaceous. *Geology*, 14(10), pp. 868–870. doi: 10.1130/0091-7613(1986)14<868:DFAABT>2.0.CO;2.

Shine, R. (1983) Reptilian reproductive modes: the oviparity-viviparity continuum. *Herpetologica*, 39(1), pp. 1–8.

Signor, P. W. and Lipps, J. H. (1982) Sampling bias, gradual extinction patterns and catastrophes in the fossil record. *Geological Society of America Special Papers*, 190, pp. 291–296.

Da Silva, F. O. et al. (2018) The ecological origins of snakes as revealed by skull evolution. *Nature Communications*, 9(1), pp. 1–11. doi: 10.1038/s41467-017-02788-3.

Simões, T. R. et al. (2018) The origin of squamates revealed by a Middle Triassic lizard from the Italian Alps. *Nature*, 557(7707), pp. 706–709. doi: 10.1038/s41586-018-0093-3.

Smith, A. B. (2001) Large-scale heterogeneity of the fossil record: implications for Phanerozoic biodiversity studies. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 356(1407), pp. 351–67. doi: 10.1098/rstb.2000.0768.

Smith, A. B. (2007) Intrinsic versus extrinsic biases in the fossil record: contrasting the fossil record of echinoids in the Triassic and early Jurassic using sampling data, phylogenetic analysis, and molecular clocks. *Paleobiology*, 33(2), pp. 310–323. doi: 10.1666/06073.1.

Smith, A. B. and Benson, R. B. J. (2013) Marine diversity in the geological record and its relationship to surviving bedrock area, lithofacies diversity, and original marine shelf area. *Geology*, 41, pp. 171–174. doi: 10.1130/G33773.1.

Smith, A. B., Lloyd, G. T. and McGowan, A. J. (2012) Phanerozoic marine diversity: rock record modelling provides an independent test of large-scale trends. *Proceedings of the Royal Society B: Biological Sciences*, 279(1746), pp. 4489–4495. doi: 10.1098/rspb.2012.1793.

Smith, A. B. and McGowan, A. J. (2007) The shape of the Phanerozoic marine palaeodiversity curve: how much can be predicted from the sedimentary rock record of western Europe? *Palaeontology*, 50(4), pp. 765–774. doi: 10.1111/j.1475-4983.2007.00693.x.

Smith, A. G., Smith, D. G. and Funnell, B. M. (2004) *Atlas of Mesozoic and Cenozoic Coastlines*. Cambridge: Cambridge University Press.

Smith, K. T. (2006) A diverse new assemblage of late Eocene squamates (Reptilia) from the Chadron Formation of North Dakota, U.S.A. *Palaeontologia Electronica*, 9(2), pp. 1–44.

Smith, K. T. (2009) A new lizard assemblage from the earliest Eocene (Zone Wa0) of the Bighorn Basin, Wyoming, USA: biogeography during the warmest interval of the Cenozoic. *Journal of Systematic Palaeontology*, 7(3), pp. 299–358. doi: 10.1017/S1477201909002752.

Smith, K. T. (2011) The evolution of mid-latitude faunas during the Eocene: late

Eocene lizards of the Medicine Pole Hills reconsidered. *Bulletin of the Peabody Museum of Natural History*, 52(1), p. 2011. doi: 10.3374/014.052.0101.

Smith, K. T. and Wuttke, M. (2012) From tree to shining sea: taphonomy of the arboreal lizard Geiseltaliellus maarius from Messel, Germany. *Palaeobiodiversity and Palaeoenvironments*, 92(1), pp. 45–65. doi: 10.1007/s12549-011-0064-2.

Spiess, A. N. and Ritz, C. (2014) qpcR: modelling and analysis of real-time PCR data. Available at: http://cran.r-project.org/package=qpcR.

Starrfelt, J. and Liow, L. H. (2016) How many dinosaur species were there? Fossil bias and true richness estimated using a Poisson sampling model. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371, p. 20150219. doi: 10.1017/CBO9781107415324.004.

Stehlin, H. G. (1909) Remarques sur les faunules de mammifères des couches éocènes et oligocènes du Bassin de Paris. *Bulletin de la Societe Geologique de France*, 9, pp. 488–520.

Sterli, J. (2008) A new, nearly complete stem turtle from the Jurassic of South America with implications for turtle evolution. *Biology Letters*, 4(3), pp. 286–289. doi: 10.1098/rsbl.2008.0022.

Sues, H.-D. and Hopson, J. A. (2010) Anatomy and phylogenetic relationships of Boreogomphodon jeffersoni (Cynodontia: Gomphodontia) from the Upper Triassic of Virginia. *Journal of Vertebrate Paleontology*, 30(4), pp. 1202–1220. doi: 10.1080/02724634.2010.483545.

Sues, H. D. and Olsen, P. E. (1990) Triassic vertebrates of Gondwanan aspect from the Richmond Basin of Virginia. *Science*, 249(4972), pp. 1020–1023. doi:

10.1126/science.249.4972.1020.

Sullivan, R. M. (1985) A new middle Paleocene (Torrejonian) rhineurid amphisbaenian, Plesiorhineura tsentasi new genus, new species, from the San Juan Basin, New Mexico. *Journal of Paleontology*, 59(6), pp. 1481–1485.

Sun, J., Ni, X., Bi, S., Wu, W., Ye, J., Meng, J. and Windley, B. F. (2014) Synchronous turnover of flora, fauna, and climate at the Eocene–Oligocene Boundary in Asia. *Scientific Reports*, 4, p. 7463. doi: 10.1038/srep07463.

Tennant, J. P., Mannion, P. D. and Upchurch, P. (2016) Environmental drivers of crocodyliform extinction across the Jurassic/Cretaceous transition. *Proceedings of the Royal Society B: Biological Sciences*, 283, p. 20152840. doi: 10.1098/rspb.2015.2840.

Tennant, J. P., Mannion, P. D. and Upchurch, P. (2016) Sea level regulated tetrapod diversity dynamics through the Jurassic/Cretaceous interval. *Nature Communications*. Nature Publishing Group, 7, p. 12737. doi: 10.1038/ncomms12737.

Thomas, E. and Gooday, A. J. (1996) Cenozoic deep-sea benthic foraminifers: tracers for changes in oceanic productivity? *Geology*, 24(4), pp. 355–358. doi: 10.1130/0091-7613(1996)024<0355:CDSBFT>2.3.CO;2.

Toljagic, O. and Butler, R. J. (2013) Triassic–Jurassic mass extinction as trigger for the Mesozoic radiation of crocodylomorphs. *Biology Letters*, 9(March), pp. 1–4.

Tong, H., Claude, J., Li, C.-S., Yang, J. and Smith, T. (2017) Wutuchelys eocenica n. gen. n. sp., an Eocene stem testudinoid turtle from Wutu, Shandong Province, China. *Geological Magazine*, pp. 1–14. doi: 10.1017/S0016756817000905.

Trapletti, A. and Hornik, K. (2017) tseries: time series analysis and computational finance. Available at: https://cran.r-project.org/package=tseries.

Tutin, S. L. and Butler, R. J. (2017) The completeness of the fossil record of plesiosaurs, marine reptiles from the Mesozoic. *Acta Palaeontologica Polonica*, 62(3), pp. 563–573. doi: 10.4202/app.00355.2017.The.

Uetz, P., Freed, P. and Hošek, J. (2018) *The Reptile Database*. Available at: http://www.reptile-database.org (Accessed: 29 August 2018).

Urban, M. C., Richardson, J. L. and Freidenfelds, N. A. (2014) Plasticity and genetic adaptation mediate amphibian and reptile responses to climate change. *Evolutionary Applications*, 7, pp. 88–103. doi: 10.1111/eva.12114.

Verrière, A., Brocklehurst, N. and Fröbisch, J. (2016) Assessing the completeness of the fossil record: comparison of different methods applied to parareptilian tetrapods (Vertebrata: Sauropsida). *Paleobiology*, 42(4), pp. 680–695. doi: 10.1017/pab.2016.26.

Vidal, N. and Hedges, S. B. (2004) Molecular evidence for a terrestrial origin of snakes. *Proceedings of the Royal Society B: Biological Sciences*, 271, pp. S226–S229. doi: 10.1098/rsbl.2003.0151.

Vlachos, E. (2018) A Review of the Fossil Record of North American Turtles of the Clade *Pan-Testudinoidea*. *Bulletin of the Peabody Museum of Natural History*, 59(1), pp. 3–94. doi: 10.3374/014.058.0201.

Waterson, A. M., Schmidt, D. N., Valdes, P. J., Holroyd, P. A., Nicholson, D. B., Farnsworth, A. and Barrett, P. M. (2016) Modelling the climatic niche of turtles: a deep-time perspective. *Proceedings of the Royal Society B: Biological Sciences*, 283(1839), p. 1839. doi: 10.1098/rspb.2016.1408.

Weems, R. E. and Knight, J. L. (2012) A new species of Bairdemys (Pelomedusoides: Podocnemididae) from the Oligocene (Early Chattian) Chandler Bridge Formation of

South Carolina, USA, and its paleobiogeographic implications for the genus. in Brinkman, D. B., Holroyd, P. A., and Gardner, J. D. (eds) *Morphology and Evolution of Turtles*. Dordrecht: Springer, pp. 289–303. doi: 10.1007/978-94-007-4309-0_18.

Werner, J. and Griebeler, E. M. (2014) Allometries of maximum growth rate versus body mass at maximum growth indicate that non-avian dinosaurs had growth rates typical of fast growing ectothermic sauropsids. *PLoS ONE*, 9(2). doi: 10.1371/journal.pone.0088834.

Whiteside, D. I., Duffin, C. J., Gill, P. G., Marshall, J. E. A. and Benton, M. J. (2016) The late Triassic and early Jurassic fissure faunas from Bristol and South Wales: stratigraphy and setting. *Palaeontologia Polonica*, 67(February), pp. 257–287. doi: 10.4202/pp.2016.67_257.

Wiens, J. J. (2007) Global patterns of diversification and species richness in amphibians. *The American Naturalist*, 170(2), p. S86. doi: 10.2307/4541092.

Wiens, J. J., Brandley, M. C. and Reeder, T. W. (2006) Why does a trait evolve multiple times within a clade? Repeated evolution of snakelike body form in squamate reptiles. *Evolution*, 60(1), pp. 123–141. doi: 10.1111/j.0014-3820.2006.tb01088.x.

Wiens, J. J. and Lambert, S. M. (2014) The phylogeny of lizard families. in Rheubert, J. L., Siegel, D. S., and Trauth, S. E. (eds) *Reproductive Biology and Phylogeny of Lizards and Tuatara*. CRC Press, pp. 27–42.

Willig, M. R., Kaufman, D. M. and Stevens, R. D. (2003) Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. *Annual Review of Ecology, Evolution, and Systematics*, 34(1), pp. 273–309. doi: 10.1146/annurev.ecolsys.34.012103.144032.

Wilson, G. P. (2013) Mammals across the K/Pg boundary in northeastern Montana, USA: dental morphology and body-size patterns reveal extinction selectivity and immigrant-fueled ecospace filling. *Paleobiology*, 39(3), pp. 429–469. doi: 10.1666/12041.

Woodburne, M. O., Gunnell, G. F. and Stucky, R. K. (2009) Climate directly influences Eocene mammal faunal dynamics in North America. *Proceedings of the National Academy of Sciences of the United States of America*, 106(32), pp. 13399–13403. doi: 10.1073/pnas.0906802106.

Wright, N., Zahirovic, S., Müller, R. D. and Seton, M. (2013) Towards community-driven paleogeographic reconstructions: integrating open-access paleogeographic and paleobiology data with plate tectonics. *Biogeosciences*, 10(3), pp. 1529–1541. doi: 10.5194/bg-10-1529-2013.

Wu, X.-C. (1994) Late Triassic-Early Jurassic sphenodontians from China and the phylogeny of the Sphenodontida. in Fraser, N. C. and Sues, H.-D. (eds) *In the Shadow of the Dinosaurs: Early Mesozoic Tetrapods*. Cambridge: Cambridge University Press, pp. 38–69.

Zachos, J. C., Dickens, G. R. and Zeebe, R. E. (2008) An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature*, 451(7176), pp. 279–283. doi: 10.1038/nature06588.

Zachos, J., Pagani, M., Sloan, L., Thomas, E. and Billups, K. (2001) Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science*, 292(5517), pp. 686–693. doi: 10.1126/science.1059412.

Zheng, Y. and Wiens, J. J. (2016) Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. *Molecular Phylogenetics and Evolution*. Elsevier Inc., 94, pp. 537–547. doi: 10.1016/j.ympev.2015.10.009.

Zhou, Z. (2014) The Jehol Biota, an Early Cretaceous terrestrial Lagerstätte: New discoveries and implications. *National Science Review*, 1(4), pp. 543–559. doi: 10.1093/nsr/nwu055.