Phylogenetic Interpretations of Macroevolution in Deep-Time

Laura Clare Soul

Thesis submitted to the University of Oxford
for the degree of Doctor of Philosophy
in Earth Sciences

Worcester College
University of Oxford
Trinity, 2015

Supervised by Dr. Matt Friedman

Declaration

This thesis is entirely my own work except where otherwise indicated and has not been submitted, either wholly or substantially, for another degree of this University or any other institution.

Acknowledgements

Thanks should go first and foremost to Matt Friedman for his patience, enthusiasm and heroic level of availability and engagement through teaching, travelling, researching and major life events. Thanks to Matt particularly for understanding (encouraging even) the inordinate amount of time I spent doing ‘other things’, it paid off in the end!

I owe Roger Benson a great deal; his encouragement and confidence in my abilities were a large factor in my doing a PhD in the first place, and since then his support and insight have been invaluable.

Graeme Lloyd answered endless questions, resolved confusions, and put up with me during the last weeks of writing up. He and the rest of the Oxford Palaeobiology lab have been an ideal group of people to work with; smart, friendly and always willing to help. It’s been great to see the group grow during my time as a member.

Thanks to many friendly faces at museums; Eliza Howlett, Matt Riley, Erin Maxwell, Mark Evans, Kate Sherburn, David Gelsthorpe and Davit Vasilyan for access to specimens.

I’ve had a great time at Oxford, which has been down to the people I have been lucky enough to spend my time with here. Thanks to PG11 and Worcester MCR, as well as all my friends from other places, for collectively providing endless coffee, gossip, misadventures, stress relief and generally more fun than I ought to have had.

Thanks to Frances, Michael and Jonathan Soul for their love and support. For having the patience to answer upwards of 18 years of questions, all the tea table science, and for making sure that I knew I could do whatever I liked the most.

Thank you finally and most of all to Chris.

Sic transimus, cum mundi revera perpetuus.

Phylogenetic Interpretations of Macroevolution in Deep-Time

Laura Soul

Worcester College

Trinity 2015

The fossil record yields information on macroevolutionary patterns that remains inaccessible from the study of extant organisms alone, presenting a natural laboratory for us to test hypotheses about the long-term drivers and processes of evolution. Fossil data are therefore increasingly incorporated into evolutionary analyses, both on their own and in combination with neontological data. Phylogeny (an explicit hypothesis of the evolutionary relationships between taxa) can be used as a framework to enable direct comparison of results of comparative methods across many different timescales and taxa, and is now commonly used in investigations of fossil data. This represents an important step towards a unified approach, however, it is not yet fully understood what the effect of using fossil data is on the results of downstream phylogenetic comparative methods, which were originally developed with only living taxa in mind. In this thesis I explore the validity of phylogenetic interpretations of fossil record data. I begin with only taxonomic classification and show that this can in some cases substitute for a cladistically inferred phylogeny in phylogenetic comparative methods, without biasing results. Moving on to scenarios where a timescaled phylogeny is available I investigate the relationship between phylogeny and extinction in the geological past, show that phylogenetic clustering of extinction was common in tetrapods, and present a summary of the ways in which fossil data biases this measurement. Finally, with timescaled phylogenies and a detailed continuous trait dataset available, I interrogate the fossil record of Sauropterygia to uncover the processes of evolutionary change in this highly labile clade. By comparing the results of a suite of phylogenetic comparative methods I demonstrate that neck length evolved through changing vertebral counts rather than somite growth; that the clade experienced a release in evolutionary constraint at the Triassic-Jurassic boundary; and that evidence does not support evolution towards a stationary adaptive peak as a suitable model for phenotypic change in the clade.

Extended Abstract

Phylogenetic comparative methods are a powerful tool for dissecting macroevolutionary patterns on many different timescales and across the tree of life. Their use in evolutionary biology is now extremely widespread. Palaeobiologists are following suit and there are now a wide range of these methods that can be used to critically interrogate a variety of fossil records. However, there are important differences between phylogenies of living and fossil taxa. Although some of the hurdles (such as incomplete sampling, presence of ancestor-descendent pairs, topological uncertainty and error in estimation of divergence dates) are all relevant to modern datasets, these issues are particularly acute when dealing with fossil data. In this thesis, I examine some of the current possibilities provided by phylogenetic comparative methods to quantitatively assess macroevolutionary patterns in completely extinct clades. Fossil records vary significantly in their quality of preservation and resolution through time and across clades. There is therefore no 'one size fits all' approach for palaeobiologists, but with careful preparation of datasets phylogeny can provide a way in which the problem of incomplete sampling can be statistically overcome, to facilitate detection of macroevolutionary patterns in the fossil record.

Taxonomy and phylogeny

Many extinct taxa with an extensive fossil record and a mature taxonomic classification have not yet been the subject of formal phylogenetic analysis. I test whether the taxonomies available for such groups represent useful (i.e., non-misleading) substitutes for trees derived from matrix-based phylogenetic analyses. I collected data for 52 animal clades that include fossil representatives, and for which a recent cladogram and pre-cladistic taxonomy were available. I quantified the difference between the time-scaled phylogenies implied by taxonomies and cladograms using the matching cluster distance metric. I simulated phenotypic trait values and used them to estimate a series

of commonly used, phylogenetically explicit measures (phylogenetic signal [Blomberg's K], phylogenetic generalised least squares [PGLS], mode of evolution [Brownian versus Ornstein-Uhlenbeck] and phylogenetic clustering of extinction [Fritz and Purvis' D]) in order to determine the degree to which they co-varied on taxonomic and cladistic trees.

With respect to topology and branch length, taxonomies are good approximations of the underlying evolutionary relationships as recorded in inferred cladograms. Detection of phylogenetic structuring of extinction could not be properly assessed. For all other evolutionary analyses, results from taxonomy-based phylogenies co-varied with those from cladogram-based phylogenies, but individual comparisons could be misleading. The relative length of terminal branches (influenced by sampling rate) is a key control on the shared information between, and therefore the relative performance of, taxonomy and cladogram-based phylogenies. Collectively these results suggest that under particular circumstances and after careful consideration, some taxonomies could be used in place of a formal analytical solution, but users must be cautious. This opens parts of a large and previously inaccessible section of the fossil record to interrogation within an explicit comparative framework, which will help to test many classical macroevolutionary hypotheses formulated for groups that currently lack formal phylogenetic estimates.

Phylogenetic clustering of extinction

Biological correlates of extinction risk are of great interest to evolutionary biologists as they attempt to develop predictive models of species extinction and risk. It is possible to analyse the relationship between extinction risk and various traits in groups of extant species, but the fossil record presents an opportunity to measure actual extinctions and investigate the effects of environmental scenarios which have only previously occurred on geological (rather than human) timescales. Unfortunately, many of the traits that have been shown to be good predictors of extinction risk frequently covary, and may not fossilise (e.g. narrow ecological tolerance, limited dispersal ability). One way around this issue is to measure the phylogenetic structure of extinction itself, by treating it as a binary trait.

Extinction is not an evolving trait but can be modelled as such based on the assumption that it depends upon a series of underlying continuously evolving traits.

I use two methods of measuring the phylogenetic clustering of extinction (Moran's I and Fritz and Purvis' D) to investigate the relationship between phylogeny and extinction in three large clades of tetrapods across the Permo-Triassic mass extinction boundary. I vary tree topology, timeslice length and method of timescaling trees to test the robustness of results to variations in the input trees.

For part of the period of time under investigation extinctions in each of the three clades were phylogenetically clustered, but not consistently so. This result is largely robust to variation in input trees, but the method of timescaling can have important effects on results. Other aspects of the evolutionary history of the clades such as variation in species richness through time, and the ecological impact of extinction, seem to be decoupled from phylogenetic selectivity. These results support the idea that future extinctions will be non-random with respect to phylogeny and therefore cause a higher than expected loss in biodiversity.

Simulating phylogenetic patterns in extinction

Some characteristics of phylogenies violate the assumptions of phylogenetic comparative methods, for example the inclusion of ancestor-descendant pairs as sister taxa in a tree, or incomplete or non-random sampling. Although these problems are present in almost any phylogeny, they are more acute in phylogenies of fossil taxa where long timescales amplify the effect. Many phylogenetic comparative methods are model based, and as such strong macroevolutionary trends should be detectable even from incomplete data. Macroevolutionary patterns in extinction however, are perhaps more vulnerable to bias caused by factors such as topological misspecification or poor phylogenetic resolution. Correcting the data for these issues before analyses is not possible, and so it is important to understand the effect they are likely to have on results.

With this in mind, I use a process-based birth-death model to simulate evolutionary histories of clades, including mass extinction events. I use these simulated histories to construct ’true’ phylogenies and measure the phylogenetic clustering of extinction on these true trees. I then sample the evolutionary histories at different rates to represent fossil occurrences and use these to reconstruct palaeontological phylogenies or ‘fossil’ trees. I measure the phylogenetic clustering of extinction on the fossil trees and make a comparison with the results from the true tree to infer the presence or direction of bias under different evolutionary and analytical scenarios.

I show that the nature of input data, as well as subsequent data treatment, have important effects on the outcome of extinction clustering analyses and that the sampling rate of the clade is the most important factor. At low sampling rates the *Hedman* method of timescaling is the most effective for making accurate estimates of measures of phylogenetic structure in extinction, but at higher rates the *cal3* method is preferable. At low sampling rates the analyses is biased towards underestimating the strength of clustering, but the opposite is true at high sampling rates. These biases are likely caused by a combination of branch length extensions back in time across boundaries (‘pseudosurvivals’) and pseudoextinctions. In some cases these effects balance one another out, in other scenarios one dominates, leading to the bias in estimation.

Macroevolution in Sauropterygians

Sauropterygia was a highly successful clade of marine reptiles that survived throughout the whole of the Mesozoic. They exhibit a broad range of body plans, including a large variation in vertebral counts, neck to trunk and skull proportions, and body size varying from small semi-aquatic taxa with a hard carapace to the enormous elasmosaurids with three meter long necks. The generating processes of such variety are of great interest, and there have been many qualitative theories about constraints or drivers of evolution in the clade. As such they represent an ideal opportunity to explore the possibilities of using modern phylogenetic comparative methods to rigorously assess macroevolutionary patterns and mechanisms.

I present a dataset of axial body plan, vertebral count and limb length measurements from 120 taxa representing all the major divisions within the clade. I use this dataset to test a series of explicit hypotheses about sauropterygian evolution: 1) Neck length evolved via somitogenetic (change in vertebral number) and homeotic (change in axial regionalisation) rather than by post-embryonic differential somitic growth. 2) Somitogenetic effects were more important as a generating mechanism of axial body plan change than homeotic effects. 3) Somitogenetic and homeotic effects were decoupled during sauropterygian evolution. 4) Sauropterygians underwent a release in evolutionary constraint on body proportions after the Triassic-Jurassic boundary, coinciding with a transition to open water habits. 5) The pliosauromorph body plan (long skull, short neck, long and large body) was an adaptive peak in the evolutionary landscape towards which members of different clades converged through time. To do this, I first compare the rates of evolution of presacral vertebral count to rates of evolution in the ratio of cervical to dorsal vertebrae using a Bayesian approach appropriate for incompletely sampled non-ultrametric trees. I then apply a maximum likelihood modelling approach that requires *a priori* specification of regimes for tip taxa, and estimates model support for a variety of evolutionary models. These include Brownian motion and Ornstein-Uhlenbeck models with one or no shifts in evolutionary regime. Finally, I apply a further modelling approach, which can be used to test for convergence to assess whether the pliosauromorph body plan represents an evolutionary peak.

My results indicate that rates of change in presacral vertebral count were high across most of the group, but slowed down in particular nested clades. Overall, homeotic effects dominated the change in sauropterygian axial body plan through time. Somitic growth was not an important factor and somitogenesis seemed mainly to be involved in generating trait change at the extremes of trait space (e.g. the extremely long necks in derived elasmosaurids), although this occurred over long timescales. I show that a model including a change in functional constraint on limb proportions at the Jr-Tr boundary is consistent with the data. In addition, results suggest that neck length ratios were under low selective constraint throughout the evolutionary history of Sauropterygia, but the transition to open water increased the area of morphospace available for this trait. Results from analysis of trait optima suggest that some taxa with a pliosauromorph body plan did converge upon a similar area of morphospace, but do not provide strong evidence for an adaptive peak towards which all pliosauromorph taxa were converging.

Contents

List of Figures

List of Tables

Chapter 1

Phylogenetic comparative methods and palaeontological data

"Three billion years the Tree has grown
From replicators' first seed sown
To branches rich with progeny
The wonder of phylogeny"

- David R. Maddison *The Tree of Life*

The fossil record yields information on macroevolutionary patterns that remains inaccessible from the study of extant organisms alone (Smith and Turner, 2005; Waller, 2006; Flynn et al., 2010). Palaeontological data expand our understanding of evolution beyond the 'snapshot' of the present day to encompass long-term trends and processes that have acted over the history of life on Earth (Sepkoski et al., 1981; Raup and Sepkoski, 1984; Jablonski, 2007; Slater and Harmon, 2013; Wagner and Marcot, 2013). Early observations of patterns and apparent trends in the fossil record (Cope, 1887; Osborn, 1917; Simpson, 1944) inspired development of many of the phylogenetic comparative methods presently used to quantitatively assess evolutionary hypotheses. However, the degree to which some of these methods are applicable to palaeontological data has not yet been rigorously assessed.

Here I present an overview of what estimated phylogeny (that might be used in an analysis) means in the context of the fossil record. I discuss the comparative methods available to investigate patterns in evolution and extinction under a phylogenetic framework, and specifically outline the methods available to timescale cladograms of fossil taxa, as these are integral to the rest of the thesis.

1.1 Phylogenies of fossil taxa

1.1.1 Definition of a palaeontological phylogeny

It has long been noted that the fossil record is incomplete (Darwin, 1859; Raup, 1972). More recently it has been demonstrated that it is not "hopelessly so" (Gauthier et al., 1988; Benton and Storrs, 1994) in terms of using it to understand evolutionary patterns, but it is clear that fossil records will be missing a large amount of information, and sampling intensity will be non-random due to the special sets of circumstances required for fossil preservation. Figure 1.1 shows a hypothetical evolutionary history of a clade, and what we might expect to have sampled for a good record; many records would be less complete even than this (Foote and Raup, 1996).

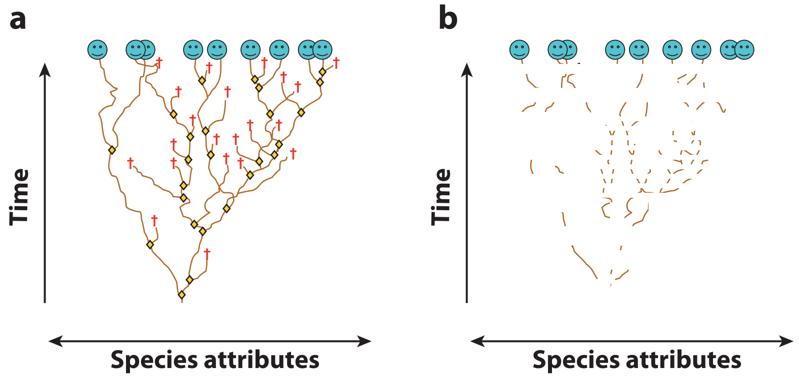


Figure 1.1: A clade evolving through time. a) shows the true evolutionary history of the clade, complete with speciation and extinction ages and full knowledge of phenotype and relationships. b) shows the rather degraded picture of what we might be able to sample for a good record. Figure adapted from Purvis (2008).

Historically there have been two main schools of thought on how to accommodate sampling problems to make evolutionary inferences. One is to use numbers of taxa and geological information to try and correct for incompleteness, the other is to reconstruct phylogenetic relationships of the taxa that are available (Wagner, 2000b). Due to the increasingly pervasive use of phylogenetic comparative methods in evolutionary biology, and a move towards unified approaches across biological and paleobiological disciplines, phylogenetic approaches to analysing fossil data are beginning to dominate over taxic approaches. Phylogeny is an important tool in correcting for the evolutionary history of taxa when modelling trait data (Felsenstein, 1985; Revell, 2010; Freckleton et al., 2011), but can also be used in its own right as positive information to test hypotheses about trait evolution (Mooers et al., 1999; Pennell and Harmon, 2013). By inferring the evolutionary relationships between taxa, unsampled lineages can also be hypothesised (Lane et al., 2005), and model based evolutionary inference can be performed on incomplete trees (Eastman et al., 2011).

As used in this thesis, the term phylogeny refers to an explicit hypothesis of the evolutionary relationships between a set of taxonomic units, represented by a branching network. The terminations of branches represent instantaneous populations of different taxa (for fossil data these are usually at genera or family level), and the lengths of edges are scaled to correspond to some measure of evolutionary distance. The edge length between a pair of tips represents their evolutionary divergence, and the last node which subtends the two can be considered to be a hypothetical taxonomic unit representing their last common ancestor. They are therefore a description of the amount of shared and unshared evolutionary history between the included taxa.

If the phylogeny is rooted this root node defines the branching order of taxa; when the edge lengths are then scaled to time it represents the timing of the origin of the clade. The branch length between the root node and a tip is proportional to the evolutionary variance of the taxon at that tip; the branch length shared between two taxa is proportional to the evolutionary covariance of those taxa. In phylogenies of living taxa all the tips are contemporaneous, ending at the present day, forming an ultrametric tree. Phylogenies which include extinct taxa as tips are non-ultrametric: the branches terminate at different times. This means that in a phylogeny of extant taxa the covariance of taxon A with its sister taxon B, is equal to the covariance of B with A. Conversely in a non-ultrametic tree the covariance of A with B can be different to that of B with A because they may have different branch lengths between the speciation event that separated them, and their tips. This has important implications for the way that some phylogenetic comparative methods (particularly those involving branch length transformations) are performed on non-ultrametric trees (Slater, 2014).

The phylogenetic variance-covariance structure is integral to many phylogenetic comparative methods (PCMs). These methods therefore require a tree with robust branch lengths, as these are a key control on this structure. This has previously been a major hurdle in the use of palaeontological phylogenies in PCMs because their topologies are constructed using morphology based cladistic inference, and so have to be scaled to time following the estimation of relationships. Often there is a relatively poor constraint on the ages of fossil specimens, in some cases they are only known to stage or epoch level (e.g. Chapter 3). Developing methods of estimating robust branch lengths for trees of fossil taxa has therefore been a recent focus of palaeobiology. Fortunately several have now been developed (outlined in section 1.4), unfortunately their comparative merits remain relatively unexplored.

1.1.2 Ancestors in samples of fossil taxa

Phylogenies are meant to represent the branching order of monophyletic clades of sister taxa. This is not possible for phylogenies of fossil taxa as many extinct species will not have been sampled and there is a high probability that sampled taxa are ancestral to one another (Foote, 1996b). Although sampling of ancestral populations occurs in phylogenies of living taxa (Funk and Omland, 2003; Ross, 2014), the longer timescales in clades of fossil taxa make this a particular problem, especially when estimating divergence dates

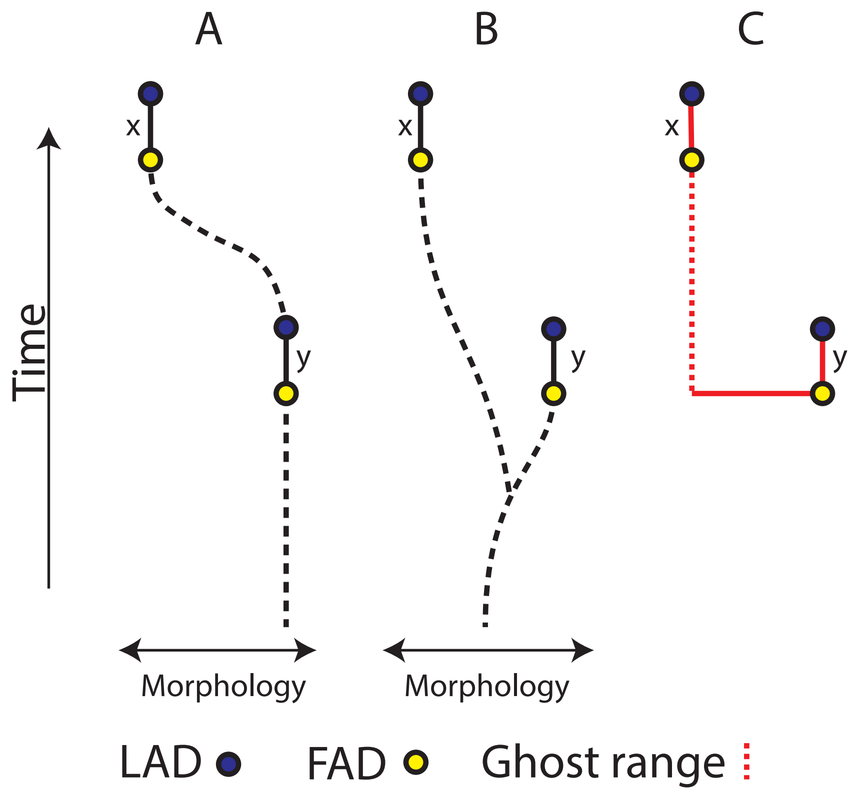


Figure 1.2: Inferred phylogenetic trees are ambiguous with respect to the issue of sampled ancestors. Morphotaxa x and y are based on fossil samples. C shows how the relationship between these two taxa would be reconstructed by a cladistic analysis and scaled to time. A ghost range is inferred between the first appearance of the younger morphotaxon and the first appearance of the older. They are reconstructed as sister clades based on their close, but not identical, morphology. A and B show two possible ‘real’ scenarios that would both lead to the phylogenetic reconstruction shown in C. In A morphotaxon y undergoes a shift in phenotype in the lineage through time to become morphotaxon x. In scenario B there is a speciation event giving rise to taxon x and both taxa undergo a shift in phenotype. These are indistinguishable when all we have is the information in C.

and patterns in extinction.

Figure 1.2 shows the classic scenario in which sampling of an ancestor-descendant pair confounds estimation of evolutionary relationships. Part C shows the inferred sister-clade relationship, including an unsampled ghost range for taxon x, that would be reconstructed from either of the real evolutionary histories A and B. It is not possible to distinguish between A and B if all we have available is C. This is clearly a problem for PCMs. Even if scenario A were true, current approaches to phylogenies in the fossil record would assume scenario B. This inaccurate scenario brings with it two principle features at odds with genuine evolutionary history: extinction of taxon y, and a more extensive ghost range between this taxon and x. If this were common across a whole phylogeny of fossil taxa it is easy to see that by distorting branch lengths and speciation and extinction times this could bias, or at least introduce noise to, the results of PCMs.

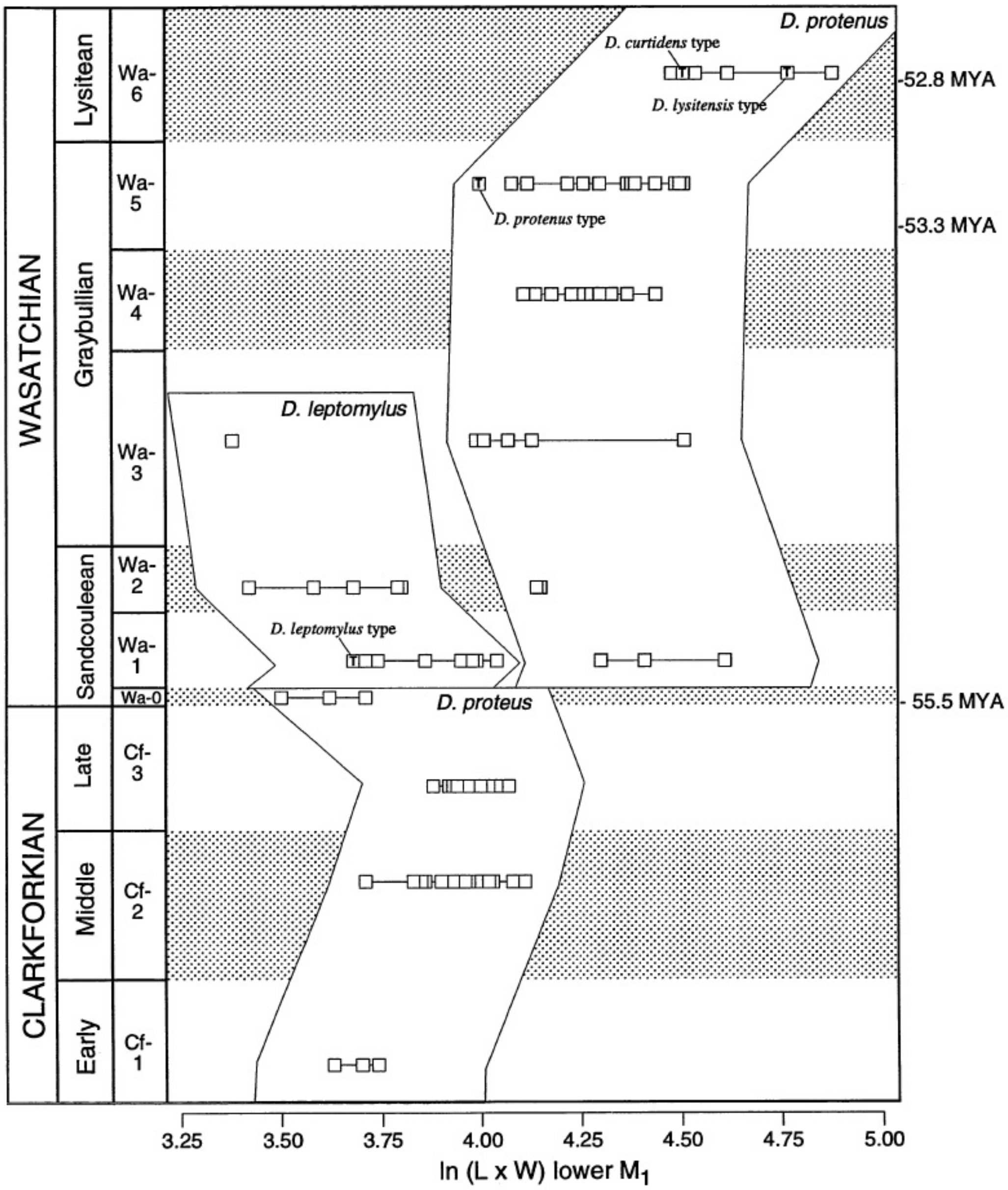


Figure 1.3: A stratocladistic phylogeny from Polly (1997). This was constructed by initially performing a normal cladistic analysis of a character dataset for these three taxa. The relationships inferred in this analysis are then combined with stratigraphic data and through this possible lineages are inferred. The x-axis shows measurements of a trait associated with the first molar and the y-axis is stratigraphic position - a proxy for time. Successive rows of data points within a white band belong to the same species and are thought to be from populations ancestral to one another. There was a speciation event near the beginning of the Sandcouleean. Stratophenetic methods follow a similar process but use clustering techniques on trait data to establish relationships, rather than cladistic inference.

Methods via which this problem could be resolved have been proposed by several authors (Dayrat, 2005). These use stratigraphic information to reconstruct the ancestor-descendant relationships. Either through stratocladistics (e.g. Polly, 1997; Fisher et al., 2002, Fig. 1.3) or stratophenetics (e.g. Roopnarine, 2005). Such trees have been constructed and used to make inferences about speciation and evolutionary rates (e.g. Aze et al., 2011) for some high-resolution records. A further method that can be used on cladistically inferred trees relies on the assumption that a polytomy in a cladistically inferred tree of fossil taxa indicates an ancestor, and selects this from the polytomy using stratigraphic information (Wagner, 2000a; Bapst, 2013). Still newer methods allow ancestral relationships during estimation of topology and branching order during Bayesian inference of cladistic character datasets (Gavryushkina et al., 2014).

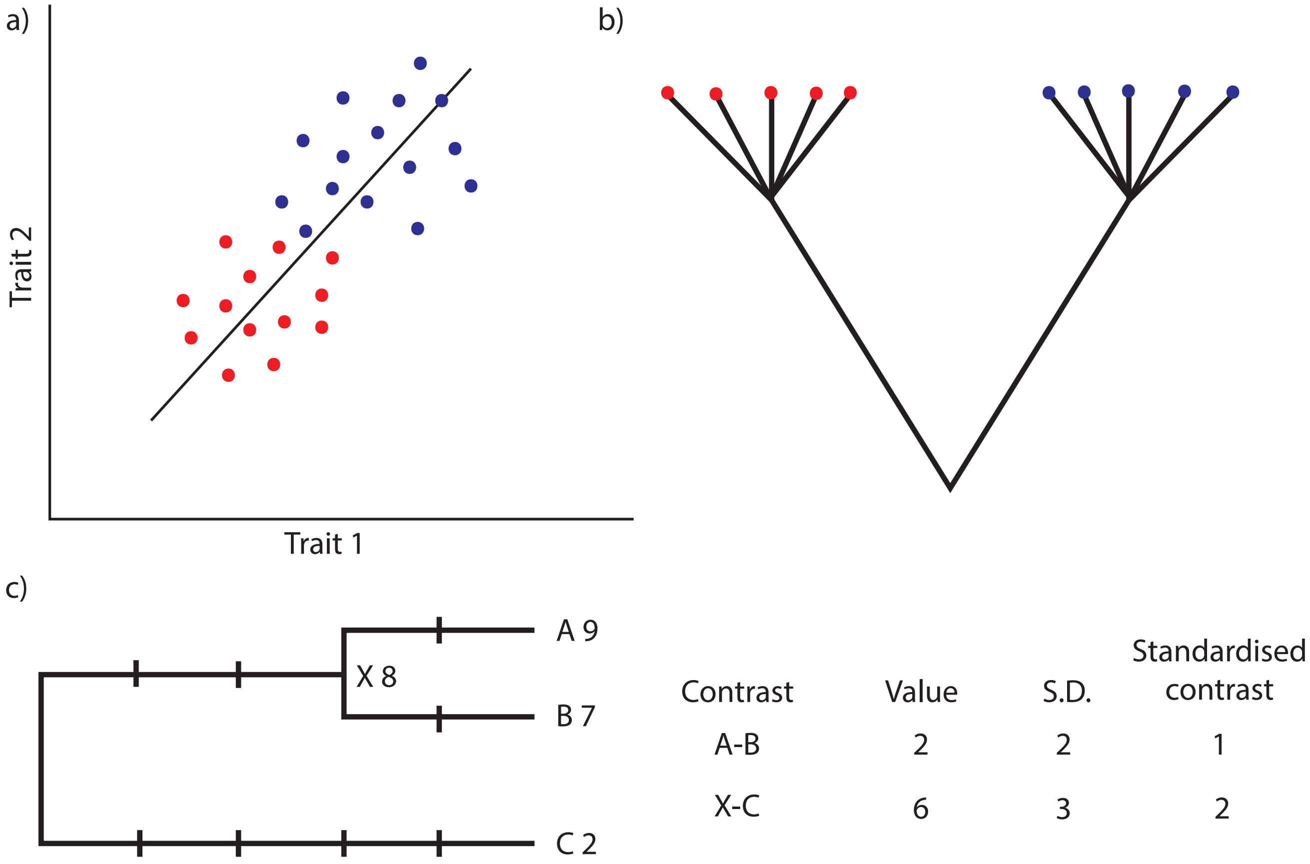
Although these methods are a step in the right direction, and could be effective with high-resolution records, they are not problem free. This is most obvious for methods that treat polytomies in inferred trees as the ‘ghosts’ of ancestor-descendant pairs. The way in which morphology based cladistics of fossil specimens is performed means that there are a wide variety of reasons, other than a sampled ancestor, that unresolved relationships could be found. For example cladistic character datasets usually include many more synapomorphies than apomorphies as these are what are used for inference. This stems from the fact that autapomorphies - features present in only a single terminal taxon - are uninformative in parsimony-based phylogenetic inference, the dominant means of estimating trees using morphology. As such, these characters tend to be excluded from matrices of anatomical traits. Unfortunately, apomorphies are what would distinguish a descendant taxa from a genuine sister taxon in a polytomy.

For many fossil datasets reconstruction of all ancestor-descendant relationships will always remain infeasible. For records where it might be possible, continued development of methods both to infer the relationships, and to analyse the resulting phylogenies will be important (Dayrat, 2005). In the meantime, two alternative approaches are possible: first to run analyses over large sets of trees to ascertain how robust results are to variation in tree structure (Pol and Norell, 2006; Bapst, 2013), and second to use simulations to model the possible effects of these problems on downstream analyses (Bapst, 2014a). Both of these approaches are incorporated into this thesis.

1.2 Phylogenetic comparative methods in the fossil record

Phylogenetic comparative methods have become commonplace in the past few decades. This followed from the realisation that phylogeny is required for valid macroevolutionary inferences (Felsenstein, 1985) and was brought to the fore by the publication of Harvey and Pagel's landmark text *The Comparative Method in Evolutionary Biology* (Harvey and Pagel, 1991). Evolutionary palaeobiology took longer to catch up - in large part due to a relative lack of timescaled phylogenies - but also due to a resistance to formal cladistic inference approaches to establishing species interrelationships, which remains difficult for some marine invertebrate clades. However, palaeobiology now regularly co-opts these methods for analysis of the evolution of extinct clades (Smith, 1994).

This set of methods uses independent estimates of the evolutionary history of a clade in order to make inferences about a wide variety of micro- and macroevolutionary patterns. Most commonly these include methods to correct for the shared evolutionary history of taxa when looking at patterns of trait correlation (Felsenstein, 1985; Garland and Ives, 2000, Fig. 1.4); to model the pattern in evolution of phenotypic traits (Hansen, 1997; Hunt, 2012; Fitzjohn, 2010); to infer speciation and extinction rates (Rabosky et al., 2013; Hagen et al., 2015) and to detect trends in trait evolution through time (Hunt, 2007; Benson et al., 2014a), among a huge variety of other potential uses. Some models of trait evolution such as the 'early-burst' and directional trends, were inspired by patterns observed in the fossil record and reported by palaeontologists (Cope, 1887; Osborn, 1917; Simpson, 1944). There is a growing literature on these kinds of macroevolutionary patterns, based on analyses of palaeontological data.



Contrast	Value	S.D.	Standardised contrast
A-B	2	2	1
X-C	6	3	2

Figure 1.4: The classic scenario where phylogenetic correction is necessary when performing a comparative analysis of trait data. a) A regression of trait data for two traits, the regression line shows the slope if the data points are treated as independent. However the colours of the points show their phylogenetic relationship according to the hypothetical phylogeny in b). The data are not independent, violating a key assumption of the regression analysis and leading to a false result of a statistically significant correlation between the two traits. c) shows the procedure for phylogenetic independent contrasts, a method via which trait data are corrected for the evolutionary history of the taxa being analysed. S.D. is the square root of the sum of corrected branch lengths, standardised contrast is the contrast value divided by S.D. and corrected branch lengths are calculated as (daughter branch 1 x daughter branch 2)/(daughter branch 1 + daughter branch 2). Figures are adapted from Felsenstein (1985) and a figure made available online by Theodore Garland Jr.

The range of questions that are being addressed is broad but some examples have included: how does shape correlate with ecology? (Pierce et al., 2009); how do rates of evolution relate to the success of a clade? (Benson et al., 2014a); how have constraint and adaptation controlled the evolutionary trajectories of clades? (Friedman, 2012; Hopkins and Lidgard, 2012); and how has extinction shaped the evolutionary history of clades? (Roy et al., 2009; Hardy et al., 2012). It has also been demonstrated using comparative methods that some patterns cannot be detected without the use of fossil data (Finarelli and Flynn, 2006; Slater and Harmon, 2013). These analyses represent only a small sample of the work that has already been done.

The basic principle of PCMs is that patterns within species in a clade (i.e. along a lineage or branch of a phylogeny) can be related to variation across species in that clade (Pennell et al., 2014, Fig. 1.4). This means that these methods are highly versatile in terms of application to different fields and have the potential to provide results that are comparable across disparate datasets, from quantitative genetics to community ecology. Each type of data has a set of special considerations that should be dealt with in order to achieve such comparability.

Evolutionary biologists increasingly use information from fossils to inform their models. There are moves towards using fossils not just as node age priors in molecular phylogenies, but as ancestral state priors and to a truly integrated approach of joint estimation of macroevolutionary parameters (Slater et al., 2012; Slater and Harmon, 2013; Simpson et al., 2011). However, these methods are in their infancy (Pennell and Harmon, 2013). The main path which is open to palaeobiologists is therefore to continue to apply PCMs to trees of fossil taxa, and to gain a better understanding on the relationship between characteristics of input phylogenies and data, and the outcome of those PCMs.

1.3 Phylogeny and extinction in the fossil record

Extinction and origination are the two major factors that sculpt biodiversity throughout time (Jablonski, 2008a). In the face of the Earth's current biodiversity crisis an understanding of macroevolutionary patterns in extinction will be important to inform strategies for conservation and maintenance of biodiversity (Finnegan et al., 2008; Barnosky et al., 2011; Harnik et al., 2012b). The fossil record is an obvious place to turn to for data and insight into the processes that govern extinction. It provides a natural laboratory in which we can test hypotheses of how organisms might respond to projected shifts in climate, and an opportunity to study actual large scale extinctions, rather than extinction risk.

There is a long history of palaeontologists investigating patterns and causes of extinction in the geological past; they were in fact responsible for the idea of extinction (Cuvier, 1818). The palaeontological record of extinction is certainly interesting in its own right, but in order to inform current questions about biodiversity loss palaeontologists must provide results that can inform or be combined with present day analyses of extinctions and extinction risk. Analytic approaches to extinction, particularly those incorporating phylogeny, are a potential avenue through which this can be achieved. Early quantitative approaches to understanding extinction began with a series of papers from Raup and colleagues, investigating what simulation modelling of macroevolutioanry processes could tell us about the nature of mass extinctions, and whether or not random species loss could explain the patterns seen in the fossil record (Raup, 1982; Raup and Sepkoski, 1984; Raup, 1991). Since then many authors have attempted to assess extinction patterns in a quantitative way, in many cases by finding methods of measuring extinction rates through time and for different clades.

There are many studies of modern taxa which attempt to untangle the biological correlates of extinction risk (Balmford, 1996; Bennett and Owens, 1997; Cardillo et al., 2005, 2008). Particularly there has been a focus on the phylogenetic or taxonomic correlates of risk, because both extinction rates and correlates have been shown to vary with evolutionary relationships (Russell et al., 1998; Bielby et al., 2006; Cooper et al., 2008; Davies et al., 2008; Harnik et al., 2012a). Ideally we would be able to measure these correlates in fossil species and unite them with analysis of which taxa actually became extinct, particularly in times of high extinction intensity. Unfortunately these traits (e.g. geographic range, abundance, ecological specialism) are not often preserved in the record. There have been some successful quantitative attempts to find traits in fossil taxa that may have lead to extinction resilience or vulnerability, although these have often focussed on marine invertebrates due to their comparatively high-resolution fossil record (Fig. 1.5, Jablonski, 2008b; Kiessling and Aberhan, 2007; Jablonski, 2005; Hopkins, 2011).

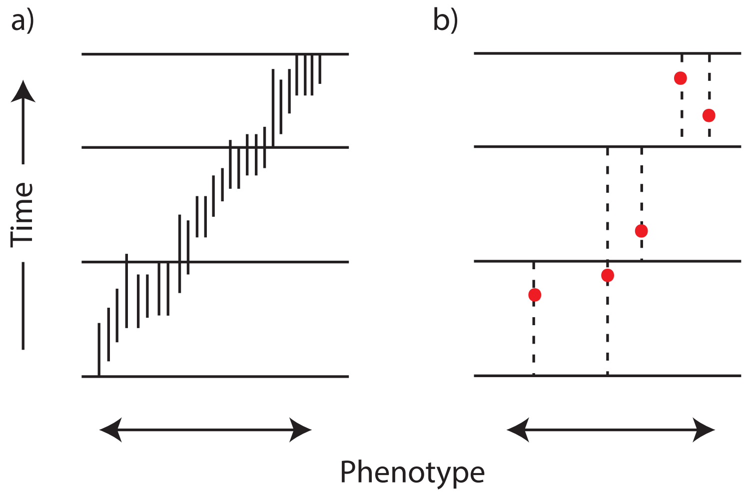


Figure 1.5: This shows the difference between a high resolution fossil record, that would be more likely to be found for marine invertebrate taxa, and a poor resolution or poorly sampled fossil record that is more likely for terrestrial vertebrates. In a) there are many known taxa and the ranges of these taxa (shown by solid black lines) have well constrained start and end dates. In b) there are not very many sampled taxa and the actual samples (shown as red circles) are just of one or a few specimens and their age would not be well known so could only be resolved to the point where it was known the sample came from somewhere along each black dashed line. The species attributes of the taxa in a) are very similar so it may be hard to find differences between them that could help in estimating the phylogenetic relationships between them. In b) on the other hand there are fewer taxa and they are more phenotypically distinct, making cladistic analysis comparatively more straight forward. Therefore for a high resolution fossil record you are likely to have good stratigraphic but poor phylogenetic control, and vice versa for a low resolution record. This is the case except for some (rare) extremely high resolution records where the integrading of phenotype from one morphospecies to another can be observed, leading to good control on both ranges and phylogenetic relationships.

Phylogenetic non-randomness in the pattern of extinction itself (i.e. treating extinction as a binary trait) offers an alternative that could help to understand extinction selectivity in the geological past. Recent papers by Roy et al. (2009), Harnik (2011) and Hardy et al. (2012) indicate that analyses of phylogenetic clustering of extinction could be a powerful tool in understanding extinction dynamics in the fossil record. This is discussed in detail in Chapters 3 and 4 of this thesis.

1.4 Age data and timescaling trees of fossil taxa

1.4.1 Range data and times of observation

For some fossil clades we have excellent resolution on the ages of fossil specimens and very complete records (Fig. 1.5 e.g. stickleback fish, Cenozoic planktonic foraminifera and Ordovician brachiopods; Bell et al., 2006; Aze et al., 2011; Tyler and Leighton, 2011). For others we have poor resolution on ages and most taxa within the clade comprise one specimen (a 'point occurrence', e.g. plesiosaurian reptiles Benson et al., 2014a). In fact there are a large variety of types of fossil record, the varying information from which should lead to different methods of data preparation (Bapst, 2014b). The subsequent chapters of this thesis focus largely on fossil tetrapods, which for the most part have poor age constraints on fossil occurrences and few representative specimens per taxon, but have well resolved and robust cladistically inferred phylogenetic trees. The following discussion on timescaling methods is therefore from the perspective of analyses of this kind of data.

An important point to note is that a phylogeny is meant only to represent the amount of *evolutionary divergence* between taxa. Once a cladogram of fossil taxa has been scaled to time, one possibility is to assume that the lengths of terminal branches represent real ranges of stable morphotaxa through time (i.e. anything sampled along this branch would have the same morphology and be classified within the same taxon). This corresponds to a punctuated equilibrium model of diversification where all trait change happens at speciation events (Eldredge and Gould, 1972; Hunt, 2007). For some methods (including that in Chapters 3 and 4 of this thesis) this assumption is integral to the analysis. This means that phylogenetic frameworks for these analyses must be timescaled using the first appearance dates of taxa, and the terminal branch lengths extended to the last appearance dates, which are used as a proxy for the time of extinction.

Whether or not punctuated equilibrium is in fact a good model for macroevolution has been highly debated since Eldredge and Gould's seminal publication (Eldredge and

Gould, 1972). On very short timescales (i.e. that we can observe in modern taxa) it seems clear that evolutionary change occurs gradually and accumulates (Simpson, 1944). However, on geological timescales, based on palaeontological evidence, the establishment of new body plans and the origin of major clades - macroevolutionary change - appears to happen in a short period of time compared to how long that clade subsequently persists for in the record (Stanley, 1998). In reality, generation of new body plans may have taken hundreds of thousands of years, it is unlikely that it was in any sense instantaneous on human timescales. Viewed on long-term geological timescales however, the timescale to which models used in this thesis are applied, punctuated equilibrium may well be an appropriate model, as it captures the nature of the speciation and extinction that can be seen in real palaeontological data.

In the normal usage of phylogenies to represent the divergence between instantaneous populations, trees should be timescaled using the time of occurrence of the observed specimen from which the trait data were measured. Obviously for a specimen whose age of occurrence can only be constrained to a geologic stage this is not possible. In these circumstances the best way forward is to run analyses on stochastically timescaled sets of trees, which gives an estimate of the error around results caused by variation in tree structure. This is done by randomly selecting an age for each taxon from the period of time to which it can be constrained and timescaling the cladogram topology using these dates, then repeating these two steps.

With a cladogram topology and an age dataset in hand (either actual ages or the best constraint available on the age), each of the following methods can be used to scale the branch lengths of a tree to generate a phylogenetic hypothesis that can then be used as the framework for a PCM.

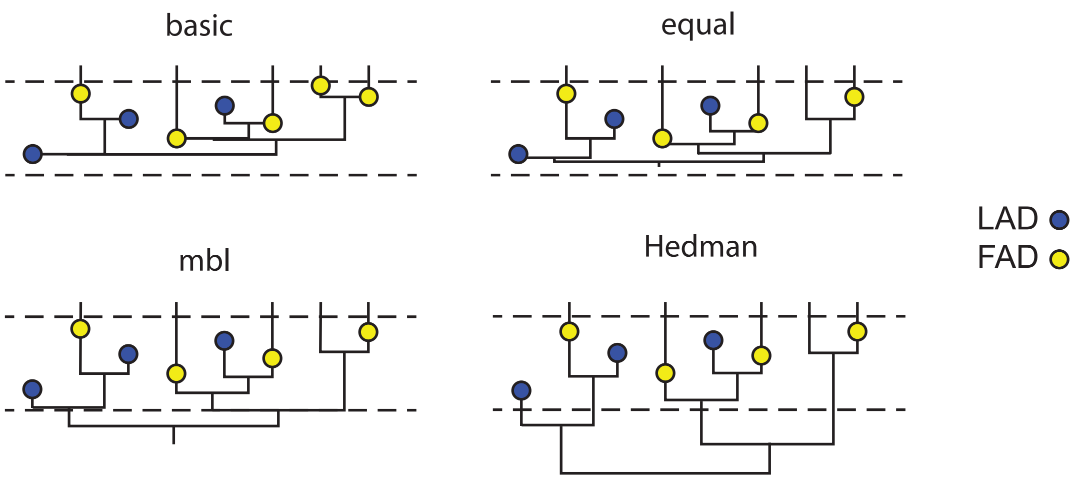


Figure 1.6: Possible effects of different timescaling methods on the lengths of branches in a tree. Basic shows a tree which includes zero length branches (no vertical length), equal, mbl and Hedman show three methods which can be used to add length to zlbs and estimate divergence dates. The other method mentioned in the text - cal3 - produces branch lengths similar to the Hedman method.

1.4.2 Basic

The *basic* method (Fig. 1.6) is the simplest way in which a cladogram can be scaled to time, involving the fewest assumptions. It relies upon only one - that the origin of a taxon must be at least as old as the origin of its sister group. In practice this means that branch lengths are extended back in time so that the start of a taxon's range matches that of its sister, and the date of divergence between the two taxa is assumed to be at the first appearance date of whichever of the pair appeared earlier in the record. This method often leads to zero-length internal branches (zlbs), when a taxon is sister to a clade with two or more taxa that appear earlier in the record. This is a problem analytically; many phylogenetic comparative methods cannot be implemented on a phylogeny that includes zlbs because this implies no evolution along the branch. There are two possible ways around these analytically intractable phylogenies, one is to collapse all zlbs to polytomies (perhaps undesirable as this discards positive information about the order of divergences), or somehow give a positive length to the zlbs, which leads on to the following two methods.

1.4.3 Equal and mbl

The *equal* (Brusatte et al., 2008) and *mbl* methods (Fig. 1.6) are similar in their aim, but different in implementation. They are both (somewhat arbitrary) ways to extend zlbs so that analyses can be run. These extensions are not based on any knowledge of how much time might have passed (or how much evolutionary change might have occurred) along the branch. In the equal method an additional length is chosen by the user, this length is divided equally among all the zlbs in the phylogeny. For *mbl* the user chooses some minimum branch length and every branch that is shorter (including those that already have some length) are extended to match this value. Generally in publications the *mbl* method has been preferred above *equal*, largely because its outcomes are more predictable. Although these methods are very easy to implement, their lack of sophistication calls into question the likelihood of either of them producing accurate divergence dates. However, Bapst (2014a) showed that for particular types of phylogenetic comparative analyses they can lead to the least bias in results.

1.4.4 Hedman

The *Hedman* method (Fig. 1.6, Hedman, 2010) is a Bayesian method used to estimate the age of a node based upon the ages of its closest outgroups. It requires that each successive outgroup moving away from the node of interest be at least as old as the one before. When implemented in this thesis it is incorporated into a larger algorithm that adjusts the time of all nodes in the cladogram, whether or not their position can be directly estimated with the Hedman algorithm. This method is most useful for fully extinct, or poorly sampled clades.

The algorithm aims to obtain posterior probabilities of the possible positions of nodes based on known values of earliest sister clade appearances. The probability distribution functions for nodes cannot be effectively modelled; they are therefore discretised and estimated based on the known fossil outgroup ages. The estimated node date is the mode of the posterior probability distribution.

The overarching function used in this thesis searches the tree for nodes which have successive fossil outgroups and dates them using the Hedman algorithm. There will usually be many nodes within the cladogram with outgroups that are not stratigraphically consistent. These nodes are selected in a random order and a date drawn from a uniform distribution between the ages of the bounding *Hedman*-scaled nodes.

1.4.5 cal3

This is a stochastic rate-based method proposed in Bapst (2013) and available in the package *paleotree* (Bapst, 2012). Due its dependence on rates of extinction, origination, and sampling, this method is best suited to groups with well-sampled records and consistently good preservation. These three input rates are used to produce a constrained distribution of possible divergence dates for each node, beginning at the base of the tree, and the dates are randomly selected from these distributions on each pass. The process is repeated to produce distributions about each node, which are effectively confidence intervals for the divergence dates. The original publication demonstrates with simulations that the distributions bracket true divergence dates (Bapst, 2013).

1.4.6 Other methods

There are additional methods that can be used when a character dataset is also available (e.g. Ruta et al., 2006; Laurin, 2004), which deal with zlbs by sharing branch lengths between branches according to the number of character changes on each branch. There are also more complex character data based approaches, generally referred to as ‘tip-dating’. These are ways in which divergence dates and topology can be simultaneously estimated using a Bayesian modelling technique. These methods are therefore useful in situations where the topology is as yet unknown and a character matrix is available. Some morphology based tip dating approaches can be used to date trees of completely extinct clades, but it is also possible to use fossil and extant taxa and a combination of morphological and molecular data in a ‘total-evidence’ approach, a technique that has been enthusiastically adopted (Pyron, 2011; Ronquist et al., 2012; Near et al., 2014; Dornburg et al., 2015). This is a major advantage of this method, as the methods outlined above can have odd effects when extant taxa are included (see Chapter 2 of this thesis and Bapst, 2014a). The downsides are that it cannot be used in the absence of a detailed character dataset and there are uncertainties about how realistic the divergence dates that they estimate are (Arcila et al., 2015).

1.5 Outline of the thesis

Chapter 1 - I introduce concepts and methods relevant to the rest of the thesis, and outline the themes and questions to be addressed.

Chapter 2 - (Data available: taxonomic or phylogenetic relationships) I make a comparison between recent cladistically inferred topologies, and historical (pre-cladistic) taxonomies for the same clade, for a set of 52 clades. I provide results of several commonly used phylogenetic comparative methods using the cladistic and taxonomic topologies as frameworks and assess whether the two topologies yield similar results. Finally I present guidelines for when and how it might be acceptable to substitute taxonomy for a formally inferred cladogram in phylogenetic comparative methods.

Chapter 3 - (Data available: phylogenetic relationships and age of taxon occurrences) I measure the phylogenetic clustering of the timings of extinctions in three major tetrapod clades (synapsids, sauropsids and temnospondyls) for a timespan encompassing the Permo-Triassic boundary. I discuss these results in the context of changes in species richness in each clade at the time.

Chapter 4 - (Data available: phylogenetic relationships and age of taxon occurrences) I test the degree to which palaeontological phylogenies produce results comparable to simulated ‘real’ phylogenies in analyses of the phylogenetic clustering of extinction. I explore the ways in which various features unique to phylogenies of fossil taxa (particularly inclusion of ancestor-descendant pairs) affect the results of the analyses.

Chapter 5 - (Data available: phylogenetic relationships and continuous univariate trait data) I present an analysis of a large dataset of Sauropterygian body plan and functionally relevant traits. I use a phylogeny of the clade and the continuous trait data to measure the comparative rates of evolution of aspects of the body plan to explore the influence of phylogenetic and environmental constraint on body plan evolution in the clade. These analyses are used to answer three key questions: 1) What mechanisms controlled the evolution of body plan in the clade? 2) Did the transition to open water habits at the end of the Triassic correspond to a shift in evolutionary constraint on body plan? 3) Do plesiosaurian morphotypes represent adaptive peaks in the evolutionary landscape?

Chapter 2

Taxonomy and phylogeny can yield comparable results in palaeontological comparative analyses

> *"We should not allow ourselves to be deluded by the hope of finding absolute standards, but it should be our object to arrange what is really known, so that it can be easiliy grasped and utilized"*
>
> \- W.G.Farlow (1898)

In this chapter I make a comparison between recent cladistically inferred tree topologies, and historical (pre-cladistic) taxonomies for the same clade, for a set of 52 clades. I provide results of several commonly used phylogenetic comparative methods using these cladistic and taxonomic topologies as frameworks, and assess whether the two topologies yield similar results. Finally I present guidelines for when and how it might be acceptable to substitute taxonomy for a formally inferred cladogram in phylogenetic comparative methods. Content from this chapter was published in Systematic Biology with Matt Friedman as co-author. Matt Friedman conceived the initial idea, I designed the study, collected the data, performed the analyses and wrote the chapter.

2.1 Taxonomy and comparative analyses

Current approaches to assessing macroevolution incorporate a temporally scaled, explicit hypothesis of relationships among taxa (an inferred phylogeny) in order to correctly estimate relationships between traits (Felsenstein, 1985; Grafen, 1989; Martins and Hansen, 1997), investigate the phylogenetic signal of traits (Fritz and Purvis, 2010; Münkemüller et al., 2012) and compare models of evolutionary mode (Harmon et al., 2003, 2010).

These comparative methods were originally developed for Recent datasets, where detailed morphological and molecular data can be obtained and used for explicit, character-based phylogenetic inference. Many clades have few or no extant species and yet are diverse and highly abundant in the fossil record, and have the potential to reveal a great deal about long-term evolutionary change and mechanisms (Jablonski, 2007, 2008a). Unfortunately these same clades have often not been subjected to formal phylogenetic analysis for a variety of reasons, the most acute of which are limited morphological data (Sansom and Wills, 2013) and the time-consuming nature of morphological phylogenetic systematics relative to molecular approaches. Consequently, such clades have not been a major focus of comparative evolutionary analysis despite the unique deep-time perspective that they provide. An increasing number of workers circumvent the problems raised by the lack of a formal phylogeny by executing analyses using phylogenetic comparative methods (PCMs) that incorporate taxonomy (the traditional method of classification, under the Linnaean hierarchy) as at least a partial substitute for an analytically inferred cladogram (Bininda-Emonds et al., 2007; Slater and Van Valkenburgh, 2008; Roy et al., 2009; Fusco et al., 2012; Hardy et al., 2012; Benson et al., 2014a) on the basis that morphologically defined taxonomic classifications closely match trees inferred from molecular data (Jablonski and Finarelli, 2009). Conversion of taxonomy to a phylogeny and use in a PCM is now easily achievable using various software packages (e.g. *ape* and *geiger* in R (Paradis et al., 2004; R Core Team, 2015; Pennell et al., 2014) each provide functions for this task), which further facilitates and encourages this practice.

The response of PCMs to various factors has largely been characterised for ultrametric trees of extant groups (Housworth and Martins, 2001; Davies et al., 2012). Phylogenetic independent contrasts (Felsenstein, 1985) and phylogenetic generalised least squares (Grafen, 1989; Martins and Hansen, 1997) in particular have been shown to be robust to inaccuracies in branch lengths (Garland and Diaz-Uriarte, 1998), and inaccuracies in topology (Symonds, 2002; Stone, 2011). Phylogenetic generalised least squares is especially robust in cases where branch lengths are known (Purvis et al., 1994). Use of taxonomy for extant datasets has previously been rejected in favour of performing PCMs on large sets of 'random' possible trees (Losos, 1994; Martins, 1996), but other workers have suggested that this approach is at best equivalent to using the raw data without a phylogenetic correction (Abouheif, 1998; Symonds, 2002). Taxonomy therefore potentially represents a viable middle ground that contains useful information on evolutionary relationships and, in the case of fossil data, information on branch lengths based on the stratigraphic distribution of fossils.

However, taxonomies could represent inaccurate topologies relative to formal solutions, and are often more poorly resolved (Harvey and Pagel, 1991; Kroh and Smith, 2010). The effects of these issues on downstream PCMs are not well understood and have not been directly tested in the context of palaeontological datasets, where the application of taxonomies in place of phylogenies is most likely to occur. It is therefore important to understand the adequacy of these topologies when they are used in this context, to assess the validity of work that has already been done, and define guidelines for the limits of acceptable use in the future. I seek to constrain this problem through a series of experiments using taxonomies and more recently published cladograms for various animal clades. My approach assumes that the results of formal phylogenetic analysis—the 'gold standard' for the application of PCMs to palaeontological data—capture the true topology and, after time-scaling, the true branch length of the underlying phylogeny. I make these assumptions because they match those made when a PCM is applied using a cladistically inferred phylogeny of fossil taxa. I simulate continuous characters on the cladistically inferred phylogeny, then estimate relevant values using both this phylogeny and a phylogeny derived from taxonomy.

The question of whether cladistically inferred phylogenies of fossil taxa are in fact suitable for use in these kinds of analyses in the first instance is not one that I address here, and is under investigation by other workers (Bapst, 2014a). Comparative studies in the fossil record are performed under the widely accepted assumption that a correction using an analytically inferred phylogeny of fossil taxa produces more reliable results than an analysis with no phylogenetic correction (Felsenstein, 1985; Laurin, 2004; Brusatte et al., 2011). Increasing numbers of studies also assume that taxonomy is a suitable (at least partial) substitute for an analytically inferred phylogeny (Slater and Van Valkenburgh, 2008; Roy et al., 2009; Turvey and Fritz, 2011; Fusco et al., 2012; Hardy et al., 2012; Benson et al., 2014a) and it is this second assumption I aim to test. I use a comparison between results from these two alternative phylogenetic frameworks to develop and present practical guidelines for incorporating taxonomy into PCMs when a formal cladistic solution is not yet available, and discuss the degree to which this approach can be applied in a palaeontological framework.

2.2 Experimental design

2.2.1 Taxonomies and formal phylogenetic solutions

I examined 52 datasets from 44 animal clades for which both a recently published explicit phylogenetic hypothesis and a traditional taxonomy that pre-dated the most recent cladogram were available. The taxonomies used here can be considered a worst-case scenario, as they are pre-cladistic, and were the (historically) earliest classification I could find that included a sufficient sample of taxa. More recent taxonomies - many based on formal cladistic solutions - are now available for these clades, so my estimate of the success of a taxonomy relative to a cladogram is a conservative one. The dataset comprised 5 invertebrate clades and 39 vertebrate clades, including 27 tetrapod groups and 12 non-tetrapod vertebrate groups; these are listed in Table 2.1.

Some cladograms were sourced from the same clade, but contained different non-overlapping selections of taxa from those clades and so were used as separate datasets (e.g. I included three different bivalve datasets). These studies are the total set of clades for which I could find an appropriate taxonomy and cladogram that contained enough matching taxa for a PCM to be run on the tree, after taxa not present in both were removed. There were taxonomies and phylogenies available for some additional clades, however they did not include enough overlapping taxa for analysis. I attempted to get as broad a taxonomic coverage as possible but the dataset is vertebrate-heavy, because vertebrates are the overwhelming focus of formal phylogenetic inference in the fossil record. Invertebrates (e.g. bivalves or marine arthropods) are comparatively under-represented in formal phylogenetic analyses and so make up a smaller component of the dataset than vertebrates, despite having a more extensive fossil record. Many of the taxonomies were taken from the same publication (Romer, 1945) as this was the most comprehensive (and for many clades the only) pre-cladistic taxonomy available. I included other sources where possible to provide a comparison to ensure there was no systematic bias in results produced using taxonomies from Romer.

All but one of the classifications are at the generic level, thereby reducing the potential influence of different authors' tendency to lump or split fossil species, as we expect that given the greater morphological distinction between taxonomic units as you move up the hierarchy the controversy over groupings reduces with each rung. However, whether authors had a tendency to lump or split should not make a difference to the outcome of this analysis. It might effect the number of taxa or the ranges of taxa in an analysis but the focus here is a comparison between results from taxonomy and phylogeny using an identical set of taxa with the same range data used for both. Theoretically however, the results produced here can be considered to apply to any level of taxonomic unit. The differences between classifications and phylogenetic hypothesis of taxa at different ranks are that lower ranks tend to have more taxa, and those taxa will tend to have shorter ranges. The effect of the number of taxa in the analysis, the longevity of taxa and the sampling of taxa are all factors that are considered in subsequent analyses. Therefore any worker wishing to use a taxonomic classification should think about the likely effect of the rank of this classification on all these factors, based on the general conclusions of this study pertaining to questions about the effect of dataset size and the stratigraphic longevity and resolution of taxa.

For each example, I investigated the degree to which evolutionary inferences drawn from taxonomy-based phylogenies (TBPs) matched those derived from cladogram-based phylogenies (CBPs). I took this approach based on the following assumption: a comparison between out-of-date taxonomies and current phylogenetic solutions is equivalent to a comparison between current taxonomies and the hypothetical phylogenetic solutions of the future for groups that do not currently have them. For the purposes of these analyses, I assume that the branching structure provided by a CBP matches that of the true phylogeny. As such, the assessment of the success of CBPs in capturing evolutionary processes and patterns should be viewed as maximally optimistic and my results should not be taken as evidence for or against the capacity of these methods to capture real patterns. In particular, because the trait data are simulated, none of the analyses provide any information on real patterns in these groups.

2.2.2 Tree construction and timescaling

All tree manipulation and analyses were performed in the R software environment (version 3.0.2; R Core Team, 2015). Topologies derived from cladograms and taxonomies were time-scaled in order to produce phylogenies (method outlined below). References and details for the source topologies are shown in Table 2.1. Taxonomic classifications, cladograms and range data are all included on the data appendix disk for this thesis. All datasets are at the generic level except that for tetraodontiform fishes, where species-level classifications and range data were available. When selecting cladograms I used whichever tree topology the original authors had applied for phylogenetic comparative analyses (if included), or the topology preferred by the original authors in the absence of further analyses within the publication. This was to ensure that the dataset included topologies that would be the most likely to be accepted for use with PCMs incorporating palaeontological data. It therefore included solutions arising from Bayesian, maximum likelihood and maximum parsimony inference. The literature used to obtain taxonomies only contained one classification scheme for each clade, and this was converted in to a tree structure as a series of nested polytomies corresponding to each taxonomic rank (Fig. 2.1).

Taxonomies by nature contain many polytomies when directly plotted as trees (e.g. if there are 5 genera contained within one family, these genera would be depicted as a single multichotomy, unless sub-familial relationships had been proposed). These were left as hard polytomies to represent the maximum amount of resolution based upon available information, except where the PCM required a fully resolved tree (mode of evolution). In preliminary analyses, executing simulations where (i) taxonomies were randomly resolved before time-scaling, or (ii) random trees used for comparison were collapsed to have the same number of internal nodes as the TBP, did not make a notable or systematic difference to the outcome. Adding polytomies to the cladograms to match the resolution of the taxonomies ought to have the same result, as this indicates that resolution itself does not have an important influence on the outcome. This is consistent with previous work showing that the inclusion of polytomies in a phylogeny for a PCM does not bias the result and has a negligible effect on the rate of type I error (Garland and Diaz-Uriarte, 1999; Stone, 2011). Housworth and Martins (2001) provide a method by which error caused by uncertainty in relationships within a polytomy can be incorporated into estimates of error bounds for the test statistic in a PCM.

Table 2.1: Sources of taxonomy and cladogram topologies

Clade	Taxonomy	Cladogram	Inference method	Range data	Includes extant
Acanthodians 1	Romer (1945)	Brazeau (2009)	Bayesian	PBDB	No
Acanthodians 2	Romer (1945)	Hanke and H (2004)	Parsimony	PBDB	No
Actinopterygians 1	Romer (1945)	Coates (1999)	Parsimony	PBDB	No
Actinopterygians 2	Romer (1945)	Gardiner et al. (2005)	Parsimony	PBDB	No
Adephagans	Beutel et al. (2008)	Beutel et al. (2008)	Supertree	PBDB	Yes
Anguimorphs	Romer (1945)	Conrad et al. (2011)	Combined parsimony	PBDB	Yes
Anomodontids	Romer (1945)	Angielczyk and Kurkin (2003)	Parsimony	PBDB	No
Arctoids	Romer (1945)	Finarelli (2008)	Combined parsimony	Finarelli and Flynn (2006)	Yes
Arthrodires	Romer (1945)	Carr and Hlavin (2010)	Parsimony	PBDB	No
Birds	Romer (1945)	Livezey and Zusi (2007)	Parsimony	PBDB	Yes
Bivalves 1	Cox et al. (1971)	Schneider (1995)	Parsimony	PBDB	Yes
Bivalves 2	Cox et al. (1971)	Giribet and Wheeler (2002)	Parsimony	PBDB	Yes
Bivalves 3	Cox et al. (1971)	Plazzi et al. (2011)	Molecular	PBDB	No
Carcharodontosaurids	Weischampel et al. (2004)	Cau et al. (2012)	Parsimony	PBDB	No
Carnivoramorphs	Romer (1945)	Spaulding and Flynn (2012)	Parsimony	PBDB	Yes
Cetartiodactyls	Romer (1945)	O'Leary and Gatesy (2008)	Parsimony	PBDB	Yes
Chondrichthyans	Romer (1945)	Coates and Sequeira (2001)	Parsimony	PBDB	No
Coelacanths 1	Romer (1945)	Forey (1998)	Parsimony	PBDB	No
Coelacanths 2	Romer (1945)	Forey (1998)	Parsimony	PBDB	No
Cricetid rodents	Romer (1945)	Maridet and Ni (2013)	Parsimony	PBDB	No
Crocodylians	Steel (1973)	Brochu (2003)	Combined parsimony	PBDB	No
Decapods	Karasawa and Kato (2003)	Karasawa and Kato (2003)	Parsimony	Karasawa and Kato (2003)	Yes
Dicynodonts	Romer (1945)	Surkov and Benton (2004)	Parsimony	PBDB	No
Early teleosts	Romer (1945)	Arratia and Tischlinger (2010)	Parsimony	PBDB	Yes
Early tetrapods	Romer (1945)	Ruta et al. (2003)	Parsimony	PBDB	No
Echinoids	Duncan (1889)	Kroh and Smith (2010)	Parsimony	PBDB	Yes
Gnathostomes	Romer (1945)	Davis et al. (2012)	Parsimony	PBDB	No

Table 2.1: Sources of taxonomy and cladogram topologies

Clade	Taxonomy	Cladogram	Inference method	Range data	Includes extant
Halecomorphs	Romer (1945)	Lambers (1995)	Parsimony	PBDB	Yes
Ichthyopterygians	Rieppel (1973)	Fröbisch et al. (2012)	Bayesian	PBDB	No
Stegocephalians	Romer (1945)	Pyron (2011)	Bayesian	PBDB	Yes
Mammals	Romer (1945)	Lloyd et al. (2012a)	Supertree	Lloyd et al. (2012a)	Yes
Mesozoic birds	Weischampel et al. (2004)	O'Connor and Zhou (2012)	Parsimony	PBDB	No
Neopterygians	Romer (1945)	Cloutier and Ahlberg (1996a)	Parsimony	PBDB	Yes
Octodontids	Romer (1945)	Verzi et al. (2014)	Parsimony	PBDB	Yes
Ornithiscians	Weischampel et al. (2004)	Coria et al. (2013)	Parsimony	PBDB	No
Osteostracians	Romer (1945)	Sansom (2008)	Parsimony	PBDB	Yes
Sphenisciformes	Romer (1945)	Ksepka et al. (2006)	Combined parsimony	PBDB	Yes
Primates	Romer (1945)	Seiffert et al. (2010)	Parsimony	PBDB	Yes
Pterosaurs	Romer (1945)	Lü et al. (2010)	Parsimony	PBDB	No
Sarcopterygians 1	Romer (1945)	Friedman et al. (2007)	Parsimony	PBDB	No
Sarcopterygians 2	Romer (1945)	Friedman (2007)	Parsimony	PBDB	No
Sarcopterygians 3	Romer (1945)	Cloutier and Ahlberg (1996b)	Parsimony	PBDB	Yes
Sauropods	Weischampel et al. (2004)	Apaldetti et al. (2013)	Parsimony	PBDB	No
Sirenians	Romer (1945)	Benoit et al. (2013)	Parsimony	PBDB	Yes
Squamates	Romer (1945)	Wiens et al. (2010)	Bayesian	PBDB	Yes
Stylonurine Eurypterids	Tollerton (1989)	Lamsdell et al. (2010)	Parsimony	PBDB	No
Temnospondyls	Romer (1945)	Ruta et al. (2007)	Parsimony	PBDB	No
Tetraodontiformes	Tyler (1980)	Santini and Tyler (2003)	Parsimony	Santini and Tyler (2003)	Yes
Theropods 1	Weischampel et al. (2004)	Lee and Worthy (2012)	Maximum Liklihood	PBDB	No
Theropods 2	Romer (1945)	Lee and Worthy (2012)	Maximum Liklihood	PBDB	No
Turtles	Romer (1945)	Sterli et al. (2013)	Parsimony	Sterli et al. (2013)	No
Tyrannosaurids	Weischampel et al. (2004)	Brusatte and Benson (2013)	Parsimony	PBDB	No

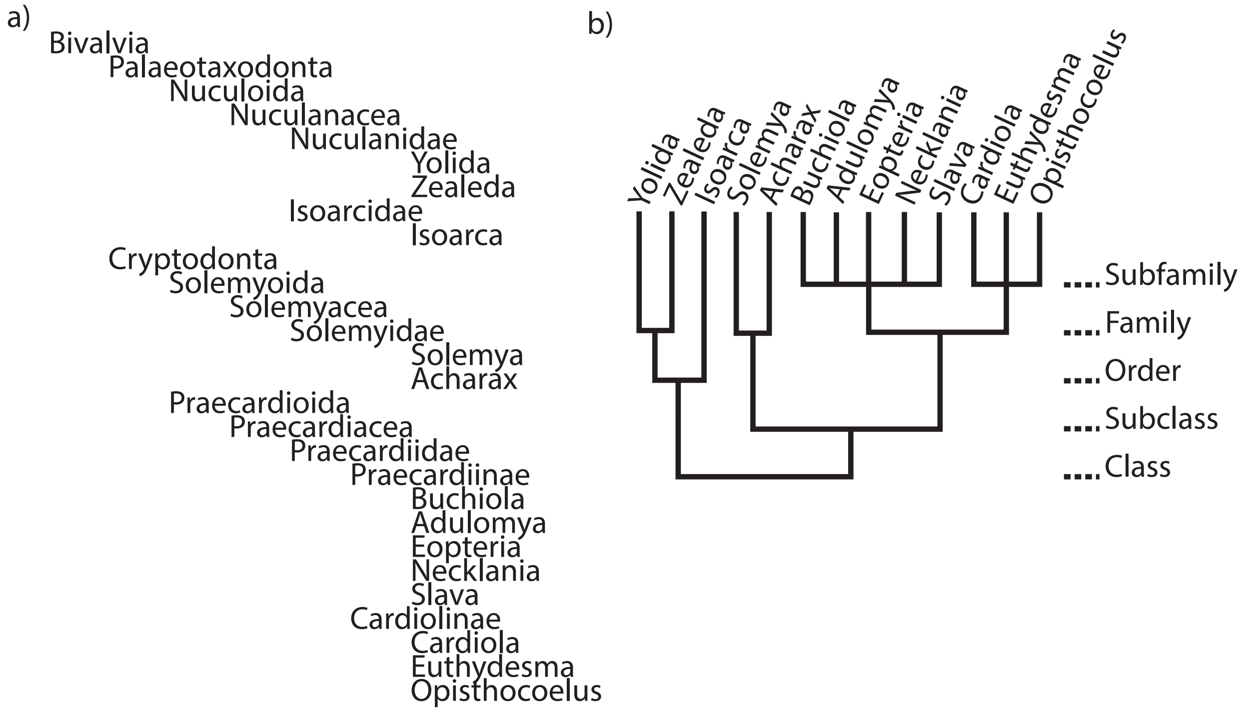


Figure 2.1: Method for converting a taxonomic classification to a cladogram that can then be time-scaled with fossil range data to make a phylogeny. Taxa that are in the same group at a particular rank are combined in a polytomy, starting at the genus level and moving towards the root of the tree. a) shows the original classification as published, b) shows the resulting cladogram after conversion, before time-scaling.

Cladograms of extinct taxa can be scaled according to the first appearance date (FAD) of each taxon to generate phylogenies with branch lengths representing the amount of time since sister taxa diverged (Chapter 1, Lloyd et al., 2012b; Bapst, 2013, 2014a). The branch lengths are estimated based on the FAD of each taxon in the fossil record, and the assumption that the divergence between two lineages must have occurred, at the latest, at the FAD of the older taxon. Some analyses also require an estimate of the last appearance date of a taxon (e.g. measuring the phylogenetic clustering of extinction) to estimate a taxon duration. First and last possible appearance dates for all taxa were principally obtained from the Paleobiology Database (at the time www.paleodb.org, now www.paleobiodb.org and www.fossilworks.org). These data were modified where the taxon was known to be extant, and where investigation of outliers showed an incorrect assignment. Where these data were not included in the PaleoBioDB, they were taken from literature detailed in table 2.1. These first and last occurrences are unlikely to correspond precisely to the real origin and extinction times, because the fossil record is not continuous. Several methods have been developed to estimate divergence dates and lineage durations that are a closer approximation to real divergences than the raw first and last occurrence dates of fossil taxa (e.g. Lloyd et al., 2012b; Bapst, 2013; Lee et al., 2014). A further complication with many of the taxa that I used is that they are point occurrences, meaning that only one specimen, or a few specimens from the same bed, have been found. In this case the date is often only known to stage-level precision. For these taxa the ages in Ma for the FADs were drawn randomly from a uniform distribution specified by the beginning and end date of the stage in which they were found. In reality, speciation events may have been more clustered than the assumption of a random distribution, however, the timescaling process compensates for overdispersal of FADs by inferring simultaneous origins of sister clades. In addition, to a certain extent the accuracy of these dates was unimportant as the important factor was that the same dates were used to timescale both the CBP and TBP.

I timescaled all the trees using the *'mbl'* (minimum branch length; Section 1.4.3) method as originally implemented in R by Lloyd et al. (2012b) and incorporated into the *paleotree* package (Bapst, 2012). This is a commonly used time-scaling method in palaeontological analysis, which avoids the problem of analytically intractable phylogenies caused by zero-length branches (ZLBs). Having zero length internal branches effectively creates a polytomy. Analyses are possible if these ZLBs are treated as hard polytomies, however, I have positive information that a bifurcation occurred and that these branch lengths were not in fact zero, and therefore should not be expressed as a polytomy. Divergences are dated according to the first appearance date of the sister group to the clade, and then any branch that is shorter than a specified length of time (for this study I used 1 Myr) is extended to this length. Significantly more complex approaches to calibrating palaeontological trees are available but the information needed to perform them is not readily available for the majority of these datasets, and would be unavailable for many of the clades for which authors may want to include taxonomic information. For example some methods require the clade to include extant taxa (Pyron, 2011; Ronquist et al., 2012; Nowak et al., 2013). The latest total evidence dating, or fossil tip-dating methods based on so-called morphological clocks require a character matrix to simultaneously estimate phylogenetic topology and branch lengths (Lee et al., 2014), which is by definition unavailable for taxonomies without formal phylogenetic solutions. Other methods require estimation of branching, extinction and sampling rates (Bapst, 2013), which is not possible with the point occurrence data that I had for many of the clades in my dataset. For some well-sampled clades it may be possible to use a more sophisticated method than *mbl* (particularly *cal3* (Bapst, 2013)), and I would encourage workers to use the most suitable time-scaling method available, as improvement in accuracy of branch length can lead to improvement of performance in analyses (Bapst, 2014a).

2.2.3 Tree similarity

I used the matching cluster (MC) distance metric (Bogdanowicz and Giaro, 2013) to measure the similarity of each TBP to its equivalent CBP based on differences in branching pattern, implemented using TreeCmp in Java (Bogdanowicz et al., 2012). The method divides the tree into clusters such that all the branches of the tree are included; a single cluster contains all the tips that descend from a particular internal node. Each cluster is matched with one from the other tree for a pairwise comparison. For each pairwise comparison the symmetric difference is calculated. This is the total number of differences in taxon positions between one cluster and the other, weighted by the proximity of each taxon to the root of the tree. This difference is summed across all matched pairs to obtain a total value. The method searches across all possible permutations of clusters and matchings to find the minimum total value possible, this is the ‘minimum-weight perfect matching’ or the matching cluster distance (Bogdanowicz and Giaro, 2013). I found the significance of this value by: 1st simulating random trees with the same number of taxa; 2nd time-scaling these random topologies using the same data with which I time-scaled the cladogram and taxonomy; and 3$^{\text{rd}}$ calculating their MC distance in order to generate a distribution against which to compare our empirical values. This method shares its basis with the commonly used Robinson-Foulds (RF) distance (Robinson and Foulds, 1981). However, the null distribution generated for MC is unimodal and close to Gaussian, as opposed to the null distribution for the RF distance which is often bimodal. Furthermore, the metric itself is more robust; RF distance can respond drastically to small changes in the position of terminal nodes (e.g. moving one terminal branch to a different position can generate the maximum RF distance (Xin and Li, 2014)), whereas MC distance does not suffer from this problem (Münkemüller et al., 2012). I also tested the correlation between the CBP and TBP variance-covariance matrices using Pearson's product moment correlation coefficient.

2.2.4 Relative Performance of Cladogram and Taxonomy-Based Trees in Phylogenetic Comparative Methods

I estimated four phylogenetically explicit measures, using a variety of packages from the statistics environment R (Table 2.2). The details of all the steps used for investigation of each method are shown in Figure 2.2.

Table 2.2: Downstream analyses tested, and the R package and function implemented

Measurement	Metric	R package - Function
Phylogenetic signal	Blomberg's K	picante - Kcalc (Kembel et al., 2010)
Clustering of extinction	Fritz and Purvis' D	caper - phylo.d (Orme et al., 2012)
Mode of evolution	Model and Akaike weights	OUwie - OUwie and OUwie.sim (Beaulieu et al., 2012)
Phylogenetic generalised least squares	PGLS slope	caper - pgls (Orme et al., 2012)

2.2.4.1 Phylogenetic signal

I simulated trait data on the CBP under a Brownian motion model with $\sigma^2 = 1$ (step variance) and $\mu = 0$ (mean) then followed the steps in Figure 2.2. After simulation I controlled the strength of signal using equation 2.1 as proposed in (Münkemüller et al., 2012) to make three sets of simulations with weak, medium and strong signal

$$trait = w.trait_{BM} + (1 - w)trait_{rand} \tag{2.1}$$

$trait_{BM}$ is a set of trait values simulated under pure Brownian motion with $\mu = 0$ and $\sigma^2 = 1$. $trait_{rand}$ is these values randomly shuffled over the tips of the phylogeny to produce a set of random trait values. The weighting factor w is then used to control which component has the larger influence, on a spectrum from completely Brownian (strong signal, $w = 1$) to white noise (weak signal, $w = 0$), with a medium signal using $w = 0.5$. I then z-transformed all values in order to standardise results.

2.2.4.2 Phylogenetic generalised least squares

I simulated sets of coevolving traits with a phylogenetic regression slope of 0.3, and used the TBP and CBT to estimate a value for the slope of each set of simulated traits, according to the steps in Figure 2.2.

2.2.4.3 Mode of evolution

I simulated data using either a Brownian or Ornstein-Uhlenbeck (OU) model of gradual change. OU models are similar to Brownian motion models of evolution, except they include an additional parameter α - the strength of selective constraint, which increases the probability that the direction of trait change will be towards an evolutionary optimum value (Hansen, 1997). I then used likelihood analysis and the Akaike information criterion with a correction for small sample sizes (AICc) (Burnham and Anderson, 2002) on the TBP and CBP with the same simulated trait data to identify which model (Brownian or OU) had greater support, and whether this corresponded to the model originally used, via the steps in Figure 2.2. For all simulations σ^2 (step variance) was set to 1, θ (the optimum trait value) was set to 0, and was the same as the ancestral trait value, α (strength of the

Figure 2.2: Flow chart of steps used for testing each method. The four methods are shown in four different colours, some of the steps apply to more than one method and the outline box for each shows this. Begin at the top of the chart and move down, where there is a black bar with x100 this indicates that the steps adjacent to this bar should be repeated 100 times.

selective constraint in an OU model) was set at 0.1 and 0.05 respectively for two different sets of simulations. I chose these α values to be biologically realistic for consistent comparisons between TBPs and CBPs. However, the ability to detect an OU model scales with tree height, so in order to make a comparison of TBP suitability across trees of different heights I also provide results of an analysis where α for each simulation corresponded to a phylogenetic half-life of $\frac{1}{4}$ of total tree height for each tree. In addition, I selected clades where the model used for simulation was consistently the model with greater support on both the CBP and TBP and recorded the parameter estimates for σ and α. This was to test not only whether the original model could be identified on the TBP, but that the values estimated for these parameters were representative of the parameter values estimated on the CBP. I chose to present AICc weights as this is the most commonly used procedure in palaeobiological studies, where often several models are being compared (Fusco et al., 2012; Sallan and Friedman, 2012; Lloyd and Friedman, 2013; Soul et al., 2013; Bapst, 2014a).

2.2.4.4 Phylogenetic clustering of extinction

Phylogenetic clustering of extinction is a measure of the strength of phylogenetic signal present in the timings of extinction of taxa. A time bin of interest is chosen and extinction is treated as a binary trait (i.e., the taxon goes extinct within the bin, or the taxon does not go extinct within the bin). I used the metric *D*, which was originally developed by Fritz and Purvis (2010) to measure phylogenetic signal strength of extinction risk in extant taxa, but it can be used to measure clustering of any binary trait (Hardy et al., 2012). This metric has been shown to be more informative and more robust to tree size and balance than other methods available to measure the phylogenetic structure of extinction. When a full timescaled phylogenetic hypothesis is available for the analysis *D* is the most powerful and consistent metric available (Hardy et al., 2012). The analysis is powerful for trees with at least 25 terminals (preferably more than 50) and if trait prevalence is not extreme.

In addition, performance is consistent only when phylogenetic resolution is greater than 70% (Fritz and Purvis, 2010). This analysis was performed with only the range data used to time-scale the trees, rather than simulated data. D assumes lower values when phylogenetic clustering is strong. A phylogenetically random trait distribution is expected to result in D=1, and D=0 is the trait distribution expected under Brownian motion. For each clade I chose the 20 Myr time interval with the most even trait distribution and measured D in that window.

2.2.4.5 Vital statistics

I collected a set of vital statistics for each clade in order to examine the degree to which these might co-vary with similarity between CBPs and TBPs: number of taxa, taxonomic resolution, taxonomic stratigraphic consistency index (Huelsenbeck, 1994), the number of ranks included in the taxonomic classification, the overall length of the tree and the average number of co-occurring lineages. Taxonomic resolution is a measure of the proportion of internal nodes that are bifurcating. Number of ranks is a count of the number of Linnaean ranks including any informal ranks the author had used (e.g. tribe, infraclass) encompassed by the taxonomy (e.g. a taxonomy that includes genus, tribe, family and order groupings has a score of 4). Stratigraphic consistency is a measure of how well the appearance and disappearance of taxa in the record matches their position in the tree. I performed regression analyses using these characteristics as independent variables to predict the values of summary statistics for the MC distance and PCM results. This provided a measure of which features of a taxonomy can be used to give an indication of whether it is likely to be suitable for use in analyses incorporating a PCM.

2.2.4.6 Extended range analysis

This analysis assessed the effect of using fossil data that includes singletons (taxa known from a single sampling horizon), versus taxa that have many occurrences through time and potentially better constrained first appearance dates. There are two possible causes of singletons: 1) the extinction rate of the lineage was high (i.e., its evolutionary duration was genuinely short), or 2) the lineage is poorly sampled. Many of the clades in our dataset are composed of singletons, and this is often the case with fossil vertebrate clades (e.g. Brusatte et al., 2008; Benson and Druckenmiller, 2014). For this reason there are few reliable estimates of sampling rates for vertebrate groups, as these estimates require some of the taxa to have more than one occurrence (although see Foote and Raup, 1996; Foote and Sepkoski, 1999; Friedman and Brazeau, 2011). However, if a clade was widely distributed, abundant, or lived in a place where preservation potential was high (e.g. shallow marine settings with a high sediment influx) higher sampling rates might be expected, leading to fewer or no singletons and better constrained divergence dates. This scenario is more in line with much of the invertebrate fossil record (Foote and Raup, 1996; Foote and Sepkoski, 1999). Many of the taxa in our datasets are singletons; to represent the alternative situation with a higher sampling rate including few or no singletons, I set the FAD of each taxon to the beginning of the stage in which it occurred, and set the sampled ‘population’ (the tip of each branch, at which trait values are measured) the end of the stage in which it occurred. This introduces more shared branch length between the CBP and TBP and is intended to represent a scenario in which there is good stratigraphic constraint on FADs and later occurrences. After simulating new trait data on these extended range CBPs, I performed the phylogenetic signal and phylogenetic generalised least squares analyses on them.

2.3 Results

Due to the publication of this work in a peer-reviewed journal, results from all simulations and analyses - including preliminary analyses and the data and R code used - were deposited in and are available to download from the Dryad data repository at http://dx.doi.org/10.5061/dryad.pt3qm. They are also contained in this thesis on the data appendix disk.

2.3.1 Tree Similarity

Taxonomy-based phylogenies (TBPs) are a good match for the most recent cladogram-based phylogenies (CBPs). The majority of matching cluster (MC) distances between the TBP and CBP are significantly ($\alpha = 0.05$) smaller than random with larger trees (more than 20 tips) showing a better-than-random match more consistently than smaller trees (Fig. 2.3). In addition there are strong significant correlations between the variance-covariance matrices of the TBP and CBP for each clade. 35 out of 52 clades have correlation coefficient (r) estimates of greater than 0.9, and the lowest value of r is 0.6937 for turtles. This shows that, as would be expected, taxonomies are good approximations of the underlying evolutionary relationships as recorded in inferred phylogenies.

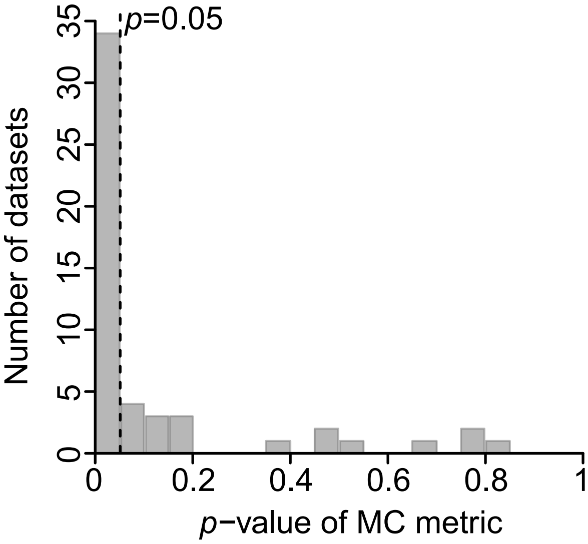


Figure 2.3: Matching cluster distance for overall tree similarity. Histogram of significance values for the matching cluster distance between the taxonomy-based phylogeny (TBP) and cladogram-based phylogeny (CBP). Values at less than p=0.05 show clades for which the TBP is significantly more similar to the CBP than random trees that have been time-scaled with the same fossil range data.

2.3.2 Phylogeny-based analyses

When considered together, the results in the following sections indicate that in some cases the variation in topology between the CBP and TBP is sufficiently small that the results of downstream analyses on a TBP are unlikely to be misleading, but can be highly misleading in others. There are some indicators of whether a taxonomy will be suitable for use in a PCM, which lead to some general heuristics, but it is not possible to propose definitive *a priori* criteria for isolating which taxonomies will be sufficient. The incorporation of taxonomies into phylogenetic comparative methods (PCMs) should be approached with caution, and assessed on a case-by-case basis.

2.3.2.1 Phylogenetic signal

Values for K obtained using TBPs (K_T) are significant ($p << 0.0001$, null hypothesis is a regression slope of 0) predictors of values for K obtained using CBPs (K_C) (major-axis R^2=0.6141, Fig. 2.4).

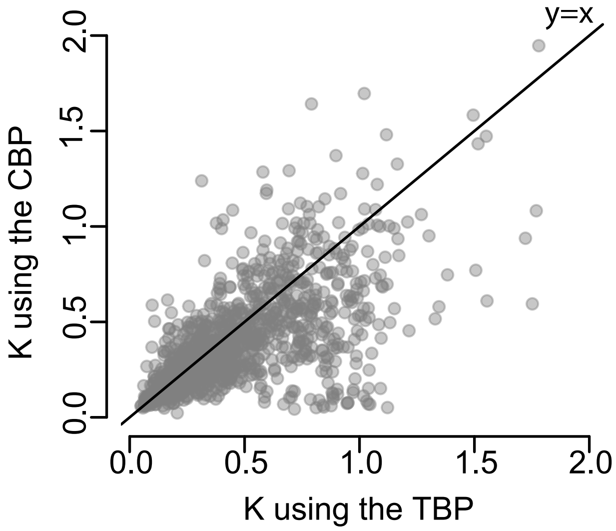


Figure 2.4: Results of the test of measuring phylogenetic signal on the TBP and CBP. This shows the results from all the iterations (tests on separate simulated datasets for each clade) for all the clades collectively. The line is where results lie if estimates of K made using the TBP match estimates of K made using the CBP.

When phylogenetic signal is strong, values of K_T are generally an underestimate of values of K_C (slope=1.3246), whereas when phylogenetic signal is weak, values of K_T are generally a small overestimate of values of K_C (slope=0.9000). p-values show that in most cases K_T is not significantly closer to K_C than K measured on random timescaled topologies of the same taxa (K_R) is to K_C (Fig. 2.5). This is true for all strengths of phylogenetic signal in the originally simulated data, although for data that is strongly phylogenetically structured the test statistic is significant more frequently, particularly for large trees. These high p-values are seen when K_R is also a good predictor of K_C. This issue is explained fully in the discussion but can occur because all topologies have been timescaled using the same range data.

2.3.2.2 Phylogenetic generalised least squares

Phylogenetic regression slope estimates can be obtained from phylogenetic generalised least squares analysis of trait data on the CBP and TBP. When CBP and TBP PGLS slope estimates are plotted against one another points lying along the line y=x would indicate that results from the TBP are not misleading compared to results from the CBP (Fig. 2.6a). Points generally lie close to the y=x line but the major axis regression slope is 0.7796, meaning that TBPs tend to produce phylogenetic regression slope estimates that are larger in magnitude than estimates from CBPs (Fig. 2.6a). Practically this indicates that on average, taxonomy-based analyses overestimate the amount of co-variation between traits, as would be expected given their comparatively poor resolution (Felsenstein, 1985). There is a significant coefficient of determination between the two estimates (R^2=0.6949, $p << 0.0001$). Phylogenetic regression slope estimates using TBPs are not always significantly better than regression slope estimates using time-scaled random topologies (Fig. 2.6b).

2.3.2.3 Mode of evolution

An Akaike weight of greater than 0.5 implies relatively greater support for that model. The weight for a particular model can be considered an estimate of the probability of that

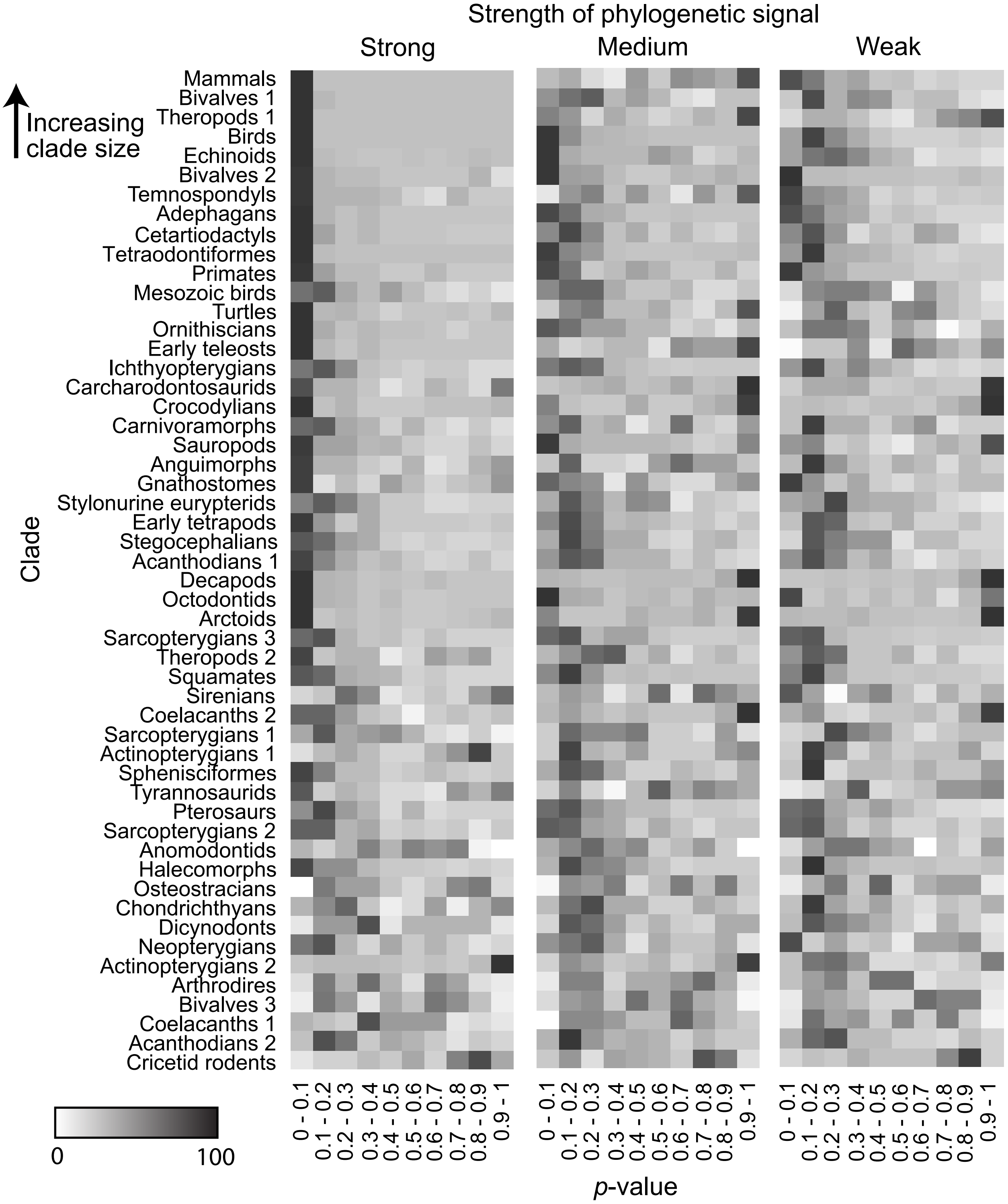


Figure 2.5: Heatmap of significance values for the difference between Blomberg's K measured on the TBP and measured on the CBP. Each row represents 100 simulations, the darker the square the more of those simulations had p-values (quantile of the test statistic, within the empirical distribution generated using timescaled random topologies) that fell within the range shown. The main feature to note is that when there is a strong signal, particularly for larger clades, there are low p-values, shown by the darkest area on the left hand side of the heatmap, particularly at the top.

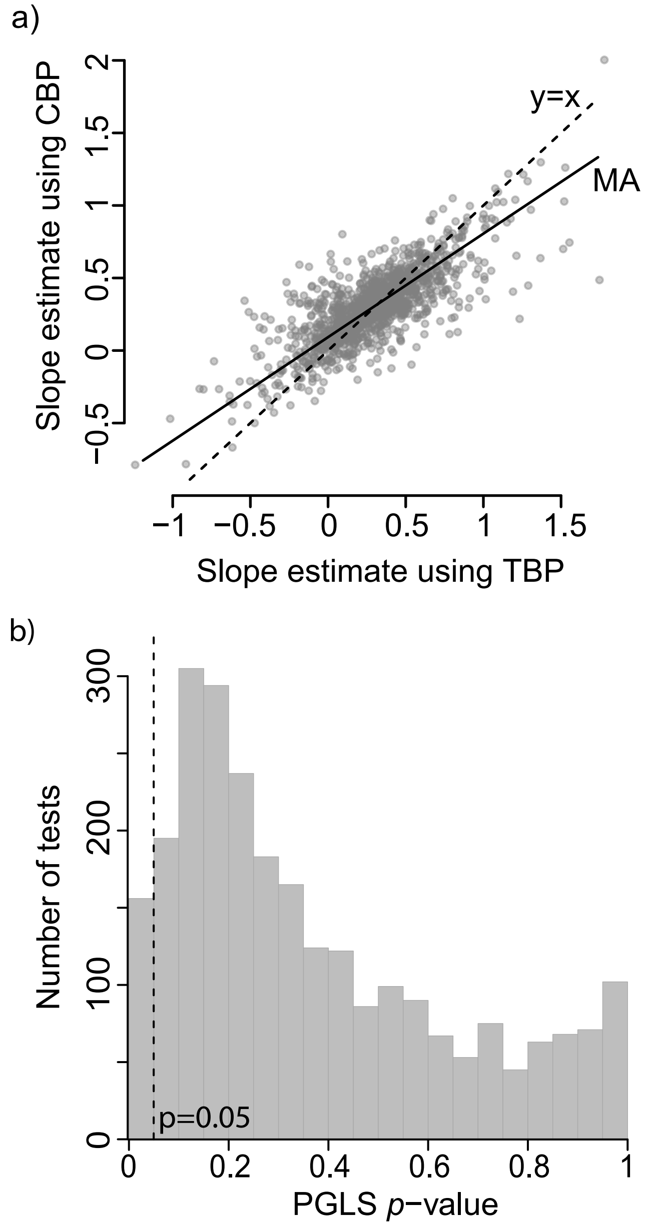


Figure 2.6: Phylogenetic generalised least squares. a) Correlation between estimated values of slope for PGLS on the TBP and CBP showing the major axis regression line (MA). b) p-values for the difference between the slope estimate using the TBP and using the CBP. These are collective results for all iterations of the test for each clade.

model being the best model for the data, out of the range of models tested (Burnham and Anderson, 2002; Bapst, 2014a). There was a broader range of model support values using the TBP than on the CBP for many data sets when the trait data were simulated under Brownian motion (BM) (Fig. 2.7). When trait data were simulated under an Ornstein-Uhlenbeck (OU) model, the range in model support values was generally the same as or smaller on the TBP than for the CBP (Figs. 2.8 and 2.9). When BM was simulated, analyses on the TBP had a tendency to incorrectly support an OU model whereas the CBP did not; this was common for larger clades and a reason for it is outlined in section 2.4.1 of the discussion (Silvestro et al., 2015). Note that as the models are nested there is a maximum possible Akaike weight for BM which depends on the number of taxa; this is clear from Figure 2.7. There were some clades where the result was directly misleading and the model used to simulate the trait data was the best supported model on the CBP but not the TBP. Examples, however, were rare (Fig. 2.7).

When OU had been used to simulate the data, number of taxa in the dataset was an important indicator of correct model identification, with OU consistently correctly identified on the CBP and TBP for clades including large numbers of taxa, but not for small clades where OU was rarely correctly identified on either tree (Figs. 2.8 and 2.9). There was a minor increase in the range of AICc weights estimated for the OU model with a smaller value of $\alpha = 0.05$, but it made no difference to the comparative performance of the TBP and CBP. In the analysis where α varied according to tree height, results were similar to OU results in the main analysis. However, there were some large clades where the TBP performed poorly compared to the CBP (Fig. 2.10). These clades were those that included many extant taxa (e.g. bivalves and modern birds); an explanation of this effect is provided in the discussion. Estimates of parameters for fitted models on the TBP were similar to – but usually slightly less accurate to original simulation parameter values – than estimates on the CBP, falling just below the line y=x. In many cases however, both trees gave parameter estimates that were not close to parameter values that were input to models when I simulated the trait data (Fig. 2.11). This was particularly a problem for estimates of α when fitting OU models (these are the points to the right side of Figure 2.11). A reason for this is outlined in the discussion.

2.3.2.4 Phylogenetic clustering of extinction

When all results are reviewed, only one of the datasets that I used falls within the acceptable range of tree size and trait prevalence outlined by Fritz and Purvis (2010) ($>$ 25 taxa in

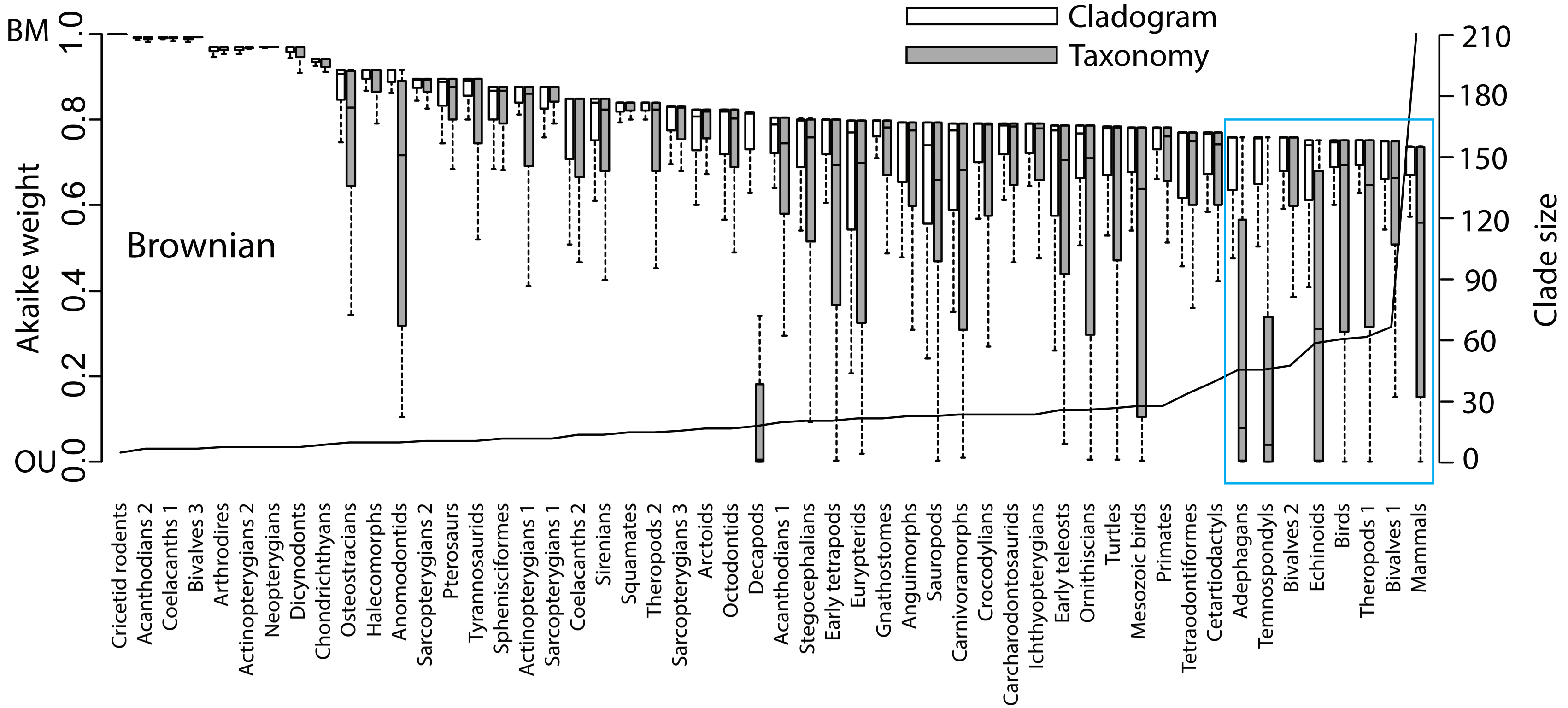


Figure 2.7: Results of analysis to fit a model of evolution to data which had been simulated under Brownian motion. Results from the CBP are shown as white bars, next to the results using the TBP for the same clade shown as grey bars. Size of each clade is plotted as thin black line and indicated on the right hand y-axis. Model support (measured as Akaike weight) for a Brownian motion (BM) model compared with an Ornstein-Uhlenbeck (OU) model, when traits were simulated using BM. Important features to note are that the top of the bars for each taxon show the maximum possible support possible for a BM model given the clade size, which decreases with clade size. In addition most of the grey bars showing TBP results are longer than their corresponding white bar showing CBP results. These features are explained in the main text in section 2.3.2.3. The blue box indicates examples of where results on the TBP tend to incorrectly support an OU model (grey bars are nearer the bottom of the plot) when results on the corresponding CBP did not (white bars are nearer the top of the plot).

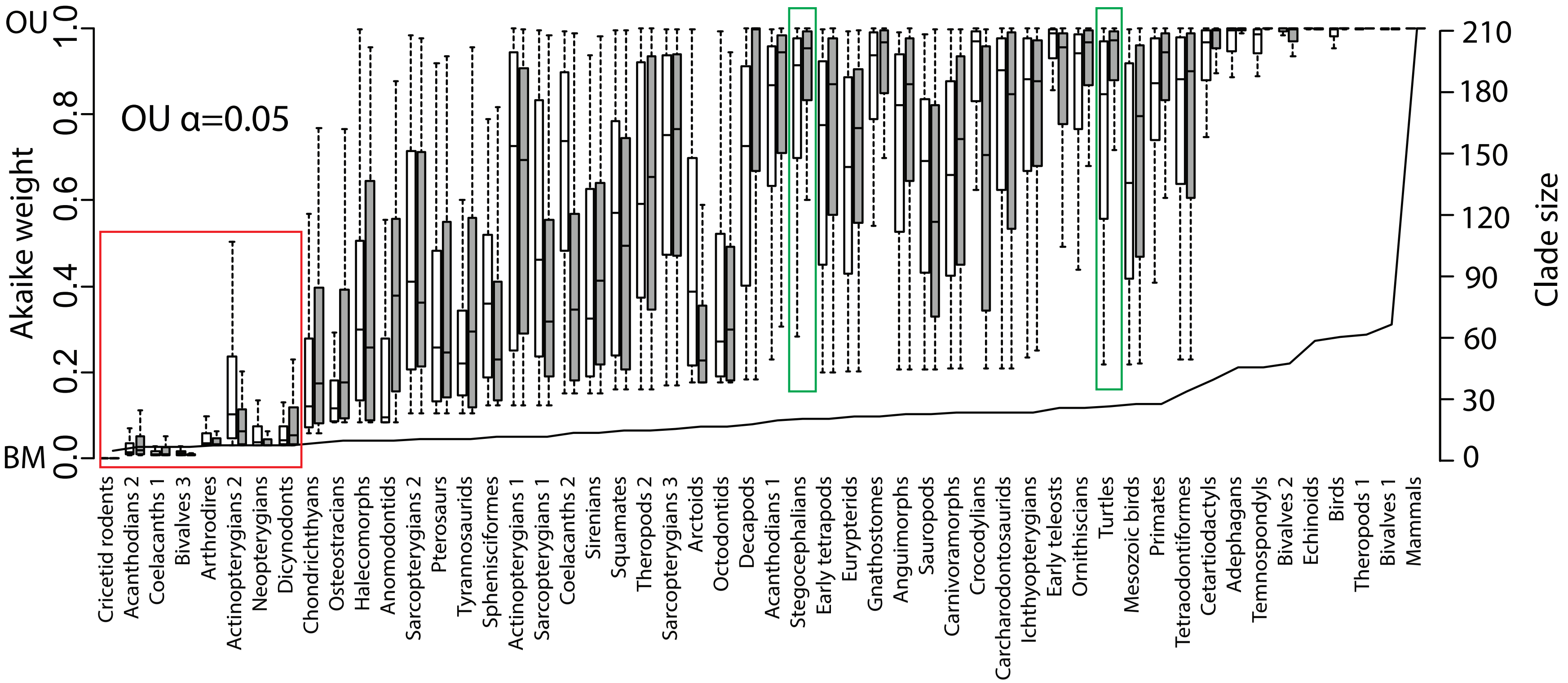


Figure 2.8: Results of analysis to fit a model of evolution to data which had been simulated under an Ornstein-Uhlenbeck model of evolution with $\alpha = 0.05$. Results from the CBP are shown as white bars, next to the results using the TBP for the same clade shown as grey bars. Size of each clade is plotted as thin black line and indicated on the right hand y-axis. Bars show the model support (measured as Akaike weight) for an OU model compared with that for a BM model for 100 sets of simulated data on the CBP. The red box shows where neither the CBP or TBP gave correct estimates of the simulated mode of evolution when used for the analysis. Green boxes indicate some examples of where the TBP provided correct estimates of mode more frequently than the CBP.

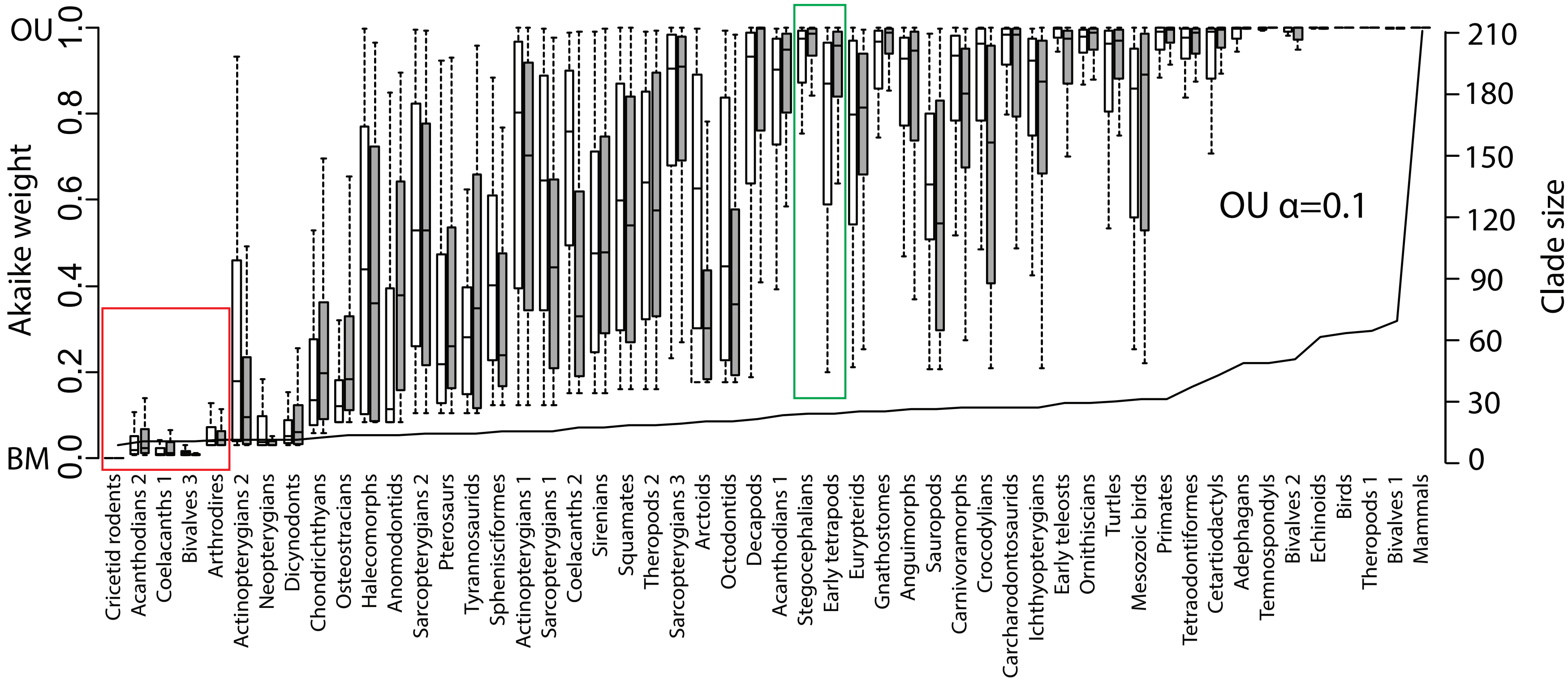


Figure 2.9: Results of analysis to fit a model of evolution to data which had been simulated under an Ornstein-Uhlenbeck model of evolution with $\alpha = 1$. Results from the CBP are shown as white bars, next to the results using the TBP for the same clade shown as grey bars. Size of each clade is plotted as thin black line and indicated on the right hand y-axis. Bars show the model support (measured as Akaike weight) for an OU model compared with that for a BM model for 100 sets of simulated data on the CBP. The red box shows where neither the CBP or TBP gave correct estimates of the simulated mode of evolution when used for the analysis. The green box indicates some examples of where the TBP provided correct estimates of mode more frequently than the CBP.

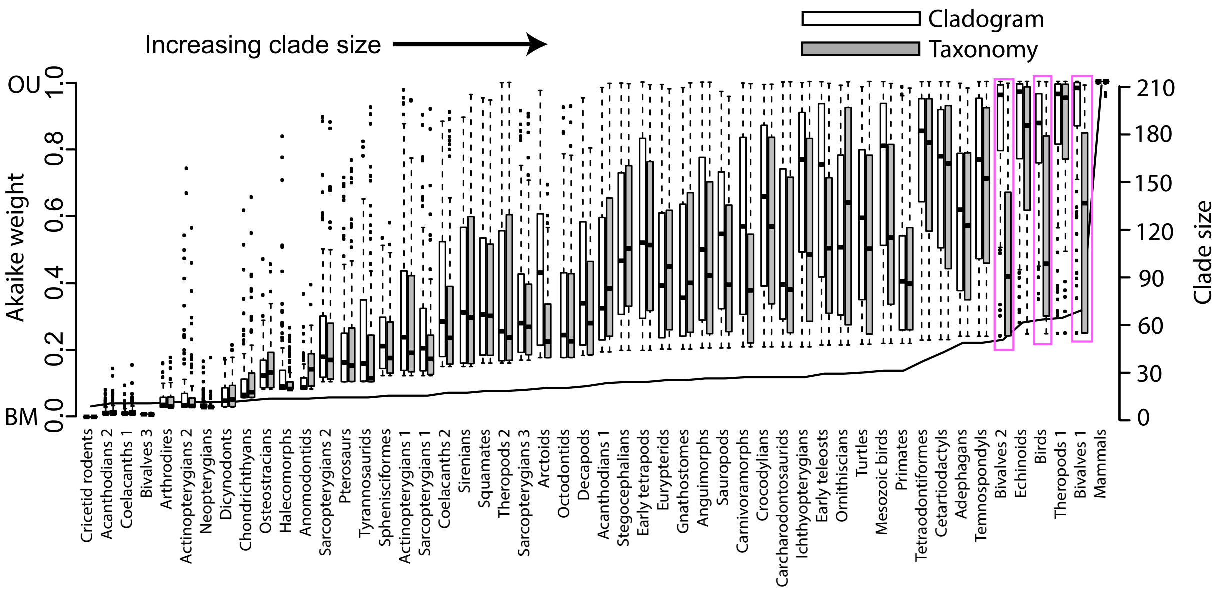


Figure 2.10: Results of analysis to fit a model of evolution where trait data were simulated using an OU model with a value of α that varied proportionally with tree height. Results from the CBP are shown as white bars, next to the results using the TBP for the same clade shown as grey bars. Size of each clade is plotted as thin black line and indicated on the right hand y-axis. Bars show the model support (measured as Akaike weight) for an OU model compared with that for a BM model for 100 sets of simulated data on the CBP. The pink boxes show large clades where the TBP performed poorly compared to the CBP.

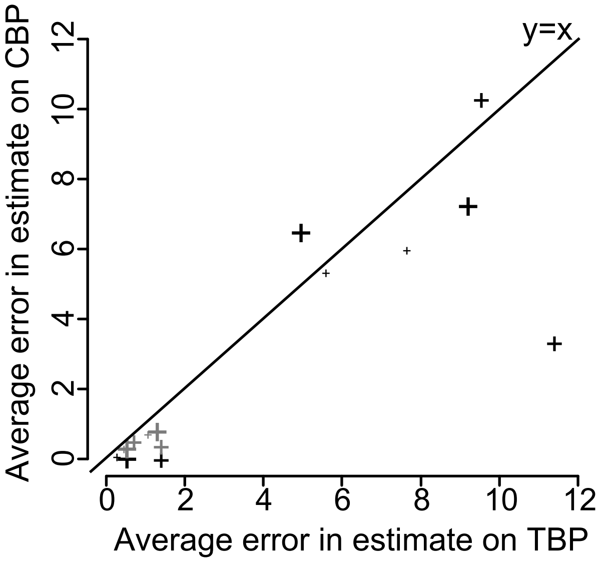


Figure 2.11: Estimated parameter values when fitting an OU or BM model for three of the largest completely extinct clades in the dataset (Large crosses: theropods; medium crosses: temnospondyls; small crosses: eurypterids). Estimates of α are shown in grey and estimates of σ are shown in black.

the timeslice and non-extreme trait prevalence between 0.2 and 0.8). Many of my datasets contain a low number of tips the following removal of taxa that are not shared between the cladogram and taxonomy. Timeslicing also reduces the number of taxa included in the

Table 2.3: Summary of regressions of vital statistics against analysis results

Summary result	Number	Resolution	SCI	Ranks	Length	Lineages
MC distance	0.0488	-0.0091	0.0513	0.2721	0.0215	0.0378
PGLS R^2	0.1223	0.3091	0.0326	-0.0095	-0.005	0.0564
PGLS p-value	0.0198	-0.0115	-0.02	0.0828	-0.0101	0.0201
K weak R^2	0.1350	0.1946	-0.0199	-0.0192	-0.0029	0.0708
K weak p-value	0.0249	0.0425	0.1221	0.1461	0.0617	0.0107
K medium R^2	-0.0196	0.0188	0.0589	0.1249	0.1001	-0.0195
K medium p-value	-0.0194	0.0295	0.1090	0.0601	0.0355	-0.0200
K strong R^2	0.0117	0.0252	0.1241	0.2439	0.134	-0.0087
K strong p-value	0.1607	0.1192	0.0992	0.1601	-0.0194	0.1360
BM Akaike weight	0.1095	0.0677	-0.0188	0.0223	0.0157	0.0694
OU1 Akaike weight	0.0584	0.0385	0.2303	0.1773	0.0145	0.0276
OU2 Akaike weight	0.0412	0.0767	0.1712	0.1492	0.0147	0.0114
BM σ value R^2	0.1350	0.1946	-0.0199	-0.0192	-0.0029	0.0708
OU1 α value R^2	-0.0200	0.001	0.0077	-0.0199	0.0404	0.0129
OU1 σ value R^2	-0.0197	0.0062	-0.0045	-0.0200	0.0484	0.0160
OU2 α value R^2	0.0164	0.0524	-0.0194	-0.0172	0.0441	0.0245
OU2 σ value R^2	0.0201	0.1170	-0.0177	-0.0199	0.0523	0.0269

R^2 values from regressions of summary results for each analysis against vital statistics of trees. The vital statistics are outlined in section 2.2. Summary results are for each analysis I performed, outlined in section 2.3. Abbreviations and symbols are MC = matching cluster; PGLS = phylogenetic generalised least squares; R^2 = adjusted coefficient of determination; K = Blomberg's K estimate of strength of phylogenetic signal; weak, medium and strong indicate the originally simulated phylogenetic signal strength; BM = Brownian motion model; OU1 = Ornstein-Uhlenbeck model with α=0.1; OU2 = Ornstein-Uhlenbeck model with α=0.05; α = strength of the selective constraint in an OU model; σ = step variance.

measurement, because only those lineages occurring within a given stratigraphic window are considered. There is good agreement between the result on the TBP and CBP for the one clade that is suitable for analysis (mammals), but further large datasets would need to be found in order to confirm the utility of taxonomy for such analyses.

2.3.2.5 Vital statistics

Some of the vital statistics of sampled phylogenies showed a small (up to $R^2 = 0.3480$) but significant correlation with the degree to which TBPs and CBPs matched for the various analyses. However, the best indicators varied between analyses, and no one statistic was consistently good for all the analyses. Table 2.3 shows which vital statistics are the best

Figure 2.12: Estimated values of phylogenetic signal (K) with CBP and TBP timescaled using extended ranges that include the whole of the stage in which the taxon occurred.

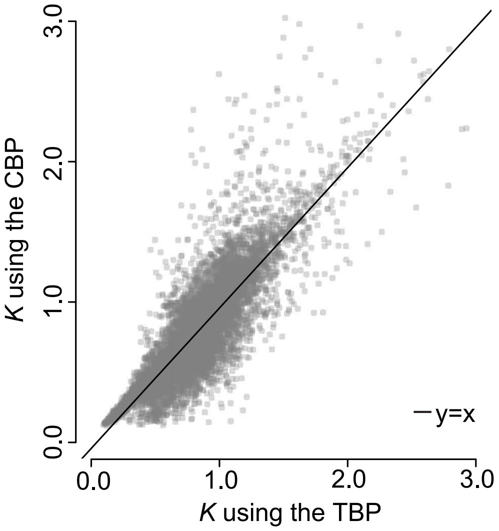
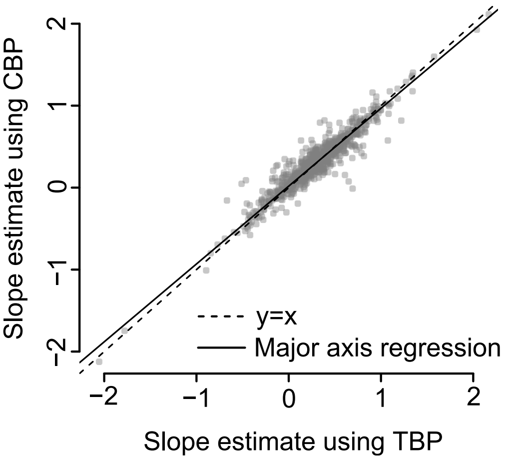


Figure 2.13: Estimated values of PGLS slope with CBP and TBP timescaled using extended ranges as in previous figure.



explanatory variables for each analysis, although in many cases there is little difference between them. All the characteristics I tested can be easily obtained from a TBP. Where there is a good correlation between a summary result and a vital statistic, larger values for vital statistics are associated with a better match between the two tree types. No one vital statistic, or indeed particular combinations of them, can be considered a definitive indicator of how well a taxonomy is likely to perform due to the low R^2 values, but if a taxonomy has high values for several of them it is more likely to produce a phylogenetic hypothesis that closely matches that derived from a formal cladogram, and thus produce similar results in an analysis.

2.3.2.6 Extended range analyses

When the ranges used were extended to include the whole of each stage in which a taxon occurred, the CBP and TBP yielded comparable results (Figs. 2.12 and 2.13). Results were a close match between the TBP and CBP more often than in the main analyses where occurrence times were selected randomly from within the stage range of each taxon. Variation in K_T explained much of the variation in K_C (R^2=0.7065). For PGLS I found a major axis regression slope of 0.9497 (R^2=0.9330) in a plot of phylogenetic regression slope estimates from the CBP, against slope estimates for the same data on the TBP. This means that TBPs tend to produce phylogenetic regression slope estimates that are slightly larger than estimates from CBPs, but are close to them.

2.4 Discussion and guidelines for incorporating taxonomy

Some of the taxonomies tested in this study can be used as a substitute for an explicit phylogenetic hypothesis, to make phylogenetically informed measurements without misleading results being produced, relative to the cladogram-based phylogeny (CBP). Others consistently performed poorly compared with the CBP, and produced results which were misleading when compared to known parameters used to simulate the data. In some cases retrieval of originally simulated values was as poor as or worse on the CBP than on the taxonomy-based phylogeny (TBP). Generally, the larger the tree the more consistent (although not always accurate) the results were, but the number of taxa included in the phylogenetic framework did not correlate with any measure of the adequacy of performance of a TBP, nor did any other characteristic of the trees. This result makes it difficult to establish or suggest general circumstances under which taxonomy might be used in PCMs. The same rigorous standards that should be used when deciding whether it is appropriate to incorporate a cladogram produced using formal phylogenetic inference into an analysis must be applied here too, and each case must be assessed on an individual basis.

2.4.1 Supertrees

A context in which taxonomic information is commonly used in the current literature is for supertrees where taxonomy is used as the 'seed' or 'backbone' for the supertree analysis (Fusco et al., 2012; Benson et al., 2014a). The majority of supertrees in the palaeontological literature are composed exclusively of cladistically inferred phylogenies and any taxonomic relationships that are incorporated are well established relationships between higher level groups. These results therefore do not suggest that using supertrees in PCMs is any more problematic than using a fully cladistically inferred tree.

Small trees generate misleading estimates using both the CBP and TBP for many of these analyses, suggesting that they require large sample sizes to be effective (Münkemüller et al., 2012). Supertrees are often used to increase taxonomic coverage of phylogeny to include more taxa. The largest taxonomy-based phylogenies performed well in analyses, again suggesting that supertrees that incorporate some taxonomic information should not be especially problematic compared to cladistically inferred trees. However, a possible exception relates to estimation of evolutionary mode. Ability to detect an OU model scales with tree height, larger trees are on average longer in this dataset, providing greater power to detect an OU model. This contributes to the higher success of correctly identifying OU evolution on larger trees. However, my results indicate that large taxonomy-based phylogenies are biased toward supporting an OU model when BM is used to simulate trait data. This makes intuitive sense as inaccuracies in TBP topology compared to the CBP could lead to tips that are sister taxa with similar trait values on the CBP having a greater phylogenetic distance between them on the TBP. Taxa with similar trait values but a larger phylogenetic distance match an OU model, where pull towards an attractor causes this pattern. In this study, bias towards an OU model is associated with topological and branch length error, but the same principle has been demonstrated by other workers for measurement error generally (Thomas et al., 2014; Silvestro et al., 2015).

This result may also have been amplified by the poorer resolution of the taxonomies.

Although the method required a fully bifurcating tree with no ZLBs, only a short length was added to the internal branches introduced within each TBP. Traits simulated under Brownian motion on the CBP would therefore have a greater co-variance than the expectation under Brownian motion evolution on the TBPs, where short internal branches lead to low co-variance. In an OU model, change in traits towards an attractor value removes co-variance over time, which matches the expectation for a Brownian motion model on a topology with short internal branches. The analysis has less power to detect OU on smaller trees, so this effect occurred less often for the clades with a low sample size.

2.4.2 Biases in individual analyses

In the mode of evolution analysis where α varied proportionally to tree height I found that when clades contained many extant taxa, analyses on the TBP were less likely to correctly support an OU model than analyses on the CBP. Taxonomies contain additional polytomies, which in these datasets may contain several extant taxa as well as older extinct taxa, causing the divergence of extant taxa within the polytomy to be dragged back in time. This causes a very ‘tippy’ tree shape where internal branches are short and terminal branches are long. If trait data are simulated via OU then they have comparatively little covariance structure. Trait data with little covariance structure also fits a Brownian motion model of evolution on the tippy topology described above, meaning that an OU model does not have additional explanatory power compared to a Brownian motion model on these trees.

Some of my results (PGLS slope and estimates of Blomberg’s K) are not statistically significant when compared to an empirical distribution of results from random topologies (Figs. 2.6 and 2.5). These statistics are based on change in continuously evolving traits, the values of which are directly related to branch lengths of the phylogeny on which they are simulated. This branch length information (derived from occurrence data) is inherent in every topology I subjected to analysis, making it a key factor in the estimates using both TBPs and random trees. p-values for K and PGLS slope using TBPs were calculated by comparison with an empirical distribution of estimates made using time-scaled random trees. Therefore when there were good estimates from random trees, TBP p-values could be high and spanned the full range of 0-1 (Fig. 2.6b). As the phylogenetic signal becomes greater, and the tree larger, the difference in topology between the TBP and random trees becomes more important and the TBP is more frequently a significant match for the CBP (Fig. 2.5).

2.4.3 Stratigraphy

These results, in combination with the results of the extended range analysis, highlight that stratigraphic control on the shared branch length between a CBP and a TBP is an important factor in their relative performance. With this sample of datasets, for analyses when ranges were constrained to begin and end at stage boundaries (resulting in 'time-stretched' terminal branches), the influence of differences in tree topology on the variance-covariance structure was small. In this scenario the majority of trait change occurred on the large shared portions of the trees, which were controlled by the stratigraphy. In the main analyses, differences in tree topology had a larger influence on the variance-covariance structure and so results from the CBP and TBP showed greater difference. In practise it is possible that a high sampling rate (resulting in better constraint on taxon durations) could lead to incorrect branch length extensions when topology is misspecified. However, my results indicate that the advantage of good stratigraphic control on first appearance dates and the ages of sampled populations overrides the influence of this problem on results.

This observation is highlighted by the datasets for which more than one cladistic hypothesis was compared to a particular taxonomy (Acanthodians, Actinopterygians, Bivalves, Coelocanths and Sarcopterygians). For these multiple comparisons, the cladistic hypotheses varied in age of publication (and therefore presumably accuracy) and yet there was not a systematic difference in how well results from the TBP matched results from the CBPs for these clades. Taxonomies were not necessarily a better match for older cladograms. Essentially, when the stratigraphic control on branch lengths was good, small variations in topology between successive cladistic analyses were not important.

Many of the clades I tested have a sparse record, and the TBP did not yield comparable results to the CBP. However, many of the clades for which there are not yet cladistically inferred phylogenies available have higher sampling rates with better stratigraphic constraint on occurrences and divergence dates (e.g. brachiopods and crinoids (Foote and Sepkoski, 1999)). They should therefore be expected to provide results less influenced by inaccuracy in topology.

2.4.4 Guidelines

Bearing all these factors in mind, workers wishing to incorporate taxonomy into a framework for making phylogenetically informed measurements should carefully consider whether the clade in question is in fact suitable for this use. If many new fossils in the clade are still being discovered and are leading to changes in classification, then the taxonomy is unlikely to give results that would match results from a cladistically inferred phylogeny of the group. Stable taxonomies too may be problematic if they have remained stable due to lack of recent rigorous study, rather than because extensive systematic work has led to a high level of confidence in the proposed evolutionary relationships. In addition to consideration of the vital statistics of the taxonomy, consultation with a systematist familiar with the group who can indicate the maturity of the taxonomy would aid in ascertaining its suitability. If the evolutionary relationships proposed within the group are well established and supported by continued study, the ranges are accurately known, and the dataset is large, well resolved and stratigraphically consistent then use of a taxonomy can be considered. This supports some of the previous studies where a taxonomic classification has been applied (e.g. Roy et al. (2009); Hardy et al. (2012)), since the biological groups in question—ammonites and bivalves—appear to be ideal candidates for such an approach.

Here I used pre-cladistic taxonomies to test a ‘worst case scenario’. Modern taxonomies are often phylogenetically informed, at least in part, and in theory should therefore give a result even closer to the ‘correct’ one. Taxonomies which include phylogenetically informed relationships should therefore be preferentially used if available.

Each taxonomy should be assessed on a case-by-case basis as it is clear that they are not always suitable for incorporation into a PCM. The following checklist of criteria for candidate classifications can be used as a guide to appropriate use of taxonomies in PCMs:

1. It is large (no less than 20 tips, usually more depending on the method), preferably in combination with a high number of taxonomic ranks, high resolution and good stratigraphic consistency.
2. It contains long established taxa that have been stable for a significant period of time and are supported by continued study.
3. The fossil taxa are from a range of time intervals, not a small number of key sampling horizons.

If the most recent estimate of phylogeny is closer to the true course of evolution than a taxonomy, it should produce results that are more reliable. Palaeobiologists work under the assumption that cladistically inferred palaeontological phylogenies provide estimates of phylogenetically informed measurements that reliably represent reality. If this is the case, my results demonstrate that in some cases taxonomic information can also provide suitable estimates of these measurements.

Although use of taxonomy must be carefully considered, these results open up certain previously inaccessible sections of the fossil record to interrogation within an explicit comparative framework, which will help to test many classical macroevolutionary hypotheses that have been based on groups for which formal phylogenetic hypotheses remain lacking.

Chapter 3

Phylogenetic clustering of extinction in tetrapods across the Permo-Triassic boundary

> *"We have experienced, so far, only 0.0001 percent of the history of life. In fact, it is rather arrogant to think that our 0.0001 percent of life's history should yield a complete sample of nature's processes. Perhaps the past is actually key to the present (and future)."*
>
> - David M. Raup *Bad Genes or Bad Luck?*

In this chapter I measure the phylogenetic clustering of the timings of extinctions in three major tetrapod clades (synapsids, sauropsids and temnospondyls) for a timespan encompassing the Permo-Triassic boundary. I do so in order to test phylogenetic methods for understanding extinction in the hope that such methods can provide a means by which a variety of types of data can be compared. I test how robust the results are to variation in factors such as cladogram timescaling methods and the metric used to measure the clustering of extinction. Finally I discuss these results in the context of changes in species richness in each clade at the time, and the response of fauna to mass extinction events.

3.1 Phylogeny and extinction

Palaeontological and modern data on extinction are difficult to combine due to differences in timescale, the volume and quality of available data, approaches to analysis, and the intensity with which different fossil and living groups have been studied (Jablonski, 2008a; Fritz et al., 2013; Slater and Harmon, 2013; De Vos et al., 2014). These problems are especially acute for vertebrate groups, which are of considerable interest to neontologists but often have a poorly sampled palaeontological record (Foote and Raup, 1996; Foote and Sepkoski, 1999). Despite these problems the fossil record offers a natural laboratory for testing our hypotheses about how extinction dynamics might change or be maintained in times of extreme ecological stress (Jablonski, 1994, 2005). This deep-time perspective is becoming increasingly important as extinction rates increase and biodiversity declines (McKinney, 1997; Erwin, 2009; Barnosky et al., 2011).

Phylogeny (a hypothesis of the precise evolutionary relationships between taxa) provides a framework that permits assessment of macroevolutionary patterns in extinction in the geological past. This approach avoids many of the biases introduced by imperfect sampling, while simultaneously providing results that can be directly compared to those from analyses of living and recently extinct taxa (Purvis, 2008; Fritz et al., 2013; Harnik et al., 2014). Few previous studies of extinction in the fossil record have considered phylogeny (but see Smith and Roy, 2006; Friedman, 2009); the majority have targeted marine invertebrate clades for which formal cladistic solutions are as yet unavailable. However, it is now recognised that incorporating an estimate of phylogeny into analyses of the fossil record is often vital in obtaining meaningful results (Felsenstein, 1985; Grafen, 1989; Norell, 1992; Rabosky, 2010).

Intuitively, we would expect that extinction is selective with respect to the relationships between taxa, based on the idea that some traits may make taxa vulnerable or resistant to extinction, and these traits might be phylogenetically conserved. Early authors speculated on the possibility of taxonomically selective extinction, mainly based on observations of the complete disappearance of clades (e.g. non-avian dinosaurs, non-lingulid brachiopods). In a review of the then current literature Raup (1981) began analytical discussion of the problem, coining the term ‘field of bullets extinction’ for extinction that is truly taxonomically random. In a series of simulation studies he investigated whether various non-selective models (for example random birth-death or biogeographic) were consistent with the observed proportion and taxonomic impact of extinction in the fossil record (Raup, 1982, 1991). These studies indicated that at particular taxonomic levels none of these models produced simulation outcomes consistent with the proportion of higher taxa actually lost during mass extinction events.

Since then it has been shown that extinction risk in the modern biota is selective with respect to several different biological, life history or emergent traits (e.g. geographic range size) (Bielby et al., 2006; Cardillo et al., 2008; Turvey and Fritz, 2011; Harnik et al., 2012b). Ideally it would be possible to use the deep-time fossil record to identify traits which may confer extinction resistance, or leave taxa vulnerable (Jackson and Erwin, 2006; Purvis, 2008; Fritz et al., 2013). Some high-resolution fossil records have been used to investigate selection against a particular trait or vulnerability to a particular threat. Previous studies have shown extinction selectivity related to body size (Harnik, 2011; Tomiya, 2013), feeding strategy (Jeffery, 2001), geographic range (Kiessling et al., 2007; Payne and Finnegan, 2007; Jablonski, 2008b), morphology (Liow, 2007; Friedman, 2009) or clade richness (Smith and Roy, 2006), among others. However, the majority of these studies lack a phylogenetic context (although see Smith and Roy, 2006; Friedman, 2009). In addition these traits frequently co-vary, making untangling the various factors difficult, and analyses of modern taxa have shown that many non-fossilising traits can also be important in extinction selectivity (Purvis, 2000; Purvis et al., 2005; Cardillo et al., 2008; Harnik et al., 2012b).

Assessment of the phylogenetic clustering of extinction itself represents a potential solution to these issues (Purvis, 2008; Roy et al., 2009; Hardy et al., 2012). This is a measure of the phylogenetic signal in the timings of extinctions among taxa. It is based upon the assumption that although extinction itself is not an evolving trait, it does depend upon a series of underlying continuously evolving traits (Fritz and Purvis, 2010). If taxa that are closely related are more likely to become extinct at the same time across the phylogeny, this results in a strong signal of phylogenetic clustering (Fig. 3.1). Different extinction drivers, or types of environmental stress, are likely to cause selective extinction of taxa with phenotypic traits that render them vulnerable to that specific threat. If these traits are conserved over a phylogeny then clustering will be high. This measurement can provide a more nuanced picture of extinction, beyond rate increase or decrease through time, and can act as a proxy for non-fossilising extinction-linked traits, particularly if those traits co-vary strongly (Lockwood et al., 2002; Jablonski, 2005; Roy et al., 2009; Fritz and Purvis, 2010; Hardy et al., 2012). It can be measured for any phylogeny in both the Recent and the

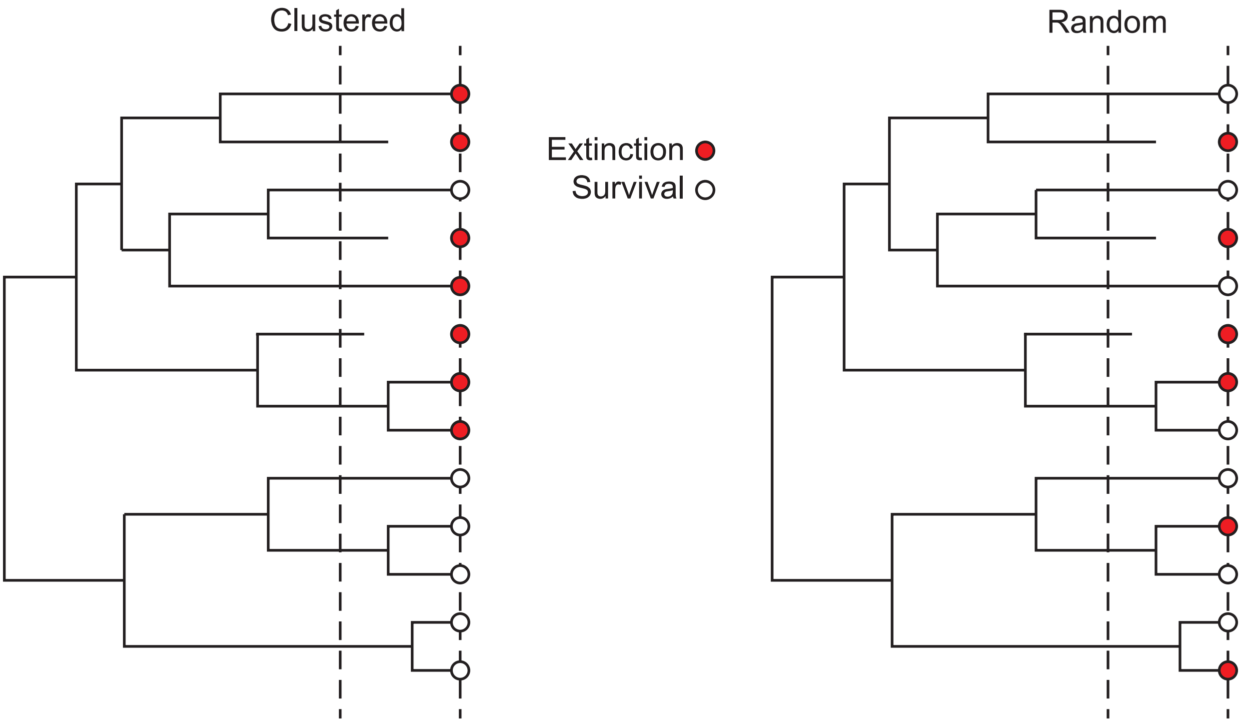


Figure 3.1: Hypothetical phylogenies as examples of phylogenetically clustered (left) extinction, and phylogenetically random (right) extinction. A measurement is made for each timeslices, which are shown by dashed lines. An extinction is any that occurs during or at the end of the timeslice, a survival is any taxon which survives through the end of the timeslice.

geological past. It has the advantage of explicitly correcting for the shared evolutionary history between taxa, preventing false interpretations of data. Variation in clustering has begun to be explored in some fossil clades (e.g. bivalves and ammonites: Roy et al., 2009; Krug et al., 2010; Hardy et al., 2012), but not in any vertebrate or terrestrial groups.

3.2 Extinction at the Permo-Triassic

As the Earth's largest mass extinction, the Permo-Triassic event (P/Tr) has received a great deal of research attention. There have been numerous studies measuring changes in richness, and diversification rates, in the periods leading up to and following the boundary (Ruta and Benton, 2008; Lucas, 2009; Ruta et al., 2011). A great deal of previous work focussed on marine invertebrates (Alroy et al., 2001; Benton and Twitchett, 2003), but more recent studies have included terrestrial tetrapods, particularly those from the Karoo of South Africa (Retallack et al., 2003; Rubidge, 2005; Ruta et al., 2011; Fröbisch, 2013). This work provides a more detailed picture of the ways in which different groups were affected during this time period, with a key aim of understanding the mechanisms – not just the patterns – of macroevolutionary change (McGhee et al., 2004; Fröbisch, 2008). There have been varying interpretations of the data that have been collected to elucidate the effect of the P/Tr on tetrapods. Several workers have suggested a huge drop in richness of terrestrial tetrapods at the boundary, with almost complete loss of tetrapod fauna in some areas, and complete ecosystem collapse (Benton et al., 2004; Ruta and Benton, 2008; Bernard et al., 2010; Fröbisch, 2013). At the other extreme is a complex story of prolonged replacement of some tetrapod clades with others; the rising dominance of archosaurs, long-term declines in richness before the boundary itself and several smaller scale extinctions in various clades (Lucas, 2009; Ruta et al., 2011).

New robust phylogenetic hypotheses and comprehensive fossil occurrence data are available for many tetrapod groups during this interval (e.g. Ruta et al., 2007; Nesbitt, 2011; Brocklehurst et al., 2013). Therefore the tetrapod fossil record in the period up to, during, and following the P/Tr mass extinction represents a unique opportunity to test hypotheses relating to extinction selectivity in a quantitatively rigorous way. These data permit analyses on fully time-scaled phylogenies, rather than on taxonomies or ultrametric trees as has been the case in previous studies (Roy et al., 2009; Hardy et al., 2012).

Although a great deal of work has been done, many questions relating to responses in diversity of taxa and the nature of extinction selectivity during mass extinctions remain (Harnik et al., 2014). In order to begin to address them I measured the phylogenetic clustering of extinction in three fossil tetrapod clades (non-mammalian synapsids, sauropsids and temnospondyls) that are the extinct relatives of major extant clades (mammals, reptiles and amphibians respectively). The analysis spans a period from the late Carboniferous to the Late Triassic, which includes a time of extreme ecological stress as well as times of comparative stability. I present sensitivity tests to demonstrate whether the results are robust to various factors, including the method used to measure clustering of extinction, and the method used to scale the branch lengths to time. I assess whether my results are consistent with, or contradictory to the statements outlined below. Here I also provide an explanation of the way in which my results contribute to the assessment of these statements.

- The phylogenetic clustering of extinction is a general rule across time and clades. These measurements have not previously been investigated in vertebrates, or for the end Permian. My results provide a dataset which can be compared to previous research on invertebrate taxa in different time periods.

- Mass extinctions represent a fundamentally different extinction regime to background times. My analysis includes the largest mass extinction as well as low extinction intensity periods, facilitating comparison of the two.

- Phylogenetic patterns in extinction can be maintained over the long term. My dataset has a long timespan, providing a comparison with the variation in extinction clustering seen on shorter time scales which have previously been investigated using invertebrate taxa.

- A measurement of phylogenetic clustering of extinction can be made using comparatively poor resolution data from terrestrial vertebrate taxa and these measurements are robust to variation in the method used to obtain them. By running sensitivity tests for my results I establish whether or not performing this analysis on the quality of data available for terrestrial vertebrate groups is valid and assess how robust the results are to various input parameters.

3.3 Methods

I quantified the phylogenetic clustering of extinctions in the fossil record of three major tetrapod clades (sauropsids, temnospondyls and non-mammalian synapsids) using two different metrics: Fritz and Purvis' *D* (Fritz and Purvis, 2010) and Moran's *I* (Moran, 1950; Gittleman and Kot, 1990). The period of time over which I measured these metrics ran from the Late Carboniferous to the Late Triassic, divided into ten timeslices of similar length comprising one or two stages. I performed sensitivity analyses by varying the length of timeslices and the method of scaling phylogenetic branch length to time. All analyses were performed in R (R Core Team, 2015).

3.3.1 Data

I selected the most recently published phylogenetic hypotheses for each clade. The topology for temnospondyls is a supertree taken directly from Ruta et al. (Ruta et al., 2007). The topologies for sauropsids and synapsids are composite trees constructed by combining higher-level topologies for each clade, with the most recently available species-level topologies from various other studies. Source phylogenies are detailed in Table 3.1 and example scaled topologies are shown in figures 3.3, 3.4 and 3.5. The full set of 100 trees used for the analysis is included on the data appendix disk. Occurrence data for each taxon were taken primarily from the Paleobiology Database (PaleoBioDB.org), except where these data were available from the author of the published topology (Ruta et al., 2011, - Parareptiles).

Table 3.1: Clades used for composite trees

Clade	**Source Publication**
Temnospondylii	Ruta et al. (2007)
Sauropsida	
Parareptilia	Ruta et al. (2011)
Captorhinidae	Reisz et al. (2011a)
Basal Diapsida	Reisz et al. (2011b)
Archosauramorpha	Nesbitt (2011)
Lepidosauramorpha	Jones et al. (2013)
Basal Sauropterygia	Rieppel (2000)
Pistosauria	Benson and Druckenmiller (2014)
Synapsida	
Basal Synapsida	Brocklehurst et al. (2013)
Basal Therapsida, Therocephalia	Huttenlocker (2009)
Cynodontia	Liu and Olsen (2010)

Separate clades incorporated into each phylogenetic topology used and the publication from which each was obtained.

3.3.2 Time-scaling

The dates of occurrences of fossil taxa in tetrapod groups during the Paleozoic and Mesozoic are usually only known to stage level resolution. To account for uncertainly in the actual times of appearance and disappearance of taxa in the record, a set of 50 stochastically generated fossil ranges was made for each taxon. First and last appearances were selected from a uniform distribution between the beginning and end of the stages from which each taxon is known. The analysis required phylogenies with branch lengths scaled

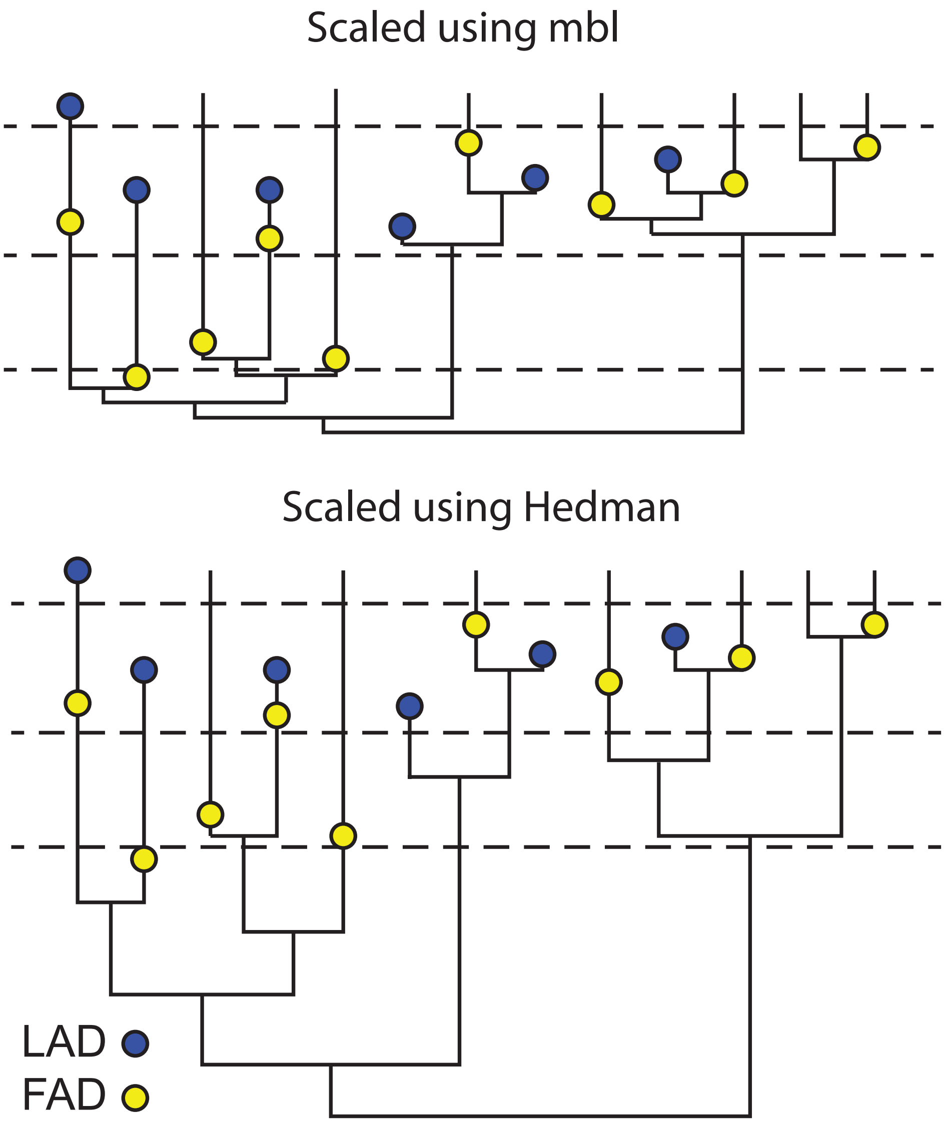


Figure 3.2: The effect on a phylogeny and the extinction clustering measurement of using different timescaling methods. The first and last appearances of taxa are shown by yellow and blue circles respectively (i.e. the first and last recorded instances of fossils that have been collected and assigned to the taxon). Hypothetical branch lengths corresponding to common features of phylogenies scaled using the *mbl* (top) and *Hedman* (bottom) methods are shown. Note that the *Hedman* method tends to extend branches further back in time, pushing back divergences and increasing the number of taxa crossing timeslice boundaries, in other words it means there are more survivals in each timeslice than there are when the *mbl* method is used.

to time, based on each set of fossil ranges. This was done using the Hedman algorithm (Hedman, 2010), which provides a distribution of estimates for the position of each internal node in the tree, based on the ages of the earliest representatives of the next taxa toward the root of the tree. This is a new method which I performed in R using code written by Graeme Lloyd and provided online. Through simulation it has been shown that this method produces results largely in line with results from other recently developed methods for time-scaling trees of fossil taxa such as cal3 (Bapst, 2013), and that the longer internal branches generated by this method are a more appropriate model of the evolution of clades than those provided by the older and widely used '*mbl*' and '*equal*' methods (G.T. Lloyd, pers comm). By completing extinction clustering analyses on these sets of 50 stochastically generated phylogenies for each clade, results have a distribution which represents the uncertainty introduced by occurrence and divergence dates that are not known precisely.

To establish whether the Hedman time-scaling method was systematically biasing results I conducted the analyses on two additional sets of 50 phylogenies for each clade. One set was time-scaled using the *equal* method of Brusatte et al. (2008) as originally implemented in R by Lloyd et al. (2012b) and incorporated into the paleotree package (Bapst, 2012). In the *equal* method divergences are dated according to the first appearance date of the sister group to the clade, and then data from the next most basal branch with a positive duration is shared equally between branches joined to the associated internal node. The second additional set of 50 trees was time-scaled using the minimum branch length ('*mbl*') method, also originally implemented in R by Lloyd et al. (2012b), and incorporated into paleotree. Divergences are dated according to the first appearance date of the sister group to the clade, and then any branch that is shorter than a specified length of time (I used 1 Myr) is extended to this length. The time-scaling method may have an important effect on the results as it controls how far back in time each internal node is placed and therefore the amount of ghost range included in the phylogeny. It therefore controls which taxa are included in each timeslice (see next section), as well as the phylogenetic distance between taxa (Fig. 3.2).

3.3.3 Time-slicing

To translate extinction to a binary trait the time-scaled phylogeny of interest is divided into successive timeslices of approximately the same length. If a taxon's last appearance falls

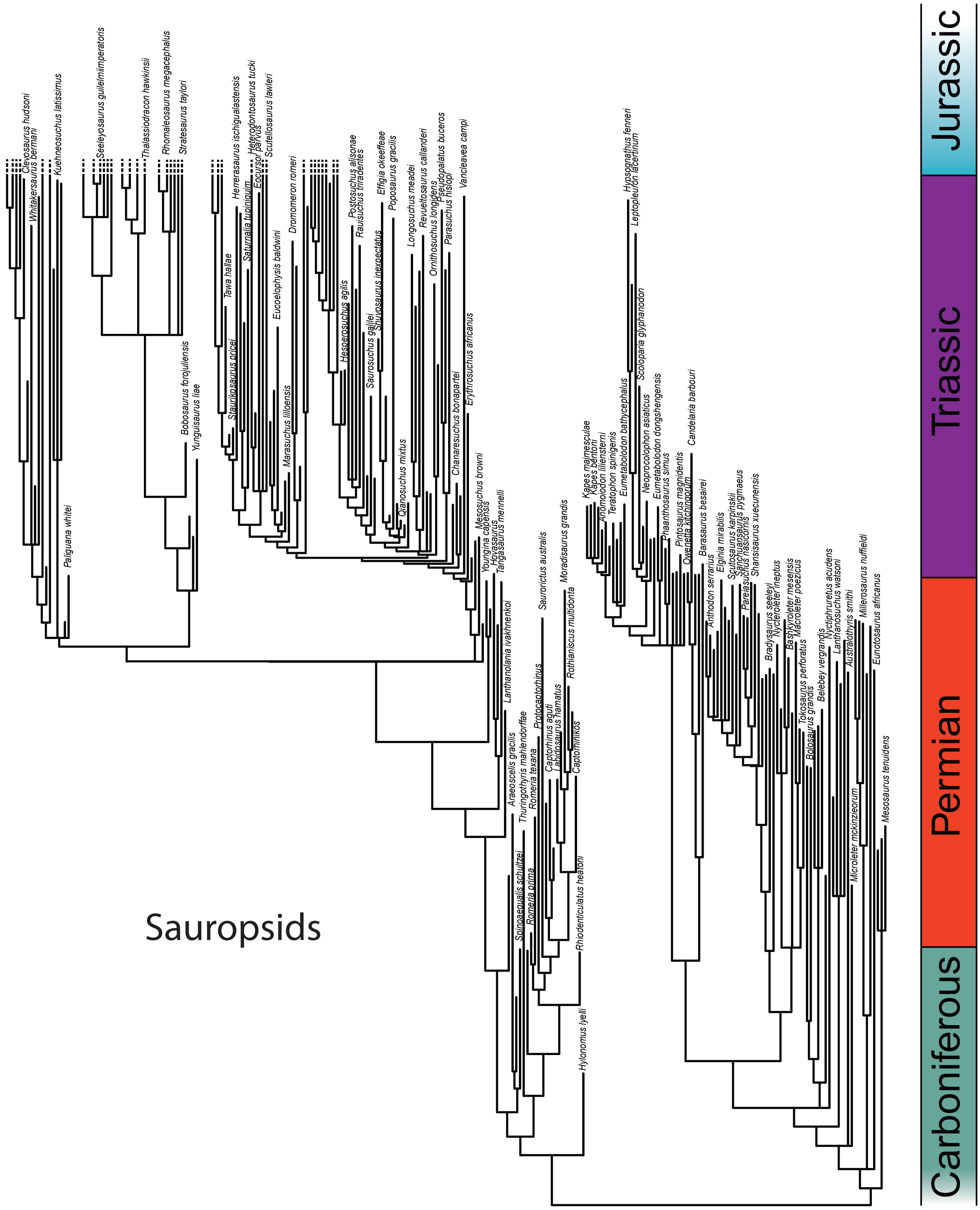


Figure 3.3: An example of one of the 100 timescaled (using the *Hedman* method) phylogenies of the clade Sauropsida which was used in the extinction clustering analyses. The last time bin in the analyses was the end of the Triassic, dashed tips indicate where the lineage continued past this point. Some taxon names were removed for legibility. The full taxon list is available on the data appendix disk.

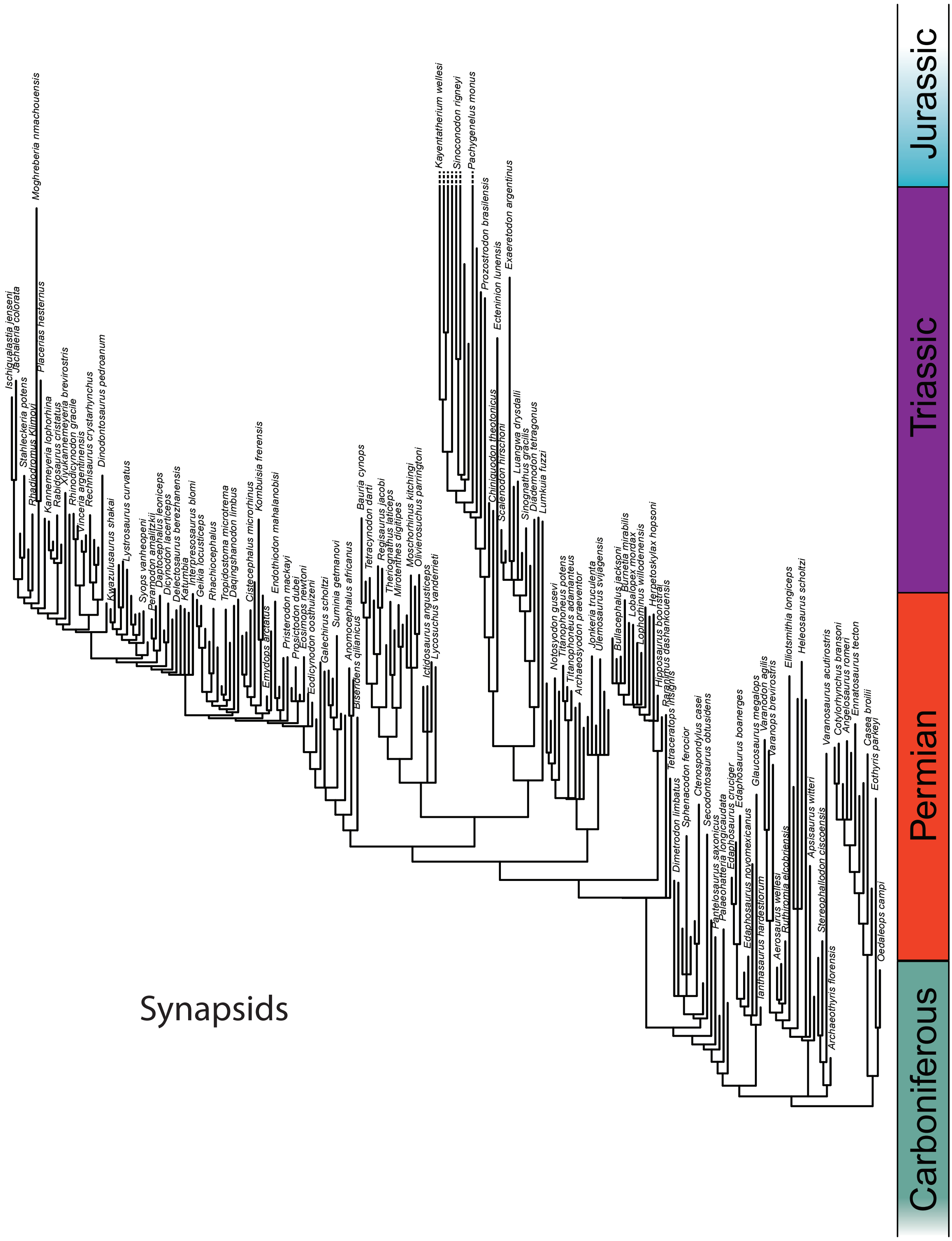


Figure 3.4: An example of one of the 100 timescaled (using the *Hedman* method) phylogenies of the clade Synapsida which was used in the extinction clustering analyses. The last time bin in the analyses was the end of the Triassic, dashed tips indicate where the lineage continued past this point. Some taxon names were removed for legibility. The full taxon list is available on the data appendix disk.

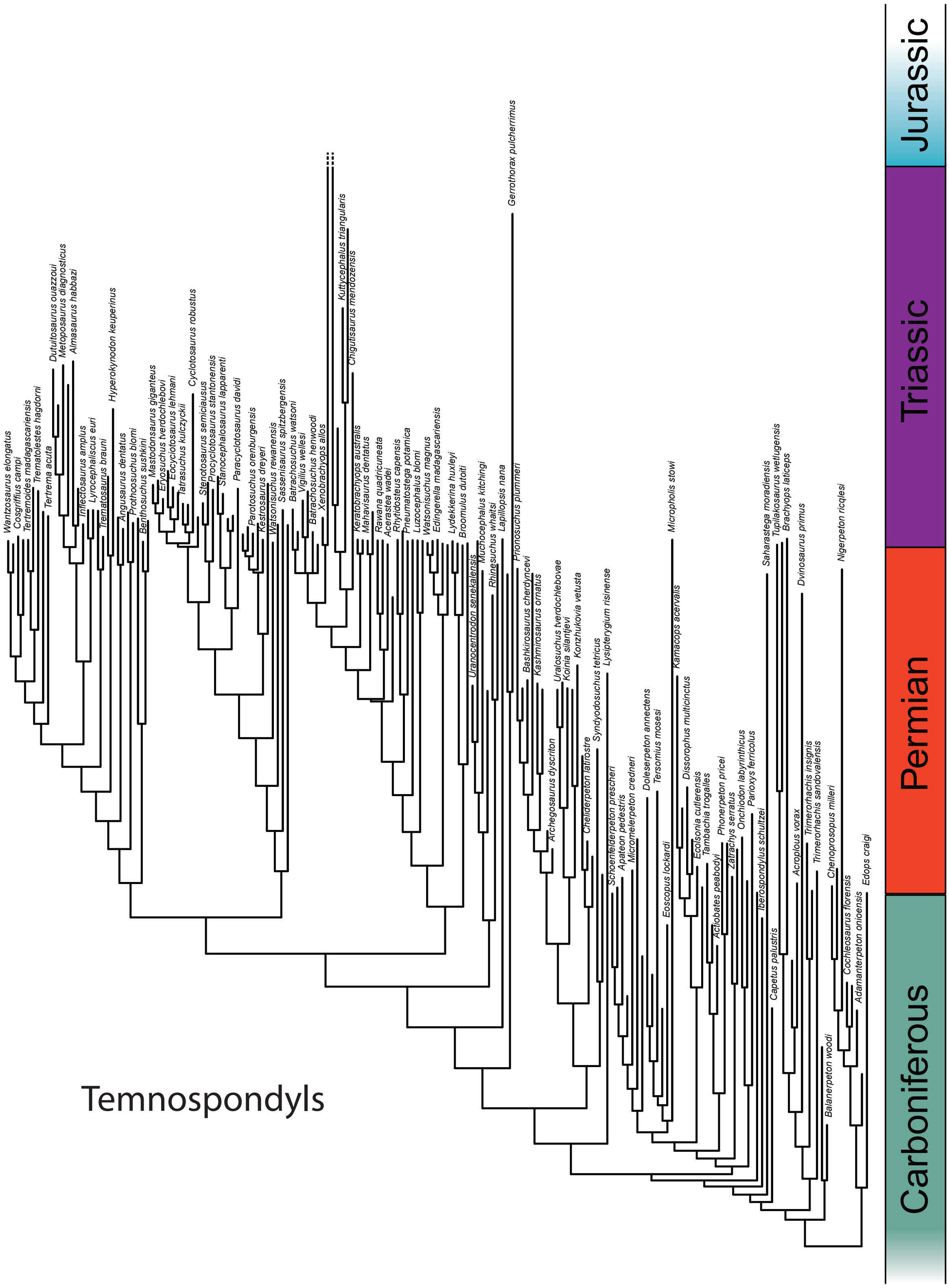


Figure 3.5: An example of one of the 100 timescaled (using the *Hedman* method) phylogenies of the clade Temnosponylii which was used in the extinction clustering analyses. The last time bin in the analyses was the end of the Triassic, dashed tips indicate where the lineage continued past this point. Some taxon names were removed for legibility. The full taxon list is available on the data appendix disk.

within any one timeslice this is classified as an extinction; if the taxon's range includes the end of the timeslice this is a survival because the taxon is present within the slice but survives into at least the next one. For the main analysis I used timeslices that began and ended at the start and end of geological stages, but combined some consecutive stages into one bin to generate bins of a more even length. These combined stages are shown in Fig. 3.7. This allowed investigation of the signal at key times in the geological record, particularly times that have been suggested as extinction events, which are usually at the end of stages. It has been demonstrated previously that the intensity of the signal can be sensitive to temporal resolution of the timeslices (Hardy et al., 2012). Therefore, to test the effect of the length and timing of the timeslices I also ran analyses using timeslices of an exactly equal duration of 10 Myr and of 15 Myr. The start and end dates for each of these sets of timeslices is provided in Appendix B.

3.3.4 Phylogenetic clustering of extinction

To compare the strength of phylogenetic clustering of extinction in different periods of time, the timings of extinctions of taxa can be treated as a binary trait. This binary trait is then compared to the phylogenetic distance between taxa. There are three main methods available: the Pearson correlation coefficient,]Moran's *I* (Moran, 1950) and Fritz and Purvis' *D* (Fritz and Purvis, 2010). Here I focus on *D* (Fritz and Purvis, 2010). This metric is scaled to random and Brownian motion expectations of trait distribution. Because of this scaling, for trees containing more than 50 tips the metric is robust to tree shape, tree size and trait prevalence, unlike the two alternative metrics. It can therefore be used to reliably compare values through time, and between clades, and for this reason provides an advantage over alternative methods. I also repeated all analyses using Moran's *I* (a test for spatial autocorrelation (Moran, 1950), which was adapted for use to measure phylogenetic signal by Gittleman and Kot (1990)) to establish whether the same signal could be detected via this alternative method.

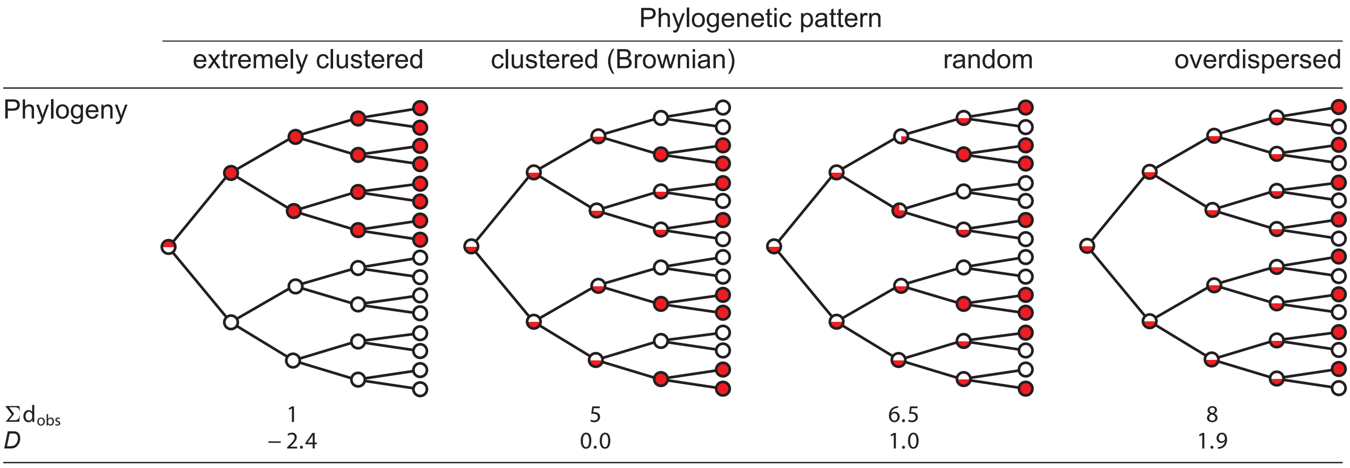


Figure 3.6: Adapted from Table 1 in Fritz and Purvis (2010). Hypothetical phylogenies showing the calculation of the metric *D*. The far left and far right phylogenies show a strongly clustered and an overdispersed distribution of extinctions across the tips of the tree. The two central phylogenies show examples of the trait distributions for the expectations used to scale the metric. $\sum d_{obs}$ is the sum of sister clades, and *D* is the resulting value after scaling.

The basis for calculating *D* is the sum of sister clade differences (Fig. 3.6). The first step in the method is comparable to the well-known method for deriving phylogenetically independent contrasts (Felsenstein, 1985; Harvey and Pagel, 1991): a nodal trait value is calculated as the average trait value between sister taxa descending from that node, weighted inversely by the length of the branches leading to those taxa. This is continued across the whole tree and summed to obtain the observed sum of sister clade differences. This value obviously depends heavily on the overall size and shape of the tree, not just the trait distribution, and this must be corrected for so that the resulting value can be meaningfully compared between trees. The next step involves calculation of the Brownian expectation of trait distribution for the tree. This is done by simulating a continuous trait over the tree, then selecting a threshold value to transform it to a binary trait with the same trait prevalence as the observed data. The sum of sister clade differences is calculated for this new simulated trait, the simulation is repeated (I used 1000 iterations) and the mean found. The random expectation is more straightforward to calculate: the observed binary trait is shuffled randomly over the tips of the tree and the sum of sister clade differences found. As with the Brownian expectation, this is repeated 1000 times and the mean is calculated. The observed value is then scaled to the Brownian and random values by the following equation:

$$D = \frac{\sum d_{obs} - mean(\sum d_b)}{mean(\sum d_r) - mean(\sum d_b)} \tag{3.1}$$

where d_{obs}, is the observed value of D and d_b and d_r are the Brownian and random values of D for each iteration. Once the value has been scaled, $D = 1$ corresponds to a random trait distribution, and $D = 0$ corresponds to a Brownian, or clustered, trait distribution. It is also possible to find a p-value for D, by comparing it to the distributions of values generated during the iterations to find mean($\sum d_b$) and mean($\sum d_r$).

If D falls outside the upper 95% of the random iteration values, it can be considered significantly different from the random expectation. In reality, the final value of D can fall between the random and Brownian distributions, so interpretation of the value in the context of these distributions is important.

Moran's I is a metric for spatial autocorrelation that was developed to measure the degree to which similar objects cluster together in space (Moran, 1950). It can be adapted for purpose here to measure the degree to which a binary trait (extinction) clusters in phylogenetic space (phylogenetic distance between taxa) (Gittleman and Kot, 1990; Lockwood et al., 2002). It is calculated with the following equation:

$$I = \frac{n \sum_i \sum_j z_i z_j w_{ij}}{\sum_i \sum_j w_{ij} \sum_i z_i^2} \tag{3.2}$$

where n is the number of observations, w_{ij} is a weighting which is calculated as 1 divided by the cophenetic distance between two species i and j, and z_i is the normalised value of the trait for the species i (Lockwood et al., 2002). In many previous studies Moran's I correlograms have been used, which is possible when the taxonomic or phylogenetic distance is also a binary trait (e.g. species belonging to the same genus, or taxa within a particular phylogenetic distance threshold of one another). I chose to maintain the phylogenetic distance as a continuous trait because my phylogenies have branch lengths scaled with respect to time and so contain additional information about the evolutionary distances between taxa. This information is removed when threshold values are used. In addition the choice of the value for the threshold is arbitrary. The generalised method for Moran's *I* also has the advantage of providing one value for the whole tree (Hardy et al., 2012).

3.4 Results

The raw results and all code used for the analyses is included on the data appendix disk. Extinction was phylogenetically clustered in all three clades during the majority of the time periods investigated (Fig. 3.7), and fell within the distribution for the Brownian expectation. There tends to be a greater spread in *D* values for a bin where the phylogenetic patterning is weak or random, showing that in these cases variation in both the topology and branch lengths of the tree has more of an effect on the result. All three clades show relatively random extinction in their early history, although it is not clear whether this is a genuine signal, or bias caused by proximity to the root of the tree or a small sample size. Extinctions are then consistently clustered in the last three timeslices of the Permian. Following the P/Tr boundary, sauropsids transition to a period of more phylogenetically random extinction, particularly in the Carnian. Temnospondyls also show a decrease in clustering, but to a lesser degree. Synapsids maintain a highly clustered signal into the Mesozoic (Fig. 3.7). There does not seem to be an overall trend in changes in extinction clustering. In addition it is not more likely for a decrease in signal strength between timeslices to follow an increase, or vice versa.

Extinction intensity does not correlate with strength of phylogenetic clustering for any of the clades (Pearson product moment correlation: sauropsids r=-0.6936 p=0.08; synapsids r=-0.5596 p=0.1915, temnospondyls r=0.2281 p=0.6228).

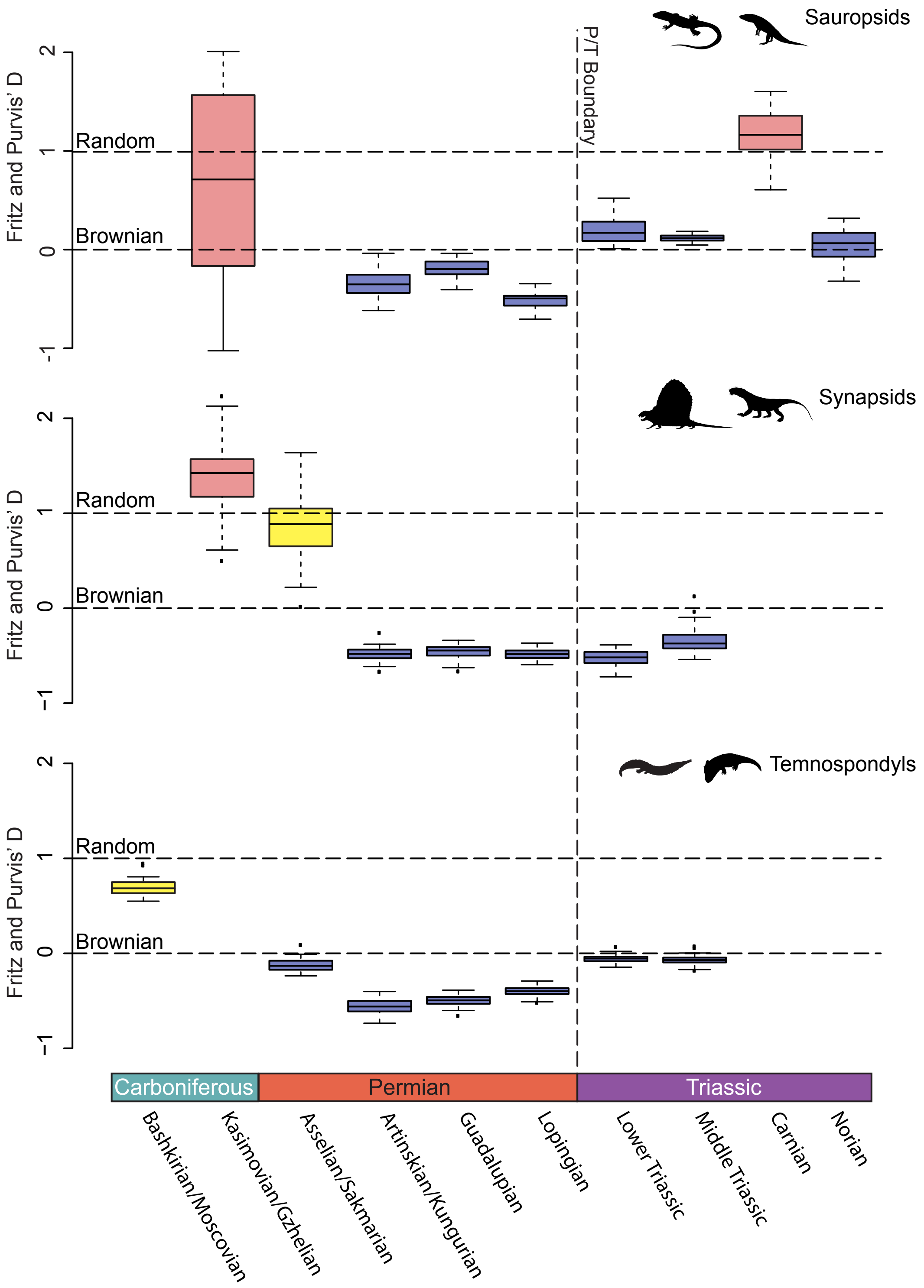


Figure 3.7: Measurement of D through time on a set of 100 phylogenies timescaled using the *Hedman* method. Boxes correspond to the first and third quantile, the line in each box is the median. Red shading indicates the values are within the distribution of the random expectation. Yellow shading indicates the values fall within both the random and normal expectations and blue that the values fall only within the Brownian expectation (i.e. the results indicates extinction was phylogenetically clustered in the timeslices shaded blue). Where there is a blank space for a particular timeslice rather than a box this means that the measurement for that timeslice did not fulfil the requirements of the method for D to provide a robust result, so less than 25 tips, trait prevalence of less than 20% or more than 80%, or poor resolution.

3.4.1 Moran's I

When the phylogenetic autocorrelation of extinctions was measured using Moran's *I*, for Sauropsids and Synapsids the results were very similar to the results obtained using *D* (Fig. 3.8). Sauropsids show the highest clustering of extinction by this measure in the late Permian timeslice, and the lowest in the Carnian. The clustering signal is significantly different from random in the final three timeslices of the Permian, the Early Triassic and the Norian. Synapsids show clustered extinction in the final three timeslices of the Permian and the first two timeslices of the Triassic. For *D*, there was less clustering during the Middle Triassic than during the preceding timeslice, however, for Moran's *I* the Middle Triassic shows the most clustered signal. This may be a consequence of the large proportion of extinctions in that bin (0.72). Values of Moran's *I* for temnospondyls show a different pattern from results of *D* in some timeslices. The end Permian timeslice does not have a significantly clustered signal, and the two first timeslices of the Triassic show the most clustered signal. Again this can most likely be attributed to the relative proportions of extinction in these timeslices. For temnospondyls only 12% of taxa become extinct during the Lopingian, in the two timeslices following the P/Tr this figure was 70% and 63%. In fact, the extinction intensity in temnospondyls correlates with the test statistic for *I* (r = 0.8295, p=0.02). Extinction intensity is more even across timeslices for sauropsids and synapsids and does not correlate significantly with results for *I*.

3.4.2 Timeslice length

The length of timeslices does not correlate with phylogenetic clustering (Pearson's r: sauropsids r=-0.1518 p=0.774; synapsids r=0.2469 p=0.5935, temnospondyls r=0.1034 p=0.8254) indicating that length of timeslice is not biasing the results. *D* measured for timeslices of 15 Myr and 10 Myr in length is broadly similar to *D* obtained using combinations of stages as timeslices, although fewer results could be obtained because many timeslices did not include enough taxa (Fig. 3.9). Many have a wider spread of results than the results from timeslices corresponding to stages in the main analysis, a reason for this is outlined in the discussion. Collectively these results demonstrate that whichever length timeslices are chosen, they largely capture the overall change in signal through time.

For sauropsids both the 10 and 15 Myr timeslice results showed more dispersed extinction in the early to mid-Permian than the main analyses showed. The same strongly clustered signal is present at the P/Tr boundary, followed by more phylogenetically random extinctions in the Triassic, particularly in the timeslice encompassing the Carnian, when extinction clustering was significantly less than the Brownian expectation.

Synapsids show an increasingly clustered signal towards the P/Tr, followed by a slightly less clustered period in the first part of the Triassic for the 15 Myr timeslices. The 10 Myr timeslices show a similar result to the main analyses except that the timeslice that includes the Guadalupian has a higher value of *D*, this is likely caused by the extinctions in the last 3 Myr of the Guadalupian being included in the timeslice with the Lopingian extinctions. In the Carnian only one taxon makes its last appearance so none of the sets of timeslices could provide a robust result for this stage. In the Late Triassic there were not enough taxa in each bin to get a value for *D*.

Temnospondyls show a consistent signal across all alternative sets of timeslices. There is phylogenetically clustered extinction during the latter half of the Permian, followed by a less clustered value in the early Triassic, and not enough taxa to get a measurement for the subsequent timeslices.

3.4.3 Time-scaling method

Results are broadly similar when a different method is used to assign branch durations to the trees (Fig. 3.10). If there is a difference in values of *D* between trees, *equal* and *mbl* scaled trees show a less clustered signal more frequently than a more clustered signal, relative to the Hedman scaled trees, but the difference is always small. A reason for this is outlined in the discussion. However, there is no consistent bias towards more or less clustering from any one method. The overall signal of variation in clustering is maintained, as is the strongly clustered extinction at the P/Tr boundary. There is a greater variation of results within each time bin when the trees have been scaled using the *equal* or *mbl* methods, than when the Hedman method was used.

Table 3.2: Summary of average results across the set of trees for each clade

Sauropsids

Timebin	1	2	3	4	5	6	7	8	9	10	11
D	NA	NA	NA	NA	-0.34	-0.19	-0.52	0.20	0.12	1.17	0.05
D std	NA	NA	NA	NA	0.14	0.09	0.08	0.13	0.03	0.26	0.16
Intensity	NA	NA	NA	NA	0.29	0.36	0.39	0.28	0.41	0.19	0.37
Taxa	NA	NA	NA	NA	61.94	73.34	70.42	69.90	80.94	62.42	67.94
R p-value	NA	NA	NA	NA	0.0000	0.0000	0.0000	0.0003	0.0000	0.6642	0.0010
B p-value	NA	NA	NA	NA	0.7952	0.7089	0.9491	0.3102	0.3793	0.0115	0.4762
I	NA	NA	NA	NA	0.0920	0.1106	0.2913	0.0898	0.1009	-0.0085	0.0618
I p-value	NA	NA	NA	NA	0.0044	0.0009	0.0000	0.0007	0.0000	0.7031	0.0072

Synapsids

Timebin	1	2	3	4	5	6	7	8	9	10	11
D	NA	NA	1.40	0.85	-0.48	-0.46	-0.49	-0.53	-0.33	NA	NA
D std	NA	NA	0.34	0.30	0.08	0.07	0.06	0.08	0.13	NA	NA
Intensity	NA	NA	0.25	0.29	0.36	0.53	0.62	0.29	0.72	NA	NA
Taxa	NA	NA	35.56	35.28	58.76	110.44	80.74	45.06	44.28	NA	NA
R p-value	NA	NA	0.7708	0.3624	0.0000	0.0000	0.0000	0.0000	0.0008	NA	NA
B p-value	NA	NA	0.0303	0.0969	0.8752	0.9061	0.8979	0.8930	0.7501	NA	NA
I	NA	NA	-0.0270	0.0276	0.1482	0.0970	0.1354	0.1427	0.2246	NA	NA
I p-value	NA	NA	0.6912	0.1763	0.0000	0.0000	0.0000	0.0009	0.0027	NA	NA

Temnospondyls

Timebin	1	2	3	4	5	6	7	8	9	10	11
D	NA	0.69	NA	-0.12	-0.56	-0.50	-0.40	-0.06	-0.07	NA	NA
D std	NA	0.08	NA	0.07	0.08	0.06	0.05	0.04	0.05	NA	NA
Intensity	NA	0.20	NA	0.45	0.12	0.18	0.12	0.70	0.63	NA	NA
Taxa	NA	54.68	NA	62.78	57.84	72.18	84.58	91.58	33.34	NA	NA
R p-value	NA	0.1590	NA	0.0001	0.0002	0.0000	0.0000	0.0000	0.0003	NA	NA
B p-value	NA	0.0673	NA	0.6436	0.8272	0.8809	0.8402	0.6022	0.5542	NA	NA
I	NA	0.0268	NA	0.0686	0.0302	0.1177	0.0338	0.1762	0.2348	NA	NA
I p-value	NA	0.2280	NA	0.0030	0.1376	0.0021	0.1444	0.0000	0.0000	NA	NA

These tables show mean values across the set of 100 trees for each clade. The numbered timebins correspond to the combinations of stages shown across the x-axis in Fig. 3.7. Where there is a column of NAs this indicates that there were not enough taxa or a good enough trait distribution to get a robust result, these correspond to the gaps in the results graphs. *D* = strength of clustering using Fritz and Purivs' metric, *D* std = standard deviation of *D*, Intensity = the proportion of extinctions, Taxa = the number of taxa in the timeslice, R p-value = position of *D* within the distribution for the random expectation, B p-value = position of *D* within the distribution for the Brownian expectation, *I* = alternative strength of clustering using Moran's metric, *I* p-value = p-value for the test statistic for *I*.

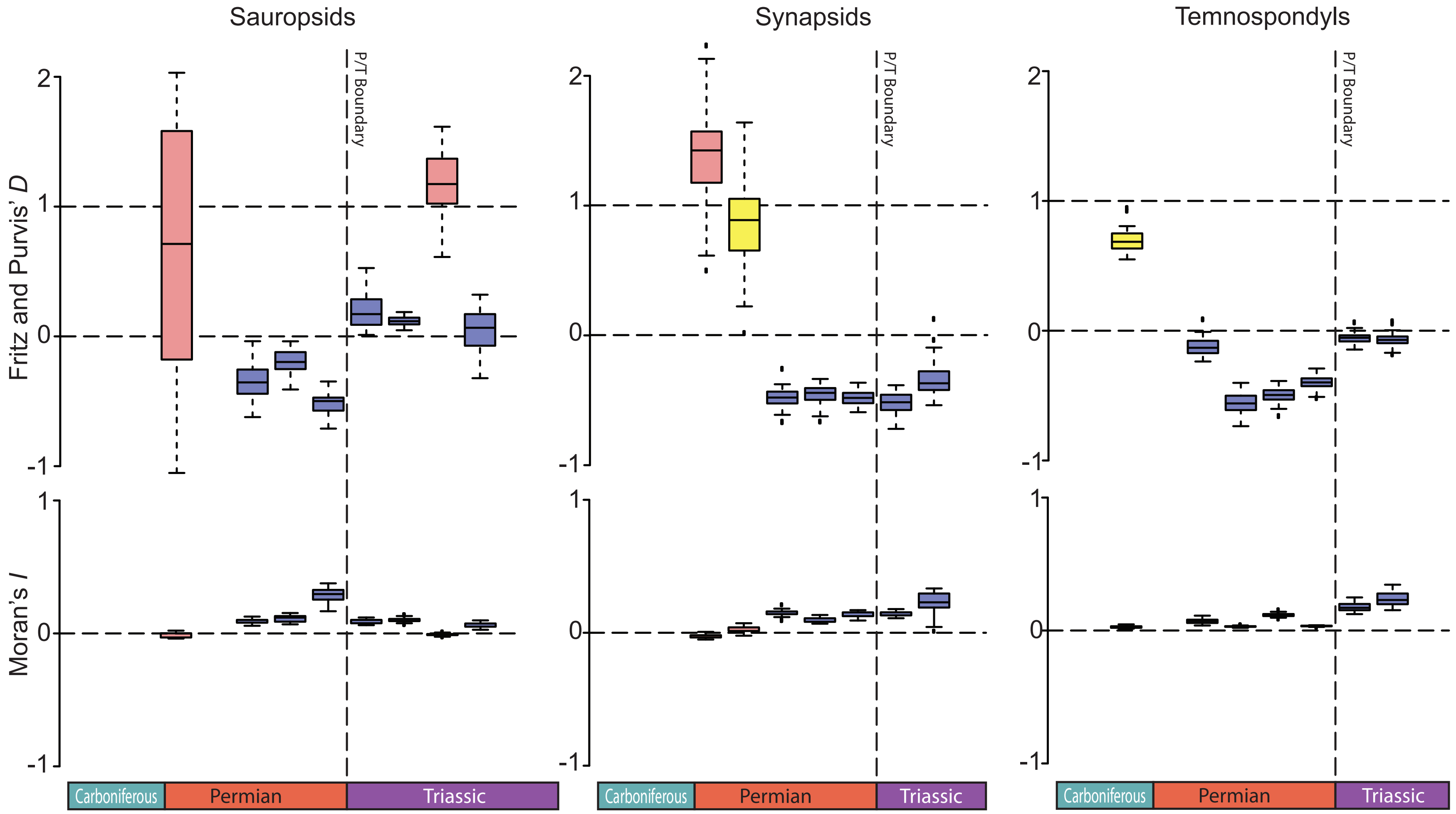


Figure 3.8: Measurement of *D* through time when an alternative metric of clustering is used - Moran's *I* - with the results from the main analysis plotted again for comparison. Red shading indicates the values are within the distribution of the random expectation. Yellow shading indicates the values fall within both the random and normal expectations and blue that the values fall only within the Brownian expectation (i.e. the results indicates extinction was phylogenetically clustered in the timeslices shaded blue). As with Fig. 3.7 a blank space indicates the requirements of the data for a robust measurement of *D* in that timeslice were not met.

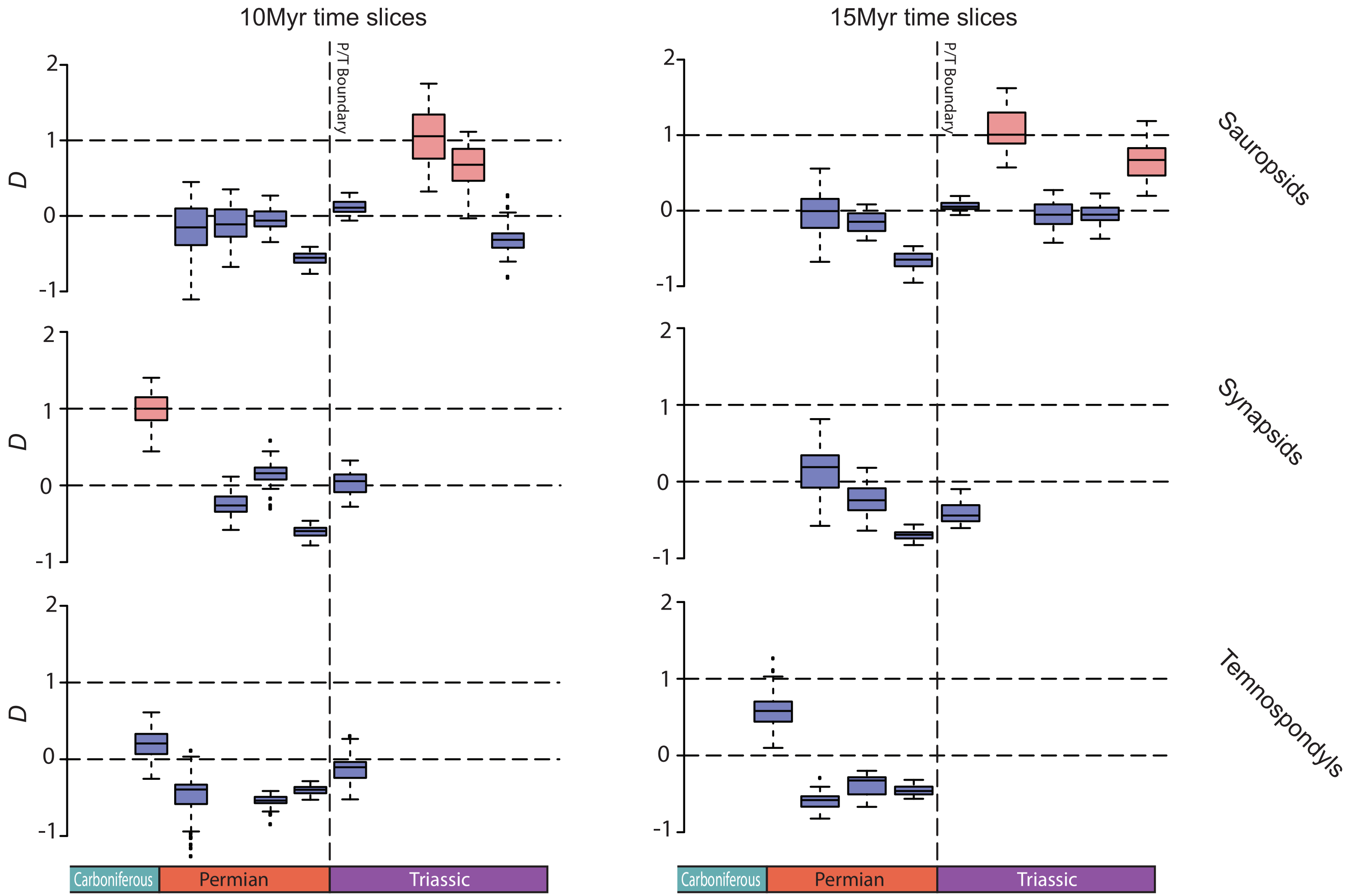


Figure 3.9: Measurement of *D* through time within different lengths of timeslice, 10Ma and 15Ma, with the main results provided for comparison. Red shading indicates the values are within the distribution of the random expectation. Yellow shading indicates the values fall within both the random and normal expectations and blue that the values fall only within the Brownian expectation (i.e. the results indicates extinction was phylogenetically clustered in the timeslices shaded blue). As with Fig. 3.7 a blank space indicates the requirements of the data for a robust measurement of *D* in that timeslice were not met.

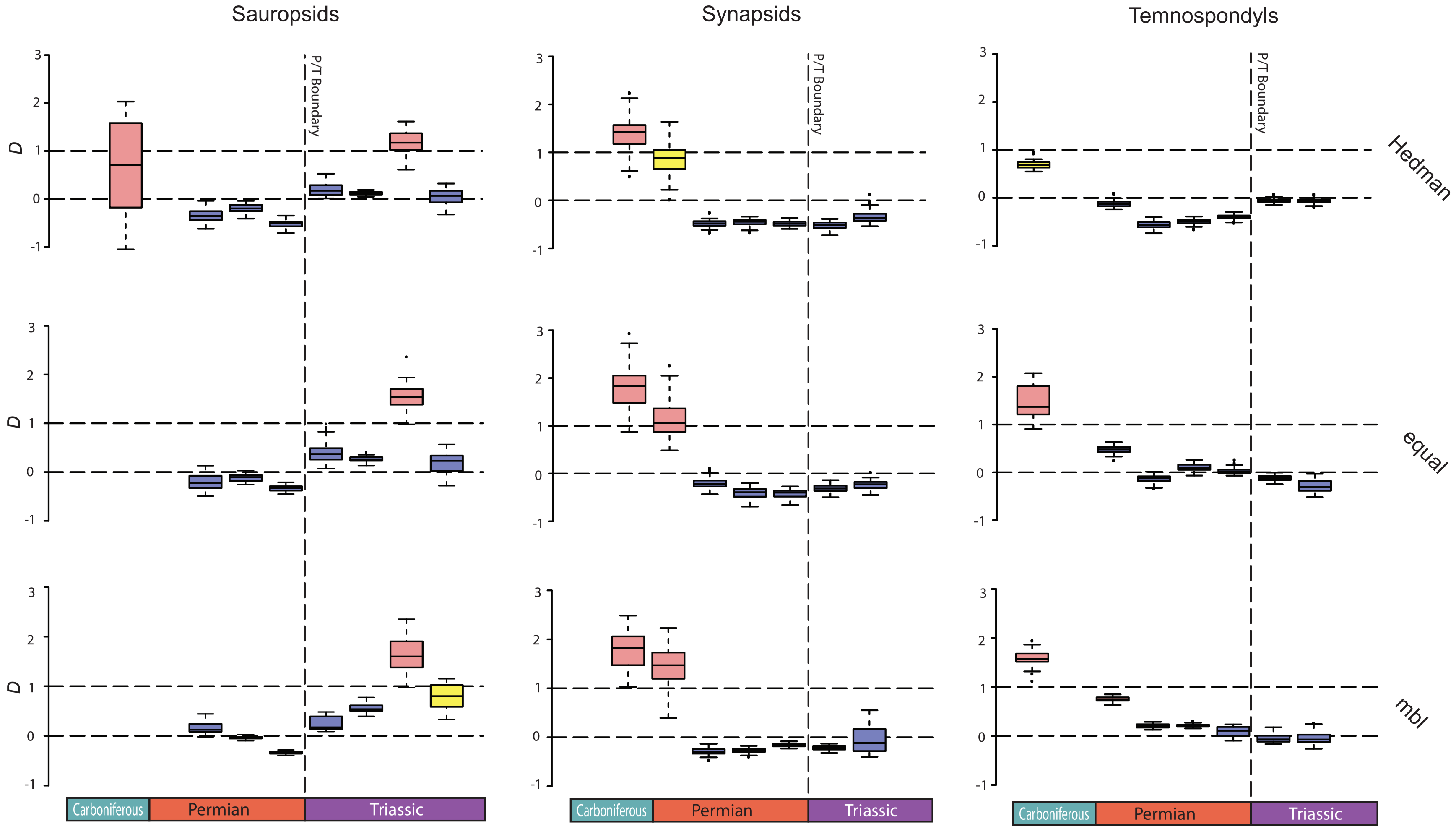


Figure 3.10: Measurement of *D* through time for different methods of timescaling the cladogram topologies - *equal* and *mbl* - with the main results that were timescaled using the *Hedman* method for comparison. Red shading indicates the values are within the distribution of the random expectation. Yellow shading indicates the values fall within both the random and normal expectations and blue that the values fall only within the Brownian expectation (i.e. the results indicates extinction was phylogenetically clustered in the timeslices shaded blue). As with Fig. 3.7 a blank space indicates the requirements of the data for a robust measurement of *D* in that timeslice were not met.

3.5 Discussion and implications

These results demonstrate that tetrapod extinctions were phylogenetically clustered during the majority of the early evolution of the clade. This corroborates previous research on shorter timescales, and on extinction risk in modern taxa, which indicates that phylogenetic signal is a common feature of extinctions regardless of timescale, and can be considered a general rule (McKinney, 1997; Janevski and Baumiller, 2009; Roy et al., 2009). All three clades show a highly clustered signal during the final timeslice of the Permian (Lopingian), which includes the losses of the P/Tr mass extinction event. However, there are many other timeslices that also show phylogenetically clustered extinctions indicating that the P/Tr was not unique by this measure. The strength of clustering varies through time and the variation in the signal cannot be explained solely by extinction intensity (Hardy et al., 2012), or by mass extinctions having a regime that is fundamentally different to the processes driving extinction in background times (Jablonski and Raup, 1995; Jablonski, 2005).

Over the long time-scales and global range of this study, there would have been geographically linked variation in the intensity of threat, and type of threat. It is probable that this variation would have correlated to some extent with phylogenetic distance between taxa. This is because closely related taxa may be more likely to occupy ranges in close proximity, meaning that palaeogeography is a possible conflating factor (Lieberman, 2003). This phylogenetic signal in geographic distribution may not occur in all taxa, particularly not in generalists with good dispersal ability. However, many taxa are restricted to particular habitats or temperate zones, to which close relatives with whom they share a recent evolutionary history are more likely to also be restricted, as they will share similar traits. In addition, without key adaptations such as flight, dispersal takes a long time and may be blocked by physical barriers. These factors all combine to suggest that although geographic and phylogenetic distance may not always be linked, there are periods of time, particularly in the early history of a taxon, where this is likely to be the case. Whether or not closely related taxa were in geographic proximity in any particular clade could be tested by finding if there is a positive correlation between geographic and phylogenetic distance of sampled taxa within the clade.

On smaller scales where variation in threat intensity is not a conflating factor, extinction (or extinction risk) is still strongly phylogenetically structured (Roy et al., 2009; Hardy et al., 2012). Thus the two factors which have control over a species' vulnerability to extinction on the timescales in this study – its phenotype and the threat it experiences – are both expected to be phylogenetically conserved to some degree. Consequently, extinction itself should be phylogenetically conserved, and the evidence presented here is consistent with this expectation. This conclusion indicates that correction for phylogenetic signal in extinction will be important when trying to identify individual traits that might confer extinction vulnerability (Harnik et al., 2014). In addition it indicates that future extinctions should be expected to erase a disproportionately large amount of evolutionary history in comparison to the number of taxa becoming extinct (Fig. 3.5, Davies and Yessoufou, 2013).

In the publication proposing *D*, the authors offer suggestions on how different signals should be interpreted (Fritz and Purvis, 2010). They suggest that a clustered signal implies susceptibility to extinction based upon biological traits of the taxa, which are strongly phylogenetically conserved, and that extrinsic factors such as variation in threat intensity in different geographic regions would lead to a random signal. The signals obtained in this study should be interpreted differently, or at least cautiously, due to the much longer timescale involved. When considering the snapshot of the modern day, or higher-resolution fossil records, it is reasonable to expect that the same mechanism was driving extinction of a group of related taxa that became extinct (or are at a high risk of extinction). This is because an individual mechanism could be sustained over a geologically short period of time. This is unlikely to be the case for the duration of the 5-20 Myr timeslices in this study. This indicates therefore that there are particular phylogenetically

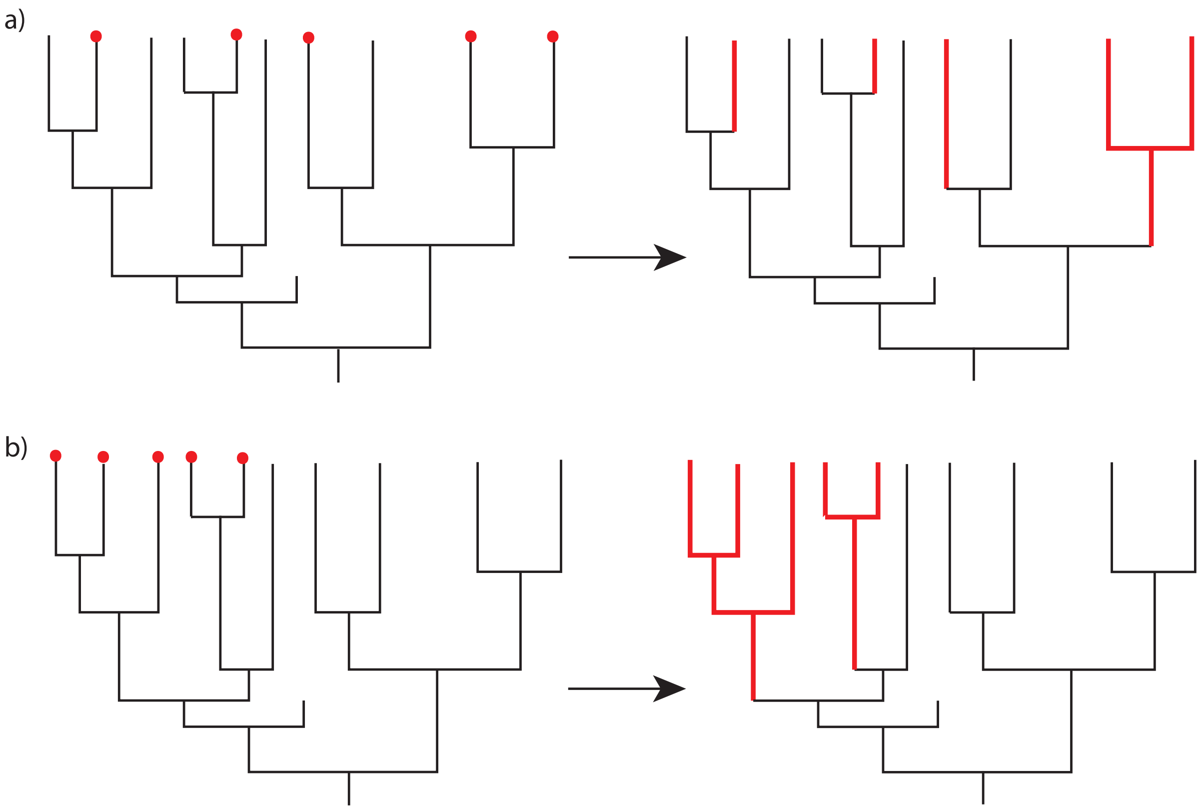


Figure 3.11: Schematic of the way in which clustered extinction causes a disproportionately large loss of evolutionary history. The 'lost' evolutionary history of a clade can be considered to be the internal branches of the phylogeny that are not represented by a living taxon. Red circles represent an extinction and red branches represent parts of the phylogeny that are not represented by any of the surviving taxa. In a) extinction is random with respect to phylogeny, leading to the loss of terminal lineages but few internal branches of the tree. In b) extinction is clustered with respect to phylogeny and the proportion of internal branch length lost is greater.

conserved traits which render a species vulnerable to extinction independently of the type of threat it experiences. This result is consistent with previous studies that show that traits such as large body size, ecological specialism, or low fecundity always confer vulnerability (Cardillo et al., 2005; Purvis et al., 2005), and demonstrates that this remains the case even with variation in threat type and intensity in the long-term.

Previously some workers have suggested that the P/Tr represented a severe extinction and period of complete ecosystem restructuring for terrestrial tetrapods (Benton et al., 2004; Fröbisch, 2013). Highly phylogenetically clustered extinction has a disproportionately large effect on biodiversity compared to random extinction (Davies and Yessoufou, 2013), perhaps allowing for or requiring major ecosystem change. However, in the results presented here extinction is not more clustered at the end-Permian than it is in many other timeslices, indicating that phylogenetic selectivity may be decoupled from both extinction intensity and ecological change (Droser et al., 2000).

Additionally there is a large body of work showing that the nature and degree of extinctions during the P/Tr event itself was different in each clade in measures such as diversity. As mentioned above all three clades show clustered extinction throughout most of the time period studied here, and this includes very similar values in all three clades for the strength of signal at the end Permian itself, despite these previous studies indicating a difference in clade dynamics between them at this time in particular. For sauropsids, other work has suggested that there was not a large drop in any measure of diversity at the P/Tr boundary (Ruta et al., 2011). Conversely, temnospondyls increased in species diversity through the Permian then experienced a large drop at the boundary itself (Ruta and Benton, 2008). For synapsids, prior to the end of the early Permian pelycosaurs were relatively successful and experienced comparatively few extinctions. At the end of the Early Permian many pelycosaur taxa became extinct, followed by periodic extinctions of other synapsid clades throughout the rest of the Permian, including at the P/Tr (Fröbisch, 2008; Lucas, 2009). This observation for synapsids, of a continuing pattern of periodic extinctions of other clades could explain why a strong clustering signal can be seen in the early Triassic for this clade, unlike the other two clades. In combination with my results this information indicates that variation in phylogenetic selectivity and variation in extinction intensity are not directly related, but may share a common driver of extreme values.

3.5.1 Sensitivity tests

The sensitivity tests indicated that in some cases the method employed to perform the various steps required to obtain a result had an important influence on the observed signal, but these effects were small. Changing the algorithm used to time-scale the trees was important because different time-scaling methods extend lineages back in time, and add duration to internal branch lengths, to varying degrees (Fig. 3.2). The *equal* method adds very little duration to any branch, which leads to very ‘tippy’ trees where the terminal branches are on average much longer than internal branches. The *mbl* method also generates this kind of tree, although by setting the minimum branch length to a comparatively large number the internal branches can be extended further. Both these methods assign an arbitrary additional branch length, and when used this is often set to be quite small so as not to assume unsampled lineage history for which there is no evidence. The Hedman method assigns branch durations in a less ad-hoc manner and so tends to extend internal branches proportionally more than the other methods. This has the effect of drawing a greater number of lineages back into the timeslice prior to the slice in which fossil occurrences of the lineage are found. This increases the number of survivals in each timeslice. It also has the effect of increasing the trait co-variance between taxa in a Brownian motion simulation. Whether an increased number of survivals increases or decreases the value measured for clustering depends on which lineages are extended. For the datasets analysed here it causes a more clustered signal to occur more frequently than a less clustered signal, when compared to the signal measured on differently time-scaled trees. The *equal* and *mbl* methods can be considered conservative approaches as they minimise the amount of ‘made-up’ evolutionary history for which evidence is lacking. However, the branch durations in the trees time-scaled this way are unlikely to represent the true timings of lineage divergences accurately. First, because in a poorly sampled record first appearance dates are unlikely to be close to actual divergence dates. Second, these tree shapes imply an extreme version of the early burst model, where the large majority of speciation events occur early in the clade’s history, which has been shown consistently to be a poorly supported model of evolution (Harmon et al., 2010).

Implementing an alternative method to measure clustering also gave a slightly different result, particularly for temnospondyls. However, the strong link between extreme values of trait prevalence and extreme values of I indicates that the method is not particularly robust to variation in trait prevalence, unlike Fritz and Purvis' *D*, which can give results comparable across timeslices even when they have a very high or low proportion of extinctions.

The length of timeslices has an effect because of the stage level resolution of the data. This means that when the timeslices correspond to combinations of stages (as they do in the main analysis), a taxon will always go extinct in the same timeslice, even if its divergence date is in different stages for different iterations of the time-scaling algorithm. This is not the case when 10 and 15 Myr timeslices are used (except for the Lopingian because this boundary is used in all the alternative sets of timeslices). An occurrence date within a stage could fall either side of the end of a timeslice if it does not correspond to a stage boundary. This leads to variation in which extinctions occur in which timeslice.

3.5.2 Ancestor-descendant relationships

One problem that cannot be tested using the dataset analysed here is that of ancestor-descendant pairs. Over such long timescales the likelihood of taxa included in the analysis being ancestral to others is greatly increased (Foote and Raup, 1996). At the very least the likelihood of additional 'pseudoextinctions' being measured as actual extinctions is high. Pseudoextinctions are where a taxon appears to have gone extinct at a particular time because it is not seen in the fossil record of later stages, but in fact a speciation event (or several) could have occurred meaning that in reality the lineage continued.

A method of time-scaling trees of fossil taxa has been developed that allows for estimation of ancestor-descendant pairs within the reconstructed topology (Section 1.4.5 Bapst, 2013), however this method requires estimation of speciation and sampling rates, which is not possible with the taxa I analysed here because the majority of them occur only in one stage, or are point occurrences. In order to jointly test how influential the lack of good estimation of the position of ancestors and the effect of unsampled taxa is, I provide a simulation study in the next chapter.

3.6 Summary

The time-scales considered here are very different to those under investigation when modern extinction risk or recent extinctions are measured, but these results demonstrate that phylogenetic clustering of extinctions is common on all scales, and that affects in the short-term scale up over time to result in similar patterns in the long-term. I have demonstrated that it is possible, using a non-ultrametric tree of fossil taxa scaled to time, to measure values for a metric that can also be measured in phylogenies of recent and modern taxa to obtain results that are unbiased by the time over which they are measured.

In the modern biota it is possible to identify the increased risk associated with various different phenotypic or life history traits, but the influence of these must be untangled. In the absence of adequate data to do this in the geological record, phylogeny can act as a proxy for the effect of selection (or lack thereof) against the combination of these traits in a species. My results suggest that in the long term extinctions in the modern will not be fundamentally different to extinctions in the deep-time record, and that we can expect extinctions to continue to show phylogenetic signal into the future. In the past extinction has not been consistent with a field of bullets model, and there is no reason to expect it to be so in the future.

The phylogenetic signal in extinction provides important additional information on the patterns of tetrapod extinctions in deep time, complementing what can be understood from extinction rate alone. Extinction was phylogenetically clustered in the geological past, but not consistently so, and measuring this signal can tell us about what kind of environmental stress taxa may have been under. It will be informative to combine phylogenies of these three clades, as well as extending them further by adding clades such as mammals and non-avian dinosaurs to ascertain how patterns varied across the entire tetrapod tree of life, as well as within clades into the Mesozoic including two further mass extinctions: the End Triassic and the End Cretaceous.

Chapter 4

Bias in quantitative analyses of extinction in the fossil record

> *"For such a model there is no need to ask the question 'Is the model true?'. If 'truth' is to be the 'whole truth' the answer must be 'No'. The only question of interest is 'Is the model illuminating and useful?'."*
>
> - George E. P. Box *Robustness in the strategy of scientific model building (1979)*

In this chapter I test the degree to which palaeontological phylogenies produce results comparable to real phylogenies in analyses of the phylogenetic clustering of extinction. As we do not know the real phylogeny for any clade I test this using simulated evolutionary relationships and a simulated sampling and cladogram reconstruction process. I then explore the ways in which various features unique to phylogenies of fossil taxa (particularly inclusion of ancestor-descendant pairs) affect the results of the analyses. I provide recommendations for the methods that should be used to do this analysis, depending on the type of data available, and what biases to expect in results.

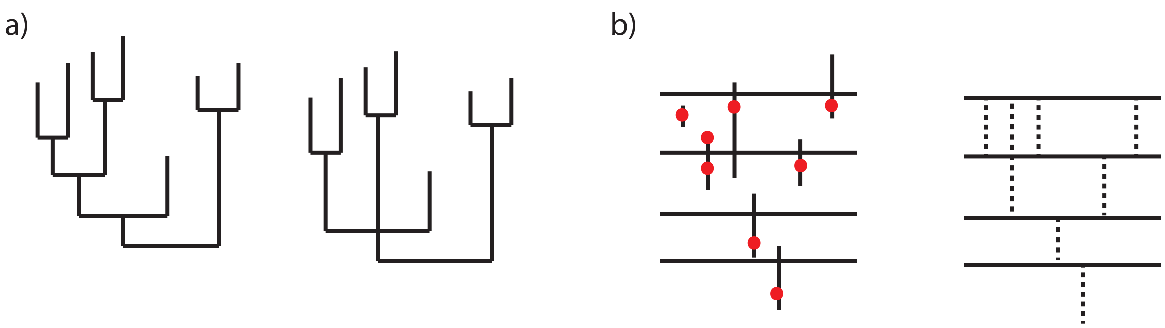
4.1 Bias and simulation

Fossil data are important in making accurate phylogenetic interpretations of macroevolutionary patterns (Finarelli and Flynn, 2006; Slater et al., 2012; Pennell and Harmon, 2013).

In return, phylogeny is important for valid interpretations of macroevolutionary signals in palaeontological data (Felsenstein, 1985; Martins and Hansen, 1997), and can overcome some of the biases introduced by analysing taxonomic data only (Harnik et al., 2014). Phylogenetic comparative methods (PCMs) are therefore now commonly used in palaeobiological studies (Hunt, 2012; Slater and Harmon, 2013; Soul et al., 2013; Bapst, 2014a), with an aim to provide results that can be compared across time and taxa (e.g. Chapter 3). A challenge to palaeobiology has been that the majority of PCMs were originally formulated by biologists to be used with living taxa (Harvey and Pagel, 1991; O'Meara, 2012). There have been studies on the effect on downstream analyses of several of the features that are more acute in phylogenies of fossil taxa than those of extant groups (e.g. uncertain divergence dates (Bapst, 2014b), missing character data causing tree misspecification (Stone, 2011), and a higher proportion of soft polytomies (Garland and Diaz-Uriarte, 1999; Housworth and Martins, 2001; Davies et al., 2012)). However, the effect on particular analyses of the overall 'degraded' nature of a palaeontological phylogeny has not yet been fully investigated. Phylogeny can be used to avoid bias introduced by the shared evolutionary history of taxa, but is not without its own bias (e.g. models or divergence dates (Uyeda et al., 2015; Bapst, 2014b)), which must be understood in order for the results of PCMs using fossil data to be meaningfully interpreted. The following issues are not exclusive to trees of extinct taxa, but are likely to be more prevalent in them, and potentially highly problematic in macroevolutionary analyses:

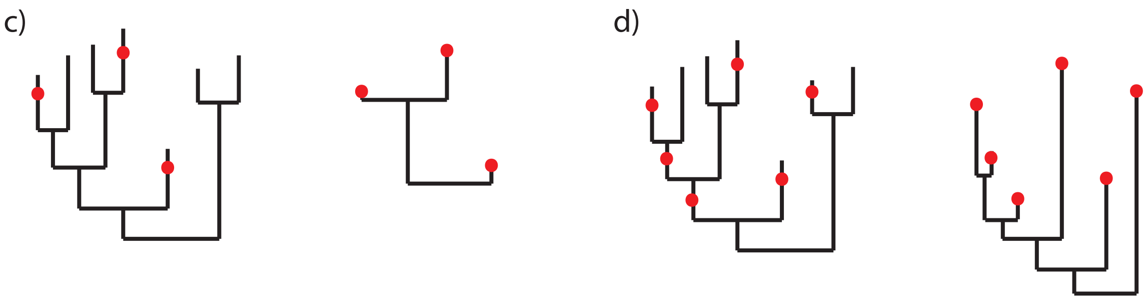
1. Imperfectly known relationships (polytomies and topology misspecification)
2. Imprecise occurrence data (age data are often coarsely binned)
3. Missing taxa (highly incomplete sampling)
4. Incorrect, or no, assignment of ancestor-descendant pairs

If any of these factors, or a combination of them, cause a systematic bias in the results of phylogenetic comparative analyses, any macroevolutionary interpretations based upon



Unresolved relationships/polytomies

Poor age resolution



Poor sampling/missing taxa

No ancestral reconstruction

Figure 4.1: Four common problems when reconstructing phylogenetic relationships which could have an effect on the results of downstream analyses. Each example has a ‘real’ set of evolutionary relationships or taxon ranges on the left hand side and a reconstructed set, showing a particular problem on the right hand side, sampling events are shown as red circles. In a) one of the sets of relationships between three internal branches cannot be resolved so this becomes a polytomy. In b) the ranges of the taxa are not continuously sampled and the resolution possible on the ages of samples is only to the level of the length of the time bins shown, so it is only known that the taxon existed for at least some of the duration of the dashed ranges, but not how long or which part of that range. In c) many of the taxa have not been sampled so the reconstructed tree is quite different to the original. In d) taxa which are ancestor descendent pairs have been sampled, and in a cladistic analysis these will get reconstructed as sister taxa, again resulting a tree topology quite different to the original.

them would at best be incompatible with results from analyses of extant taxa, and at worst completely incorrect.

Perhaps the most immediately obvious issue with palaeontological phylogenies is that of missing taxa (Fig. 4.1c). When performing analyses on living taxa it is assumed that the sample is a monophyletic clade including all the descendants of the last common ancestor, or with some more recent methods that the taxa are sampled randomly with respect to phylogeny (Fitzjohn, 2010). Although this assumption often does not hold (new living species are regularly discovered, recently extinct species may remain unknown, many species are paraphyletic (Funk and Omland, 2003; Ross, 2014)), for fossil taxa this is clearly never true. The proportion of lineages sampled for the majority of fossil clades is less than 5% (Foote and Sepkoski, 1999). There are therefore large numbers of taxa of which we have no knowledge. Many will be recovered through future field excavations and re-interpretation of collected material, but even so, most species are unlikely to have been preserved in the rock record in the first place.

In a worst case scenario, control on sampling or preservation of extinct taxa might have been biologically related to a macroevolutionary measurement. In this case a bias in the results of the measurement should be expected. For example, if an analysis focuses on measuring rates of speciation and extinction and includes a clade containing many short-lived species, it is likely that due to their short ranges many of these species will not be preserved, leading to an underestimate of turnover rates.

The issue of polytomies and topological misspecification (Fig. 4.1a) is one from which analyses of extant taxa also suffer (although probably to a lesser degree) and is therefore the most fully investigated of the above problems in the existing literature (e.g. Garland and Diaz-Uriarte, 1999). The effects of various treatments of polytomies is quite well characterised (Housworth and Martins, 2001; Stone, 2011). Some of the likely effects of polytomies and incorrect estimations of evolutionary relationships were explored in Chapter 2 of this thesis, particularly for model estimation. Likewise the positioning of nodes in phylogenies of fossil taxa is one which has been assessed in previous literature, including a simulation study (Bapst, 2014a). The effects of these two problems are therefore not explicitly assessed via simulation in the current chapter.

A fundamental problem in palaeontology is the difficulty of identifying ancestor descendant pairs (section 1.1.2, Fig. 4.1d). In phylogenies of extant metazoan taxa, much of the time it can be assumed that none of the taxa sampled in the analysis are the ancestor of any of the others, due to the comparatively short timescales involved (although see Funk and Omland, 2003; Ross, 2014). This is not the case with fossils, where long gaps in time between occurrences of taxa mean that it is probable that one taxon in the analysis could have descended from another (Foote, 1996b; Aldous and Popovic, 2005; Wagner, 2000a). This is problematic in part because it distorts speciation and extinction rates by introducing pseudoextinctions (section 1.1.2). Methods to estimate ancestral relationships accurately and consistently have not yet matured (although see Gavryushkina et al., 2014); characterising the downstream effects of these relationships in analyses is one way to understand their effect without explicitly identifying where they occur in a dataset.

Simulations have long been used as a method in palaeobiology to explore the effect of different variations in data on comparisons with a hypothetical modelled 'true' evolutionary history (Raup et al., 1973; Raup, 1982; Foote, 1996a; Alroy, 2010). By simulating a phylogeny of taxa using a birth-death model under controlled conditions, then degrading it to the kind of phylogeny we might expect to be able to construct from fossil occurrence data, the outcome of an analysis can be compared when conducted on the true tree, and when conducted on the degraded fossil tree (e.g. Bapst, 2014a).

The focus in Chapter 3 of this thesis was on the phylogenetic structure of extinction in the deep-time fossil record. Analyses relating to extinction are likely to be particularly vulnerable to bias because incorrectly estimating the presence or timings of extinctions (caused by pseudoextinctions or imprecise time constraints on fossil data) directly affect the results of this measurement. Here I focus on characterising the bias in extinction metrics under a variety of evolutionary and analytical conditions. This is in order to directly assess the probable validity of the results obtained in Chapter 3 of this thesis, as well as to provide a more general guide for anyone wishing to use extinction clustering metrics on fossil data.

4.2 Simulation set-up

Several packages are available in R (R Core Team, 2015) which include functions designed to simulate phylogenetic trees via different methods (e.g. `sim.bdtree` in *geiger*; Pennell et al. (2014)). The majority of these are discrete time birth-death simulators. An algorithm keeps track of a set of taxa and increments in time. During each time step there is a probability of a birth (meant to represent a speciation event, resulting in a new taxon/branch of the phylogeny) and of a death (meant to represent an extinction, resulting in the branch terminating). These probabilities can be set at different values depending on the required tree structure, and can be varied through time, or for different sections of the tree. The majority of existing studies including a phylogenetic simulation use this kind of function (Raup, 1981).

A more recent alternative (but still a birth-death model) is available in the package *paleotree* (Bapst, 2012), as the function `simFossilTaxa`. Rather than moving through the simulation one step at a time and assessing whether a birth or death has occurred for each taxon, the *paleotree* simulation functions draw a waiting time to the next event from a Poisson distribution in continuous time based upon instantaneous per-lineage speciation and extinction rates. It also incorporates explicit models of morphological differentiation. Taxa in the fossil record are defined by their morphology, the taxon units in `simFossilTaxa` are designed to represent this, and speciation can be set to a model corresponding to budding cladogenesis, bifurcating cladogenesis, anagenesis or a combination of these. This method of simulation is advantageous for the purposes of the analysis in this chapter as it can be used to faithfully represent the way in which we interpret relationships between morphotaxa in the fossil record.

The speciation models (Fig. 4.2) can be defined as follows: budding cladogenesis leads to the origination of one new taxon, while its parent taxon continues. Under this model of speciation, the end of a taxon's range represents a real extinction. Bifurcating cladogenesis leads to the origination of two new taxa at one event, and the discontinuation of the

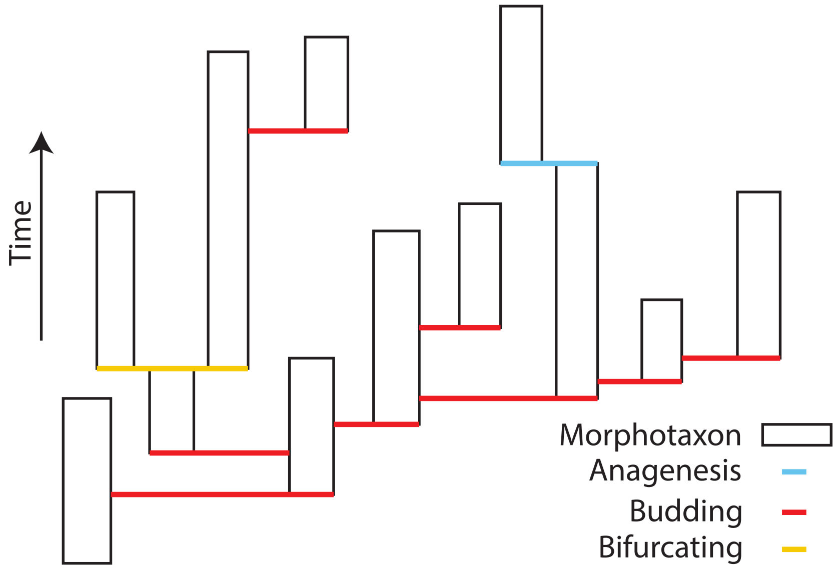


Figure 4.2: Schematic of types of morphotaxon differentiation, representing speciation, which can be accommodated in the simulation function that was used in this chapter to generate morphotaxa - `simFossilTaxa` from the *paleotree* package in R. Arbitrary x axis. Adapted from Bapst (2013)

parent taxon. In this model the end of a taxon's range does not necessarily represent a true extinction. Anagenesis represents a scenario where one new morphotaxon appears and the parent taxon discontinues. Again, range terminations do not necessarily represent real extinctions under this model. All these models effectively simulate a punctuated equilibrium model of evolution where all phenotypic change is concentrated at speciation events, rather than along branches. The validity of using such a model in this kind of analysis is outlined in Chapter 1, section 1.4.1.

I used `simFossilTaxa` with a variety of input parameters based on one simulation time unit representing 1Myr, but in addition introduced 'mass extinction' events at periodic intervals. These extinctions were either phylogenetically clustered or phylogenetically random. This was achieved by allowing a regular simulation to run for a set period of time (50Myrs to begin then several 20Myr intervals), then pausing it. The morphotaxa were converted to a phylogeny and a continuous trait was simulated over this phylogeny with a Brownian motion model. A threshold value was chosen which allowed for the desired proportion of extinction at this boundary. When clustered extinction was required a low proportion of lineages with a trait value below the threshold were terminated, and a high proportion of those taxa with a trait value above the threshold were terminated. For phy-

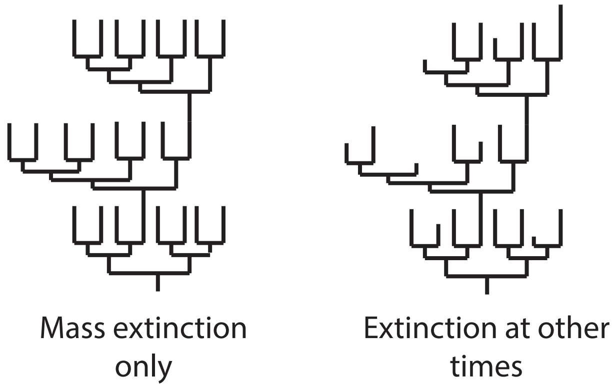


Figure 4.3: Schematic of the two options for extinction. On the left hand side all extinction occurs at periodic mass extinction events (this makes it easier to control the strength of clustering in the simulated extinctions) and on the right hand side there is also randomly timed extinction throughout the simulation, not just at mass extinctions.

logenetically random extinction the same overall proportion of lineages were terminated but terminations were selected randomly across the tree.

Following this, simulations of morphotaxa were restarted from those that had survived at the mass extinction boundary. This process was repeated until 110 simulation time units had elapsed, including a mass extinction event every 20 time units. In some simulations extinctions only occurred at these period events but in others they were also allowed to occur at other times (Fig. 4.3). A final ‘true’ phylogeny of evolutionary relationships between all the morphotaxa was estimated using the function `taxa2phylo` (Bapst, 2012). The morphotaxa were also sampled (at a specified preservation rate in per-lineage time units (ltu^{-1})) to represent the process of preservation in the rock record and sampling/discovery. These sets of sampling events were translated to ranges of fossil morphotaxa, from which a cladogram can be reconstructed. This cladogram was then timescaled using one of three methods that were outlined in Chapter 1 (*Hedman* section 1.4.4, *cal3* section 1.4.5, *mbl* section 1.4.3). The method that was used for timescaling in each case was dictated by the set of parameters under investigation, and is outlined in Table 4.1.

The parameters for the baseline simulation set (simulation set 1) were a budding model of cladogenesis with clustered extinction and extinction concentrated at mass extinction events. This baseline was chosen to most closely represent the data and method used in the previous chapter for real data.

Variations on the baseline simulation set were also tested, and these are shown in Table

4.1. Seven variations on the input simulation parameters were used to simulate the true phylogenies. These are shown as 1 to 7 in the 'Simulation' column of Table 4.1. Each of these seven simulation sets contained five 'true' phylogenies. Some of them are repeated in the table because for some tests the same simulation set was used but it were then treated differently to get results (e.g. different subsequent timescaling methods were used).

Each of the five true phylogenies in each simulation set (i.e. each row in the table) was then sampled 50 times at three different rates each, 0.01, 0.1 and 0.5 per-lineage time units (ltu^{-1}). The baseline output method of reconstruction was to use the *Hedman* method without reconstructing ancestors. Again this was to represent the previous treatment of real data in this thesis. Variations were tested by reconstructing phylogeny based on the sampled trees in different ways, yielding 13 sets of results with varying parameters. Overall there were 35 simulated true phylogenies and 1950 simulated phylogenies of sampled fossil taxa; processing was done in parallel on 16 cores. This was equivalent to approximately 22 months single processor computing time. These variations were chosen from the huge number of possible combinations in order to directly address questions of interest, the way in which they do so is outlined in Table 4.1 and subsequent text.

Following the reconstruction of the phylogeny based on fossil occurrences I measured Fritz and Purvis' *D* (Fritz and Purvis, 2010) for the same timeslice in both the true phylogeny and the reconstructed fossil phylogeny. I then varied different parameters in the original simulation of the true phylogeny, as well as sampling rate and the method of cladogram timescaling.The strength and significance of the phylogenetic clustering of extinction measured on each fossil phylogeny was then compared to the same measurement on the true phylogeny from which it had been sampled. This allowed assessment of which parameters were the most important controls on the ability of palaeobiological studies of this metric to recover the true signal.

Each of the combinations of parameters shown in Table 4.1 was chosen to address the effect of a particular potential bias or data treatment method on the ability to recover the

Table 4.1: Parameters used in simulating true phylogenies and in reconstructing them after sampling

Output	Simulation	Branching	Extinction	Clustered	Timescaling	Ancestors
1	1	Budding	No	Yes	Hedman	No
2	1	Budding	No	Yes	**mbl**	No
3	1	Budding	No	Yes	**cal3**	No
4	2	Budding	No	**No**	Hedman	No
5	3	Budding	**Yes**	Yes	Hedman	No
6	4	**Bifurcating**	No	Yes	Hedman	No
7	4	**Bifurcating**	No	Yes	**cal3**	No
8	4	**Bifurcating**	No	Yes	**cal3**	**Half**
9	4	**Bifurcating**	No	Yes	**cal3**	**Yes**
10	4	**Bifurcating**	No	Yes	**mbl**	No
11	5	**Bifurcating**	**Yes**	Yes	**mbl**	No
12	6	**Bifurcating**	No	**No**	**mbl**	No
13	7	**Both**	No	Yes	Hedman	No

The baseline parameters are in output 1, for each of the other configurations the differences from the baseline parameters are highlighted in bold. Additional detail on why each parameter set was chosen is provided in the main text below. The first column shows which set up was used to simulate the true data. The column titled Branching shows the model used to simulate the way in which morphotaxa diverge. The column titled Extinction shows **No** when extinctions were restricted to mass extinction events and **Yes** when extinction also occurred randomly via the birth-death process at times other than mass extinction events. The column titled Clustered shows whether or not extinction at events was constrained to be phylogenetically clustered. The column titled Timescaling shows the method use to scale the branch lengths of the reconstructed fossil cladograms to time. The column titled Ancestors shows whether or not ancestor-descendant relationships were estimated and included in the reconstructed timescaled cladogram.

'true' clustering signal. Output sets 1-3 were to investigate the effect of using a different method to timescale the reconstructed phylogeny after sampling. Output sets 4-6 were to test whether the pattern of the simulated 'true' speciation and extinction affected recovery of the correct result. Sets 7-9 were to test whether the *cal3* method of timescaling the phylogeny and allowing for ancestral reconstruction lead to a more accurate result. Sets 10-12 were to test whether the pattern of 'true' extinction affected the result when a bifurcating model of speciation was used (the *mbl* method was used to timescale these three sets because the processing time using *Hedman* for these bifurcating trees was prohibitively long.) Set 13 was to test what affect using different models of speciation to simulate the 'true' phylogenies had.

The method used here can be viewed as optimistic as only two factors (missing taxa and sampled ancestors) are being investigated. It is assumed in the reconstruction of cladogram topologies that the ideal evolutionary relationships are recovered, which is unlikely to be the case with real data. It is also assumed that there is no uncertainty in the ages of the fossil specimens, when in reality these are often only known to stage level precision (see previous chapters for examples). This analysis is designed to test for likely bias, and the cause of that bias, in analyses of extinction clustering. The results are not a perfect representation of how accurately clustering signals can be obtained, but an indication of where problems are likely to arise, and should be read as such. The code for all simulations and analyses can be found on the data appendix disk, with this the experiment can be fully repeated as there are no real data used.

4.3 Results

With the exception of the first results figure (Fig. 4.4) all of the boxplots in this section show the difference between the value of *D* estimated on the 'real' simulated tree and the values of *D* estimated on each of the 50 'fossil' trees that were reconstructed from sampling that one real tree. This was repeated 5 times and the results combined so each box in a boxplot represents 250 differences, i.e. how wrong the estimate from each of those fossil trees was, compared to the estimate made on the real tree that each was sampled on.

4.3.1 Sampling rate

The baseline simulation demonstrates clearly that accurate recovery of the degree of phylogenetic clustering of extinction is not guaranteed, whether or not extinction is clustered in (simulated) reality (Figs. 4.4a and b). At low sampling rates the value of *D* is on average higher (less clustered) than, or close to, the originally simulated value, and at high and medium sampling rates the value of *D* is consistently lower than the simulated value. In the case where extinction was not significantly clustered in the simulation, this corresponds to a high type I error rate. Only sets of results which provided interesting information above that which was found from the baseline simulation are presented.

Correct recovery of the strength of phylogenetic clustering of extinction depends heav-

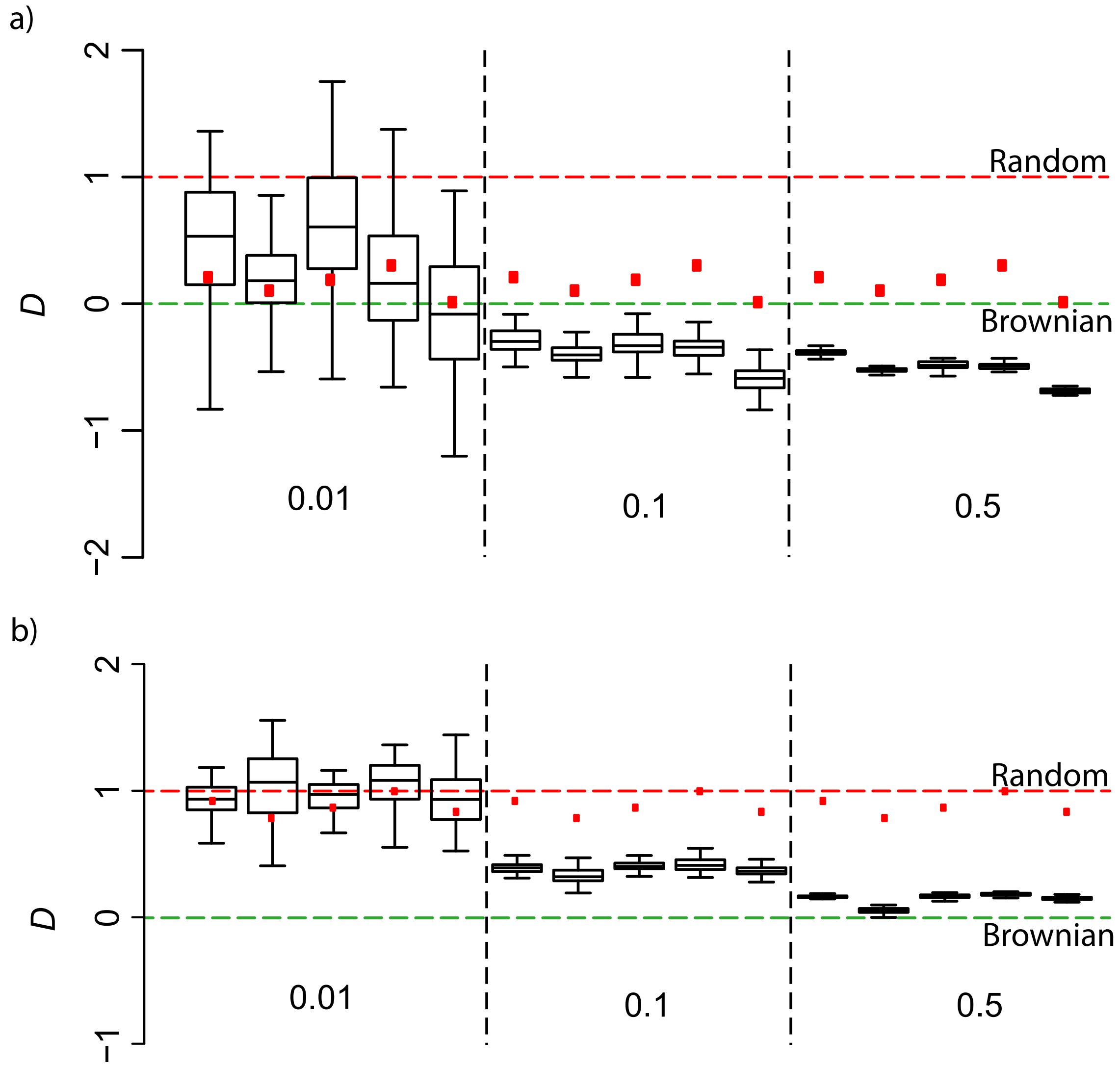


Figure 4.4: Results of simulation set 1 and 4. Five simulated phylogenies were sampled at three different sampling rates (0.01, 0.1 and 0.5 indicated at the bottom of the plot), red rectangles are the true values for *D* for each phylogeny. Box and whisker plots show the range of values of *D* measured on 50 timescaled cladograms for each box. Part a) shows results for when extinction in the simulation was phylogenetically clustered, and b) shows results for extinction that was phylogenetically random.

ily on sampling rate (Fig. 4.5a). For results when the sampling rate was 0.01 ltu^{-1}, on average the strength of the phylogenetic signal in extinction (*D*) was underestimated. A sampling rate of 0.1 ltu^{-1} usually resulted in overestimates of *D*, and a sampling rate of 0.5 ltu^{-1} always lead to overestimates of the strength of clustering of extinction.

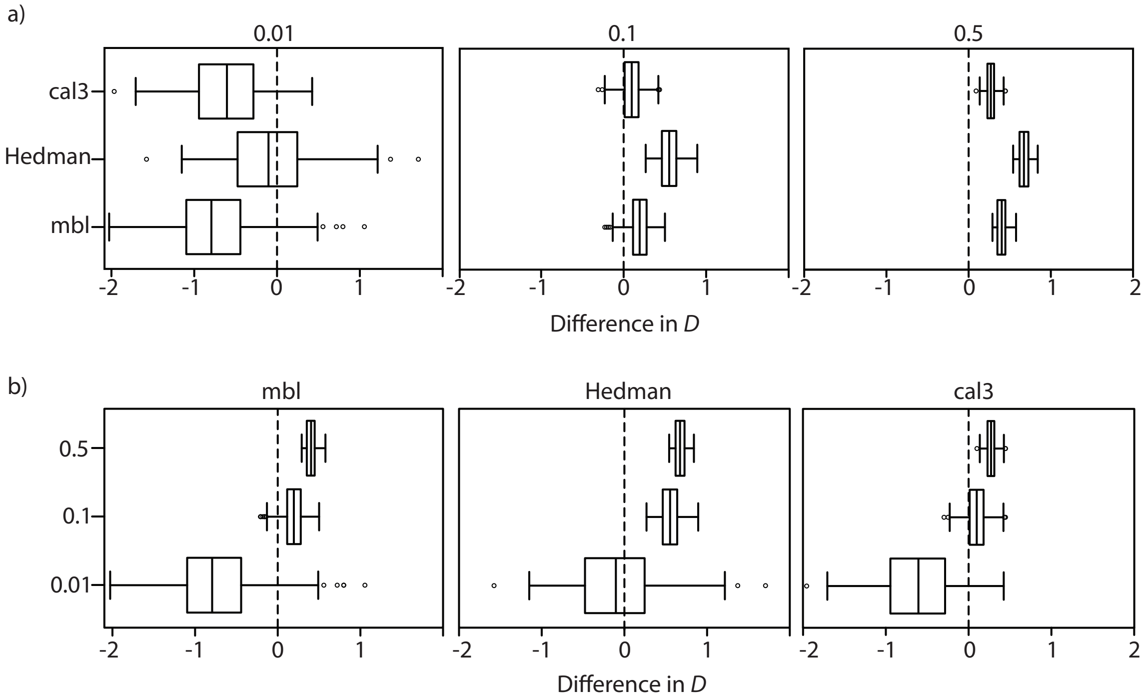


Figure 4.5: Results of simulation sets 1, 2 and 3. Part a) shows the median and interquartile ranges of the difference in estimated value of D from the true value of D for three different sampling rates from left to right, using three different methods to timescale the cladogram. They are plotted to highlight the influence of sampling rate. Part b) shows the same data but arranged to highlight the influence of timescaling method. The methods increase in complexity and amount of input data required from left to right. Values close to the dashed line at 0 on the plots indicate that good estimates were made on the timescaled cladograms, with reference to the simulated true phylogeny. The narrower a box is, the more consistent results across the many iterations of cladograms were.

4.3.2 Timescaling method

The method used to timescale the trees of fossil taxa also had an important influence on recovery of accurate estimates of D (Fig. 4.5b). At 0.01 ltu^{-1}, when the trees were timescaled using *mbl* and *cal3* this lead to large underestimates of clustering. When the trees were timescaled using *Hedman*, estimates at 0.01 ltu^{-1} were close to estimates of D from the real tree but showed a large variance across measurements from different topologies, with a median slightly underestimating the strength of clustering. At higher sampling rates *Hedman* timescaled trees gave D values which implied a far greater degree of clustering than the original simulated phylogeny. When the trees were timescaled using *cal3*, estimates were more accurate overall, although low and high sampling rates did lead to a slight underestimate and overestimate of clustering respectively. Trees scaled using *mbl*

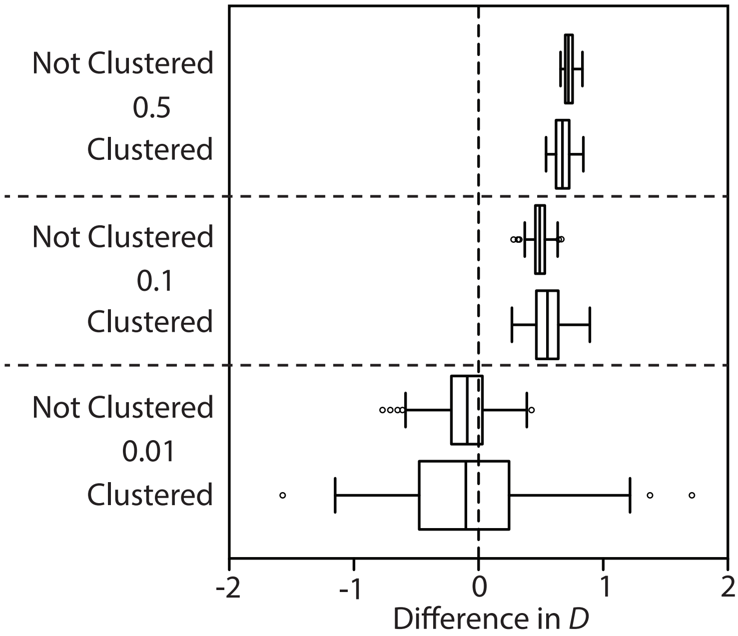


Figure 4.6: Results of simulation sets 1 and 4. The accuracy of estimates of *D* on the fossil trees compared to *D* on the true trees, at three different sampling rates when the simulated mass extinction events were, or were not, phylogenetically clustered when measured on the true tree.

did not give the most accurate estimates at any sampling rate, but were better than *Hedman* at the two higher sampling rates.

4.3.3 Strength of clustering and model

A factor that made a small difference in how effective the method was at producing accurate estimates of *D*, was whether or not extinction in the simulation was phylogenetically clustered (Fig. 4.6). When extinctions were phylogenetically clustered there was a larger variance in estimates from fossil trees than when extinction in the simulation was phylogenetically random. Medians of estimates for clustered and non-clustered extinctions showed approximately the same difference from the true value of *D*.

The model used to simulate the phylogeny in the first place had a small effect (Fig. 4.7). At 0.01 ltu^{-1} the most accurate results were obtained when morphotaxon differentiation occurred by budding cladogenesis, with a median inaccuracy close to 0, but a large variance in results across the tree set. Bifurcating and combination models lead to a slightly more negative median difference. At 0.1 ltu^{-1} the most accurate results were obtained when differentiation had occurred by bifurcating cladogenesis, the median was almost equal to real estimates of clustering and there was much lower variation in results from different trees. The effect of sampling rate was consistent with all the models, where low rates lead to more variation in results. The implications of this particular result are complex because the variable that was changed is one that would have to be independently estimated. This is discussed in section 4.4.2.

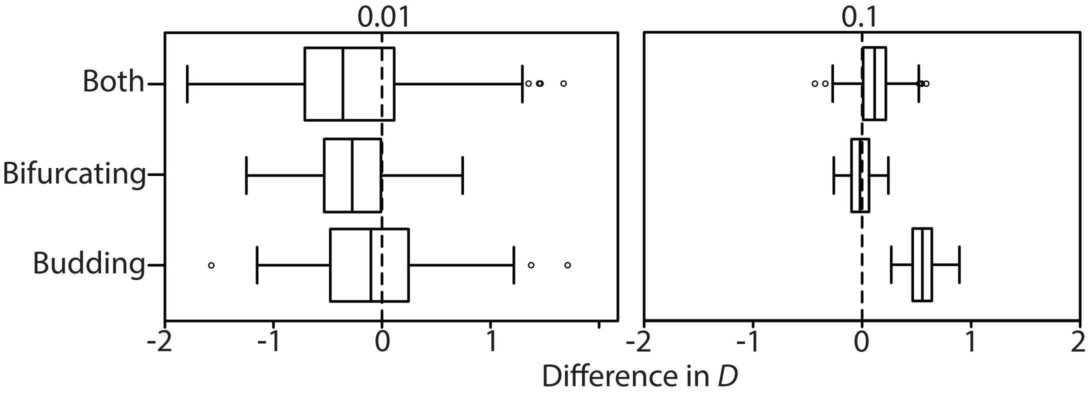


Figure 4.7: Results of sets 1, 6 and 13. A comparison between results when the models used to simulate the phylogenies used budding, bifurcating and a combination of both types of cladogenesis for simulating morphotaxon differentiation. A sampling rate of 0.5 is not included here because the processing time for the analysis was prohibitively long.

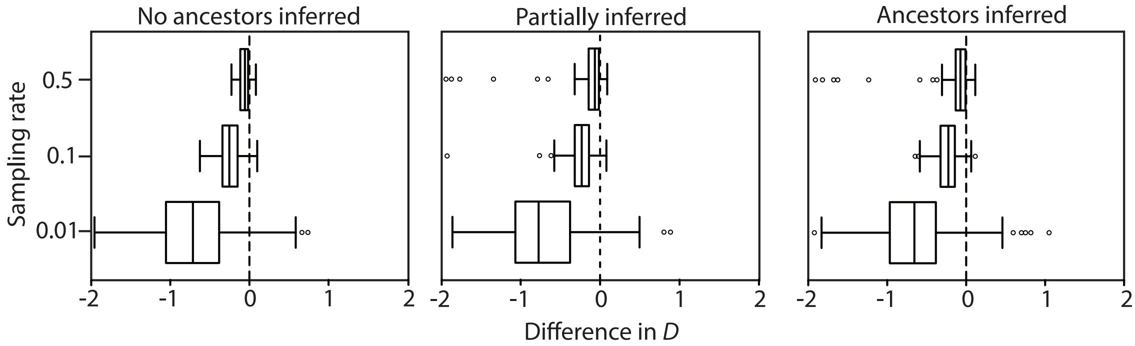


Figure 4.8: Results of sets 7, 8 and 9. A comparison between sets of simulations where the true phylogenies were the same but with an increasingly high probability of inferring ancestors from left to right, when the timescaled cladograms were estimated.

4.3.4 Ancestors

The *cal3* timescaling algorithm includes the possibility of inferring which sampled taxa might be ancestral to others and incorporates this into the resulting phylogeny. It does this by examining the stratigraphic position of taxa within polytomies. As ancestral relationships were hypothesised to be a potential cause of bias I ran simulations allowing full and partial ancestral inference. There is very little difference between the three sets of results (Fig. 4.8). This indicates either that ancestor descendant relationships are not an important bias in this analysis, or that the manner in which *cal3* reconstructs ancestral relationships fails to resolve problems associated with measurements of extinction clustering.

4.4 Discussion

These results indicate that there are several important factors that need to be considered when interpreting the outcomes of an analysis of phylogenetic clustering of extinction. Different methods of treating and analysing the data are effective depending on the type of data being used for the analysis. The way in which taxa in the clade under investigation evolved and became extinct also has an effect on the accuracy and precision of results, so caution must be taken when drawing conclusions from any one analysis. The results provided above can be used as a reference to find out what the likely biases are in a dataset once an analysis has been done (see section 4.4.2). Although many factors have an influence on the bias in simulation outcomes, the sampling rate has the largest effect. In addition once the sampling rate is known, the biases introduced by other factors can be avoided, or at least anticipated.

4.4.1 Causes of bias

Results suggest that the main bias at high sampling rates - towards overestimation of the strength of phylogenetic clustering - is a result of reconstruction of ancestors as sister taxa to their descendants, in combination with missing taxa, as these are the two problems that were introduced in the analyses. However, this bias does not seem to be due to the pseudoextinctions that the reconstruction causes, but instead to lineage extensions causing pseudosurvivals. The lineage extensions result in an increased number of survivals at the end of each timeslice. False survivals have the potential to bias measurements of D in either direction, but results indicate that overall they lead to clumps of closely related taxa surviving the end of timeslices; the way in which this could occur is shown in figure 4.9. Extinctions and survivals are symmetrical in the calculation of D, so an increase in survivals, where those survivals are in closely related taxa, has the same effect as an increase of extinctions in closely related taxa.

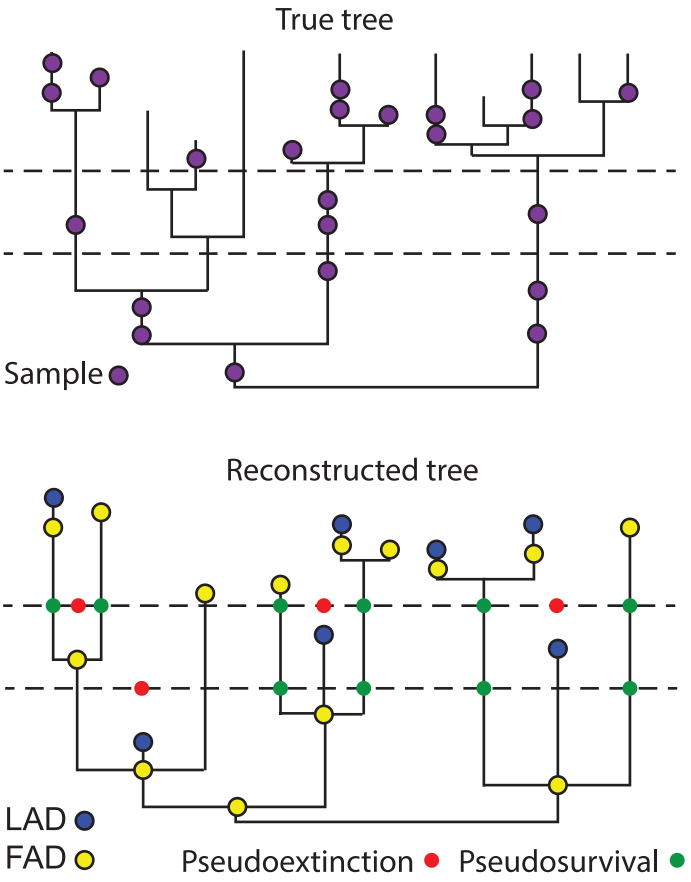


Figure 4.9: The effect of ancestors and descendants on the reconstruction of a cladogram and consequently when extinctions and survivals are inferred. In the top schematic black lines show the true history of lineage diversification of the clade and purple circles indicate hypothetical sampling events. In the bottom schematic black lines represent lineage durations and relationships based upon the sampled taxa. Yellow and blue circles indicate the ranges of morphotaxa, and red and green circles indicate inferred extinctions and survivals respectively, that did not actually occur.

This effect is seen for the higher sampling rates as these have an increased probability of sampling ancestors. At low sampling rates the signal is rarely measured as significantly clustered. At lower sampling rates pseudoextinctions are the more important cause of bias. Fewer samples means a lower incidence of sampling ancestor-descendant pairs (Foote and Raup, 1996), reducing the frequency of pseudosurvivals. The lineage extension effect is also present at low sampling rates, but because sampling is relatively even across the tree the signal of very closely related taxa surviving or becoming extinct at the same time is lost. For a poorly sampled tree the most closely related taxa that have actually been sampled will not necessarily have been closely related in absolute terms.

In order to confirm this interpretation of the cause of the biases I completed additional simulations based on set 10, but with a modification. In the simulations based on bifurcating cladogenesis it was possible to delete ancestral taxa that had been sampled from the tree, both before or after timescaling. When these were removed after timescaling (which removed pseudoextinctions but not pseudosurvivals) the signal measured actually shifted to a larger overestimate; when they were removed before timescaling (removing both pseudoextinctions and pseudosurvivals) the signal at high sampling rates shifted from an overestimate of the strength of clustering to a small underestimate (Fig. 4.10). This confirms my interpretation of the causes of signal bias, and in addition indicates that to a small degree pseudoextinctions act to bias the results in the opposite direction to the strong bias caused by pseudosurvivals, somewhat correcting for it. In general, pseudoextinctions bias the signal towards an underestimate of the strength of clustering, whereas lineage extensions bias it towards overestimation of the strength of clustering.

There is some ambiguity in the interpretations of models based on budding cladogenesis as it is not clear what the best way to accurately represent taxa which are related via budding might be. It is also not possible to remove ancestral taxa from the data because under a budding model an ancestral taxon survives the origin of its descendant and so its extinction is a real extinction. Measurements made using simulation sets based on budding cladogenesis showed more bias than those based on bifurcating cladogenesis, this is potentially due to the same effect described in the previous paragraph, and shown in figure

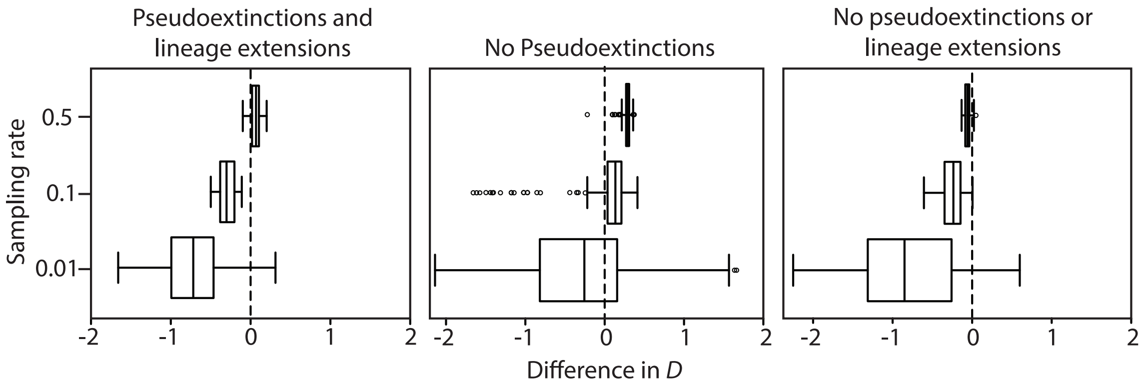


Figure 4.10: Results of simulation set 10 and two additional simulations where ancestors were removed from the phylogenies at different points in the analysis. Removing sampled ancestors after timescaling the cladogram results in removal of pseudoextinctions (centre), removing sampled ancestors before timescaling results in removal of pseudoextinctions and lineage extensions (right).

4.9. There is no *a priori* reason to expect that budding models would experience lineage extensions less frequently than bifurcating models. They should have less frequent pseudoextinctions however, because ancestors continue after speciation of their descendants under a budding model, so when they become extinct it is a real extinction. Perhaps the presence of pseudosurvivals and a reduced number of pseudoextinctions means one bias does not compensate for the other in the same way that they do under the bifurcating model. Unfortunately this cannot be tested in the same way.

4.4.2 Methodological recommendations

The correct way to implement and interpret measurements of the phylogenetic clustering of extinction is evidently a complex question. The nature of the data used for the analysis is important, as well as the way the data are subsequently treated. At high sampling rates the analysis is consistently prone to overestimation of the strength of clustering, which means that a significantly clustered signal could be found that is in fact an artefact of the analysis. What does seem clear is that sampling rate can often be estimated (to the correct order of magnitude), a timescaling method can be selected, but all other factors are either not possible to control, or difficult to estimate. With this in mind the most appropriate procedure is to estimate sampling rate, choose an appropriate timescaling method, and then interpret results in the context of the other biases that are likely to have occurred. For example if data are found to have a low sampling rate and significantly clustered extinction, then this result can be expected to have a large error. If the data shows random extinction at a low sampling rate it can be considered more reliable (Fig. 4.6).

At very low sampling rates, cladograms should be timescaled using the Hedman method to reduce bias, whereas at higher rates the *cal3* method should be used. Conveniently *cal3* is a method more suited to clades with higher sampling rates as it requires additional information (speciation, extinction and sampling rates) that can be more accurately measured for clades with a high sampling rate. Conversely the *Hedman* method has been shown to be more effective when sampling is low and the additional information on rates is not available.

In reality it is likely that models of budding and bifurcating cladogenesis do not adequately represent the complexity of real speciation processes, but given this complexity it is probably closer to some kind of combination of the two rather than one or the other. Although the model of speciation used does have an effect on the bias in the results, the direction of the bias is consistent for each sampling rate. In addition, inferring an appropriate model for speciation of a clade is not currently possible. What is important is the way in which speciation relates to sampled ancestors (section 4.4.1), which is in turn related to where pseudoextinctions and survivals are inferred. There is not a methodology available by which these issues can be avoided, but their likely effect must be considered a caveat of any extinction clustering analysis.

Ideally, to obtain an unbiased estimate of D, the phylogeny would be reconstructed using a method by which ancestors can be reliably inferred. *cal3* can in theory do this, but its ancestor reconstruction algorithm had little effect on recovery of simulated values of D (Fig. 4.8). Very new methods (e.g. Gavryushkina et al., 2014) which use Bayesian MCMC algorithms to model and allow for sampled taxa to be directly ancestral to others in the estimated phylogeny hold possibilities for ancestral estimation. These could be implemented to be used for phylogenetic comparative methods in the future, but the data required for this kind of inference are unavailable for the clades studied in Chapter 3, and for many other clades of fossil taxa. Although the response of downstream analyses has not been quantified for phylogenies inferred in this way, there is great potential for PCMs that are particularly vulnerable to bias caused by sampled ancestors, such as the method used here.

4.4.3 Implications for Chapter 3

Due to the large proportion of point occurrences in the datasets the sampling rate could not be directly estimated for the clades investigated in Chapter 3. It is likely though, given this information, that the sampling rate for all three clades was very low. The literature indicates that sampling rates which translate to the region of 0.1 lmy^{-1} can be expected for marine invertebrate taxa, for Neogene mammals and well preserved marine invertebrate records a rate of 0.5 lmy^{-1} might be possible (Foote and Sepkoski, 1999; Alba et al., 2001). For the majority of terrestrial vertebrate clades, particularly those which include many point occurrences, the sampling rate is likely to be on the order of 0.01 lmy^{-1}, or lower. Consequently the simulation results at 0.01 lmy^{-1} are the most representative of the bias that can be expected in my results in the previous chapter.

Results at 0.01 ltu^{-1} are on average biased towards underestimating the strength of clustering. However, there is a large variation from results that were measured on different samples of the record, which includes estimates of strong clustering. Although given the number of timeslices that show clustered extinction suggests it is unlikely, it is possible that each of these estimates represents the ‘whisker’ in the variation and that extinction was not in fact phylogenetically clustered.

More uncertain than the results for the many clustered timeslices is the estimate for the stage that showed random extinction for sauropsids (the Carnian). This simulation analysis suggests that there should be a bias towards a signal of dispersed extinction. In addition there is only one timeslice that is random, and given the number of tests performed we would expect at least one of the bins to show phylogenetically random extinction. It is therefore unclear whether or not this estimate is a good one or is an artefact.

4.5 Further work

An area for further investigation relates to the comparison between simulations and real results. In the original simulations of true phylogenies, extinction that was significantly phylogenetically clustered could only be achieved by setting the proportion of taxa ‘killed off’, with a trait value above the chosen threshold, to close to 1. This strict of a cut off is probably biologically unrealistic because inherited traits confer extinction vulnerability, not certainty. Even so it was not possible to simulate very low (highly significantly clustered) values of D. These values were only produced when fossil trees derived from records with high sampling rate were tested. At low sampling rates significant clustering was rarely observed. This calls into question how, at the low sampling rates seen for terrestrial vertebrate clades, significant clustering was so commonly found. One possibility might be that there is a taphonomic bias; for example during a period of relatively localised high ecological stress there might be high extinction intensity across unrelated taxa, but a bias in preservation potential towards those with a particular habitat (e.g. aquatic). With a sparse record these preserved aquatic taxa might only be sampled at this point in time, and their less closely related terrestrial contemporaries never sampled at all. This would mean our measurement would be based on a biased sample of a set of closely related taxa that went extinct at the time, and missing the more distantly related taxa that also became extinct. A further possibility is that the birth-death model used to simulate the true phylogenies is not a good model for the evolutionary process (Hagen et al., 2015). This is discussed in more detail in the conclusions section of the thesis, in the context of future work in combination with other chapters.

Chapter 5

Phylogenetic interpretations of fossil trait data: a case study of Sauropterygia

> *"...but herein a palaeontologist's mode of work is like that of antiquaries of another order, who read inscriptions on roman buildings by the nail-marks when the letters themselves have been wrenched off..."*
>
> - Richard Owen *A History of British Fossil Reptiles*

In this chapter I present an analysis of a large dataset of Sauropterygian body plan and functionally relevant traits. I use a phylogeny of the clade and the continuous trait data to measure the comparative rates of evolution of aspects of the body plan to explore the influence of phylogenetic and environmental constraint on body plan evolution in the clade. I also use various modelling techniques to map shifts in evolutionary regime through time which affected sauropterygians. I use these analyses to answer three key questions: 1) What mechanisms controlled the evolution of body plan in the clade? 2) Did the transition to open water habits at the end of the Triassic correspond to a shift in evolutionary constraint on body plan? 3) Do plesiosaurian morphotypes represent adaptive peaks in the evolutionary landscape?

The majority of work for this chapter was completed by me. An additional contributor - Roger Benson - provided a set of MPTs for Plesiosauria + pistosaurs based on an unpublished matrix of 102 taxa (275 characters), extending the taxon list of Benson and

Druckenmiller (2014), details of the algorithm and method used for the cladistic analysis of this matrix, and photographs of 27 of the specimens used. Character data for the additional taxa are included in the data appendix disk.

Terminology: For convenience, unless explicitly stated I use ‘basal’ to refer to the paraphyletic group of all sauropterygians outside Plesiosauria. All other clade names correspond to the monophyletic groups defined in Benson and Druckenmiller (2014).

5.1 Background and data collection

5.1.1 Phenotypic evolution

In this chapter I explore the possibilities of phylogeny-based analysis of continuous morphological trait data from fossils. Phylogenetic frameworks both enable researchers to correct for the evolutionary relationships between taxa, and allow us to make positive evolutionary inferences, thereby offering a means by which fossil data can be rigorously analysed (Pennell and Harmon, 2013). There are now a wide range of phylogenetic comparative methods available that can be implemented on non-ultrametric trees. Assessment of whether or not these are suitable for such trees has revealed where incorrect or biased results may occur and will continue to be important as new methods are developed (Slater (2014), Chapters 2 and 4 of this thesis). These methods open the door to quantitatively testing some of our many qualitative theories on the presence or drivers of macroevolutionary patterns, that may have been based on fossil or geological data but not yet statistically assessed (Kemp, 1999). The fossil record provides an opportunity to study the evolution of phenotypes over long timescales and through major environmental and evolutionary transitions (Hunt and Rabosky, 2014).

The period of time over which new advantageous phenotypic traits spread throughout populations has been shown to be orders of magnitude shorter than the time between samples in the majority of fossil time-series (Bell et al., 2006; Hunt, 2007; Tyler and

Leighton, 2011; Azzurro et al., 2014). However, long-term trends across large clades, and relative rates of trait evolution within such clades, have the potential to shed light on many unresolved questions in macroevolutionary theory (e.g. niche filling (Rabosky and Hurlbert, 2015); biological versatility (Vermeij, 1973); morphological complexity (Adamowicz et al., 2008); ecological innovation (Benson et al., 2014a)).

Sauropterygia represents a classic example of a clade about which there are many qualitative theories on how phenotype evolved (Andrews, 1922; Bakker, 1993) as these were some of the first articulated fossil reptiles found and described, during the 1800's (Conybeare, 1824; Owen, 1840; Hector, 1874). Since then the enormous variety in body proportions within Sauropterygia and observations of apparent morphological convergence (Storrs, 1993) has also inspired quantitative analysis of the evolution of their body plan (Caldwell, 2002; O'Keefe, 2002; O'Keefe and Carrano, 2005). The clade survived the Triassic-Jurassic boundary, where low sea level may have reduced richness to as few as one boundary-crossing lineage, then underwent successive radiations in an open water habit (Benson and Druckenmiller, 2014). Sauropterygia therefore provides an opportunity to test interesting macroevolutionary hypotheses relating to development, constraint and ecology, and convergence.

This chapter comprises three main sections. In this section, the first, I provide an overview of the history of Sauropterygia and patterns of extinction in the clade as well as outlining the type of data used in the subsequent analyses and how it was collected. In the second section I use comparisons of the rate of evolution of traits associated with axial body plan to assess the effects of developmental processes modulating vertebral counts and body region patterning in macroevolution of body shape. This includes testing three hypotheses:

1. Sauropterygian axial body plan changed through somitogenesis and homeotic effects rather than differential somitic growth

2. Somitogenetic and homeotic effects were decoupled in Sauropterygia

3. Extreme body plans (e.g. very long necks in elasmosaurids) evolved via high rates of pleomerism

In the third section I use a variety of analytical methods to establish the effects of major environmental transitions, such as to pelagic life at the the Jurassic-Triassic boundary, on rates and constraints of phenotypic evolution. This includes testing two hypotheses

1. There was a change in evolutionary regime (best fitting model of evolution) at the Tr/Jr boundary

2. The pliosauromorph body plan represents a temporally extensive adaptive peak available in open water habitats after the Tr/Jr boundary

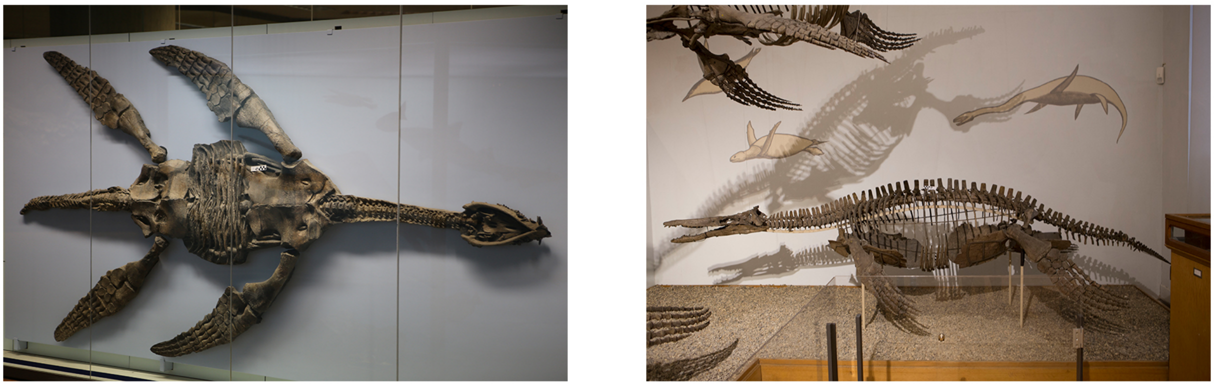
I finish the chapter by summarising the results obtained as a whole.

5.1.2 Sauropterygia

Sauropterygians were a highly abundant and long-lived monophyletic clade of marine reptiles, spanning a period of around 180 Myr during the Mesozoic (Motani, 2009). The group originated in the Early Triassic with several near-shore radiations including placodonts, nothosaurs, pachypleurosaurs, cymatosaurs and pistosaurs (Rieppel, 2000; Neenan et al., 2013). Diversity of these early groups peaked around the early Ladinian before eustatic sea-level fall during the Late Triassic coinciding with extinction of all but a minimum of one lineage within Pistosauria (Rieppel, 2000; Benson and Butler, 2011; Kelley et al., 2012). These subsequently diversified during the Jurassic and Cretaceous to form Plesiosauria (Benson and Druckenmiller, 2014). Sporadic minor extinctions occurred throughout the rest of the Mesozoic, including a protracted period of high turnover at the Jurassic-Cretaceous boundary (Bakker, 1993; Benson and Druckenmiller, 2014), until all remaining representatives went extinct during the Cretaceous-Paleogene mass extinction event (Mulder et al., 2000; Vincent et al., 2011).

a) *Placodus* - placodont

b) *Lariosaurus* - nothosaur



c) *Meyerasaurus* - plesiosauromorph

d) *Liopleurodon* - pliosauromorph

Figure 5.1: Examples of the variety of form within Sauropterygia. In a) *Placodus* and in b) *Lariosaurus* general body plan of some basal taxa living in the Triassic are shown. In c) and d) *Meyerasaurus* and *Liopleurodon* represent the two major 'morphs' of Jurassic and Cretaceous taxa that evolved in several lineages. In these two taxa the characteristics outlined in Table 5.1 can be seen. *Meyerasaurus* has a long neck, small skull, a short ischium (the front flat bone in the pelvis area, immediately posterior to the ribcage) and the front limbs are slightly longer than the back limbs. The opposite characteristics can be seen in *Liopleurodon*, particularly a much longer ischium relative to the length of the trunk.

Due to what is thought to be convergent evolution of overall body plan (plesiosauromorph and pliosauromorph: Table 5.1, Figure 5.1) the taxonomy of Plesiosauria remained ambiguous into the 1990s (O'Keefe, 2002; O'Keefe and Carrano, 2005). With the increasing availability of large character datasets, cladistic analyses are now in agreement that these morphotypes evolved multiple times within different lineages, and large inter-species variation in many traits (e.g. cervical count, neck length, skull size) suggests that the clade was highly evolutionary labile (O'Keefe, 2002; O'Keefe and Wahl, 2003).

The evolutionary history of Sauropterygia includes a major ecological transition. Triassic non-plesiosaurian taxa inhabited near-shore environments and a marine regression prior to the end of the Triassic may have caused the extinction of these taxa (Benson and Butler, 2011; Kelley et al., 2012). An expansion of fossil occurrences of Pistosauria to deep-water facies (Rieppel and Hagdorn, 1997) and the modification of limbs to form flippers (Storrs, 1993; Caldwell, 2002) suggests either that lineages that had already made the transition to open water environments were the only survivors, or that the regression forced surviving lineages into these habitats. It has been proposed that combinations of feeding and locomotion strategies – inferred from axial, skull and limb configurations – that evolved following this transition in Jurassic and Cretaceous Plesiosauria corresponded to adaptive peaks in the macroevolutionary landscape (O'Keefe, 2002). The 'plesiosauromorph' and 'pliosauromorph' body plans (Table 5.1, Figure 5.1) are character combinations which represent these peaks (O'Keefe, 2002). However, body plans within these two 'morphs' show variation and whether an adaptive peak model adequately describes this has not yet been rigorously assessed.

Table 5.1: Key contrasting features of plesiosauromorph and pliosauromorph body plan

Pliosauromorph	Plesiosauromorph
Short neck	Long neck
Large skull	Small skull
Long ischium	Short ischium
Femur longer than humerus	Femur shorter than humerus

Traits of the neck (cervical vertebral count and neck length compared to body size) show the largest variation across taxa (Rieppel, 2000; O'Keefe, 2002). Early groups such as the armoured placodonts have comparatively short necks and low vertebral counts (e.g. six cervicals and neck one seventh of trunk length in *Placodus gigas*: Rieppel (1995)), later radiations like the elasmosaurids had necks much longer than the body and high cervical counts (e.g. 60 cervicals and a neck nearly three times body length in *Hydrotherosaurus alexandrae* Sato (2002)). Due to this high variation in axial body plan ratios and verte-

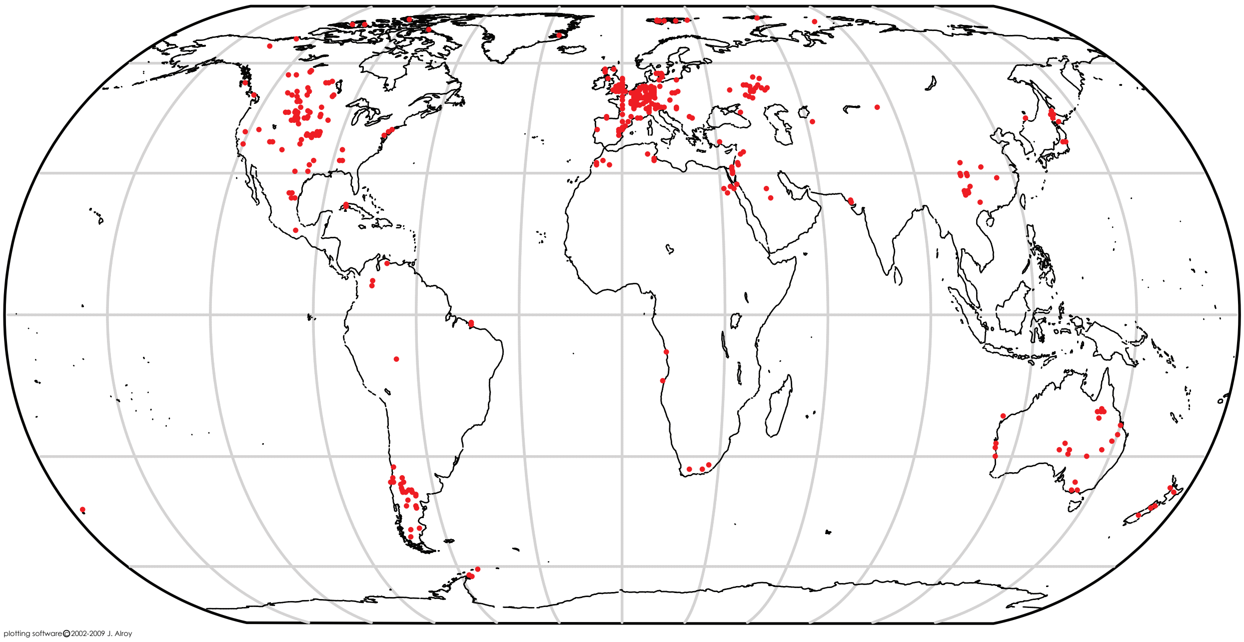


Figure 5.2: Map downloaded from fossilworks (www.fossilworks.org) showing as red points every recorded occurrence of a Sauropterygian fossil in deposits of any age. This shows the distribution of rocks bearing these taxa.

bral counts, sauropterygians represent a unique opportunity to study in detail the possible mechanisms of body plan evolution and its relationship with functional and developmental constraint. It has been suggested that adaptations to extreme habitats such as marine life can impose constraints on axial body plan evolution, leading to homogeneity across lineages, especially in taxa with ‘fish-like’ body plans such as ichthyosaurs and cetaceans (Müller et al., 2010). Deviation from fish-like morphologies in sauropterygians, and the apparent high plasticity in sauropterygian body plan indicates that this may not always be the case. An assessment of the underlying mechanisms of these patterns will help to elucidate whether they were due to different adaptive optima in the evolutionary landscape, inherent plasticity in development, or to release from functional constraint.

Fossils of sauropterygians are globally distributed in shoreline and marine facies (Fig. 5.2), but several key horizons have yielded significant numbers of specimens. Areas in western continental Europe provided many early discoveries of non-plesiosaurian (basal) taxa from the Triassic Muschelkalk units (Haas, 1980; Rieppel and Kebang, 1995; Klein, 2009; Klein and Scheyer, 2013). The Oxford Clay Formation, Kimmeridge Clay Formation and Lias Group exposed across the UK contain many well preserved Jurassic Plesiosaurian specimens (Owen, 1861, 1865; Seeley, 1865; Andrews, 1910a, 1913; Vincent and Benson, 2013; Benson and Bowdler, 2014). A wide age range of taxa have been found in the basins of the USA (Sander et al., 1997; McKean, 2012; Albright et al., 2007). Two relatively recent developments are discovery of many basal species from China, where many new taxa are being described (Rieppel, 1999; Chun and Rieppel, 2002; Zhao et al., 2008), and excavation of Jurassic and Cretaceous plesiosauroids from permafrost areas in Svalbard (Knutsen et al., 2012; Hurum et al., 2012). Differences in faunal composition through time are thought to be real, rather than a facies bias, supporting the hypothesis that disappearance of genera and families from the fossil record represents actual extinction rather than taphonomic bias (Rieppel, 2000). Some recent cladistic analyses find the clade to be the sister group to lepidosauromorphs or at the base of Lepidosauramorpha (e.g. Laurin and Reisz, 1995; Rieppel, 1998), others find them nested within other marine reptile taxa as the sister group to Sauria (archosaurs and lepidosaurs) (Neenan et al., 2013). In either case they fall well within Diapsida.

5.1.3 Dataset

Phenotypic data were collected from a range of mostly complete specimens, which I either measured directly, or from photographs using the software ImageJ when direct measurement was not possible (Abràmoff et al., 2004). I collected a series of body plan and skull measurements (Fig. 5.3). Body plan measurements were vertebral counts in the major axial regions (cervical, dorsal, sacral and caudal); lengths of the whole body (snout to distal caudals), neck, trunk, tail, the proximal three segments of each limb (humerus, radius, ulna, femur, tibia and fibula); and lengths of some girdle elements (ischium, pubis, corocoid, scapula). Skull measurements were length of the cranium, mandible, mandibular symphysis, retroarticular process, pre and post orbit, jaw depth and cranial height. I measured 102 specimens in total, accurate to an error of 5 mm. The majority of taxa were missing some of the measurements, the most frequently available measurement was cervical vertebral count and this lead to a total dataset of 81 taxa that could also be reliably phylogenetically placed.

For the vertebral counts, cervicals were defined as those vertebrae functionally belonging to the neck, when the position of the pectoral girdle and ribs allowed this to be determined. The atlas-axis complex was counted as two elements. When the pectoral girdle was not present or had moved from life position, the first dorsal was counted as the first element where the rib formed a contact with both the centrum and the neural arch (i.e. the first pectoral vertebra). The sacrals were those which connected to short, sturdy ribs that would have formed attachement to the pelvic girdle, and the caudal vertebrae as all those more posterior to the sacrals, bearing a rib facet on the centrum. This resulted in some counts which differ from those previously reported in the literature (e.g. cervical vertebral count for *Yunguisaurus*: Sato et al. (2010), cervical vertebral counts in O'Keefe (2002)).

It was possible to get a few more measurements by interpolating them from measurements of other taxa. This was done for 5 individual measurements overall, which are shown in bold in Appendix C. Where a species was represented by additional specimens possessing the missing trait I calculated the ratio of the trait measurement to a closely associated trait present in both (e.g. ratio of radius to humerus), and then used this ratio to estimate the missing trait measurement scaled from the trait measurement present in both. If a conspecific was unavailable but there was a closely related congeneric which was morphologically similar I used measurements from this to scale the estimated values. There were not many species for which more than one specimen was available so although this did increase taxa with complete trait data somewhat, the total number of taxa for which all the traits of interest were available was 28 (matching the number of taxa for which a measurement of the ischium was available).

Measurement data for all taxa that was used in the thesis is included in Appendix C

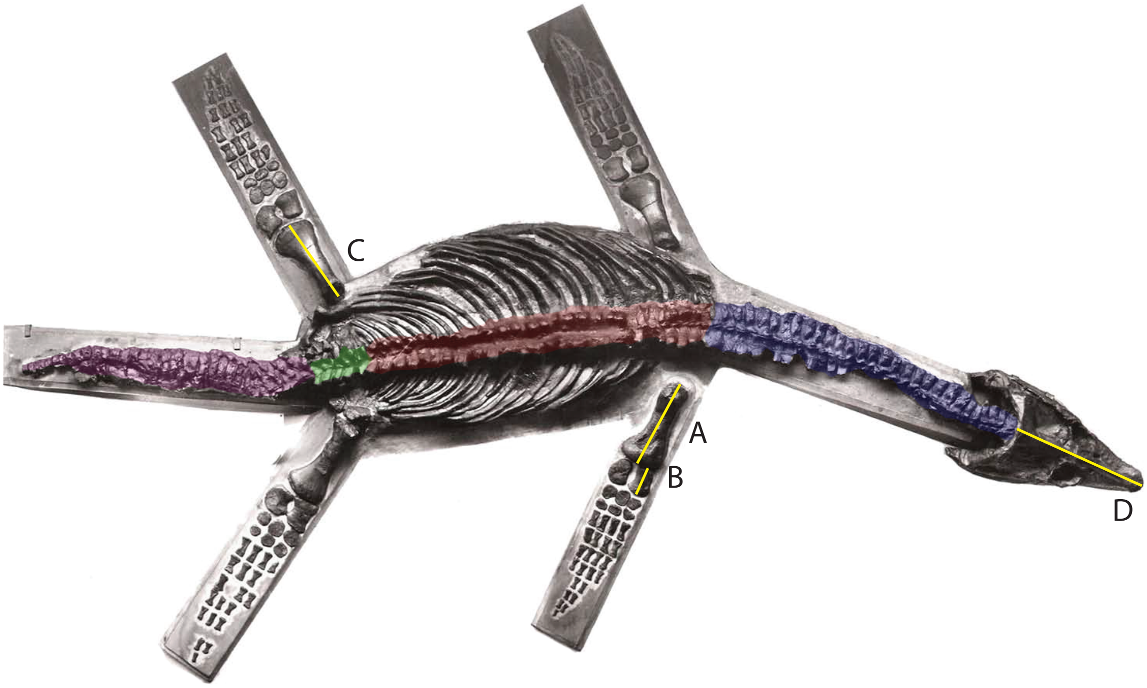


Figure 5.3: Schematic of the body plan data collected from each specimen and then used in analysis of evolutionary rates, evolutionary model fitting, and trait optimum identification. A) length of humerus B) length of radius C) length of femur D length of cranium. Coloured shaded areas indicate vertebral counts: blue=cervicals, red=dorsals, green=sacrals, pink=caudals.

and the full dataset including all the measurements made and metadata is included on the data appendix disk.

5.1.4 Phylogeny

Initial attempts to assess relationships among sauropterygian taxa were confounded by apparent convergent evolution of similar morphotypes (Andrews, 1910a, 1913; Owen, 1840; Persson, 1963). Despite this it was recognised relatively early on that the Triassic nothosaurs and pachypleurosaurs were close relatives of the larger ‘plesiosaurus’ and ‘pliosaurus’ taxa of the Jurassic and Cretaceous (Lydekker, 1889). The placement of placodonts within Sauropteryia, with their hard carapace, took longer to establish (Rieppel, 2000). Understanding of the phylogenetic relationships between taxa has advanced rapidly in recent years, particularly for plesiosaurians. The branching order between basal taxa still remains uncertain, due particularly to debate over whether they originated in the eastern or western Tethys (Neenan et al., 2013), and awaits a large scale genus level cladistic analysis. However, formal cladistic solutions to interclade relationships have been proposed (Rieppel, 2000; Neenan et al., 2013).

To construct the phylogeny used in this chapter I took an updated, more inclusive, version of the relationships proposed in the most recent cladistic analysis of plesiosaurian taxa, using the same character list but an expanded taxon list as Benson and Druckenmiller (2014). Tree searches were performed in PAUP* 4.0b10 for Macintosh. Initial exploration for shortest-length tree islands was conducted using four independent randomisations of the Parsimony Ratchet implemented by PAUPRat. The resulting subset of most parsimonious trees (MPTs) was then used as the starting point for TBR (tree bisection and reconnection) branch swapping. 'Wildcard' taxa were identified by inspection of the Adams consensus and pruned from the set of most parsimonious cladograms.

I randomly selected 100 of these most parsimonious cladograms and combined them with the most recent comprehensive hypotheses of non-plesiosaurian relationships (Rieppel, 2000; Neenan et al., 2013) to generate a composite tree. The influence of topological uncertainty on results was assessed by completing all the following analyses on this set of 100 trees containing 139 taxa. A representative topology is shown Fig. 5.4 and the tree set and range data are included on the data appendix disk. I timescaled the set of trees using the Hedman (2010) method outlined in Chapter 1 and implemented in Chapters 3 and 4, the resolution of the range data for taxa was stage level.

5.1.5 Phylogenetic clustering of extinction

Extinction and diversification in Sauropterygia seems to show three major radiations, one in each period of the Mesozoic. Basal taxa appear in the Early Triassic and almost all become extinct at the Triassic-Jurassic boundary (Neenan et al., 2013). Plesiosauria make their first appearance in the Jurassic, represented by rhomaleosaurids, pliosaurids, microcleidids and cryptoclidids, which were then replaced by elasmosaurids and leptocleidians

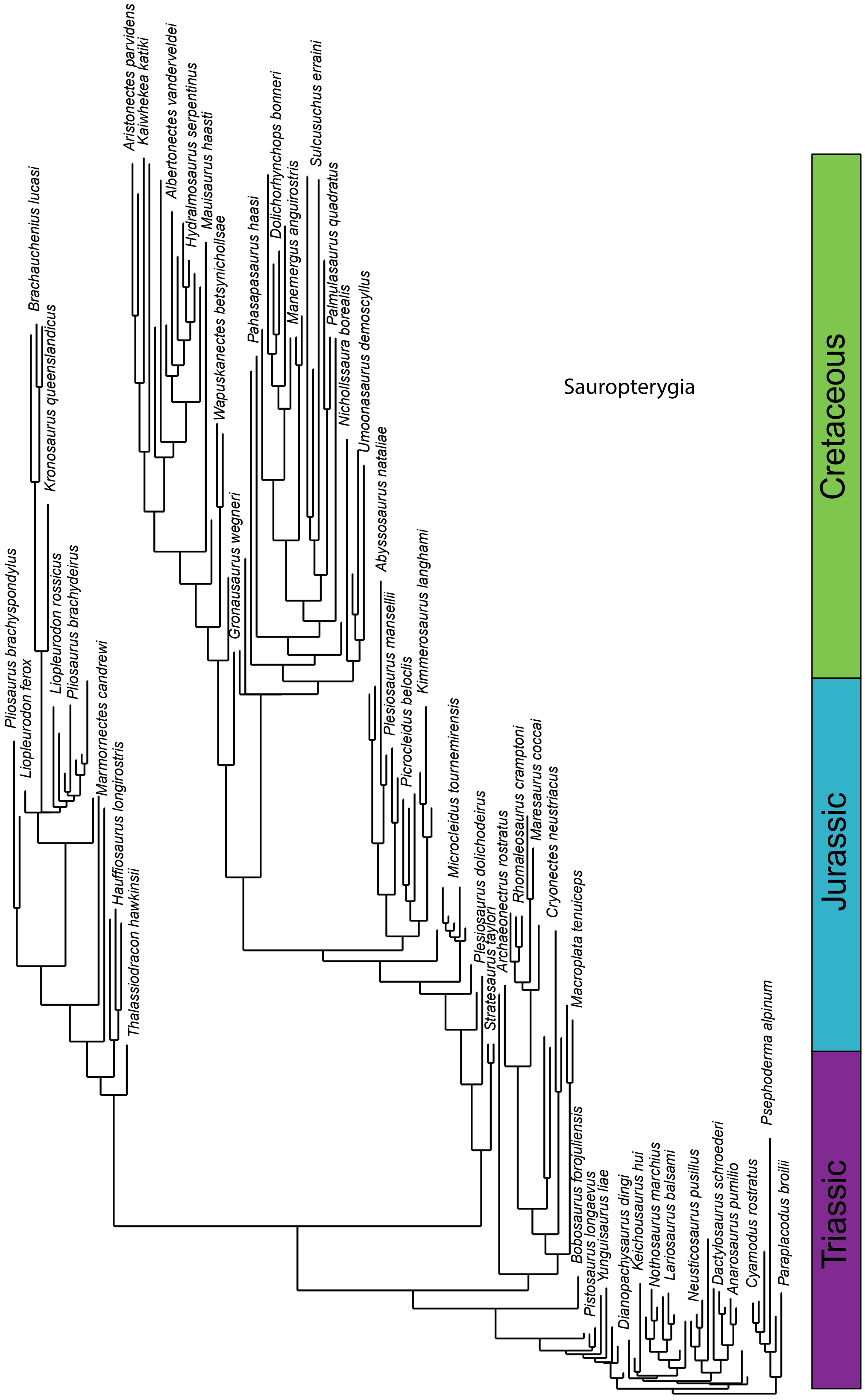


Figure 5.4: An example of one of the 100 timescaled phylogenies of the clade Sauropterygia which were used in all the following analyses. Some taxon names were removed for legibility. The full taxon list is available on the data appendix disk.

(including leptocleidids and polycotylids) in the Cretaceous (Benson and Druckenmiller, 2014). Ecological selectivity against extreme morphologies (e.g. very short neck and long head of derived pliosaurids) leading to a higher probability of extinction has been proposed, tying in with the idea that the observed body morphologies appeared repeatedly to fill in an adaptive peak when it had been vacated through an extinction event (O'Keefe, 2002). Following the method laid out in Chapter 3 I measured the phylogenetic clustering of extinction in the clade using the metric *D* (Fritz and Purvis, 2010). This approach was applied to a subsample of 100 trees of Sauropterygia.

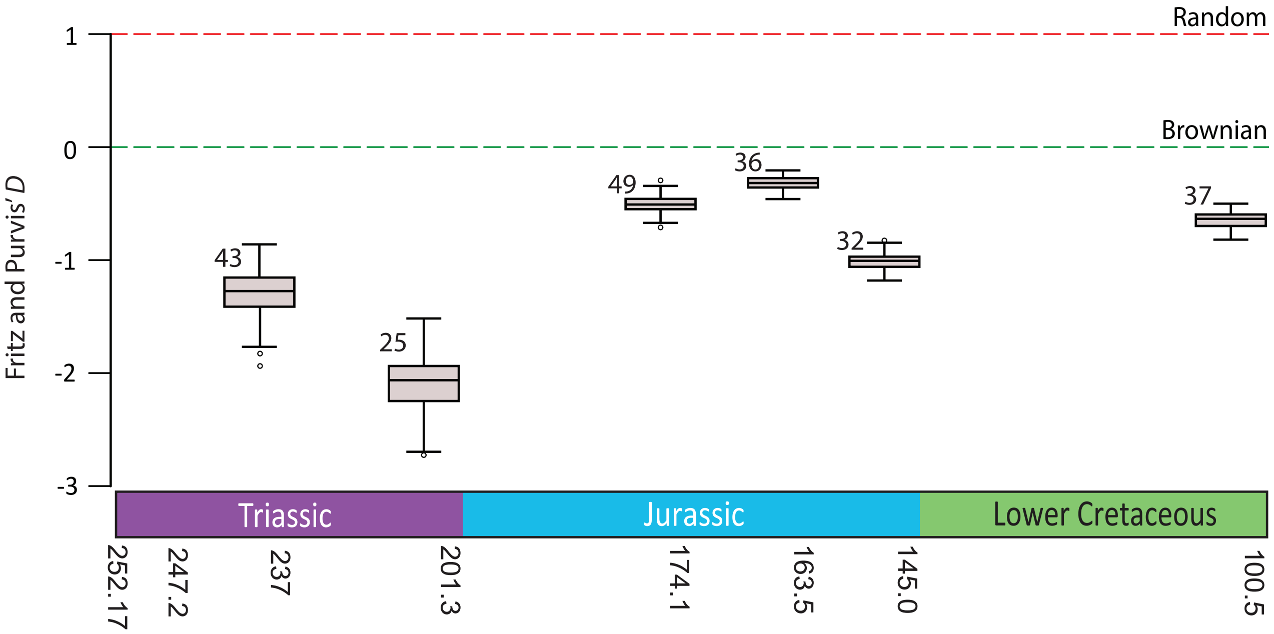


Figure 5.5: Values of Fritz and Purvis' D for Mesozoic epochs (Middle and Upper Triassic, Lower, Middle and Upper Jurassic and the Lower Cretaceous), with reference to the value for *D* expected for a Brownian trait distribution (0) and a random trait distribution (1). Boxes represent results for each timeslice from the set of 100 trees and numbers to left of each box are average sample size for the timeslice. All results were within the distribution for the Brownian expectation and more clustered than the random expectation, although did vary in the strength of clustering through time.

Results of this analysis indicate that, as might be expected given the loss of basal taxa, extinction at the Tr/Jr boundary was highly selective, falling in the top 0.5% quantile of the distribution for the Brownian expectation (Fig. 5.5). The measurement for every timeslice is significantly different from random and placed high within the Brownian distribution. Therefore, turnover within Sauropterygia was generally selective at a time resolution of around 30 Myr throughout their evolutionary history. In addition to the lowest value (greatest clustering) at the Tr/Jr boundary, the value for the Jurassic-Cretaceous boundary is also very low, supporting the three radiation model for the clade.

Although sample sizes in each timeslice are above the threshold for powerful analyses suggested by the method's authors (Fritz and Purvis, 2010) (particularly as resolution is high at $< 95\%$ and trait prevalence is even, falling between $30 - 80\%$ in all time bins), they do all contain a limited number of taxa so whilst it is likely that extinction was indeed phylogenetically clustered (Chapter 4), the results should be interpreted cautiously.

5.2 Rates of axial body plan evolution

5.2.1 Background

Palaeontologists rarely have available a fossil record of sufficiently high resolution to measure driven trends in phenotypic change on microevolutionary timescales (although see; Bell et al., 2006; Tyler and Leighton, 2011). However, observations of evolutionary transitions in the fossil record, supported by knowledge of the ages of taxa provided by stratigraphy, mean that comparative rates of macroevolutionary change in morphology have long been of great interest (Niklas, 1978; Hunt, 2012; Benson et al., 2014b). Many new methods have recently been developed to assess change in evolutionary rate through time and among lineages (e.g. Butler and King, 2004; O'Meara et al., 2006). Comparisons between relative rates across clades for different univariate traits may help to identify answers to questions about evolutionary mechanisms.

An area of current interest to evolutionary biology is the process of formation of repeated elements during vertebrate embryonic development. Somitogenesis is the process by which repeated axial segments are generated in the embryo, by budding from the anterior mesoderm layer. The process is controlled by a molecular oscillator that 'ticks', periodically triggering budding of a new somite (Dequéant and Pourquié, 2008). If the clock is fast then a high number of relatively small somites will be produced, as opposed to fewer large somites if the clock is slow (Gomez et al., 2008). These somites form the presacral vertebral elements, but which type of element (dorsal or cervical) they become is controlled in a separate process by the timing of activation of *Hox* genes (homeotic effects; Iimura et al., 2009). Due to homeotic effects two taxa can have identical pre-sacral vertebral counts, but this count might comprise two thirds cervicals in one and only one fifth in the other. The formation of caudal vertebral elements occurs at a different budding site (Polly et al., 2001), therefore allowing cervical and dorsal ratios and number to be analysed independently of them. Rates of change in total count is referred to as ‘somitogenetic’ and change in ratios of one region to another as ‘homeotic’, herein. Increase in overall body size through addition of vertebral elements is referred to as pleomerism. A further control on lengths of axial elements is somatic growth, which occurs in the post-embryonic stage. Differential growth of somites among body regions could result in evolutionary change in body proportions in the absence of pleomerism.

A previous study of fossil and modern taxa indicated that somitogenetic and homeotic effects were uncorrelated across Amniota (Müller et al., 2010), supported by the observation that somitogenesis and axial regionalisation occur at different times during development and are induced through separate regions of the pre-somitic mesoderm. It has also been suggested (based on the high vertebral count in some clades) that sauropterygians evolved long necks through pleomerism, and changed their body proportions through change in vertebral counts rather than the size of vertebrae. This is in contrast to some other clades where constraint on vertebral numbers has lead to differential somitic growth underpinning major change in body proportions. For example 7 cervical vertebrae are conserved across mammals from short-necked echidnas to long-necked giraffes, with few exceptions (Asher et al., 2011). In mammals the *Hox* genes also perform a function in the proliferation of cell lines in the embryo, this is thought to be the cause of the observed conservation as a *Hox* mutation increases the likelihood of embryonic cancer. In addition where a *Hox* mutation causes a cervical vertebra to have a thoracic rib this can result in reduced functionality of the shoulder region and nerve problems. In mammalian species where a deviation from seven cervicals is observed, it is accompanied by pleomerism (Asher et al., 2011).

In this section I use two main methods - standardised phylogenetic independent contrasts and *AUTEUR* - to estimate and compare rates in traits associated with the neck and vertebrae of sauropterygians: ratio of neck length to trunk length, number of presacrals, and cervical to presacral count ratio. I used these methods to test three hypotheses as outlined in section 5.1.1.

1. I tested whether Sauropterygian body plan evolved via somitogenetic and homeotic effects rather than by differential somitic growth. This hypothesis would be supported if change in count ratio and change in length ratio were linked across the tree (rather than relative lengths changing independently of count ratio).

2. I tested whether somitogenetic and homeotic effects were decoupled across Sauropterygia. This hypothesis would be supported if rates of change in overall presacral count and rates of change in the ratio of cervical to dorsal vertebral count were not correlated and if shifts in the rate of presacral count evolution did not correspond to shifts in the rate of count ratio evolution.

3. I tested whether extreme body morphologies were evolved via high rates of pleomerism. This hypothesis would be supported if the taxa with proportionally extremely long necks or long trunks showed higher than average rates of presacral count evolution with normal rates of count ratio evolution.

5.2.2 Methods

To test the first two hypotheses I performed an ordinary least squares regression of standardised phylogenetic independent contrasts (hereafter ‘standardised contrasts’; Felsenstein, 1985) of neck to trunk ratio against standardised contrasts of cervical to dorsal count ratio (Fig. 1.4). Phylogenetic independent contrasts (PIC) is a fundamental method in phylogenetic comparative analysis and is the basic way in which raw trait data can be corrected for the influence of the shared evolutionary history of taxa. The method is outlined in detail in section 1.2 of Chapter 1 of this thesis. In this section standardised PIC (where the contrasts are weighted according to distance from the root of the tree) was performed in R (R Core Team, 2015) using the function `pic` in the package *ape* (Paradis et al., 2004). The results of PIC correspond to nodes of the tree, and each can be considered a point estimate of the evolutionary rate at that node, including a direction of the shift in trait value which occurred.

To test the third hypothesis and as an additional test of the second I used a method via which rates of continuous trait evolution can be measured as a parameter within a model of evolution (Hunt, 2012). These models (Hansen models; Hansen, 1997) can include additional parameters relating to trait mean and the strength of selection through time (Section 5.3) but all have a basic rate parameter σ^2 which is the step variance. In each time-step of the model the magnitude and direction of trait change is randomly drawn from a normal distribution with mean the same as the trait mean and variance σ^2. A large σ^2 allows larger steps in trait change on average, and thus a faster rate of evolution of the trait.

AUTEUR (Accommodating Uncertainty in Trait Evolution Using R) is a method originally presented by Eastman et al. (2011) and now incorporated into the *geiger* package in R (Pennell et al., 2014). It is a Bayesian sampling approach, which estimates the generating process of trait evolution under Brownian motion across a phylogeny, without the user having to specify the positions of rate changes on the tree in advance. By iteratively splitting and merging rates on adjacent branches, and testing AIC (Akaike information criterion, Burnham and Anderson (2002)), it optimises branch rates. The advantage of this method for phylogenies of fossil taxa is that it does not require the tree to be fully sampled, or ultrametric. I assessed convergence of the mcmc chains using the *CODA* package in R (Plummer et al., 2006) and adjusted input parameters until the chains reached stationarity.

5.2.3 Results

5.2.3.1 Hypothesis one

There was a good correlation between length ratio and count ratio (R^2=0.4948, p <0.0001). This indicates that somitogenetic or homeotic effects dominated the evolution of diversity in sauropterygian axial body proportions, rather than differential somitic growth, although some of the residual variation could be explained by this factor. Nothosaurs have the largest average residuals (yellow circles in Fig. 5.6a), suggesting that this clade in particular experienced some differential somitic growth.

If variation in length ratios was controlled mainly by differential somitic growth the contrast plot would show either consistent count ratio contrasts with variation in length ratio contrasts, or no correlation between the two. Figures 5.6b and c show traitgrams - plots of the change in each ratio through time on the phylogeny. If change in axial length ratios was dominated by change in somitic growth rather than somitogenesis and homeotic effects, a different pattern would be seen in each traitgram, with taxon branches going in different directions for the same taxa, or no correspondence at all. In fact the patterns in the traitgrams are broadly similar, with the direction of change matching for taxa, although the contrast in rate of trait change (slope of lines) between Triassic and post-Triassic taxa is larger for length ratios than count ratios (Fig. 5.6b and c).

5.2.3.2 Hypotheses two and three

Continuous body size data are usually log transformed before analysis, on the basis that characters change in size proportionally to body size. Vertebral counts are more complicated because they are a meristic trait and so whether or not the data should be transformed depends on whether change in number of vertebrae is proportional to the number of ver-

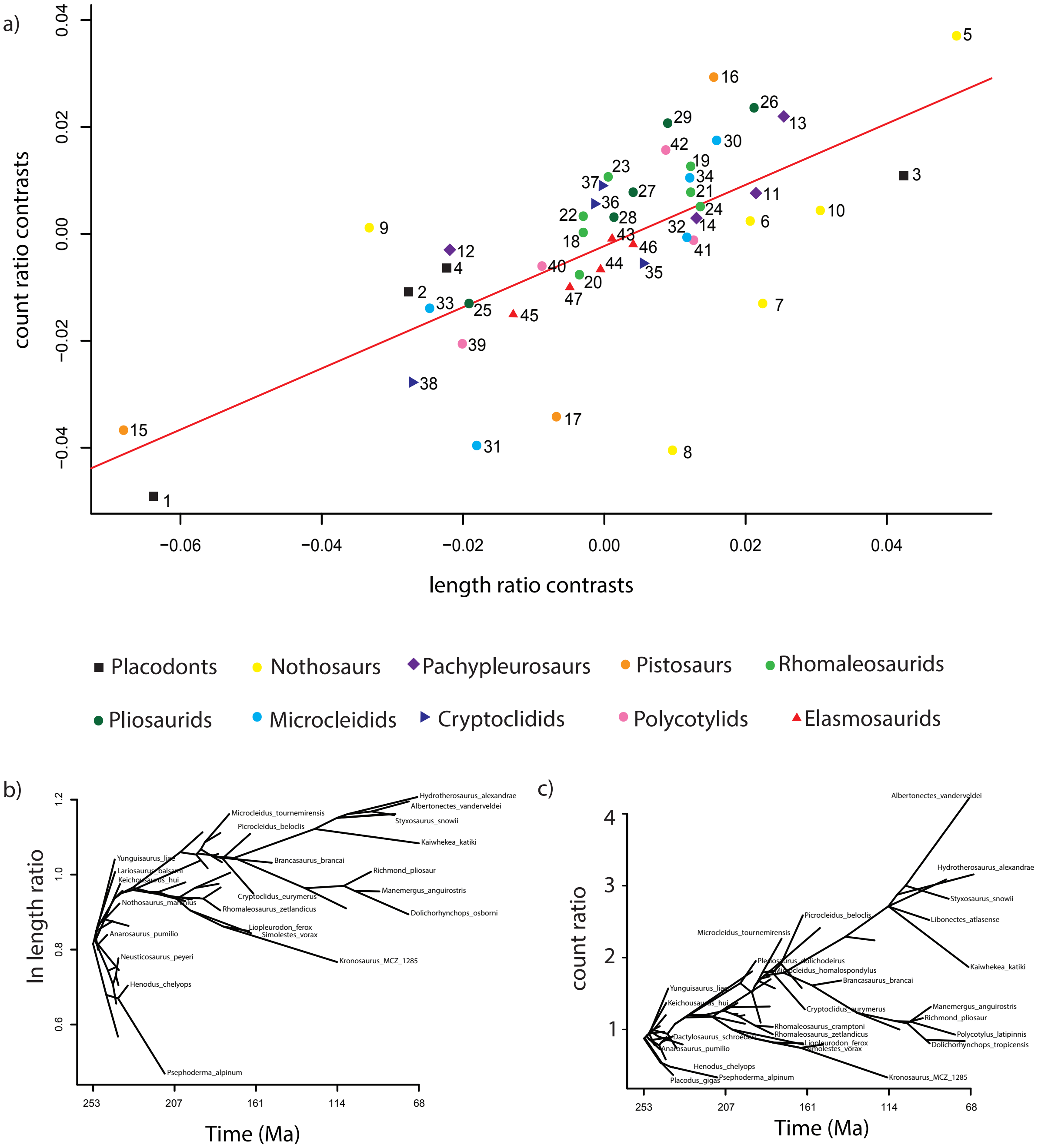


Figure 5.6: Raw data used for these plots was the ratio of neck length to trunk length (length ratio), and the ratio of cervical to dorsal vertebral counts (count ratio). Part a) shows standardised phylogenetic independent contrasts for each ratio and the ordinary least squares regression line of length ratio against count ratio. Points are coloured and shaped according to clade membership. The numbers next to each point indicate the node of the phylogeny that was used for the analysis, 1 is the root node, the lowest node number within each subclade is the ‘root’ node of that group and other nodes are nested. Yellow circles represent nodes nested within *Nothosauria* and this subclade has the largest average residuals (largest distance from the regression line). b) and c) show traitgrams - the phylogeny plotted into a space defined by trait values of length ratio and count ratio on the y axis, and time on the x axis.

tebrae itself. Before conducting the complete rates analysis I performed 20 preliminary analyses on raw presacral count and extracted the terminal branch rates for each taxon. I then performed a least-squares regression of presacral count against corresponding tip rate. There were no significant correlations between these two values with an average R^2 of -0.0056 and p-values all higher than 0.2 across the 20 trees. This indicates that ease of gaining or losing vertebrae does not depend on the number of vertebrae a taxon has. In addition, I plotted traitgrams with node values calculated using standardised contrasts for raw and natural-log transformed counts (Fig. 5.7). These plots showed that the pattern of trait change is robust in raw and log transformed data. I therefore proceeded with raw vertebral counts.

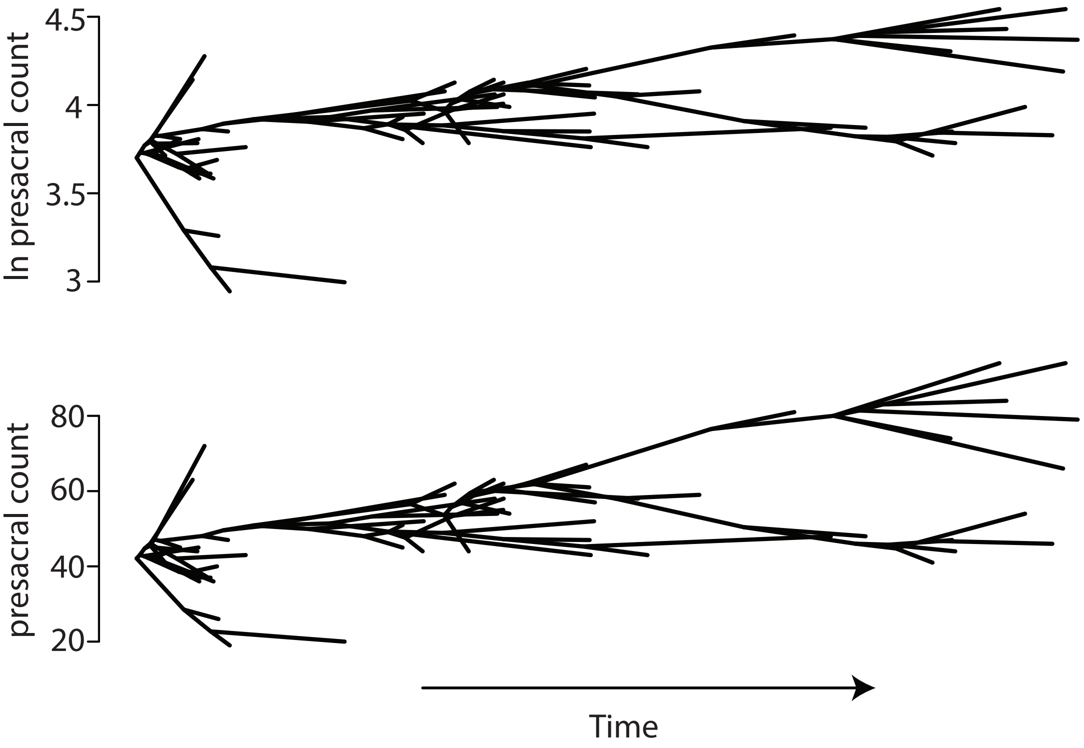


Figure 5.7: Traitgrams (phylogenies plotted in trait space) with node values reconstructed using phylogenetic independent contrasts, of log transformed presacral counts (top) and raw presacral counts (bottom).

Rates of evolution of vertebral count ratio and presacral count are different in some parts of the tree but similar in others (Fig. 5.8 and Table 5.2). Basal taxa show the largest difference; rates for vertebral count ratio are low at the root of the tree and this is inherited

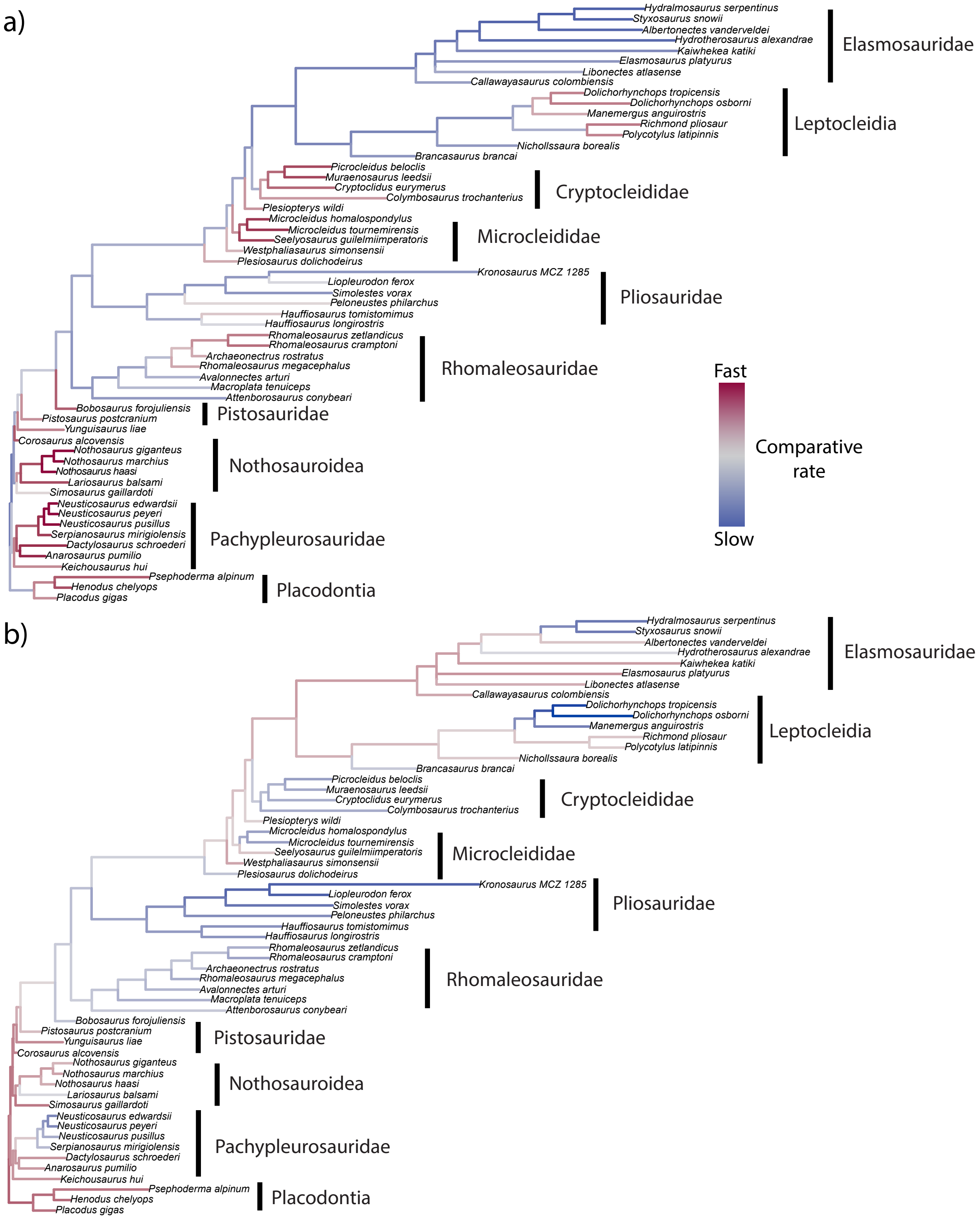


Figure 5.8: Trees representative of results across all trees for estimation of evolutionary rates in vertebral count ratio and presacral count. Colours of the branches correspond to relative rates of phenotypic change scaled to the median. Red corresponds to a high rate and blue to a low. The darker the colour the more extreme the rate value. a) Rates for vertebral count ratio. b) Rates for presacral count.

in internal branches down the whole tree with various nested clades showing high rates, especially the basal groups (Fig. 5.8). Presacral count rates show almost the opposite pattern with high rates at the root followed by a slowing down of rates in individual descendant clades, although the internal branches also show a period of slightly decreased rate at the base of Plesiosauria. Rates in both traits are low in pliosaurids and rhomaleosaurids, although more derived rhomaleosaurids show increasing rates of ratio change. In microcleidids and cryptoclidids vertebral count ratio rates are high, whereas presacral count rates are around average. Within elasmosaurids evolutionary rates in ratio are very low. Presacral count rates in elasmosaurids are high but slow down in the most highly nested taxa.

Table 5.2: Differential rates of vertebral evolution for subclades

Clade	Ratio Rate	Count Rate
Placodonts	High	High
Pachypleurosaurs	High	Mixed/low
Nothosaurs	High	Low
Pistosaurs	High	High
Rhomaleosaurids	Low, High in more nested taxa	Low
Pliosaurids	Low, lower in more nested taxa	Low, lower in more nested taxa
Microcleidids	High	Mixed
Cryptoclidids	High	Low
Leptocleidians	Low, high in more nested taxa	High, low in more nested taxa
Elasmosaurids	Low	High, low in more nested taxa

A summary of the overall pattern of high and low rates for each subclade within Sauropterygia, comparing the rate of change of vertebral count ratio to the rate of change in presacral count.

The pattern of opposite relative rates in vertebral count ratio and presacral count in many parts of the tree indicates that homeotic and somitogenetic effects can be decoupled, although there are some parts of the tree where this may not be the case (e.g. the most highly nested pliosaurids and elasmosaurids). Following Müller et al. (2010) I used phylogenetic generalised least squares regression of presacral counts against the ratio of cervical to presacral counts and found a weak but significant correlation (R^2=0.2250, $p<0.0001$, red line in 5.9), indicating that the two measurements are coupled across the tree to some degree. Standardised contrasts can be considered to be a point estimate of rate at a node (Freckleton and Harvey, 2006) and indicate where large or small shifts in rates occur.

When the contrasts are plotted according to clade membership it can be seen that the nodes that are not nested within clades, but are on the ‘spine’ of the tree (Fig. 5.9; brown points) show a greater correspondence between the two contrasts. To investigate this further I performed regressions on the nested and un-nested node contrast sets separately. The un-nested nodes showed a strong significant correlation (R^2=0.7838, p <0.0001, blue line in figure 5.9) whereas for the remaining nodes that fall within a clade there was no correlation (R^2=-0.0256, p=0.869, green line in figure 5.9). Along the spine of the tree rates of ratio change remain relatively low and rates of count change remain relatively high, but both are stable. Where they deviate from this the change tends to be in the same direction in both, for example the shift to higher rates at node 16 between *Yunguisaurus liae* and plesiosaurians (brown point labelled 16 in top right part of graph in Fig. 5.9). This suggests that vertebral count ratio and presacral count are not completely decoupled across Sauropterygia, in contrast with the results of Müller et al. (2010), which found they were decoupled across the whole tree of amniotes and for all subclades.

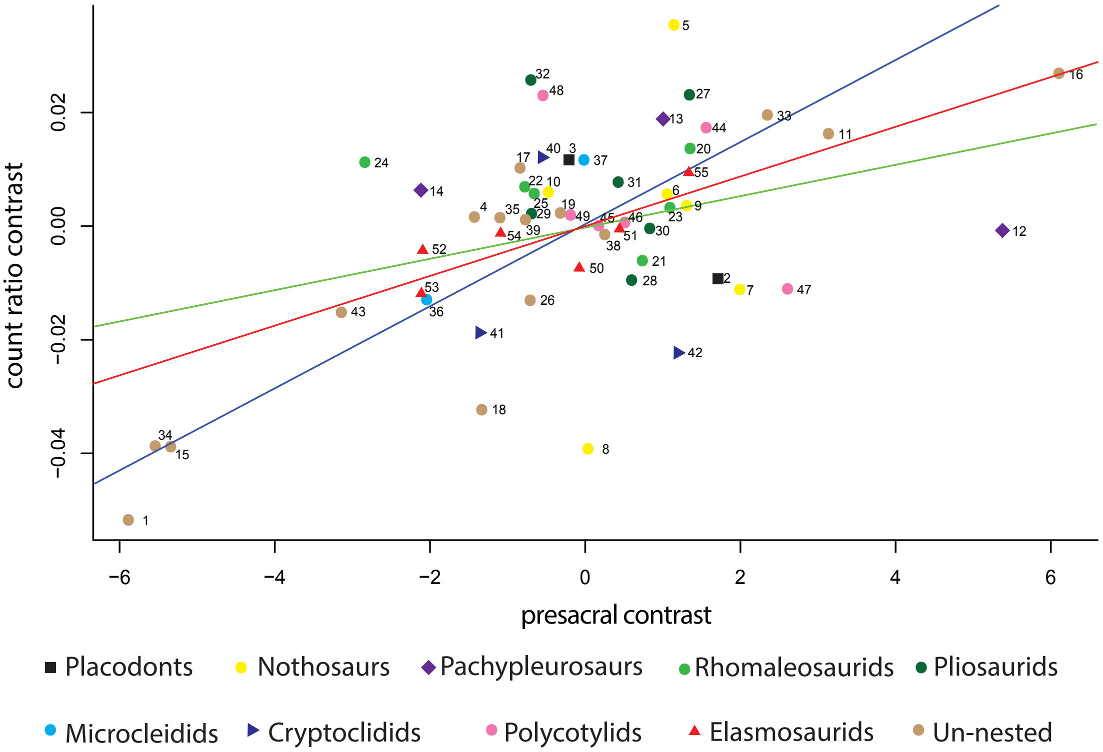


Figure 5.9: Standardised phylogenetic independent contrasts of vertebral count ratio against presacral count. Colours correspond to clade membership. Red line is the regression model for all the data, the green line is for only taxa nested within clades and the blue line is for un-nested taxa.

5.2.4 Discussion

Rates of change in vertebral count ratio and presacral count were decoupled in some parts of the tree, but not others. The root of the tree shows high branch rates for presacral count but low rates in vertebral count ratio change. This means that vertebrae were gained or lost rapidly, but the number of cervical vertebrae relative to the number of dorsal vertebrae remained similar. This might be inherited across internal branches of the tree; although presacral rates and count ratio rates slow after the transition to open water, presacral rates pick up again at the base of Plesiosauroidea and remain high and stable all the way through to the elasmosaurids. Other lineages show greater shifts to higher rates in count ratio and lower rates of presacral count change. It is possible that this maintenance of higher rates of presacral count change in the upper part of the tree (and consequently the capacity to change axial body plan rapidly) facilitated the survival of these cryptic internal lineages into the Cretaceous when other sister taxa whose rates had decreased became extinct. However, this would need to be constrained further with tests of plausible ecological hypotheses of a causal link between body plan and survivorship. When there is very little variation in rate (e.g. the nodes along the spine of the tree as outlined above) both count and ratio remain stable and are thus also coupled.

The largest absolute contrast in both presacral count and count ratio is seen near the base of the tree. Placodonts had very short necks compared to their bodies and these results suggest they acquired this body plan by adding dorsal vertebrae; shifting the boundary between cervicals and dorsals to maintain short neck length. This implies that both homeotic and somitogenetic effects were important in this part of the tree. Pistosaurs show a high rate of change in both vertebral count ratios and presacral count on branches, indicating that when there are large differences in count over a short phylogenetic distance both somitogenetic and homeotic effects are involved in order to generate this large change and are therefore coupled. In other basal taxa there is a shift from low root and internal branch rates to very high rates in vertebral count ratio for nothosaurs and pachypleurosaurs, and a slowing of rates in presacral count. This indicates their body plan evolved via change in regionalisation.

Pliosaurids underwent a decrease in neck length across the clade. The rate of presacral count change is low, and rate of change in vertebral count ratio becomes increasingly slow in more derived taxa. This is suggestive of slow shortening of the neck via homeotic effects and the decrease from an already low ancestral rate possibly represents a bound caused by a functional limit to how short the neck could become.

Cryptoclidids and microcleidids each independently derived an increase in neck length. In combination with the fast rates in vertebral count ratio change and lower rates of change in the number of presacrals this suggests that homeotic effects were the driving mechanism of relative elongation of the neck in these two groups.

Polycotylid leptocleidians derived a decrease in neck length. In combination with a higher rate of vertebral count ratio change and low rate of presacral change this implies that homeotic effects dominated their change in axial body plan and that cervical vertebrae switched to dorsal vertebrae through shifting regionalisation. A similar pattern is seen in the more highly nested rhomaleosaurids.

Elasmosaurids derived an increased relative neck length. This was accompanied by high rates of change in presacral count and low rates in change of ratio, over long branches on the tree. This suggests that pleomerism generated the high presacral counts (i.e. the basic material required for a long neck) but that slow change in regionalisation over a long period of time lead to the eventual extreme vertebral count ratios. Thus both somitogenetic and homeotic effect were important in elasmosaurids. There is a decrease in presacral evolutionary rate in the most derived elasmosaurids. These are the taxa with some of the highest overall number of presacrals; there are other taxa which had slightly higher presacral counts that still show high rates (e.g. *Elasmosaurus platyurus*), but they have shorter necks in terms of absolute length. It is possible therefore that this slowing represents arrival at an upper bound in the trend of increased total body length based on functional

Table 5.3: Developmental processes controlling evolution of body plan in subclades

Clade	Dominant effect
Spine taxa	Both
Placodontia	Both
Pachypleurosauridae	Homeotic
Nothosauroidea	Homeotic
Pistosauridae	Both
Rhomaleosauridae	Homeotic
Pliosauridae	Homeotic
Microclididae	Homeotic
Cryptocleididae	Homeotic
Leptocleidia	Hometotic
Elasmosauridae	Somitogenetic

This shows the dominant developmental process that was changing within each subclade to produce the resulting changes in body plan. Spine taxa indicates nodes which do not lead directly to tips and are not nested within a subclade.

constraint, but this would have to be tested further.

Overall these results indicate that for more generalised body morphologies, changes in axial vertebral proportions were achieved by homeotic effects, but in the parts of the tree that show the greatest difference in neck compared to the body length (either a very short neck or a very long neck; placodonts [e.g. 7 cervicals to 19 dorsals in *Placodus gigas*] and elasmosaurids [e.g. 60 cervicals and 19 dorsals in *Hydrotherosaurus alexandrae*]), somitogenetic effects also played a part or dominated.

The results are consistent with the hypothesis that change in relative axial body lengths occurred mainly through embryonic processes, rather than post-embryonic differential somite growth. Somitogenetic and homeotic effects were coupled on nodes along the spine of the tree located between branches that showed stable rates over time. Where change in body proportions did occur it involved changes in regionalisation, but for extreme body morphologies somitogenetic effects also contributed to the pattern. This suggests that the generation of body plans corresponding to higher level groups was generated by coupled somitogenesis and homeotic effects. Modification from these basic body plans within clades was mainly achieved through changes in regionalisation, and this was decoupled from somitogenesis. High rates of shifting regionalisation could occur without a simultaneous high rate of change in vertebral count, but not vice versa. One possible explanation for this could be that changes in *Hox* patterning have a subsequent effect on the somitogenetic clock, as regions are established in the pre-somitic mesoderm before somitogenesis begins, but the details of these biological mechanisms have not yet been fully uncovered.

5.3 Modelling macroevolution across the Triassic-Jurassic boundary

5.3.1 Background

A key hypothesis that has been proposed is that sauropterygians experienced a change in evolutionary constraint at the Triassic-Jurassic transition. This is based upon an expansion of range of limb proportions, body size and axial proportions observed in Plesiosauria, compared to Triassic taxa, and through time in the Jurassic (Benson et al., 2012). One possible explanation for the observed increase in body plan disparity in Plesiosauria as compared to basal taxa is that basal sauropterygians were under environmental functional constraint imposed by terrestrial ability or retention of the features associated with it, and that following the transition to open water this constraint was released. An alternative explanation is that developmental constraints within the clade prevented rapid rates of evolution, until this was overcome in the lineage leading to Plesiosauria, which then radiated into the many niches available. In this scenario either the radiation was facilitated by high rates of change in various aspects of body plan, or radiation into new niches facilitated high rates.

Previous workers have shown that there were correlated trends in plesiosaurian body size and locomotion, as well as a correlation between limb proportions and axial body length ratios (O'Keefe and Carrano, 2005). The main trend shown was towards large body size in tandem with increased constraint (i.e. a restriction on the variety of relative proportions) on the ratio of segment lengths within each limb. These correlations resulted in the 'plesiosauromorph' and 'pliosauromorph' morphotypes seen in the Plesiosauria (Table 5.1), as well as the original confusion in taxonomic classifications based on gross body morphology (Andrews, 1910b, 1913; Persson, 1963). In addition it has been suggested that limb proportions in basal sauropterygians were constrained by the requirements of terrestrial locomotion imposed by a semi-aquatic habit, which were features such as jointedness and relatively equal segment lengths that allowed for body mass support (Storrs, 1993).

Here I test whether change in rate of limb proportions (again using AUTEUR), represented by ratios of the humerus to the radius and humerus to the femur, is consistent with the hypothesis that the evolutionary transition to open water during the end of the Triassic and beginning of the Jurassic represented a change in constraint, facilitating evolution of new limb configurations. I then use evolutionary model fitting to test the hypothesis that there was a release in constraint during the Tr/Jr boundary and transition to open water, and attempt to identify whether patterns are consistent with functional, developmental, or a combination of both types of constraint.

An additional claim that has been made in relation to sauropterygian (and plesiosaurian specifically) evolution is that the pliosauromorph body plan represents an adaptive peak in the evolutionary landscape, towards which different lineages have repeatedly evolved (O'Keefe and Carrano, 2005). This body plan appears separately in three lineages: polycotylids within leptocleidians, rhomaleosaurids and pliosaurids. The key features of a pliosauromorph body plan, implied as the optima for this adaptive peak, are outlined in Table 5.1. Body plans of basal taxa can also be defined by these characteristics and it may be interesting to investigate how basal body plans relate to more derived body plans based upon ratios.

1. I tested whether there was a change in evolutionary regime for sauropterygians at the Triassic-Jurassic boundary. This hypothesis would be supported if rates of evolution showed a difference either side of the boundary, and if Hansen models that incorporate a regime change at the boundary are a better fit to the data than single regime models.

2. I tested whether the pliosauromorph body plan represented an adaptive peak throughout Sauropterygian evolutionary history. This hypothesis would be supported if results from the adaptive peak analysis showed that a model where taxa with this body plan were convergent was the best fit to the data.

5.3.2 Methods

The first method I used to address this second set of hypotheses was the same as one used in the previous section - *AUTEUR* - to estimate evolutionary rates of continuous traits along branches of a phylogeny, as a parameter in a Hansen model. In addition to this I used three other methods outlined below. The set of phylogenies used for this analysis was the same as the set of 100 trees used for the analyses in the previous section, except for when a particular measurement could not be made for a taxon, in which case this taxon was dropped from the phylogeny and the analysis.

5.3.2.1 Phylomorphospace analysis

I plotted tip and node values through time on a phylogeny, to get an overview of change in phylomorphospace occupation through time. To do this, initially I performed phylogenetic size correction on neck, cranium, humerus and radius lengths using `phy.resid` in R package *phytools* (Revell, 2012). This requires calculating residuals of phylogenetic generalised least squares regression of each variable against trunk length, as a proxy for body size, these residuals are then used in subsequent analyses (Revell, 2009). I then estimated ancestral states in neck-skull and humerus-radius traitspace under Brownian motion using a maximum likelihood estimator. The relationships between these relative lengths are a major defining characteristic of plesiosaurian taxa (O'Keefe, 2002) and also varied within basal groups, so can provide a helpful indication of body plan variety.

5.3.2.2 Regime change modelling

Following the phylomorphospace analysis I performed a more rigorous analysis to establish whether change in phenotypic traits of sauropterygians through time was consistent with a change in evolutionary regime at the Triassic/Jurassic boundary. To do this I fit seven different Hansen models to various individual traits and then compared Akaike weights (calculated using the Akaike information criterion - AIC) to find the model with the best fit to the data. A Hansen model is a model of evolutionary change with parameters σ^2 (step variance), θ (the optimum trait value) and α (strength of selective constraint) (Hansen, 1997; Butler and King, 2004). Brownian motion is a special case of a Hansen model where α is zero. Akaike weights are a method of model selection, where better fitting models have a greater weight, but additional parameters within the model incur a penalty (Burnham and Anderson, 2002). Akaike weights for a set of models fit to the same data sum to unit and support for each model can therefore be expressed as a percentage and considered an estimate of the probability of that model being the best model for the data, out of the range of models tested (Burnham and Anderson, 2002; Bapst, 2014a).

I chose to fit models to univariate data so that patterns could be compared between traits and potentially be brought to bear on the question of whether functional or developmental constraint was more important. The four traits I modelled were the ratio of neck length to trunk length, the ratio of the cervical to presacral count, the ratio of the humerus to the femur and the ratio of the humerus to the radius. The first two traits contribute to the question of how axial body plan of Sauropterygians evolved through the Mesozoic and whether there was a change in constraint on the axis. The second two traits contribute to the question of whether functional constraint was imposed by the shallow water habit of Triassic sauropterygians, and then released when plesiosaurians made the evolutionary transition to an open water habit after the start of the Jurassic.

I fit seven models to the data, these were 1) single rate Brownian motion (BM1), 2) Brownian motion with one shift in rate (BMS), 3) a single rate Ornstein-Uhlenbeck model

(OU1), 4) OU with one shift in trait mean (OUM) 5) OU with one shift in trait mean and rate (OUMV) 6) OU with one shift in trait mean and strength of selection (OUMA) and 7) OU with one shift in trait mean, rate and strength of selection (OUMVA) (Beaulieu et al., 2012). These models were chosen because when compared they can be used to identify whether a one or two regime model is a better fit to the data, and which parameter (step variance, trait optima or strength of selective constraint) was most important in explaining the shift in regime, if there was one.

This analysis, unlike the previous rates analysis, requires *a priori* specification of evolutionary regimes of the tips. This means it can be used to test a specific hypothesis of a regime change at a particular point in the phylogeny - in this case the Jurassic-Triassic boundary and the division between plesiosaurians and all other sauropterygian taxa. If any of the two regime models (models 2, 4, 5, 6 and 7) have the best support under AIC for a trait, this would indicate that a model with a shift in evolutionary regime at the boundary is better supported by the data than a model where there is no shift in evolutionary regime at the boundary (models 1 and 3). This would therefore support the hypothesis of a shift in regime at the Triassic-Jurassic boundary.

5.3.2.3 Adaptive peak modelling

To test whether the appearance of pliosauromorph body types is consistent with an adaptive peak model, rather than passive exploration of available trait space at different points in time, I used an iterative Hansen modelling process available in the package *surface* in R (Ingram and Mahler, 2013). The purpose of this function is to identify branches on the phylogeny that share trait optima. There is no prior definition of expected lineages sharing the same optimum; it can therefore be used to find lineages in different parts of a tree which converge. The modelling process makes a forward and backward pass; during the forward pass it iteratively tests fit of different OU models to the data by merging and splitting branches' trait optima and comparing AICs. Once it has converged on a low AIC solution it moves back down the tree collapsing branches with one trait optimum to share the same as another and maintaining the change if it produces a lower AIC. Any taxa that have converged upon the same adaptive peak would show as the same branch colour/optimum after the final iteration, even if they are in different parts of the phylogeny. In this way the function can detect evolutionary convergence of phenotype. If the pliosauromorph body type does indeed represent an adaptive peak, taxa with this body type should share trait optima.

I conducted the analysis on features of body plan that are classically associated with the difference between pliosauromophs and plesiosauromorphs; neck to body length ratio, skull to body length ratio, humerus to femur ratio, humerus to radius ratio and ischium to trunk length ratio. I tested each trait individually and then combined them and tested the multivariate dataset.

5.3.3 Results and discussion

5.3.3.1 Hypothesis one

For the neck-skull phylomorphospace basal taxa occupy a slightly smaller area in comparison with their younger relatives, but plesiosaurian taxa occupy a different part of the space suggesting that they were able to explore areas unavailable to earlier taxa. This is consistent with either a pattern of release in constraint or a shift to a new trait optimum, or both (Fig. 5.10).

In humerus-radius phylomorphospace there is a pattern of restricted ratio between the two traits in basal taxa, with points occupying quite a large area but along a narrow region of traitspace. Later taxa diverge into this previously unoccupied space, suggesting a release in constraint on these traits relative to one another.

The ratio of the humerus to the radius shows a marked change in rates between basal clades and plesiosaurians (Fig. 5.11a). Rates are consistently low at the base of the tree, in all non-plesiosaurian taxa, implying that the ancestral ratios were constrained; this is

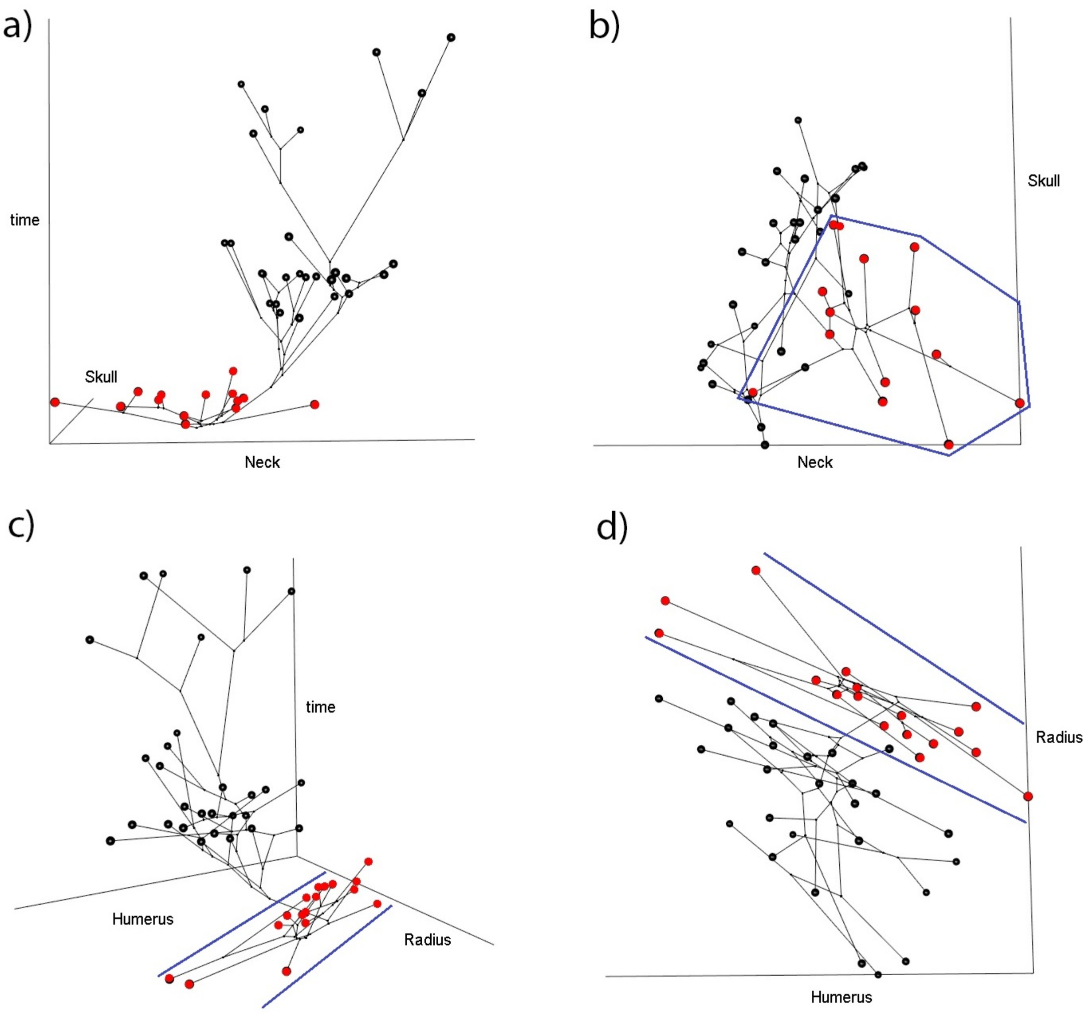


Figure 5.10: Different views of a) and b) phylomorphospace plot of phylogenetically size corrected neck length and skull length on a phylogeny through time and c) and d) phylomorphospace plot of phylogenetically size corrected humerus and radius length on a phylogeny through time. Each plot has three axes - two traits and time, where one is not labelled this axis is running directly into the page. Triassic taxa are the red points and Jurassic and Cretaceous taxa are shown in black. The blue lines highlight the general are where taxa from the Triassic lie in traitspace in plots b) c) and d).

consistent with the idea that strong functional constraint on limb proportions was present for these taxa.

The transition to open water occurred on the phylogeny between *Augustasaurus hagdorni* and the plesiosaurian stem. This is the point in the tree where there is an initial shift to fast rates of evolution in limb proportions. Although this decreases again in early plesiosauroids, including Microcleididae, it is maintained across the rest of the clade. This implies that the functional constraint that basal sauropterygians were under was released

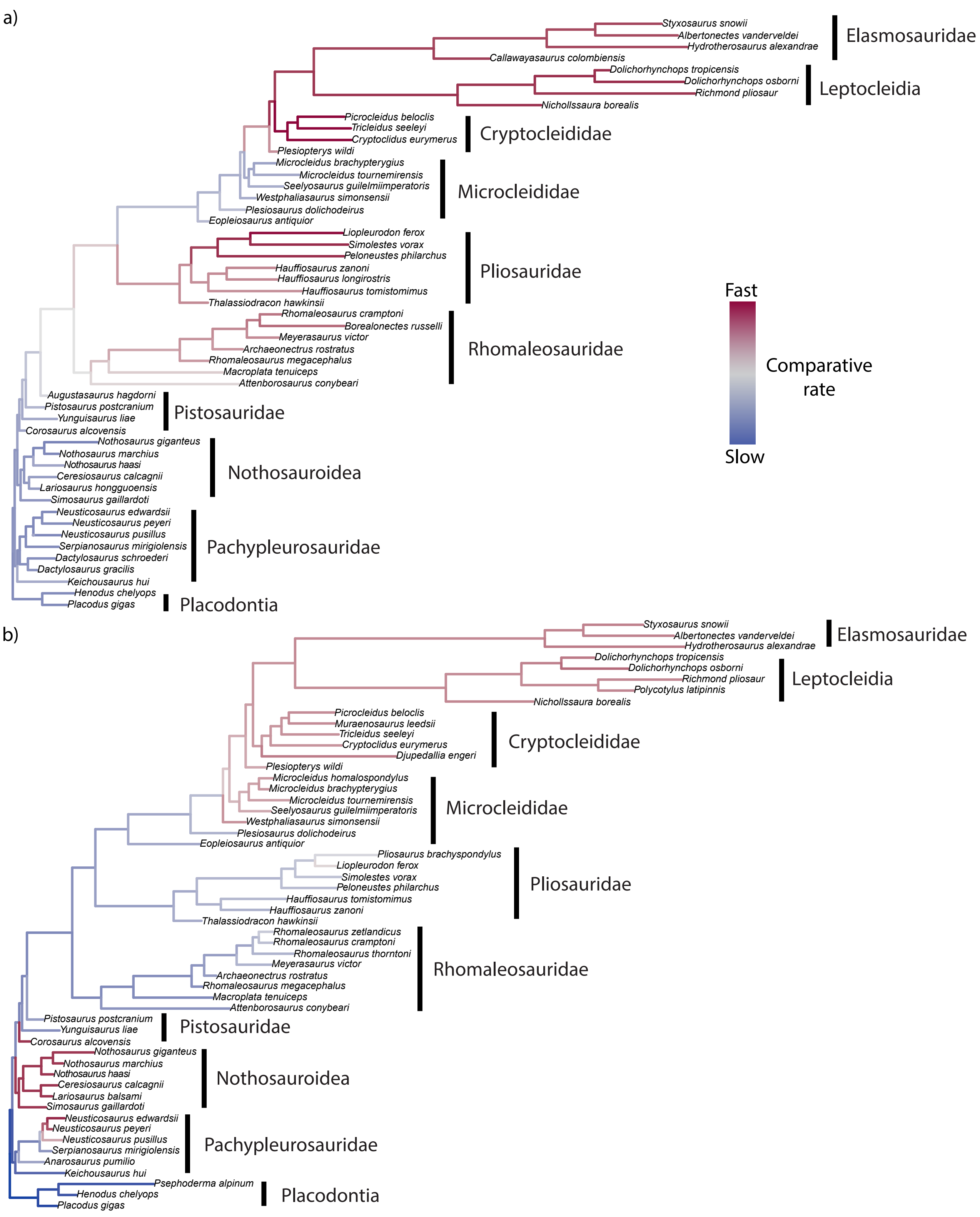


Figure 5.11: Representative trees showing rates of evolution of limb ratios. Colours represent the same as in figure 5.8. a) Rate of change in ratio of humerus to radius. b) Rate of change in ratio of humerus to femur.

and that plesiosaurian taxa evolved different limb configuration strategies corresponding to different locomotive styles (Caldwell, 2002).

The rates of the ratio of humerus to femur show a slightly different pattern. Within basal taxa the nothosaurs show very high rates. This could indicate that nothosaurs had higher evolutionary plasticity in locomotory style, however, the functional aspects of their limb configurations have yet to be investigated so this cannot be directly tested. A further difference is that the rate remains low in the Jurassic within rhomaleosaur and pliosaur taxa before shifting to higher rates in plesiosauroidea. Equal humerus and femur lengths imply that both limbs are used for locomotion (either on land or for propulsion in the water). Some plesiosauroids show an increase in the length of the hindlimb relative to the forelimb and others a decrease. These patterns are associated with the evolution of the two morphs, with some leptocleidians gaining longer hindlimbs and a pliosauromorph body plan and elasmosaurids showing reduction in hindlimbs (O'Keefe and Carrano, 2005). The fast rates seen in these analyses may have been driven by these two contrasting body plan trends, but further tests are required to confirm this and can be found in the next section (section 5.3.3.2).

Overall the results in Figures 5.8, 5.10 and 5.11 give a good indication that there was a release in constraint at the Tr/Jr boundary. Figure 5.11a in particular shows results that indicate release of constraint as there is a clear increase in average rate between basal and plesiosaurian taxa in the ratio of the humerus to the radius. Other traits do not show as distinct a rate increase but do show changes in the total area or position of morphospace that taxa occupy.

Moving on to the results of fitting Hansen models to the data. Cervical to presacral count has the best model support for model 6 (OUMA, Figure 5.12), this means that the average value of the trait mean and the strength of selection shifted between the two predefined regimes. In this case the trait mean increased and the strength of selection decreased (Table 5.4).

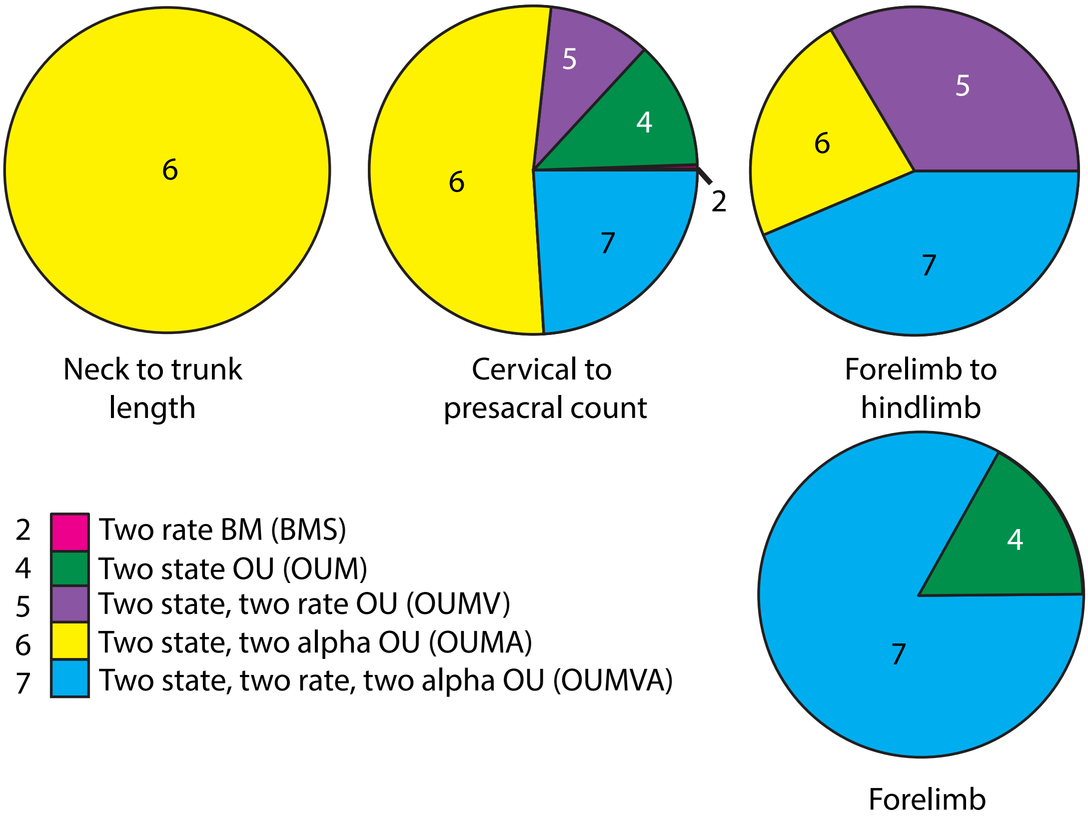


Figure 5.12: Pie charts showing average percentage model support for seven different Hansen models across a set of 100 phylogenies of Sauropterygia. Models included a shift in parameters (a regime change) at the Triassic-Jurassic boundary. Model support was calculated using the Akaike information criterion. Details of each model (identified in the text by the number to the left of the key) and Akaike weights are outlined in section 5.3.2.2. Some of the seven models fit to each dataset did not receive any support (the single regime models) and are therefore not included in the pie charts or the key. Acronyms in the key correspond to which model parameters were allowed to change at the regime shift boundary. BM=Brownian motion, OU=Ornstein-Uhlenbeck. For the BM model S=trait mean. For the OU models M=trait mean, V=rate, A=strength of selection.

For the length of the neck compared to the trunk model 6 (OUMA) is by far the best supported model (Fig. 5.12) as none of the other models were supported at all. The change in parameters associated with this shift were a decrease in α from an already low value to close to zero (i.e. almost no selection), and a large decrease in the trait mean. This is consistent with availability and exploitation of new, previously restricted areas of morphospace becoming available.

For humerus to radius ratio model 7 had the best support and 4 had some support. This strongly suggests that there was a change in the strength of selection acting on this trait.

Table 5.4: Results of the Hansen model analysis

Trait	Best model	Mean	Sigma	Alpha
Length ratio	OUMA	Decrease	NA	Decrease
Count ratio	OUMA	Increase	NA	Decrease
Forelimb ratio	OUMVA	Decrease	Decrease	Decrease
Forelimb to hindlimb	OUMVA	Slight increase	Slight decrease	Similar

This shows which model was the best supported across the 100 trees for each of the different traits, as well as what direction of shift in each of the model parameters the best model represented.

Average α values across trees decrease from 0.18 to 0.15 and there was a shift to very low values for σ^2.

Humerus to femur ratio had approximately equal support from models 5,6 and 7, which corresponded to a very small decrease in α and σ^2 and an increase in trait mean, when each of these was allowed to vary. Model 7 (OUMVA) was a bit better supported and in this model although α changed between regimes, on average over all the trees it stayed approximately the same but it had very high values for both regimes for this trait at 1.38 and 1.37 before and after the Triassic-Jurassic boundary.

This good support for two regime models (none of the single regime models had any support for and of the traits) and a consistent reduction in α (selective strength) across these traits strongly supports the hypothesis that there was a release in constraint on sauropterygian morphology after the Tr/Jr boundary. However, what type of constraint this corresponded to is more difficult to establish. Fast rates and some variety in axial body proportions (e.g. short neck and long body with many vertebrae in *Placodus*) in basal taxa demonstrate that earlier clades had the capacity to evolve disparate axial body proportions, and did so under low constraint on this feature. Limb ratios, particularly the humerus to femur ratio were highly constrained before the transition to open water, and although α decreased after this transition it remained high when the trait mean shifted. Overall this evidence is consistent with high functional constraint on humerus to femur values but for different locomotory styles before and after the boundary, and a release in environmental constraint in humerus to radius ratios. Axial body proportion results indicate low selection on this trait through time, but potentially a small release in constraint associated with the transition to open water.

In summary, my results provide good evidence in support of hypothesis one, that there was a release in constraint on Sauropterygian body proportions at the Tr/Jr boundary, but what the nature of this constraint was is as yet unclear.

5.3.3.2 Hypothesis two

In Figure 5.10 it is possible to see that although many taxa do overlap in neck-head traitspace through time, there does not at first pass appear to be evolution directed towards a peak, but rather that various clades expanded into the trait space made available after the Triassic/Jurassic transition. For the main adaptive peak analysis using *surface* none of the analyses on single traits strongly supported the hypothesis that pliosauromorph taxa were converging upon the same trait optimum. In most of the analyses the ancestral trait optimum was conserved across the tree, or shifted early in the clade's history and was then maintained (Fig. 5.13). There are a few examples of convergence between basal taxa within nothosaurs and pachypleurosaurs but none within plesiosaurian clades for univariate traits (Fig. 5.13, femur, skull and ischium).

In the analysis for the ratio of neck length to body length the pliosauromorph configuration is the ancestral optimum (red) and both microcleidids and elasmosaurids shift away from this optimum (blue and purple respectively). Skull to body length ratio and humerus to radius ratio both have a consistent trait optimum across most of the tree with a few deviations. Ischium to trunk ratio showed almost no change across the tree indicating that there may not have been enough data to parametrise the model effectively, particularly as there are large differences in ischium to trunk ratios between sauropterygian subclades (O'Keefe and Carrano, 2005).

When all five traits were modelled simultaneously in a multivariate analysis it resulted in a convergence in trait optimum between *D. osborni* and the derived pliosaurids (red). Although this is consistent with a pliosauromorph adaptive peak, this tree only contained

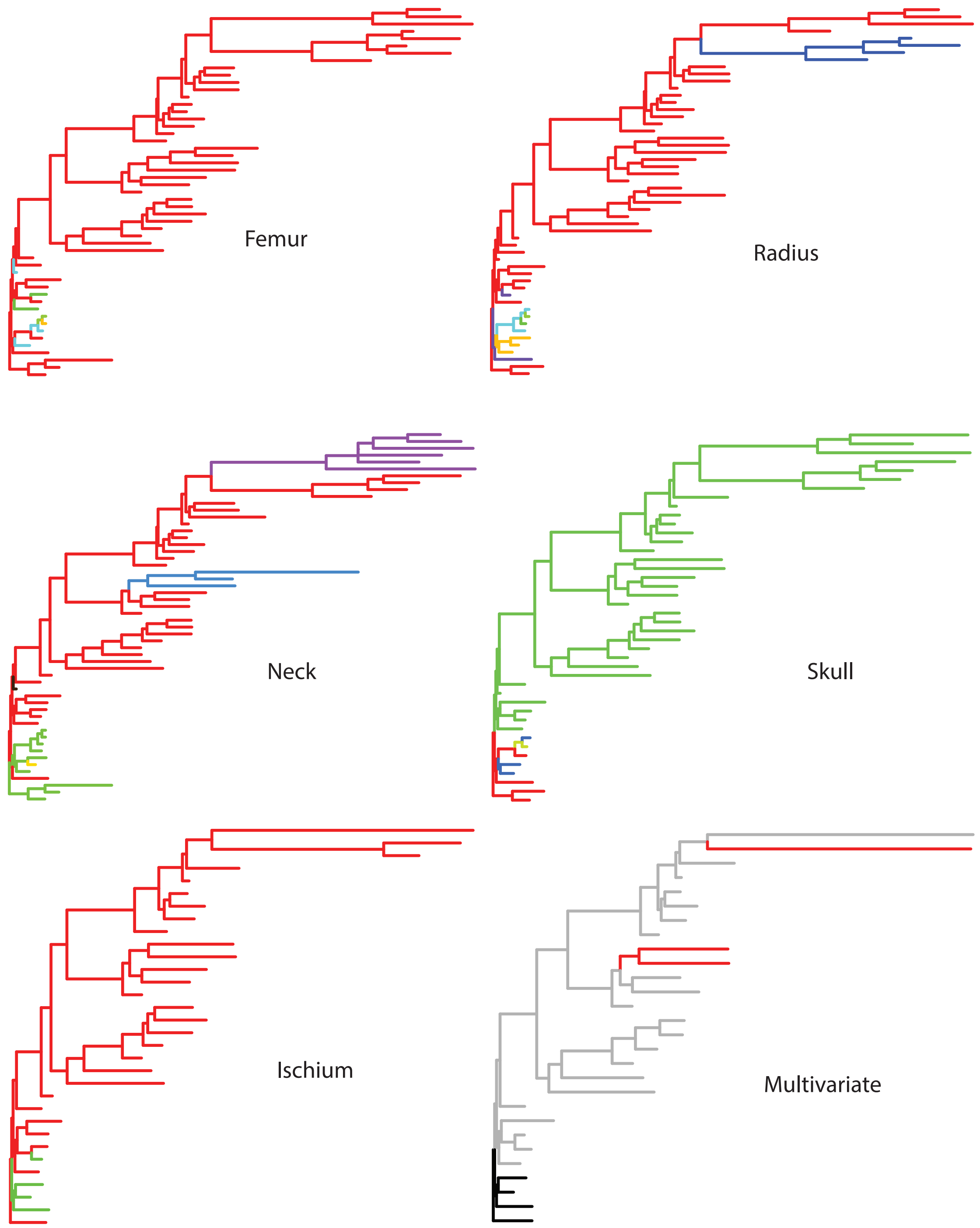


Figure 5.13: Results of the *surface* analysis which uses maximum likelihood to identify convergence. The phylogenies have branches coloured according to matching trait optima. There are five trees corresponding to five univariate traits, indicated next to each tree, and a multivariate analysis including all five of these traits. The colours are arbitrary, but branches that have the same colour on one tree have been inferred to be convergent for the trait(s) analysed for that tree. Where two lineages on the same tree have the same colour this indicates they converged upon the same trait optimum. Colours do not correspond between different trees, they only distinguish between optima on single trees. Each terminal branch represents one taxon but taxon names have been removed for clarity, important features of the plots are explained in the text.

one Polycotylid taxon and so although this provides evidence for convergence between this taxon and the pliosaurids, it doesn't provide evidence for a shared adaptive peak between polycotylids and pliosaurids generally. Repeated runs did produce the same results, but with few taxa the model has low power to detect convergence and results should therefore be interpreted cautiously. If the pliosauromorph body plan really was an adaptive peak we would expect to see convergence upon a set of trait optima in all taxa which are considered to be exemplars of this morph, which would include some rhomaleosaurids.

Overall these results are consistent with a scenario where the pliosauromorph body form does not represent an adaptive peak towards which there was a driven trend of evolution. It may represent a region of traitspace that was explored repeatedly by different taxa, and evolution may have been restricted to certain areas of traitspace due to the functional problems of others (for example a large head on a long neck seems biologically implausible). This would be in agreement with the pattens seen in analyses in section 5.3.

Given the vast timescales over which these morphs emerge it might be possible that the deviation between groups that reach the pliosauromorph body plan through time could represent shifts in a single adaptive peak in the evolutionary landscape (Hunt and Rabosky, 2014), leading to convergence in general form but different trait optima. However, this is speculative as the lack of overall variation in optima in the *surface* analysis does suggest that there may not have been enough data available in some traits for the algorithm to effectively estimate convergence. These results should therefore not be taken as definite evidence against an adaptive peak model of sauropterygian evolution.

5.4 Summary

These analytical methods represent a powerful set of tools for understanding the pattens and mechanisms of long-term evolutionary change. By making comparisons between traits, ratios of traits, and results of different methods it is possible to build a detailed picture of evolutionary change within a clade.

Results show that changes in axial length proportions were generated by homeotic and somitogenetic effects, rather than by post-embryonic differential growth, although some differential growth may have occurred to generate variation about the mean relationship. In addition they demonstrate that the mechanism of change in body proportions in sauropterygians varied across time. Often it was controlled by homeotic effects but to achieve the most extreme morphologies in the group somitogenetic effects were important, even if these acted on long timescales. Change in total presacral count and change in the ratio of cervical to presacral vertebrae was decoupled across nested clades of the phylogeny with a few key outliers corresponding to high evolutionary rates. Change in axial and limb ratios confirms that there was a shift in evolutionary regime at the Triassic/Jurassic boundary, consistent with a relaxation of functional constraint, supported by a two regime model of evolution. Available evidence on the presence of an adaptive peak corresponding to the plesiosauromorph body plan suggests although there may have been some convergence from different lineages towards similar trait configurations, for the most part this body plan was acquired in multiple lineages by passive exploration of the available trait space, which was functionally restricted, rather than by driven evolution towards an adaptive peak.

Many new sauropterygian taxa are currently being unearthed, being described, or waiting to be described. This presents an exciting opportunity to greatly expand the size of the dataset, particularly of Triassic taxa. This will be useful in many respects, not only will it lead to more accurate timescaling of the phylogeny, but it will enable greater statistical power in analyses, some of which are currently in the region of the lower limit of sample size for a valid analysis.

Chapter 6

Conclusions

The potential to use palaeontological data to understand macroevolutionary patterns is rapidly expanding with the development of new methods and the refinement of old. Palaeobiologists and neontologists are making genuine steps towards unified approaches that capitalise on both the vast record of the history of life available from fossils, and the extensive detail available from examining modern taxa (Slater and Harmon, 2013; Pennell and Harmon, 2013; Gavryushkina et al., 2014). However, these unified approaches are not yet mature, and there are many clades of fossil taxa to which they cannot currently be applied. These clades still hold interesting information about macroevolution and extinction on long time scales. In light of this it is important to assess the ways in which such data can be quantitatively analysed in order to make meaningful inferences (Bapst, 2014a). Phylogeny is a framework under which these inferences might be made, but is not without its own challenges (Foote, 1996b; Symonds, 2002). In this thesis I have explored some of the possibilities of analytical interrogation of palaeontological data in a phylogenetic context, beginning with only taxonomic classification, continuing with cladistically inferred phylogenies and finishing with phylogenies and continuous trait data. Here I outline and discuss my conclusions.

6.1 Taxonomy in phylogenetic comparative methods

Methods that incorporate the influence of shared ancestry, via a phylogenetic framework, in an analysis of taxon data are now regularly used by palaeobiologists. There is a wide variety in the perceived ‘quality’ of such phylogenies (for example good resolution and strongly supported relationships are considered preferable), and there is a significant body of research on the effects of defects such as polytomies or missing taxa on the results of PCMs (e.g. Garland and Diaz-Uriarte, 1999; Housworth and Martins, 2001; Stone, 2011; Davies et al., 2012). Cladistic inference methods such as parsimony or Bayesian estimation can sometimes prove difficult with palaeontological data (for example in very speciose clades with few distinct characters). It would therefore be desirable to be able to use the taxonomy of such clades as a substitute for a cladistically inferred phylogeny and incorporate it into PCMs (Bapst, 2014b).

In the second chapter of this thesis I demonstrate that taxonomic information can indeed provide a useful substitute for cladistically inferred phylogenies as a framework for phylogenetic comparative methods (PCMs), in certain circumstances. This conclusion validates previous studies where taxonomy was used in this way, and provides an opportunity for the rich macroevolutionary information contained within some fossil clades to be accessed. The taxonomy being used should be well studied, relatively stable, and derived from a well sampled record. These conditions apply to phylogenetic frameworks generally, and taxonomy should be held to the same rigorous standards as the phylogenies that have previously been considered suitable. The most important thing to note is that stratigraphic information (the comparative ages of taxa) provides an additional control on the tree structure. This often reduces the influence of problematic features such as poor resolution, which may be present in any phylogenetic hypothesis, but can be particularly prevalent in hypotheses based on taxonomy.

My results showed that large clades produced more consistent outcomes between phylogeny and taxonomy. This suggests that it may be particularly advantageous to generate ‘composite’ trees for use in PCMs, where cladistically inferred relationships or taxonomic hypotheses within subclades are grafted onto an overall hypothesis of higher level relationships. In this way taxa for which relationships have been cladistically inferred can be analysed in combination with taxa for which there is not currently a cladogram, by ‘filling in the gaps’ with taxonomic information. Taxonomy will be useful for expanding the temporal and taxonomic breadth of phylogenetic comparative analyses.

6.2 Phylogenetic clustering of extinction

In Chapter 2 I investigated a possible solution to the problem of clades for which no cladistically inferred phylogenetic hypothesis is available. This situation is most often the case for invertebrate clades which are highly speciose, well sampled and may have good stratigraphic constraint, but have a low number of identifiable morphological characters compared to the number of taxa, making cladistic analysis difficult. In Chapters 3 and 4 I investigated a common alternative type of palaeontological data where taxa are poorly sampled (often singletons) with poor stratigraphic constraint, but have many characters and so have robust cladistically inferred phylogenetic hypotheses. This is most often the case with clades of extinct vertebrates.

Using phylogenetic and range data for three extinct clades of tetrapods, I measured phylogenetic clustering of extinction during the Permian and beginning of the Mesozoic. I chose this analysis because the measurement can be made when a phylogeny is available, but detailed data on potential individual extinction linked traits are lacking. My results demonstrate that this measurement can provide a useful proxy for non-fossilising extinction linked traits, particularly for clades which have a robust phylogeny but a low sampling rate.

The results of the analysis of real data showed that extinction during the Permian and Early Triassic the three clades (synapsids, temnospondyls and sauropsids) was often phylogenetically clustered but varied across time bins and between clades. The degree of clustering was not correlated with extinction intensity. The outcomes of analyses of the phylogenetic clustering of extinction were contingent on many different factors, including the nature of the evolutionary relationships of the taxa, the sampling rate and the method used to timescale cladograms. Despite this, when a strong signal was present this was robust to variation in the input data and methods.

In addition to analysis of real data I performed a simulation study to establish the effect of factors such as sampling rate, and degree of clustering, on the frequency with which an accurate estimation of clustering was retrieved, with reference to a measurement made on a simulated ‘true’ phylogeny. The most important result was that the inclusion of ancestor-descendant pairs in phylogenies used to measure the phylogenetic clustering of extinction biases results, due to the pseudoextinctions and pseudosurvivals that it produces. These two biases can cancel each other out to a certain degree but pseudosurvivals dominate at high sampling rates, causing an overestimation of how clustered the true extinctions were.

A particularly interesting outcome, when the results from the analysis of real and simulated data are considered together, is that the strongly clustered signal seen for most of the evolutionary history of tetrapods should not be possible to detect given that the sampling rate of the tetrapod fossil record is very low. At low sampling rates the measurement is biased towards showing random or overdispersed extinction. This suggests there are additional factors effecting the results from real data, two possibilities are that preservation was phylogenetically linked, or that geographic distribution of extinctions was correlated with phylogeny. These are possible areas for future investigation (see section 6.4).

6.3 Evolutionary analyses of Sauropterygia

Following on from the previously outlined analyses that are possible in the absence of trait data, in Chapter 5 I went on to investigate macroevolutionary patterns relating to the body plan of sauropterygians, incorporating methodology from previous chapters with detailed continuous trait data.

From these analyses I could conclude that continuous trait data and a phylogeny can be used to test hypotheses about the processes, and perhaps the mechanisms, governing evolutionary change in the body plan of a large clade over long timescales. This can be done by making comparisons between results across different analyses and different traits.

From this analysis of evolutionary rates and constraints in Sauropterygia I made several key findings. First, change in their axial body proportions was controlled by somitogenetic and homeotic effects during embryonic development rather than differential post-somitic growth. This stands in contrast to other clades (for example mammals) where a longer or shorter neck with reference to body size is achieved by lengthening or shortening of individual vertebrae, and not variation in the number or regionalisation of vertebrae. This suggests that certain types of developmental constraint present in clades like mammals, were absent in Sauropterygia.

Second, changing ratios in sauropterygian vertebral count occurred in nested clades where vertebral count ratio and presacral count were decoupled, rather than on the stem lineages. This suggests that the establishment of higher level body plans occurred via both pleomerism and homeotic effects, but modifications from this within nested subclades occurred mainly through changes in regionalisation. Where extreme morphologies occurred within the subclades somitogenetic effects also contributed, for example during the slow evolutionary acquisition of extremely long necks in elasmosaurids, and very long dorsal regions in Triassic taxa.

Third, a change in evolutionary regime and a consequent release in functional constraint at the Triassic-Jurassic boundary allowed diversification in limb proportions which had been restricted to a particular area of morphospace throughout the Triassic. A variety of evolutionary models fit to the data strongly supported an evolutionary regime change coincident with the extinction of a majority of early sauropterygians such as placodonts and nothosaurs, and a shift in habitat from shallow near shore environments to deeper open water.

Finally, an analysis to identify adaptive peaks for sauropterygian body plan did not strongly support continued attraction towards particular axial and limb configurations. However, given the paucity of data available this did not necessarily rule out the possibility of the presence of an adaptive peak representing a ‘pliosaurian’ body type, to which multiple lineages converged.

6.4 Future Work

The work I present here hints at some interesting patterns which should be followed up in future analyses. An important focus for future work, relevant to the whole thesis, will be to develop more appropriate models for simulating evolutionary histories of fossil clades. Although the birth-death process based models used in this thesis do provide a good approximation of the patterns seen in real data there are aspects of real palaeontological phylogenies that they do not yet fully capture. This is evidenced by the results of Chapters 3 and 4 where real data often show a highly phylogenetically clustered signal at low sampling rates, which is difficult to consistently produce via simulation. One way in which this could be done is to develop a process-based birth-death model that controls speciation and extinction events on lineages with input rates across the whole tree or per clade, rather than per lineage as was used in this thesis. This would allow ‘real-time’ simulation of phylogenetically clustered extinction via either heterogeneous extinction rates across clades and through time, or by controlling the extinction probability of a morphotaxon based on its evolutionary proximity to other recently extinct morphotaxa.

Updated models such as these would also help to confirm and constrain results in the analyses of taxonomic classification as a substitute for cladistically inferred phylogenies (Chapter 2). It would be informative to develop a model approximation of how the real evolutionary relationships between taxa are represented in taxonomies and then, through a similar method to that used in Chapter 4, compare the outcome of results conducted on these simulated taxonomy trees and simulated phylogenies. Although as evidenced by Chapter 2 taxonomy does contain phylogenetic information, the way in which it relates to real evolutionary relationships is not straightforward and probably varies depending on both the taxonomist and the original purpose of the classification. Appropriate models would therefore have to be carefully considered, and may benefit from the incorporation of simulated morphological trait data.

These developments in simulation method could help to uncover the cause of the disparity between simulation study results and results from real data at low sampling rates, found in Chapters 3 and 4. This could be achieved by updating the simulation models of speciation and extinction, and adding in model parameters to create heterogeneity in preservation through time. In addition to this, one interesting avenue to extend the work in Chapter 3 would be to construct a larger, more inclusive dataset of tetrapod (or even vertebrate) taxa that continues through the Mesozoic and into the Cenozoic, therefore including additional extinction events and large range of climatic variability. This may provide a means to access the relationship between the phylogenetic structure of extinction and environmental volatility, or with the ecological impact of extinctions. A prediction based on the evidence that phylogenetically selective extinctions disproportionately impact biodiversity, is that the ecological impact of highly clustered extinction in the fossil record should be greater than for times of equal intensity but less clustered extinction. This hypothesis is not currently supported by the results presented in this thesis but could be tested by a comparison of clustering through time and across clades, by applying a quantitative metric of ecological impact such as lifestyle disparity (Christie et al., 2013).

In Chapter 5 I made progress in the quantitative assessment of the patterns and processes of sauropterygian macroevolution. With addition of more Sauropterygian taxa (particularly from the Triassic) there is an opportunity for further examination of the adaptive peak model in macroevolution. This could also be related to the phylogenetic structure of extinction in the clade to rigorously assess a niche-refilling model, where selective extinctions against taxa with specialised body plans (e.g. the pliosauromorph body plan), in times of high ecological stress vacates adaptive peaks that are then refilled in subsequent radiations.

An extension to this work with the sauropterygian dataset could be comparisons of vertebral count ratios and axial body proportions in a clade-by-clade analyses of many other groups to map variation among amniote clades. A particularly interesting comparison would be with mammals as their cervical, but not dorsal, vertebral counts are highly constrained. It may be that the relationship between count ratio and length ratio, and between count ratios and presacral count varies considerably across clades. A comparison could be used to test hypotheses about the nature of constraints that control axial body plans within each individual clade.

The first parts of this thesis provide the means by which previously inaccessible extinct clades might be mined for knowledge on general patterns and mechanisms of macroevolution, under a phylogenetic framework, and the latter parts provide examples of how this might be done. Hopefully this will open up many avenues for future investigation of macroevolutionary patterns by specialists in a wide variety of clades.

Bibliography

Abouheif, E. (1998). Random trees and the comparative method: a cautionary tale. *Evolution*, 52(4):1197–1204.

Abràmoff, M. D., Magalhães, P. J., and Ram, S. J. (2004). Image processing with imageJ. *Biophotonics International*, 11(7):36–41.

Adamowicz, S. J., Purvis, A., and Wills, M. a. (2008). Increasing morphological complexity in multiple parallel lineages of the Crustacea. *Proceedings of the National Academy of Sciences of the United States of America*, 105(12):4786–4791.

Alba, D. M., Agustí, J., and Moyà-Solà, S. (2001). Completeness of the mammalian fossil record in the Iberian Neogene. *Paleobiology*, 27(1):79–83.

Albright, L. B., Gillette, D. D., and Titus, A. L. (2007). Plesiosaurs from the Upper Cretaceous (Cenomanian-Turonian) tropic shale of Southern Utah, part 1: new records of the pliosaur Brachauchenius lucasi. *Journal of Vertebrate Paleontology*, 27(1):31–40.

Aldous, D. and Popovic, L. (2005). A critical branching process model for biodiversity. *Advances in Applied Probability*, 37(4):1094–1115.

Alroy, J. (2010). Fair sampling of taxonomic richness and unbiased estimation of origination and extinction rates. In Alroy, J. and Hunt, G., editors, *Quantitative Methods in Paleobiology*, pages 55–80. The Paleontological Society.

Alroy, J., Marshall, C. R., Bambach, R. K., Bezusko, K., Foote, M., Fursich, F. T., Hansen, T. A., Holland, S. M., Ivany, L. C., Jablonski, D., Jacobs, D. K., Jones, D. C., Kosnik, M. A., Lidgard, S., Low, S., Miller, A. I., Novack-Gottshall, P. M., Olszewski, T. D., Patzkowsky, M. E., Raup, D. M., Roy, K., Sepkoski, J. J., Sommers, M. G., Wagner, P. J., and Webber, A. (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):6261–6266.

Andrews, C. W. (1910a). *A descriptive catalogue of the marine reptiles of the Oxford Clay, Part I.* British Museum (Natural History), London.

Andrews, C. W. (1910b). The skeleton of Peloneustes. *Geology Magazine*, pages 110–111.

Andrews, C. W. (1913). *A descriptive catalogue of the marine reptiles of the Oxford Clay, Part II.* British Museum (Natural History), London.

Andrews, C. W. (1922). Description of a new plesiosaur from the Weald Clay of Berwick (Sussex). *Quarterly Journal of the Geological Society*, 78(3):285–298.

Angielczyk, K. D. and Kurkin, A. A. (2003). Phylogenetic analysis of Russian Permian dicynodonts (Therapsida: Anomodontia): Implications for Permian biostratigraphy and Pangaean biogeography. *Zoological Journal of the Linnean Society*, 139(2):157–212.

Apaldetti, C., Pol, D., and Yates, A. (2013). The postcranial anatomy of Coloradisaurus brevis (Dinosauria: Sauropodomorpha) from the Late Triassic of Argentina and its phylogenetic implications. *Palaeontology*, 56(2):277–301.

Arcila, D., Alexander Pyron, R., Tyler, J. C., Ortí, G., and Betancur-R., R. (2015). An evaluation of fossil tip-dating versus node-age calibrations in tetraodontiform fishes (Teleostei: Percomorphaceae). *Molecular Phylogenetics and Evolution*, 82:131–145.

Arratia, G. and Tischlinger, H. (2010). The first record of Late Jurassic crossognathiform fishes from Europe and their phylogenetic importance for teleostean phylogeny. *Fossil Record*, 13(2):317–341.

Asher, R. J., Lin, K. H., Kardjilov, N., and Hautier, L. (2011). Variability and constraint in the mammalian vertebral column. *Journal of Evolutionary Biology*, 24(5):1080–1090.

Aze, T., Ezard, T. H. G., Purvis, A., Coxall, H. K., Stewart, D. R. M., Wade, B. S., and Pearson, P. N. (2011). A phylogeny of Cenozoic macroperforate planktonic foraminifera from fossil data. *Biological reviews of the Cambridge Philosophical Society*, 86(4):900–27.

Azzurro, E., Tuset, V. M., Lombarte, A., Maynou, F., Simberloff, D., Rodríguez-Pérez, A., and Solé, R. V. (2014). External morphology explains the success of biological invasions. *Ecology Letters*, 17:1455–1463.

Bakker, R. T. (1993). Plesiosaur extinction cycles - events that mark the beginning, middle and end of the Cretaceous. In Caldwell, W, G, E. and Kauffman, E, G., editors, *Evolution of the Western Interior Basin*, pages 641–664. Geological association of Canada.

Balmford, A. (1996). Extinction filters and current resilience: the significance of past selection pressures for conservation biology. *Trends in ecology & evolution*, 11(5):193–196.

Bapst, D. W. (2012). paleotree: an R package for paleontological and phylogenetic analyses of evolution. *Methods in Ecology and Evolution*, 3(5):803–807.

Bapst, D. W. (2013). A stochastic rate-calibrated method for time-scaling phylogenies of fossil taxa. *Methods in Ecology and Evolution*, 4(8):724–733.

Bapst, D. W. (2014a). Assessing the effect of time-scaling methods on phylogeny-based analyses in the fossil record. *Paleobiology*, 40(3):331–351.

Bapst, D. W. (2014b). Preparing Palaeontological Datasets for Phylogenetic Comparative Methods. In Garamszegi, L. Z., editor, *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology*, chapter 22, pages 515–544. Springer-Verlag, Berlin, Heidelberg.

Barnosky, A. D., Matzke, N., Tomiya, S., Wogan, G. O. U., Swartz, B., Quental, T. B., Marshall, C., McGuire, J. L., Lindsey, E. L., Maguire, K. C., Mersey, B., and Ferrer, E. a. (2011). Has the Earth's sixth mass extinction already arrived? *Nature*, 471(7336):51–7.

Beaulieu, J. M., Jhwueng, D.-C., Boettiger, C., and O'Meara, B. C. (2012). Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. *Evolution*, 66(8):2369–2383.

Bell, M. A., Travis, M. P., and Blouw, D. M. (2006). Inferring natural selection in a fossil threespine stickleback. *Paleobiology*, 32(4):562–577.

Bennett, P. M. and Owens, I. P. F. (1997). Variation in extinction risk among birds: chance or evolutionary predisposition? *Proceedings of the Royal Society B: Biological Sciences*, 264:401–408.

Benoit, J., Adnet, S., El Mabrouk, E., Khayati, H., Ben Haj Ali, M., Marivaux, L., Merzeraud, G., Merigeaud, S., Vianey-Liaud, M., and Tabuce, R. (2013). Cranial Remain from Tunisia Provides New Clues for the Origin and Evolution of Sirenia (Mammalia, Afrotheria) in Africa. *PLoS ONE*, 8(1).

Benson, R. B. J. and Bowdler, T. (2014). Anatomy of Colymbosaurus megadeirus (Reptilia, Plesiosauria) from the Kimmeridge Clay Formation of the U.K., and high diversity among Late Jurassic plesiosauroids. *Journal of Vertebrate Paleontology*, 34(5):1053–1071.

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):191–208.

Benson, R. B. J., Campione, N. E., Carrano, M. T., Mannion, P. D., Sullivan, C., Upchurch, P., and Evans, D. C. (2014a). Rates of dinosaur body mass evolution indicate

170 million years of sustained ecological innovation on the avian stem lineage. *PLoS Biology*, 12(5):e1001853.

Benson, R. B. J. and Druckenmiller, P. S. (2014). Faunal turnover of marine tetrapods during the Jurassic-Cretaceous transition. *Biological Reviews*, 89(1):1–23.

Benson, R. B. J., Evans, M., and Druckenmiller, P. S. (2012). High diversity, low disparity and small body size in plesiosaurs (Reptilia, Sauropterygia) from the Triassic-Jurassic boundary. *PloS one*, 7(3):e31838.

Benson, R. B. J., Frigot, R. a., Goswami, A., Andres, B., and Butler, R. J. (2014b). Competition and constraint drove Cope's rule in the evolution of giant flying reptiles. *Nature communications*, 5:3567.

Benton, M. J. and Storrs, G. W. (1994). Testing the quality of the fossil record: paleontological data is improving. *Geology*, 22:111–114.

Benton, M. J., Tverdokhlebov, V. P., and Surkov, M. V. (2004). Ecosystem remodelling among vertebrates at the Permian-Triassic boundary in Russia. *Nature*, 432:97–100.

Benton, M. J. and Twitchett, R. J. (2003). How to kill (almost) all life: The end-Permian extinction event. *Trends in Ecology and Evolution*, 18(7):358–365.

Bernard, E. L., Ruta, M., Tarver, J. E., and Benton, M. J. (2010). The fossil record of early tetrapods: worker effort and the end-Permian mass extinction. *Acta Palaeontologica Polonica*, 55(2):229–239.

Beutel, R., Ribera, I., and Binindaemonds, O. (2008). A genus-level supertree of Adephaga (Coleoptera). *Organisms Diversity & Evolution*, 7(4):255–269.

Bielby, J., Cunningham, A. A., and Purvis, A. (2006). Taxonomic selectivity in amphibians: ignorance, geography or biology? *Animal Conservation*, 9:135–143.

Bininda-Emonds, O. R. P., Cardillo, M., Jones, K. E., MacPhee, R. D. E., Beck, R. M. D., Grenyer, R., Price, S. A., Vos, R. A., Gittleman, J. L., and Purvis, A. (2007). The delayed rise of present-day mammals. *Nature*, 446(7135):507–512.

Bogdanowicz, D. and Giaro, K. (2013). On a matching distance between rooted phylogenetic trees. *International Journal of Applied Mathematics and Computer Science*, 23(3):669–684.

Bogdanowicz, D., Giaro, K., and Wróbel, B. (2012). TreeCmp: comparison of trees in polynomial time. *Evolutionary Bioinformatics*, 8:475–487.

Brazeau, M. D. (2009). The braincase and jaws of a Devonian 'acanthodian' and modern gnathostome origins. *Nature*, 457(7227):305–308.

Brochu, C. A. (2003). Phylogenetic approaches toward Crocodylian history. *Annual Review of Earth and Planetary Science*, 31:357–397.

Brocklehurst, N., Kammerer, C. F., and Fröbisch, J. (2013). The early evolution of synapsids, and the influence of sampling on their fossil record. *Paleobiology*, 39(3):470–490.

Brusatte, S. I. and Benson, R. B. J. (2013). The systematics of Late Jurassic tyrannosauroid theropods from Europe and North America. *Acta Palaeontologica Polonica*, 58(1):47–54.

Brusatte, S. L., Benton, M. J., Ruta, M., and Lloyd, G. T. (2008). Superiority, competition, and opportunism in the evolutionary radiation of dinosaurs. *Science*, 321:1485–1488.

Brusatte, S. L., Montanari, S., Yi, H.-y., and Norell, M. A. (2011). Phylogenetic corrections for morphological disparity analysis: new methodology and case studies. *Palaeobiology*, 37(1):1–22.

Burnham, K. P. and Anderson, D. R. (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer, New York, 2nd edition.

Butler, M. A. and King, A. A. (2004). Phylogenetic comparative analysis: A modeling approach for adaptive evolution. *The American Naturalist*, 164(6):683–695.

Caldwell, M. W. (2002). From fins to limbs to fins: Limb evolution in fossil marine reptiles. *American Journal of Medical Genetics*, 112:236–249.

Cardillo, M., Mace, G. M., Gittleman, J. L., Jones, K. E., Bielby, J., and Purvis, A. (2008). The predictability of extinction: biological and external correlates of decline in mammals. *Proceedings of the Royal Society B: Biological Sciences*, 275(1641):1441–1448.

Cardillo, M., Mace, G. M., Jones, K. E., Bielby, J., Bininda-Emonds, O. R. P., Sechrest, W., Orme, C. D. L., and Purvis, A. (2005). Multiple causes of high extinction risk in large mammal species. *Science*, 309:1239–1241.

Carr, R. K. and Hlavin, W. J. (2010). Two new species of Dunkleosteus Lehman, 1956, from the Ohio Shale Formation (USA, Famennian) and the Kettle Point Formation (Canada, Upper Devonian), and a cladistic analysis of the Eubrachythoraci (Placodermi, Arthrodira). *Zoological Journal of the Linnean Society*, 159(1):195–222.

Cau, A., Vecchia, F. M. D., and Fabbri, M. (2012). Evidence of a new Carcharodontosaurid from the Upper Cretaceous of Morocco. *Acta Palaeontologica Polonica*, 57(3):661–665.

Christie, M., Holland, S. M., and Bush, A. M. (2013). Contrasting the ecological and taxonomic consequences of extinction. *Paleobiology*, 39(4):538–559.

Chun, L. I. and Rieppel, O. (2002). A new cyamodontoid placodont from Triassic of Guizhou, China. *Chinese Science Bulletin*, 47(5):403–407.

Cloutier, R. and Ahlberg, P. E. (1996a). Interrelationships of basal Neopterygians. In Stiassny, M. L. J., Parenti, L. R., and Johnson, G. D., editors, *Interrelationships of fishes*, pages 117–146. Academic Press, San Diego.

Cloutier, R. and Ahlberg, P. E. (1996b). Morphology, characters and the interrelationships of basal sarcopterygians. In Stiassny, M. L. J., Parenti, L. R., and Johnson, G. D., editors, *Interrelationships of fishes*, pages 445–480. Academic Press, San Diego.

Coates, M. I. (1999). Endocranial preservation of a Carboniferous actinopterygian from Lancashire, UK, and the interrelationships of primitive actinopterygians. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 354(1382):435–462.

Coates, M. I. and Sequeira, S. E. K. (2001). A new stethacanthid chondrichthyan from the lower Carboniferous of Bearsden, Scotland. *Journal of Vertebrate Paleontology*, 21(3):438–459.

Conrad, J. L., Ast, J. C., Montanari, S., and Norell, M. a. (2011). A combined evidence phylogenetic analysis of Anguimorpha (Reptilia: Squamata). *Cladistics*, 27(3):230–277.

Conybeare, W. D. (1824). Discovery of an almost perfect specimen of Plesiosaurus. *Transactions of the Geological Society of London*.

Cooper, N., Bielby, J., Thomas, G. H., and Purvis, A. (2008). Macroecology and extinction risk correlates of frogs. *Global Ecology and Biogeography*, 17(2):211–221.

Cope, E. D. (1887). *The origin of the fittest*. Appleton, New York.

Coria, R. A., Moly, J. J., Reguero, M., Santillana, S., and Marenssi, S. (2013). A new ornithopod (Dinosauria; Ornithischia) from Antarctica. *Cretaceous Research*, 41:186–193.

Cox, L. R., Moore, R., and Tiechert, C. (1971). *Treatise on Invertebrate Palaeontology Part N, Mollusca 6, Bivalvia*. Geological Society of America, Boulder, Colorado.

Cuvier, G. (1818). *Essay on the theory of the Earth*. Kirk and Mercien, New York.

Darwin, C. R. (1859). *On the origin of species by means of natural selection*. John Murray, London.

Davies, T. J., Cooper, N., Felizola Diniz-Filho, J. A., Thomas, G. H., and Meiri, S. (2012). Using phylogenetic trees to test for character displacement: A model and an example from a desert mammal community. *Ecology*, 93(8):44–51.

Davies, T. J., Fritz, S. A., Grenyer, R., Orme, C. D. L., Bielby, J., Bininda-emonds, O. R. P., Cardillo, M., Jones, K. E., Gittleman, J. L., Mace, G. M., and Purvis, A. (2008). Phylogenetic trees and the future of mammalian biodiversity. *PNAS*, 105:11556–11563.

Davies, T. J. and Yessoufou, K. (2013). Revisiting the impacts of non-random extinction on the tree-of-life. *Biology letters*, 9(4):20130343.

Davis, S. P., Finarelli, J. A., and Coates, M. I. (2012). Acanthodes and shark-like conditions in the last common ancestor of modern gnathostomes. *Nature*, 486(7402):247–250.

Dayrat, B. (2005). Ancestor-descendant relationships and the reconstruction of the Tree of Life. *Paleobiology*, 31(3):347–353.

De Vos, J. M., Joppa, L. N., Gittleman, J. L., Stephens, P. R., and Pimm, S. L. (2014). Estimating the normal background rate of species extinction. *Conservation Biology*.

Dequéant, M.-L. and Pourquié, O. (2008). Segmental patterning of the vertebrate embryonic axis. *Nature reviews. Genetics*, 9(5):370–382.

Dornburg, A., Friedman, M., and Near, T. J. (2015). Phylogenetic analysis of molecular and morphological data highlights uncertainty in the relationships of fossil and living species of Elopomorpha (Actinopterygii: Teleostei). *Molecular Phylogenetics and Evolution*, 89:205–218.

Droser, M. L., Bottjer, D. J., Sheehan, P. M., and McGhee G.R., J. (2000). Decoupling of taxonomic and ecologic severity of Phanerozoic marine mass extinctions. *Geology*, 28(8):675–678.

Duncan, P. M. (1889). A revision of the genera and great groups of the Echinoidea. *Zoological Journal of the Linnean Society*, 23(141-144):1–311.

Eastman, J. M., Alfaro, M. E., Joyce, P., Hipp, A. L., and Harmon, L. J. (2011). A novel comparative method for identifying shifts in the rate of character evolution on trees. *Evolution*, 65(12):3578–3589.

Eldredge, N. and Gould, S. J. (1972). Punctuated equilibria: an alternative to phyletic gradualism. In Schopf, T. J. M. and Cooper, F., editors, *Models in Paleobiology*, pages 82–115. Freeman, Cooper and Company.

Erwin, D. (2009). A call to the custodians of deep time. *Nature*, 462:282–283.

Farlow, W. G. (1898). The conception of species as affected by recent investigations on Fungi. *Science*, 8(196):423–435.

Felsenstein, J. (1985). Phylogenies and the comparative method. *The American naturalist*, 125(1):1–15.

Finarelli, J. A. (2008). A total evidence phylogeny of the Arctoidea (Carnivora: Mammalia): Relationships among basal taxa. *Journal of Mammalian Evolution*, 15(4):231–259.

Finarelli, J. A. and Flynn, J. J. (2006). Ancestral state reconstruction of body size in the Caniformia (Carnivora, Mammalia): the effects of incorporating data from the fossil record. *Systematic biology*, 55(2):301–313.

Finnegan, S., Anderson, S. C., Harnik, P. G., Simpson, C., Tittensor, D. P., Byrnes, J., Finkel, Z. V., Lindberg, D. R., Liow, L. H., Lockwood, R., Lotze, H. K., McClain,

C. R., McGuire, J. L., O'Dea, A., and Pandolfi, J. M. (2008). Paleontological baselines for evaluating extinction risk in the modern oceans. *Science*, 348(6234):567–570.

Fisher, D., Foote, M., Fox, D. L., and Leighton, L. R. (2002). Stratigraphy in Phylogeny Reconstruction—Reply To Smith (2000). *Journal of Paleontology*, 76(4):585–586.

Fitzjohn, R. G. (2010). Quantitative traits and diversification. *Systematic Biology*, 59(6):619–633.

Flynn, J. J., Finarelli, J. A., and Spaulding, M. (2010). Phylogeny of the Carnivora and Carnivoramorpha, and the use of the fossil record to enhance understanding of evolutionary transformations. In Goswami, A. and Friscia, A., editors, *Carnivoran Evolution*, chapter 2, pages 25–59. Cambridge University Press, 1 edition.

Foote, M. (1996a). Models of morphological diversification. In Jablonski, D., Erwin, D., and Lipps, J. H., editors, *Evolutionary Paleobiology*, chapter 1.3, pages 62–81. University of Chicago Press, Chicago.

Foote, M. (1996b). On the Probability of Ancestors in the Fossil Record. *Paleobiology*, 22(2):141–151.

Foote, M. and Raup, D. M. (1996). Fossil preservation and the stratigraphic ranges of taxa. *Palaeobiology*, 22(2):121–140.

Foote, M. and Sepkoski, J. J. (1999). Absolute measures of the completeness of the fossil record. *Nature*, 398(April):415–417.

Forey, P. (1998). *History of the Coelacanth fishes*. Springer.

Freckleton, R. P., Cooper, N., and Jetz, W. (2011). Comparative methods as a statistical fix: the dangers of ignoring an evolutionary model. *The American naturalist*, 178(1):E10–E17.

Freckleton, R. P. and Harvey, P. H. (2006). Detecting non-brownian trait evolution in adaptive radiations. *PLoS Biology*, 4(11):2104–2111.

Friedman, M. (2007). Styloichthys as the oldest coelacanth: implications for early osteichthyan interrelationships. *Journal of Systematic Palaeontology*, 5(April):289–343.

Friedman, M. (2009). Ecomorphological selectivity among marine teleost fishes during the end-Cretaceous extinction. *Proceedings of the National Academy of Sciences of the United States of America*, 106(13):5218–23.

Friedman, M. (2012). Parallel evolutionary trajectories underlie the origin of giant suspension-feeding whales and bony fishes. *Proceedings of the Royal Society B: Biological Sciences*, 279(1730):944–51.

Friedman, M. and Brazeau, M. D. (2011). Sequences, stratigraphy and scenarios: what can we say about the fossil record of the earliest tetrapods? *Proceedings of the Royal Society B: Biological Sciences*, 278(1704):432–439.

Friedman, M., Coates, M. I., and Anderson, P. (2007). First discovery of a primitive coelacanth fin fills a major gap in the evolution of lobed fins and limbs. *Evolution and Development*, 9:329–337.

Fritz, S. A. and Purvis, A. (2010). Selectivity in Mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. *Conservation Biology*, 24(4):1042–1051.

Fritz, S. A., Schnitzler, J., Eronen, J. T., Hof, C., Böhning-Gaese, K., and Graham, C. H. (2013). Diversity in time and space: wanted dead and alive. *Trends in ecology & evolution*, 28(9):509–16.

Fröbisch, J. (2008). Global taxonomic diversity of anomodonts (Tetrapoda, Therapsida) and the terrestrial rock record across the Permian-Triassic boundary. *PLoS ONE*, 3(11).

Fröbisch, J. (2013). Vertebrate diversity across the end-Permian mass extinction — Separating biological and geological signals. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 372:50–61.

Fröbisch, N. B., Fröbisch, J., Sander, P. M., Schmitz, L., and Rieppel, O. (2012). Macropredatory ichthyosaur from the Middle Triassic and the origin of modern trophic networks. *Proceedings of the National Academy of Sciences of the United States of America*, 110(4):1393–1397.

Funk, D. J. and Omland, K. E. (2003). Species level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics*, 34(1):397–423.

Fusco, G., Garland, T., Hunt, G., and Hughes, N. C. (2012). Developmental trait evolution in trilobites. *Evolution*, 66(2):314–329.

Gardiner, B. G., Schaeffer, B., and Masserie, J. A. (2005). A review of the lower actinopterygian phylogeny. *Zoological Journal of the Linnean Society*, 144(4):511–525.

Garland, T. and Diaz-Uriarte, R. (1998). Effects of branch length errors on the performance of phylogenetically independent contrasts. *Systematic biology*, 47(4):654–672.

Garland, T. and Diaz-Uriarte, R. (1999). Polytomies and phylogenetically independent contrasts: examination of the bounded degrees of freedom approach. *Systematic Biology*, 48(3):547–558.

Garland, T. and Ives, A. R. (2000). Using the past to predict the present: confidence intervals for regression equations in phylogenetic comparative methods. *The American Naturalist*, 155(3):346–364.

Gauthier, J., Kluge, A. G., and Rowe, T. (1988). Amniote phylogeny and the importance of fossils. *Cladistics*, 4(2):105–209.

Gavryushkina, A., Welch, D., Stadler, T., and Drummond, A. (2014). Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. *PLoS Computational Biology*, 10(12):e1003919.

Giribet, G. and Wheeler, W. (2002). On bivalve phylogeny: a high-level analysis of the Bivalvia (Mollusca) based on combined morphology and DNA sequence data. *Invertebrate Biology*, 121:271–324.

Gittleman, J. L. and Kot, M. (1990). Adaptation: statistics and a null model for estimating phylogenetic effects. *Systematic Zoology*, 39(3):227–241.

Gomez, C., Ozbudak, E. M., Wunderlich, J., Baumann, D., Lewis, J., and Pourquié, O. (2008). Control of segment number in vertebrate embryos. *Nature*, 454(7202):335–339.

Grafen, A. (1989). The phylogenetic regression. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 326(1233):119–157.

Haas, G. (1980). Ein Nothosaurier-Schädel aus dem Muschelkalk des Wadi Ramon (Negev, Israel). *Annalen des Naturhistorischen Museums in Wien*, 83:119–125.

Hagen, O., Hartmann, K., Steel, M., and Stadler, T. (2015). Age-dependent speciation can explain the shape of empirical phylogenies. *Systematic Biology*, 64(3):432–440.

Hanke, G. F. and H, W. M. V. (2004). New teleostome fishes and acanthodian systematics. In Arratia, G., Wilson, M. V. H., and Cloutier, R., editors, *Recent advances in the origin and early radiation of vertebrates*, pages 189–216. Verlag Dr. Friedrich Pfeil, Munchen.

Hansen, T. F. (1997). Stabilising selection and the comparative analysis of adaptation. *Evolution*, 51(5):1342–1351.

Hardy, C., Fara, E., Laffont, R., Dommergues, J., Meister, C., and Neige, P. (2012). Deep-time phylogenetic clustering of extinctions in an evolutionarily dynamic clade (Early Jurassic Ammonites). *PloS one*, 7(5):e37977.

Harmon, L. J., Losos, J. B., Davies, J. T., Gillespie, R. G., Gittleman, J. L., Jennings, B. W., Kozak, K. H., McPeek, M. A., Moreno-Roark, F., Near, T. J., Purvis, A., Ricklefs, R. E., Schluter, D., Schulte II, J. A., Seehausen, O., Sidlauskas, B. L., Torres-Carvajal, O., Weir, J. T., and Mooers, A. O. (2010). Early bursts of body size and shape evolution are rare in comparative data. *Evolution*, 64(8):2385–2396.

Harmon, L. J., Schulte, J. A., Larson, A., and Losos, J. B. (2003). Tempo and mode of evolutionary radiation in iguanian lizards. *Science*, 301(5635):961–964.

Harnik, P. G. (2011). Direct and indirect effects of biological factors on extinction risk in fossil bivalves. *PNAS*, 108(33):13594–13599.

Harnik, P. G., Fitzgerald, P. C., Payne, J. L., and Carlson, S. J. (2014). Phylogenetic signal in extinction selectivity in Devonian terebratulide brachiopods. *Paleobiology*, 40(4):675–692.

Harnik, P. G., Lotze, H. K., Anderson, S. C., Finkel, Z. V., Finnegan, S., Lindberg, D. R., Liow, L. H., Lockwood, R., McClain, C. R., McGuire, J. L., O'Dea, A., Pandolfi, J. M., Simpson, C., and Tittensor, D. P. (2012a). Extinctions in ancient and modern seas. *Trends in Ecology and Evolution*, 27(11):608–617.

Harnik, P. G., Simpson, C., and Payne, J. L. (2012b). Long-term differences in extinction risk among the seven forms of rarity. *Proceedings of the Royal Society B: Biological Sciences*, 279(1749):4969–4976.

Harvey, P. H. and Pagel, M. D. (1991). *The comparative method in evolutionary biology*. Oxford University Press, Oxford.

Hector, J. (1874). On the fossil reptilia of New Zealand. *Transactions of the New Zealand Institute*.

Hedman, M. M. (2010). Constraints on clade ages from fossil outgroups. *Paleobiology*, 36(1):16–31.

Hopkins, M. J. (2011). How species longevity, intraspecific morphological variation, and geographic range size are related: a comparison using late cambrian trilobites. *Evolution*, 65(11):3253–3273.

Hopkins, M. J. and Lidgard, S. (2012). Evolutionary mode routinely varies among morphological traits within fossil species lineages. *Pnas*, 109:20520–20525.

Housworth, E. A. and Martins, E. P. (2001). Random sampling of constrained phylogenies: conducting phylogenetic analyses when the phylogeny is partially known. *Systematic biology*, 50(5):628–639.

Huelsenbeck, J. P. (1994). Comparing the stratigraphic record to estimates of phylogeny. *Paleobiology*, 20(4):470–483.

Hunt, G. (2007). The relative importance of directional change, random walks, and stasis in the evolution of fossil lineages. *Proceedings of the National Academy of Sciences of the United States of America*, 104:18404–18408.

Hunt, G. (2012). Measuring rates of phenotypic evolution and the inseparability of tempo and mode. *Paleobiology*, 38(3):351–373.

Hunt, G. and Rabosky, D. L. (2014). Phenotypic evolution in fossil species: pattern and process. *Annual Review of Earth and Planetary Sciences*, 42:421–441.

Hurum, J. r. H., Nakrem, H. A., Hammer, O. y., Knutsen, E. M., Druckenmiller, P. S., Hryniewicz, K., and Novis, L. K. (2012). An Arctic lagerstätte – the Slottsmøya Member of the Agardhfjellet Formation (Upper Jurassic – Lower Cretaceous) of Spitsbergen. *Norwegian Journal of Geology*, 92:55–64.

Huttenlocker, A. (2009). An investigation into the cladistic relationships and monophyly of therocephalian therapsids (Amniota: Synapsida). *Zoological Journal of the Linnean Society*, 157(4):865–891.

Iimura, T., Denans, N., and Pourquié, O. (2009). Establishment of Hox vertebral identities in the embryonic spine precursors. *Current Topics in Developmental Biology*, 88:201–234.

Ingram, T. and Mahler, D. L. (2013). SURFACE: Detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. *Methods in Ecology and Evolution*, 4(5):416–425.

Jablonski, D. (1994). Extinctions in the fossil record. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 344:11–17.

Jablonski, D. (2005). Mass extinctions and macroevolution. *Paleobiology*, 31(2):192–210.

Jablonski, D. (2007). Scale and hierarchy in macroevolution. *Palaeontology*, 50(1):87–109.

Jablonski, D. (2008a). Biotic interactions and macroevolution: extensions and mismatches across scales and levels. *Evolution*, 62:715–739.

Jablonski, D. (2008b). Extinction and the spatial dynamics of biodiversity. *PNAS*, 105:11528–11535.

Jablonski, D. and Finarelli, J. A. (2009). Congruence of morphologically-defined genera with molecular phylogenies. *PNAS*, 106(20):8262–8266.

Jablonski, D. and Raup, D. M. (1995). Selectivity of End-Cretaceous marine bivalve extinctions. *Science*, 268:389–391.

Jackson, J. B. C. and Erwin, D. H. (2006). What can we learn about ecology and evolution from the fossil record? *Trends in Ecology and Evolution*, 21(6):322–328.

Janevski, G. A. and Baumiller, T. K. (2009). Evidence for extinction selectivity throughout the marine invertebrate fossil record. *Paleobiology*, 35(4):553–564.

Jeffery, C. H. (2001). Heart urchins at the Cretaceous/Tertiary boundary: a tale of two clades. *Paleobiology*, 27(1):140–158.

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*, 13:208.

Karasawa, H. and Kato, H. (2003). The family Goneplacidae MacLeay, 1838 (Crustacea:Decapoda: Brachyura) systematics, phylogeny, and fossil records. *Paleontological Research*, 7(2):129–151.

Kelley, N. P., Motani, R., Jiang, D.-y., Rieppel, O., and Schmitz, L. (2012). Selective extinction of Triassic marine reptiles during long-term sea-level changes illuminated by seawater strontium isotopes. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 400:9–16.

Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., Blomberg, S. P., and Webb, C. O. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, 26(11):1463–1464.

Kemp, T. S. (1999). *Fossils and Evolution*. Oxford University Press, New York.

Kiessling, W. and Aberhan, M. (2007). Geographical distribution and extinction risk: Lessons from Triassic-Jurassic marine benthic organisms. *Journal of Biogeography*, 34:1473–1489.

Kiessling, W., Aberhan, M., Brenneis, B., and Wagner, P. J. (2007). Extinction trajectories of benthic organisms across the Triassic–Jurassic boundary. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 244(1-4):201–222.

Klein, N. (2009). Skull morphology of Anarosaurus heterodontus (Reptilia: Sauropterygia: Pachypleurosauria) from the Lower Muschelkalk of the Germanic Basin (Winterswijk, The Netherlands). *Journal of Vertebrate Paleontology*, 29(3):665–676.

Klein, N. and Scheyer, T. M. (2013). A new placodont sauropterygian from the Middle Triassic of the Netherlands. *Acta Palaeontologica Polonica*.

Knutsen, E. M., Druckenmiller, P. S., and Hurum, J. r. H. (2012). A new plesiosauroid (Reptilia: Sauropterygia) from the Agardhfjellet Formation (Middle Volgian) of central Spitsbergen, Norway. *Norwegian Journal of Geology*, 92:213–234.

Kroh, A. and Smith, A. B. (2010). The phylogeny and classification of post-Palaeozoic echinoids. *Journal of Systematic Palaeontology*, 8(2):147–212.

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).

Ksepka, D. T., Bertelli, S., and Giannini, N. P. (2006). The phylogeny of the living and fossil Sphenisciformes (penguins). *Cladistics*, 22(5):412–441.

Lambers, P. H. (1995). The monophyly of the caturidae (pisces, actinopterygii) and the phylogeny of the halecomorphi. *Geobios*, 28(2):201–203.

Lamsdell, J. C., Braddy, S. J., and Tetlie, O. E. (2010). The systematics and phylogeny of the Stylonurina (Arthropoda: Chelicerata: Eurypterida). *Journal of Systematic Palaeontology*, 8(1):49–61.

Lane, A., Janis, C. M., and Jr, J. J. S. (2005). Estimating paleodiversities: A test of the taxic and phylogenetic methods. *Paleobiology*, 31(1):21–34.

Laurin, M. (2004). The evolution of body size, Cope's rule and the origin of amniotes. *Systematic biology*, 53(4):594–622.

Laurin, M. and Reisz, R. R. (1995). A reevaluation of early amniote phylogeny. *Zoological Journal of the Linnean Society*, 113(2):165–223.

Lee, M. S. Y., Cau, A., Naish, D., and Dyke, G. J. (2014). Morphological clocks in paleontology, and a mid-Cretaceous origin of crown Aves. *Systematic biology*, 63(3):442–449.

Lee, M. S. Y. and Worthy, T. H. (2012). Likelihood reinstates Archaeopteryx as a primitive bird. *Biology Letters*, 8(2):299–303.

Lieberman, B. S. (2003). Paleobiogeography: The relevance of fossils to biogeography. *Annual Review of Ecology, Evolution, and Systematics*, 34:51–69.

Liow, L. H. (2007). Does versatility as measured by geographic range, bathymetric range and morphological variability contribute to taxon longevity? *Global Ecology and Biogeography*, 16:117–128.

Liu, J. and Olsen, P. (2010). The phylogenetic relationships of Eucynodontia (Amniota: Synapsida). *Journal of Mammalian Evolution*, 17(3):151–176.

Livezey, B. C. and Zusi, R. L. (2007). Higher-order phylogeny of modern birds (Theropoda, Aves: Neornithes) based on comparative anatomy. II. Analysis and discussion. *Zoological Journal of the Linnean Society*, 149(1):1–95.

Lloyd, G. T. and Friedman, M. (2013). A survey of palaeontological sampling biases in fishes based on the Phanerozoic record of Great Britain. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 372:5–17.

Lloyd, G. T., Friedman, M., and Bell, M. (2012a). Confidence intervals on node age estimates in vertebrate phylogeny. In *72nd Annual Meeting Society of Vertebrate Paleontology*.

Lloyd, G. T., Wang, S. C., and Brusatte, S. L. (2012b). Identifying heterogeneity in rates of morphological evolution: discrete character change in the evolution of lungfish (Sarcopterygii; Dipnoi). *Evolution*, 66(2):330–348.

Lockwood, J. L., Russell, G. J., Gittleman, J. L., Daehler, C. C., McKinney, M. L., and Purvis, A. (2002). A metric for analyzing taxonomic patterns of extinction risk. *Conservation Biology*, 16(4):1137–1142.

Losos, J. B. (1994). An approach to the analysis of comparative data when a phylogeny is unavailable or incomplete. *Systematic biology*, 43(1):117–123.

Lü, J., Unwin, D. M., Jin, X., Liu, Y., and Ji, Q. (2010). Evidence for modular evolution in a long-tailed pterosaur with a pterodactyloid skull. *Proceedings of the Royal Society B: Biological Sciences*, 277:383–389.

Lucas, S. G. (2009). Timing and magnitude of tetrapod extinctions across the Permo-Triassic boundary. *Journal of Asian Earth Sciences*, 36(6):491–502.

Lydekker, R. (1889). The orders Ichthyopterygia and Sauropterygia. In *Catalogue of the fossil Reptilia and amphibia of the British Museum*. Taylor and Francis.

Maridet, O. and Ni, X.-J. (2013). A new cricetid rodent from the early Oligocene of Yunnan, China, and its evolutionary implications for early Eurasian cricetids. *Journal of Vertebrate Paleontology*, 33:185–194.

Martins, E. P. (1996). Conducting phylogenetic comparative studies when the phylogeny is not known. *Evolution*, 50(1):12–22.

Martins, E. P. and Hansen, T. F. (1997). Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *The American naturalist*, 149(4):646–667.

McGhee, G. R., Sheehan, P. M., Bottjer, D. J., and Droser, M. L. (2004). Ecological ranking of Phanerozoic biodiversity crises: ecological and taxonomic severities are decoupled. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 211:289–297.

McKean, R. S. (2012). A new species of polycotylid plesiosaur (Reptilia: Sauropterygia) from the Lower Turonian of Utah: etending the stratigraphic range of Dolichorhynchops. *Cretaceous Research*, 34:184–199.

McKinney, M. L. (1997). Extinction vulnerability and selectivity: combining ecological and paleontological views. *Annual Review of Ecology and Systematics*, 28(1):495–516.

Mooers, A. O., Vamosi, S. M., and Schluter, D. (1999). Using Phylogenies to Test Macroevolutionary Hypotheses of Trait Evolution in Cranes (Gruinae). *The American Naturalist*, 154(2):249–259.

Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37(3):17–23.

Motani, R. (2009). The evolution of marine reptiles. *Evolution: Education and Outreach*, 2:224–235.

Mulder, E. W. A., Bardet, N., Godefroit, P., and M, J. (2000). Elasmosaur remains from the Maastrichtian type area, and a Review of latest Cretaceous Elasmosaurs (Reptilia, Plesiosauroidea). *Bulletin de L'Institut Royal des Sciences Naturelles de Belguique*, 70:161–178.

Müller, J., Scheyer, T. M., Head, J. J., Barrett, P. M., Werneburg, I., Ericson, P. G. P., Pol, D., and Sánchez-Villagra, M. R. (2010). Homeotic effects, somitogenesis and the evolution of vertebral numbers in recent and fossil amniotes. *Proceedings of the National Academy of Sciences of the United States of America*, 107(5):2118–2123.

Münkemüller, T., Lavergne, S., Bzeznik, B., Dray, S., Jombart, T., Schiffers, K., and

Thuiller, W. (2012). How to measure and test phylogenetic signal. *Methods in Ecology and Evolution*, 3(4):743–756.

Near, T. J., Dornburg, A., and Friedman, M. (2014). Phylogenetic relationships and timing of diversification in gonorynchiform fishes inferred using nuclear gene DNA sequences (Teleostei: Ostariophysi). *Molecular phylogenetics and evolution*, 80(August):297–307.

Neenan, J. M., Klein, N., and Scheyer, T. M. (2013). European origin of placodont marine reptiles and the evolution of crushing dentition in Placodontia. *Nature communications*, 4:1621.

Nesbitt, S. J. (2011). The Early Evolution of Archosaurs: Relationships and the Origin of Major Clades. *Bulletin of the American Museum of Natural History*, 352(352):1–292.

Niklas, K. J. (1978). Coupled evolutionary rates and the fossil record. *Brittonia*, 30(3):373–394.

Norell, M. A. (1992). Taxic origin and temporal diversity: the effect of phylogeny. In Novacek, M. J. and Wheeler, Q. D., editors, *Extinction and Phylogeny*, pages 89–118. Columbia University Press.

Nowak, M. D., Smith, A. B., Simpson, C., and Zwickl, D. J. (2013). A simple method for estimating informative node age priors for the fossil calibration of molecular divergence time analyses. *PloS one*, 8(6):e66245.

O'Connor, J. K. and Zhou, Z. (2012). A redescription of Chaoyangia beishanensis (Aves) and a comprehensive phylogeny of Mesozoic birds. *Journal of Systematic Palaeontology*, 11(7):889–906.

O'Keefe, F. and Wahl, W. (2003). Current taxonomic status of the plesiosaur Pantasaurus striatus from the Upper Jurassic Sundance Formation, Wyoming. *Paludicola*, 4(2):37–46.

O'Keefe, F. R. (2002). The evolution of Plesiosaur and Pliosaur morphotypes in the Plesiosauria (Reptilia: Sauropterygia). *Paleobiology*, 28(1):101–112.

O'Keefe, F. R. and Carrano, M. T. (2005). Correlated trends in the evolution of the plesiosaur locomotor system. *Paleobiology*, 31(4):656.

O'Leary, M. A. and Gatesy, J. (2008). Impact of increased character sampling on the phylogeny of Cetartiodactyla (Mammalia): combined analysis including fossils. *Cladistics*, 24:397–442.

O'Meara, B. C. (2012). Evolutionary inferences from phylogenies: a review of methods. *Annual Review of Ecology, Evolution, and Systematics*, 43:267–285.

O'Meara, B. C., Ané, C., Sanderson, M. J., and Wainwright, P. C. (2006). Testing for different rates of continuous trait evolution using likelihood. *Evolution; international journal of organic evolution*, 60(5):922–933.

Orme, D., Freckleton, R. P., Thomas, G., Petzoldt, T., and Fritz, S. A. (2012). caper: comparative analyses of phylogenetics and evolution in R. *R Foundation for Statistical Computing*.

Osborn, H. F. (1917). *The Origin and Evolution of Life on the theory of action reaction and interaction of energy*. Charles Scribner's Sons, New York, first edition.

Owen, C. B. (1865). A monograph of the fossil reptilia of the Liassic formation.

Owen, R. (1840). Report on British fossil reptiles. Technical report, British Association for the Advancement of Science, Birmingham.

Owen, R. (1861). Monograph of the British fossil Reptilia from the Kimmeridge Clay. Technical report.

Paradis, E., Claude, J., and Strimmer (2004). APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, 20(2):289–290.

Payne, J. L. and Finnegan, S. (2007). The effect of geographic range on extinction risk during background and mass extinction. *PNAS*, 104(25):10506–10511.

Pennell, M. W., Eastman, J. M., Slater, G. J., Brown, J. W., Uyeda, J. C., FitzJohn, R. G., Alfaro, M. E., and Harmon, L. J. (2014). Geiger V2.0: an Expanded Suite of Methods for Fitting Macroevolutionary Models To Phylogenetic Trees. *Bioinformatics*, 30(15):2216–8.

Pennell, M. W. and Harmon, L. J. (2013). An integrative view of phylogenetic comparative methods: Connections to population genetics, community ecology, and paleobiology. *Annals of the New York Academy of Sciences*, 1289:90–105.

Persson, P. O. (1963). A revision of the classification of the Plesiosauria, with a synopsis of the stratigraphical and geographical distribution of the group. *Lunds Universitets Arsskrift N.f. Avd 2*, 59(1):1–59.

Pierce, S. E., Angielczyk, K. D., and Rayfield, E. J. (2009). Shape and mechanics in thalattosuchian (Crocodylomorpha) skulls: Implications for feeding behaviour and niche partitioning. *Journal of Anatomy*, 215:555–576.

Plazzi, F., Ceregato, A., Taviani, M., and Passamonti, M. (2011). A molecular phylogeny of bivalve mollusks: ancient radiations and divergences as revealed by mitochondrial genes. *PLoS ONE*, 6(11).

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News*, 6:7–11.

Pol, D. and Norell, M. a. (2006). Uncertainty in the age of fossils and the stratigraphic fit to phylogenies. *Systematic biology*, 55(3):512–521.

Polly, P. D. (1997). Ancestry and species definition in paleontology: a stratocladistic analysis of Paleocene-Eocene Viverravidae (Mammalia, Carnivora) form Wyoming. *Contributions from the Museum of Paleontolgy, University of Michigan*, 30(1):1–53.

Polly, P. D., Head, J. J., and Cohn, M. J. (2001). *Beyond heterochrony: The evolution of development*. Wiley, New York.

Purvis, A. (2000). Nonrandom extinction and the loss of evolutionary history. *Science*, 288:328–330.

Purvis, A. (2008). Phylogenetic approaches to the study of extinction. *Annual Review of Ecology, Evolution, and Systematics*, 39:301–319.

Purvis, A., Cardillo, M., Greyner, R., and Collen, B. (2005). Correlates of extinction risk: phylogeny, biology, threat and scale. In Purvis, A., Gittleman, J. L., and Brooks, T., editors, *Phylogeny and Conservation*, pages 295–316. Cambridge University Press, Cambridge, 1 edition.

Purvis, A., Gittleman, J. L., and Luh, H. (1994). Truth or consequences: effects of phylogenetic accuracy on two comparative methods. *Journal of theoretical biology*, 167:293–300.

Pyron, R. A. (2011). Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia. *Systematic biology*, 60(4):466–481.

R Core Team (2015). R: A language and environment for statistical computing.

Rabosky, D. L. (2010). Extinction rates should not be estimated from molecular phylogenies. *Evolution*, 64(6):1816–24.

Rabosky, D. L. and Hurlbert, A. H. (2015). Species Richness at Continental Scales Is Dominated by Ecological Limits. *The American Naturalist*, 185(5):572–583.

Rabosky, D. L., Santini, F., Eastman, J., Smith, S. a., Sidlauskas, B., Chang, J., and Alfaro, M. E. (2013). Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nature Communications*, 4:1–8.

Raup, D. M. (1972). Taxonomic diversity during the phanerozoic. *Science*, 177(4504):1065–1071.

Raup, D. M. (1981). Extinction: bad genes or bad luck? *Acta Geologica Hispanica. Concept and method in Paleontology*, 16(1-2):25–33.

Raup, D. M. (1982). Biogeographic extinction: a feasibility test. *Geological Implications of Impacts of Large Asteroids and Comets on the Earth*, Special Pa:277–281.

Raup, D. M. (1991). *Extinction: bad genes or bad luck?* W. W. Norton & Company, NY/London, 1st edition.

Raup, D. M., Gould, S. J., Schopf, T. J. M., and Simberloff, D. S. (1973). Stochastic models of phylogeny and the evolution of phylogeny. *The Journal of Geology*, 81(5):525–542.

Raup, D. M. and Sepkoski, J. J. (1984). Periodicity of extinctions in the geologic past. *PNAS*, 81(3):801–805.

Reisz, R. R., Liu, J., Li, J. L., and Müller, J. (2011a). A new captorhinid reptile, Gansurhinus qingtoushanensis, gen. et sp. nov., from the Permian of China. *Naturwissenschaften*, 98(5):435–441.

Reisz, R. R., Modesto, S. P., and Scott, D. M. (2011b). A new Early Permian reptile and its significance in early diapsid evolution. *Proceedings of the Royal Society B: Biological Sciences*, 278(1725):3731–3737.

Retallack, G. J., Smith, R. M. H., and Ward, P. D. (2003). Vertebrate extinction across Permian-Triassic boundary in Karoo Basin, South Africa. *Geological Society of America Bulletin*, 115(9):1133–1152.

Revell, L. J. (2009). Size-correction and principal components for interspecific comparative studies. *Evolution*, 63(12):3258–3268.

Revell, L. J. (2010). Phylogenetic signal and linear regression on species data. *Methods in Ecology and Evolution*, 1(4):319–329.

Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3(2):217–223.

Rieppel, O. (1998). Corosaurus alcouensis Case and the phylogenetic interrelationships of Triassic stem-group Sauropterygia. *Zoological Journal of the Linnean Society*, 124:1–41.

Rieppel, O. (1999). The Sauropterygian genera Chinchenia, Kwangsisaurus, and Sanchiaosaurus from the Lower and Middle Triassic of China. *Journal of Vertebrate Paleontology*, 19(2):321–337.

Rieppel, O. (2000). *Handbuch der Paläoherpetologie, Part 12A Sauropterygia I.* Verlag Dr. Friedrich Pfeil, Munchen, 2nd edition.

Rieppel, O. and Hagdorn, H. (1997). Paleobiogeography of Middle Triassic Sauropterygia in central and western Europe. In Callaway, J. M. and Nicholls, E. L., editors, *Ancient Marine Reptiles*, chapter 5, pages 121–144. Academic Press.

Rieppel, O. C. (1973). *Handbuch der Paläoherpetologie, Part 12A Sauropterygia I.* Verlag Dr. Friedrich Pfeil, Munchen, 1st edition.

Rieppel, O. C. (1995). The genus Placodus: systematics, morphology, paleobiogeography and paleobiology. *Fieldiana: Geology*, 31:1–44.

Rieppel, O. C. and Kebang, L. (1995). Pachypleurosaurs (Reptilia:Sauropterygia) from the Lower Muschelkalk, and a review of the Pachypleurosauroidea. *Fieldiana: Geology*, 32.

Robinson, D. F. and Foulds, L. R. (1981). Comparison of phylogenetic trees. *Mathematical Biosciences*, 53:131–141.

Romer, A. S. (1945). *Vertebrate Paleontology*. University of Chicago Press, Chicago, 2nd edition.

Ronquist, F., Klopfstein, S., Vilhelmsen, L., Schulmeister, S., Murray, D. L., and Rasnitsyn, A. P. (2012). A total-evidence approach to dating with fossils, applied to the early radiation of the hymenoptera. *Systematic biology*, 61(6):973–99.

Roopnarine, P. D. (2005). The likelihood of stratophenetic-based hypotheses of genealogical succesion. *Special Papers in Palaeontology*, 73:1–15.

Ross, H. A. (2014). The incidence of species-level paraphyly in animals: A re-assessment. *Molecular Phylogenetics and Evolution*, 76(1):10–17.

Roy, K., Hunt, G., and Jablonski, D. (2009). Phylogenetic conservatism of extinctions in marine bivalves. *Science*, 325:733–737.

Rubidge, B. (2005). Middle-late Permian tetrapod faunas from the South African Karoo and their biogeographic significance. *New Mexico Museum of Natural History and Science Bulletin*, 30:292–294.

Russell, G. J., Brooks, T. M., Kinney, M. M. M. C., and Anderson, C. G. (1998). Present and future taxonomic selectivity in bird and mammal extinctions. *Conservation Biology*, 12(6):1365–1376.

Ruta, M. and Benton, M. J. (2008). Calibrated diversity, tree topology and the mother of mass extinctions: the lesson of temnospondyls. *Palaeontology*, 51(1977):1261–1288.

Ruta, M., Cisneros, J. C., Liebrecht, T., Tsuji, L. A., and Mu, J. (2011). Amniotes through mahor biological crises: faunal turnover among parareptiles during the end-Permian mass extinction. *Paleontology*, 54(5):1117–1137.

Ruta, M., Coates, M. I., and Quicke, D. L. J. (2003). Early tetrapod relationships revisited. *Biological reviews of the Cambridge Philosophical Society*, 78:251–345.

Ruta, M., Pisani, D., Lloyd, G. T., and Benton, M. J. (2007). A supertree of Temnospondyli: cladogenetic patterns in the most species-rich group of early tetrapods. *Proceedings of the Royal Society B: Biological Sciences*, 274:3087–3095.

Ruta, M., Wagner, P. J., and Coates, M. I. (2006). Evolutionary patterns in early tetrapods. I. Rapid initial diversification followed by decrease in rates of character change. *Proceedings of the Royal Society B: Biological Sciences*, 273(1598):2107–2111.

Sallan, L. C. and Friedman, M. (2012). Heads or tails: staged diversification in vertebrate evolutionary radiations. *Proceedings of the Royal Society B: Biological Sciences*, 279(1735):2025–2032.

Sander, P. M., Rieppel, O. C., and Bucher, H. (1997). A new pistosaurid (Reptilia: Sauropterygia) from the Middle Triassic of Nevada and its implications for the origin of the plesiosaurs. *Journal of Vertebrate Paleontology*, 17(3):526–533.

Sansom, R. S. (2008). The origin and early evolution of the Osteostraci (Vertebrata): A phylogeny for the Thyestiida. *Systematic Palaeontology*, 6(3):317–332.

Sansom, R. S. and Wills, M. A. (2013). Fossilization causes organisms to appear erroneously primitive by distorting evolutionary trees. *Scientific reports*, 3:2545.

Santini, F. and Tyler, J. C. (2003). A phylogeny of the families of fossil and extant tetraodontiform fishes (Acanthomorpha, Tetraodontiformes), Upper Cretaceous to Recent. *Zoological Journal of the Linnean Society*, 139:565–617.

Sato, T. (2002). *Description of plesiosaurs (Reptilia: Sauropterygia) from the Bearpaw formation (Campanian - Maastrichtian) and a phylogenetic analysis of the Elasmosauridae*. PhD thesis, University of Calgary.

Sato, T., Cheng, Y.-n., Wu, X.-c., and Li, C. (2010). Osteology of Yunguisaurus Cheng et al., 2006 (Reptilia; Sauropterygia), a Triassic Pistosauroid from China. *Paleontological Research*, 14(3):179–195.

Schneider, J. A. (1995). Phylogeny of the Cardiidae (Mollusca, Bivalvia): Protocardiinae, Laevicardiinae, Lahilliinae, Tulongocardiinae subfam. N. and Pleuriocardiinae subfam. N. *Zoologica Scripta*, 24(4):321–346.

Seeley, H. G. (1865). On two new plesiosaurs, from the Lias. *Annals And Magazine of Natural History*.

Seiffert, E. R., Simons, E. L., Boyer, D. M., Perry, J. M. G., Ryan, T. M., and Sallam, H. M. (2010). A fossil primate of uncertain affinities from the earliest late Eocene of Egypt. *Proceedings of the National Academy of Sciences of the United States of America*, 107(21):9712–9717.

Sepkoski, J. J., Bambach, R. K., Raup, D. M., and Valentine, J. W. (1981). Phanerozoic marine diversity and the fossil record. *Nature*, 293:435–437.

Silvestro, D., Kostikova, A., Litsios, G., Pearman, P. B., and Salamin, N. (2015). Measurement Errors Should Always Be Incorporated in Phylogenetic Comparative Analysis. *Methods in Ecology and Evolution*.

Simpson, C., Kiessling, W., Mewis, H., Baron-Szabo, R. C., and Müller, J. (2011). Evolutionary diversification of reef corals: A comparison of the molecular and fossil records. *Evolution*, 65(11):3274–3284.

Simpson, G. G. (1944). *Tempo and mode in evolution*. Columbia University Press, New York, first edition.

Slater, G. J. (2014). Correction to 'Phylogenetic evidence for a shift in the mode of mammalian body size evolution at the Cretaceous-Palaeogene boundary', and a note on fitting macroevolutionary models to comparative paleontological data sets. *Methods in Ecology and Evolution*, 5(7):714–718.

Slater, G. J. and Harmon, L. J. (2013). Unifying fossils and phylogenies for comparative analyses of diversification and trait evolution. *Methods in Ecology and Evolution*, 4:699–702.

Slater, G. J., Harmon, L. J., and Alfaro, M. E. (2012). Integrating fossils with molecular phylogenies improves inference of trait evolution. *Evolution*, 66(12):3931–3944.

Slater, G. J. and Van Valkenburgh, B. (2008). Long in the tooth: evolution of sabertooth cat cranial shape. *Palaeobiology*, 34(3):403–419.

Smith, A. B. (1994). *Systematics and the Fossil Record: documenting evolutionary patterns*. Blackwell Science, Oxford.

Smith, J. T. and Roy, K. (2006). Selectivity during background extinction: Plio-Pleistocene scallops in California. *Paleobiology*, 32(3):408–416.

Smith, N. and Turner, A. (2005). Morphology's role in phylogeny reconstruction: perspectives from paleontology. *Systematic Biology*, 54(1):166–173.

Soul, L. C., Benson, R. B. J., and Weisbecker, V. (2013). Multiple regression modeling for estimating endocranial volume in extinct Mammalia. *Paleobiology*, 39(1):149–162.

Spaulding, M. and Flynn, J. J. (2012). Phylogeny of the Carnivoramorpha: the impact of postcranial characters. *Journal of Systematic Palaeontology*, 10(4):653–677.

Stanley, S. M. (1998). *Macroevolution, pattern and process*. The Johns Hopkins University Press, Baltimore, 2 edition.

Steel, R. (1973). *Handbuch der Paläoherpetologie, Part 16 Crocodylia*. Verlag Dr. Friedrich Pfeil, Munchen, 1st edition.

Sterli, J., Pol, D., and Laurin, M. (2013). Incorporating phylogenetic uncertainty on phylogeny-based palaeontological dating and the timing of turtle diversification. *Cladistics*, 29(3):233–246.

Stone, E. A. (2011). Why the phylogenetic regression appears robust to tree misspecification. *Systematic biology*, 60(3):245–260.

Storrs, G. W. (1993). Function and phylogeny in Sauropterygian (Diapsida) evolution. *American Journal of Science*, 293(A):63–90.

Surkov, M. V. and Benton, M. J. (2004). The basicranium of dicynodonts (Synapsida) and its use in phylogenetic analysis. *Palaeontology*, 47:619–638.

Symonds, M. R. E. (2002). The effects of topological inaccuracy in evolutionary trees on the phylogenetic comparative method of independent contrasts. *Systematic biology*, 51(4):541–553.

Thomas, G. H., Cooper, N., Venditti, C., Meade, A., and Freckleton, R. P. (2014). Bias and measurement error in comparative analyses: a case study with the Ornstein Uhlenbeck model. *bioRxiv*.

Tollerton, V. P. J. (1989). Morphology, taxonomy and classification of the order Eurypterida Burmeister, 1843. *Journal of Paleontology*, 63(5):642–657.

Tomiya, S. (2013). Body size and extinction risk in terrestrial mammals above the species level. *The American naturalist*, 182(6):E196–214.

Turvey, S. T. and Fritz, S. A. (2011). The ghosts of mammals past: biological and geographical patterns of global mammalian extinction across the Holocene. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 366:2564–2576.

Tyler, C. L. and Leighton, L. R. (2011). Detecting competition in the fossil record: support for character displacement among Ordovician brachiopods. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 307:205–217.

Tyler, J. C. (1980). Osteology, phylogeny and higher classification of the fishes of the order Plectognathi. Technical report, U.S. Deparment of Commerce.

Uyeda, J. C., Caetano, D. S., and Pennell, M. W. (2015). Comparative analysis of principal components can be misleading. *Sytematic Biology*, page Advance Access.

Vermeij, G. J. (1973). Biological versatility and earth history. *Proceedings of the National Academy of Sciences of the United States of America*, 70(7):1936–1938.

Verzi, D., Olivares, A. I., and Morgan, C. C. (2014). Phylogeny, evolutionary patterns and timescale of South American octodontoid rodents: The importance of recognising morphological differentiation in the fossil record. *Acta Palaeontologica Polonica*, 59(4):757–769.

Vincent, P., Bardet, N., Pereda, X., Bouya, B., Amaghzaz, M., and Meslouh, S. (2011). Zarafasaura oceanis, a new elasmosaurid (Reptilia: Sauropterygia) from the Maastrichtian phosphates of Morocco and the palaeobiogeography of latest Cretaceous plesiosaurs. *Gondwana Research*, 19(4):1062–1073.

Vincent, P. and Benson, R. B. J. (2013). Anningasaura, a basal plesiosaurian (Reptilia, Plesiosauria) from the Lower Jurassic of Lyme Regis, United Kingdom. *Journal of Vertebrate Paleontology*, 32(5):1049–1063.

Wagner, P. J. (2000a). Likelihood tests of hypothesized durations: determining and accommodating biasing factors. *Paleobiology*, 26(3):431–449.

Wagner, P. J. (2000b). The quality of the fossil record and the accuracy of phylogenetic inferences about sampling and diversity. *Systematic biology*, 49(1):65–86.

Wagner, P. J. and Marcot, J. D. (2013). Modelling distributions of fossil sampling rates over time, space and taxa: assessment and implications for macroevolutionary studies. *Methods in Ecology and Evolution*, 4(8):703–713.

Waller, T. R. (2006). Phylogeny of families in the Pectinoidea (Mollusca: Bivalvia): importance of the fossil record. *Zoological Journal of the Linnean Society*, 148(3):313–342.

Weischampel, D. B., Dodson, P., and Osmolska, H. (2004). *The Dinosauria*. University of California press, Berkeley and Los Angeles, 2nd edition.

Wiens, J. J., Kuczynski, C. A., Townsend, T., Reeder, T. W., Mulcahy, D. G., and Sites, J. W. (2010). Combining phylogenomics and fossils in higher-level squamate reptile phylogeny: Molecular data change the placement of fossil taxa. *Systematic Biology*, 59(6):674–688.

Xin, X. and Li, X. (2014). Sensitivity Analysis of Bipartition Dissimilarity under Tree Rearrangement Operations. In *Proceedings of the International Conference on Logistics, Engineering, Management and Computer Science*, Paris, France. Atlantis Press.

Zhao, L., Sato, T., and Li, C. (2008). The most complete pistosauroid skeleton from the Triassic of Yunnan, China. *Acta Geologica Sinica*, 82(2):283–286.

Appendix A

Publication

Paper published in Systematic Biology based on the research completed in Chapter 2 of the thesis.

Syst. Biol. 0(0):1–13, 2015

DOI:10.1093/sysbio/syv015

Taxonomy and Phylogeny Can Yield Comparable Results in Comparative Paleontological Analyses

LAURA C. SOUL* and MATT FRIEDMAN

Department of Earth Sciences, University of Oxford, South Parks Road, Oxford OX1 3AN, UK
**Correspondence to be sent to: Department of Earth Sciences, University of Oxford, South Parks Road, Oxford, OX1 3AN UK;*
E-mail:laura.soul@earth.ox.ac.uk.

Received 26 November 2013; reviews returned 29 January 2015; accepted 18 March 2015
Associate Editor: Adrian Paterson

Abstract.—Many extinct taxa with extensive fossil records and mature taxonomic classifications have not yet been the subject of formal phylogenetic analysis. Here, we test whether the taxonomies available for such groups represent useful (i.e., non-misleading) substitutes for trees derived from matrix-based phylogenetic analyses. We collected data for 52 animal clades that included fossil representatives, and for which a recent cladogram and pre-cladistic taxonomy were available. We quantified the difference between the time-scaled phylogenies implied by taxonomies and cladograms using the matching cluster distance metric. We simulated phenotypic trait values and used them to estimate a series of commonly used, phylogenetically explicit measures (phylogenetic signal [Blomberg's *K*], phylogenetic generalized least squares [PGLS], mode of evolution [Brownian vs. Ornstein–Uhlenbeck], and phylogenetic clustering of extinction [Fritz and Purvis' *D*]) in order to determine the degree to which they co-varied on taxonomic and cladistic trees. With respect to topology taxonomies are good approximations of the underlying evolutionary relationships as recorded in inferred cladograms. Detection of phylogenetic clustering of extinction could not be properly assessed. For all other evolutionary analyses, results from taxonomy-based phylogenies (TBPs) co-varied with those from cladogram-based phylogenies (CBPs), but individual comparisons could be misleading. The relative length of terminal branches (influenced by stratigraphy and sampling rate) is a key control on the shared information between, and therefore the relative performance of, TBP and CBP. Collectively these results suggest that under particular circumstances and after careful consideration some taxonomies, or composite trees that incorporate taxonomic information, could be used in place of a formal analytical solution, but workers must be cautious. This opens certain parts of a previously inaccessible section of the fossil record to interrogation within an explicit comparative framework, which will help to test many classical macroevolutionary hypotheses formulated for groups that currently lack formal phylogenetic estimates. [Evolutionary mode; fossil record; PGLS; phylogenetic signal; phylogeny; taxonomy.]

The fossil record yields information on macroevolutionary patterns that remains inaccessible from the study of extant organisms alone (Smith and Turner 2005; Waller 2006; Flynn et al. 2010). Paleontological data expand our understanding of evolution beyond the ‘snapshot’ of the present day to encompass long-term trends and processes that have acted over the history of life on Earth (Sepkoski et al. 1981; Raup and Sepkoski 1984; Jablonski 2007; Slater and Harmon 2013; Wagner and Marcot 2013). Early observations of patterns and apparent trends in the fossil record (Cope 1887; Osborn 1917; Simpson 1944) inspired development of many of the phylogenetic comparative methods (PCMs) presently used to quantitatively assess evolutionary hypotheses. Current approaches incorporate a temporally scaled, explicit hypothesis of relationships among taxa (an inferred phylogeny) in order to correctly estimate relationships between traits (Felsenstein 1985; Grafen 1989; Martins and Hansen 1997), investigate the phylogenetic signal of traits (Fritz and Purvis 2010; Münkemüller et al. 2012), and compare models of evolutionary mode (Harmon et al. 2003, 2010).

These comparative methods were originally developed for recent data sets, where detailed morphological and molecular data can be obtained and used for explicit, character-based phylogenetic inference. Many clades have few or no extant species and yet are diverse and highly abundant in the fossil record, and have the potential to reveal a great deal about long-term evolutionary change and mechanisms (Jablonski 2007, 2008). Unfortunately these same clades have often not been subjected to formal phylogenetic analysis for a variety of reasons, the most acute of which are limited morphological data (Sansom and Wills 2013) and the time-consuming nature of morphological phylogenetic systematics relative to molecular approaches. Consequently, such clades have not been a major focus of comparative evolutionary analysis despite the unique ‘deep-time’ perspective that they provide. An increasing number of workers circumvent the problems raised by the lack of a formal phylogeny by executing analyses using PCMs that incorporate taxonomy (the traditional method of classification, under the Linnaean hierarchy) at least as a partial substitute for an analytically inferred cladogram (Slater and Van Valkenburgh 2008; Roy et al. 2009; Fusco et al. 2012; Hardy et al. 2012; Benson et al. 2014) on the basis that morphologically defined taxonomic classifications closely match trees inferred from molecular data (Jablonski and Finarelli 2009). Conversion of taxonomy to a phylogeny and use in a PCM is now easily achievable using various software (e.g., the *ape* and *geiger* packages in R [Paradis et al. 2004; R Core Team 2013; Pennell et al. 2014] each provide functions for this task), which further facilitates and encourages this practice.

The response of PCMs to various factors has largely been characterized for ultrametric trees of extant groups (Housworth and Martins 2001; Davies et al. 2012). Phylogenetic independent contrasts (Felsenstein 1985) and phylogenetic generalized least squares (PGLS) (Grafen 1989; Martins and Hansen 1997) in particular have been shown to be robust to inaccuracies in branch lengths (Garland and Diaz-Uriarte 1998), and inaccuracies in topology (Symonds 2002; Stone 2011) especially in cases where branch lengths are known (Purvis et al. 1994). Use of taxonomy for extant data sets has previously been rejected in favor of performing PCMs on large sets of random possible trees (Losos 1994; Martins 1996), but other workers have suggested that this approach is at best equivalent to using the raw data, without a phylogenetic correction (Abouheif 1998; Symonds 2002). Taxonomy therefore potentially represents a viable middle ground that contains useful information on evolutionary relationships and, in the case of fossil data, information on branch lengths based on the stratigraphic distribution of fossils.

However, taxonomies could represent inaccurate topologies and are more poorly resolved relative to formal solutions (Harvey and Pagel 1991; Kroh and Smith 2010). The effects of these issues on downstream PCMs are not well understood and have not been directly tested in the context of paleontological data sets, where the application of taxonomies in place of phylogenies is most likely to occur. It is therefore important to understand the adequacy of these topologies when they are used in this context, to assess the validity of work that has already been done, and define guidelines for the limits of acceptable use in the future. We seek to constrain this problem through a series of experiments using taxonomies and more recently published cladograms for various animal clades. Our approach assumes that the results of formal phylogenetic analysis—the 'gold standard' for the application of PCMs to paleontological data—capture the true topology and, after time-scaling, the true branch length of the underlying phylogeny. We make these assumptions because they match those made when a PCM is applied using a cladistically inferred phylogeny of fossil taxa. We simulate continuous characters on the cladistically inferred phylogeny, then estimate relevant values using both this phylogeny and a phylogeny derived from taxonomy.

The question of whether cladistically inferred phylogenies of fossil taxa are in fact suitable for use in these kinds of analyses in the first instance is not one that we address here, and is under investigation by other workers (Bapst 2014). Comparative studies in the fossil record are performed under the widely accepted assumption that a correction using an analytically inferred phylogeny of fossil taxa produces more reliable results than an analysis with no phylogenetic correction (Felsenstein 1985; Laurin 2004; Brusatte et al. 2011). Increasing numbers of studies also assume that taxonomy is a suitable (at least partial) substitute for an analytically inferred phylogeny (Slater and Van Valkenburgh 2008; Roy et al. 2009; Turvey and Fritz 2011; Fusco et al. 2012; Hardy et al. 2012; Benson et al. 2014) and it is this second assumption we aim to test. We use a comparison between results from these two alternative phylogenetic frameworks to develop and present practical guidelines for incorporating taxonomy into PCMs when a formal cladistic solution is not yet available, and discuss the degree to which this approach can be applied in a paleontological framework.

MATERIALS AND METHODS

Taxonomies and Formal Phylogenetic Solutions

We examined 52 animal groups for which both a recently published explicit phylogenetic hypothesis and a traditional taxonomy that predated the most recent cladogram were available. The taxonomies used here can be considered a 'worst-case scenario', as they are precladistic, and were the (historically) earliest classification we could find that included a sufficient sample of taxa. More recent taxonomies—many based on formal cladistic solutions—are now available for these clades, so our estimate of the success of a taxonomy relative to a cladogram is a conservative one. The data set comprised seven invertebrate clades and 45 vertebrate clades, including 28 tetrapod groups and 17 non-tetrapod vertebrate groups; these are listed in online Appendix 1, available on Dryad at http://dx.doi.org/10.5061/dryad.pt3qm. Some cladograms were sourced from the same clade, but contained different taxa and so were used as separate data sets (e.g., we included three different bivalve data sets). These studies are the total set of clades for which we could find an appropriate taxonomy and cladogram that contained enough matching taxa for a PCM to be run on the tree, after taxa not present in both were removed. This has led to a vertebrate-heavy data set, as this clade is the overwhelming focus of formal phylogenetic inference in the fossil record. Invertebrates (e.g., bivalves or marine arthropods) are comparatively underrepresented in formal phylogenetic analyses and so make up a smaller component of our data set than vertebrates, despite having a more extensive fossil record. Many of the taxonomies were taken from the same publication (Romer 1945) as this was the most comprehensive (and for many clades the only) precladistic taxonomy available. We included other sources where possible to provide a comparison to ensure there was no systematic bias in results produced using taxonomies from Romer. All but one of the classifications are at the generic level, thereby reducing the potential influence of different authors tendency to 'lump or split' fossil species.

For each example, we investigated the degree to which evolutionary inferences drawn from taxonomy-based phylogenies (TBPs) matched those derived from cladogram-based phylogenies (CBPs). We took this approach based on the following assumption: a comparison between out-of-date taxonomies and current phylogenetic solutions is equivalent to a comparison between current taxonomies and the hypothetical phylogenetic solutions of the future for groups that do not currently have them. For the purposes of our analyses, we assume that the branching structure provided by a CBP matches that of the true phylogeny. As such, our assessment of the success of CBPs in capturing evolutionary processes and patterns should be viewed as maximally optimistic and our results should not be taken as evidence for or against the capacity of these methods to capture real patterns. In particular, because our trait data are simulated, none of the analyses provide any information on real patterns in these groups.

Conversion of Taxonomies and Cladograms into Phylogenies: Tree Construction and Time-Scaling

All tree manipulation and analyses were performed in the R software environment (3.0.2; R Core Team 2013). Topologies derived from cladograms and taxonomies were time-scaled in order to produce phylogenies (method outlined below). References and details for the source topologies are shown in online Appendix 1, available on Dryad at http://dx.doi.org/10.5061/dryad.pt3qm. All data sets are at the generic level except that for tetraodontiform fishes, where species-level classifications and range data were available. When selecting cladograms we used whichever tree topology the original authors had applied for phylogenetic comparative analyses (if included), or the topology preferred by the original authors in the absence of further analyses within the publication. This was to ensure that our data set included topologies that would be the most likely to be accepted for use with PCMs incorporating paleontological data. Our data set therefore included solutions arising from Bayesian, maximum-likelihood and maximum parsimony inference. The literature used to obtain taxonomies only contained one classification scheme for each clade, and this was converted in to a tree structure as a series of nested polytomies corresponding to each taxonomic rank (Fig. 1).

Taxonomies by nature contain many polytomies when directly plotted as trees (e.g., if there are five genera contained within one family, these genera would be depicted as a single multichotomy, unless subfamilial relationships had been proposed). These were left as hard polytomies to represent the maximum amount of resolution based on available information, except where the PCM required a fully resolved tree (mode of evolution). In preliminary analyses (Supplementary Material: Results, available on Dryad at http://dx.doi.org/10.5061/dryad.pt3qm), executing simulations where (i) taxonomies were randomly resolved before time-scaling or (ii) random trees used for comparison were collapsed to have the same number of internal nodes as the TBP did not make a notable or systematic difference to the outcome. This is consistent with previous work showing that the inclusion of polytomies in a phylogeny for a PCM does not bias the result and has a negligible effect on the rate of type I error (Garland and Diaz-Uriarte 1999; Stone 2011). Housworth and Martins (2001) provide a method by which error caused by uncertainty in relationships within a polytomy can be incorporated into estimates of error bounds for the test statistic in a PCM.

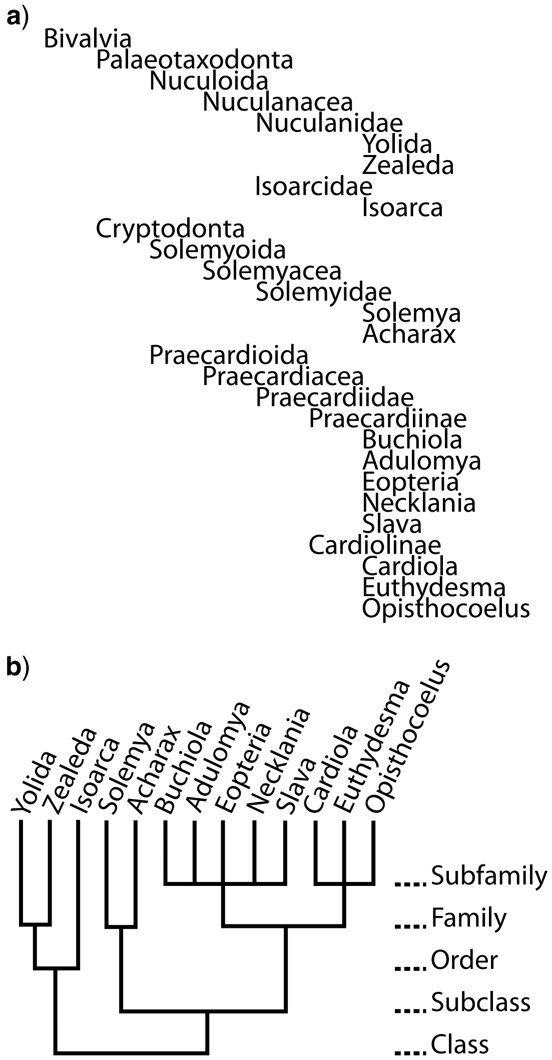


FIGURE 1. Method for converting a taxonomic classification to a cladogram that can then be time-scaled with fossil range data to make a phylogeny. Taxa that are in the same group at a particular rank are combined in a polytomy, starting at the genus level and moving toward the root of the tree. a) The original classification as published. b) The resulting cladogram after conversion, before time-scaling.

Cladograms of extinct taxa can be scaled according to the first appearance date (FAD) of each taxon to generate phylogenies with branch lengths representing the amount of time since sister taxa diverged (Lloyd et al. 2012; Bapst 2013, 2014). The branch lengths are estimated based on the FAD of each taxon in the fossil record, and the assumption that the divergence between two lineages must have occurred, at the latest, at the FAD of the older taxon. Some analyses also require an estimate of the last appearance date of a taxon (e.g., measuring the phylogenetic clustering of extinction) to estimate a taxon duration. First and last possible appearance dates for all taxa derived principally from the Paleobiology Database (PaleoBioDB; www.paleobiodb.org last accessed March 30, 2015). These data were modified where the taxon was known to be extant, and where investigation of outliers showed an incorrect assignment. Where these data were not included in the PaleoBioDB, they were taken from literature detailed in online Appendix 1, available on Dryad at http://dx.doi.org/10.5061/dryad.pt3qm. These first and last occurrences are unlikely to correspond precisely to the real origin and extinction times, because the fossil record is not continuous. Therefore, several methods have been developed to estimate divergence dates and lineage durations that are a closer approximation to real divergences than the raw first and last occurrence dates of fossil taxa (e.g., Lloyd et al. 2012; Bapst 2013; Lee et al. 2014). A further complication with many of the taxa that we used is that they are point occurrences, meaning that only one specimen, or a few specimens from the same bed, have been found. In this case the date is often only known to stage-level precision. For these taxa the ages in Ma for the FADs were drawn randomly from a uniform distribution specified by the beginning and end date of the stage in which they were found.

We time-scaled all the trees using the minimum branch length (mbl) method as originally implemented in R by Lloyd et al. (2012) and incorporated into the *paleotree* package (Bapst 2012). This is a commonly used time-scaling method in paleontological analysis that avoids the problem of analytically intractable phylogenies caused by zero-length branches (ZLBs). Having zero length internal branches effectively creates a polytomy. Analyses are possible if these ZLBs are treated as hard polytomies, however, we have positive information that a bifurcation occurred and that these branch lengths were not in fact zero, and therefore should not be expressed as a polytomy. Divergences are dated according to the FAD of the sister group to the clade, and then any branch that is shorter than a specified length of time (for this study we used 1 myr) is extended to this length. Significantly more complex approaches to calibrating paleontological trees are available but the information needed to perform them is not readily available for the majority of our data sets, and would be unavailable for many of the clades for which authors may want to include taxonomic information. For example, some methods require the clade to include extant taxa (Pyron 2011; Ronquist et al. 2012; Nowak et al. 2013). The latest total evidence dating, or fossil tip dating methods based on so-called morphological clocks require a character matrix to simultaneously estimate phylogenetic topology and branch lengths (Lee et al. 2014), which is by definition unavailable for taxonomies without formal phylogenetic solutions. Other methods require estimation of branching, extinction, and sampling rates (Bapst 2013), which is not possible with the point occurrence data that we have for many of the clades in our data set. For some well-sampled clades it may be possible to use a more sophisticated method than mbl (particularly *cal3* [Bapst 2013]), and we encourage workers to use the most suitable time-scaling method available, as improvement in accuracy of branch length can lead to improvement of performance in analyses (Bapst 2014).

Tree Similarity

We used the matching cluster (MC) distance metric (Bogdanowicz and Giaro 2013) to measure the similarity of each TBP to its equivalent CBP based on differences in branching pattern, implemented using TreeCmp in Java (Bogdanowicz et al. 2012). The method divides the tree into clusters such that all the branches of the tree are included; a single cluster contains all the tips that descend from a particular internal node. Each cluster is matched with one from the other tree for a pairwise comparison. For each pairwise comparison the "symmetric difference" is calculated. This is the total number of differences in taxon positions between one cluster and the other, weighted by the proximity of each taxon to the root of the tree. This difference is summed across all matched pairs to obtain a total value. The method searches across all possible permutations of clusters and matchings to find the minimum total value possible, this is the "minimum-weight perfect matching" or the MC distance (Bogdanowicz and Giaro 2013). We found the significance of this value by: (i) simulating random trees with the same number of taxa; (ii) time-scaling these random topologies using the same data with which we time-scaled the cladogram and taxonomy; and (iii) calculating their MC distance in order to generate a distribution against which to compare our empirical values. This method shares some of its basis with the commonly used Robinson–Foulds (RF) distance (Robinson and Foulds 1981). However, the null distribution generated for MC is unimodal and close to Gaussian, as opposed to the null distribution for the RF distance which is often bimodal. Furthermore, the metric itself is more robust; RF distance can respond drastically to small changes in the position of terminal nodes (e.g., moving one terminal branch to a different position can generate the maximum RF distance [Xin and Li 2014]), MC distance does not suffer from this problem (Münkemüller et al. 2012). We also tested the correlation between the CBP and TBP variance–covariance matrices using Pearson's product moment correlation coefficient.

Relative Performance of Cladogram and Taxonomy-Based Trees in PCMs

We estimated four phylogenetically explicit measures, using a variety of packages from the statistics computing environment R (Table 1). The details of all the steps used for investigation of each method are shown in online Appendix 2, available on Dryad at http://dx.doi.org/10.5061/dryad.pt3qm.

Phylogenetic signal.—We simulated trait data on the CBP under a Brownian motion (BM) model with $\sigma^2=1$ (step variance) and $\mu=0$ (mean) then followed

TABLE 1. Downstream analyses tested, and the R package and function implemented

Measurement	Metric	R package—function
Phylogenetic signal	Blomberg's K	picante—Kcalc (Kembel et al. 2010)
Clustering of extinction	Fritz and Purvis' D	caper—phylo.d (Orme et al. 2012)
Mode of evolution	Model and Akaike weights	OUwie—OUwie and OUwie.sim (Beaulieu et al. 2014)
Phylogenetic generalized least squares	PGLS slope	caper—pgls (Orme et al. 2012)

the steps in online Appendix 2, available on Dryad at http://dx.doi.org/10.5061/dryad.pt3qm. After simulation we controlled the strength of signal using equation (1) as proposed in Münkemüller et al. (2012) to make three sets of simulations with weak, medium, and strong signal:

$$\text{trait} = w \cdot \text{trait}_{\text{BM}} + (1-w)\text{trait}_{\text{rand}} \tag{1}$$

where trait_{BM} is a set of trait values simulated under pure BM with $\mu = 0$ and $\sigma = 1$. $\text{trait}_{\text{rand}}$ is the same values randomly shuffled over the tips of the phylogeny to produce a set of random trait values. The weighting factor w is then used to control which component has the larger influence, on a spectrum from completely Brownian (strong signal, $w = 1$) to white noise (weak signal, $w = 0$), with a medium signal using $w = 0.5$. We then z-transformed all values in order to standardize results.

Phylogenetic generalized least squares.—We simulated sets of coevolving traits with a phylogenetic regression slope of 0.3, and used the TBP and CBT to estimate a value for the slope of each set of simulated traits, according to the steps in online Appendix 2, available on Dryad at http://dx.doi.org/10.5061/dryad.pt3qm.

Mode of evolution.—We simulated data using either a Brownian or Ornstein–Uhlenbeck (OU) model of gradual change. We then used likelihood analysis and the Akaike information criterion (with a correction for small sample sizes [AICc] [Burnham and Anderson 2002]) on the TBP and CBP with the same simulated trait data to identify which model (Brownian or OU) had greater support, and whether this corresponded to the model originally used, via the steps in online Appendix 2, available on Dryad at http://dx.doi.org/10.5061/dryad.pt3qm. For all simulations σ^2 (step variance) was set to 1, θ (the optimum trait value) was set to 0, and was the same as the ancestral value, α (strength of the selective constraint in OU model) was set at 0.1 and 0.05, respectively, for two different sets of simulations. We chose these α values to be biologically realistic for consistent comparisons between TBPs and CBPs. However, the ability to detect an OU model scales with tree height, so in order to make a comparison of TBP suitability across trees of different heights we also provide results of an analysis where α for each simulation corresponded to a phylogenetic half-life of one-fourth of total tree height for each tree. In addition, we selected clades where the model used for simulation was consistently the model with greater support on both the CBP and TBP and recorded the parameter estimates for σ and α. This was to test not only whether the original model could be identified on the TBP, but that the values estimated for these parameters were representative of the parameter values estimated on the CBP. We chose to present AICc weights as this is the most commonly used procedure in paleobiological studies, where often several models are being compared (Fusco et al. 2012; Sallan and Friedman 2012; Lloyd and Friedman 2013; Soul et al. 2013; Bapst 2014). However, as we were only comparing two nested models, we also performed likelihood ratio tests using the χ^2 distribution in line with some other studies (Finarelli 2010; Friedman and Brazeau 2011) and provided the results in Supplementary Material, available on Dryad at http://dx.doi.org/10.5061/dryad.pt3qm, for reference.

Phylogenetic clustering of extinction.—Detection of phylogenetic clustering of extinction could not be properly assessed. The method requires time-slicing each phylogeny, thereby reducing the number of taxa included. This rendered the majority of our data sets too small to obtain a robust result. However, we have included a full description of the method and our results in online Appendix 3, available on Dryad at http://dx.doi.org/10.5061/dryad.pt3qm.

Vital statistics.—We collected a set of vital statistics for each clade in order to examine the degree to which these might co-vary with similarity between CBPs and TBPs: number of taxa, taxonomic resolution, taxonomic stratigraphic consistency index (Huelsenbeck 1994), the number of ranks included in the taxonomic classification, the overall length of the tree and the average number of co-occurring lineages. Taxonomic resolution is a measure of the proportion of internal nodes that are bifurcating. Number of ranks is a count of the number of Linnaean ranks including any informal ranks the author had used (e.g., tribe, infraclass) encompassed by the taxonomy (e.g., a taxonomy that includes genus, tribe, family, and order groupings has a score of 4). Stratigraphic consistency is a measure of how well the appearance and disappearance of taxa in the record matches their position in the tree. We performed regression analyses using these characteristics as independent variables to predict the values of summary statistics for the MC distance and PCM results. This provided a measure of which features of a taxonomy can be used to give an indication of whether it is likely to be suitable for use in analyses incorporating a PCM.

Extended range analysis.—This analysis assessed the effect of using fossil data that includes 'singletons' (taxa known from a single sampling horizon), versus taxa that have many occurrences through time and potentially better constrained FADs. There are two possible causes of singletons: (i) the extinction rate of the lineage was high (i.e., its evolutionary duration was genuinely short) or (ii) the lineage is poorly sampled. Many of the clades in our data set are composed of singletons, and this is often the case with fossil vertebrate clades (e.g., Brusatte et al. 2008; Benson and Druckenmiller 2013). For this reason there are few reliable estimates of sampling rates for vertebrate groups, as these estimates require some of the taxa to have more than one occurrence (although see Foote and Sepkoski 1999; Foote et al. 1999; Friedman and Brazeau 2011). However, if a clade was widely distributed, abundant, or lived in a place where preservation potential was high (e.g., shallow marine settings with a high sediment influx), we might expect higher sampling rates leading to fewer or no singletons and better constrained divergence dates. This scenario is more in line with much of the invertebrate fossil record (Foote and Raup 1996; Foote and Sepkoski 1999). Many of the taxa in our data sets are singletons; to represent the alternative situation with a higher sampling rate including few or no singletons, we made the FAD of each taxon the beginning of the stage in which it occurred, and made the sampled population (the tip of each branch, at which trait values are measured) the end of the stage in which it occurred. This introduces more shared branch length between the CBP and TBP and is intended to represent a scenario in which we have good stratigraphic constraint on FADs and later occurrences. We performed the phylogenetic signal and PGLSs analyses on these trees with greater shared branch length.

RESULTS

Results from all simulations and analyses, including preliminary analyses and the data and R code used can be found in the Supplementary Material, available on Dryad at http://dx.doi.org/10.5061/dryad.pt3qm.

Tree Similarity

TBPs are a good match for the most recent CBPs. The majority of MC distances between the TBP and CBP are significantly ($\alpha = 0.05$) smaller than random with larger trees (more than ~20 tips) showing a better-than-random match more consistently than smaller trees (Fig. 2). In addition there are strong significant correlations between the variance–covariance matrices of the TBP and CBP for each clade; 35 out of 52 clades have correlation coefficient (r) estimates of greater than 0.9, and the lowest value of r is 0.6937 for turtles (Supplementary Material: Results, available on Dryad at http://dx.doi.org/10.5061/dryad.pt3qm). This shows that, as would be expected, taxonomies are good approximations of the underlying evolutionary relationships as recorded in inferred phylogenies.

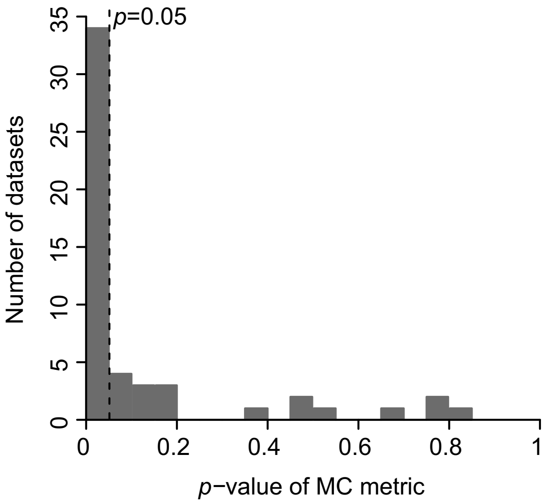


FIGURE 2. MC distance for overall tree similarity. Histogram of significance values for the MC distance between the TBP and CBP. Values at less than $p = 0.05$ show clades for which the TBP is significantly more similar to the CBP than random trees that have been time-scaled with the same fossil range data.

Phylogeny-Based Analyses

When considered together, the results in the following sections indicate that in some cases the variation in topology between the CBP and TBP is sufficiently small that the results of downstream analyses on a TBP are unlikely to be misleading, but can be highly misleading in others. There are some indicators of whether a taxonomy will be suitable for use in a PCM, which lead to some general heuristics, but we cannot propose definitive *a priori* criteria for isolating which taxonomies will be sufficient. We therefore advise that the incorporation of taxonomies into PCMs should be approached with caution, and assessed on a case-by-case basis.

Phylogenetic signal.—Values for K obtained using TBPs (K_T) are significant ($P \ll 0.0001$, null hypothesis is a regression slope of 0) predictors of values for K obtained using CBPs (K_C) (major-axis $R^2 = 0.6141$, Fig. 3). When phylogenetic signal is strong, values of K_T are generally an underestimate of values of K_C (slope = 1.3246), whereas when phylogenetic signal is weak, values of K_T are generally a small overestimate of values of K_C (slope = 0.9000). Pvalues show that in most cases K_T is not significantly closer to K_C than K measured on random time-scaled topologies of the same taxa (K_R) is to K_C (Supplementary Material: additional figures available

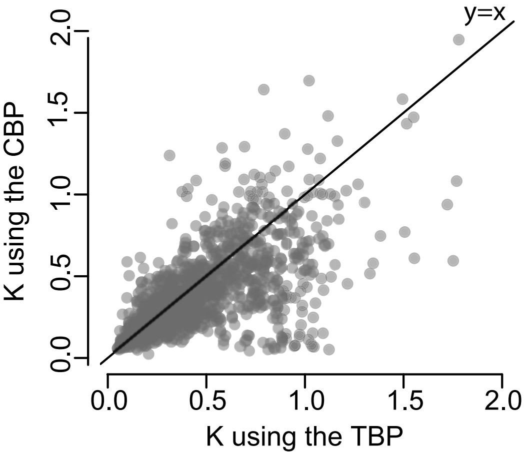


FIGURE 3. Phylogenetic signal. Correlation between estimated values of strength of phylogenetic signal on CBP and TBP.

on Dryad at http://dx.doi.org/10.5061/dryad.pt3qm). This is true for all strengths of phylogenetic signal in the originally simulated data, although for data that are strongly phylogenetically structured the test statistic is significant more frequently, particularly for large trees. These high *p*-values are seen when K_R is also a good predictor of K_C. This issue is explained fully in the discussion but can occur because all topologies have been time-scaled using the same range data.

Phylogenetic generalized least squares.—Phylogenetic regression slope estimates can be obtained from PGLSs analysis of trait data on the CBP and on the TBP. When CBP and TBP PGLS slope estimates are plotted against one another points lying along the line $y=x$ would indicate that results from the TBP are not misleading compared with results from the CBP (Fig. 4a). Points generally lie close to the $y=x$ line but the major axis regression slope is 0.7796, meaning that TBPs tend to produce phylogenetic regression slope estimates that are larger in magnitude than estimates from CBPs (Fig. 4a). Practically this indicates that on average, taxonomy-based analyses overestimate the amount of co-variation between traits, as would be expected given their comparatively poor resolution (Felsenstein 1985). There is a significant coefficient of determination between the two estimates ($R^2 = 0.6949$, $P \ll 0.0001$). Phylogenetic regression slope estimates using TBPs are not always significantly better than regression slope estimates using time-scaled random topologies (Fig. 4b).

Mode of evolution.—An Akaike weight of greater than 0.5 implies relatively greater support for that model. The weight for a particular model can be considered an estimate of the probability of that model being the best model for the data, out of the range of models tested (Burnham and Anderson 2002; Bapst 2014). There was a broader range of model support values using the TBP than on the CBP for many data sets when the trait data were simulated under BM (Fig. 5a). When trait data were simulated under an OU model, the range in model support values was generally the same as or smaller on the TBP than for the CBP (Fig. 5b,c). When BM was simulated, analyses on the TBP had a tendency to incorrectly support an OU model whereas the CBP did not; this was common for larger clades and a reason for it is outlined in the discussion (Silvestro et al. Forthcoming 2015). Note that as the models are nested there is a maximum possible Akaike weight for BM which depends on the number of taxa; this is clear from Figure 5a. There were some clades where the result was directly misleading, and the model used to simulate the trait data was the best supported model on the CBP but not the TBP. Examples, however, were rare (Fig. 5). When OU had been used to simulate the data, number of taxa in the data set was an important indicator

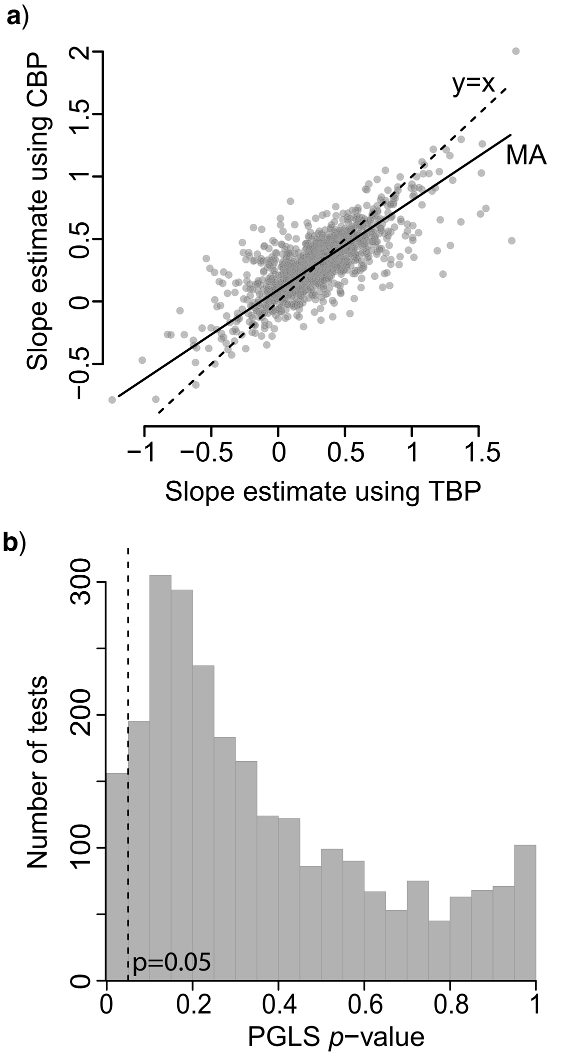


FIGURE 4. Phylogenetic generalized least squares. a) Correlation between estimated values of slope for PGLS on the TBP and CBP showing the major axis regression line (MA). b) *p*-values for the difference between the slope estimate using the TBP and using the CBP. These are collective results for all iterations of the test for each clade.

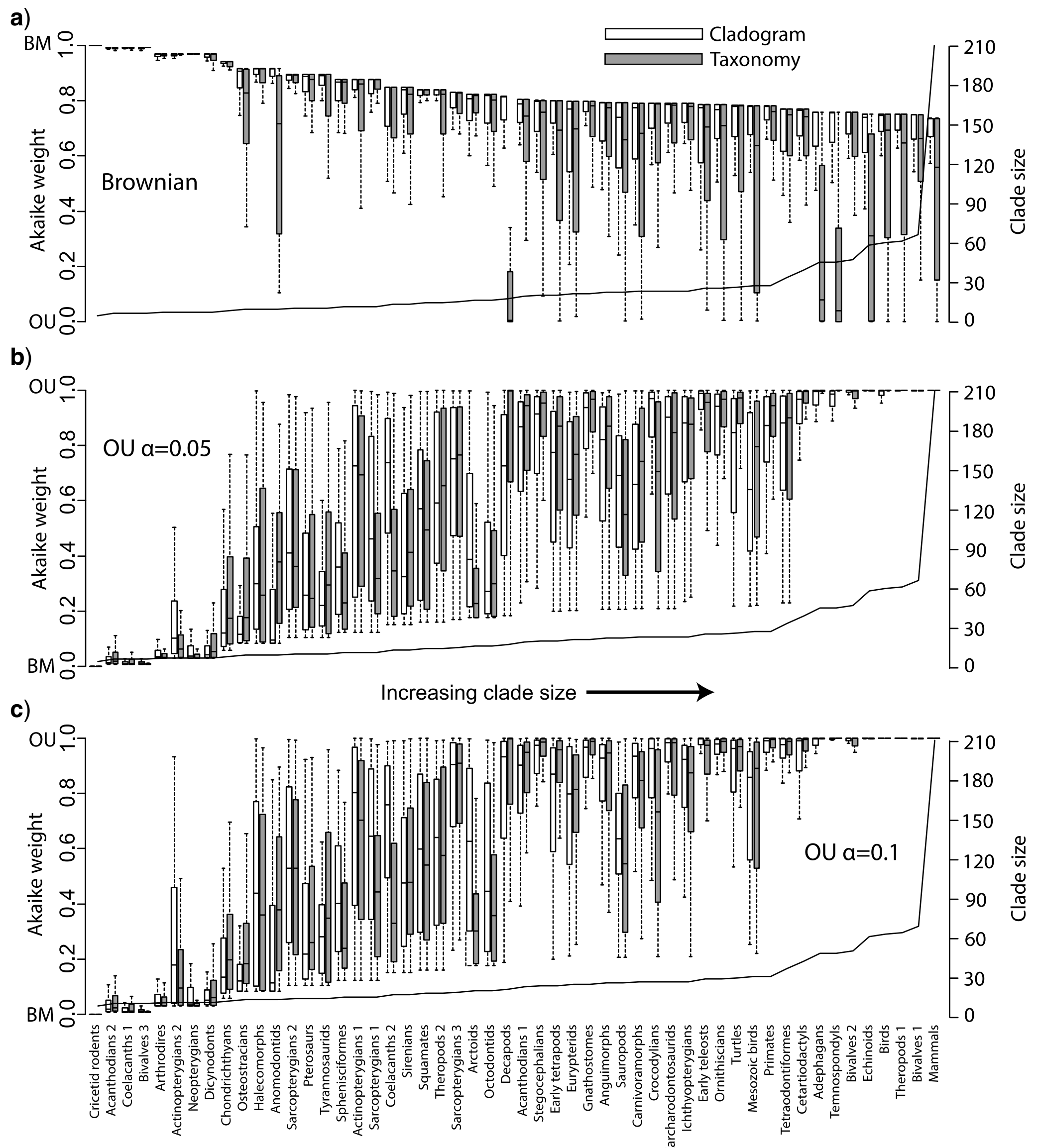


FIGURE 5. Mode of evolution. Results from the CBP are shown as white bars, next to the results using the TBP for the same clade shown as gray bars. Size of each clade is plotted as thin black line. a) Model support (measured as Akaike weight) for a BM model compared with an OU model, when traits were simulated using BM with $\sigma = 1$. b) Model support for an OU model compared with a BM model when traits were simulated using OU with $\alpha = 0.05$ and $\sigma = 1$. c) Model support for an OU model compared with a BM model when traits were simulated using OU with $\alpha = 0.1$ and $\sigma = 1$.

of correct model identification, with OU consistently correctly identified on the CBP and TBP for clades including large numbers of taxa, but not for small clades where OU was rarely correctly identified on either tree (Fig. 5b,c). There was a minor increase in the range of AICc weights estimated for the OU model with a smaller value of $\alpha = 0.05$, but it made no difference to the comparative performance of the TBP and CBP. Results from χ^2 tests do not differ qualitatively from the AICc results (Supplementary Material: Results, available on Dryad at http://dx.doi.org/10.5061/dryad.pt3qm). In the analysis where α varied according to tree height, results were similar to OU results in the main analysis. However, there were some large clades where the TBP performed poorly compared with the CBP (Supplementary Material: additional figures, available on Dryad at http://dx.doi.org/10.5061/dryad.pt3qm). These clades were those that included many extant taxa (e.g., bivalves and modern birds); an explanation of this effect is provided in the discussion.

Estimates of parameters for fitted models on the TBP were similar to—but usually slightly less accurate to original simulation parameter values—than estimates on the CBP, falling just below the line $y = x$. In many cases however, both trees gave parameter estimates that were not close to parameter values that were input to models when we simulated the trait data (Fig. 6). This was particularly a problem for estimates of σ when fitting OU models (these are the points to the right side on Fig. 6). A reason for this is outlined in the discussion.

Vital Statistics.—Some of the vital statistics of sampled phylogenies showed a small (up to $R^2 = 0.3480$) but significant correlation with the degree to which TBPs and CBPs matched for the various analyses (Table 2). However, the best indicators varied between analyses, and no one statistic was consistently good for all the analyses (see Supplementary Material: Results, available on Dryad at http://dx.doi.org/10.5061/dryad.pt3qm, for which vital statistics are the best explanatory variables for each analysis, although in many cases there is little difference between them). The summary results for each of the analyses can be found in Supplementary Material: Results, available on Dryad at http://dx.doi.org/10.5061/dryad.pt3qm (R^2 values for correlations and medians for p-values). All the characteristics we tested can be easily obtained from a TBP. Where there is a good correlation between a summary result and a vital statistic, larger values for vital statistics are associated with a better match between the two tree types. No one vital statistic, or indeed particular combinations of them, can be considered a definitive indicator of how well a taxonomy is likely to perform due to the low R^2 values, but if a taxonomy has high values for several of them it is more likely to produce a phylogenetic hypothesis that closely matches that derived from a formal cladogram, and thus produce similar results in an analysis.

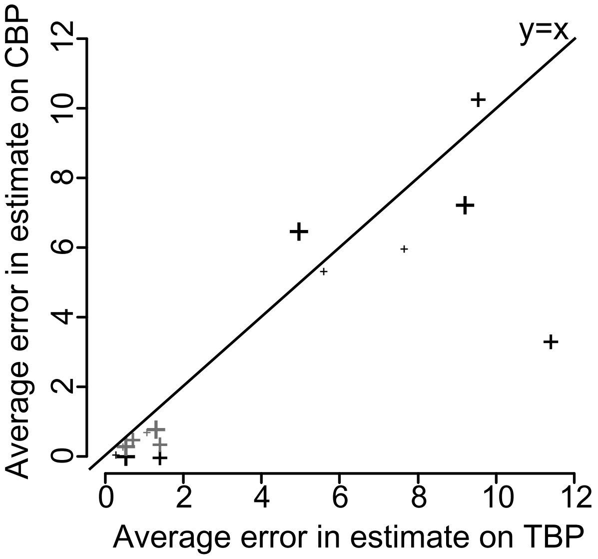


FIGURE 6. Estimated parameter values when fitting an OU or BM model for three of the largest completely extinct clades in the data set (large crosses: theropods; medium crosses: temnospondyls; small crosses: eurypterids). Estimates of α are shown in grey and estimates of σ are shown in black.

Extended range analyses.—When the ranges used were extended to include the whole of each stage in which a taxon occurred, the CBP and TBP yielded comparable results (Supplementary Material: additional figures, available on Dryad at http://dx.doi.org/10.5061/dryad.pt3qm). Results were a close match between the TBP and CBP more often than in the main analyses where occurrence times were selected randomly from within the stage range of each taxon. Variation in K_T explained much of the

TABLE 2. R^2 values from regressions of summary results for each analysis against vital statistics of trees

Summary result	Number	Resolution	SCI	Ranks	Length	Lineages
MC distance	0.0488	−0.0091	0.0513	0.2721	0.0215	0.0378
PGLS R^2	0.1223	0.3091	0.0326	−0.0095	−0.0050	0.0564
PGLS p-value	0.0198	−0.0115	−0.0200	0.0828	−0.0101	0.0201
K weak R^2	0.1350	0.1946	−0.0199	−0.0192	−0.0029	0.0708
K weak p-value	0.0249	0.0425	0.1221	0.1461	0.0617	0.0107
K medium R^2	−0.0196	0.0188	0.0589	0.1249	0.1001	−0.0195
K medium p-value	−0.0194	0.0295	0.1090	0.0601	0.0355	−0.0200
K strong R^2	0.0117	0.0252	0.1241	0.2439	0.1340	−0.0087
K strong p-value	0.1607	0.1192	0.0992	0.1601	−0.0194	0.1360
BM Akaike weight	0.1095	0.0677	−0.0188	0.0223	0.0157	0.0694
OU1 Akaike weight	0.0584	0.0385	0.2303	0.1773	0.0145	0.0276
OU2 Akaike weight	0.0412	0.0767	0.1712	0.1492	0.0147	0.0114
BM σ value R^2	0.1350	0.1946	−0.0199	−0.0192	−0.0029	0.0708
OU1 α value R^2	−0.0200	0.0010	0.0077	−0.0199	0.0404	0.0129
OU1 σ value R^2	−0.0197	0.0062	−0.0045	−0.0200	0.0484	0.0160
OU2 α value R^2	0.0164	0.0524	−0.0194	−0.0172	0.0441	0.0245
OU2 σ value R^2	0.0201	0.1170	−0.0177	−0.0199	0.0523	0.0269

Notes: The vital statistics are outlined in the Methods section. Summary results are summaries of the results for each analysis we performed, outlined in the Results section. Abbreviations and symbols in the summary results are MC, matching cluster; PGLS, phylogenetic generalized least squares; R^2, adjusted coefficient of determination; K, Blomberg's K estimate of strength of phylogenetic signal; weak, medium, and strong indicate the originally simulated phylogenetic signal strength; BM, Brownian motion model; OU1, Ornstein–Uhlenbeck model with $\alpha = 0.1$; OU2, Ornstein–Uhlenbeck model with $\alpha = 0.05$; α, strength of the selective constraint in an OU model; σ, step variance.

variation in K_C ($R^2 = 0.7065$). For PGLS we found a major axis regression slope of 0.9497 ($R^2 = 0.9330$) in a plot of phylogenetic regression slope estimates from the CBP, against slope estimates for the same data on the TBP. This means that TBPs tend to produce phylogenetic regression slope estimates that are slightly larger than estimates from CBPs, but are close to them.

Discussion

Some of the taxonomies tested in this study can be used as a substitute for an explicit phylogenetic hypothesis, to make phylogenetically informed measurements without misleading results being produced, relative to the CBP. Others consistently performed poorly compared with the CBP, and produced results which were misleading when compared with known parameters used to simulate the data. In some cases retrieval of originally simulated values was as poor as or worse on the CBP than on the TBP. Generally speaking, the larger the tree the more consistent (although not always accurate) the results were, but the number of taxa included in the phylogenetic framework did not correlate with any measure of the adequacy of performance of a TBP, nor did any other characteristic of the trees. This result makes it difficult to establish or suggest general circumstances under which taxonomy might be used in PCMs. The same rigorous standards that should be used when deciding whether it is appropriate to incorporate a cladogram produced using formal phylogenetic inference into an analysis must be applied here too, and each case must be assessed on an individual basis.

A context in which taxonomic information is commonly used in the current literature is for supertrees where taxonomy is used as the ‘seed’ or ‘backbone’ for the supertree analysis (Bininda-Emonds et al. 2007; Benson et al. 2014). The majority of supertrees in the paleontological literature are composed exclusively of cladistically inferred phylogenies and any taxonomic relationships that are incorporated are well established relationships between higher level groups. Our results therefore do not suggest that using supertrees in PCMs is any more problematic than using a fully cladistically inferred tree.

Small trees generate misleading estimates using both the CBP and TBP for many of our analyses, suggesting that they require large sample sizes to be effective (Münkemüller et al. 2012). Supertrees are often used to increase taxonomic coverage of phylogeny to include more taxa. The largest TBPs performed well in our analyses, again suggesting that supertrees which incorporate some taxonomic information should not be especially problematic compared with cladistically inferred trees. However, a possible exception relates to estimation of evolutionary mode. Our results indicated that large TBPs were biased toward supporting an OU model when BM had been used to simulate trait data. This makes intuitive sense as inaccuracies in TBP topology compared with the CBP could lead to tips that are sister taxa with similar trait values on the CBP having a greater phylogenetic distance between them on the TBP. Taxa with similar trait values but a larger phylogenetic distance match an OU model, where pull toward an attractor causes this pattern. In this study, bias toward an OU model is associated with topological and branch length error, but the same principle has been demonstrated by other workers for measurement error generally (Thomas et al. 2014; Silvestro et al. 2015). This result may also have been amplified by the poorer resolution of the taxonomies. Although the method required a fully bifurcating tree with no ZLBs, only a short length was added to the internal branches introduced within each TBP. Traits simulated under BM on the CBP would therefore have a greater co-variance than the expectation under BM evolution on the TBPs, where short internal branches lead to low co-variance. In an OU model, change in traits toward an attractor value removes co-variance over time, which matches the expectation for a BM model on a topology with short internal branches. The analysis has less power to detect OU on smaller trees, so this effect occurred less often for the clades with a low sample size.

In the mode of evolution analysis where α varied proportionally to tree height we found that when clades contained many extant taxa, analyses on the TBP were less likely to correctly support an OU model than analyses on the CBP. Taxonomies contain additional polytomies, which in our data sets may contain several extant taxa as well as older extinct taxa, causing the divergence of extant taxa within the polytomy to be dragged back in time. This causes a very ‘tippy’ tree shape where internal branches are short and terminal branches are long. If trait data are simulated via OU then they have comparatively little covariance structure. Trait data with little covariance structure also fits a BM model of evolution on the tippy topology described above, meaning that an OU model does not have additional explanatory power compared with a BM model on these trees.

Some of our results (PGLS slope and estimates of Blomberg’s K) were not statistically significant when compared with an empirical distribution of results from random topologies (Fig. 4b and Supplementary Material: additional figures, available on Dryad at http://dx.doi.org/10.5061/dryad.pt3qm). These statistics are based on change in continuously evolving traits, the values of which are directly related to branch lengths of the phylogeny on which they are simulated. This branch length information (derived from occurrence data) is inherent in every topology we subjected to analysis, making it a key factor in the estimates using both TBPs and random trees. p-values for K and PGLS slope using TBPs were calculated by comparison with an empirical distribution of estimates made using time-scaled random trees. Therefore, when there were good estimates from random trees, TBP p-values could be high and spanned the full range of 0–1 (Fig. 4b). This stands in contrast to analyses conducted with all branch lengths in the random trees set to 1, when the TBP always gave a significantly (P <0.05) closer result to the CBP result than random expectation. As the phylogenetic signal becomes greater, and the tree larger, the difference in topology between the TBP and random trees becomes more important and the TBP is more frequently a significant match for the CBP (Supplementary Material: additional figures, available on Dryad at http://dx.doi.org/10.5061/dryad.pt3qm).

This result, in combination with the results of the extended range analysis, highlights that stratigraphic control on the shared branch length between a CBP and a TBP is an important factor in their relative performance. With our sample of data sets, for analyses when ranges were constrained to begin and end at stage boundaries (resulting in 'time-stretched' terminal branches), the influence of differences in tree topology on the variance–covariance structure was small. In this scenario the majority of trait change occurred on the large shared portions of the trees, which were controlled by the stratigraphy. In the main analyses, differences in tree topology had a larger influence on the variance–covariance structure and so results from the CBP and TBP showed greater difference. In practice it is possible that a high sampling rate (resulting in better constraint on taxon durations) could lead to incorrect branch length extensions when topology is misspecified. However, our results indicate that the advantage of good stratigraphic control on FADs and the ages of sampled populations override the influence of this problem on results.

Many of the clades we tested here have a sparse record, and the TBP did not yield comparable results to the CBP. However, many of the clades for which we do not have cladistically inferred phylogenies have higher sampling rates with better stratigraphic constraint on occurrences and divergence dates (e.g., brachiopods and crinoids [Foote and Sepkoski 1999]). They should therefore be expected to provide results less influenced by inaccuracy in topology.

Bearing all these factors in mind, workers wishing to incorporate taxonomy into a framework for making phylogenetically informed measurements should carefully consider whether the clade in question is in fact suitable for this use. If many new fossils in the clade are still being discovered and are leading to changes in classification, then the taxonomy is unlikely to give results that would match results from a cladistically inferred phylogeny of the group. Stable taxonomies too may be problematic if they have remained stable due to lack of recent rigorous study, rather than because extensive systematic work has led to a high level of confidence in the proposed evolutionary relationships. In addition to consideration of the vital statistics of the taxonomy, consultation with a systematist familiar with the group who can indicate the maturity of the taxonomy would aid in ascertaining its suitability. If the evolutionary relationships proposed within the group are well established and supported by continued study, the ranges are accurately known, and the data set is large, well resolved and stratigraphically consistent then use of a taxonomy can be considered. This supports some of the previous studies where a taxonomic classification has been applied (e.g., Roy et al. 2009; Hardy et al. 2012), since the biological groups in question—ammonites and bivalves—appear to be ideal candidates for such an approach. Here, we used pre-cladistic taxonomies to test a worst-case scenario. Modern taxonomies are often phylogenetically informed, at least in part, and in theory should therefore give a result even closer to the 'correct' one. Taxonomies that include phylogenetically informed relationships should therefore be preferentially used if available.

Each taxonomy should be assessed on a case-by-case basis as it is clear that they are not always suitable for incorporation into a PCM. As a guide to appropriate use of taxonomies in PCMs we provide the following checklist of criteria for candidate classifications:

1. It is large (no less than 20 tips, usually more depending on the method), preferably in combination with a high number of taxonomic ranks, high resolution, and good stratigraphic consistency.
2. It contains long established taxa that have been stable for a significant period of time and are supported by continued study.
3. The fossil taxa are from a range of time intervals, not a small number of key sampling horizons.

If our most recent estimate of phylogeny is closer to the true course of evolution than a taxonomy, it should produce results that are more reliable. We are working under the assumption that cladistically inferred paleontological phylogenies provide estimates of phylogenetically informed measurements that reliably represent reality. If this is the case, our results demonstrate that in some cases taxonomic information can too provide suitable estimates of these measurements.

Although use of taxonomy must be carefully considered, this study opens up certain previously inaccessible sections of the fossil record to interrogation within an explicit comparative framework, which will help to test many classical macroevolutionary hypotheses that have been based on groups for which formal phylogenetic hypotheses remain lacking.

SUPPLEMENTARY MATERIAL

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.pt3qm.

ACKNOWLEDGMENTS

The authors thank Roger Benson and Graeme Lloyd for helpful comments and discussion, and four anonymous reviewers whose suggestions greatly improved the manuscript. This work is Paleobiology Database official publication 226.

FUNDING

This work was supported by the Natural Environment Research Council UK doctoral training grant NE/J500045/1.

REFERENCES

Abouheif E. 1998. Random trees and the comparative method: a cautionary tale. Evolution. 52:1197–1204.

Bapst D.W. 2012. paleotree: an R package for paleontological and phylogenetic analyses of evolution. Methods Ecol. Evol. 3:803–807.

Bapst D.W. 2013. A stochastic rate-calibrated method for time-scaling phylogenies of fossil taxa. Methods Ecol. Evol. 4:724–733.

Bapst D.W. 2014. Assessing the effect of time-scaling methods on phylogeny-based analyses in the fossil record. Paleobiology. 40:331–351.

Beaulieu J.M., O'Meara B. 2014. OUwie: Analysis of evolutionary rates in an OU framework. R package version 1.43. http://CRAN.R-project.org/package=OUwie (last accessed March 30, 2015).

Benson R.B.J., Campione N.E., Carrano M.T., Mannion P.D., Sullivan C., Upchurch P., Evans D.C. 2014. Rates of dinosaur body mass evolution indicate 170 million years of sustained ecological innovation on the avian stem lineage. PLoS Biol. 12: e1001853.

Benson R.B.J., Druckenmiller P.S. 2013. Faunal turnover of marine tetrapods during the Jurassic–Cretaceous transition. Biol. Rev. 89:1–23.

Bininda-Emonds O.R.P., Cardillo M., Jones K.E., MacPhee R.D.E., Beck R.M.D., Grenyer R., Price S.A., Vos R.A., Gittleman J.L., Purvis A. 2007. The delayed rise of present-day mammals. Nature. 446:507–512.

Bogdanowicz D., Giaro K., Wróbel B. 2012. TreeCmp: comparison of trees in polynomial time. Evol. Bioinformatics. 8:475–487.

Bogdanowicz D., Giaro K. 2013. On a matching distance between rooted phylogenetic trees. Int. J. Appl. Math. Comput. Sci. 23:669–684.

Brusatte S.L., Benton M.J., Ruta M., Lloyd G.T. 2008. Superiority, competition, and opportunism in the evolutionary radiation of dinosaurs. Science. 321:1485–1488.

Brusatte S.L., Montanari S., Yi H., Norell M.A. 2011. Phylogenetic corrections for morphological disparity analysis: new methodology and case studies. Palaeobiology. 37:1–22.

Burnham K.P., Anderson D.R. 2002. Model selection and multimodel inference: a practical information-theoretic approach. New York: Springer.

Cope E.D. 1887. The origin of the fittest. New York: Appleton.

Davies T.J., Kraft N.J.B., Salamin N., Wolkovich E.M. 2012. Incompletely resolved phylogenetic trees inflate estimates of phylogenetic conservatism. Ecology. 93:242–247.

Felsenstein J. 1985. Phylogenies and the comparative method. Am. Nat. 125:1–15.

Finarelli J.A. 2010. Does encephalization correlate with life history or metabolic rate in Carnivora? Biol. Lett. 6:350–353.

Flynn J.J., Finarelli J.A., Spaulding M. 2010. Phylogeny of the Carnivora and Carnivoramorpha, and the use of the fossil record to enhance understanding of evolutionary transformations. In: Goswami A., Friscia A., editors. Carnivoran evolution. Cambridge University Press. p. 25–59.

Foote M., Hunter J.P., Janis C.M., Sepkoski J.J. 1999. Evolutionary and preservational constraints on origins of biologic groups: divergence times of eutherian mammals. Science. 283:1310–1314.

Foote M., Raup D.M. 1996. Fossil preservation and the stratigraphic ranges of taxa. Palaeobiology. 22:121–140.

Foote M., Sepkoski J.J. 1999. Absolute measures of the completeness of the fossil record. Nature. 398:415–417.

Friedman M., Brazeau M.D. 2011. Sequences, stratigraphy and scenarios: what can we say about the fossil record of the earliest tetrapods? Proc. R. Soc. B Biol. Sci. 278:432–439.

Fritz S.A., Purvis A. 2010. Selectivity in Mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. Conserv. Biol. 24:1042–1051.

Fusco G., Garland T. Jr., Hunt G., Hughes N.C. 2012. Developmental trait evolution in trilobites. Evolution. 66:314–329.

Garland T., Diaz-Uriarte R. 1998. Effects of branch length errors on the performance of phylogenetically independent contrasts. Syst. Biol. 47:654–672.

Garland T., Diaz-Uriarte R. 1999. Polytomies and phylogenetically independent contrasts: examination of the bounded degrees of freedom approach. Syst. Biol. 48:547–558.

Grafen A. 1989. The phylogenetic regression. Philos. Trans. R. Soc. B Biol. Sci. 326:119–157.

Hardy C., Fara E., Laffont R., Dommergues J., Meister C., Neige P. 2012. Deep-time phylogenetic clustering of extinctions in an evolutionarily dynamic clade (Early Jurassic Ammonites). PLoS One. 7:e37977.

Harmon L.J., Losos J.B., Davies J.T., Gillespie R.G., Gittleman J.L., Jennings B.W., Kozak K.H., McPeek M.A., Moreno-Roark F., Near T.J., Purvis A., Ricklefs R.E., Schluter D., Schulte J.A. II, Seehausen O., Sidlauskas B.L., Torres-Carvajal O., Weir J.T., Mooers A.Ø. 2010. Early bursts of body size and shape evolution are rare in comparative data. Evolution. 64:2385–2396.

Harmon L.J., Schulte J.A., Larson A., Losos J.B. 2003. Tempo and mode of evolutionary radiation in iguanian lizards. Science. 301:961–964.

Harvey P.H., Pagel M.D. 1991. The comparative method in evolutionary biology. Oxford: Oxford University Press.

Housworth E.A., Martins E.P. 2001. Random sampling of constrained phylogenies: conducting phylogenetic analyses when the phylogeny is partially known. Syst. Biol. 50:628–639.

Huelsenbeck J.P. 1994. Comparing the stratigraphic record to estimates of phylogeny. Paleobiology. 20:470–483.

Jablonski D., Finarelli J.A. 2009. Congruence of morphologically-defined genera with molecular phylogenies. Proc. Natl Acad. Sci. USA. 106:8262–8266.

Jablonski D. 2007. Scale and hierarchy in macroevolution. Palaeontology. 50:87–109.

Jablonski D. 2008. Extinction and the spatial dynamics of biodiversity. Proc. Natl Acad. Sci. USA. 105:11528–11535.

Kembel S.W., Cowan P.D., Helmus M.R., Cornwell W.K., Morlon H., Ackerly D.D., Blomberg S.P., Webb C.O. 2010. Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26:1463–1464.

Kroh A., Smith A.B. 2010. The phylogeny and classification of post-Palaeozoic echinoids. J. Syst. Palaeontol. 8:147–212.

Laurin M. 2004. The evolution of body size, Cope's rule and the origin of amniotes. Syst. Biol. 53:594–622.

Lee M.S.Y., Cau A., Naish D., Dyke G.J. 2014. Morphological clocks in paleontology, and a mid-Cretaceous origin of crown Aves. Syst. Biol. 63:442–449.

Lloyd G.T., Friedman M. 2013. A survey of palaeontological sampling biases in fishes based on the Phanerozoic record of Great Britain. Palaeogeogr. Palaeoclimatol. Palaeoecol. 372:5–17.

Lloyd G.T., Wang S.C., Brusatte S.L. 2012. Identifying heterogeneity in rates of morphological evolution: discrete character change in the evolution of lungfish (Sarcopterygii; Dipnoi). Evolution. 66:330–348.

Losos J.B. 1994. An approach to the analysis of comparative data when a phylogeny is unavailable or incomplete. Syst. Biol. 43:117–123.

Martins E.P., Hansen T.F. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. Am. Nat. 149:646–667.

Martins E.P. 1996. Conducting phylogenetic comparative studies when the phylogeny is not known. Evolution. 50:12–22.

Münkemüller T., Lavergne S., Bzeznik B., Dray S., Jombart T., Schiffers K., Thuiller W. 2012. How to measure and test phylogenetic signal. Methods Ecol. Evol. 3:743–756.

Nowak M.D., Smith A.B., Simpson C., Zwickl D.J. 2013. A simple method for estimating informative node age priors for the fossil calibration of molecular divergence time analyses. PLoS One. 8:e66245.

Osborn H.F. 1917. The origin and evolution of life on the theory of action reaction and interaction of energy. New York: Charles Scribner's Sons.

Orme D., Freckleton R.P., Thomas G., Petzoldt T., Fritz S.A. 2012. caper: Comparative Analyses of Phylogenetics and Evolution in R. R package version 0.5.2. http://CRAN.R-project.org/package=caper (last accessed March 30, 2015).

Paradis E., Claude J., Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics. 20: 289–290.

Pennell M.W., Eastman J.M., Slater G.J., Brown J.W., Uyeda J.C., FitzJohn R.G., Alfaro M.E., Harmon L.J. 2014. Geiger V2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. Bioinformatics. 30:2216–2218.

Purvis A., Gittleman J.L., Luh H. 1994. Truth or consequences: effects of phylogenetic accuracy on two comparative methods. J. Theor. Biol. 167:293–300.

Pyron R.A. 2011. Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia. Syst. Biol. 60: 466–481.

R Core Team. 2013. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria.

Raup D.M., Sepkoski J.J. 1984. Periodicity of extinctions in the geologic past. Proc. Natl Acad. Sci. USA. 81:801–805.

Robinson D.F., Foulds L.R. 1981. Comparison of phylogenetic trees. Math. Biosci. 53:131–141.

Romer A.S. 1945. Vertebrate paleontology. Chicago: University of Chicago Press.

Ronquist F., Klopfstein S., Vilhelmsen L., Schulmeister S., Murray D.L., Rasnitsyn A.P. 2012. A total-evidence approach to dating with fossils, applied to the early radiation of the hymenoptera. Syst. Biol. 61:973–99.

Roy K., Hunt G., Jablonski D. 2009. Phylogenetic conservatism of extinctions in marine bivalves. Science. 325:733–737.

Sallan L.C., Friedman M. 2012. Heads or tails: staged diversification in vertebrate evolutionary radiations. Proc. R. Soc. B Biol. Sci. 279:2025–2032.

Sansom R.S., Wills M.A. 2013. Fossilization causes organisms to appear erroneously primitive by distorting evolutionary trees. Sci. Rep. 3:2545.

Sepkoski J.J., Bambach R.K., Raup D.M., Valentine J.W. 1981. Phanerozoic marine diversity and the fossil record. Nature. 293: 435–437.

Silvestro D., Kostikova A., Litsios G., Pearman P.B., Salamin N. 2015. Measurement errors should always be incorporated in phylogenetic comparative analysis. Methods Ecol. Evol. 6:340–346.

Simpson G.G. 1944. Tempo and mode in evolution. New York: Columbia University Press.

Slater G.J., Harmon L.J. 2013. Unifying fossils and phylogenies for comparative analyses of diversification and trait evolution. Methods Ecol. Evol. 4:699–702.

Slater G.J., Van Valkenburgh B. 2008. Long in the tooth: evolution of sabertooth cat cranial shape. Palaeobiology. 34:403–419.

Smith N., Turner A. 2005. Morphology's role in phylogeny reconstruction: perspectives from paleontology. Syst. Biol. 54: 166–173.

Soul L.C., Benson R.B.J., Weisbecker V. 2013. Multiple regression modeling for estimating endocranial volume in extinct Mammalia. Paleobiology. 39:149–162.

Stone E.A. 2011. Why the phylogenetic regression appears robust to tree misspecification. Syst. Biol. 60:245–260.

Symonds M.R.E. 2002. The effects of topological inaccuracy in evolutionary trees on the phylogenetic comparative method of independent contrasts. Syst. Biol. 51:541–553.

Thomas G.H., Cooper N., Venditti C., Meade A., Freckleton R.P. 2014. Bias and measurement error in comparative analyses: a case study with the Ornstein Uhlenbeck model. bioRxiv.

Turvey S.T., Fritz S.A. 2011. The ghosts of mammals past: biological and geographical patterns of global mammalian extinction across the Holocene. Philos. Trans. R. Soc. Lond. B Biol. Sci. 366:2564–2576.

Wagner P.J., Marcot J.D. 2013. Modelling distributions of fossil sampling rates over time, space and taxa: assessment and implications for macroevolutionary studies. Methods Ecol. Evol. 4:703–713.

Waller T.R. 2006. Phylogeny of families in the Pectinoidea (Mollusca: Bivalvia): importance of the fossil record. Zool. J. Linn. Soc. 148:313–342.

Xin X., Li X. 2014. Sensitivity analysis of bipartition dissimilarity under tree rearrangement operations. Proceedings of the International Conference on Logistics Engineering, Management and Computer Science. Atlantis Press.

Appendix B

Chapter 3 timeslices

Tables of dates for the beginning and end of timeslices used in the extinction clustering analyses in Chapter 3. There are three tables, corresponding to the timeslices for the main analysis which were combinations of stages, the 15Ma timeslices and the 10Ma timeslices. All the dates are in million years before present.

Corresponding to stages - main analysis

Timeslice	Start	End
1 Vis-Serp	346.7	323.2
2 Bash-Mosc	323.2	307
3 Kasi-Gzhe	307	298.9
4 Asse-Sakm	298.9	290.1
5 Arti-Kung	290.1	272.3
6 Guadalupian	272.3	259.8
7 Lopingian	259.8	252.17
8 Lower-Triassic	252.17	247.2
9 Middle-Triassic	247.2	237
10 Carnian	237	227
11 Norian	227	208.5
12 Rhaetian	208.5	201.3
13 Hett-Sine	201.3	190.8
14 Plie-Toar	190.8	174.1

15Ma timeslices

Timeslice	Start	End
slice1	327.17	312.17
slice2	312.17	297.17
slice3	297.17	282.17
slice4	282.17	267.17
slice5	267.17	252.17
slice6	252.17	237.17
slice7	237.17	222.17
slice8	222.17	207.17
slice9	207.17	192.17
slice10	192.17	177.17

10Ma timeslices

Timeslice	Start	End
slice1	322.17	312.17
slice2	312.17	302.17
slice3	302.17	292.17
slice4	292.17	282.17
slice5	282.17	272.17
slice6	272.17	262.17
slice7	262.17	252.17
slice8	252.17	242.17
slice9	242.17	232.17
slice10	232.17	222.17
slice11	222.17	212.17
slice12	212.17	202.17
slice13	202.17	192.17
slice14	192.17	182.17
slice15	182.17	172.17
slice16	172.17	162.17

Appendix C

Chapter 5 data

Dataset of measurements of sauropterygian body plan and skull. The following data tables show the species names, museum location and specimen number for all the specimens measured for these analyses and all of the measurements made. ‘Uncat’ means the specimen has not yet been catalogued. There are 5 trait values that were interpolated from measurements from conspecific or congeneric specimens, these are shown in bold. Vertebral counts are given in whole numbers, all other measurements are in cm. Neck and Trunk are the overall lengths of these body parts. Cerv=cervical, Dors=dorsal, Pec=pectoral, Sac=sacral all correspond to vertebral counts. Hum=humerus, Rad=radius, Fem=femur, Tib=tibia, Skull=cranium and Isch=ischium. The full dataset including measurements that were not used in these analyses and all the metadata for each taxon can be found on the data appendix disk.

Taxon	Museum	Specimen	Neck	Trunk	Cervical	Dorsal	Humerus	Radius	Femur	Skull	Ischium
Albertonectes vanderveldi	TMP	2007.011.0001	700	240	76	18	40	10.5	38	NA	NA
Anarosaurus heterodontus	SIPG	R594	8.35	NA	15	NA	3	NA	NA	3.65	NA
Anarosaurus pumilio	SMNS	59073	NA	NA	19	26	2.9	NA	3.5	3.4	2
Anningasaura lymense	NHMUK	OR49202	NA	NA	NA	NA	NA	NA	NA	29	NA
Archaeonectrus rostratus	NHMUK	38525	90	121	24	20	24.2	7.9	24.1	51.1	**19.8**
Attenborosaurus conybeari	NHMUK	R1339	181	NA	38	21	37.5	14	36	46	19.5
Augustasaurus hagdorni	FMNH	PR 1974	105	NA	38	NA	16.6	9.3	NA	33	NA
Avalonnectes arturi	NHMUK	14550	61	66.5	26	19	NA	NA	20.6	NA	NA
Bobosaurus forojuliensis	MFSN	27285	89	109	20	27	26	NA	NA	NA	12.5
Borealonectes russelli	CMN	40729	NA	NA	28	NA	34.4	8.2	NA	41	NA
Brancasaurus brancai	GWWU	A3.B4	138	99	37	22	20.2	4.5	20	22.7	13.8
Callawayasaurus colombiensis	UCMP	38349	NA	NA	56	25	44.9	17.2	33	36.9	NA
Ceresiosaurus calcagnii	GPIT	01889	50.5	84.5	26	NA	18.5	9	12.5	19.2	7
Colymbosaurus megadeirus	CAMSM	J29596/59743	215	NA	41	NA	NA	NA	42	NA	NA
Colymbosaurus megadeirus	NHMUK	31787	200	170	41	17	55.2	NA	NA	NA	NA
Colymbosaurus megadeirus	CAMSM	J63919	234	165	NA	17	NA	NA	NA	NA	NA
Corosaurus alocovensis	UW	5485	25	56	18	23	9	5.5	13.5	10.5	NA
Cryptocleidus eurymerus	GPIT	03183	112	144	32	25	32	6.5	36	28	19
Cryptoclidus eurymerus	NHMUK	R2860	104	NA	32	25	27.5	7.7	26.3	27	19
Dactylosaurus gracilis	GPIT	01867	7.7	6.2	NA	NA	2.5	1.2	NA	2.2	NA
Dactylosaurus schroederi	GPIT	Uncat	5.1	8.9	17	19	2.1	1.1	NA	NA	NA
Djupedalia engeri	PMO	216.839	200	NA	54	NA	30	NA	37.5	NA	NA
Dolichorhynchops osborni	FHSM	VP404	70	NA	21	25	34.5	NA	34	52	40
Dolichorhynchops osborni	KUVP	1300	62.5	102	21	25	21.2	3.4	27.5	58	28.2
Dolichorhynchops tropicensis	MNA	V10046	68.5	**122**	21	26	26	5.3	26	56.7	28.8
Edgarosaurus muddi	MOR	751	146	NA	26	9+	NA	4.8	NA	47.5	NA
Elasmosaurus platyurus	ANSP	10081	667	270	71	23	NA	NA	NA	NA	NA
Eoplesiosaurus antiquior	TTNCM	8348	135	96	38	NA	21.9	7.9	22.1	NA	14.4
Hauffiosaurus longirostris	MCZ	1033	178	203	33	25	37	13	NA	68	NA

Taxon	Museum	Specimen	Neck	Trunk	Cervical	Dorsal	Humerus	Radius	Femur	Skull	Ischium
Hauffiosaurus tomistomimus	MMUM	LL 8004	152	148	34	21	40	13.5	38	44	14
Hauffiosaurus zanoni	HAUF	Uncat	93	105	30	NA	31.4	9.2	33.9	38.6	13.7
Henodus chelyops	GPIT	07290	17	56	6	13	10.5	5.1	10.8	12.5	NA
Hydralmosaurus serpentius	AMNH	5835	NA	NA	63	19	NA	NA	44.7	40.4	30.7
Hydrotherosaurus alexandrae	UCMP	33912	471	164	60	19	41.5	10.3	37.6	33	31
Kaiwhekea katiki	OU	12649	308	198	53	23	NA	NA	43	47.4	NA
Keichosaurus hui	SMNS	81682	6.2	6.5	26	19	2.1	1.6	2.2	2.8	NA
Keichosaurus hui	GXD	7613	8.8	7.9	25	19	2.5	1.3	1.8	NA	NA
Kronosaurus queenslandicus	MCZ	1285	130	571	12	32	NA	NA	106	NA	103
Lariosaurus balsami	GPIT	01887	18	32	20	23	6	3.1	7.3	8	NA
Lariosaurus balsami	GPIT	01861	6.5	7.7	21	NA	2	1.1	2.5	4	NA
Lariosaurus balsami	SMNS	81693	16	15.7	24	20	3.7	NA	4.1	5.3	NA
Lariosaurus balsami	NHMUK	R2736	12.5	7.5	NA	NA	2	1.1	2.4	3.9	NA
Lariosaurus hongguoensis	GPIT	Uncat	NA	NA	NA	NA	3.1	1.8	NA	7.8	NA
Libonectes atlasense	SMNK	3978	400	175	53	21	NA	NA	NA	38.5	NA
Libonectes morgani	SMUSMP	69210	562	NA	62	NA	NA	NA	NA	42.4	NA
Liopleurodon ferox	GPIT	03184	92.5	206.5	21	26	39.5	6.5	51.5	89	54
Macroplata tenuiceps	NHMUK	R5488	115	161	27	25	33	12.5	34.5	54.5	33
Manemergus anguirostris	SMNK	3861	61	74	25	19	NA	NA	13.4	32	8.8
Maresaurus coccai	MOZ	4386	NA	NA	NA	NA	NA	NA	NA	89.9	NA
Megacephalosaurus eulerti	FHSM	VP-321	NA	NA	NA	NA	NA	NA	NA	154	NA
Meyerasaurus victor	SMNS	12478	86	128	27	31	42	13.8	39	42	21
Microcleidus brachypterygius	SMNS	51143	142	71	36	18	23	NA	22	NA	10.5
Microcleidus brachypterygius	SMNS	51141	85	52	37	22	10	NA	12	16	NA
Microcleidus homalospondylus	MMUM	LL 7135	251	164	40	24	39	15	NA	NA	21
Microcleidus homalospondylus	NHMUK	36184	210	123	40	22	33.7	**13**	34.5	22.7	**15.8**
Microcleidus indet.	SMNS	51945	136.5	81.5	42+	20	15	3.8	17.3	NA	NA
Microcleidus tournemirensis	MMM	J.T.86-100	186	90	43	19	20.5	6.9	21.8	19	17.8
Muraenosaurus leedsii	NHMUK	R2421	NA	NA	44	22	28.3	9.1	26.3	NA	8.1

Taxon	Museum	Specimen	Neck	Trunk	Cervical	Dorsal	Humerus	Radius	Femur	Skull	Ischium
Muraenosaurus leedsii	NHMUK	R.2863	195	NA	41	26	30.5	NA	29.5	NA	20
Neusticosaurus edwardsii	PIMUZ	3460	12.7	36.8	17	20	8	4.3	4.9	NA	NA
Neusticosaurus peyeri	PIMUZ	T3615	3.6	5.2	17	19	1.1	0.67	1.17	1.9	NA
Neusticosaurus pusillus	GPIT	01241	6	10.5	17	23	2.6	1.2	2.3	NA	NA
Neusticosaurus pusillus	NHMUK	R5737	4.1	8.6	19	21	1.5	0.8	1.6	1.8	NA
Neusticosaurus pygmaeus	SMNS	53932	3.8	8.5	16	24	1.3	NA	1.4	2.4	NA
Nichollssaura borealis	TMP	94.122.01	69	105	24	24	19.8	3.7	22.7	30.7	NA
Nothosaurus giganteus	PIMUZ	T4829	69	135	19	24	23.5	11.7	25.6	48.6	16.1
Nothosaurus mirabalis	SMNS	Uncat	NA	NA	NA	NA	NA	NA	NA	41	NA
Nothosaurus raabi	SMNS	331	21.5	27.8	17	20	6.5	3.3	7.5	NA	4.5
Pachypleurosaurus sp	SMNS	81781	7	7.7	26	19	2.3	1	1.8	2.1	NA
Peloneustes philarchus	GPIT	03182	74	164	19	24	40	8.5	45	74	33
Peloneustes philarchus	SMNS	10113	NA	NA	NA	NA	NA	NA	NA	62	NA
Peloneustes philarchus	CAMSM	J46913	88	127	18	21	37	8	43	NA	NA
Picrocleidus sp	LEICT	G18.1996	170.5	103	44	17	26.5	6.8	25.2	NA	NA
Pistosaurus longaevus	SMF	R4041	NA	NA	23	22	16.6	11	17.8	NA	NA
Pistosaurus strunzi	GPIT	01886	78	NA	22	22	NA	NA	NA	NA	NA
Placodus gigas	SMNS	55751	12.5	85.5	7	19	14.5	7	15	15	7.8
Plesiopterys wildi	SMNS	16812	88.5	70	40	23	15	5.5	15.8	10.9	8.2
Microcleidus brachypterygius	GPIT	03185	115	82	33	NA	21	7.5	20	15	NA
Plesiosaurus dolichodeirus	NHMUK	22656	115	71	41	21	18.3	7.2	17.4	19.5	11.5
Plesiosaurus sp.	SMNS	59653	NA	39	NA	20	NA	NA	10	NA	6.5
Pliosaurus brachyspondylus	CAMSM	35991	NA	NA	19	26	40	NA	45.5	NA	NA
Polycotylus latipinnis	YPM	1125	112	NA	26	28	42	NA	47	NA	63
Psephoderma alpinum	MCSNM	V527	4.7	27.1	5	15	41.6	NA	40.4	NA	NA
Rhomaleosaurus cramptoni	NMING	F8785	187	225	28	27	53	16	53.4	93	**22**
Rhomaleosaurus megacephalus	LEICT	G221.1851	139	202	28	23	35.3	15.2	39.7	61.5	**27.8**
Rhomaleosaurus thorntoni	NHMUK	R4853	NA	275	NA	25	72.5	NA	67.2	90	33.9
Rhomaleosaurus zetlandicus	YORYM	G503	120	199	20	30	51	NA	50.8	70.5	NA

Taxon	Museum	Specimen	Neck	Trunk	Cervical	Dorsal	Humerus	Radius	Femur	Skull	Ischium
Richmond pliosoaur	QMF	1609	138	133	22	19	40	8.9	40.7	89	NA
Seelyosaurus guilelmiimperatoris	SMNS	51015	126	108	33	21	27	10.5	26	17	12
Serpianosaurus mirigiolensis	PIMUZ	3931	7.6	14.8	14	24	3.6	2.10	3.5	5.9	NA
Simolestes vorax	NHMUK	R3319	134	NA	22	30	42.6	14.2	53.5	NA	59.9
Simosaurus ”juvenilus”	GPIT	01888	62	99	NA	NA	18	NA	18.5	16.5	9.5
Simosaurus gaillardoti	SMNS	14733/10360	49	85	31	32	26.2	12.8	20.3	29	12
Styxosaurus snowii	SDSM	451	600	250	62	22	38	12.5	36.5	37.5	53.7
Thalassiodracon hawkinsii	NHMUK	2018	64	53	31	NA	14.9	5.3	14.2	13	9.90
Thalassiodracon hawkinsii	NHMUK	2020	68	63.6	30	NA	13.4	3.9	15	14.5	12
Thililua longicollis	MHNGr	11710	34.9	NA	30	NA	NA	NA	NA	66	NA
Tricleidus seeleyi	NHMUK	R3539	85	NA	27	NA	21.4	3.9	22.2	26	25
Wangasaurus brevirostris	GMPKU	P1529	87.7	59.8	33	NA	16.8	9.2	16.4	20.3	NA
Westphaliasaurus simonsensii	GWWU	P58091	150	138	23	21	22	7.8	26.4	NA	NA
Yunguisaurus liae	NMNS	4529/F003862	73.2	62	44	28	9.9	5.1	9.6	11.9	NA

Appendix D

Data appendix disk

Contents of the data appendix disk sorted by chapter, showing the file name and contents of each file on the disk. With the files provided here, any one of the analyses within this thesis can be repeated by anyone with access to R. The disk is located on the inside back cover of the thesis.

Chapter 2

- Chapter2Code.txt - Script for all of the analyses performed in R.
- Datasets.zip - File containing newick files for each of the phylogenies and text tables for each of the taxonomies. Includes a key to these files as a spreadsheet.
- tax2tree.txt - Script specifically for the function used to convert a table of taxonomic classifications to a cladogram.
- Results.xls - Excel workbook containing the results of all of the analyses, preliminary analyses and extended range analyses

Chapter 3

- TreeSet3.RData - The set of trees used for the analysis, in ape format to be read into R.

- Chapter3Code.txt - Script for all of the analyses performed in R.
- Results.RData - All results for the different timeslices, metrics and clades in matrix format to be read into R.
- Chapter3Ranges.txt - Range data used to timescale all the cladograms

Chapter 4

- Chapter4Code.txt - Script for all of the simulations and analyses performed in R.

Chapter 5

- Chapter5Code.txt - Script for all of the analyses performed in R.
- Plesiosauria.dat - Character matrix for plesiosaurian taxa.
- Data.xls - Full dataset of all the measurements made including metadata for each taxon.
- TreeSet5.RData - The set of trees used for the analysis, in ape format to be read into R.