The Importance of Phylogenetic Model Assessment for Macroevolutionary Inference

David Alejandro Duchêne Garzón

Research School of Biology

Australian National University

January 2016

A thesis submitted for the degree of

Doctor of Philosophy of the Australian National University

Statement of originality

I, David Alejandro Duchêne Garzón, hereby state that the work presented in this thesis is original and my own work, except where due reference is given in the text. I am the principal contributor and the corresponding author of all chapters. The subject used in several chapters is "we" instead of "I" because they are collaborative projects with multiple authors. Author contributions and publication details are included in the title page for each chapter. The formatting differs among chapters in order to meet the requirements of different journals. No part of this thesis has been submitted for any previous degree.

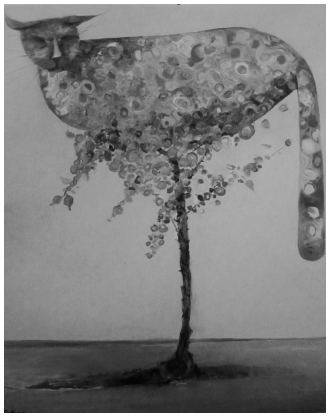
_____

David Alejandro Duchêne Garzón

January 2016

Acknowledgements

Many were involved in the making of this thesis, but five people stand out for teaching me most of what I have learnt. My primary supervisor, Marcel Cardillo, has a remarkable equanimity and positivity, which provided a critical balance to any hurdles I had to overcome. His humbleness is humbling, and he has a balance between compromise and independence towards his work and family that I deeply admire. I hope



I have learnt from these attributes and from his kindness, but I know that I am yet to learn a lot more from him. My co-primary supervisor, Lindell Bromham, is one of the most admirable people I will ever meet. Every moment of her presence is a moment to cherish. Her fire and vitality are the only the scaffold of an historic, even mythical, character that embodies wisdom and power. Lindell and Marcel believed in me and gave me the opportunity to start this degree, for which my gratitude is immense. I hope I never loose contact so I can continue to learn from them, forever. Lindell gave me a million key pieces of advice, but one that should be written here is “never look sideways”. My co-supervisor and mentor, Simon Ho, has continuously given me support and believed in my ideas and style. For a personality of such distinction to have such a gesture has been extraordinary. He is a naturally and effortlessly remarkable, well-rounded individual. I look forward with excitement to the next period of my life learning from him.

My father is completely aware and pragmatic about the fact that he is a nutcase, and he has taught me all the tricks I know to be the same. He has helped build in me a unique and highly functional understanding of life, which I think was crucial to get me where I am. My brother’s insights play a central role in most of this thesis. I think I will never know enough about myself to describe what he has done for me and taught me. He is

also a nutcase. Sadly, my mother was not a protagonist in this chapter of my life. Our relationship and what she would have taught me will keep being a recurrent theme in my dreams.

This brings me to the girls that have helped build me as an adult during this degree. They have encouraged me to live life with love and compassion. The role of women has been critical during my PhD, helping me to become more confident in my work, while encouraging me to be as sweet as I possibly can. I find this was critical to my wellbeing and that of others during my PhD and for the future. These girls have included Bananatime McLean, Crocket Moray, Zo Reynolds, Jess Hopf, Silvia Castro, Laura Aristizabal, Jesse Au, Marta Vidal, the muse Euterpe, Frances Jacomb, Xia, Bo, Lynne Van-Herwerden, Maider.

During this degree I developed and nurtured my love for playing music on the fiddle. I believe that the music I have learnt should be a chapter in this thesis if it were relevant. Euterpe accompanied me every day of this degree, she feeds the thoughts of wanting to make the best of myself, and without her my life is empty. For teaching me and playing music with me I want to thank Simon Kravis, Jeanette Mill, Alan Rumsey, Thomas Rowell, and Hannah Windley.

Thanks to the people at EEG, especially my convener Adrienne Nicotra, and ANU broadly. EEG is an extraordinary department in an extraordinary university. I am extremely proud to have been trained here and to have interacted with such a friendly and academically outstanding group of people. I want to thank people like Edward Holmes, Hannah Kokko, and John Welch for reinforcing in me the idea that individuality characterises beautiful minds.

Last, but not least, thanks to several good blokes that offered a helping hand and friendship in times of need during this degree. These include Daniel Bishop, Scott Hansen, Graham Templeton, Trev, Keaghan, and a bunch of other special fellas at EEG. Cheers.

Table of contents

Abstract

Several fields in biology rely on inference of evolutionary timescales using phylogenetics. As more data become available, estimates of phylogeny and evolutionary timescales can be used to answer long-standing questions in biology. Some examples include the resolution of deep taxonomic relationships or the causes of geographic gradients in species richness. Importantly, answering these questions largely depends on phylogenetic and timescale estimation methods that can reliably recover the molecular evolutionary process. If the methods used in phylogenetics suffer from systematic bias, the inferences that are now routinely made in several fields of biology might be misleading. For this reason, it is critical to identify the processes that can bias phylogenetic inference, and to propose solutions that can be used in practice. In this thesis, I apply empirical tests and simulation analyses to explore the way that molecular processes at the level of DNA sequences link with the inference of phylogeny and evolutionary timescales, and with broad macroevolutionary and macroecological patterns of biodiversity.

In chapter 2 I use a published estimate of the phylogeny of the birds to address one of the major questions in biogeography: the causes of the latitudinal diversity gradient. I find support for the hypothesis that dispersal across latitudes is limited and higher latitudes are likely to contain younger clades, such that they have had less time available to accumulate diversity compared to tropical clades. Chapter 2 provides an example of the use of phylogenetic estimates to test hypotheses in macroevolution. However, macroevolutionary processes themselves might have an effect on the inference of phylogenetic timescales. In chapter 3, I use a data set for the plant family Proteaceae to demonstrate a link between the rate of diversification and the rate of molecular evolution. I find in chapter 4 that this link between the rate of diversification and the

rate of molecular evolution could cause systematically biased estimates of evolutionary timescale. I also find in chapter 5 that phylogenetic imbalance, a phylogenetic pattern that arises from variation in macroevolutionary processes across lineages, can also be a source of systematic bias in estimates of evolutionary timescales. Finally, in chapter 6 I propose a method to assess the absolute performance of phylogenetic methods to estimate evolutionary timescale, as opposed to assessing the relative performance among methods. In this thesis, I show that despite the prominence and progress in methods to estimate phylogeny and evolutionary timescales, there is work to be done towards accounting for the effect of possible sources of bias. Methods to assess absolute model performance might provide a fruitful way forward to improve phylogenetic and evolutionary timescale estimates. A promising approach for the near future is to use the regions in the genome that are reasonably described by the existing models for phylogenetics.

Chapter 1 – General introduction

Parts of this chapter were published as:

Duchêne D, Bromham L, (2013) Molecular Dating of Evolutionary Events, in *Encyclopaedia of Scientific Dating Methods*, Springer Netherlands, Dordrecht. DOI 10.1007/978-94-007-6326-5_18-1

Author contributions:

DD wrote the paper and LB revised drafts of the paper.

Phylogenetics plays a key role in many fields of biology. In the basic framework of modern phylogenetics, evolutionary models are used to infer relationships among organisms (e.g. Felsenstein 1983; Tavaré 1986). These relationships are most often represented in the form of a fully or partly bifurcating tree, sometimes with branch lengths proportional to the amount of time or evolution among samples. Some fields of research now frequently rely on phylogenetic inference, such as taxonomy, comparative analysis among organisms (Felsenstein 1985), macroevolution (Dial & Marzluff 1989; Harvey *et al.* 1994), epidemiology (Drummond *et al.* 2003), and historical biogeography (e.g. Mittelbach *et al.* 2007). One component of phylogenetic inference that is now fundamental to many of these fields of research is molecular dating of evolutionary events. This thesis focuses on three topics surrounding the importance of molecular dating of evolutionary events for the field of macroevolution, which is the study of the processes that lead to the emergence and extinction of taxa through time and space (Dial & Marzluff 1989; Harvey *et al.* 1994): (i) the reliance of macroevolution research on inference of phylogeny and evolutionary timescales; (ii) sources of bias in methods to infer molecular phylogenies and evolutionary timescales,

in particular the bias that might arise from macroevolutionary processes themselves; and (iii) approaches to alleviate bias in estimates of phylogeny and evolutionary timescales.

1.1 The rise of phylogenetics in macroevolution

The structure of phylogenies has been used to make inferences about macroevolutionary processes for many decades. The branching process in phylogenies was initially studied in the field of palaeontology. For example, simulations of branching, extinction, and persistence were used to propose theories of macroevolution under stochastic evolutionary processes (Raup & Gould 1974). With the increasing availability of molecular data, however, evolutionary biologists increasingly used models of molecular evolution to make phylogenetic inferences and study macroevolution (e.g. Nee *et al.* 1992; Pybus & Harvey 2000; Rabosky & Lovette 2008; Etienne *et al.* 2012). The use of molecular phylogenies is now common practice in evolutionary biology and an integral component of analyses of macroevolution.

Analysis of phylogenetic trees shows evidence that some lineages undergo a greater net accumulation of species compared to others (Mooers & Heard 1997). This disparity can be quantified by measuring the extent of phylogenetic imbalance. In its simplest form, the amount of imbalance describes the number of species accumulated in a lineage relative to its sister lineage. Severe imbalance results in a phylogenetic tree with a comb-like or pectinate structure. Phylogenetic imbalance has been shown to be common, indicating that the rates of speciation and extinction vary greatly among taxa (Blum & François 2006). Measuring phylogenetic imbalance can be used to make inferences of macroevolution (Heard 1992; Page 1993; Chan & Moore 2002). For instance, severe phylogenetic imbalance has been used to suggest that purifying

selection is an important force driving extinction, and is important to describe the macroevolution of viruses (Volz *et al.* 2013).

Macroevolutionary events can also be studied using phylogenetic estimates of the ages of bifurcation events (e.g. Pybus & Harvey 2000; Rabosky & Lovette 2008; Morlon *et al.* 2010; Etienne *et al.* 2012). Information about timescale in a phylogenetic tree can be used to describe changes in the rate of accumulation of species. For instance, if the ages of bifurcations are restricted to the beginning of a clade's evolutionary history, such that the more recent branches are long relative to those that are older, this might be interpreted as evidence for an explosive radiation with subsequent slowdown in the rate of accumulation of species (Pybus & Harvey 2000). However, timescale estimates alone are unable to determine the relative importance of speciation and extinction in shaping the observed diversity (e.g. Rabosky 2009). Recent studies have made progress in this regard by assessing the likelihood of highly specific macroevolutionary models, each of which gives varying importance to each of speciation and extinction (Morlon *et al.* 2010; Etienne *et al.* 2012).

Other methods in macroevolution use phylogenetic timescale estimates to test whether specific clades underwent changes in net diversification rates (e.g. Alfaro *et al.* 2009; Rabosky 2014). Clade-specific estimates of net diversification rate can be calculated and compared between clades. This kind of approach has been used, for example, to identify major events of speedup in diversification rates in jawed vertebrates, including the fast radiations of coral reef fishes, birds, lizards, and mammals (Alfaro *et al.* 2009). A more recent approach can further assume that the monophyletic clades that have a particular class of diversification rate might have also undergone gradual changes in diversification rates through time (Rabosky 2014). The methods described so far

provide a brief overview of the substantial amount of work and enthusiasm in the past decades to study macroevolution using phylogenetic estimates.

Chapters 2 to 5 in this thesis address two assumptions that could be made when using phylogenetic timescale estimates to study macroevolution. The first is that speciation and extinction are the primary drivers of species richness. While phylogenetic estimates can provide some insight about macroevolution, it is becoming increasingly apparent that dispersal can be a primary driver of local species richness (e.g. Wiens & Donoghue 2004). Chapter 2 uses phylogenetic inferences to assess the importance of dispersal in driving species richness across latitudes. The second assumption that is common in studies of macroevolution is that the methods for phylogenetic inference provide reliable estimates of divergence times. Critically, research in macroevolution is becoming increasingly reliant on the accurate estimation of evolutionary divergence times using phylogenetics. Meanwhile, there are several potential sources of bias to these estimates that have not been investigated in depth. Some sources of biased estimates of divergence times might arise from macroevolutionary processes themselves, which is the subject of chapters 3, 4, and 5.

1.2 Molecular evolutionary models in phylogenetics

Inferences in macroevolution often rely on molecular phylogenies, which in turn frequently rely on models for describing the molecular evolutionary process. These models are used to optimize or approximate the likelihood of data given a set of parameters, including topology and branch lengths. Misleading inferences from molecular evolutionary models can be described in terms of a trade-off between bias and variance (Wertheim *et al.* 2009). Biased or misleading estimates can occur when a

model is missing important parameters for the inferences of interest. Increased variance or uncertainty can occur when there are parameters that have little power to inform inferences from the data. Chapters 3, 4, and 5 in this thesis assess primarily possible sources of bias in molecular phylogenetics. Bias is of particular interest because it can lead to misleading estimates in downstream analyses, most critically in macroevolution.

Two of the components of phylogenetic inference that can cause bias in estimates of evolutionary divergence times are the model of molecular substitution and the model of rate variation across lineages, also known as the "clock model". This thesis focuses primarily on the possible biases of models of rate variation across lineages. Substitution models are critical to phylogenetic inference, however, and might frequently be subject to bias (Ripplinger & Sullivan 2008). Although not of primary focus, the importance of substitution models is emphasized in chapter 6, and there is a large body of literature that explores their sources of bias (Lemmon & Moriarty 2004; Revell *et al.* 2005), methods for model selection (Posada & Crandall 2001; Huelsenbeck *et al.* 2004; Posada & Buckley 2004; Lanfear *et al.* 2012), methods for assessment of adequacy (Goldman 1993; Bollback 2002; Foster 2004; Rodrigue *et al.* 2009; Brown 2014a), and ways towards improving their accuracy (Yang 1993; Foster 2004; Jayaswal *et al.* 2014).

1.2.1 Models of rate variation across lineages

A basic and popular method to estimate evolutionary divergence times in phylogenetics is to make draws evolutionary divergence times and rates of molecular change to be assigned across lineages. Values of divergence times and molecular rates are optimised or approximated using the phylogenetic likelihood according to a given model of substitution. The simplest model of rates across lineages assumes that data is "clocklike", such that the rate is the same in all lineages and through time (Zuckerkandl

& Pauling 1965). There are several methods, such as likelihood ratio tests, to determine whether variation in rates is statistically significant (Langley & Fitch 1973; Takezaki *et al.* 1995; Brown & Yang 2011), and it is common to select the data that behave in a clocklike fashion for inference (e.g. Jarvis *et al.* 2014). Importantly, while some datasets may approximate uniform rates, most display some departure from clocklike behaviour.

Perhaps the most intuitive way to account for rate variation across lineages is to assume that the rate evolves through the tree (Thorne *et al.* 1998), but several other methods have been proposed. For example, one way to account for rate variation across lineages is to impose multiple rate categories in the same analysis. For example, sections of the data set may have their own "local clock" (Hasegawa *et al.* 1985; Kishino & Hasegawa 1990; Rambaut & Bromham 1998; Yoder & Yang 2000; Drummond & Suchard 2010). Other methods account for rate variation across lineages by "relaxing" the molecular clock constraint on all branches, so every branch in a phylogeny has a different rate (e.g. Drummond *et al.* 2006; Lepage *et al.* 2007; reviewed by Ho & Duchêne 2014).

Some analyses of molecular evolutionary divergence times assume a single value for the rates of molecular evolution across lineages. However, given that rates of molecular evolution vary across the genome and between lineages, it is preferable to use independent temporal information to calibrate rates of molecular evolution (Welch & Bromham 2005). The most common way of calibrating rates is to use the date of one or more divergence events in the phylogeny, inferred from an independent source of information such as from fossils or geological events (Ho & Phillips 2009; Sauquet *et al.* 2012; Ho *et al.* 2015b).

Calibration information may correspond to the earliest possible age of a node in the phylogeny. The dates of divergence events are rarely known with certainty, and the confidence limits on calibrations can be very large. In the case of fossil evidence, part of

the uncertainty comes from the determination of the age of the fossil itself, drawn from stratigraphy and isotopic composition of the fossil (Benton & Donoghue 2007). However, the exact relationship between the fossil taxon and a specific divergence event in the phylogeny is also typically unknown (Sauquet *et al.* 2012). A fossil taxon is unlikely to be identifiable as a member of a particular lineage until some time after the origin of the lineage, when key diagnostic characters have had time to evolve. So fossil dates typically represent minimum ages of lineages: they provide evidence that a lineage must have originated some time before that date (Bromham *et al.* 1999; Ho & Phillips 2009).

1.2.2 Biases to models of rate variation across lineages

There are three primary sources of bias when estimating evolutionary divergence times using molecular phylogenetics: the model of the evolutionary process, the temporal calibration data, and the sampling or form of collection of genetic data. The most basic source of bias in estimates of divergence times arises from failure to account parameters of the evolutionary process that are important for inferences of phylogeny and divergence times (Steel 2005). For example, bias can occur when using an overly simple or incorrect model of molecular evolution across lineages (e.g. Lemmon & Moriarty 2004; Revell *et al.* 2005; Drummond *et al.* 2006; Wertheim *et al.* 2009; Ho *et al.* 2015a). Inferences also largely depend on the quality of temporal calibrations. For instance, a small number of calibrations in nodes that are "shallow" or close to the tips will provide less information about rates and might lead to misleading inferences compared to using a large number of calibrations close to the root (Duchêne *et al.* 2014). The quality of the genetic data can also lead to a bias in estimates of the phylogenetic divergence times. For example, biased estimates can arise from a

substitution model being unable to account for excessive variation in the rate of molecular evolution across lineages (Wertheim *et al.* 2012).

The importance of using appropriate molecular evolutionary models is a primary focus of this thesis. The drivers of the rate of molecular evolution are difficult to predict and model (Welch & Bromham 2005; Bromham 2009), current models of rate variation across lineages frequently generalize the distribution of rates across lineages (e.g. Drummond *et al.* 2006), instead of explicitly describing the processes that drive molecular evolution. This is important because there is a range of processes that have influenced molecular data sets that we are only beginning to understand, and which have a substantial influence on estimates of evolutionary divergence times. If these processes are sources of systematic bias, this might have implications for previous findings in macroevolution using molecular phylogenetic estimates.

1.3 Assessing model robustness in phylogenetic inference

1.3.1 Methods to assess model adequacy in molecular phylogenetics

Phylogenetic statistical methods involve using statistical models that describe molecular evolutionary processes. A desirable model is not one that can capture all the processes that lead to the observed data, and it is not necessarily one that can be used to predict future observations of the data (Holland 2013). Instead, a good model in phylogenetics aims to be simple, while capturing the properties of the data that are relevant to the inferences of interest (Steel 2005). For many fields of biology, including macroevolution, the inferences of interest are often the tree topology and branch lengths in units of time. The aim of a thorough model selection procedure should be to find

whether the currently available models fulfil these properties of simplicity and descriptive power.

The field of phylogenetic model selection has seen major developments in recent years. Methods for Bayesian model selection have increased the accuracy of estimates of the marginal likelihood, which is used as an indicator of model fit. Specifically, practice has drifted away from using the harmonic mean estimator of the marginal likelihood, which is frequently biased (Beerli & Palczewski 2010), to using the more accurate methods of path sampling (Lartillot & Philippe 2006) and stepping stone sampling (Xie *et al.* 2011). In addition, there has been progress towards methods for Bayesian model averaging for models of both substitution (Huelsenbeck *et al.* 2004; Posada & Buckley 2004), and rate variation across lineages (Li & Drummond 2012). Model averaging is a method that can facilitate the process of model selection by assessment of statistical fit. The posterior of an analysis using model averaging includes samples from every model in a set of candidates in proportion to their probability given the data and the prior. Therefore, the inferences using these approaches are considered to be averaged across models.

Methods to assess absolute model performance, or model adequacy, have been widely developed for substitution models, but less progress has been made for assessing models of rate variation across lineages. Methods that assess model adequacy are useful in that they provide insight into the absolute performance of a model; meanwhile, methods for model selection aim to assess a set of candidate models relative to each other. This provides the benefit of falsifiability of all the models in the set. The ability to falsify all the available models has long been considered fundamental for methods in phylogenetic inference (Penny *et al.* 1992); however, the existing methods to assess

substitution models have not become standard practice (Brown 2014b). Chapter 6 of this thesis makes progress towards assessing model adequacy for models of rate variation across lineages.

1.4 Overview

In chapter 2 of this thesis, I explored a novel approach for explaining the drivers of the latitudinal diversity gradient in birds. Multiple studies have tested the hypothesis of a latitudinal gradient in diversification rates, largely using methods described in section 1.1 (e.g. Cardillo 1999; Cardillo *et al.* 2005; Mittelbach *et al.* 2007; Jetz *et al.* 2012). However, multiple studies have suggested that the processes of dispersal and time available for diversification are important drivers of the gradient and have been understudied (Wiens & Donoghue 2004; Pyron & Wiens 2013; Kerkhoff *et al.* 2014). Chapter 2 aims to complement analyses of diversification rates across latitudes by considering the process of dispersal and the time available for diversification in the birds. This can be done by combining components of biogeography and community phylogenetics to study macroevolutionary processes (Kerkhoff *et al.* 2014).

The biogeographic component involves calculating the rate of dispersal events across geographic regions using estimates of ancestral geographic ranges. The component of community phylogenetics involves comparing the summed amount of macroevolutionary history (sum of branch lengths in terms of time for a given group or community), standardised by species richness, among latitudinal bands. A low amount of summed evolutionary history is indicative of recent diversification (Kerkhoff *et al.* 2014), which might be expected if cool temperate environments appeared only recently (Wiens & Donoghue 2004). In chapter 2, I assess the hypotheses that bird taxa have

limited dispersal across latitudes, and that taxa from higher latitudes are younger, such that they have had less time to accumulate diversity (Wiens & Donoghue 2004; Hawkins *et al.* 2006).

Analyses in macroevolution and macroecology such as those in chapter 2 rely on phylogenetic branch lengths being an accurate representation of the distribution of speciation events in time. However, this reliance could be misleading if diversification rates are linked to the rate of molecular evolution. In chapters 3 and 4, I explore this association, whereby lineages with faster rates of diversification have faster rates of molecular evolution. Such an association has been identified in some organisms (Webster *et al.* 2003; Davies *et al.* 2004; Lanfear *et al.* 2010; Eo & DeWoody 2010), but not in others (Webster *et al.* 2003; Goldie *et al.* 2011), such that its universality remains unknown. In chapter 3, I assess the universality of a link between the rate of diversification and the rate of molecular evolution in the Proteaceae, a large family of plants. If the link between the rate of molecular evolution and the rate of diversification is common, then it is critical to assess the possible consequences for estimates of divergence times using current methods.

Specifically, the positive association between the rate of diversification and the rate of molecular evolution can cause an over-representation of molecular substitutions in periods with fast diversification. An accurate estimate of divergence times would ideally recover periods of fast diversification as having short branch lengths, because speciation events occur close together in time. However, an over-representation of substitutions that are not accounted for in the model might lead to misleadingly long branches, and therefore underestimation of diversification rates. In chapter 4, I use

simulation to assess the potential bias on macroevolutionary estimates caused by the link between the rate of diversification and the rate of molecular evolution.

Another pattern that can arise in macroevolution and which might have an effect on evolutionary divergence time estimates is phylogenetic imbalance, described in section 1.1. This pattern is known to be common in a wide range of taxa and arises primarily from clades having different opportunities to have sampled taxa (Blum & François 2006). An imbalanced macroevolutionary process has the property that some clades had a disproportionate probability of surviving to the present compared to others, or some clades have a disproportionate amount of sampled taxa compared to others. In chapter 5, I explore the power of current methods to recover accurate divergence time estimates under conditions of phylogenetic imbalance.

In chapter 6 I focus on a general assessment of the adequacy of models of rate variation across lineages. In recent years, there have been developments to improve the assessment of substitution model adequacy. While several methods have been proposed for assessing particular assumptions made by substitution models (e.g. Goldman 1993; Foster 2004), a recent study proposed a method for directly assessing inferences of topology and branch lengths (Brown 2014a). Furthermore, one study developed an approach to assess the coalescent model (Reid *et al.* 2014), which is another component of Bayesian models used for phylogenetics, independent from models of substitution or rate variation across lineages. Approaches to assess model adequacy in phylogenetics share that they use simulation to describe the expectation from the model (Goldman 1993; Bollback 2002). Inspired by the recent progress to assess model adequacy in phylogenetics, in chapter 6 I propose a method to assess models of rate variation across lineages in a Bayesian framework using simulation.

In my concluding chapter, I synthesise the information gained in this thesis, focusing on the practical implications of the findings and approaches used. I make the argument that genomic data can be filtered using methods that assess model adequacy. One example of a method that can be used for data filtering is the one proposed in chapter 6. In this context, data filtering involves selecting the loci for which available models are adequate. If data can be selected from large data sets such that they do not violate the assumptions made by current phylogenetic models, this might improve the reliability of phylogenetic inferences. Critically, this practice might also improve the reliability of downstream analyses, such as those made in modern studies of macroevolution.

1.4 References

Alfaro ME, Santini F, Brock C *et al.* (2009) Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 13410–13414.

Beerli P, Palczewski M (2010) Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics*, **185**, 313–326.

Benton MJ, Donoghue PCJ (2007) Paleontological evidence to date the tree of life. *Molecular Biology and Evolution*, **24**, 26–53.

Blum M, François O (2006) Which random processes describe the tree of life? A large-scale study of phylogenetic tree imbalance. *Systematic Biology*, **55**, 685–691.

Bollback JP (2002) Bayesian model adequacy and choice in phylogenetics. *Molecular Biology and Evolution*, **19**, 1171–1180.

Bromham L (2009) Why do species vary in their rate of molecular evolution? *Biology Letters*, **5**, 401–404.

Bromham L, Phillips M, Penny D (1999) Growing up with dinosaurs: molecular dates and the mammalian radiation. *Trends in Ecology & Evolution*, **14**, 113–118.

Brown JM (2014a) Detection of implausible phylogenetic inferences using posterior predictive assessment of model fit. *Systematic Biology*, **63**, 334–348.

Brown JM (2014b) Predictive approaches to assessing the fit of evolutionary models. *Systematic Biology*, **63**, 289–292.

Brown RP, Yang Z (2011) Rate variation and estimation of divergence times using strict and relaxed clocks. *BMC Evolutionary Biology*, **11**, 271.

Cardillo M (1999) Latitude and rates of diversification in birds and butterflies. *Proceedings of the Royal Society B: Biological Sciences*, **266**, 1221–1225.

Cardillo M, Orme C, Owens I (2005) Testing for latitudinal bias in diversification rates: An example using new world birds. *Ecology*, **86**, 2278–2287.

Chan KMA, Moore BR (2002) Whole-Tree Methods for Detecting Differential Diversification Rates. *Systematic Biology*, **51**, 855–865.

Davies TJ, Savolainen V, Chase MW, Moat J, Barraclough TG (2004) Environmental energy and evolutionary rates in flowering plants. *Proceedings of the Royal Society B: Biological Sciences*, **271**, 2195–2200.

Dial KP, Marzluff JM (1989) Nonrandom diversification with in taxonomic assemblages. *Systematic Biology*, **38**, 26–37.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biology*, **4**, e88.

Drummond AJ, Pybus OG, Rambaut A, Forsberg R, Rodrigo AG (2003) Measurably evolving populations. *Trends in Ecology & Evolution*, **18**, 481–488.

Drummond AJ, Suchard MA (2010) Bayesian random local clocks, or one rate to rule them all. *BMC Biology*, **8**, 114.

Duchêne S, Lanfear R, Ho SYW (2014) The impact of calibration and clock-model choice on molecular estimates of divergence times. *Molecular Phylogenetics and Evolution*, **78**, 277–289.

Eo SH, DeWoody JA (2010) Evolutionary rates of mitochondrial genomes correspond to diversification rates and to contemporary species richness in birds and reptiles.

Proceedings of the Royal Society B: Biological Sciences, **277**, 3587–3592.

Etienne RS, Haegeman B, Stadler T *et al.* (2012) Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 1300–1309.

Felsenstein J (1983) Statistical inference of phylogenies. *Journal of the Royal Statistical Society Series A*, 246–272.

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

Foster P (2004) Modeling compositional heterogeneity. *Systematic Biology*, **53**, 485–495.

Goldie X, Lanfear R, Bromham L (2011) Diversification and the rate of molecular evolution: no evidence of a link in mammals. *BMC Evolutionary Biology*, **11**, 286.

Goldman N (1993) Simple diagnostic statistical tests of models for DNA substitution. *Journal of Molecular Evolution*, **37**, 650–661.

Harvey PH, May RM, Nee S (1994) New uses for new phylogenies. *Evolution*, **48**, 523–529.

Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, **22**, 160–174.

Hawkins B, Diniz-Filho JAF, Jaramillo C, Soeller S (2006) Post-Eocene climate change, niche conservatism, and the latitudinal diversity gradient of New World birds. *Journal of Biogeography*, **33**, 770–780.

Heard SB (1992) Patterns in tree balance among cladistic, phenetic, and randomly generated phylogenetic trees. *Evolution*, **46**, 1818–1826.

Ho SYW, Duchêne S (2014) Molecular-clock methods for estimating evolutionary rates and timescales. *Molecular Ecology*, **23**, 5947–5965.

Ho SYW, Duchêne S, Duchêne DA (2015a) Simulating and detecting autocorrelation of molecular evolutionary rates among lineages. *Molecular Ecology Resources*, **15**, 688–696.

Ho SYW, Phillips MJ (2009) Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Systematic Biology*, **58**, 367–380.

Ho SYW, Tong KJ, Foster CSP *et al.* (2015b) Biogeographic calibrations for the molecular clock. *Biology Letters*, **11**, 20150194.

Holland BR (2013) The rise of statistical phylogenetics. *Australian & New Zealand Journal of Statistics*, **55**, 205–220.

Huelsenbeck JP, Larget B, Alfaro ME (2004) Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. *Molecular Biology and Evolution*, **21**, 1123–1133.

Jarvis ED, Mirarab S, Aberer AJ *et al.* (2014) Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, **346**, 1320–1331.

Jayaswal V, Wong TKF, Robinson J, Poladian L, Jermiin LS (2014) Mixture models of nucleotide sequence evolution that account for heterogeneity in the substitution process across sites and across lineages. *Systematic Biology*, **63**, 726–742.

Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO (2012) The global diversity of birds in space and time. *Nature*, **491**, 444–448.

Kerkhoff AJ, Moriarty PE, Weiser MD (2014) The latitudinal species richness gradient in New World woody angiosperms is consistent with the tropical conservatism hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 8125–8130.

Kishino H, Hasegawa M (1990) [34] Converting distance to time: Application to human evolution. *Methods in enzymology*, **183**, 550–570.

Lanfear R, Calcott B, Ho SYW, Guindon S (2012) Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, **29**, 1695–1701.

Lanfear R, Ho SYW, Love D, Bromham L (2010) Mutation rate is linked to diversification in birds. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 20423–20428.

Langley CH, Fitch WM (1973) The constancy of evolution: a statistical analysis of a and b haemoglobins, cytochrome c, and fibrinopeptide A. In: *Genetic Structure of Populations*. (ed Mather K), pp. 246–262. University of Hawaii Press, Honolulu.

Lartillot N, Philippe H (2006) Computing Bayes factors using thermodynamic integration. *Systematic Biology*, **55**, 195–207.

Lemmon AR, Moriarty EC (2004) The importance of proper model assumption in bayesian phylogenetics. *Systematic Biology*, **53**, 265–277.

Lepage T, Bryant D, Philippe H, Lartillot N (2007) A general comparison of relaxed molecular clock models. *Molecular Biology and Evolution*, **24**, 2669–2680.

Li WLS, Drummond AJ (2012) Model averaging and Bayes factor calculation of relaxed molecular clocks in Bayesian phylogenetics. *Molecular Biology and Evolution*, **29**, 751–761.

Mittelbach GG, Schemske DW, Cornell H V *et al.* (2007) Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecology Letters*, **10**, 315–331.

Mooers AO, Heard SB (1997) Inferring evolutionary process from phylogenetic tree shape. *Quarterly Review of Biology*, 31–54.

Morlon H, Potts MD, Plotkin JB (2010) Inferring the dynamics of diversification: a coalescent approach. *PLoS Biology*, **8**, 9.

Nee S, Mooers AO, Harvey PH (1992) Tempo and mode of evolution revealed from molecular phylogenies. *Proceedings of the National Academy of Sciences of the United States of America*, **89**, 8322–8326.

Page RDM (1993) ON DESCRIBING THE SHAPE OF ROOTED AND UNROOTED TREES. *Cladistics*, **9**, 93–99.

Penny D, Hendy MD, Steel MA (1992) Progress with methods for constructing evolutionary trees. *Trends in Ecology & Evolution*, **7**, 73–79.

Posada D, Buckley TR (2004) Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Systematic Biology*, **53**, 793–808.

Posada D, Crandall KA (2001) Selecting the best-fit model of nucleotide substitution. *Systematic Biology*, **50**, 580–601.

Pybus OG, Harvey PH (2000) Testing macro-evolutionary models using incomplete molecular phylogenies. *Proceedings of the Royal Society B: Biological Sciences*, **267**, 2267–2272.

Pyron RA, Wiens JJ (2013) Large-scale phylogenetic analyses reveal the causes of high tropical amphibian diversity. *Proceedings of the Royal Society B: Biological Sciences*, **280**, 20131622.

Rabosky DL (2009) Heritability of extinction rates links diversification patterns in molecular phylogenies and fossils. *Systematic Biology*, **58**, 629–6240.

Rabosky DL (2014) Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS One*, **9**, e89543.

Rabosky DL, Lovette IJ (2008) Density-dependent diversification in North American wood warblers. *Proceedings of the Royal Society B: Biological Sciences*, **275**, 2363–2371.

Rambaut A, Bromham L (1998) Estimating divergence dates from molecular sequences. *Molecular Biology and Evolution*, **15**, 442–448.

Raup DM, Gould SJ (1974) Stochastic simulation and evolution of morphology-towards a nomothetic paleontology. *Systematic Biology*, **23**, 305–322.

Reid NM, Hird SM, Brown JM *et al.* (2014) Poor fit to the multispecies coalescent is widely detectable in empirical data. *Systematic Biology*, **63**, 322–333.

Revell L, Harmon L, Glor R (2005) Under-parameterized model of sequence evolution leads to bias in the estimation of diversification rates from molecular phylogenies. *Systematic Biology*, **54**, 973–983.

Ripplinger J, Sullivan J (2008) Does choice in model selection affect maximum likelihood analysis? *Systematic Biology*, **57**, 76–85.

Rodrigue N, Kleinman CL, Philippe H, Lartillot N (2009) Computational methods for evaluating phylogenetic models of coding sequence evolution with dependence between codons. *Molecular Biology and Evolution*, **26**, 1663–1676.

Sauquet H, Ho SYW, Gandolfo MA *et al.* (2012) Testing the impact of calibration on molecular divergence times using a fossil-rich group: the case of Nothofagus (Fagales). *Systematic Biology*, **61**, 289–313.

Steel M (2005) Should phylogenetic models be trying to “fit an elephant”? *Trends in Genetics*, **21**, 307–309.

Takezaki N, Rzhetsky A, Nei M (1995) Phylogenetic test of the molecular clock and linearized trees. *Molecular Biology and Evolution*, **12**, 823–833.

Tavaré S (1986) Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences*, **17**, 57–86.

Thorne JL, Kishino H, Painter IS (1998) Estimating the rate of evolution of the rate of molecular evolution. *Molecular Biology and Evolution*, **15**, 1647–1657.

Volz EM, Koelle K, Bedford T (2013) Viral phylodynamics. *PLoS Computational Biology*, **9**, e1002947.

Webster AJ, Payne RJH, Pagel M (2003) Molecular phylogenies link rates of evolution and speciation. *Science*, **301**, 478.

Welch JJ, Bromham L (2005) Molecular dating when rates vary. *Trends in Ecology & Evolution*, **20**, 320–327.

Wertheim JO, Fourment M, Kosakovsky Pond SL (2012) Inconsistencies in estimating the age of HIV-1 subtypes due to heterotachy. *Molecular Biology and Evolution*, **29**, 451–456.

Wertheim JO, Sanderson MJ, Worobey M, Bjork A (2009) Relaxed molecular clocks,

the bias–variance trade-off, and the quality of phylogenetic inference. *Systematic Biology*, **59**, 1–8.

Wiens JJ, Donoghue MJ (2004) Historical biogeography, ecology and species richness. *Trends in Ecology & Evolution*, **19**, 639–644.

Xie W, Lewis PO, Fan Y, Kuo L, Chen M-H (2011) Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Systematic Biology*, **60**, 150–160.

Yang Z (1993) Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Molecular Biology and Evolution*, **10**, 1396–1401.

Yoder AD, Yang Z (2000) Estimation of primate speciation dates using local molecular clocks. *Molecular Biology and Evolution*, **17**, 1081–1090.

Zuckerkandl E, Pauling L (1965) Evolutionary Divergence and Convergence in Proteins. *Evolving genes and proteins*, **97**, 97–166.

Chapter 2 – Phylogenetic patterns in the geographic distributions of birds support the tropical conservatism hypothesis

A version of this chapter was published as:

Duchêne D, Cardillo M, (2015) Phylogenetic patterns in the geographic distributions of birds support the tropical conservatism hypothesis. *Global Ecology and Biogeography*. **24**, 1261-1268.

Author contributions:

MC devised the initial idea for the project. DD designed the experimental analyses of the study, with later contribution by MC. DD collected and analysed the data. DD wrote the paper; with input from MC.

2.1 Introduction

The latitudinal diversity gradient (LDG) – the decrease in species richness from the equator to the poles – is one of the largest geographic patterns of biodiversity, yet there is limited consensus over the processes that explain it. Although geographic variation in species richness can be predicted from contemporary environmental conditions such as geographic area, primary productivity, topography and energy availability (Jetz & Rahbek 2002; Blackburn & Gaston 2006; Evans *et al.* 2006; Storch *et al.* 2006; Davies *et al.* 2007; Jetz & Fine 2012), an understanding of the mechanisms underlying variation in diversity must be based on evolutionary and biogeographic processes: speciation, extinction and dispersal. The growing amount of phylogenetic and

geographic data available for large groups of organisms provides an opportunity to investigate how the processes that led to present-day species richness vary with latitude. Here, we use a near-complete fossil-calibrated phylogenetic estimate of the birds, together with geographic distribution data, to investigate whether patterns of dispersal across latitudes and phylogenetic clustering of species within latitudes, are consistent with the expectations of two prominent hypotheses for the LDG.

Ultimately, the LDG must be shaped by the three processes that can change species richness: speciation, extinction and dispersal (Mittelbach *et al.* 2007), but the relative importance of these processes is poorly understood. Furthermore, the processes of diversification and dispersal have not received equal attention; numerous studies have investigated how rates of speciation and extinction might vary across latitudes (Jablonski 1993; Buzas *et al.* 2002; Davies *et al.* 2004; Cardillo *et al.* 2005; Hawkins *et al.* 2006; Ricklefs 2006; Wiens 2007; Weir & Schluter 2007; Soria-Carrasco & Castresana 2012; Jetz *et al.* 2012; Jansson *et al.* 2013; Pyron & Wiens 2013; Rolland *et al.* 2014), but relatively few studies have explicitly investigated the patterns of dispersal events across latitudes (Jablonski *et al.* 2006; Jansson *et al.* 2013; Pyron & Wiens 2013; Kerkhoff *et al.* 2014). Nonetheless, describing patterns of dispersal across latitudes could help distinguish two prominent hypotheses to explain the LDG. Under the "out of the tropics" model (OTM), rates of origination (speciation) are highest in tropical regions, and there is a net movement of species from tropical to extra-tropical regions, as tropical clades continually produce descendant lineages that adapt to temperate conditions (Jablonski *et al.* 2006). Most of the support for this model comes from paleontology of marine organisms, but the OTM has been suggested to apply generally (Jansson *et al.* 2013). Support for higher tropical origination as suggested by the OTM has been mixed, with some studies providing support (Jablonski *et al.* 2006, 2013;

Jansson *et al.* 2013), but others failing to do so (Smith *et al.* 2012; Pyron & Wiens 2013; Kerkhoff *et al.* 2014). Under the tropical niche conservatism hypothesis (TCH), high tropical diversity has arisen as a consequence of the tropical origins of most major groups of organisms (Wiens & Donoghue 2004). Wiens & Donoghue (2004) argued that many major taxa originated and began diversifying in tropical environments, which occurred primarily because the climatic conditions that are currently within the tropics extended far into temperate zones until the mid-Tertiary period (Behrensmeyer 1992). Phylogenetic conservatism in adaptations to tropical environments means that dispersal events into temperate zones are rare. Under this scenario, the comparatively high diversity of the tropics is the result of the longer time available for tropical assemblages to diversify.

Hence, the OTM and TCH make distinct predictions regarding the patterns of evolutionary dispersal across latitudes. Under the OTM, dispersal is frequent and asymmetric; tropical to temperate dispersal is frequent, and temperate to tropical dispersal is rare. On the other hand, under the TCH dispersal is generally rare, and there is no expectation of asymmetry in the direction of dispersal. The two models also differ in the expected phylogenetic structure of temperate assemblages (Kerkhoff *et al.* 2014). The OTM proposes that taxa from across the phylogeny can readily disperse out of the tropics, so that temperate assemblages are either phylogenetically overdispersed compared to a null model (i.e., species are less closely-related to one another than expected), or that phylogenetic structure is indistinguishable from a null model. Under the TCH, dispersal events into temperate zones are rare and can give rise to endemic and phylogenetically clustered temperate radiations (i.e., temperate species are more closely-related to each other than expected). Such a pattern of phylogenetic clustering was recently found by Kerkhoff *et al.* (2014) in New World angiosperms, and

interpreted as evidence in support of the TCH.

Although evidence consistent with the TCH is accumulating (Wiens & Donoghue 2004; Hawkins *et al.* 2006, 2007; Wiens *et al.* 2009; Hawkins & DeVries 2009; Buckley *et al.* 2010; Condamine *et al.* 2012; Smith *et al.* 2012; Kerkhoff *et al.* 2014), there has been no reconstruction of dispersal events across latitudes for the global bird fauna to our knowledge, and explicit comparison of the expectations under OTM and TCH. In this study, we make use of a phylogenetic estimate for birds and a database of geographic distributions. We use these data to reconstruct ancestral latitudinal zones and infer the frequency of dispersal events into and out of the tropics. As an additional means of distinguishing the OTM and TCH, we follow the approach by Kerkhoff *et al.* (2014) to analyze the phylogenetic structure of bird assemblages across latitudes.

2.2 Methods

2.2.1 Phylogenetic and geographic data

We downloaded the geographic ranges of 9118 bird species (approximately 90% of known species) including both breeding and winter distributions from Birdlife International (NatureServe 2012). As the phylogenetic estimate we used 100 randomly-selected trees sampled from the stage 2 posterior distribution of phylogenies presented by Jetz *et al.* (2012).

From the geographic range maps we calculated the latitudinal centroid of each bird species. In a recent study, Kerkhoff *et al.* (2014) used the maximum and minimum latitudes of species distributions to calculate an index of "tropicality". We applied a modified form of their index that maintains the distinction between northern and

southern temperate zones, which allowed us to recover the patterns associated with the historical biogeography of bird clades, rather than simply their association with tropical environments. We refer to this measure as the "latitude index" (LI). It is calculated as the proportion of the latitudinal range that falls north of the tropics minus the proportion of the latitudinal range that falls south of the tropics. The result is a continuous variable ranging from -1, for entirely southern species, through 0 for entirely tropical species, to 1 for entirely northern species. We used 23.5° north and south as boundaries for the tropics.

To categorize species into latitudinal zones, we refer to the species with a proportion 25-75% of their latitudinal range either south or north of the tropics as southern subtropical (SST) or northern subtropical (NST) respectively ($-0.75 \leq LI \leq -0.25$ and $0.25 \leq LI \leq 0.75$), species with >75% of their range outside the tropics as either southern temperate (ST) or northern temperate (NT; $-0.75 \geq LI \geq 0.75$), and species with <25% of their range outside the tropics as "tropical" ($-0.25 < LI < 0.25$). Of course, geographic summary measures based solely on latitude will not fully capture the environmental affinities of some species, like high-altitude tropical species. At a global scale, however, we believe that latitude is correlated strongly enough with major climatic variables, such that latitudinal range should be a reasonable proxy of environmental affinity.

2.2.2 Inferring ancestral latitudes

To infer dispersal events across latitudes, we reconstructed ancestral latitudinal zones for each of the internal nodes in the phylogenies. We then produced a matrix of dispersal events between each of the latitudinal zones. We combined the matrices for the 100 phylogenies by calculating mean values for the frequencies of dispersal events. Under the OTM we expect an excess of tropical to temperate events, whereas the

expectation under the TCH is that there is no asymmetry in the direction of dispersal. To test whether dispersal has occurred with similar frequency across latitudinal zones, we used a χ^2 test of matrix homogeneity on the combined matrix of dispersal events. Another expectation under the two models is that dispersal events are common under the OTM, but uncommon under TCH. There is no simple objective way to define "common" and "uncommon" in order to test these predictions statistically, but in general, we expect that if a great majority of phylogenetic branches show no latitudinal shift, it would more strongly support the TCH.

We used two methods to estimate ancestral latitudinal zones. The first method treated latitudinal zones as a categorical variable, using models of geographic range evolution implemented in the R package BioGeoBEARS (Matzke 2013). To account for the uncertainty in the best model of geographic range evolution, we used the results based on the best-fitting model for each of the 100 trees sampled from the posterior. We chose to test two models that describe processes likely to have occurred during the radiation of birds. The simplest model was the DEC model (Ree *et al.* 2005; Ree & Smith 2008), which allows lineages to either remain in their ancestral range, to split from part of their ancestral range, or to evenly split their ancestral range in two. We tested this model against the more complex DEC+J model, which includes an additional parameter for founder-speciation events, meaning that lineages are allowed to undergo migrate to a new region after speciation (Matzke 2014). We used the reconstructed nodal values of LI with the highest estimated probabilities to infer latitudinal transition frequencies.

The second method to estimate ancestral latitudinal zones was based on LI as a continuous trait. We first tested four alternative models for the temporal pattern of evolution of LI using the R package GEIGER v2.0 (Harmon *et al.* 2008). The models

we selected to test included (i) the Ornstein-Uhlenbeck, which is plausible if dispersal away from the ancestral range is penalized by decreased fitness; (ii) Pagel's λ-transformed random walk, which is high when dispersal is very extensive and low when it is minimal, and describes a scenario where northern and southern limits of species distributions drift randomly and non-directionally through time; the (iii) trend model, in which evolution occurs under a Brownian motion process with a trend through time; and (iv) early burst model, which also describes a process in which the rate of evolution varies through time and is plausible if rapid geographic expansion occurs early in a clade's history, as expected under an adaptive radiation scenario. For each of the 100 phylogenies, the model with the lowest AICc score was selected, and all further analyses were performed with that model (Appendix – Chapter 2). To reflect the pattern of LI evolution on the phylogeny, we made the branch lengths to be proportional to the amount of trait change according to the best-fitting model and estimated parameters. This allowed us to calculate the maximum likelihood ancestral estimate of LI for each node in the tree using the sequential re-rooting method (Garland *et al.* 1999; Garland Jr & Ives 2000), implemented in the R package PHYTOOLS v0.4 (Revell 2012). This method re-roots the phylogeny at every node and calculates the phylogenetically independent contrast for the root node (Felsenstein 1985), taking advantage of the fact that this value is the maximum likelihood estimate for that node.

A potential issue with estimating ancestral latitudes is that the great excess of tropical compared to temperate species may bias the reconstructed latitudes towards the tropics. For this reason, we carried out an additional analysis in which we randomly sub-sampled each of the 100 phylogenies such that each of the five LI categories contained the same number of species as the least species-rich category. The least species-rich category was ST with 458 species, so the sub-sampled phylogenies from the posterior

had 2290 species. We repeated the analyses using BioGeoBEARS and PHYTOOLS for each of the sub-sampled phylogenies, and present these results in the Appendix – Chapter 2.

2.2.3 Quantifying phylogenetic clustering within latitudinal zones

To quantify the degree of phylogenetic clustering of bird assemblages within latitudinal bands of 10° width, we used the Net Relatedness Index (NRI; Webb et al., 2002), a standardized measure of the mean patristic distance among pairs of species. We assigned each species to the latitudinal band that contained its centroid latitude, so that each species was only included once in the analysis. We performed this analysis separately for species belonging to the New World (n=3730) and the Old World (n=5205), excluding species that occurred in both regions and entirely marine species. Because larger communities are likely to have a greater mean distance separating species than smaller communities, NRI is standardized with respect to a null assemblage, which we generated by randomizing species identities 999 times among all latitudinal zones, maintaining the observed number of species per zone. This randomization method to obtain a null distribution is also known as the independent swap algorithm (Gotelli 2000). In this way we obtained values of NRI that are proportional to the size of the null assemblage, such that they are comparable across assemblages, and indicate the degree of phylogenetic clustering or dispersion in the context of the entire fauna of the New World and Old World, respectively. NRI values were calculated using functions in the R package PICANTE v1.6 (Kembel *et al.* 2010).

2.3 Results

The reconstructed latitudinal zones for the deepest nodes in the phylogeny of birds,

including the crown node for all birds, are inferred to have a distribution in the southern part of the tropical zone (Figure 2.1). Hence, all the internal and tip-nodes reconstructed with temperate distributions are nested within older tropical clades. There is asymmetry in the distribution of node ages between the northern and southern hemispheres. Diversification into southern temperate regions began around 70-80Mya, while widespread diversification into northern temperate regions did not begin until around 50Mya. The spread into far northern regions is likely to have accelerated during the first half of the Miocene Epoch, less than 20Mya (Figure 2.1b). Results based on phylogenies subsampled to an equal number of species per zone show similar results (Appendix – Chapter 2).

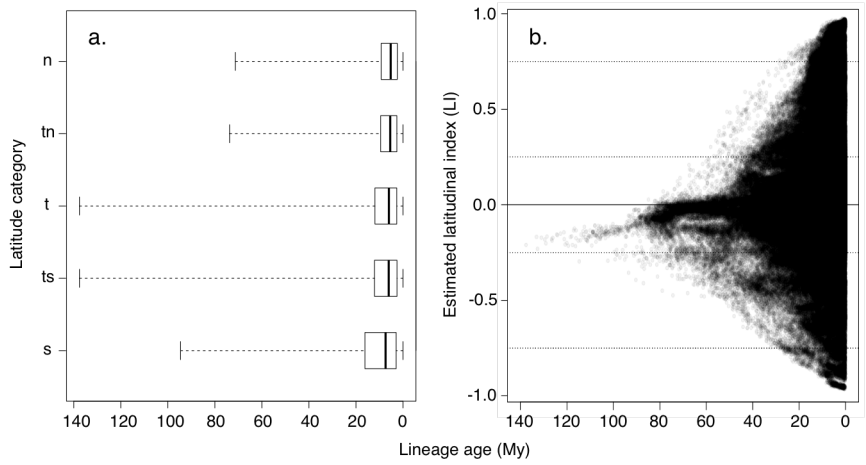


Figure 2.1. Ancestral estimates of LI for nodes across node age estimates (My) of the nodes of 100 posterior trees, made using the best fitting model of (a) geographic models of range evolution and (b) models of continuous-trait evolution.

The matrix of the frequency of evolutionary transitions across latitudes is not homogeneous. This result was consistent when reconstructed ancestral latitudes are based on categorical range evolution models (χ^2 = 26,933.75, df = 16, $p < 0.001$) and on a continuous measure of latitude (χ^2 = 24,011.25, df = 16, $p < 0.001$). Overall, a disproportionately high number of nodes descended from nodes found in the same

latitudinal zone (mean of 67% for analyses based on categorical range models, Figure 2.2a; mean of 79% for analyses based on continuous trait models, Figure 2.2b). Latitudinal conservatism is particularly high in the northern and southern temperate zones, with >90% of nodes from that zone remaining in the same zone, in the continuous trait models (Figure 2.2b). There is also a net movement of species into the tropics from the subtropical regions, which can be observed in the low rates of transition out of the tropics (≤5%). These results were similar when the data were sub-sampled to contain an equal number of species in each latitudinal category. In sub-sampled data, transitions across latitudes were more frequent, with average conservatism of 47% and 76% in analyses using models of discrete geographic and continuous-trait evolution, respectively (Appendix – Chapter 2). A greater rate of transitions in sub-sampled data is to be expected, however, because sub-sampling tips will increase the amount of time between ancestor and descendant.

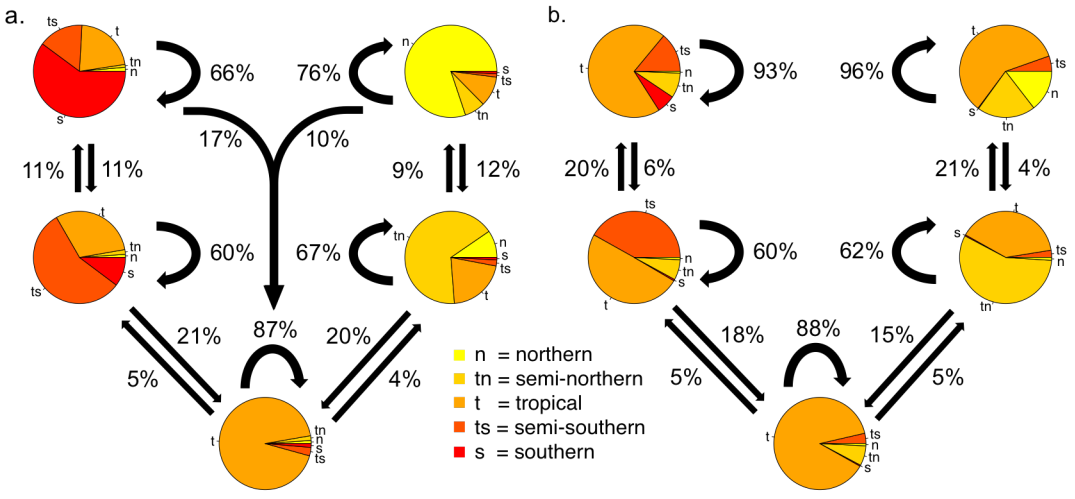


Figure 2.2. Pies show the proportion of phylogenetic branches in a latitudinal band that remained in that region or moved into it from other regions (shown for 100 posterior trees). South is shown on the left, north on the right, and the tropics in the centre. Arrows show the proportion of migration evetns from each given region to another. Values come from ancestral reconstructions using (a) geographic models of range evolution and (b) models of continuous-trait evolution. Transitions percentages below 1% are not shown.

Analyses of phylogenetic structure show that none of the mean values of NRI indicate significant clustering or significant overdispersion (Figure 2.3), with respect to the entire terrestrial avifauna of the New World (Figure 2.3c) or Old World (Figure 2.3d). Nonetheless, the degree of phylogenetic clustering varies considerably among latitudinal zones, with some zones showing a far greater degree of clustering than others. The pattern of phylogenetic clustering among zones does not vary systematically with species richness or latitude. Instead, some of the regions with the highest clustering are in latitudes corresponding to tropical and temperate forests that have produced large avian radiations, including the Amazon rainforests (latitudes -20 to 0 of the New World; Figure 2.3c), north American temperate forests (latitudes 35 to 40 of the New World; Figure 2.3c), sub-Saharan Africa and Indonesia/Papua-New Guinea (latitudes -5 to 5 of the Old world; Figure 2.3d), and South-East Asia (latitudes 25 to 30 of the Old World; Figure 2.3d).

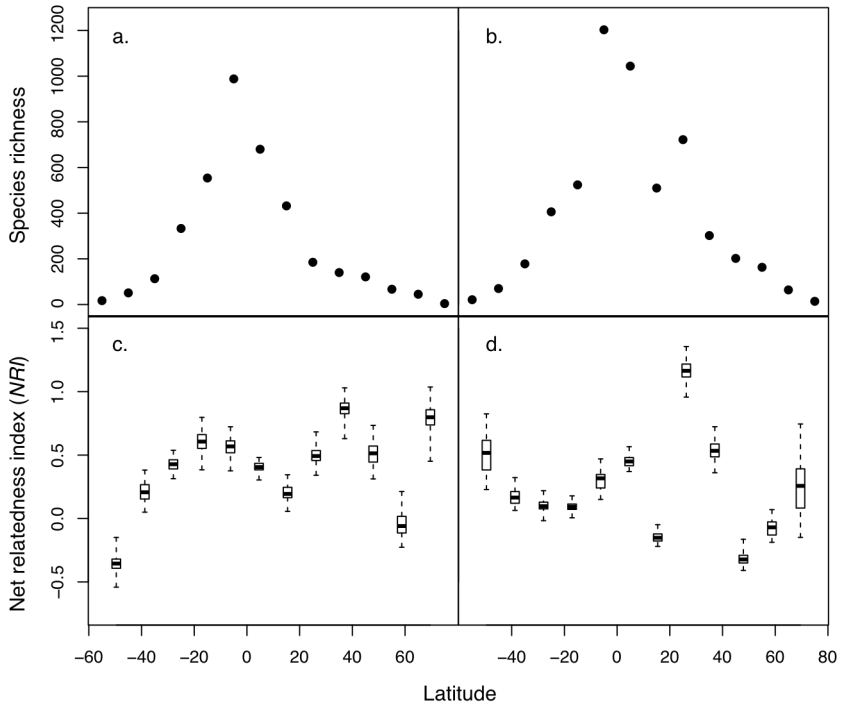


Figure 2.3. Species richness (upper panels) and standardized net relatedness index (lower panels) across latitudes for avifauna of the (a, c) New World and (b, d) the Old World, excluding marine species. The uncertainty in net relatedness index at each latitudinal band is derived from the results of 100 posterior trees.

2.4 Discussion

Both the TCH and the OTM rest on the assumption that the major clades of birds originated in tropical environments, and that temperate clades are derived from older tropical clades. The TCH explicitly links the timing of tropical-temperate transitions with the Eocene-Oligocene Climate Transition around 34Mya. During this period, average temperatures in high latitudes dropped by around 5 degrees (Liu *et al.* 2009), leading to the emergence of large new temperate climate regimes. In contrast, the OTM suggests continuous dispersal of lineages into temperate latitudes, and is less explicit about the timing of tropical-temperate transitions (Jablonski *et al.* 2006, 2013). Our results from inferred latitudinal zones for nodes of the avian phylogeny support the nestedness of temperate within tropical clades. We also find that the bird clades from temperate zones originated almost entirely after the Eocene-Oligocene Climate Transition. This is consistent with some of the evidence from the fossil record that the appearance of high-latitude bird clades occured largely after the Oligocene (Manegold *et al.* 2004; Mayr 2004). Because this pattern is explicit about the timing of diversification, it is consistent with the expectations of the TCH; a similar pattern was found recently in New World angiosperms (Kerkhoff *et al.* 2014). However, our results using discrete models of geographic evolution sugest that diversification out of the tropics has occurred as far back as 90Mya. This result is reasonable since the fossil record shows some bird diversity was present at high latitudes before the Oligocene. This might reflect the conservativeness of continuous-trait ancestral state reconstruction, which might show a late colonization of temeprate regions due to poor survival of temperate clades or due to “averaging” across tips.

Further support for the TCH is provided by the patterns of evolutionary transitions

across latitudes. While the TCH predicts a generally low rate of transition across latitudes, making no explicit predictions about the relative frequency of transitions into and out of the tropics, the OTM suggests that transitions are both frequent and asymmetric, with a higher rate of dispersal out of the tropics. Although it is difficult to define a transition rate that would distinguish the OTM and TCH models objectively, our results indicate that the majority of nodes (>60%) occupy the same latitudinal zone as their immediate ancestral node, supporting the TCH prediction of widespread latitudinal conservatism. Furthermore, the asymmetry in the transitions into and out of the tropical zone is biased in favor of dispersal into the tropics, the opposite of the expectation under the OTM. This provides support for the TCH and mirrors the pattern for New World plants (Kerkhoff *et al.* 2014).

Our results for the phylogenetic structure of bird assemblages do not support one model over the other. We find no evidence for significant phylogenetic clustering of assemblages at high latitudes, in contrast with the findings in New World angiosperms (Kerkhoff *et al.* 2014). This result might be caused by limited power in our tests to reject null models in the direction of phylogenetic clustering. This is because we standardized empirical values of NRI by drawing random samples of species from all latitudinal zones, and calculating their mean pairwise distance (MPD). These samples of the null distribution will typically include close relatives from large tropical radiations (such as the New World suboscines), leading to high MPD values for random samples. Even endemic temperate radiations derived from only a few lineages are unlikely to be sufficiently closely related to be significantly clustered. On the other hand, the metric for quantifying phylogenetic clustering used by Kerkhoff *et al.* (2014), the standardized effect size of phylogenetic diversity (PD_z), probably suffers from the opposite problem. The phylogenetic diversity of an assemblage is the sum of the branch lengths

connecting the species in the assemblage. A random sample of species is likely to traverse the root of the tree, while small, nonrandom subsets of the fauna are likely not to traverse the root. In this way, PD_z will frequently indicate significant phylogenetic clustering. NRI is less sensitive to traverses through the root of the tree because it is based on mean pairwise branch lengths compared to total branch lengths. For this reason, we believe NRI provides a more reasonable metric for phylogenetic clustering than PD_z. Although our NRI estimates do not reject the null models, they indicate substantial variation in the degree of phylogenetic clustering among latitudinal zones. We find that NRI does not vary in a monotonic fashion with latitude, suggesting that clustering is more likely related to the patterns of avian radiations in different biomes. This is consistent with evidence that the history of bird diversification is closely linked to the emergence and expansion of global floristic regions (Hawkins *et al.* 2005, 2007).

We conclude that our findings broadly support the expectations of the TCH, but are less consistent with the expectations of the OTM. At the same time, our results do not contradict the third major model for the high diversity of the tropics: faster diversification rates in the tropics. Although some previous analyses based on smaller subsets of bird fauna have supported faster diversification at lower latitudes (e.g. Cardillo, 1999; Cardillo *et al.*, 2005), two recent analyses of the global bird fauna have failed to do so (Jetz *et al.*, 2012; Rabosky and Huang 2015). Instead, these studies found that diversification rates vary idiosyncratically among clades, with some of the highest diversification rates associated with radiations in temperate regions, or on islands (Jetz *et al.* 2012). This places the time available for diversification and latitudinal conservatism mechanisms at the forefront of explanations for the latitudinal diversity gradient in birds.

2.5 References

Behrensmeyer AK (1992) *Terrestrial ecosystems through time: evolutionary paleoecology of terrestrial plants and animals*. University of Chicago Press, Chicaco, IL.

Blackburn TM, Gaston KJ (2006) Spatial patterns in the species richness of birds in the New World. *Ecography*, **19**, 369–376.

Buckley LB, Davies TJ, Ackerly DD *et al.* (2010) Phylogeny, niche conservatism and the latitudinal diversity gradient in mammals. *Proceedings of the Royal Society B: Biological Sciences*, **277**, 2131–2138.

Buzas MA, Collins LS, Culver SJ (2002) Latitudinal difference in biodiversity caused by higher tropical rate of increase. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 7841–7843.

Cardillo M (1999) Latitude and rates of diversification in birds and butterflies. *Proceedings of the Royal Society B: Biological Sciences*, **266**, 1221–1225.

Cardillo M, Orme C, Owens I (2005) Testing for latitudinal bias in diversification rates: An example using new world birds. *Ecology*, **86**, 2278–2287.

Condamine FL, Sperling FAH, Wahlberg N, Rasplus J-Y, Kergoat GJ (2012) What causes latitudinal gradients in species diversity? Evolutionary processes and ecological constraints on swallowtail biodiversity. *Ecology Letters*, **15**, 267–277.

Davies RG, Orme CDL, Storch D *et al.* (2007) Topography, energy and the global distribution of bird species richness. *Proceedings of the Royal Society B: Biological Sciences*, **274**, 1189–1197.

Davies TJ, Savolainen V, Chase MW, Moat J, Barraclough TG (2004) Environmental energy and evolutionary rates in flowering plants. *Proceedings of the Royal Society B: Biological Sciences*, **271**, 2195–2200.

Evans KL, James NA, Gaston KJ (2006) Abundance, species richness and energy availability in the North American avifauna. *Global Ecology and Biogeography*, **15**, 372–385.

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

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

Garland T, Midford PE, Ives AR (1999) An introduction to phylogenetically based statistical methods, with a new method for confidence intervals on ancestral values. *American Zoologist*, **39**, 374–388.

Gotelli NJ (2000) Null model analysis of species co-occurrence patterns. *Ecology*, **81**, 2606–2621.

Harmon LJ, Weir JT, Brock CD, Glor RE, Challenger W (2008) GEIGER: Investigating evolutionary radiations. *Bioinformatics*, **24**, 129–131.

Hawkins BA, DeVries PJ (2009) Tropical niche conservatism and the species richness gradient of North American butterflies. *Journal of Biogeography*, **36**, 1698–1711.

Hawkins B, Diniz-Filho JAF, Jaramillo C, Soeller S (2006) Post-Eocene climate change, niche conservatism, and the latitudinal diversity gradient of New World birds. *Journal of Biogeography*, **33**, 770–780.

Hawkins BA, Diniz-Filho JAF, Jaramillo CA, Soeller SA (2007) Climate, niche conservatism, and the global bird diversity gradient. *The American Naturalist*, **170**, 16–27.

Hawkins BA, Diniz-Filho JAF, Soeller SA (2005) Water links the historical and contemporary components of the Australian bird diversity gradient. *Journal of Biogeography*, **32**, 1035–1042.

Jablonski D (1993) The tropics as a source of evolutionary novelty through geological time. *Nature*, **364**, 142–144.

Jablonski D, Belanger CL, Berke SK *et al.* (2013) Out of the tropics, but how? Fossils, bridge species, and thermal ranges in the dynamics of the marine latitudinal diversity gradient. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 10487–10494.

Jablonski D, Roy K, Valentine JW (2006) Out of the tropics: evolutionary dynamics of the latitudinal diversity gradient. *Science*, **314**, 102–106.

Jansson R, Rodríguez-Castañeda G, Harding L (2013) What can multiple phylogenies say about the latitudinal diversity gradient? A new look at the tropical conservation, out-of-the-tropics and diversification rate hypothesis. *Evolution*, **67**, 1741–1755.

Jetz W, Fine PVA (2012) Global gradients in vertebrate diversity predicted by historical area-productivity dynamics and contemporary environment. *PLoS Biology*, **10**, e1001292.

Jetz W, Rahbek C (2002) Geographic range size and determinants of avian species richness. *Science*, **297**, 1548–1551.

Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO (2012) The global diversity of birds in space and time. *Nature*, **491**, 444–448.

Kembel SW, Cowan PD, Helmus MR *et al.* (2010) PICANTE: R tools for integrating phylogenies and ecology. *Bioinformatics*, **26**, 1463–1464.

Kerkhoff AJ, Moriarty PE, Weiser MD (2014) The latitudinal species richness gradient in New World woody angiosperms is consistent with the tropical conservatism

hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 8125–8130.

Liu Z, Pagani M, Zinniker D *et al.* (2009) Global cooling during the eocene-oligocene climate transition. *Science*, **323**, 1187–1190.

Manegold A, Mayr G, Mourer-Chauviré C, Nelson DA (2004) Miocene songbirds and the composition of the European passeriform avifauna. *The Auk*, **121**, 1155–1160.

Matzke NJ (2013) BioGeoBEARS: Biogeography with Bayesian (and likelihood) evolutionary analysis in R scripts. *R package, version 0.2*, **1**.

Matzke NJ (2014) Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic Biology*, **63**, 951–970.

Mayr G (2004) Old World fossil record of modern-type hummingbirds. *Science*, **304**, 861–864.

Mittelbach GG, Schemske DW, Cornell H V *et al.* (2007) Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecology Letters*, **10**, 315–331.

NatureServe BI and (2012) *Bird species distribution maps of the world*. Cambridge, UK and Arlington, USA.

Pyron RA, Wiens JJ (2013) Large-scale phylogenetic analyses reveal the causes of high tropical amphibian diversity. *Proceedings of the Royal Society B: Biological Sciences*, **280**, 20131622.

Rabosky DL & Huang H (2015). Minimal effects of latitude on present-day speciation rates in New World birds. *Proceedings of the Royal Society B: Biological Sciences,* **282**, 20142889.

Ree RH, Moore BR, Webb CO, Donoghue MJ (2005) A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution*, **59**, 2299–2311.

Ree RH, Smith SA (2008) Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology*, **57**, 4–14.

Revell LJ (2012) PHYTOOLS: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, **3**, 217–223.

Ricklefs RE (2006) Global variation in the diversification rate of passerine birds. *Ecology*, **87**, 2468–2478.

Rolland J, Condamine FL, Jiguet F, Morlon H (2014) Faster speciation and reduced extinction in the tropics contribute to the Mammalian latitudinal diversity gradient. *PLoS Biology*, **12**, e1001775.

Smith BT, Bryson RW, Houston DD, Klicka J (2012) An asymmetry in niche conservatism contributes to the latitudinal species diversity gradient in New World vertebrates. *Ecology Letters*, **15**, 1318–1325.

Soria-Carrasco V, Castresana J (2012) Diversification rates and the latitudinal gradient of diversity in mammals. *Proceedings of the Royal Society B: Biological Sciences*, rspb20121393.

Storch D, Davies RG, Zajícek S *et al.* (2006) Energy, range dynamics and global species richness patterns: reconciling mid-domain effects and environmental determinants of avian diversity. *Ecology Letters*, **9**, 1308–1320.

Webb C, Ackerly D (2002) Phylogenies and community ecology. *Annual Review of Ecology and Systematics*, **33**, 475–505.

Weir JT, Schluter D (2007) The latitudinal gradient in recent speciation and extinction rates of birds and mammals. *Science*, **315**, 1574–1576.

Wiens J (2007) Global Patterns of Diversification and richness in amphibians. *The American Naturalist*, **170**, 86–106.

Wiens JJ, Donoghue MJ (2004) Historical biogeography, ecology and species richness. *Trends in Ecology & Evolution*, **19**, 639–644.

Wiens JJ, Sukumaran J, Pyron RA, Brown RM (2009) Evolutionary and biogeographic origins of high tropical diversity in old world frogs (Ranidae). *Evolution*, **63**, 1217–1231.

Chapter 3 – Rates of molecular evolution and diversification in plants: chloroplast substitution rates correlate with species-richness in the Proteaceae

A version of this chapter was published as:

Duchêne D, Bromham L, (2013) Rates of molecular evolution and diversification in plants: chloroplast substitution rates correlate with species-richness in the Proteaceae. *BMC Evolutionary Biology*, **13**, 65

Author contributions:

LB devised the idea and experimental design for the study. DD collected and analysed the data. LB checked the alignments. DD wrote the paper; with input from LB.

3.1 Introduction

Present biodiversity has come about through processes of diversification and extinction of species, and the mechanisms that drive these processes are a central focus in evolutionary biology (e.g. Huey 2000; Allen Orr 2001; Rieseberg & Wendel 2003; Wu & Ting 2004). One intriguing relationship that has been revealed through studies of branch lengths on molecular phylogenies is a link between the rate of molecular evolution and the net diversification rate. A correlation between evolutionary rates and species diversity has been found in several groups including flowering plants (Barraclough *et al.* 1996; Barraclough & Savolainen 2001), reptiles (Eo & DeWoody 2010), birds (Lanfear *et al.* 2010a; Eo & DeWoody 2010), and other metazoan phyla, orders, and classes (Fontanillas *et al.* 2007).

However, not all the datasets analysed have provided evidence for a link between diversification rates and rates of molecular evolution. One study of a large number of phylogenies found a relationship between root-to-tip branch lengths and net diversification in around half of the phylogenies tested, but it is not clear whether this was due to low power or lack of a common trend (Webster *et al.* 2003). Another study examined genetic data within the mammals and found no evidence of an association between molecular rates and net diversification (Goldie *et al.* 2011). Accordingly, the universality and causes of the link remain uncertain.

There are three possible causes for the association between net diversification and the rate of molecular evolution (Barraclough *et al.* 1996; Barraclough & Savolainen 2001). One is that the process of diversification drives changes in the rate of molecular evolution. Speciation might influence the rate of molecular evolution through positive selection on particular genes associated with adaptation to novel niches (Rieseberg & Blackman 2010). Speciation could also cause genome-wide increases in substitution rate if speciation is typically associated with population subdivision (Pagel *et al.* 2006; Venditti & Pagel 2010). This is because a reduction in effective population size (N_e) can cause a higher rate of fixation of nearly neutral mutations (e.g. Woolfit & Bromham 2003), leading to a faster substitution rate (Venditti & Pagel 2010).

Conversely, a higher rate of molecular evolution may increase the diversification rate. A faster mutation rate may hasten differentiation between populations and promote reproductive incompatibility (Orr & Turelli 2001). For example, it has been suggested that higher standing genetic diversity in populations at low latitude may contribute to faster diversification in the tropics (Eo *et al.* 2008). Increased standing genetic variation

may produce more raw material for adaptation (Schluter & Conte 2009) or reduce the likelihood of extinction (Spielman *et al.* 2004). However, a recent study of orchids found no evidence for a link between population genetic variability and net diversification rate (Kisel *et al.* 2012). A higher rate of molecular evolution may increase the rate of diversification by accelerating the formation of hybrid incompatibility, occurring through the accumulation of genetic incompatibilities between the genomes of the diverging populations (Orr & Turelli 2001).

Alternatively, there might be a third factor that influences both the rate of molecular evolution and diversification rate, creating an indirect link between diversification and molecular evolution. For example, environmental energy (temperature and UV light) has been associated with both the diversification rate and the rate of molecular evolution in angiosperms (Davies *et al.* 2004). Other potential third factors are life history features, such as size or generation length, which are linked with the rate of molecular evolution and diversification rates of angiosperms and several metazoan taxa (Verdú 2002; Fontanillas *et al.* 2007; Smith & Donoghue 2008; Gaut *et al.* 2011). It has also been suggested that both morphological and molecular rates of change may be connected to diversification rate (Gaut *et al.* 2011). Whether the correlation between rate of molecular evolution and net diversification has a causal or indirect effect needs more investigation.

One way of disentangling the potential causes of the observed relationship between diversification rate and rate of molecular evolution is to partition substitutions in protein-coding genes into synonymous and non-synonymous substitutions. Synonymous mutations do not change the amino acid sequence of a protein and hence are expected to behave as neutral. If so, then the synonymous substitution rate (dS) should reflect only

the mutation rate (Kimura 1985). Nonsynonymous mutations are expected to have a range of fitness effects, including neutral, positive and negative, so may be subject to both drift and selection. An increase in the nonsynonymous substitution rate (*dN*) relative to the synonymous rate (*dS*) can occur through positive selection promoting the fixation of nonsynonymous mutations, or through a reduction in population size increasing the rate of fixation of nearly neutral mutations by drift.

The link between rate of molecular evolution and diversification rate has been attributed to the action of selection during speciation, or to a reduction in average population size in taxa undergoing diversification (Venditti & Pagel 2010), both of which would be expected to increase the relative rate of nonsynonymous substitutions. However, studies in angiosperms (Barraclough & Savolainen 2001), reptiles (Eo & DeWoody 2010), and birds (Lanfear *et al.* 2010a; Eo & DeWoody 2010) have found a correlation between synonymous substitutions and net diversification, leading to the suggestion that the link between molecular rates and net diversification may be driven by the mutation rate.

Here, we focus on the rate of molecular evolution in chloroplast genes. Genetic changes in chloroplast genomes have been implicated in the process of speciation in plants. Coevolution between organelle and nuclear genomes has been recognized as an important factor in plant diversification (Greiner *et al.* 2011). Plastome-genome incompatibility can cause hybrid sterility or inviability, by disrupting sexual reproduction, leaf morphologies, and machineries for photosynthesis or respiration (Kirk & Bassett 1967; Stubbe 1989; Levin 2003). Some of the genetic events in chloroplasts that produce these aberrations are gene duplications, loss of gene complexes and genome rearrangements (Xiong *et al.* 2009; Greiner *et al.* 2011; Wicke *et al.* 2011). The resulting incompatibilities are probably generalized phenomena in

plants, and the evolutionary consequence is that they can enhance post-zygotic barriers during speciation (Levin 2003; Herrmann *et al.* 2003; Johnson 2010; Greiner *et al.* 2011). It seems possible, then, that variation in rates of molecular evolution of chloroplasts could also influence the speed of genetic isolation, and hence the diversification rate of plant taxa.

Using a phylogenetic comparative analysis of sister pairs (Lanfear *et al.* 2010b), we investigated the relationship between rates of molecular evolution and net diversification in chloroplast genes of the plant family Proteaceae. This highly diverse family is mostly restricted to the Southern Hemisphere. It contains 79 recognized genera and around 1600 species, and some of its most diverse groups are the Australian genus *Banksia* and the African genus *Protea*. The high diversity of Proteaceae makes it a particularly attractive case study for diversification (e.g. Cowling & Lamont 1998; Valente *et al.* 2010; Prunier & Holsinger 2010). In addition, the family has stark contrasts in species-richness between genera even within its biodiversity hotspots (Sauquet *et al.* 2009). Of particular interest to this study are the numerous cases of monophyletic sister clades with remarkable differences in number of species. For example, the genus *Protea* has 112 species, while its sister genus *Faurea* has 15, and the *Banksia* taxon (including the dryandras) has 169 species while its sister taxon of the genera *Austromuellera* and *Musgravea* contains only 4.

We focus on the rates of evolution of six chloroplast genes available for a genus level phylogeny of the family Proteaceae (Sauquet *et al.* 2009). We use three protein-coding genes to estimate and contrast rates of synonymous (*dS*) and non-synonymous (*dN*) substitutions. Comparing *dN*, *dS*, and ω (*dN*/*dS*) to species-richness of clades allows us to separate the effect of mutation rate on net diversification from the effect of selection

and effective population size. In this way, we aim to provide insight into the factors underlying the correlation between rates of molecular evolution and net diversification.

3.2 Methods

3.2.1 Sister pairs

We used a phylogenetic analysis of the family Proteaceae by Sauquet *et al.* (2009) that includes the 79 recognized genera and the species-richness, being the number of species in each genus, was compiled from the literature (references are also available in Weston & Barker 2006). For the present study, monophyletic pairs of sister taxa that display differences in current species-richness were chosen from Sauquet *et al*'s phylogeny. The main criterion to select pairs was that the pair was monophyletic, so the two sister taxa had the same amount of time to accumulate species diversity and substitutions, and each sister pair was phylogenetically independent from all other such pairs.

We chose one genus to represent each sister taxon in order to avoid bias in branch length estimation due to the node density effect (Hugall & Lee 2007). The chosen genus was the one with the greatest gene coverage. If the genera of a sister taxon had equal genetic coverage the genus was chosen at random. Using only one sequence per genus may reduce the power of the test, which may obscure a weak pattern; but using only one randomly selected species per sister taxon is unlikely to generate any systematic biases in rates, making this approach conservative for testing an association between rates of molecular evolution and net diversification (see Lanfear *et al.* 2010a; Goldie *et al.* 2011). In some cases a sister taxon is represented by combining sequences from several closely related genera (shown subsequently as a taxon with a "+"). This practice increases our power to resolve the shared history of that taxon since its divergence from

the common ancestor of the sister pair, and this is unlikely to create any systematic biases in rate estimations (Appendix – Chapter 3).

3.2.2 Molecular dataset

Branch length estimation was critical for comparative analyses in this study, so the genetic dataset required unambiguous genetic alignments and the maximum gene coverage of the species analysed. With these criteria we included six genes of chloroplast origin (*atpB*, *atpB-rbcL*, *matK*, *rbcL*, *trnL* intron, and *trnL-trnF*) from the data by Sauquet et al. 2009 and available in the GenBank repository (Appendix – Chapter 3). These were then aligned using the MUSCLE algorithm, checked by eye, and manually corrected using the program SeaView v4 (Gouy *et al.* 2010). This resulted in a 6278bp alignment with 62 taxa, 4457bp of exons, and 1821bp of introns (Appendix – Chapter 3).

3.2.3 Phylogenetic estimation

Each gene alignment was tested for the most appropriate model of substitutions using likelihood estimation and comparison with the Bayesian Information Criterion as implemented in the package "ape" (Paradis *et al.* 2004) in R (The R Project - www.r-project.org/). Applying a partition by genes with the models selected (Appendix – Chapter 3), a maximum likelihood analysis was run using GARLI v2.0 (Zwickl 2006). One thousand replicates of this analysis were run in GARLI with different random seeds to avoid reaching a local optimum. The resulting tree was then used to extract the branch length values of the sister pairs (Figure 3.1). If any of our chosen sister pairs were not monophyletic in our phylogeny they were excluded from the analysis. Twenty sister pairs of the initial twenty-two chosen from Sauquet at al's phylogeny were monophyletic in our estimates (Figure 3.1).

The inferred phylogenies (including those inferred for synonymous and non-synonymous substitutions; see next section) were examined for significant overall variation in branch lengths. To do this, we estimated the likelihoods of both a constant rates model and a free rates model (where there is one rate per branch) in the program HYPHY v2.1 (Pond *et al.* 2005), and compared them using a likelihood ratio test. If the free rates model provides a better fit for the data, this suggests significant variation in rates of molecular evolution across the phylogeny.

3.2.4 dN and dS trees

To examine the potential link between synonymous substitutions (*dS*), non-synonymous substitutions (*dN*), and ω (*dN*/*dS*) and species-richness, we estimated branch-wise *dN* and *dS* rates using an alignment of the coding genes (*atpB*, *matK*, and *rbcL*) in the program HYPHY v2.1 (Pond *et al.* 2005) with the MG94 model of codon evolution (Muse & Gaut 1994). For the estimation of *dN* and *dS* trees, the MG94 model can be combined with any of the nucleotide substitution models nested in GTR+G+I. To choose the best combination we first gave HYPHY v2.0 a notation to estimate the codon frequencies (the frequency of each of the four bases in each of the three codon positions), which was a 3x4 matrix. Then the likelihood of each of the 203 possible models (possible partitions of rates in the model transition matrix) was estimated and one was chosen according to the Akaike Information Criterion. The model chosen had four parameters, where $\theta_{AG} = \theta_{CT}$ and $\theta_{CG} = \theta_{GT}$, as implemented in HYPHY v2.0. Finally, these parameters were optimized with maximum likelihood, constraining the topology to that estimated from the full six-gene dataset (Appendix – Chapter 3). The output included the branch-wise *dN* and *dS* substitutions per site (Figure 3.1), which were used to extract the branch lengths of sister pairs. Given that HYPHY v2.0 estimates the values of *dN* and *dS* as the expected number of substitutions per

nucleotide per site, the values for ω were calculated as the ratio between the two estimates (*dN*/*dS*).

3.2.5 Statistics

The total species-richness and the estimates of branch lengths (for all substitutions, *dN* and *dS*) and ω were collected for each of the two taxa in the sister pairs (Table 3.1). As the sister taxa had the same amount of time to accumulate species and substitutions, we assumed that the branch length is proportional to the rate of molecular evolution of chloroplasts (reviewed in Lanfear *et al.* 2010b). Similarly, we assumed that species-richness of each sister clade reflects the net diversification (speciation minus extinction) of that taxon since the last common ancestor of the sister pair.

We performed a one-tailed Wilcoxon Signed-Ranks test in R, which resembles the standard sign test but accounts for the magnitudes of the differences between matched taxa (Wilcoxon *et al.* 1963). This test sets a sign to each pair by subtracting branch lengths in the direction from species-rich to species poor; we did not include the sister pairs with equal species-richness as these cannot be accommodated in the Wilcoxon Signed-Ranks test. Then, the absolute difference between the two values was used to rank the pairs (lowest difference has rank 1 and the highest rank is the number of pairs). Tied values receive as a rank the mean of the ranks they span. The ranks are then given the sign of the pair and then added to produce a W statistic (Wilcoxon *et al.* 1963).

3.3 Results

A model where every branch in the phylogeny had an independent rate of substitutions had a significantly higher likelihood than the constant rates model in all the rates

estimations (all substitutions, *dN*, *dS*), and ω (P value < 0.01 for all tests; see Methods section), indicating that the rate of molecular evolution of the chloroplast genes analysed varies significantly between taxa of the family Proteaceae.

Species-rich taxa had significantly longer branch lengths in the phylogeny estimated from the full 6-gene dataset (one-tailed Wilcoxon Signed-Rank test, W = 175, P = 0.0036). This is evidence of a positive association between net diversification and the rate of molecular evolution of chloroplasts in the family Proteaceae. We also found significant differences in estimates of synonymous (*dS*: W = 152, P = 0.041), and non-synonymous rates (*dN*: W = 165, P = 0.012). However, we did not find a significant differences in estimates of ω between species-rich and species-poor sister taxa (W = 100, P = 0.14).

3.4 Discussion

We found a significant positive association between the rate of molecular evolution in chloroplast genes and species-richness in the plant family Proteaceae. There were significant associations between both synonymous and non-synonymous rates of substitutions and net diversification, but not between ω (*dN*/*dS*) and diversification. The pattern of correlations in this study are consistent with other studies of angiosperms (Barraclough & Savolainen 2001; Lancaster 2010), reptiles, and birds (Webster *et al.* 2003; Lanfear *et al.* 2010a; Eo & DeWoody 2010). Importantly, our results give some insight into the cause of this relationship. The variation in both synonymous and non-synonymous substitution rates between taxa may reflect a role for the rate of production of mutations in the chloroplast genome in the process of diversification in the Proteaceae. Because we fail to detect an increase in ω in species-rich clades, our

analysis provides no tangible evidence for a role of selection or population size change in driving the relationship between substitution rates and diversification rates in this group.

3.4.1 Synonymous substitutions and net diversification

Synonymous substitution rates are typically interpreted to reflect the rate of production of mutations. Mutation rates are known to vary between taxa for a range of reasons. For example, species with shorter generation times tend to have faster mutation rates (Smith & Donoghue 2008), presumably due to the accumulation of DNA replication errors (Bromham 2009). Mutation rates can also vary across the genome, which may be at least in part due to differences in base composition or gene length (Yang & Gaut 2011; Gaut *et al.* 2011).

Since synonymous substitutions are commonly assumed to be functionally neutral, they are often used to provide a window into variation of mutation rates. However, bias in codon use can influence the synonymous rate if, for example, there is selection for efficiency in the process of translation (Ikemura 1985). This type of bias has been found in angiosperm mitochondrial genes although with selection that is so weak that it is considered not to affect estimations of mutation rates (Sloan & Taylor 2010). The chloroplast genome of angiosperms also has minimal codon bias and weak selection for translation efficiency (Morton 1998). Therefore, in this study, we consider that the relationship between synonymous substitution rate and net diversification is telling us something about the link between mutation and diversification, whether it reflects differences in the absolute mutation rate per unit time or in the differences in the distribution of fitness effects of synonymous mutations between taxa.

One explanation for the link between synonymous substitutions and net diversification is that higher mutation rates could cause faster genetic divergence between taxa. In this case, genes of chloroplast origin may be important because they can drive reproductive isolation in plants by interacting with nuclear alleles (Kimura 1985). Reproductive barriers can occur due to the failure of interactions between nuclear and cytoplasmic gene complexes, for example cytoplasmic male sterility (Fishman & Willis 2006). An increased mutation rate may generate more molecular changes that cause these phenomena, known as Bateson-Dobzhansky-Muller (BDM) incompatibilities, and so might accelerate post-zygotic isolation (Orr & Turelli 2001; Welch 2004; Rieseberg & Willis 2007).

Some studies have found that taxon-specific variation in rates of molecular evolution are consistent across the nuclear, mitochondrial, and chloroplast genomes (Eyre-Walker & Gaut 1997), so it may be that the increase in substitution rates that we detected also apply to the nuclear genomes of species-rich taxa in the Proteaceae. In this case, higher rates of substitution in the nuclear genome may be contributing to the formation of incompatibilities between diverging populations, either by generating BDM incompatibilities between the nuclear genomes or through interactions between the nuclear and organelle genomes.

Therefore, the association between the synonymous rate of chloroplast genes and diversification rate reported here may reflect the acceleration in the formation of post-zygotic reproductive isolation. This is also consistent with our finding of an association between non-synonymous rates and net diversification because an increase in the mutation rate should also result in more effectively neutral non-synonymous substitutions going to fixation.

3.4.2 Indirect links between diversification and the rate of molecular evolution

An indirect relationship between the rate of molecular evolution and diversification could arise if some factor influenced both. For example, it has been suggested that tropical taxa have a higher rate of molecular evolution than their temperate counterparts (Wright *et al.* 2006). This correlation might reflect a direct effect of temperature or UV light on mutagenesis (Allen *et al.* 2002), or an indirect effect if higher environmental energy leads to further growth rates and more rapid generation turnover, which could influence the mutation rate through accumulation of replication errors (Smith & Donoghue 2008). If higher growth rates also lead to faster diversification (Rohde 1992), then this could create an indirect link between the mutation rate and diversification. This may also explain the patterns in a study on angiosperms that investigated the correlations between species-richness, the rate of molecular evolution, and three energy variables (temperature, UV light, and evapotransportation), but which found no support for the mutation rate as the direct mediator of species-richness (Davies *et al.* 2004). However, it is interesting to note that the Proteaceae do not appear to have higher rates of diversification in the tropics. Instead, much of their radiation has occurred in Mediterranean climate hotspots (Sauquet *et al.* 2009).

Life history variation provides another possible indirect link between rates of molecular evolution and diversification. Several studies have suggested that annual plants have a faster rate of molecular evolution than perennials, a pattern generally attributed to the generation time effect (see Bromham 2009; Gaut *et al.* 2011). The potential for interactions between mechanisms that influence species-richness and the rate of molecular evolution has a broad scope and remains to be studied in detail.

3.4.3 Net diversification and ω

It has been suggested that processes associated with speciation drive the link between rates of substitution and net diversification (Webster *et al.* 2003; Pagel *et al.* 2006; Venditti & Pagel 2010), including diversifying selection and changes in effective population size. A reduction in effective population size (N_e) may be caused by a speciation event that changes the population structure, such as vicariant or peripatric speciation (Charlesworth 2009). This could lead to new adaptive pressures (Lee 2002), or high levels of genetic drift in population bottlenecks (Carson & Templeton 1984). These processes could increase the rate of fixation on non-synonymous substitutions, which may be reflected in an increase in ω (dN/dS; Ohta 1992; Charlesworth 2009; Woolfit 2009).

Two studies on large numbers of phylogenies found a recurrent correlation between root-to-tip distances and the number of speciation events (Webster *et al.* 2003; Pagel *et al.* 2006). This result was interpreted as evidence that clades with more speciation events have a faster rate of molecular evolution, which they attributed to punctuational change associated with the founder-effect model of speciation. However, while these phylogenetic tests reveal an association between rates of evolution and number of phylogenetic nodes, they are not able to localise those changes to the nodes rather than the edges of the phylogeny, so cannot distinguish between two alternative explanations, that speciation events increase the substitution rate or that higher substitution rate increases diversification. One possible way to separate these models is in their predicted effects on the patterns of substitutions. If population divisions associated with speciation events have significant effects on rates of substitution, either through change in selection or reduction in effective population size, it should result in a relative increase in the nonsynonymous rate, reflected in an increase in dN/dS (ω).

We did not detect any association between ω and net diversification (see also Lanfear *et al.* 2010a). This may be because net diversification is not associated with consistent effects on population size (Rieseberg & Willis 2007), or diversification does affect effective population size, but the effect on ω is overwritten by other population fluctuations (Rieseberg & Willis 2007). Alternatively, the effect on reduction in effective population size may be too small to be detected or may be affected by the method of estimation of ω (Eyre-Walker 2006). In theory, N_e is an adequate representation of genetic drift in large populations and when the population size has been consistent for a long enough time (Charlesworth 2009). It has even been shown that following transient increases in N_e there can be an increase in the rate of substitutions due to slightly advantageous mutations, which is the opposite of the predicted effect (Charlesworth & Eyre-Walker 2007). Therefore, although N_e is likely to have a significant effect on the rate of substitutions, predicting the form of the effect is far from a simple task (Woolfit 2009). Therefore, failure to detect an effect of population size changes of ω in this study does not imply that N_e is unaffected by diversification; however, it does suggest that changes in N_e during diversification are unlikely to explain the differences in substitution rates that we observe in these data.

3.4.4 Molecular evolution and diversification in plants

Many studies have focussed on identifying the genetic loci underlying speciation. These can be genes that contribute to the genetic isolation of populations, genes that drive differential ecological adaptation, and "magic traits" that do both (e.g. Nosil & Schluter 2011; Servedio *et al.* 2011). Genome-wide scans are increasingly being used to identify outlier loci that show signatures of selection, including loci that differ between pairs that are associated with floral traits, climatic factors, and sterility (Strasburg *et al.*

2012). This study takes a different, and complementary, approach to analysing the role that the genome-wide generation of genetic change plays in diversification.

It is possible that higher mutation rates may create a greater pool of standing variation from which adaptive substitutions can be derived. This increased pool of variation can lead to either higher rates of speciation, through adaptation to foreign environments, or lower rates of extinction, through adaptation to changing conditions. The assumption that chloroplast genes do not play a direct role in ecological adaptation has now been challenged: for example, values of ω above 1 have been estimated in rbcL (Kapralov & Filatov 2007) and MatK sequences (Hao *et al.* 2010) for some linages, which was interpreted as a signal of positive selection. However, we did not find evidence of higher ω in more diverse clades for the loci analysed in this study, and the relationship between the amount of standing variation and diversification in plants is not clear. For example, studies have found that diversification in orchids is not associated with greater genetic diversity at the population level (Kisel *et al.* 2012; Phillips *et al.* 2012).

Another scenario is that higher mutation rates contribute to the rate at which the genomes of different populations diverge and become gradually incompatible, making hybrids between the populations less fit. Bateson-Dozhansky-Muller (BDM) incompatibilities may arise from selection in different populations, but they might also be unconnected to ecological or behavioural divergence, in other words they may be "incidental on other acquired differences'' (Darwin 1859). For a mutation to go to substitution in one population, it must be broadly compatible with other common alleles in that population. But it will not have been "tested by natural selection" against alleles in isolated populations, and bringing those unharmonised alleles together may result in a maladapted individual (Presgraves 2010). The more unique substitutions each

population has acquired, the greater the chance that a hybrid zygote will contain at least one pair of incompatible alleles. The steady increase in hybrid incompatibility with time in many species has been taken as evidence that many loci may contribute to BDM incompatibilities (Welch 2004). Under the BDM model, the rate of speciation may increase as the mutation rate increases (Gavrilets 2003). Since the substitutions underlying BDM incompatibilities do not have to occur evenly in both taxa, a higher mutation rate in one taxon should drive divergence between them (Welch 2004). Debate continues over the rate at which hybrid incompatibility accumulates, particularly concerning the prediction that BDM incompatibilities should "snowball", accelerating relative to the substitution rate (Gourbière & Mallet 2010; Matute *et al.* 2010).

Importantly, incompatibilities between populations can involve both organelle and nuclear genomes. Just as alleles within the nuclear genome must be able to work together to produce viable offspring, genomes of chloroplasts must be co-adapted to nuclear genome to allow normal development (Moison *et al.* 2010). For example, alleles that cause cytoplasmic male sterility may be countered by suppression genes in the nuclear genome that restore male function, so a hybrid that inherited the organelle genome without the corresponding nuclear allele would be male sterile (Johnson 2010). While cytonuclear conflict has been more frequently studied between mitochondrial and nuclear genomes, there is increasing evidence that incompatibilities between chloroplast and nuclear genomes contribute to hybrid incompatibility in many plant species (Greiner *et al.* 2011).

Polyploidy is another important factor in the diversification of many plant taxa (Ainouche & Jenczewski 2010; Abbott *et al.* 2013), but by focussing only on the chloroplast we minimized the impact of genome duplication on our analyses.

Chloroplasts typically have uniparental inheritance, which simplifies the interpretation of the effects of genetic changes on divergence. However, the mode of inheritance of chloroplasts, whether inherited paternally or maternally, can vary between taxa, which can influence their levels of genetic diversity (Petit *et al.* 2005). Chloroplast sequences should also limit the impact of "divergence hitchhiking", where linked neutral loci go to fixation through being linked to a locus under selection (Strasburg *et al.* 2012). Lastly, while chloroplasts use recombination for genome repair (Maréchal & Brisson 2010), hybridization of chloroplasts from different taxa does not appear to be common (Bock 2010; Greiner *et al.* 2011).

3.4.5 Conclusions

We show a significantly faster rate of molecular evolution in chloroplast genes of species-rich taxa of the family Proteaceae. These results offer evidence for the influence of the rate of molecular evolution on diversification. This does not imply that the rate of molecular evolution explains the process of diversification, because this complex and heterogeneous process can be influenced by many mechanisms such as hybridization (Mallet 2007), polyploidy (Soltis *et al.* 2009), allopatric events (Venditti & Pagel 2010), and the duplication of genes (Xiong *et al.* 2009). However, the results do suggest that the substitution rate in chloroplasts may be one of these influences on the speed at which populations diverge, thus influencing the probability of populations becoming separate species (Orr & Turelli 2001). Notably, studies like this one contain several sources of error, such as stochastic error in molecular branch lengths, and error from sampling of taxa from each genus. Future studies on this topic would benefit from investigating the effects of these sources of bias on our conclusions.

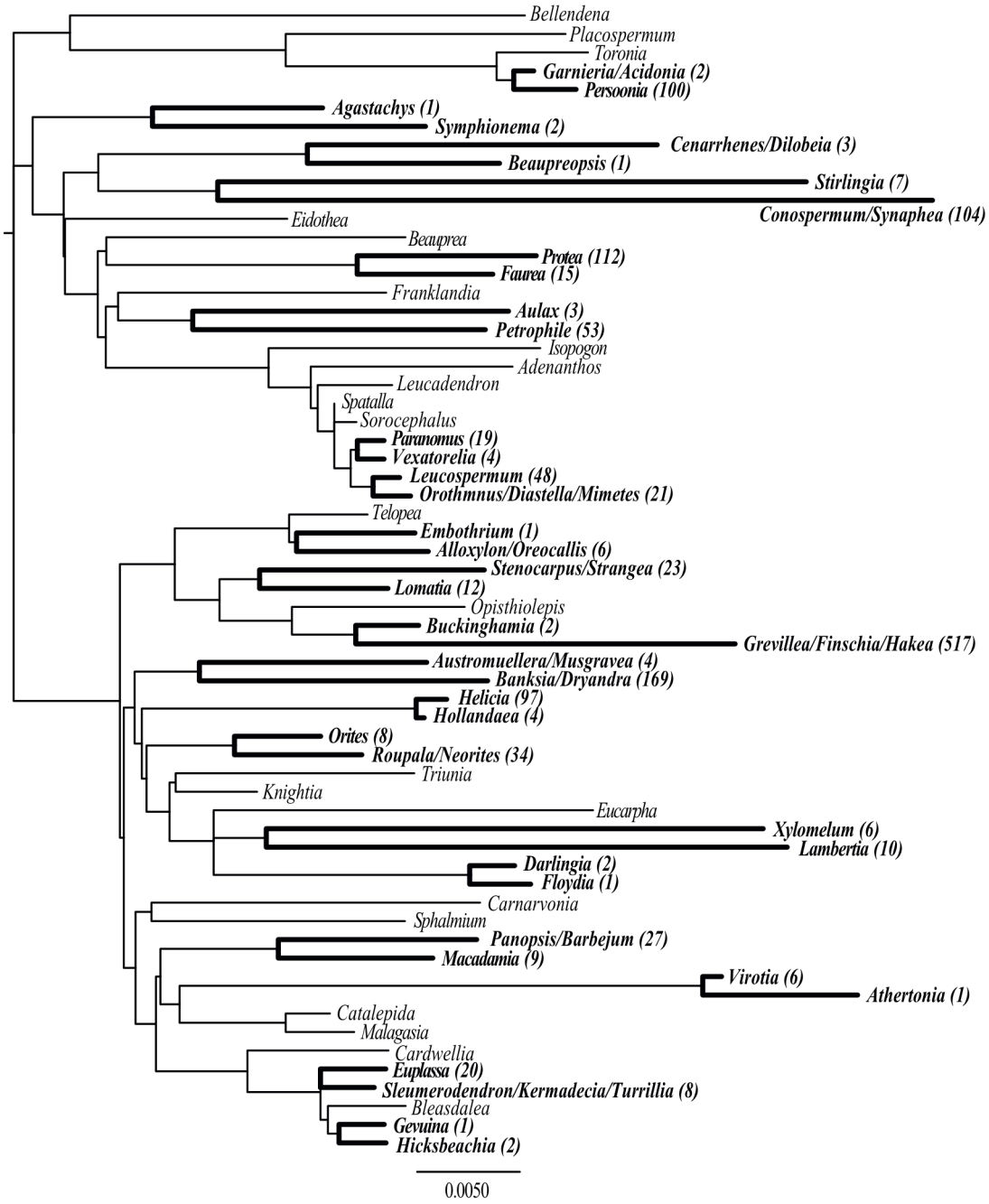


Figure 3.1. Molecular phylogeny of the family Proteaceae including the 20 sister pair groups used for the present analyses in bold. The branch lengths shown are proportional to the estimated number of substitutions as indicated by the whole dataset of six chloroplast genes. The species richness of the clades compared is shown in parentheses, with species numbers taken from Sauquet *et al.* (2009). The scale bar indicates number of mutations per site.

Table 3.1. Sister comparisons included in this analysis. For each pair, we list the two clades compared (a "+" indicates where more than one genus were combined as one sister taxon of a comparison). The taxon from which the sequence was taken is listed, with "spp" indicating that several congeneric sequences were combined (see Table A3.1 in Appendix – Chapter 3 for details). The species richness for each sister taxon is taken from Sauquet *et al.* (2009), and the number of available gene sequences for each comparison, out of the six chloroplast genes analysed in this study (See Table A3.1 in Appendix – Chapter 3). The estimated substitution rates for each sister taxon are given for Total branch lengths (all substitutions) as well as synonymous (*dS*) and non-synonymous (*dN*) substitutions, and estimates of ω (*dN*/*dS*).

Pair	Sister clades	Taxa	Species richness	Genes available	Total branch lengths	*dN* branch lengths	*dS* branch lengths	*dN*/*dS*
1	Persoonia	*Persoonia spp.*	100	2	0.00303	0.00331	0.00299	1.10704
	Garnieria + Acidonia	*Garnieria spathulaefolia*	2		0.00097	0.00105	0.00170	0.61562
2	Symphionema	*Symphionema montanum*	2	4	0.01329	0.00990	0.01713	0.57818
	Agastachys	*Agastachys odorata*	1		0.00828	0.00403	0.02123	0.18971
3	Cenarrhenes + Dilobeia	*Cenarrhenes nitida*	3	1	0.01703	0.00379	0.02849	0.13315
	Beaupreopsis	*Beaupreopsis paniculata*	1		0.00936	0.00315	0.01317	0.23892
4	Conospermum + Synaphea	*Conospermum spp.*	104	4	0.03480	0.01312	0.06744	0.19454
	Stirlingia	*Stirlingia latifolia*	7		0.02865	0.01164	0.05841	0.19922
5	Protea	*Protea cynaroides*	112	5	0.00872	0.00588	0.01364	0.43081
	Faurea	*Faurea spp.*	15		0.00660	0.00589	0.01052	0.55959
6	Petrophile	*Petrophile spp.*	53	3	0.01424	0.00723	0.02834	0.25512
	Aulax	*Aulax spp.*	3		0.01537	0.01279	0.02951	0.43337
7	Paranomus	*Paranomus spp.*	19	2	0.00130	0.00187	1.00E-09	1.87E+06
	Vexatorella	*Vexatorella alpina*	4		0.00130	0.00047	0.00445	0.10504
8	Leucospermum	*Leucospermum spp.*	48	2	0.00129	0.00095	0.00292	0.32577
	Orothamnus + Diastella + Mimetes	*Mimetes spp.*	21		0.00183	0.00187	0.00296	0.63319
9	Alloxylon + Oreocallis	*Alloxylon spp.*	6	4	0.00641	0.00418	0.00931	0.44871
	Embothrium	*Embothrium coccineum*	1		0.00577	0.00323	0.00625	0.51730
10	Stenocarpus	*Stenocarpus*	26	6	0.01094	0.00602	0.01962	0.30693

	Strangea	*salignus*						
	Lomatia	*Lomatia spp.*	12		0.00625	0.00395	0.01304	0.30245
11	Grevillea + Finschia + Hakea	*Grevillea spp.*	515	6	0.01847	0.01027	0.03031	0.33881
	Buckinghamia	*Buckinghamia spp.*	2		0.00306	0.00118	0.00916	0.12925
12	Virotia	*Virotia leptophylla*	6	1	0.00091	0.00097	1.00E-09	9.71E+05
	Athertonia	*Athertonia diversifolia*	1		0.00754	0.00535	0.01243	0.43028
13	Panopsis+ Brabejum	*Panopsis spp.*	27	4	0.00966	0.00577	0.01995	0.28920
	Macadamia	*Macadamia spp.*	9		0.00752	0.00623	0.01371	0.45473
14	Hicksbeachia	*Hicksbeachia pinnatifolia*	2	1	0.00229	0.00261	0.00095	2.75046
	Gevuina	*Gevuina avellana*	1		0.00220	0.00098	0.00309	0.31894
15	Euplassa	*Euplassas occidentalis*	20	3	0.00318	0.00281	0.00448	0.62802
	Sleumerodendron + Kermadecia + Turrillia	*Sleumerodendron austrocaledonicum + Kermadecia pronyensis*	8		0.00262	0.00199	0.00619	0.32064
16	Banksia + Dryandra	*Banksia spp.*	169	6	0.01403	0.00893	0.02906	0.30745
	Austromuellera + Musgravea	*Austromuellera trinervia*	4		0.01107	0.00525	0.02498	0.21017
17	Roupala + Neorites	*Roupala montana + Roupala monosperma + Neorites kevediana*	34	6	0.00620	0.00286	0.01367	0.20891
	Orites	*Orites spp.*	8		0.00420	0.00229	0.00483	0.47439
18	Darlingia	*Darlingia darlingiana*	2	2	0.00219	0.00236	0.00295	0.79752
	Floydia	*Floydia praealta*	1		0.00297	0.00132	0.00571	0.23042
19	Lambertia	*Lambertia spp.*	10	4	0.02536	0.01116	0.05726	0.19491
	Xylomelum	*Xylomelum spp.*	6		0.02419	0.01036	0.04448	0.23296
20	Helicia	*Helicia spp.*	97	3	0.00149	0.00094	0.00304	3.10E-01
	Hollandaea	*Hollandaea riparia*	4		0.00039	0.00000	0	0

3.5 References

Abbott R, Albach D, Ansell S *et al.* (2013) Hybridization and speciation. *Journal of Evolutionary Biology*, **26**, 229–246.

Ainouche ML, Jenczewski E (2010) Focus on polyploidy. *The New Phytologist*, **186**, 1–4.

Allen AP, Brown JH, Gillooly JF (2002) Global biodiversity, biochemical kinetics, and the energetic-equivalence rule. *Science*, **297**, 1545–1548.

Barraclough TG, Harvey PH, Nee S (1996) Rate of rbcL gene sequence evolution and species diversification in flowering plants (Angiosperms). *Proceedings of the Royal Society B: Biological Sciences*, **263**, 589–591.

Barraclough T, Savolainen V (2001) Evolutionary rates and species diversity in flowering plants. *Evolution*, **55**, 677–683.

Bock R (2010) The give-and-take of DNA: horizontal gene transfer in plants. *Trends in Plant Science*, **15**, 11–22.

Bromham L (2009) Why do species vary in their rate of molecular evolution? *Biology Letters*, **5**, 401–404.

Carson HL, Templeton AR (1984) Genetic revolutions in relation to speciation phenomena: The founding of new populations. *Annual Review of Ecology and Systematics*, **15**, 97–132.

Charlesworth B (2009) Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, **10**, 195–205.

Charlesworth J, Eyre-Walker A (2007) The other side of the nearly neutral theory, evidence of slightly advantageous back-mutations. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 16992–16997.

Cowling RM, Lamont BB (1998) On the nature of Gondwanan species flocks: diversity of Proteaceae in mediterranean south-western Australia and South Africa. *Australian Journal of Botany*, **46**, 335.

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

Davies TJ, Savolainen V, Chase MW, Moat J, Barraclough TG (2004) Environmental energy and evolutionary rates in flowering plants. *Proceedings of the Royal Society B: Biological Sciences*, **271**, 2195–2200.

Eo SH, DeWoody JA (2010) Evolutionary rates of mitochondrial genomes correspond to diversification rates and to contemporary species richness in birds and reptiles. *Proceedings of the Royal Society B: Biological Sciences*, **277**, 3587–3592.

Eo SH, Wares JP, Carroll JP (2008) Population divergence in plant species reflects latitudinal biodiversity gradients. *Biology Letters*, **4**, 382–384.

Eyre-Walker A (2006) The genomic rate of adaptive evolution. *Trends in Ecology & Evolution*, **21**, 569–575.

Eyre-Walker A, Gaut BS (1997) Correlated rates of synonymous site evolution across plant genomes. *Molecular Biology and Evolution*, **14**, 455–460.

Fishman L, Willis JH (2006) A cytonuclear incompatibility causes anther sterility in Mimulus hybrids. *Evolution*, **60**, 1372–1381.

Fontanillas E, Welch JJ, Thomas JA, Bromham L (2007) The influence of body size and net diversification rate on molecular evolution during the radiation of animal phyla. *BMC Evolutionary Biology*, **7**, 95.

Gaut B, Yang L, Takuno S, Eguiarte LE (2011) The patterns and causes of variation in plant nucleotide substitution rates. *Annual Review of Ecology, Evolution, and Systematics*, **42**, 245–266.

Gavrilets S (2003) Models of speciation: What have we learned in 40 years? *Evolution*, **57**, 2197–2215.

Goldie X, Lanfear R, Bromham L (2011) Diversification and the rate of molecular evolution: no evidence of a link in mammals. *BMC Evolutionary Biology*, **11**, 286.

Gourbière S, Mallet J (2010) Are species real? The shape of the species boundary with exponential failure, reinforcement, and the "missing snowball". *Evolution*, **64**, 1–24.

Gouy M, Guindon S, Gascuel O (2010) SEAVIEW version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution*, **27**, 221–224.

Greiner S, Rauwolf U, Meurer J, Herrmann RG (2011) The role of plastids in plant speciation. *Molecular Ecology*, **20**, 671–691.

Hao DC, Chen SL, Xiao PG (2010) Molecular evolution and positive Darwinian selection of the chloroplast maturase matK. *Journal of Plant Research*, **123**, 241–247.

Herrmann RG, Maier RM, Schmitz-Linneweber C (2003) Eukaryotic genome evolution: rearrangement and coevolution of compartmentalized genetic information. *Philosophical transactions of the Royal Society of London Series B: Biological Sciences*, **358**, 87–97.

Huey RB (2000) Rapid evolution of a geographic cline in size in an introduced fly. *Science*, **287**, 308–309.

Hugall AF, Lee MSY (2007) The likelihood node density effect and consequences for evolutionary studies of molecular rates. *Evolution*, **61**, 2293–2307.

Ikemura T (1985) Codon usage and tRNA content in unicellular and multicellular organisms. *Molecular Biology and Evolution*, **2**, 13–34.

Johnson NA (2010) Hybrid incompatibility genes: remnants of a genomic battlefield? *Trends in Genetics*, **26**, 317–325.

Kapralov M V, Filatov DA (2007) Widespread positive selection in the photosynthetic Rubisco enzyme. *BMC Evolutionary Biology*, **7**, 73.

Kimura M (1985) *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.

Kirk JTO, Bassett RAE (1967) *The plastids. Their chemistry, structure, growth and inheritance. 608 pp*. WH Freeman & Co., London & San Francisco.

Kisel Y, Moreno-Letelier A, Bogarín D (2012) Testing the link between population genetic differentiation and clade diversification in Costa Rican orchids. *Evolution*, **66**, 3035–3052.

Lancaster LT (2010) Molecular evolutionary rates predict both extinction and speciation in temperate angiosperm lineages. *BMC Evolutionary Biology*, **10**, 162.

Lanfear R, Ho SYW, Love D, Bromham L (2010a) Mutation rate is linked to diversification in birds. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 20423–20428.

Lanfear R, Welch JJ, Bromham L (2010b) Watching the clock: studying variation in rates of molecular evolution between species. *Trends in Ecology & Evolution*, **25**, 495–503.

Lee CE (2002) Evolutionary genetics of invasive species. *Trends in Ecology & Evolution*, **17**, 386–391.

Levin D (2003) The cytoplasmic factor in plant speciation. *Systematic Botany*, **28**, 5–11.

Mallet J (2007) Hybrid speciation. *Nature*, **446**, 279–83.

Maréchal A, Brisson N (2010) Recombination and the maintenance of plant organelle genome stability. *The New Phytologist*, **186**, 299–317.

Matute DR, Butler IA, Turissini DA, Coyne JA (2010) A test of the snowball theory for the rate of evolution of hybrid incompatibilities. *Science*, **329**, 1518–1521.

Moison M, Roux F, Quadrado M *et al.* (2010) Cytoplasmic phylogeny and evidence of cyto-nuclear co-adaptation in Arabidopsis thaliana. *The Plant Journal*, **63**, 728–738.

Morton BR (1998) Selection on the codon bias of chloroplast and cyanelle genes in different plant and algal lineages. *Journal of Molecular Evolution*, **46**, 449–459.

Muse S, Gaut B (1994) A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Molecular Biology and Evolution*, **11**, 715–724.

Nosil P, Schluter D (2011) The genes underlying the process of speciation. *Trends in Ecology & Evolution*, **26**, 160–167.

Ohta T (1992) The nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematics*, **23**, 263–286.

Orr AH (2001) The genetics of species differences. *Trends in Ecology & Evolution*, **16**, 343–350.

Orr H, Turelli M (2001) The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. *Evolution*, **55**, 1085–1094.

Pagel M, Venditti C, Meade A (2006) Large punctuational contribution of speciation to evolutionary divergence at the molecular level. *Science*, **314**, 119–121.

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

Petit RJ, Duminil J, Fineschi S *et al.* (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology*, **14**, 689–701.

Phillips R, Dixon K, Peakall R (2012) Low population genetic differentiation in the Orchidaceae: implications for the diversification of the family. *Molecular Ecology*, **21**, 5208–5220.

Pond SLK, Frost SDW, Muse S V (2005) HYPHY: hypothesis testing using phylogenies. *Bioinformatics*, **21**, 676–679.

Presgraves DC (2010) Speciation genetics: search for the missing snowball. *Current Biology*, **20**, 1073–1074.

Prunier R, Holsinger KE (2010) Was it an explosion? Using population genetics to explore the dynamics of a recent radiation within Protea (Proteaceae L.). *Molecular Ecology*, **19**, 3968–3980.

Rieseberg LH, Blackman BK (2010) Speciation genes in plants. *Annals of Botany*, **106**, 439–455.

Rieseberg LH, Wendel J (2003) Plant speciation - rise of the poor cousins. *New Phytologist*, **161**, 3–8.

Rieseberg LH, Willis JH (2007) Plant speciation. *Science*, **317**, 910–914.

Rohde K (1992) Latitudinal gradients in species diversity: The search for the primary cause. *Oikos*, **65**, 514–527.

Sauquet H, Weston PH, Anderson CL *et al.* (2009) Contrasted patterns of hyperdiversification in Mediterranean hotspots. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 221–225.

Schluter D, Conte GL (2009) Genetics and ecological speciation. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 9955–9962.

Servedio MR, Van Doorn GS, Kopp M, Frame AM, Nosil P (2011) Magic traits in speciation: "magic" but not rare? *Trends in Ecology & Evolution*, **26**, 389–397.

Sloan DB, Taylor DR (2010) Testing for selection on synonymous sites in plant mitochondrial DNA: the role of codon bias and RNA editing. *Journal of Molecular Evolution*, **70**, 479–491.

Smith SA, Donoghue MJ (2008) Rates of molecular evolution are linked to life history in flowering plants. *Science*, **322**, 86–89.

Soltis DE, Albert VA, Leebens-Mack J *et al.* (2009) Polyploidy and angiosperm diversification. *American Journal of Botany*, **96**, 336–348.

Spielman D, Brook BW, Frankham R (2004) Most species are not driven to extinction before genetic factors impact them. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 15261–15264.

Strasburg JL, Sherman NA, Wright KM *et al.* (2012) What can patterns of differentiation across plant genomes tell us about adaptation and speciation? *Philosophical transactions of the Royal Society of London Series B: Biological sciences*, **367**, 364–373.

Stubbe W (1989) Oenothera—An ideal system for studying the interactions of genome and plastome. *Plant Molecular Biology Reporter*, **7**, 245–257.

Valente LM, Reeves G, Schnitzler J *et al.* (2010) Diversification of the African genus Protea (Proteaceae) in the Cape biodiversity hotspot and beyond: equal rates in different biomes. *Evolution*, **64**, 745–760.

Venditti C, Pagel M (2010) Speciation as an active force in promoting genetic evolution. *Trends in Ecology & Evolution*, **25**, 14–20.

Verdú M (2002) Age at maturity and diversification in woody angiosperms. *Evolution*, **56**, 1352–1361.

Webster AJ, Payne RJH, Pagel M (2003) Molecular phylogenies link rates of evolution and speciation. *Science*, **301**, 478.

Welch J (2004) Accumulating Dobzhansky-Muller incompatibilities: reconciling theory and data. *Evolution*, **58**, 1145–1156.

Weston P, Barker N (2006) A new suprageneric classification of the Proteaceae, with an annotated checklist of genera. *Telopea*, **11**, 314–344.

Wicke S, Schneeweiss GM, DePamphilis CW, Müller KF, Quandt D (2011) The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Molecular Biology*, **76**, 273–297.

Wilcoxon F, Katti SK, Wilcox RA (1963) *Critical values and probability levels for the Wilcoxon rank sum test and the Wilcoxon signed rank test*. Lederle Laboratories Division, American Cyanamid Company.

Woolfit M (2009) Effective population size and the rate and pattern of nucleotide substitutions. *Biology Letters*, **5**, 417–420.

Woolfit M, Bromham L (2003) Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. *Molecular Biology and Evolution*, **20**, 1545–1555.

Wright S, Keeling J, Gillman L (2006) The road from Santa Rosalia: a faster tempo of evolution in tropical climates. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 7718–7722.

Wu C-I, Ting C-T (2004) Genes and speciation. *Nature Reviews Genetics*, **5**, 114–122.

Xiong A-S, Peng R-H, Zhuang J *et al.* (2009) Gene duplication, transfer, and evolution in the chloroplast genome. *Biotechnology Advances*, **27**, 340–347.

Yang L, Gaut BS (2011) Factors that contribute to variation in evolutionary rate among Arabidopsis genes. *Molecular Biology and Evolution*, **28**, 2359–2369.

Zwickl DJ (2006) GARLI: genetic algorithm for rapid likelihood inference. *See http://www.bio.utexas.edu/faculty/antisense/garli/Garli.html*.

Chapter 4 – The link between molecular evolution and macroevolution causes a bias in phylogenetic estimates of diversification rates

Author contributions:

DD devised the idea for the study. DD, LB, and Xia Hua designed the study. Xia Hua designed the theoretical background for simulations. David A. Duchêne wrote the simulations framework, and collected and analysed the data. David A. Duchêne wrote the paper; with input from Lindell Bromham and Xia Hua.

4.1 Introduction

Molecular phylogenies are increasingly being used to estimate patterns of diversification through time. Several biological processes have been studied using estimates of diversification rate dynamics, including adaptive radiations (Rabosky & Lovette 2008; McPeek 2008; Etienne *et al.* 2012), diversification mediated by the environment (e.g. Lovette and Bermingham 1999; Hughes and Eastwood 2006; Antonelli et al. 2009; Pigot et al. 2010), or protracted speciation events (Etienne & Rosindell 2012; Etienne *et al.* 2014). Inferences of diversification rate dynamics have been popular in the last decade and are often biologically important, but they rely on accurate estimates of phylogenetic node heights. In this study, we focus on how among-lineage variation in the rate of diversification and molecular evolution affects estimates of rates of diversification through time.

Despite the importance of methods to estimate diversification rate dynamics in biology, the estimates have several potential sources of bias. For example, a misleading signal of

diversification rate slowdown can be caused in several scenarios, including non-random sampling of species (Cusimano & Renner 2010), phylogenies that are extremely large (Pennell *et al.* 2012), or under-parameterised models of substitution (Revell *et al.* 2005). In the latter scenario, for instance, an overly simplistic model of substitution fails to identify substitutions in deep nodes of the phylogeny. Meanwhile, such a simple model will identify gross differences between close relatives. This results in longer branches near the present compared to the past, and therefore causes a misleading pattern of slow diversification near the present (Revell *et al.* 2005).

Models of molecular rates among lineages can also be a source of bias; for example, in data sets that have large amounts of molecular rate variation among lineages (Wertheim *et al.* 2012), or if analyses have inappropriate node-age priors (Condamine *et al.* 2015). However, the literature estimating diversification rate dynamics has placed little attention to the potential biases introduced during the estimation of phylogeny. Here we ask whether variation in the rate of molecular evolution can influence the accuracy of reconstructed patterns of diversification rates over time, particularly the inference of diversification rate speedup or slowdown over time. We focus on a particular pattern of rate variation that might be of concern to researchers using phylogenies to reconstruct diversification rates: the positive association between the rate of diversification and the rate of molecular evolution (e.g. Webster *et al.* 2003; Lanfear *et al.* 2010).

The possible causes for this link include a direct effect of the speciation process on molecular evolution (Webster *et al.* 2003; Pagel *et al.* 2006; Venditti & Pagel 2010), molecular evolution as a driver of diversification (Lanfear *et al.* 2010), or a third factor that independently drives both the rate of diversification and the rate of molecular evolution (e.g. Davies et al. 2004; Allen et al. 2006). Despite the debate over what

drives the link between the rate of diversification and the rate of molecular evolution, there has been little attention to the implications of the link for evolutionary analyses that rely on estimates using molecular phylogenies. For example, timescale estimates can be compromised when there is a time period with exceptionally high rates of diversification and high rates of molecular evolution, and therefore a large number of molecular substitutions compared to other parts of the evolutionary history. The additional molecular substitutions can lead to an overestimation of the timescale in these lineages, leading to a failure to identify a period of fast diversification.

Here we ask whether variation in the rate of molecular evolution can influence the accuracy of reconstruction of patterns of diversification rate over time, particularly the inference of diversification speedup or slowdown over time. We focus on a particular pattern of rate variation that might be of concern to researchers using phylogenies to reconstruct diversification rate: the positive association between rate of diversification and rate of molecular evolution. Therefore, we use simulations to explore whether commonly used molecular phylogenetic estimates of diversification rate are influenced by an association between substitution rate and diversification rate. We simulate the evolution of DNA sequences under a number of models that allow both speciation rate and substitution rate to vary. We use the resulting DNA sequence to reconstruct the phylogeny, which we then used in a number of common methods for analysing diversification rates through time (Appendix – Chapter 4). Then we ask whether these commonly used phylogenetic methods can correctly recover patterns in the timing of diversification events.

4.2 Methods

Our aim in this study is to ask whether variation in rate of molecular evolution might influence the detection of patterns of diversification events over time from molecular phylogenetic analysis. We restrict our investigation to a defined set of macroevolutionary models that describe the relationship between rates of molecular evolution and patterns of diversification. In this section, we first describe these models, then we explain how we parameterised those models using data from the literature in order to provide realistic simulations. Then we describe how we simulated the evolution of DNA sequences under different models of rate variation and temporal patterns of diversification. We use these simulated sequences to reconstruct the evolutionary history using commonly employed phylogenetic methods, and compare the reconstructions to the known history of the sequences. Because we wish to ask whether these models that link diversification rates to molecular evolution rates could change our view of macroevolutionary processes, we use a number of popular methods for detecting the pattern of diversification events over time, and ask whether the reconstructed data gives an accurate picture of the true history of the sequences.

4.2.1 Macroevolutionary models

Using a birth-death model with constant extinction, we simulated phylogenies with six models of variation in the rate of substitutions and the rate of speciation through time:

(A) *Variable rates.* Speciation rate and substitution rate both change stochastically but independently of each other.

(B) *Linked rates.* Speciation rate and substitution rate both change stochastically, but increase or decrease together according to a positive relationship between them.

(C) *Slowdown.* Speciation rates tend to decrease over time, but the substitution rate varies stochastically and is independent from the speciation rate.

(D) *Speedup.* Speciation rates tend to increase over time, but the substitution rate varies stochastically and is independent from the speciation rate.

(E) *Linked slowdown.* Speciation rates have a tendency to decrease over time, and decreases are associated with decrease in substitution rates.

(F) *Linked speedup.* Speciation rates have a tendency to increase over time, and increases are associated with increase in substitution rates.

4.2.2 Simulation of phylogenies and sequences

In order to make our simulations representative of typical macroevolutionary studies, we selected parameters that are reasonable representations of published studies. To determine an appropriate number of taxa per phylogeny, sequence length, and reconstruction methods, we sampled one hundred published studies from the last ten years that used phylogenetic analyses to detect changes in diversification rate (Appendix – Chapter 4). Based on the median values from this sample, we set the size of our simulated dataset to be 150 sequences of length 4000 bases. We also selected two of the most commonly used phylogeny reconstruction methods from the studies included in our survey: nonparametric rate smoothing (NPRS; Sanderson 1997) and Bayesian "relaxed clock" phylogenies, as implemented in BEAST (Drummond *et al.* 2006).

Our simulations require a set value for extinction rate and a starting value for speciation rate and molecular evolution rate. In order to condition our simulations to reasonable values, we used published estimates of these parameters for birds, as an example taxon. In all simulations we included a constant background extinction rate of 0.01 per time

step per lineage, which is approximately the mean extinction rate reported in the past for birds (Jetz *et al.* 2012). Speciation rates are drawn from a distribution with a maximum of 4.64 per million years per lineage, which represents the highest rate of diversification estimated for family-level bird lineages (Jetz *et al.* 2012). The minimum speciation rate was 0.05. The maximum rate of molecular evolution was set to 0.01 substitutions per site per million years, based on estimates for avian nuclear genes (van Tuinen & Hedges 2001), and the minimum rate of molecular evolution was set to 1×10^{-7} substitutions per site per million years.

All simulations began with the same initial rate of molecular evolution, 2.47×10^{-3} substitutions per site per million years, which is typical substitution rate estimates for avian nuclear genes (van Tuinen & Hedges 2001), and is similar to some mitochondrial gene estimates for birds (Pereira and Baker 2006). We chose a starting speciation rate of 0.10, which is approximately the median speciation rate for the birds according to the estimates by Jetz *et al.* (2012). Each of our simulations follows a stochastic process forwards in time, with discrete time steps of 0.1 million years. At each time step, sequentially, we used a probability distribution to determine whether a speciation, extinction, or substitution event occurred. Speciation and substitution rates were sampled from a multivariate distribution with variables on a log-log scale, which has been used to describe the relationship between rates of diversification and rates of molecular evolution (Lanfear *et al.* 2010). With these parameters, our final alignments had between 200 and 1000 variable sites. The following equations describe the total rate of an event of any kind, ρ, and the probability for each of the events occurring at a given time step of size *dt*:

$$\rho = \lambda + \varepsilon + sl$$

$$P(event) = \rho e^{-\rho dt}$$

$$P(speciation \mid event) = \frac{\lambda}{\rho}$$

$$P(extinction \mid event) = \frac{\varepsilon}{\rho}$$

$$P(substitution \mid event) = \frac{sl}{\rho}$$

Where λ is the speciation rate, ε is the extinction rate, *s* is the substitution rate, and *l* is the sequence length. The probability of each of these events was defined at each time step until the desired number of extant taxa was reached. The probability for each kind of substitution followed a general time reversible model using empirically derived parameters (Murphy *et al.* 2001). This simulation procedure stops when the desired number of species is reached. This means that it ignores the possibility that another combination of birth and death parameters would lead back to the desired number of species. However, accounting for this possibility would require a different threshold to stop the simulations. Our procedure generates a phylogeny and a sequence alignment; the code is available from GitHub (https://github.com/duchene/moldivlink/).

4.2.3 Estimation of phylogenies

For each simulated alignment of 150 sequences and 4000 base pairs, we reconstructed a time-calibrated phylogeny under two common models of rate variation. The first was an autocorrelated rates model, where rates are similar among related branches (Sanderson 2002). This is implemented using non-parametric rate smoothing (NPRS) in a maximum likelihood framework using the R package "ape" (Paradis *et al.* 2004). In this method, an estimate of the optimal smoothing of rates is obtained from a sequence-based cross-validation procedure. For each analysis, we selected a value of the smoothing parameter between minus one and six on a $\log_{10}$ scale in increments of 1 with the lowest cross-validation score (Sanderson 2002). It was also necessary to estimate branch lengths in substitutions per site independently beforehand, which was

done in the R package PHANGORN (Schliep 2011). We used a root calibration with a confidence interval of ±5% of the true root-node age.

The second model of rate variation was an uncorrelated rates model, where rates are independent and identically distributed using a lognormal distribution (Drummond *et al.* 2006), implemented in a Bayesian framework using the software BEAST 2.2 (Bouckaert *et al.* 2014). We maintained the true topology for each analysis constant. We used a constant rates birth-death tree prior and a normally distributed root calibration prior with a standard deviation of 5% of the true root-node age. All Bayesian analyses had a chain length of 20 million steps, sampled every two thousand steps. We discarded the burn-in after 2 million steps and visually checked for convergence of the likelihood. We also calculated effective sample sizes using the R package CODA (Plummer *et al.* 2006), and only considered analyses where all the parameters had values above 200. In addition, we ran 10 simulations for every scheme four times, and assessed that the inferences of molecular rates were similar to confirm convergence to a stationary distribution. We used the maximum clade credibility tree from each posterior distribution for subsequent analyses. We chose the general time-reversible substitution model for all estimation analyses, which is the best-suited model to the substitution matrix used to simulate the data.

4.2.4 Macroevolutionary metrics from estimated phylogenies

We had a true tree (the known history of the sequences) and a reconstructed tree (the inferred history of the sequences given the phylogenetic reconstruction method) for one hundred simulated data sets for each of the six models. We then wanted to ask whether commonly used macroevolutionary analysis methods can describe the pattern of diversification that gave rise to those sequences. Although a wide variety of methods

are used to detect changes in diversification rate from phylogenies, we chose two representative methods from our sample of publications, plus one more recent method for comparison.

The first method is a metric-based method. The gamma statistic (used in one third of the sampled studies) describes the relative distribution of nodes in a phylogeny. Since the gamma statistic describes the ages of nodes in a phylogeny, we can also use it to assess the accuracy of methods of inference to estimate the overall timing of divergence times in a given phylogeny. This statistic is usually compared to phylogeny with a constant diversification process with no extinction (Yule process), and this comparison is often referred to as the constant rates (CR) test (Pybus & Harvey 2000). Under a constant speciation rate with no extinction (Yule process), nodes will accumulate exponentially from the root to the tip, which give a gamma value distributed around zero. A slowdown in speciation rate over time is expected to produce a phylogeny with an excess of nodes in the early part of the history. Since gamma lies on a standard normal distribution, gamma values below the 95% of total density of values will be below -1.645. Similarly, a speedup in speciation rate will result in more nodes near the tips, such that scenarios with speedup yield values of gamma above 1.645. These criteria to determine the underlying diversification process have been used in a large number of studies in recent years (Supplementary information). We calculated gamma for both the true tree (the expected value of gamma, γ_{exp}) and the reconstructed tree (the observed value of gamma, γ_{obs}) using the R package "ape" (Paradis *et al.* 2004).

The second method is a maximum likelihood model-fitting procedure. There are a range of methods but we selected the most commonly used implementation for these tests, from the R package 'laser' (Rabosky 2006b), which was used in over a quarter of the

sampled studies. This approach starts by fitting alternative diversification models: a birth-death model (bd) with constant rates of speciation and extinction over the phylogeny, a slowdown model with decreasing rates of speciation (dec), and a speedup model with increasing rates of speciation (inc). Then, the method tests whether the model with constant rates can be rejected in favour of either of the other two models (Rabosky 2006a). This is done by approximating the distribution of the Akaike Information Criterion (AIC) under the model with constant rates. The constant rates model is rejected if the AIC for a variable rates model falls outside the 95th percentile of the approximated distribution (Rabosky 2006a).

The third method is a Bayesian model-fitting procedure. We selected a recently proposed approach implemented in the R package 'TESS' (Höhna *et al.* 2015). We used this method to assess the same three models as in 'laser'. We used stepping stone sampling to estimate marginal likelihoods for each model, which is an estimate of model fit (Xie *et al.* 2011). This model fitting requires the user to provide priors for the parameters, which can be informative about the expected parameter values. We aimed to give prior information the minimum opportunity to override the signal from the data, so we used broad priors that provide little information about each parameter. The constant rates birth-death model was used with two parameters: diversification rate (prior exp{rate = 10}) and extinction rate (prior exp{rate = 10}). The other models were used with three parameters: the extinction rate (prior exp{rate = 0.1}), the initial rate of speciation (prior exp{rate = 10}), and the rate of decay of speciation rates through time (prior exp{rate = 0.1}). Following standard practice in Bayesian statistics, we rejected the constant rates birth-death model if the ratio between its marginal likelihood and the competing model (the bayes factor) was >5 (Kass & Raftery 1995; Lartillot & Philippe 2006).

Because we wish to know whether the phylogenetic estimates of the pattern of diversification accurately capture the true history of the sequences, we applied the three methods for estimating diversification process to both the true tree (produced by the simulation) and the reconstructed tree (produced by analysing the simulated sequences) for each simulated dataset. We then ask whether the reconstructed phylogeny accurately reflects the underlying diversification process. More specifically, we asked: (i) is the underlying diversification process identified correctly in the simulated (true) phylogenies? (ii) Is the underlying process identified correctly in the reconstructed phylogenies (from NPRS and BEAST)? And (iii) do we identify the underlying process with similar accuracy when the rates of molecular substitution and rates of diversification are linked (models D and F)?

In models A and B, the speciation rate varies randomly over the tree (Table 4.1), so we do not expect to reject the constant rates birth-death process. We expect an average gamma value with no speedup or slowdown, so that it falls between -1.645 and 1.645, and we expect that model selection using laser or TESS will not reject the constant rates birth-death process. For the slowdown models, C and D, speciation rates have a trend of decreasing over time (Table 4.1). We expect in these cases a gamma statistic less than -1.645, and we expect that model selection with laser and TESS will reject the constant rates birth-death model in favour of speciation rate slowdown. For the speedup models, E and F, speciation rates have a trend of increasing over time. We expect in these cases a value for gamma above 1.645, and we expect that model selection with laser and TESS to reject the constant rates birth-death model in favour of speciation rate speedup.

For each simulation and reconstruction scenario (true trees, and reconstructions under NPRS and BEAST), we report how often each method correctly detects the expected model (i.e. the proportion of times that the expected model is detected). To compare these proportions across scenarios of simulation or reconstruction, we calculated the confidence interval using a normal approximation for sample proportions. We considered a difference to be significant when there was no overlap in the confidence intervals of two given scenarios.

Table 4.1: Macroevolutionary models used to generate simulated datasets.

Model	Description		Speciation rates	Molecular rate
A	Stochastic-Unlinked	STU	Stochastic variation	Stochastic variation
B	Stochastic-Linked	STL	Stochastic variation	Linked to speciation rate
C	Slowdown-Unlinked	SLU	Decrease over time	Stochastic variation
D	Slowdown-Linked	SLL	Decrease over time	Linked to speciation rate
E	Speedup-Unlinked	SPU	Increase over time	Stochastic variation
F	Speedup-Linked	SPL	Increase over time	Linked to speciation rate

4.3 Results

Our first step was to investigate how often each of the three macoevolutionary methods identified the expected process in the simulated (true) phylogenies. In these analyses on the true trees, we find that the three methods frequently fail to detect the expected model of speciation rates through time. When speciation rates vary stochastically, under models A and B, gamma reaches the expected outcome ~80% of the time (Table 4.2). Under models A and B, laser produced the expected outcome only 50% of the time, while TESS always produced the expected outcome. When speciation rates slowdown over time, under models C and D, gamma and laser reject constant rates birth-death in favour of slowdown around 70% of the time, while TESS only produced the expected result around 10% of the time. Speciation rate speedup over time, under models E and F, was the best identified by the three methods. Gamma and laser identified speedup around 70% of the time, while TESS always identified it correctly. From these results we conclude that, within the parameters explored in this simulation study, the macroevolutionary methods applied to reconstructed trees can detect slowdown and speedup in speciation rate with low power, and this performance is different depending on the method.

Next, we wished to know how often the macroevolutionary methods can identify the expected process in the phylogenies reconstructed using NPRS or BEAST. Here, we examined the results from scenarios where speciation and molecular substitution rates vary, but were not forced to have an association between them (models A, C, and E; Table 4.1). We find that the expected macroevolutionary process is identified less often in reconstructed trees than it is when using the true tree (Table 4.2). This difference is often not significant, suggesting that reconstruction does not always distort the

underlying process. Gamma and laser identified the expected process for models A between 51% and 71% of the time, and half of the time for models C and E. Interestingly, when using the method in laser, speedup was correctly identified significantly more often in NPRS reconstructions than in those from BEAST, and slowdown was correctly identified more often in BEAST reconstructions than in those from NPRS. Although the latter comparison was not significant, the macroevolutuonary inferences using gamma show similar patterns, suggesting that different reconstruction methods might have different power to identify specific macroevolutionary processes. TESS always identified the expected process for models A and E. Under model C with speciation rate slowdown, however, the expected process was only detected by TESS less that 25% of the time, which is similar to the results using true trees and echoes previous findings that this pattern is common in trees with <200 tips (Höhna 2014).

The last question we asked was whether a link between the rate of speciation and molecular substitution caused a change in our ability to detect the underlying diversification process. When speciation rates varied stochastically under models A and B, there was no significant difference across macroevolutionary methods and phylogenetic reconstruction methods in how often the expected process was identified (Table 4.2). This suggests that the association between rates of speciation and molecular substitution alone does not change our ability to infer diversification rate patterns through time. However, when speciation rates had a tendency to decrease or increase through time, the underlying process was identified significantly less often across macroevolutionary methods and phylogenetic reconstruction methods (Table 4.2). From these results we can conclude that the power to detect changes in diversification rate is greatly reduced when molecular rates and speciation rates are linked. In fact, the association between rates of speciation and molecular substitution can even lead to the

detection of an opposite macroevolutionary pattern to that which is expected from the underlying process (Appendix – Chapter 4).

Table 4.2. Proportion of time that the expected model (see Methods) is selected for each of the macroevolutionary methods (gamma, laser, and TESS), for each of the six models of simulation of the link between the rates of speciation and molecular substitutions (Table 4.1), and each of the three sources of phylogeny (simulated or true trees, NPRS, and BEAST).

Gamma statistic

Underlying model	Simulated	NPRS	BEAST
A – Stochastic Unlinked	0.87 (0.066)	0.62 (0.095)	0.55 (0.098)
B – Stochastic Linked		0.59 (0.096)	0.64 (0.094)
C – Slowdown Unlinked	0.72 (0.088)	0.53 (0.098)	0.54 (0.098)
D – Slowdown Linked		0.27 (0.089)	0.14 (0.068)
E – Speedup Unlinked	0.7 (0.09)	0.65 (0.093)	0.46 (0.098)
F – Speedup Linked		0.33 (0.092)	0.19 (0.077)
Laser			
A – Stochastic Unlinked	0.5 (0.098)	0.71 (0.089)	0.55 (0.098)
B – Stochastic Linked		0.54 (0.098)	0.51 (0.098)
C – Slowdown Unlinked	0.71 (0.089)	0.58 (0.097)	0.66 (0.093)
D – Slowdown Linked		0.48 (0.098)	0.31 (0.091)
E – Speedup Unlinked	0.65 (0.093)	0.56 (0.097)	0.28 (0.088)
F – Speedup Linked		0.3 (0.09)	0.15 (0.07)
TESS			
A – Stochastic Unlinked	1 (0)	1 (0)	1 (0)
B – Stochastic Linked		1 (0)	1 (0)
C – Slowdown Unlinked	0.12 (0.064)	0.13 (0.066)	0.24 (0.084)
D – Slowdown Linked		0 (0)	0 (0)
E – Speedup Unlinked	1 (0)	1 (0)	1 (0)
F – Speedup Linked		0 (0)	0 (0)

4.4 Discussion

Estimates of phylogenetic node heights are used routinely to estimate diversification rate dynamics. For this reason it is critical to use realistic simulation conditions to examine the behaviour of methods to estimate both phylogeny (e.g. Drummond *et al.* 2006; Duchêne *et al.* 2014) and diversification rate dynamics (e.g. Revell et al. 2005; Cusimano and Renner 2010; Pennell et al. 2012). Studies that have assessed the methods to estimate diversification rate dynamics have largely assumed that the phylogeny is known without error. However, this approach excludes the potential biases that occur in phylogenetic estimation using molecular data. Our analyses show that estimates of diversification rate dynamics can be misleading when there is a link between the rate of speciation and the rate of substitution. This can have substantial impact on the interpretation of previous studies and on future practice to assess and use methods in macroevolution using phylogenies.

The association between the rate of speciation and the rate of substitution causes a loss in the signal of diversification rate slowdown, as well as a loss in the signal of diversification rate speedup. One likely explanation is that periods with high rates of diversification will contain an overrepresentation of substitutions relative to other branches. Without a time-calibration to inform the model about this change in the rate of molecular evolution, the branches with fast diversification will be given underestimated rates. This means that the branch lengths in periods of fast diversification will be overestimated, reducing the signal of a change in diversification rates. This phenomenon is in effect a kind of heterotachy, where the model is unable to account for the form of among-lineage rate variation in the data.

It is less common for empirical studies to detect a general trend of diversification rate speedup than a trend of slowdown. In fact, there are several known biases that can cause a signal of diversification rate slowdown (e.g. Revell et al. 2005; Cusimano and Renner 2010; Pennell et al. 2012), instead of removing it as observed in our results. One implication of our analyses is that despite the large number of studies assessing diversification rate dynamics, the true prevalence of diversification rate slowdown and speedup is probably unknown for many taxa. One example is the birds, for which slowdown has been inferred to occur frequently (e.g. Phillimore & Price 2008). The difficulty in detecting speedup in our results can have critical implications for studies of data that frequently have increasing rates of diversification, like data from pathogens (e.g. Koelle et al. 2006) and populations (e.g. Gilbert et al. 2008). However, the methods used in this study are different to those usually used in analyses of pathogens and population-level data. Further investigation is needed to understand whether other estimation methods also present difficulty in detecting diversification rate slowdown or speedup.

As there are multiple potential sources of bias when inferring diversification rate dynamics, including the phenomena shown here, special caution is necessary when interpreting many of the inferences from previous studies. A critical recommendation for future studies is to assess the adequacy of the substitution model instead of simply selecting the model with the best fit, which might prevent the bias from using under-parameterised evolutionary models (Revell *et al.* 2005). Regarding our findings, however, it is also recommended to assess the molecular rate variation model and time-calibrations for conflict with the molecular data. More research into methods that assess model performance is necessary, and their usage is yet to become mainstream practice when estimating phylogeny for studies in macroevolution. Furthermore, future studies

that assess the performance of the various existing molecular clock to infer divergence times would be interesting, and shed light into the source of the bias caused by the link between the rate of molecular evolution and the rate of diversification. Ideally, however, the methods to estimate timescales and diversification rates through time should not be independent, but should be integrated in a fully Bayesian framework. These methods should also integrate the link between the rate of molecular evolution and the rate of diversification. This integrated approach is more appropriate because it avoids the compounding of multiple methodological sources of error become intractable.

It is important to consider that several factors that we have controlled in our analyses can influence the performance of phylogenetic and macroevolutionary analyses. These factors include the quality of calibration data (Duchêne *et al.* 2014), or the adequacy of the clock and substitution models (Revell *et al.* 2005; Phillips 2009). Similarly, the methods to make macroevolutionary inferences that we have examined have been largely replaced by more sophisticated approaches (e.g. Morlon *et al.* 2010; Etienne *et al.* 2012), and the data sets in empirical studies have a tendency to increase in size. However, our analyses make it clear that even new methods should be scrutinized carefully before being used widely for studies in evolutionary biology.

Methods to infer phylogenetic timescales and diversification rate dynamics have become an integral component of the tools for macroevolutionary analysis. These methods should be assessed regularly for robustness to new insights about molecular evolution. Common approaches to estimate diversification rate dynamics during the last decade are often not robust to the link between the rate of speciation and rate of substitution. As inferences of diversification rate dynamics are susceptible to several sources of bias, significant caution is necessary when interpreting the results from these

methods. We recommend that simulations to assess macroevolutionary methods be performed in the form of genetic alignment instead of phylogeny; in other words, that by simulating processes on alignments instead of directly on phylogenies, as is common practice. This is because a simulated phylogeny represents an estimate that contains error. Using simulated alignments will allow for more detailed assessment of the sensitivity of methods to each of the components of the evolutionary process, like variation in evolutionary rates across lineages, or across sites in the alignment (e.g. Revell et al. 2005; Phillips 2009).

4.5 References

Allen AP, Gillooly JF, Savage VM, Brown JH (2006) Kinetic effects of temperature on rates of genetic divergence and speciation. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 9130–9135.

Antonelli A, Nylander JAA, Persson C, Sanmartín I (2009) Tracing the impact of the Andean uplift on Neotropical plant evolution. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 9749–9754.

Bouckaert R, Heled J, Kühnert D *et al.* (2014) BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, **10**, e1003537.

Condamine FL, Nagalingum NS, Marshall CR, Morlon H (2015) Origin and diversification of living cycads: a cautionary tale on the impact of the branching process prior in Bayesian molecular dating. *BMC Evolutionary Biology*, **15**, 65.

Cusimano N, Renner SS (2010) Slowdowns in diversification rates from real phylogenies may not be real. *Systematic Biology*, **59**, 458–464.

Davies TJ, Savolainen V, Chase MW, Moat J, Barraclough TG (2004) Environmental energy and evolutionary rates in flowering plants. *Proceedings of the Royal Society B: Biological Sciences*, **271**, 2195–2200.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biology*, **4**, e88.

Duchêne S, Lanfear R, Ho SYW (2014) The impact of calibration and clock-model choice on molecular estimates of divergence times. *Molecular Phylogenetics and Evolution*, **78**, 277–289.

Etienne RS, Haegeman B, Stadler T *et al.* (2012) Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 1300–1309.

Etienne RS, Morlon H, Lambert A (2014) Estimating the duration of speciation from phylogenies. *Evolution*, **68**, 2430–2440.

Etienne RS, Rosindell J (2012) Prolonging the past counteracts the pull of the present: protracted speciation can explain observed slowdowns in diversification. *Systematic Biology*, **61**, 204–213.

Gilbert MTP, Drautz DI, Lesk AM *et al.* (2008) Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 8327–8332.

Höhna S (2014) Likelihood inference of non-constant diversification rates with incomplete taxon sampling. *PLoS one*, **9**, e84184.

Höhna S, May MR, Moore BR (2015) TESS: an R package for efficiently simulating phylogenetic trees and performing Bayesian inference of lineage diversification rates. *Bioinformatics*, **32**, 789–791.

Hughes C, Eastwood R (2006) Island radiation on a continental scale: exceptional rates of plant diversification after uplift of the Andes. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 10334–10339.

Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO (2012) The global diversity of birds in space and time. *Nature*, **491**, 444–448.

Kass RE, Raftery AE (1995) Bayes Factors. *Journal of the American Statistical*

Association, **90**, 773–795.

Koelle K, Cobey S, Grenfell B, Pascual M (2006) Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans. *Science*, **314**, 1898–903.

Lanfear R, Ho SYW, Love D, Bromham L (2010) Mutation rate is linked to diversification in birds. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 20423–20428.

Lartillot N, Philippe H (2006) Computing Bayes factors using thermodynamic integration. *Systematic Biology*, **55**, 195–207.

Lovette IJ, Bermingham E (1999) Explosive speciation in the New World Dendroica warblers. *Proceedings of the Royal Society B: Biological Sciences*, **266**, 1629–1636.

McPeek M a (2008) The ecological dynamics of clade diversification and community assembly. *The American Naturalist*, **172**, 270–84.

Morlon H, Potts MD, Plotkin JB (2010) Inferring the dynamics of diversification: a coalescent approach. *PLoS Biology*, **8**, 9.

Murphy WJ, Eizirik E, O'Brien SJ *et al.* (2001) Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science*, **294**, 2348–2351.

Pagel M, Venditti C, Meade A (2006) Large punctuational contribution of speciation to evolutionary divergence at the molecular level. *Science*, **314**, 119–121.

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

Pennell MW, Sarver BAJ, Harmon LJ (2012) Trees of unusual size: biased inference of early bursts from large molecular phylogenies. *PloS One*, **7**, e43348.

Phillimore AB, Price TD (2008) Density-dependent cladogenesis in birds. *PLoS Biology*, **6**, e71.

Phillips MJ (2009) Branch-length estimation bias misleads molecular dating for a vertebrate mitochondrial phylogeny. *Gene*, **441**, 132–140.

Pigot AL, Phillimore AB, Owens IPF, Orme CDL (2010) The shape and temporal dynamics of phylogenetic trees arising from geographic speciation. *Systematic Biology*, **59**, 660–673.

Plummer M, Best N, Cowles K, Vines K (2006) CODA: convergence diagnosis and output analysis for MCMC. *R News*, **6**, 7–11.

Pybus OG, Harvey PH (2000) Testing macro-evolutionary models using incomplete molecular phylogenies. *Proceedings of the Royal Society B: Biological Sciences*, **267**, 2267–2272.

Rabosky DL (2006a) Likelihood methods for detecting temporal shifts in diversification rates. *Evolution*, **60**, 1152–1164.

Rabosky DL (2006b) LASER: a maximum likelihood toolkit for detecting temporal shifts in diversification rates from molecular phylogenies. *Evolutionary Bioinformatics Online*, **2**, 273–276.

Rabosky DL, Lovette IJ (2008) Density-dependent diversification in North American wood warblers. *Proceedings of the Royal Society B: Biological Sciences*, **275**, 2363–2371.

Revell L, Harmon L, Glor R (2005) Under-parameterized model of sequence evolution

leads to bias in the estimation of diversification rates from molecular phylogenies. *Systematic Biology*, **54**, 973–983.

Sanderson MJ (1997) A nonparametric approach to estimating divergence times in the absence of rate constancy. *Molecular biology and evolution*, **14**, 1218–1231.

Sanderson MJ (2002) Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology and Evolution*, **19**, 101–109.

Schliep KP (2011) PHANGORN: phylogenetic analysis in R. *Bioinformatics*, **27**, 592–593.

van Tuinen M, Hedges SB (2001) Calibration of avian molecular clocks. *Molecular Biology and Evolution*, **18**, 206–213.

Venditti C, Pagel M (2010) Speciation as an active force in promoting genetic evolution. *Trends in Ecology & Evolution*, **25**, 14–20.

Webster AJ, Payne RJH, Pagel M (2003) Molecular phylogenies link rates of evolution and speciation. *Science*, **301**, 478.

Wertheim JO, Fourment M, Kosakovsky Pond SL (2012) Inconsistencies in estimating the age of HIV-1 subtypes due to heterotachy. *Molecular Biology and Evolution*, **29**, 451–456.

Xie W, Lewis PO, Fan Y, Kuo L, Chen M-H (2011) Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Systematic Biology*, **60**, 150–160.

Chapter 5 – Tree imbalance causes a bias in phylogenetic estimation of evolutionary timescales using heterochronous sequences

A version of this chapter was published as:

Duchêne D, Duchêne S, Ho SYW (2015) Tree imbalance causes a bias in phylogenetic estimation of evolutionary timescales using heterochronous sequences. *Molecular Ecology Resources*. **15**, 785-794

Author contributions:

DD devised the idea for the study. DD, SD, and SYWH designed the simulations framework and analyses. DD made the simulations, collected and analysed the data. DD wrote the paper; with input from SD and SYWH.

5.1 Introduction

Phylogenetic analyses of genetic data are widely used to estimate evolutionary rates and divergence times. This can be done using methods based on molecular clocks, calibrated using independent information about time. The resulting estimates of dated evolutionary tree, or chronograms, form the basis of a wide range of subsequent analyses and interpretations. Accordingly, understanding the performance and behaviour of molecular-clock methods is important for reliable evolutionary and ecological inferences.

Estimates of evolutionary rates and timescales can be affected by various sources of error, such as misspecification of the calibrations, substitution model, or clock model (e.g. Ho *et al.* 2005b; Revell *et al.* 2005; Lepage *et al.* 2007; Duchêne *et al.* 2014b). Long branches present a well known source of bias in phylogenetic inference (Felsenstein 1978; Hendy & Penny 1989; Huelsenbeck 1995; Parks & Goldman 2014), but their impact on molecular-clock analyses remains unclear. It is critical to investigate the features of datasets that might lead to biases in phylogenetic estimates of timescales. Examining the effects on estimation error is also important, because large variances can reduce the power to test evolutionary and biogeographic hypotheses (Crisp *et al.* 2011).

Tree imbalance is a property of phylogenetic trees that has received considerable attention, although not in the context of estimating molecular evolutionary timescale. The property of tree imbalance refers to the relative number of tips descending from internal nodes in the tree. In a completely balanced tree, each of the two lineages descending from any internal node leads to the same number of tips. A completely imbalanced tree has a pectinate or comb-like arrangement of branches. Imbalanced ultrametric trees tend to have more long branches than do balanced trees, because some of the lineages will have few descendants. The imbalance of nodes is widely used to estimate variations in net diversification rates (Ricklefs 2007). In practice, however, many trees exhibit levels of imbalance that exceed those predicted on the basis of biologically plausible differences in diversification rates (Heard 1996). This might be due to stochastic events during macroevolution, such as mass extinctions (Heard & Mooers 2002; Blum & François 2006).

Understanding the effects of tree imbalance on phylogenetic inference is important because many studies of ecology and evolution rely on accurate estimates of phylogenetic relationships and divergence times (Mooers & Heard 1997; Blum & François 2006). Imbalance has been shown to increase with the tree depth in several metazoan and plant groups (Purvis & Agapow 2002). This is partly attributed to biological drivers of diversification-rate variation, and to the cumulative effect of differences in diversification rates between lineages (Holman 2005). In clades with a history of variable diversification rates, incomplete taxon sampling also tends to increase tree imbalance (Mooers 1995; Heath *et al.* 2008). In addition, tree imbalance can arise from biased in taxon sampling (Mooers & Heard 1997; Blum & François 2006), and from particular implementations of methods for tree inference (Holton *et al.* 2014).

Tree imbalance can affect estimates of evolutionary timescales in two different ways. First, tree imbalance might cause over- or underestimation of evolutionary rates or the amount of genetic change along different branches. A recent study identified that clock-model selection can be inaccurate for trees simulated under a Yule speciation process compared with those that are completely balanced (Baele *et al.* 2013). Although the causes of this discrepancy remain unclear, a consequence is that existing clock models will tend to have higher statistical fit for balanced trees. Another study investigated the effect of long branches on estimates of timescales by placing simulated sequences at different points along a long branch, allowing the long branch to be broken up (Magallón 2010). Although the effect on estimates of divergence times was not substantial, the study only considered a single long branch. The effect of a consistent source of long branches, such as tree imbalance, has received little attention in the context of molecular-clock analyses.

The second possible consequence of tree imbalance is that it might improve estimation of evolutionary timescales because of its effects on the relative placement of calibrating nodes. Calibrations at deep nodes are more informative in imbalanced trees because they contain a greater number of descendent branches than they do in balanced trees (Linder *et al.* 2011; Duchêne *et al.* 2014b). Therefore, in an imbalanced tree with calibrations at deep nodes, most of the nodes with unknown age are descendent from the calibrating nodes. This phylogenetic relationship between calibrated and non-calibrated nodes might be particularly important when evolutionary rates are autocorrelated, because rates along branches that are not descendent from a calibrating node can be very different from those that are nested within the calibration (Duchêne *et al.* 2014b).

The effectiveness of the calibration scheme depends on whether the phylogenetic tree is calibrated at internal nodes (isochronous datasets) or at the tips (heterochronous datasets). In isochronous datasets, some or all of the branches in the tree descend from calibrating nodes. In heterochronous datasets, all branches are ancestral to the calibrating nodes, which are the tips of the tree (Rambaut 2000). The calibrations in these two types of dataset are fundamentally different, such that they might have disparate susceptibilities to the effects of tree imbalance. The uncertainty in the calibrations can be taken into account by using probability distributions (Ho & Phillips 2009), but this is rarely done with heterochronous datasets (Molak *et al.* 2013, 2014). In heterochronous data from viruses, there is often precise information about the sample collection time.

Any impacts of tree imbalance on methods for estimating evolutionary timescales from molecular data will have consequences for studies of molecular ecology and evolution.

A broad range of studies have been based on imbalanced phylogenetic trees, including investigations of population differentiation (e.g. Clark & Vogler 2009; Nater *et al.* 2011), the timing of dispersal at the population level (e.g. Rajabi-Maham *et al.* 2008; Korsten *et al.* 2009), the role of dispersal and geography on diversification (e.g. Carstens & Knowles 2007; Goldberg *et al.* 2011), and the causes of patterns of macroevolution (e.g. Sheng *et al.* 2014; Worobey *et al.* 2014). In addition to recovering imbalanced trees, these studies included datasets that spanned multiple evolutionary timescales and were based on various clock calibrations. The prevalence of imbalanced phylogenetic trees in these studies underscores the importance of evaluating the impact of tree balance on molecular estimates of evolutionary timescales.

Analyses of data generated via simulation offer a means of identifying the effect of tree imbalance on phylogenetic estimates. Such an approach provides information about the performance of a method under idealised conditions, and about the robustness of a method to violations of its assumptions (Huelsenbeck 1995). Here we evaluate the reliability of timescale estimates from simulated datasets with various levels of tree imbalance. We investigate this effect using different calibration schemes involving isochronous and heterochronous datasets. To examine the effects of tree imbalance in empirical data, we also present analyses of isochronous data from primates and heterochronous data from the African swine fever virus.

5.2 Methods

5.2.1 Data from simulations

We simulated the evolution of nucleotide sequences under three models of molecular rate evolution and three levels of tree imbalance. Simulations allowed us to investigate the reliability of timescale estimates while knowing the true evolutionary history of the data. Using the R package APE (Popescu *et al.* 2012), we generated chronograms with 50 tips and with a root age of 50 time units. Using a constant-size coalescent model, we generated 2000 trees, of which 1000 had isochronous tips and 1000 had heterochronous tips. We calculated the level of imbalance using the Colless index (Colless 1982). This index is based on the difference in taxon richness between the two lineages descending from each node in the phylogeny. The Colless index is a popular metric of imbalance and has been shown to have more power than other metrics, only matched in power by the Sakin index (Kirkpatrick & Slatkin 1993; Agapow & Purvis 2002).

For each of the two sets of 1000 trees we took the 100 most balanced, the 100 least balanced, and the 100 with the median balance values. This represented a total of 600 simulated trees for further analyses. To specify a desired depth in the samples of heterochronous phylogenies, we first simulated isochronous trees. We then cut the longest terminal branch in each tree such that the age of the tip was changed to the maximum calibration age chosen, which varied across three calibration schemes: deep (25 time units), medium (15 time units), and shallow (5 time units). We then halved the lengths of all but one of the other terminal branches.

We generated distributions of branch-specific substitution rates for the 600 chronograms using the R package NELSI (Ho et al. 2014), under three models of rate variation: a strict clock (SC; $\mu = 10^{-3}$), an uncorrelated lognormal relaxed clock (UCLN; log mean = -4.6, sd = 0.1; Drummond *et al.* 2006), and an autocorrelated lognormal relaxed clock (ACLN; mean = 10^{-3}, $v = 0.001$; Thorne & Kishino 2002). All rates are given in substitutions per site per time unit. In NELSI, the branch lengths of the chronograms (in units of time) are multiplied by the rate to generate phylograms (with branch lengths given in substitutions/site). We simulated the evolution of nucleotide sequences on these phylograms according to the Jukes-Cantor substitution model (Jukes & Cantor 1969), using the R package PHANGORN (Schliep 2011).

5.2.2 Empirical datasets

Analyses of simulated data are useful for testing the sensitivity of methods, but they can fail to predict how the methods will perform when presented with data that have evolved under more complex conditions. To assess the impact of phylogenetic tree imbalance in empirical data, we carried out two case studies: mitogenomic sequences from 83 primate taxa (Finstermeier *et al.* 2013) and an African swine fever virus (ASFV) dataset used in previous studies, with 50 samples obtained over a period of 50 years. We selected empirical datasets with reliable calibrations; the performance of the calibrations associated with the ASFV data has been evaluated previously (Duchêne *et al.* 2014a).

We randomly pruned the primate dataset to 50 taxa, and the ASFV dataset to 31 taxa, making sure that we kept at least a single lineage from both basal clades so that the same root node was always represented. We repeated this process 1000 times, while

ensuring that the pruned primate trees retained all of the calibrating nodes. In the ASFV dataset we always retained the youngest and oldest tips. For both datasets we selected the most balanced and the most imbalanced of the pruned trees, according to their Colless index. We analysed the datasets associated with these balanced and imbalanced trees and compared the estimates with those from the analyses of the complete datasets. The accession numbers for the empirical datasets used are available in the Appendix – Chapter 5 and the alignments have been made available in Dryad (doi:10.5061/dryad.q4545).

5.2.3 Phylogenetic estimation of timescales

To estimate the evolutionary timescale from each simulated dataset, we used Bayesian inference in the software BEAST v2.1.1 (Bouckaert *et al.* 2014). Analyses were performed using an uncorrelated lognormal relaxed clock, which performs well under different levels of among-lineage rate variation (Drummond *et al.* 2006; Lepage *et al.* 2007). We matched the substitution model to that used to simulate the data and fixed the tree to the simulated topology to exclude the effects of phylogenetic uncertainty. Three calibration schemes were used for the isochronous and heterochronous datasets. For the heterochronous datasets we chose three sampling times: at 10%, 30%, and 50% of the age of the root. For the datasets with isochronous tips, we used three different calibration depths: the two deepest nodes, the two shallowest nodes, or two nodes at depths of 1/3 and 2/3 of the age of the root. In total, we analysed data that were simulated under three levels of tree imbalance, three clock models, using three calibration schemes, and with isochronous or heterochronous sampled tips. For each of these settings we conducted 100 simulation replicates, for a total of 5400 datasets.

We selected the best-fitting substitution models according to the Bayesian information criterion for the primate and ASFV datasets. The analyses of the primate datasets were calibrated using four fossil calibrations with the following mean values: most recent common ancestor (MRCA) of Anthropoidea at 43 million years ago (Mya; $\sigma = 4.5$), MRCA of Lorsiformes at 40Mya ($\sigma = 3.0$), MRCA of Catarrhini at 29Mya ($\sigma = 6.0$), and MRCA of Papionini at 7Mya ($\sigma = 1$; Perelman *et al.* 2011). The analyses of the ASFV dataset were calibrated using the dated tips.

Each analysis was run for 50 million MCMC steps, with samples drawn every 50 thousand steps. The first 10% of the samples were discarded. We assessed sufficient sampling by verifying that the effective sample sizes of the parameter estimates were over 200. For the analyses of the primate and ASFV data, we used the maximum-clade-credibility trees to summarize the phylogenetic estimate. To evaluate the accuracy of the node-age estimates from the simulated data, we calculated the difference between the estimated and simulated node ages and the coverage probabilities of the true ages of all nodes. The coverage probability is the proportion of true node ages that are included within the 95% credible intervals of the estimates. We compared coverage probabilities across simulation schemes and among nodes through time. We also evaluated the precision in the estimates by dividing the width of the 95% credible interval by the mean age estimate for each of the nodes. The R code and some example data files are publicly available online (https://github.com/duchene/imbaltime1).

5.3 Results

In analyses of simulated data, the estimates from isochronous datasets produced coverage probabilities that were close to 1.00 for all simulation schemes (Table 5.1). The differences between the estimated and the simulated root node ages were less than 10% (Appendix – Chapter 5). Analyses of heterochronous data yielded estimates that had low coverage and were up to 30% younger than the simulated values when using shallow calibrations and when the tree was imbalanced (Table 5.1; Appendix – Chapter 5). Analyses with a strict clock were an exception, where coverage was close to 1.00, except in analyses of datasets with imbalanced trees and calibrations of median (0.91) and shallow (0.89) depths. When simulations were conducted using uncorrelated or autocorrelated relaxed-clock models and in the presence of tree imbalance, the coverage probabilities of the estimates were consistently low. The lowest coverage probabilities were observed in analyses of data with imbalanced trees and shallow calibrations based on uncorrelated (0.60) and autocorrelated (0.62) relaxed-clock models.

The pattern of precision was similar to that of accuracy, with a decrease in precision when using shallow calibrations and in the presence of tree imbalance. In analyses of isochronous datasets, however, we only observed a decrease in precision when using shallow calibrations (Table 5.2). The node-age estimates using shallow calibrations were more precise in analyses of heterochronous data than in analyses of isochronous data, but the opposite pattern was seen when using deep and median calibrations. This reflects our implementation of calibrating information; deep and median calibrations tend to be shallower for tip-calibrations than for internal-node calibrations, such that these tip-calibrations are less informative and lead to lower precision in the estimates.

On the other hand, when the isochronous and heterochronous datasets have shallow calibrations of similar depth, tip-calibrations result in highly precise estimates because they have zero uncertainty; the internal-node calibrations have an uncertainty of 10% of the true node age (Table 5.2).

	Calibrations \ Balance	Isochronous datasets			Heterochronous datasets		
		High	Medium	Low	High	Medium	Low
Strict clock	Deep	1.00	1.00	1.00	0.99	0.99	0.96
	Median	1.00	1.00	1.00	0.96	0.96	0.91
	Shallow	1.00	1.00	1.00	0.95	0.96	0.89
Uncorrelated lognormal clock	Deep	1.00	1.00	1.00	0.98	0.95	0.87
	Median	0.99	1.00	0.99	0.97	0.94	0.80
	Shallow	1.00	1.00	0.99	0.93	0.81	0.60
Autocorrelated lognormal clock	Deep	0.99	1.00	1.00	0.99	0.92	0.90
	Median	0.99	1.00	1.00	0.95	0.91	0.84
	Shallow	1.00	1.00	1.00	0.97	0.81	0.62

Table 5.1. The mean proportions of node-age estimates in isochronous (left) and heterochronous (right) data analyses for which the simulated value lies within the 95% credible interval for each simulation scheme. Results are shown for analyses with simulations performed under each of the three clock models: strict clock (top), relaxed uncorrelated lognormal clock (middle), and relaxed autocorrelated lognormal clock (bottom). Darker shades of grey indicate lower coverage probabilities.

		Isochronous datasets			Heterochronous datasets		
Balance / Calibrations		High	Medium	Low	High	Medium	Low
Strict clock	Deep	0.37	0.38	0.37	0.48	0.54	0.78
	Median	0.43	0.43	0.49	0.67	0.64	0.91
	Shallow	1.84	1.83	1.87	1.27	1.24	1.66
Uncorrelated lognormal clock	Deep	0.37	0.39	0.38	0.68	0.53	0.86
	Median	0.43	0.47	0.50	0.66	0.87	0.95
	Shallow	1.76	1.90	1.84	1.35	1.46	1.63
Autocorrelated lognormal clock	Deep	0.33	0.38	0.35	0.64	0.40	0.93
	Median	0.39	0.37	0.40	0.78	0.91	0.99
	Shallow	1.96	2.03	2.02	1.33	1.40	1.55

Table 5.2. The mean widths of the 95% credible intervals of node-age estimates in analyses of isochronous (left) and heterochronous (right) data, expressed as a proportion of node-age for each simulation scheme. Results are shown for analyses with simulations performed under each of the three clock models: strict clock (top), relaxed uncorrelated lognormal clock (middle), and relaxed autocorrelated lognormal clock (bottom). Darker shades of grey indicate wider 95% credible intervals.

In some analyses of heterochronous datasets, we found that coverage probabilities were the lowest for the estimates of the ages of nodes in deep and middle sections of the tree (Figure 5.1). This was apparent in the analyses of datasets with low overall coverage probabilities, such as those with tree imbalance, shallow calibrations, and simulated under the uncorrelated or autocorrelated relaxed-clock models. Analyses of datasets with imbalanced trees tended to underestimate the ages of most nodes (Figure 5.2), with ages as low as ~23% of the real age when calibrations were shallow (Figure 5.1).

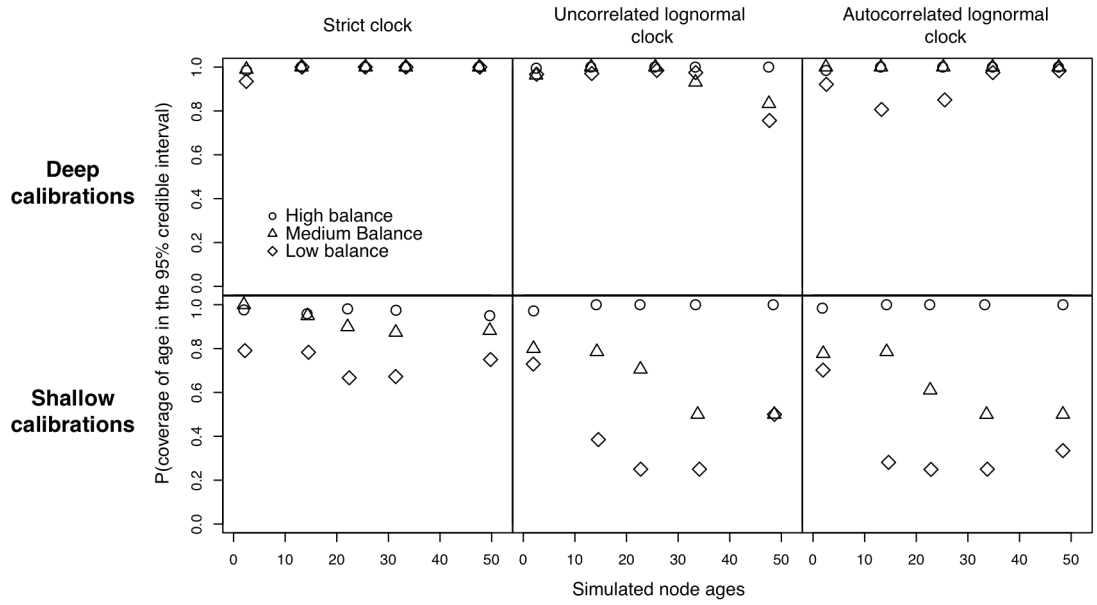


Figure 5.1. The mean proportions of node-age estimates in heterochronous data analyses for which the simulated value lies within the 95% credible interval across time depths. Nodes have been binned by their simulated age by dividing the total tree depth into five sections. Results are shown for simulation schemes with deep (top) and shallow (bottom) calibrations, and for analyses with simulations performed under each of the three clock models: strict clock (left), relaxed uncorrelated lognormal clock (centre), and relaxed autocorrelated lognormal clock (right).

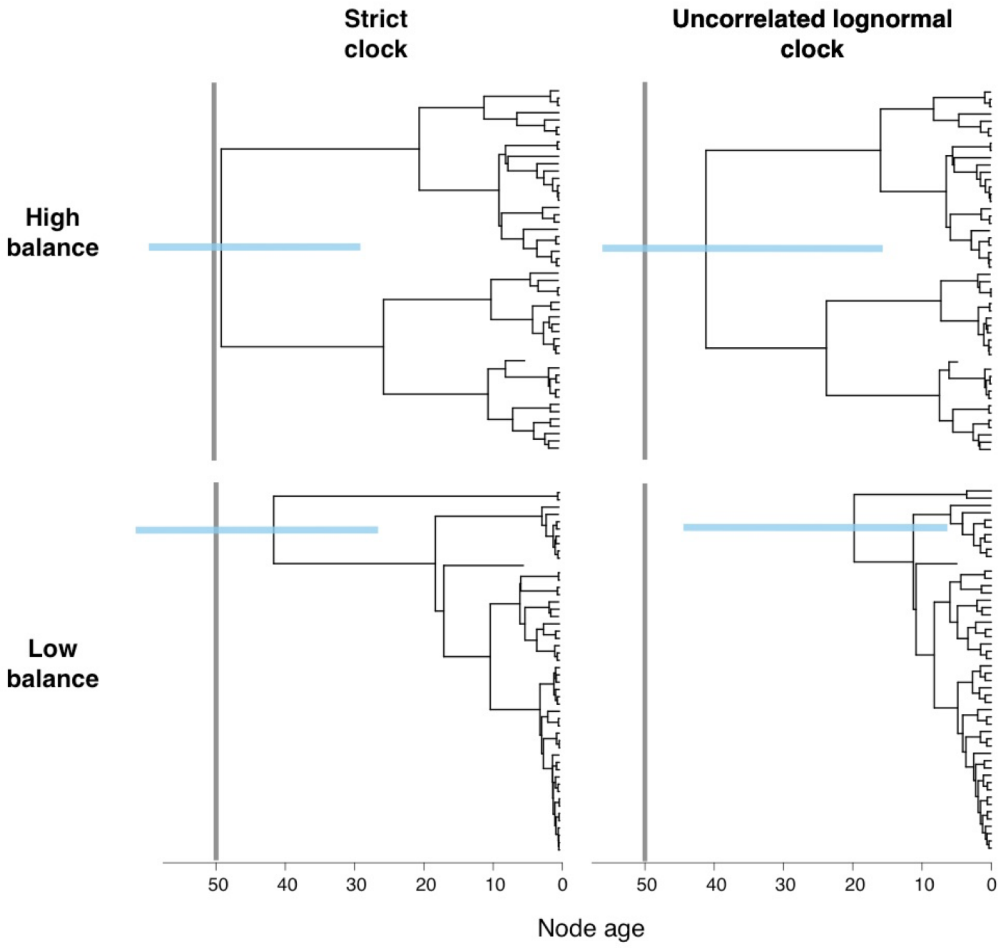


Figure 5.2. Examples of heterochronous maximum-clade-credibility trees for four simulation schemes. The blue line shows the 95% credible interval for the estimate of the root age, while the vertical grey line indicates the simulated root age. Results are shown for balanced (top) and imbalanced trees (bottom), and datasets simulated under a strict clock (left) and a relaxed uncorrelated lognormal clock (right).

The results from the analyses of the empirical datasets were consistent with those of our simulations. The estimated mean age of the root of the primate tree was similar for the datasets with balanced and imbalanced trees. These estimates were also similar to those obtained with the complete dataset, and to those of previous studies (Figure 5.3; Finstermeier *et al.* 2013). The ASFV dataset with tree imbalance produced a median estimate for the age of the root that was two orders of magnitude younger that that obtained with the complete dataset. In contrast, the estimate of the dataset with a

balanced tree was similar to that from the whole dataset (Figure 5.3). For analyses of primate and ASFV data, there was some overlap in the 95% credible intervals of the root age.

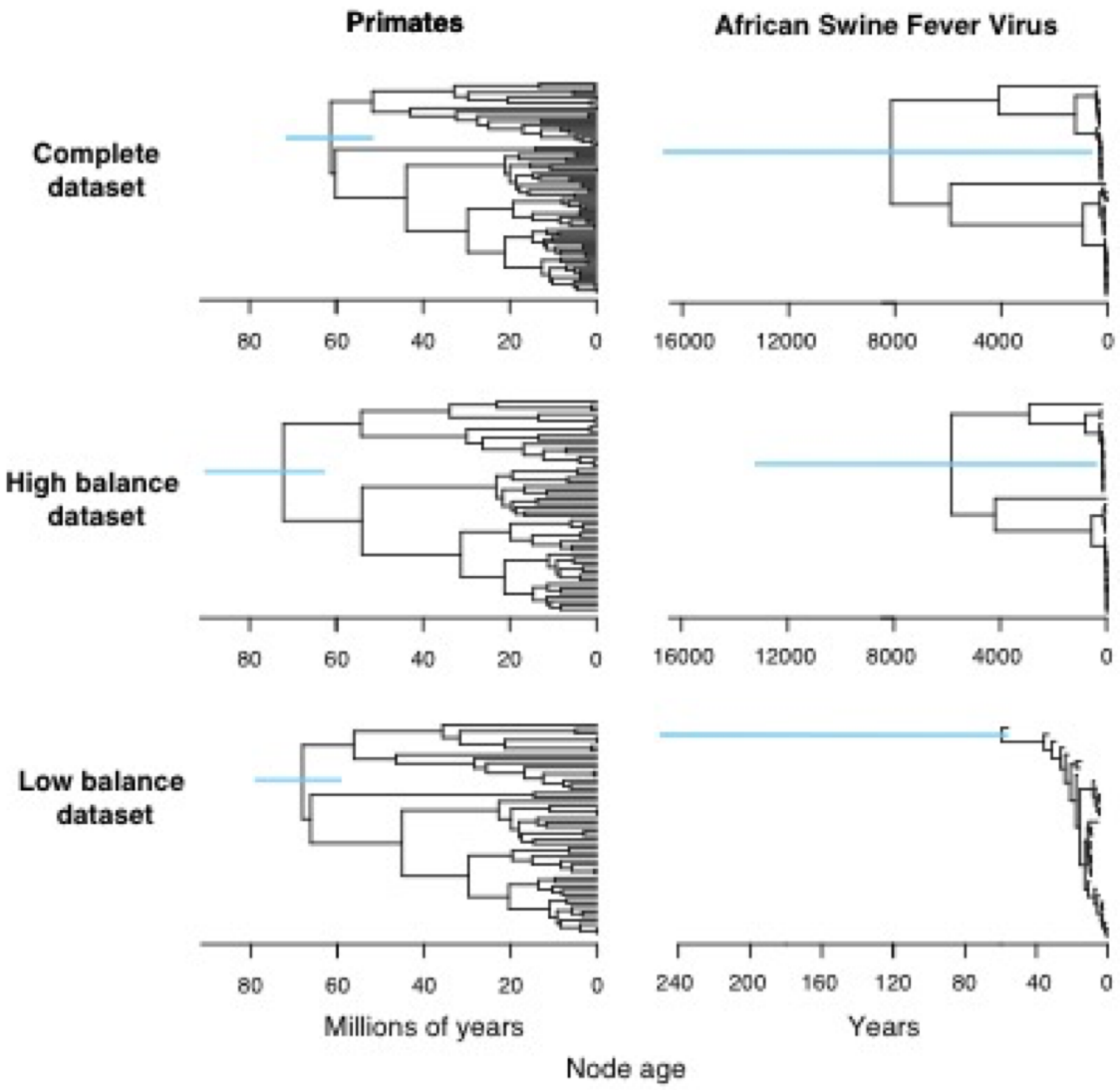


Figure 5.3. The maximum-clade-credibility trees for the six analyses of data from primates (left) and African swine fever virus (right), each using the complete dataset (top), a balanced dataset (middle), and an imbalanced dataset (bottom). The blue line shows the 95% credible interval for the estimate of the root age.

5.4 Discussion

We used simulations and empirical data to explore the effect of tree imbalance on phylogenetic estimation of divergence times. We found that tree imbalance had little impact on dates estimated from isochronous datasets, but led to underestimates of node-ages in heterochronous datasets. This effect was most severe for deep and medium-aged nodes, and for analyses involving shallow calibrations (Figure 5.1). With ineffective calibration schemes, the effect of imbalance can be severe, which we observed in the coverage probabilities of some node-age estimates that were well below 0.5. This reduction in accuracy was associated with a reduction in precision (Table 5.2). It is important to note, however, that tree imbalance might have a different impact in datasets with greater among-lineage rate variation than that in our simulations and in our empirical datasets. However, many of the available molecular-clock methods are unable to accommodate extreme levels of rate variation satisfactorily (e.g., Wertheim *et al.* 2012).

The findings from our analyses of empirical data were consistent with those of our simulations. The estimate of the age of the root of the primate dataset was similar between the two subsampled datasets with high and low levels of tree imbalance. These estimates were also consistent with those from the complete dataset and from previous studies. When the ASFV dataset was pruned to produce a balanced tree, the node-age estimates were consistent with those obtained with the complete dataset. With an imbalanced tree, however, the age of the root was underestimated by two orders of magnitude. In practice, this bias can result in highly misleading biological conclusions.

The estimation biases observed in our analyses are likely to affect heterochronous datasets with high tree imbalance, including those from ancient DNA, viruses and bacteria. Many datasets from modern viruses are often associated with highly imbalanced trees. For example, the phylogenetic trees from influenza viruses are considerably imbalanced, a property that has been attributed to the particular selective constraints imposed by the immune system of their host (Grenfell *et al.* 2004; Pompei *et al.* 2012). The results of our study are relevant to analyses of these datasets, particularly if there is evidence for an ancient origin and the calibrations only encompass a very short timeframe.

The estimation biases associated with tree imbalance might also affect datasets of slow-evolving pathogens, such as the herpesviruses, which are considered to have evolutionary timescales on the order of millions of years (McGeoch *et al.* 2000). Molecular data from these viruses have mostly been collected in the past century. If these modern datasets provide informative calibration points, they are still likely to be uninformative in the presence of imbalanced phylogenies, leading to spurious estimates of the evolutionary timescale (e.g. Paraskevis *et al.* 2013). Similarly, ancient DNA has been used to calibrate the molecular clock of many groups, including bears (Knapp *et al.* 2009; Korsten *et al.* 2009), bovines (Ho *et al.* 2008), hyaenas (Sheng *et al.* 2014), and humans (Fu *et al.* 2013). Owing to the non-random nature of sample collection, ancient DNA datasets are often associated with highly imbalanced trees (e.g. Calvignac *et al.* 2008; Knapp *et al.* 2009; Hekkala *et al.* 2011), making them prone to the bias we have identified in our study. The errors caused by tree imbalance can have impacts on downstream analyses that rely on molecular estimates of evolutionary timescales.

The effect of tree imbalance when there are only shallow calibrations can interact with another existing bias, known as the time dependence of molecular rates (Ho *et al.* 2005a). This source of bias can lead to overestimation of molecular rates in deep branches of the tree when the calibrations are located at shallow nodes (Ho *et al.* 2011a), resulting in an underestimation of the overall evolutionary timescale. The interaction of this effect with tree imbalance is unknown, but might be important because a time-dependent bias in rate estimates is widespread in viruses (Duchêne *et al.* 2014a), bacteria (Comas *et al.* 2013), and ancient DNA from vertebrates (Ho *et al.* 2007, 2011b). For these reasons, efforts to obtain deep calibrations for imbalanced trees are justified.

One way to improve divergence-time estimates for datasets that display tree imbalance is to increase taxon sampling. The effect of imbalance can be likened to the node-density effect, which occurs when some sections of phylogenetic trees have taxon undersampling or have experienced a high extinction rate. These processes produce long branches, with the amount of evolutionary change along these branches subject to underestimation if the substitution model is inadequate (Hugall & Lee 2007). Trees that have deeper roots and greater sampling tend to be more imbalanced than trees that span short timescales and contain few taxa (Purvis & Agapow 2002). Therefore, increasing sampling can be beneficial, but is perhaps most appropriate when the added samples have the effect of reducing tree imbalance.

The cause of date underestimation observed for imbalanced trees is probably a form of model inadequacy, or the inability of the model to generate the data. Model adequacy has been explored for substitution models (Bollback 2002; Ripplinger & Sullivan 2010), but less attention has been paid to the adequacy of models for phylogenetic estimation

of timescales. One possible source of bias is the clock model, since models that relaxed the rate had a more severe bias as imbalance increased. But the two specific sources of bias from tree imbalance are long branches and the tree structure. In many of the analyses with unreliable estimates of timescales, we matched the assumptions of the models of substitutions, among-lineage rate variation, and branching structure to those used for simulation. Nevertheless, any of these models might perform poorly under the atypical conditions posed by tree imbalance. Other problems might exacerbate the problem caused by tree imbalance, such as high rates of substitution or the estimation of tree topology, both of which generate more imbalance than expected from the true process (Huelsenbeck & Kirkpatrick 1996). Further investigation of model adequacy in phylogenetic estimation of evolutionary timescales will be a useful avenue of research.

Methods for clock-model selection have been found to be most effective in fully balanced trees (Baele *et al.* 2013), suggesting that tree imbalance poses a challenge to the estimation of evolutionary timescales. Patterns of rate variation among lineages can be difficult to detect using different clock models (Ho *et al.* 2015), and long branches might further complicate this identification. Since many trees are necessarily imbalanced in practice, further work is required to improve molecular timescale estimation using heterochronous data. This should include advances in the implementation of tip-calibrations (Molak *et al.* 2013, 2014), continued development of molecular-clock models (Worobey *et al.* 2014a; Ho & Duchêne 2014), and improvements in our understanding of how model misspecification leads to underestimation of branch lengths (Phillips 2009).

5.5 References

Agapow P-M, Purvis A (2002) Power of eight tree shape statistics to detect nonrandom diversification: A comparison by simulation of two models of cladogenesis. *Systematic Biology*, **51**, 866–872.

Baele G, Li WLS, Drummond AJ, Suchard MA, Lemey P (2013) Accurate model selection of relaxed molecular clocks in bayesian phylogenetics. *Molecular Biology and Evolution*, **30**, 239–243.

Blum M, François O (2006) Which random processes describe the tree of life? A large-scale study of phylogenetic tree imbalance. *Systematic Biology*, **55**, 685–691.

Bollback JP (2002) Bayesian model adequacy and choice in phylogenetics. *Molecular Biology and Evolution*, **19**, 1171–1180.

Bouckaert R, Heled J, Kühnert D *et al.* (2014) BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, **10**, e1003537.

Calvignac S, Hughes S, Tougard C *et al.* (2008) Ancient DNA evidence for the loss of a highly divergent brown bear clade during historical times. *Molecular Ecology*, **17**, 1962–1970.

Carstens BC, Knowles LL (2007) Shifting distributions and speciation: species divergence during rapid climate change. *Molecular Ecology*, **16**, 619–627.

Clark R, Vogler AP (2009) A phylogenetic framework for wing pattern evolution in the mimetic Mocker Swallowtail Papilio dardanus. *Molecular Ecology*, **18**, 3872–3884.

Colless DH (1982) Review of phylogenetics: the theory and practice of phylogenetic systematics. *Systematic Zoology*, **31**, 100–104.

Comas I, Coscolla M, Luo T *et al.* (2013) Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. *Nature Genetics*, **45**, 1176–1182.

Crisp MD, Trewick SA, Cook LG (2011) Hypothesis testing in biogeography. *Trends in Ecology & Evolution*, **26**, 66–72.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biology*, **4**, e88.

Duchêne S, Holmes EC, Ho SYW (2014a) Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates. *Proceedings of the Royal Society B: Biological Sciences*, **281**, 20140732.

Duchêne S, Lanfear R, Ho SYW (2014b) The impact of calibration and clock-model choice on molecular estimates of divergence times. *Molecular Phylogenetics and Evolution*, **78**, 277–289.

Felsenstein J (1978) Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Biology*, **27**, 401–410.

Finstermeier K, Zinner D, Brameier M *et al.* (2013) A mitogenomic phylogeny of living primates. *PLoS One*, **8**, e69504.

Fu Q, Mittnik A, Johnson PLF *et al.* (2013) A revised timescale for human evolution based on ancient mitochondrial genomes. *Current biology*, **23**, 553–559.

Goldberg EE, Lancaster LT, Ree RH (2011) Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. *Systematic Biology*, **60**, 451–465.

Grenfell BT, Pybus OG, Gog JR *et al.* (2004) Unifying the epidemiological and evolutionary dynamics of pathogens. *Science*, **303**, 327–332.

Heard SB (1996) Patterns in phylogenetic tree balance with variable and evolving speciation rates. *Evolution*, **46**, 2141–2148.

Heard SB, Mooers AØ (2002) Signatures of random and selective mass extinctions in phylogenetic tree balance. *Systematic Biology*, **51**, 889–897.

Heath TA, Zwickl DJ, Kim J, Hillis DM (2008) Taxon sampling affects inferences of macroevolutionary processes from phylogenetic trees. *Systematic Biology*, **57**, 160–166.

Hekkala E, Shirley MH, Amato G *et al.* (2011) An ancient icon reveals new mysteries: mummy DNA resurrects a cryptic species within the Nile crocodile. *Molecular Ecology*, **20**, 4199–4215.

Hendy MD, Penny D (1989) A framework for the quantitative study of evolutionary trees. *Systematic Zoology*, **38**, 297.

Ho SYW, Duchêne S (2014) Molecular-clock methods for estimating evolutionary rates and timescales. *Molecular Ecology*, **23**, 5947–5965.

Ho SYW, Duchêne S, Duchêne D (2015) Simulating and detecting autocorrelation of molecular evolutionary rates among lineages. *Molecular Ecology Resources*, **15**, 688–696.

Ho SYW, Kolokotronis S-O, Allaby RG (2007) Elevated substitution rates estimated from ancient DNA sequences. *Biology Letters*, **3**, 702–705.

Ho SYW, Lanfear R, Bromham L *et al.* (2011a) Time-dependent rates of molecular evolution. *Molecular Ecology*, **20**, 3087–3101.

Ho SYW, Lanfear R, Phillips MJ *et al.* (2011b) Bayesian estimation of substitution rates from ancient DNA sequences with low information content. *Systematic Biology*, **60**, 366–375.

Ho SYW, Larson G, Edwards CJ *et al.* (2008) Correlating Bayesian date estimates with climatic events and domestication using a bovine case study. *Biology Letters*, **4**, 370–374.

Ho SYW, Phillips MJ (2009) Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Systematic Biology*, **58**, 367–380.

Ho SYW, Phillips MJ, Cooper A, Drummond AJ (2005a) Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular Biology and Evolution*, **22**, 1561–1568.

Ho SYW, Phillips MJ, Drummond AJ, Cooper A (2005b) Accuracy of rate estimation using relaxed-clock models with a critical focus on the early metazoan radiation. *Molecular Biology and Evolution*, **22**, 1355–1363.

Holman EW (2005) Nodes in phylogenetic trees: the relation between imbalance and number of descendent species. *Systematic Biology*, **54**, 895–899.

Holton TA, Wilkinson M, Pisani D (2014) The shape of modern tree reconstruction methods. *Systematic Biology*, **63**, 436–441.

Huelsenbeck JP (1995) Performance of phylogenetic methods in simulation. *Systematic Biology*, **44**, 17–48.

Huelsenbeck JP, Kirkpatrick M (1996) Do phylogenetic methods produce trees with biased shapes? *Evolution*, 1418–1424.

Hugall AF, Lee MSY (2007) The likelihood node density effect and consequences for evolutionary studies of molecular rates. *Evolution*, **61**, 2293–2307.

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: *Mammalian Protein Metabolism* (ed Munro H), pp. 21–132. Academic Press, New York.

Kirkpatrick M, Slatkin M (1993) Searching for evolutionary patterns in the shape of a phylogenetic tree. *Evolution*, **47**, 1171–1181.

Knapp M, Rohland N, Weinstock J *et al.* (2009) First DNA sequences from Asian cave bear fossils reveal deep divergences and complex phylogeographic patterns. *Molecular Ecology*, **18**, 1225–1238.

Korsten M, Ho SYW, Davison J *et al.* (2009) Sudden expansion of a single brown bear maternal lineage across northern continental Eurasia after the last ice age: a general demographic model for mammals? *Molecular Ecology*, **18**, 1963–1979.

Lepage T, Bryant D, Philippe H, Lartillot N (2007) A general comparison of relaxed molecular clock models. *Molecular Biology and Evolution*, **24**, 2669–2680.

Linder M, Britton T, Sennblad B (2011) Evaluation of Bayesian models of substitution rate evolution--parental guidance versus mutual independence. *Systematic Biology*, **60**, 329–342.

Magallón S (2010) Using fossils to break long branches in molecular dating: a comparison of relaxed clocks applied to the origin of angiosperms. *Systematic Biology*, **59**, 384–399.

McGeoch DJ, Dolan A, Ralph AC (2000) Toward a comprehensive phylogeny for mammalian and avian herpesviruses. *Journal of Virology*, **74**, 10401–10406.

Molak M, Lorenzen ED, Shapiro B, Ho SYW (2013) Phylogenetic estimation of timescales using ancient DNA: the effects of temporal sampling scheme and uncertainty in sample ages. *Molecular Biology and Evolution*, **30**, 253–262.

Molak M, Suchard MA, Ho SYW, Beilman DW, Shapiro B (2014) Empirical calibrated radiocarbon sampler: a tool for incorporating radiocarbon-date and calibration error into Bayesian phylogenetic analyses of ancient DNA. *Molecular Ecology Resources*, **15**, 81–86.

Mooers AO (1995) Tree balance and tree completeness. *Evolution*, **49**, 379–384.

Mooers AO, Heard SB (1997) Inferring evolutionary process from phylogenetic tree shape. *Quarterly Review of Biology*, 31–54.

Nater A, Nietlisbach P, Arora N *et al.* (2011) Sex-biased dispersal and volcanic activities shaped phylogeographic patterns of extant Orangutans (genus: Pongo). *Molecular Biology and Evolution*, **28**, 2275–2288.

Paraskevis D, Magiorkinis G, Magiorkinis E *et al.* (2013) Dating the origin and dispersal of hepatitis B virus infection in humans and primates. *Hepatology*, **57**, 908–916.

Parks SL, Goldman N (2014) Maximum Likelihood Inference of Small Trees in the Presence of Long Branches. *Systematic Biology*, **63**, 798–811.

Perelman P, Johnson WE, Roos C *et al.* (2011) A molecular phylogeny of living primates. *PLoS Genetics*, **7**, e1001342.

Phillips MJ (2009) Branch-length estimation bias misleads molecular dating for a vertebrate mitochondrial phylogeny. *Gene*, **441**, 132–140.

Pompei S, Loreto V, Tria F (2012) Phylogenetic properties of RNA viruses. *PLoS One*, **7**, e44849.

Popescu A-A, Huber KT, Paradis E (2012) ape 3.0: New tools for distance-based phylogenetics and evolutionary analysis in R. *Bioinformatics*, **28**, 1536–1537.

Purvis A, Agapow P-M (2002) Phylogeny imbalance: Taxonomic level matters. *Systematic Biology*, **51**, 844–854.

Rajabi-Maham H, Orth A, Bonhomme F (2008) Phylogeography and postglacial expansion of Mus musculus domesticus inferred from mitochondrial DNA coalescent, from Iran to Europe. *Molecular Ecology*, **17**, 627–641.

Rambaut A (2000) Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies. *Bioinformatics*, **16**, 395–399.

Revell L, Harmon L, Glor R (2005) Under-parameterized model of sequence evolution leads to bias in the estimation of diversification rates from molecular phylogenies. *Systematic Biology*, **54**, 973–983.

Ricklefs RE (2007) Estimating diversification rates from phylogenetic information. *Trends in Ecology & Evolution*, **22**, 601–610.

Ripplinger J, Sullivan J (2010) Assessment of substitution model adequacy using frequentist and Bayesian methods. *Molecular Biology and Evolution*, **27**, 2790–2803.

Schliep KP (2011) PHANGORN: phylogenetic analysis in R. *Bioinformatics*, **27**, 592–593.

Sheng G-L, Soubrier J, Liu J-Y *et al.* (2014) Pleistocene Chinese cave hyenas and the recent Eurasian history of the spotted hyena, Crocuta crocuta. *Molecular Ecology*, **23**, 522–533.

Thorne JL, Kishino H (2002) Divergence time and evolutionary rate estimation with multilocus data. *Systematic Biology*, **51**, 689–702.

Wertheim JO, Fourment M, Kosakovsky Pond SL (2012) Inconsistencies in estimating the age of HIV-1 subtypes due to heterotachy. *Molecular Biology and Evolution*, **29**, 451–456.

Worobey M, Han G-Z, Rambaut A (2014a) A synchronized global sweep of the internal genes of modern avian influenza virus. *Nature*, **508**, 254–257.

Worobey M, Han G-Z, Rambaut A (2014b) Genesis and pathogenesis of the 1918 pandemic H1N1 influenza A virus. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 8107–8112.

Chapter 6 – Evaluating the adequacy of molecular clock models using posterior predictive simulations

A version of this chapter was published as:

Duchêne DA, Duchêne S, Holmes EC, Ho SYW (2015) Evaluating the adequacy of molecular clock models using posterior predictive simulations. *Molecular Biology and Evolution*. **32**, 2986-2995.

Author contributions:

DD, SD devised the idea for the study. DD, SD, ECH, and SYWH designed the study. DD made and analysed the simulations. SD collected the empirical data. DD and SD analysed the data. DD and SD wrote the paper; with input from ECH and SYWH.

6.1 Introduction

Analyses of nucleotide sequences can provide a range of valuable insights into evolutionary relationships and timescales, allowing various biological questions to be addressed. The problem of inferring phylogenies and evolutionary divergence times is a statistical one, such that inferences are dependent on reliable models of the evolutionary process (Felsenstein 1983). Bayesian methods provide a powerful framework for estimating phylogenetic trees and evolutionary rates and timescales using parameter-rich models (Huelsenbeck *et al.* 2001; Yang & Rannala 2012). Model-based phylogenetic inference in a Bayesian framework has several desirable properties: it is possible to include detailed descriptions of molecular evolution (e.g. Dutheil *et al.* 2012; Heath *et al.* 2014); many of the model assumptions are explicit (Sullivan & Joyce 2005); large parameter spaces can be explored efficiently (e.g. Nylander *et al.* 2004;

Drummond *et al.* 2006); and uncertainty is naturally incorporated in the estimates. As a consequence, the number and complexity of evolutionary models for Bayesian inference has grown rapidly, prompting considerable interest in methods of model selection (e.g. Xie *et al.* 2011; Baele *et al.* 2013).

Evolutionary models can provide useful insight into biological processes, but they are incomplete representations of molecular evolution (Goldman 1993). This can be problematic in phylogenetic inference when all of the available models are poor descriptions of the process that generated the data (Gatesy 2007). Traditional methods of model selection do not allow the rejection, or falsification, of every model in the set of candidates being considered. Gelman and Shalizi (2013) recently referred to this as a critical weakness in current practice of Bayesian statistics. A different approach to model selection is to evaluate the adequacy, or plausibility (following Brown 2014a), of the model. This involves testing whether the data could have been generated by the model in question (Gelman *et al.* 2014).

Assessment of model adequacy is a critical step in Bayesian inference in general (Gelman & Shalizi 2013), and phylogenetics in particular (Brown 2014b). One method of evaluating the adequacy of a model is to use posterior predictive checks (Gelman *et al.* 2014). Among the first of such methods in phylogenetics was the use of posterior predictive simulations, proposed by Bollback (2002). The first step in this approach is to conduct a Bayesian phylogenetic analysis of the empirical data. The second step is to use simulation to generate data sets with the same size as the empirical data, using the values of model parameters sampled from the posterior distribution obtained in the first step. The data generated via these posterior predictive simulations are considered to

represent hypothetical alternative or future data sets, but generated by the model used for inference.

If the process that generated the empirical data can be described with the model used for inference, the posterior predictive data sets should resemble the empirical data set (Gelman *et al.* 2014). Therefore, the third step in assessing model adequacy is to perform a comparison between the posterior predictive data and the empirical data. This comparison must be done using a test statistic that quantifies the discrepancies between the posterior predictive data and the empirical data (Gelman & Meng 1996). The test statistic is calculated for each of the posterior predictive data sets to generate a distribution of values. If the test statistic calculated from the empirical data falls outside this distribution of the posterior predictive values, the model in question is considered to be inadequate. This method is similar to using a parametric bootstrap, but can be used to assess model adequacy by comparing the simulated data to the empirical data.

Previous studies using posterior predictive checks of nucleotide substitution models have implemented a number of different test statistics. Some of these provide descriptions of the sequence alignments, such as the homogeneity of base composition (Huelsenbeck *et al.* 2001; Foster 2004), site frequency patterns (Bollback 2002; Lewis *et al.* 2014), and unequal synonymous versus non-synonymous substitution rates (Nielsen 2002; Rodrigue *et al.* 2009). Brown (2014b) and Reid et al. (2014) introduced test statistics based on phylogenetic inferences from posterior predictive data sets. Some of the characteristics of inferred phylogenies that can be used as test statistics include the mean tree length and the median Robinson-Foulds distance between the sampled topologies in the analysis (Brown 2014a). Although several test statistics are available for assessing models of nucleotide substitution (e.g. Brown and ElDabaje 2009; Brown

2014a; Lewis et al. 2015), there are no methods available to assess the adequacy of molecular clock models.

Molecular clocks have become an established tool in evolutionary biology, allowing the study of molecular evolutionary rates and divergence times between organisms (Kumar 2005; Ho 2014). Molecular clock models describe the pattern of evolutionary rates among lineages, relying on external temporal information (e.g., fossil data) to calibrate estimates of absolute rates and times. The primary differences among the various clock models include the number of distinct substitution rates across the tree and the degree to which rates are treated as a heritable trait (e.g. Thorne *et al.* 1998; Drummond *et al.* 2006; Drummond & Suchard 2010; for a review see Ho & Duchêne 2014). For example, the strict clock assumes that the rate is the same for all branches, whereas some relaxed-clock models allow each branch to have a different rate. We refer to models that assume a large number of rates as being more parameter-rich than models with a small number of rates (see Ho & Duchêne 2014). Although molecular clock models are used routinely, the methods of assessing their efficacy are restricted to estimating and comparing their statistical fit. For example, a common means of model selection is to compare marginal likelihoods in a Bayesian framework (Baele *et al.* 2013). However, model selection can only evaluate the relative statistical fit of the models, such that it can lead to false confidence in the estimates if all of the candidate models are actually inadequate.

In this study, we introduce a method for assessing the adequacy of molecular clock models. Using simulated and empirical data, we show that our approach is sensitive to under-parameterization of the clock model, and that it can be used to identify the branches of the tree that are in conflict with the assumed clock model. In practice, our

method is also sensitive to other aspects of the hierarchical model, such as misspecification of the node-age priors. We highlight the importance of methods of evaluating the adequacy of substitution models in molecular clock analyses.

6.2 New Approach

6.2.1 A method of assessing clock-model adequacy

To evaluate the adequacy of molecular clock models, we propose a method of generating and analysing posterior predictive data. In this method, the posterior predictive data sets are generated using phylogenetic trees inferred from branch-specific rates and times from the posterior samples (Figure 6.1). Because this method uses branch-specific estimates, it requires a fixed tree topology.

The first step in our method is to conduct a Bayesian molecular clock analysis of empirical data. We assume that this analysis obtains samples from the posterior distribution of branch-specific rates and times. These estimates are given in relative time, or in absolute time if calibration priors are used. In the second step, we take a random subset of these samples. For each of these samples, we multiply the branch-specific rates and times to produce phylogenetic trees in which the branch lengths are measured in substitutions per site (substitutions/site), known as phylograms. To assess model adequacy, we randomly select 100 samples from the posterior, excluding the burn-in. From these samples, posterior predictive data sets are generated by simulation along the phylograms and using the estimates of the parameters in the nucleotide substitution model. The third step in our approach is to use a clock-free method to estimate a phylogram from each of the posterior predictive data sets and from the

empirical data set. For this step, we find that the maximum likelihood approach implemented in phangorn (Schliep 2011) is effective.

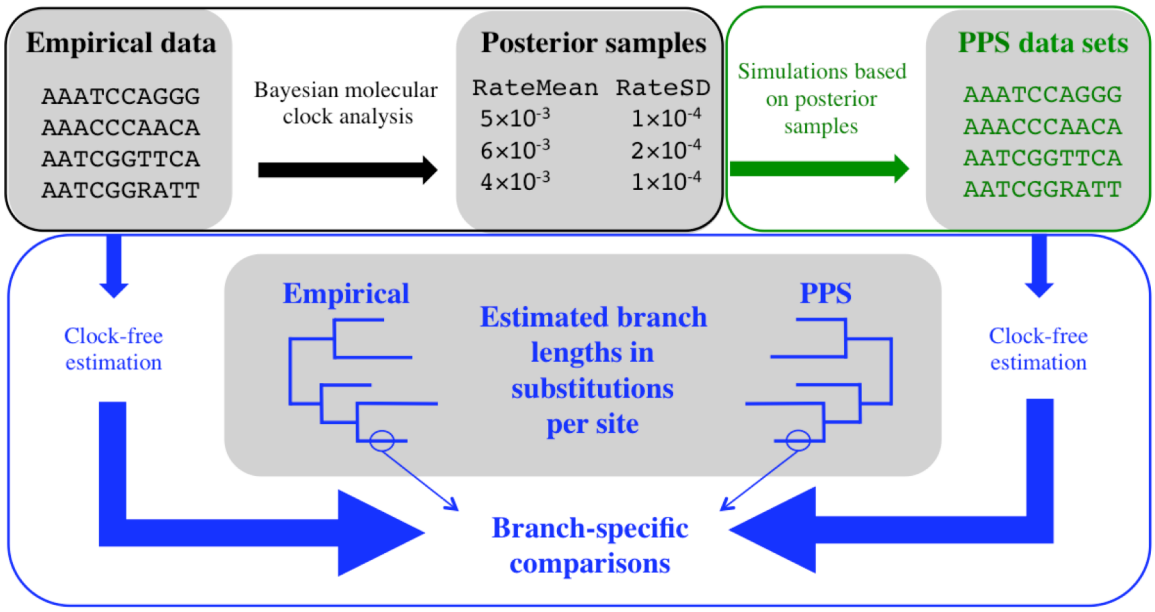


Figure 6.1. Procedure for assessing the adequacy of molecular clock models. The top-left box shows the components of a Bayesian clock analysis of empirical data, including samples from the posterior of the mean estimates and standard deviation of the substitution rates. The top-right box shows the first step in assessing model adequacy using posterior predictive simulations (PPS). In our analyses, this step is performed using branch-specific rates and times. The bottom box shows our procedure for testing the clock model, which is based on the clock-free posterior predictive distribution of the length of each branch. The thin arrows indicate that the test statistic is the posterior predictive *P*-value for each branch.

To compute our adequacy index, we consider the branch lengths estimated from the posterior predictive data sets under a clock-free method, such that there is a distribution of length estimates for each branch. We calculate a posterior predictive *P*-value for each branch using the corresponding distribution obtained with the posterior predictive data sets. This value is important for identifying the length estimates for individual branches that are in conflict with the clock model. Our index for overall assessment is the proportion of branches in the phylogram from the empirical data that have lengths

falling outside the 95% quantile range of those estimated from the posterior predictive data sets. We refer to our index as *A*, or overall plausibility of branch length estimates.

We also provide a measure of the extent to which the branch length estimates from the clock-free method differ from those obtained using the posterior predictive simulations. To do this, we calculate for each branch the absolute difference between the empirical branch length estimated using a clock-free method and the mean branch length estimate from the posterior predictive data. We then divide this value by the empirical branch-length estimated using a clock-free method. This measure corresponds to the deviation of posterior predictive branch lengths from the branch length estimated from the empirical data. For simulations and analyses of empirical data, we present the median value across branches to avoid the effect of extreme values. We refer to this measure as 'branch length deviation', of which low values represent high performance.

We also investigated the uncertainty in the estimates of posterior predictive branch lengths. This is useful because it provides insight into the combined uncertainty in estimates of rates and times. The method we used was to take the width of the 95% quantile range from the posterior predictive data sets, divided by the mean length estimated for each branch. This value, along with the width of the 95% credible interval of the rate estimate from the original analysis, can then be compared among clock models to investigate the increase in uncertainty that can occur when using complex models.

6.3 Methods

6.3.1 Analyses of simulated data

We generated 100 pure-birth trees with 50 tips and root-node ages of 50 million years (My) using BEAST v2.1 (Bouckaert *et al.* 2014). We then simulated branch-specific rates under five clock-model treatments using the R package NELSI (Ho *et al.* 2015). This program simulates rates under a given model and multiplies rates by time to produce phylogenetic trees in which the branch lengths represent substitutions/site, known as phylograms. These phylograms were then used to simulate the evolution of DNA sequences of 2000 nucleotides (nt) in the R package phangorn.

The five clock model treatments included: (i) a strict clock with a rate of 5×10^{-3} substitutions/site/My; (ii) an uncorrelated lognormal relaxed clock (Drummond *et al.* 2006), with a mean rate of 5×10^{-3} substitutions/site/My and a standard deviation of 0.1; (iii) a treatment in which a randomly selected clade with at least ten tips experienced an increase in the rate, representing a scenario with two local clocks (Yoder & Yang 2000), with rates of 1×10^{-2} and 1×10^{-3} substitutions/site/My; (iv) a treatment with rate autocorrelation with an initial rate of 5×10^{-3} substitutions/site/My and a ν parameter of 0.3 (Kishino *et al.* 2001); and (v) a treatment with rate variation that followed a beta distribution with equal shape parameters of 0.4 and centred at 5×10^{-3} substitutions/site/My, resulting in a bimodal shape. In every simulation, the mean rate was 5×10^{-3} substitutions/site/My, which is approximately the mean mitochondrial evolutionary rate in mammals, birds, non-avian reptiles, and amphibians (Pereira & Baker 2006). We selected this mean rate instead of sampling from the prior because our estimation methods involved an uninformative rate prior, and random samples from this can produce data sets with high sequence saturation or with low information content.

We used the Jukes-Cantor substitution model for simulation (Jukes & Cantor 1969). This model allows us to avoid making arbitrary parameterizations of more parameter-rich models, which is not the focus of this study.

To explore the effect of substitution model under-parameterization, we simulated additional data sets under a strict clock and a general time-reversible model with gamma distributed rates among sites, using parameters from empirical data (Murphy *et al.* 2001). We analysed these data sets using the same method as for the rest of the simulated data, including the use of the simpler Jukes-Cantor substitution model. We also explored the effect of using misleading node-age priors (i.e. node calibrations that differ from the true node ages). To do this, we placed two time-calibrations with incorrect ages. One calibration was placed in one of the two nodes descending from the root selected at random, with an age-prior of 0.1 times its true age (i.e. younger than the truth). The other calibration was placed on the most recent node in the other clade descending from the root, with an age of 0.9 of the root age (i.e. older than the truth). For this scenario, we only used trees with more than one descendant in each of the two oldest clades. We show an example of the simulated phylogeny compared with this kind of marginal prior on node ages in the Appendix – Chapter 6. Our study had 100 simulated data sets for each simulation treatment, for a total of 700 simulated alignments.

We analysed the simulated alignments using Bayesian Markov chain Monte Carlo (MCMC) sampling as implemented in BEAST. We used three different clock models to analyse each of the simulated alignments: the strict clock, uncorrelated lognormal relaxed clock (Drummond *et al.* 2006), and random local clock (Drummond & Suchard 2010). We used the same tree prior and substitution model for estimation as those used

for simulation. We fixed the age of the root to 50 My and fixed the tree topology to that used to simulate sequence evolution in every analysis. We analysed the simulated data with an MCMC chain length of 2×10^7 steps, with samples drawn from the posterior every 2×10^3 steps. We discarded the first 10% of the samples as burn-in, and assessed satisfactory sampling from the posterior by verifying that effective sample sizes for all parameters were above 200 using the R package CODA (Plummer *et al.* 2006). We performed analyses using each of the three clock models for each of the 300 simulated data sets, for a total of 900 clock analyses.

We assessed the accuracy and uncertainty of the estimates made using each of the analysis schemes (Table 6.1). To do this, we compared the simulated rates with the branch-specific rates in the posterior. Next, we tested the power of our method for assessing clock-model adequacy using the simulated data under each of the scenarios of simulation and analysis. We provide example code and results in a public repository in GitHub (https://github.com/duchene/modadclocks). We also tested the power of the multinomial test statistic to assess clock-model adequacy in each of the 900 analyses. This test statistic quantifies the frequency of site patterns in an alignment and is appropriate for testing the adequacy of models of nucleotide substitution (Bollback 2002; Brown 2014b).

6.3.2 Analyses of empirical data

We used four published data sets to investigate the performance of our method of assessing clock-model adequacy in empirical data. For each data set, we performed analyses in BEAST using each of the three clock models used to analyse the simulated data sets. To select the substitution model for each empirical data set, we used the Bayesian information criterion as calculated in the R package phangorn.

For each analysis of the empirical data sets, we ran the MCMC chain for 10^8 steps, with samples drawn from the posterior every 10^3 steps. We discarded the first 10% of the samples as burn-in and assessed satisfactory sampling from the posterior by verifying that the effective sample sizes for all parameters were above 200 using the R package CODA. We used stepping-stone sampling to estimate the marginal likelihood of the clock model (Gelman & Meng 1998; Lartillot & Philippe 2006; Xie *et al.* 2011). For each Bayesian analysis, we performed posterior predictive simulations as done for the simulated data sets, and assessed the substitution model using the multinomial test statistic. In addition, to estimate the clock-free multinomial test statistic, we analysed each of the empirical data sets using MrBayes 3.2 (Ronquist *et al.* 2012). For these analyses we used the same chain length, sampling frequency, sampling verification method, and substitution model as in the analyses using clock models.

Our empirical data sets included nucleotide sequences of coronaviruses. This data set contained 43 sequences of 638 nt of a portion of the *M* (matrix) gene, as used by Wertheim et al. (2013). These sequences were sampled between 1941 and 2011. The best-fitting substitution model for this data set was GTR+Γ. We also used a data set of the *gag* gene of simian immunodeficiency viruses (SIV), which comprised 78 sequences of 477 nt, sampled between 1983 and 2004 (Wertheim & Worobey 2009). The best-fitting substitution model for this data set was GTR+Γ. We used the Bayesian skyline demographic model (Drummond *et al.* 2005) for the analyses of both of the virus data sets, and used the sampling times for calibration.

We analysed a data set of the killer whale (*Orcinus orca*), which contained 60 complete mitochondrial genome sequences of 16,386 nt (Morin *et al.* 2010). We calibrated the

age of the root using a normal distribution with mean of 0.7 and a standard deviation of 5% of the mean, as used in the original study. The best-fitting substitution model for this data set was HKY+ Γ. Lastly, we analysed a data set of several genera of marine turtles, which comprised 24 sequences of the 13 mitochondrial protein-coding genes (Duchene *et al.* 2012), and we selected the GTR+ Γ substitution model. Following the scheme in the original study, we used calibrations at four internal nodes. The pure-birth process was used to generate the tree prior in the analyses of the killer whales and the marine turtles.

6.4 Results

6.4.1 Assessment of clock-model adequacy in simulated data

We first evaluated the accuracy and uncertainty of substitution rate estimates from simulated data. To do this, we compared the values used to generate the data with those estimated using each of three clock models; strict clock, random-local clocks (Drummond & Suchard 2010), and the uncorrelated lognormal relaxed clock (Drummond *et al.* 2006). We regarded the branch-specific rates as accurate when the rate used for the simulation was contained within the 95% credible interval. We found that rate estimates were frequently inaccurate under five circumstances: clock model under-parameterization; rate autocorrelation among branches (Kishino *et al.* 2001); uncorrelated beta-distributed rate variation among lineages; misleading node-age priors (i.e. node calibrations that differ considerably from the true node ages); and when data were generated under a strict clock but analysed with an under-parameterized substitution model (Table 6.1a). When analyses were performed using the correct or an over-parameterized clock model, more than 75% of branch rates were accurately estimated, such that the true value was contained within the 95% credible interval

(Table 6.1a). In most simulation schemes, the uncorrelated lognormal relaxed clock had high accuracy, at the expense of a small increase in the uncertainty compared with the other models (Table 6.1b). These results are broadly similar to those of Drummond *et al.* (2006), who also found that under-parameterization of the clock model resulted in low accuracy in rate estimates, whereas over-parameterization had a negligible effect on accuracy.

		Simulated													
		a. Accuracy							b. Uncertainty						
		SC	LOC	UCL	ACL	BIM	PRI	GTRG	SC	LOC	UCL	ACL	BIM	PRI	GTRG
Estimated	SC	1.00	0.18	0.31	0.04	0.03	0.10	<0.01	0.07	0.06	0.07	0.06	0.07	0.08	0.04
	RLC	0.82	0.78	0.35	0.19	0.20	0.06	0.01	0.20	0.23	0.20	0.31	0.38	0.30	0.14
	UCL	1.00	0.75	0.97	0.59	0.93	0.71	<0.01	0.16	1.40	0.31	0.88	1.45	2.74	0.23

Table 6.1. Mean values of (a) accuracy and (b) uncertainty of branch rate estimates from molecular clock analyses of simulated data. Each cell shows the results of 100 replicate analyses. Accuracy is measured as the proportion of data sets for which the rate used for simulation was contained in the 95% credible interval of the estimate. Darker shades in (a) represent high accuracy. Uncertainty is measured as the width of the 95% credible interval as a proportion of the mean rate. Dark shades in (b) represent small ranges in branch length estimates, and therefore low uncertainty. The initials stand for each of the schemes for estimation or simulation (SC, strict clock; LOC, local clock; UCL, uncorrelated lognormal relaxed clock; RLC, random local clock; ACL, autocorrelated relaxed clock; BIM, beta-distributed bimodal clock; PRI, misleading node-age prior; GTRG, data simulated under the parameter-rich general time-reversible substitution model with a

We analysed data generated by simulation to test our method of assessing the adequacy of molecular clock models. The A index was approximately proportional to the branch length deviation (Table 6.2a). We found that $A \geq 0.95$ (indicating high performance) when the model used in the analyses matched that used to generate the data, or when it was over-parameterized. When the assumed model was under-parameterized, A was

≤0.92. The uncertainty obtained using posterior predictive branch lengths was sensitive to the rate variance in the simulations. For this reason, estimates from data generated according to a strict clock or an uncorrelated lognormal relaxed clock had lower uncertainty than estimates from data generated under local clocks, regardless of the model used for analysis (Table 6.2b). Estimates made using the uncorrelated lognormal relaxed clock had a larger variance in three analysis schemes: when data were generated with autocorrelated rates across branches; when data were generated with beta-distributed rates across branches; and when there was a prior for the node ages that was not described by the actual prior used for analysis. For analyses with substitution model under-parameterization, our method incorrectly provided greater support for the more complex clock model, indicating that rate variation among lineages was overestimated (Table 6.2).

		Simulated													
		a. Accuracy							b. Uncertainty						
		SC	LOC	UCL	ACL	BIM	PRI	GTRG	SC	LOC	UCL	ACL	BIM	PRI	GTRG
Estimated	SC	0.97 (0.08)	0.82 (0.19)	0.92 (0.10)	0.72 (0.19)	0.52 (0.33)	0.46 (0.40)	0.93 (0.08)	0.84	1.12	0.84	0.84	0.88	0.97	0.71
	RLC	0.97 (0.08)	0.95 (0.10)	0.92 (0.09)	0.87 (0.12)	0.84 (0.13)	0.88 (0.10)	0.94 (0.08)	0.85	1.13	0.84	0.86	0.94	0.93	0.72
	UCL	0.98 (0.07)	0.96 (0.04)	0.98 (0.07)	0.98 (0.03)	0.98 (0.03)	0.97 (0.02)	0.99 (0.05)	0.85	1.23	0.87	0.92	1.01	1.02	0.73

Table 6.2. Mean values of (a) plausibility, *A*, and (b) uncertainty as described by the posterior predictive simulations from clock analyses of simulated data. Each cell shows the results of 100 replicate analyses. Values in parentheses are the branch length deviations, of which lower values indicate good performance. The darker shades represent higher values of *A* and less uncertainty. High values of *A* represent good performance. In the case of uncertainty, small values indicate small ranges in posterior predictive branch lengths, and therefore low uncertainty. The initials stand for each of the schemes for estimation or simulation (SC, strict clock; LOC, local clock; UCL, uncorrelated lognormal relaxed clock; RLC, random local clock; ACL, autocorrelated relaxed clock; BIM, beta-distributed bimodal clock; PRI, misleading node-age prior; GTRG, data simulated under the parameter-rich general time-reversible substitution model with among-site rate heterogeneity).

We used our simulated data and posterior predictive simulations to investigate the performance of the multinomial test statistic for evaluating the adequacy of molecular clock models. This test statistic was originally designed to assess models of nucleotide substitution (Bollback 2002) and can perform well compared with some of the other existing test statistics (Brown 2014b). The multinomial test statistic for the empirical alignment can be compared with the distribution of test statistics from posterior predictive data sets to produce a posterior predictive *P*-value. We find that the multinomial test statistic correctly identified when the substitution model was matched or under-parameterized (Table 6.3). The multinomial likelihood did not have the power to detect clock model adequacy, but it was sensitive to rate variation among lineages, primarily from the simulation involving autocorrelated rates and when the node-age

prior was misleading (i.e. it contained calibrations that differed considerably from the true node ages; Table 6.3).

6.4.2 Assessment of clock-model adequacy for empirical data

We used three clock models, as in our analyses of simulated data, to analyse a broad range of nucleotide sequence data sets: the *M* (matrix) gene of a set of coronaviruses; the *gag* gene of simian immunodeficiency virus (SIV; Wertheim and Worobey 2009); complete mitochondrial genomes of killer whales *Orcinus orca* (Morin *et al.* 2010); and 13 mtDNA protein-coding genes of marine turtles (Duchene *et al.* 2012).

		Simulated						
		SC	LOC	UCL	ACL	BIM	PRI	GTRG
Estimated	SC	0.5	0.47	0.46	0.37	0.52	0.42	<0.01
	RLC	0.5	0.53	0.46	0.36	0.49	0.39	<0.01
	UCL	0.49	0.52	0.47	0.36	0.48	0.35	<0.01

Table 6.3. Mean *P*-values of the multinomial test statistic from posterior predictive simulations from simulated data. Each cell represents 100 replicate analyses. Darker shades correspond to higher numbers. A value of 0.5 indicates that the model is adequate. The initials indicate the models for simulation and estimation: strict clock (SC), local clock (LOC), uncorrelated lognormal relaxed clock (UCL), random local clock (RLC), autocorrelated relaxed clock (ACL), beta-distributed bimodal clock (BIM), misleading node-age prior (PRI), and data simulated under the parameter-rich general time-reversible substitution model (GTRG).

The uncorrelated lognormal relaxed clock was the best-fitting clock model according to the marginal likelihood for the coronaviruses, SIV, and the killer whales (Table 6.4). For the marine turtles, the random local clock provided the best fit. In all of the analyses of empirical data sets, the uncorrelated lognormal relaxed clock had the best performance according to our *A* index. The highest *A* index was 0.78, for the SIV and

the killer whales, and the lowest uncertainty in posterior predictive branch lengths was of 0.7 for the killer whales. The uncertainty for all other data sets was above 1, indicating that it was larger than the mean of the posterior predictive branch lengths.

We calculated the multinomial test statistic for the empirical data sets using the posterior predictive data from a clock-model analysis, as well as under a clock-free method. The multinomial test statistic from both methods suggested that the substitution model was inadequate for the SIV and the marine turtles, with posterior predictive *P*-values below 0.05. The substitution model was identified as inadequate for the coronavirus data set by the multinomial test statistic estimated using posterior predictive data sets from a clock analysis ($P < 0.05$); however, it was identified as adequate when using a clock-free method ($P = 0.20$). The mitochondrial data set from killer whales represented the only case in which the substitution model was adequate according to both multinomial likelihood estimates. For the data sets from coronaviruses and killer whales, the clock models with the highest performance had *A* indices of 0.53 and 0.78, respectively (Table 6.4). These indices are substantially lower than those obtained in analyses of simulated data when the clock model used for simulation and estimation were matched. However, we evaluated the posterior predictive *P*-values for all branches in these empirical data sets and found that at least two-thirds of the incorrect estimates correspond to relatively short terminal branches (Appendix – Chapter 6).

Table 6.4. Statistical fit and performance of three molecular clock models in analyses of four empirical data sets. The clock models are the strict clock (SC), uncorrelated lognormal relaxed clock (UCL), and the random local clock (RLC). For each data set, the number of rate changes is only estimated using the RLC. For the coronaviruses and SIV the rate estimates are shown in substitutions/site/year, while those for the killer whales and marine turtles correspond to substitutions/site/My. Note that substitution model assessment under the clock-free method was conducted only once per data set.

Data set	**Clock model**	**Mean number of rate changes (95% credible interval)**	**Mean rate estimate (95% credible interval)**	**Marginal likelihood estimate**	**Multinomial test statistic**		***A* index (mean branch-wise test statistic)**	**Uncertainty**
					Clock	**Clock-free**		
Coronaviruses	SC	-	2.04×10^{-5} (5.20×10^{-7} - 7.27×10^{-5})	-14445.78	0.01	0.20	0.49 (0.24)	1.09
	RLC	0.80 (0 - 3)	2.48×10^{-5} (7.11×10^{-7} - 9.48×10^{-5})	-14771.90	0.01	0.20	0.52 (0.24)	1.14
	UCL	-	**2.18×10^{-5} (5.90×10^{-7}-7.98×10^{-5})**	**-14329.07**	**<0.01**	0.20	**0.53 (0.16)**	**1.11**
SIV	SC	-	1.10×10^{-3} (8.22×10^{-4} - 1.45×10^{-3})	-3275.23	0.01	<0.01	0.65 (0.46)	2.02
	RLC	2.43 (1 - 5)	1.10×10^{-3} (7.90×10^{-4} - 1.44×10^{-3})	-3272.53	0.03	<0.01	0.65 (0.44)	2.03

	UCL	-	**1.10×10^{-3} (7.90×10^{-4} - 1.56×10^{-3})**	**-3256.00**	**0.04**	<0.01	**0.78 (0.23)**	**2.36**
Killer whales	SC	-	3.78×10^{-3} (3.02×10^{-3} - 4.67×10^{-3})	-24240.82	0.56	0.45	0.68 (0.48)	1.96
	RLC	0.79 (0 - 3)	3.77×10^{-3} (3.01×10^{-3} - 4.66×10^{-3})	-22211.21	0.54	0.45	0.25 (0.48)	1.05
	UCL	-	**3.90×10^{-3} (2.97×10^{-3} - 5.13×10^{-3})**	**-22167.75**	**0.47**	0.45	**0.78 (0.47)**	**0.70**
Marine turtles	SC	-	1.43×10^{-3} (1.34×10^{-3} - 1.52×10^{-3})	-37505.44	<0.01	<0.01	0.66 (0.23)	4.14
	RLC	**3.11 (1 – 6)**	**1.37×10^{-3} (1.15×10^{-3} - 1.56×10^{-3})**	**-37454.97**	<0.01	<0.01	**0.68 (0.21)**	**3.96**
	UCL	-	1.66×10^{-3} (1.39×10^{-3} - 1.90×10^{-3})	-37488.56	<0.01	<0.01	0.70 (0.09)	4.17

The branch length deviation in the empirical data ranged between 0.09 for the uncorrelated lognormal relaxed clock in the turtle data to 0.48 for the killer whale data analysed with a strict clock (Table 6.4). Low values for this metric indicate small differences between the posterior predictive and the empirical branch lengths. Although scores for this metric varied considerably between data sets, they were closely associated with the *A* indices for the different models for each data set individually. For example, in every empirical data set, the lowest branch length deviation was achieved by the model with the highest *A* index (indicative of higher performance). Importantly, the branch length deviation was not directly comparable to the *A* index between data sets. This is probably because the posterior predictive branch lengths have different amounts of uncertainty. In particular, the *A* index will tend to be low if the posterior predictive branch length estimates are similar to the empirical value but have low uncertainty. This would create a scenario with a small branch length deviation but also a low *A* index. This appears to be the case for the coronaviruses, for which the all the clock models appear inadequate according to the *A* index, but with the uncorrelated lognormal relaxed clock having a small branch length deviation.

6.5 Discussion

Assessing the adequacy of models in phylogenetics is an important process that can provide information beyond that offered by traditional methods for model selection. Although traditional model selection can be used to evaluate the relative statistical fit of a set of candidates, model adequacy provides information about the absolute performance of the model, such that even the best-fitting model can be a poor predictor of the data (Gelman *et al.* 2014). There have been important developments in model-adequacy methods and test statistics in the context of substitution models (Ripplinger &

Sullivan 2010; Brown 2014a; Lewis *et al.* 2014) and estimates of gene trees (Reid *et al.* 2014). Here we have described a method that can be used for assessment of molecular clock models, and which should be used in combination with approaches for evaluating the adequacy of substitution models. The results of our analyses suggest that our method is able detect whether estimates of branch-specific rates and times are consistent with the expected number of substitutions along each branch. For example, in the coronavirus data set analysed here, the best-fitting clock model was a poor predictor of the data, as was the substitution model. Our index is sensitive to under-parameterization of clock-models and has the benefit of being computationally efficient. In addition, our metric of uncertainty in posterior predictive branch lengths is sensitive to some cases of misspecification of clock models and node-age priors, but not to substitution model misspecification, as shown for our analyses of the coronavirus data set.

Analyses based on the random local clock and the data simulated under two local clocks generally produced low accuracy (Table 6.1a), with lower *A* indices than the other models that were matched to the true model (Table 6.2a). The substandard performance of the random local clock when it is matched to the true model is surprising. A possible explanation is that our simulations of the local clock represented an extreme scenario in which the rates of the local clocks differed by an order of magnitude. Previous studies based on simulations and empirical data demonstrated that this model can be effective when the rate differences are smaller (e.g. Drummond & Suchard 2010; Dornburg *et al.* 2012). One possible reason why the random local clock has low performance is that the prior has a substantial effect on estimates of rate changes. For example, a prior that identifies minimal changes in the rate might fail to identify large and quick changes in the rate. Interestingly, the *A*-index can be used to assess whether different prior settings for a given model are more adequate to study a given data set.

In our analyses of empirical data, even the highest values of our index were lower than the minimum value obtained in our analyses of simulated data when the three models matched those used for simulation. This is consistent with the results of previous studies of posterior predictive simulations, which have suggested that the proposed threshold for a test statistic using simulations is conservative for empirical data (Bollback 2002; Ripplinger & Sullivan 2010; Brown 2014a). It is difficult to suggest a specific threshold for our index to determine whether a model is inadequate. However, the interpretation is straightforward: a low *A* index indicates that a large proportion of branch rates and times are inconsistent with the expected number of substitutions along the branches. Under ideal conditions, an *A* index of 0.95 or higher means that the clock model accurately describes the true pattern of rate variation. However, our method allows the user to inspect the particular branches with inconsistent estimates, which can be useful for identifying regions of the tree that cause the clock model to be inadequate. Measuring the effect size of differences in the branch length estimates of the posterior predictive and empirical data can also be useful for quantifying potential errors in the estimates of node times and branch-specific rates.

An important finding of our study is that over-parameterized clock-models typically have higher accuracy than those that are under-parameterized. This is consistent with a statistical phenomenon known as the bias-variance trade-off, with under-parameterization leading to high bias, and over-parameterization leading to high uncertainty. This was demonstrated for molecular clock models by Wertheim *et al.* (2009). Although our results show a bias when the model is under-parameterized, we did not detect high uncertainty with increasing model complexity. This probably occurs because the models used here are not severely over-parameterized. This is consistent

with the fact that Bayesian analyses are robust to mild over-parameterization because estimates are integrated over the uncertainty in additional parameters (Lemmon & Moriarty 2004; Huelsenbeck & Rannala 2004).

We note that our index is insensitive to the over-parameterization in our analyses. This problem is also present in some adequacy statistics for substitution models (e.g. Bollback 2002; Ripplinger & Sullivan 2010). Identifying an over-parameterized model is challenging, but a recent study proposed a method to do this for substitution models (Lewis *et al.* 2014). An equivalent implementation for clock models would also be valuable. Another potential solution is to select a pool of adequate models and to perform model selection using methods that penalize an excess of parameters, such as marginal likelihoods or information criteria.

We find that our assessment of clock-model adequacy can be influenced by other components of the analysis. For example, multiple calibrations can create a misleading node-age prior that is in conflict with the clock model (Warnock *et al.* 2012; Duchêne *et al.* 2014; Heled & Drummond 2015). Although our simulations with misleading node calibrations were done using a strict clock, our method identified this scenario as clock-model inadequacy when the models for estimation were the strict or random local clocks (Table 6.2a). In the case of the uncorrelated lognormal relaxed clock, our method identified a misleading node-age prior, in which the node-calibrations differed considerably from the true node ages, as causing an increase in uncertainty (Table 6.2b). This highlights the critical importance of selecting and using time calibrations appropriately, and we refer the reader to the comprehensive reviews of this topic (Benton & Donoghue 2007; Ho & Phillips 2009). Another component of the analysis that can have an impact on the adequacy of the clock model is the tree prior, which can

influence the estimates of branch lengths. Although one study suggested that the effect of the tree prior is not substantial (Lepage *et al.* 2007), its influence on divergence-time estimates remains largely unknown.

We found that substitution model under-parameterization led to a severe reduction in accuracy. Over-confidence in incorrect branch lengths in terms of substitutions can cause bias in divergence-time estimates (Cutler 2000). However, this form of model inadequacy is incorrectly identified by the methods we used for estimation as a form of rate variation among lineages. For our data generated using a strict clock and an underparameterized substitution model, the *A* index rejected the strict clock and supported the over-parameterized uncorrelated lognormal relaxed clock. On the other hand, the multinomial test statistic was sensitive to substitution model under-parameterization, and to some forms of rate variation among lineages. The sensitivity of the multinomial likelihood to rate variation among lineages might explain why the substitution model was rejected for the coronavirus data set when using a clock model, but not when using a clock-free method. Due to this sensitivity and the substantial impact of substitution model misspecification, we recommend the use of a clock free method to assess the substitution model before performing analyses using a clock model. Our results suggest that it is only advisable to perform a clock-model analysis when an adequate substitution model is available. Other methods for substitution model assessment that are less conservative than the multinomial likelihood represent an interesting area for further research.

We find that the *A* index is sensitive to patterns of rate variation among lineages that conflict with the clock model used for estimation. This is highlighted in the simulations of rate variation among lineages under autocorrelated and the unusual beta-distributed

rates. In these cases, the *A* index identified the uncorrelated lognormal clock as the only adequate clock model, despite an increase in uncertainty in both cases. Although other studies have also suggested that the uncorrelated lognormal relaxed clock can account for rate autocorrelation (Drummond *et al.* 2006; Ho *et al.* 2015), an increase in uncertainty can impair the interpretation of divergence-time estimates. We suggest caution when the uncertainty values are above 1, which occurs when the widths of the 95% credible intervals are greater than the mean parameter estimates.

In our analyses of the two virus data sets, the multinomial test statistic suggested that the best-fitting substitution model was inadequate. In the analyses of the SIV data, our index of clock model adequacy was 0.78, similar to that of killer whales, for which the substitution model appeared adequate. We recommend caution when interpreting estimates of evolutionary rates and timescales when the substitution model is inadequate. This typically suggests that the substitution process is not being modelled correctly, which can affect inferences of branch lengths regardless of whether a clock model is used or not. For this reason, the *A* index of 0.78 for the SIV data set might be overconfident compared with the same index obtained for the killer whale data. Previous research has also suggested that there are processes in the evolution of SIV that are not accounted for by current evolutionary models (Wertheim & Worobey 2009).

We also found that the all of the clock models were inadequate for the coronavirus sequence data. Our results might provide an explanation for the lack of consensus over the evolutionary timescale of these viruses. For example, a study of mammalian and avian coronaviruses estimated that these viruses originated at most 5,000 years ago (Woo *et al.* 2012). This result stands in contrast with a subsequent study that suggested a much deeper origin of these viruses, in the order of millions of years (Wertheim *et al.*

2013). Our results suggest that estimating the timescale of these viruses might not be feasible with the current clock models.

Our analysis of mitochondrial genomes of killer whales shows that even if the clock-model performance is not as high as that obtained in the simulations that match the models used for estimation, a large proportion of the divergence-time estimates can be useful. Examining the estimates of specific branch lengths can indicate whether many of the node-age estimates are reliable, or whether important branches provide unreliable estimates. We recommend this practice when the substitution model has been deemed adequate and when a substantial proportion of the branch lengths are consistent with the clock model (i.e., when the *A* index is high). We note that the mitochondrial genomes of killer whales have the lowest *A* index of any data set when analysed using a random local clock. This might occur because the model identified an average of 0–3 rate changes along the tree (0.79 rate changes; Table 6.4). While rate variation is likely to be higher in this data set, it might not be sufficiently high for the model to detect it.

Analyses of mitochondrial protein-coding genes from marine turtles identified the substitution model as inadequate using the multinomial test statistic. The clock model with the highest performance had an *A* index of 0.70, which might be considered sufficient to interpret the divergence-time estimates for at least some portions of the tree. Again, the fact that the substitution model is inadequate precludes further interpretation of the estimates of evolutionary rates and timescales. This is a surprising result for a mitochondrial data set with several internal-node calibrations. A potential solution is to assess substitution-model adequacy for individual genes and to conduct the molecular clock analysis using only those genes for which an adequate substitution

model is available. We believe that, with the advent of genomic data sets, this will become a feasible strategy in the near future.

Some of the reasons for the paucity of studies that assess model adequacy in phylogenetics include computational demand and the lack of available methods. In this study, we have presented a method of evaluating clock-model adequacy, using a simple test statistic that can be computed efficiently. Assessment of clock-model adequacy is an important complement to traditional methods of model selection for two primary reasons: it allows the researcher to reject all of the available models if they are inadequate; and, as implemented in this study, it can be used to identify the branches with length estimates that are implausible under the assumed model. The results of our analyses of empirical data underscore the importance of evaluating the adequacy of the substitution and clock models. In some cases, several models might be adequate, particularly when they are over-parameterized. In this respect, methods for traditional model selection are important tools because they can be used to select a single best-fitting model from a set of adequate models. Further research into methods, test statistics, and software for evaluating model adequacy is needed, both to improve the existing models and to identify data sets that will consistently provide unreliable estimates.

6.6 References

Baele G, Li WLS, Drummond AJ, Suchard MA, Lemey P (2013) Accurate model selection of relaxed molecular clocks in bayesian phylogenetics. *Molecular Biology and Evolution*, **30**, 239–243.

Benton MJ, Donoghue PCJ (2007) Paleontological evidence to date the tree of life. *Molecular Biology and Bvolution*, **24**, 26–53.

Bollback JP (2002) Bayesian model adequacy and choice in phylogenetics. *Molecular Biology and Evolution*, **19**, 1171–1180.

Bouckaert R, Heled J, Kühnert D *et al.* (2014) BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, **10**, e1003537.

Brown JM (2014a) Detection of implausible phylogenetic inferences using posterior predictive assessment of model fit. *Systematic Biology*, **63**, 334–348.

Brown JM (2014b) Predictive approaches to assessing the fit of evolutionary models. *Systematic Biology*, **63**, 289–292.

Brown JM, ElDabaje R (2009) PuMA: Bayesian analysis of partitioned (and unpartitioned) model adequacy. *Bioinformatics*, **25**, 537–538.

Cutler DJ (2000) Estimating divergence times in the presence of an overdispersed molecular clock. *Molecular Biology and Evolution*, **17**, 1647–1660.

Dornburg A, Brandley MC, McGowen MR, Near TJ (2012) Relaxed clocks and inferences of heterogeneous patterns of nucleotide substitution and divergence time estimates across whales and dolphins (Mammalia: Cetacea). *Molecular Biology and Evolution*, **29**, 721–736.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biology*, **4**, e88.

Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, **22**, 1185–1192.

Drummond AJ, Suchard MA (2010) Bayesian random local clocks, or one rate to rule them all. *BMC Biology*, **8**, 114.

Duchene S, Frey A, Alfaro-Núñez A *et al.* (2012) Marine turtle mitogenome phylogenetics and evolution. *Molecular Phylogenetics and Evolution*, **65**, 241–250.

Duchêne S, Lanfear R, Ho SYW (2014) The impact of calibration and clock-model choice on molecular estimates of divergence times. *Molecular Phylogenetics and Evolution*, **78**, 277–289.

Dutheil JY, Galtier N, Romiguier J *et al.* (2012) Efficient selection of branch-specific models of sequence evolution. *Molecular Biology and Evolution*, **29**, 1861–1874.

Felsenstein J (1983) Statistical inference of phylogenies. *Journal of the Royal Statistical Society Series A*, 246–272.

Foster P (2004) Modeling compositional heterogeneity. *Systematic Biology*, **53**, 485–495.

Gatesy J (2007) A tenth crucial question regarding model use in phylogenetics. *Trends in Ecology & Evolution*, **274**, 3–14.

Gelman A, Carlin JB, Stern HS, Rubin DB (2014) *Bayesian data analysis*. Taylor & Francis.

Gelman A, Meng X-L (1996) Model checking and model improvement. In: *Markov chain Monte Carlo in practice* (eds Gilks WR, Richardson S, Spiegelhalter DJ), pp. 189–201. Chapman and Hall, New York.

Gelman A, Meng X-L (1998) Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical Science*, **13**, 163–185.

Gelman A, Shalizi CR (2013) Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology*, **66**, 8–38.

Goldman N (1993) Statistical tests of models of DNA substitution. *Journal of Molecular Evolution*, **36**, 182–198.

Heath TA, Huelsenbeck JP, Stadler T (2014) The fossilized birth-death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 2957–2966.

Heled J, Drummond AJ (2015) Calibrated birth-death phylogenetic time-tree priors for bayesian inference. *Systematic Biology*, **64**, 369–383.

Ho SYW (2014) The changing face of the molecular evolutionary clock. *Trends in Ecology & Evolution*, **29**, 496–503.

Ho SYW, Duchêne S (2014) Molecular-clock methods for estimating evolutionary rates and timescales. *Molecular Ecology*, **23**, 5947–5965.

Ho SYW, Duchêne S, Duchêne DA (2015) Simulating and detecting autocorrelation of molecular evolutionary rates among lineages. *Molecular Ecology Resources*, **15**, 688–696.

Ho SYW, Phillips MJ (2009) Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Systematic Biology*, **58**, 367–380.

Huelsenbeck J, Rannala B (2004) Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Systematic Biology*, **53**, 904–913.

Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. *Science*, **294**, 2310–2314.

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: *Mammalian Protein Metabolism* (ed Munro H), pp. 21–132. Academic Press, New York.

Kishino H, Thorne JL, Bruno WJ (2001) Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Molecular Biology and Evolution*, **18**, 352–361.

Kumar S (2005) Molecular clocks: four decades of evolution. *Nature Reviews Genetics*, **6**, 654–662.

Lartillot N, Philippe H (2006) Computing Bayes factors using thermodynamic integration. *Systematic Biology*, **55**, 195–207.

Lemmon AR, Moriarty EC (2004) The importance of proper model assumption in bayesian phylogenetics. *Systematic Biology*, **53**, 265–277.

Lepage T, Bryant D, Philippe H, Lartillot N (2007) A general comparison of relaxed molecular clock models. *Molecular Biology and Evolution*, **24**, 2669–2680.

Lewis PO, Xie W, Chen M-H, Fan Y, Kuo L (2014) Posterior predictive Bayesian phylogenetic model selection. *Systematic Biology*, **63**, 309–321.

Morin PA, Archer FI, Foote AD *et al.* (2010) Complete mitochondrial genome phylogeographic analysis of killer whales (Orcinus orca) indicates multiple species. *Genome Research*, **20**, 908–916.

Murphy WJ, Eizirik E, O'Brien SJ *et al.* (2001) Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science*, **294**, 2348–2351.

Nielsen R (2002) Mapping mutations on phylogenies. *Systematic Biology*, **51**, 729–739.

Nylander J, Ronqiuist F, Huelsenbeck J, Nieves-Aldery J (2004) Bayesian phylogenetic analysis of combined data. *Systematic Biology*, **53**, 47–67.

Pereira SL, Baker AJ (2006) A mitogenomic timescale for birds detects variable phylogenetic rates of molecular evolution and refutes the standard molecular clock. *Molecular Biology and Evolution*, **23**, 1731–1740.

Plummer M, Best N, Cowles K, Vines K (2006) CODA: convergence diagnosis and output analysis for MCMC. *R News*, **6**, 7–11.

Reid NM, Hird SM, Brown JM *et al.* (2014) Poor fit to the multispecies coalescent is widely detectable in empirical data. *Systematic Biology*, **63**, 322–333.

Ripplinger J, Sullivan J (2010) Assessment of substitution model adequacy using frequentist and Bayesian methods. *Molecular Biology and Evolution*, **27**, 2790–2803.

Rodrigue N, Kleinman CL, Philippe H, Lartillot N (2009) Computational methods for evaluating phylogenetic models of coding sequence evolution with dependence between codons. *Molecular Biology and Evolution*, **26**, 1663–1676.

Ronquist F, Teslenko M, Van Der Mark P *et al.* (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539–542.

Schliep KP (2011) PHANGORN: phylogenetic analysis in R. *Bioinformatics*, **27**, 592–593.

Sullivan J, Joyce P (2005) Model selection in phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, 445–466.

Thorne JL, Kishino H, Painter IS (1998) Estimating the rate of evolution of the rate of molecular evolution. *Molecular Biology and Evolution*, **15**, 1647–1657.

Warnock RCM, Yang Z, Donoghue PCJ (2012) Exploring uncertainty in the calibration of the molecular clock. *Biology Letters*, **8**, 156–159.

Wertheim JO, Chu DKW, Peiris JSM, Pond SLK, Poon LLM (2013) A case for the ancient origin of coronaviruses. *Journal of Virology*, **87**, 7039–7045.

Wertheim JO, Sanderson MJ, Worobey M, Bjork A (2009) Relaxed molecular clocks, the bias–variance trade-off, and the quality of phylogenetic inference. *Systematic Biology*, **59**, 1–8.

Wertheim JO, Worobey M (2009) Dating the age of the SIV lineages that gave rise to HIV-1 and HIV-2. *PLoS Computational Biology*, **5**, e1000377.

Woo PCY, Lau SKP, Lam CSF *et al.* (2012) Discovery of seven novel mammalian and avian coronaviruses in Deltacoronavirus supports bat coronaviruses as the gene source of Alphacoronavirus and Betacoronavirus and avian coronaviruses as the gene source of Gammacoronavirus and Deltacoronavirus. *Journal of Virology*, JVI–06540.

Xie W, Lewis PO, Fan Y, Kuo L, Chen M-H (2011) Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Systematic Biology*, **60**, 150–160.

Yang Z, Rannala B (2012) Molecular phylogenetics: principles and practice. *Nature Reviews Genetics*, **13**, 303–314.

Yoder AD, Yang Z (2000) Estimation of primate speciation dates using local molecular clocks. *Molecular Biology and Evolution*, **17**, 1081–1090.

Chapter 7 – General Discussion

Statistical phylogenetics is a widely used tool to understand processes that range from molecular evolution to macroevolution. One feature in studies of molecular evolution and macroevolution in the last decade is the increasing reliance on phylogenetic inference performed in a Bayesian framework (Yang & Rannala 2012). As data sets become ever larger, inferences using phylogenetic estimation, such as inferences of divergence times or diversification rates, are allowing researchers to test long-standing questions in biology. However, many findings in evolutionary biology have become reliant on methods to estimate phylogeny and evolutionary divergence times. This thesis explores the use of phylogenetics in macroevolution and emphasises the need to be critical of the methods used to perform inferences before relying on them for further interpretation.

7.1 The biological motivation for understanding and improving phylogenetic methods for macroevolution

Identifying the processes that drive the latitudinal diversity gradient of species richness is one of the long-standing questions in biology that can be explored using phylogenetic inference. Chapter 2 of this thesis provides evidence that dispersal across latitudes has been relatively rare, such that it is likely that clades from higher latitudes have had less time to accumulate biodiversity. Phylogenetic data have been primarily used to test the hypothesis of a latitudinal gradient in rates of diversification. While some studies have suggested that indeed birds have a latitudinal gradient in diversification rates (e.g. Cardillo 1999; Cardillo *et al.* 2005; Mittelbach *et al.* 2007), the study with the most

comprehensive sampling of bird species suggested that this is not the case (Jetz *et al.* 2012). The results in chapter 2 of this thesis are consistent with recent studies in taxa such as plants and amphibians suggesting that dispersal and the time available for accumulating species at different latitudes are a primary process driving the latitudinal diversity gradient (Pyron & Wiens 2013; Kerkhoff *et al.* 2014).

Progress towards understanding the role of dispersal as a driver of the latitudinal diversity gradient will continue as models of biogeographic evolution become more accurate, and data become more comprehensive. For example, it would be useful to relax the assumption that the rate of geographic range evolution is constant across lineages. This is because the rate of geographic range evolution in birds has been found to vary depending on climatic factors (Pigot *et al.* 2010) and biotic interactions (Pigot & Tobias 2013). If the assumption of a constant rate of geographic range evolution is frequently violated, estimates of ancestral geographic ranges could be misleading (King & Lee 2015). Progress towards identifying and accounting for rate variation in the evolution of character states is likely to lead to more accurate and precise estimates of dispersal rates across latitudes.

The data used in chapter 2 also contains substantial phylogenetic uncertainty due to missing species, which were included randomly in each posterior sample of phylogeny in the original study (Jetz *et al.* 2012). Missing taxa is probably a cause of the low precision in the estimates of the timing and extent of dispersal across latitudes. Future studies are likely to increase the accuracy and precision of these inferences by improving the sampling of species, as well as increasing the amount of molecular data to a genomic scale (e.g. Jarvis *et al.* 2014; Prum *et al.* 2015).

7.2 The risks in current practice in phylogenetics

Chapter 2 shows an example of the questions that can be answered using molecular phylogenetics using large data sets, but it is also an example of the importance of using reliable methods for inference of phylogeny and divergence times. It is common practice to select methods for inference of phylogeny and evolutionary divergence times based on the statistical fit of candidate methods relative to each other. This approach might be sufficient in some cases, but it carries the risk of favouring models that poorly represent the evolutionary process that generated the data (Goldman 1993a; Ripplinger & Sullivan 2008). Even the most complex method available might describe the data very poorly if parameters that are important to describe the molecular evolutionary processes are not considered (Lemmon & Moriarty 2004; Steel 2005; Wertheim *et al.* 2010). If this is the case, the researcher could have false confidence in potentially misleading inferences (Minin *et al.* 2003; Ripplinger & Sullivan 2008; Spielman & Wilke 2015).

One example of a specific underlying evolutionary process that is not accounted for in current methods is the link between the rate of diversification and the rate of molecular evolution. Chapter 3 shows support for the hypothesis that this link is common in empirical data, in agreement with multiple previous studies (Webster *et al.* 2003; Davies *et al.* 2004; Lanfear *et al.* 2010; Eo & DeWoody 2010). It remains unlikely, however, that the link between the rate of diversification and the rate of molecular evolution is universal (Webster *et al.* 2003; Goldie *et al.* 2011). Chapter 3 further provides support that the rate of molecular evolution plays a role in driving the rate of diversification. Interestingly, there seems to be increasing support for several of the

hypotheses for the link between the rate of diversification and the rate of molecular evolution in different taxa (see Dowle *et al.* 2013 for an overview of these hypotheses). Faster diversification might in some cases drive molecular evolution (e.g. Pagel *et al.* 2006; Venditti & Pagel 2010), or vice-versa, as suggested from chapter 3 and other studies (e.g. Lanfear *et al.* 2010; Bromham *et al.* 2015), or a third factor, such as UV radiation, might frequently have an influence on both the rate of diversification and the rate of molecular evolution (e.g. Davies 2004; Gillman & Wright 2013).

Future research might provide an improved resolution of the importance of each of the possible explanations for the link between the rate of diversification and the rate of molecular evolution. For example, it would be interesting to investigate whether the results from chapter 3, that faster rates of synonymous substitutions are linked to faster net rates of diversification, are universal across the genome. This kind of analysis will be increasingly feasible as phylogenomic data become more common (e.g. Jarvis *et al.* 2014; Prum *et al.* 2015). Similarly, statistical approaches will also have more power to disentangle the directionality in the relationship between the rate of diversification and the rate of molecular evolution. An example of progress in this respect is a recent study that found that longevity can drive rates of molecular evolution in rockfish using a Poisson regression method (Hua *et al.* 2015). This new statistical advance increases the power of sister clade comparisons because it allows the inclusion of sister pairs that would normally need to be excluded due to a lack of sufficient information about substitution rates (Welch & Waxman 2008). Advances in statistical techniques such as this one might provide a better performance to identify the drivers of the link between the rate of diversification and the rate of molecular evolution compared to the traditional sister-clade comparison used in chapter 3.

One important reason to identify the causes of variation in rates of molecular evolution is that these processes can have an influence on estimates of phylogeny and divergence times (Welch & Bromham 2005). Chapter 4 shows that failing to account for the link between the rate of diversification and the rate of molecular evolution can lead to misleading inferences about macroevolutionary processes. Regions of the phylogeny that underwent fast diversification might have an over-representation of molecular substitutions in the data, while regions with slow diversification might have an under-representation of molecular substitutions. While the possible nature of this bias is intuitive, the solution might be more elusive.

It might be difficult to account for a link between the rate of diversification and the rate of molecular evolution because the outcome in the data is similar to that of the node density effect (Webster *et al.* 2003). The node density effect occurs when phylogenetic inference methods fail to account for multiple substitutions due to a lack of sampling in a particular clade (Hugall & Lee 2007). Some studies claim that the link between the rate of diversification and the rate of molecular evolution can be differentiated from the node density effect (Webster *et al.* 2003; Venditti *et al.* 2006). However, a model-based approach for phylogenetic inference that accounts for the link between the rate of diversification and the rate of molecular evolution while accounting for the confounding effects of differing node densities across lineages has not yet been proposed.

Also problematic is the bias on estimates of phylogenetic divergence times caused by phylogenetic imbalance, shown in chapter 5. It is more difficult to determine how phylogenetic imbalance operates to produce biased inferences of evolutionary divergence times, so a model-based approach to account for this bias seems elusive. After the publication of chapter 5, a study suggested that the bias caused by

phylogenetic imbalance is linked to the fact that imbalanced phylogenies tend to have more small clades than balanced phylogenies (Murray *et al.* 2015). In that study, the authors suggest that in analyses with "dated tips", in which samples are taken from different points in time, the temporal and genetic structures are often confounded because close relatives tend to share sampling times (Murray *et al.* 2015). Lineages in smaller clades are more likely to share similar sampling times by chance, so the confounding effect of genetic and temporal structure becomes pronounced as phylogenetic imbalance increases. Fortunately, it might be possible to identify this bias by testing the power of the time differences among samples to recover rates of substitution (Ramsden *et al.* 2008; Duchêne *et al.* 2015; Murray *et al.* 2015).

A question raised by the results of chapters 3, 4, and 5, is the extent to which the most popular models of molecular rate variation across lineages are sufficient to account for some of the known sources of biased inference. Instead, models that are mechanistic and describe specific processes might be more appropriate. Flexible models of molecular rate variation among lineages are frequently proposed (Drummond & Suchard 2010; Heath *et al.* 2012; To *et al.* 2016), and are generally considered to be robust. For example, one study found that a model that described rates across lineages to be heritable was difficult to detect, while a more general model with uncorrelated rates across lineages had similar statistical performance (Ho *et al.* 2015). It would be interesting to make comparisons in future studies between models of molecular rate variation across lineages that assume some extent of independent rates across lineages with the performance of models that explicitly account for specific molecular evolutionary processes, such as the process discussed in chapters 3 and 4.

7.3 Improving phylogenetic inference

One way forward is to use methods that allow the rejection of all available models when they are inadequate for describing the aspects of the data that are relevant for inferences of phylogeny and evolutionary divergence times. This kind of assessment of the absolute model performance is also known as model checking, or assessment of model adequacy or plausibility (Gelman & Meng 1996). Methods that assess absolute model performance have been advocated in phylogenetics for more than two decades (e.g. Penny *et al.* 1992; Goldman 1993b; Bollback 2002; Brown 2014), but the uptake of these methods has been minimal. The studies that have used or developed methods for selecting models based on their relative statistical fit greatly outnumber the studies involving assessment of model adequacy (Figure 7.1). The scarcity and low uptake of methods of assessing model adequacy is a critical weakness in the current practice of likelihood and Bayesian phylogenetics.

Chapter 6 proposes a new method to assess the absolute performance of models of rate variation across lineages. This method was inspired by recently developed methods for assessing the adequacy of substitution models (Brown 2014a) and the multispecies coalescent process (Reid *et al.* 2014). Although the method proposed in chapter 6 can identify many instances of model inadequacy, it is worth noting that this method relies on a reasonably realistic substitution model and calibration scheme. Similarly, inadequacies identified by the method in chapter 6 might derive from components of the analyses other than the model of rate variation across lineages, such as the prior on node ages (e.g. Yule or birth-death process; Condamine *et al.* 2015).

Despite the reliance on other components of the hierarchical model, methods such as that proposed in chapter 6 are a promising avenue for improving the reliability of phylogenetic and evolutionary divergence time estimates. Specifically, the growth of genomic data allows researchers to select the data that are adequately described by the available models. For example, one recent study on birds used this kind of data filtering to obtain the loci that presented "clocklike" behaviour, such that the rate of substitution was approximately constant across lineages (Jarvis *et al.* 2014).

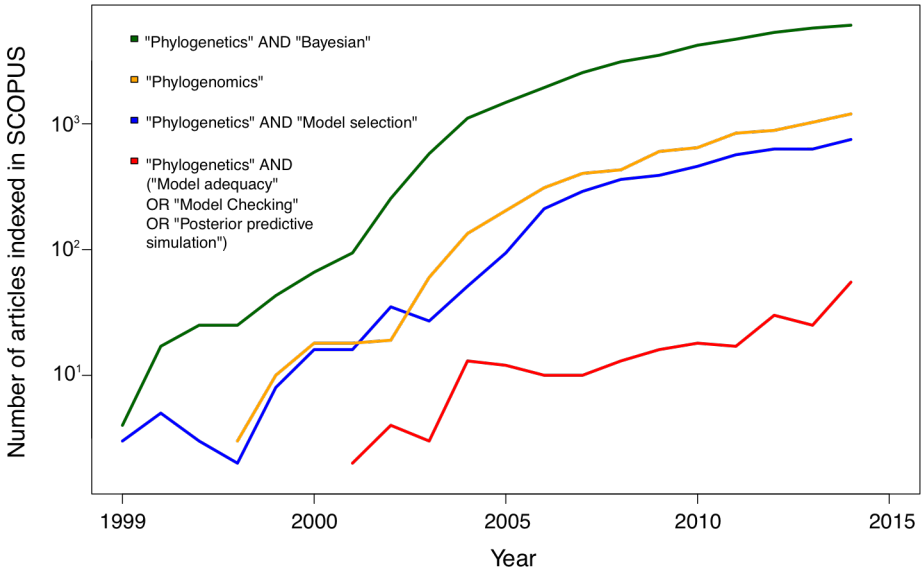


Figure 7.1. The number of publications containing the key search terms (in $\log_{10}$ scale) in SCOPUS relevant to phylogenetics, phylogenomics, and model adequacy in the past two decades.

Data filtering using objective methods, such as the model checking approach described in chapter 6, provides multiple benefits and could become an important step in phylogenomic analyses. Current evolutionary models are unlikely to be adequate for every genomic region. Genomic regions have different amounts of information (e.g. Springer & Gatesy 2015), follow different patterns in molecular rate variation across lineages (Duchêne & Ho 2015), and might be affected to different extents by processes such as the link between the rate of diversification and the rate molecular evolution. In

addition, even analysing a small subset of the genome can lead to informative and useful inferences of evolutionary parameters, as has been observed in estimates of topology (Streicher *et al.* 2015) and divergence times (Yang & Rannala 2006). Given a surplus of data, it becomes feasible to filter the genome data according to some criterion, such as model adequacy. Data filtering also brings the benefit of reducing the computational burden associated with analysing very large data sets. The remaining data, for which available models might not be adequate, can be investigated further and used to develop new models.

Assessing the absolute performance, and more generally the reliability of phylogenetic methods using simulation, can improve our confidence in the resulting phylogenetic estimates. Methods to assess whether current models are reasonable for phylogenetic inference can naturally lead to model improvement, and can adopt the following procedure: (i) an existing method or model is evaluated. For example, a method that explicitly accounts for a link between the rate of diversification and the rate of molecular evolution. (ii) If the method or model has poor performance, the data can be explored to understand the factors that are not accounted for by the method or model in question. This assessment can be done, for example, by comparing empirical data with data simulated from existing models, or by assessing the performance of existing methods using data simulated under new hypotheses. This thesis largely focuses on approaches that can be used make this assessment. (iii) A new method or model is then proposed according to the findings from the data; (iv) if the new method or model shows an improvement in performance to a certain tolerance level (e.g. using a *P*-value), implying higher absolute statistical fit, then the inferences from the data are likely to become more accurate and precise. Steps (iii) and (iv) can be repeated by adding or replacing parameters until the model achieves an acceptable performance.

This method of refinement does not entail that future methods or models must always become more complex. Instead, it is aimed at finding a set of parameters that better describe the processes that generated the data, and that are relevant to the inferences of interest such as phylogeny and evolutionary divergence times (Steel 2005). It is clear that there is room to develop a better understanding of molecular evolutionary processes, new methods of model assessment, and new models to make inferences of phylogeny and divergence times. A change in the current practice in model choice is likely to improve the reliability of phylogenetic inference and, in turn, our knowledge about evolutionary processes and the Tree of Life.

7.5 References

Bollback JP (2002) Bayesian model adequacy and choice in phylogenetics. *Molecular Biology and Evolution*, **19**, 1171–1180.

Bromham L, Hua X, Lanfear R, Cowman PF (2015) Exploring the relationships between mutation rates, life history, genome size, environment, and species richness in flowering plants. *The American Naturalist*, **185**, 507–524.

Brown JM (2014a) Detection of implausible phylogenetic inferences using posterior predictive assessment of model fit. *Systematic Biology*, **63**, 334–348.

Brown JM (2014b) Predictive approaches to assessing the fit of evolutionary models. *Systematic Biology*, **63**, 289–292.

Cardillo M (1999) Latitude and rates of diversification in birds and butterflies. *Proceedings of the Royal Society B: Biological Sciences*, **266**, 1221–1225.

Cardillo M, Orme C, Owens I (2005) Testing for latitudinal bias in diversification rates: An example using new world birds. *Ecology*, **86**, 2278–2287.

Condamine FL, Nagalingum NS, Marshall CR, Morlon H (2015) Origin and diversification of living cycads: a cautionary tale on the impact of the branching process prior in Bayesian molecular dating. *BMC Evolutionary Biology*, **15**, 65.

Davies T (2004) Environmental energy and evolutionary rates in flowering plants. *Proceedings of the Royal Society B: Biological Sciences*, 2195–2200.

Davies TJ, Savolainen V, Chase MW, Moat J, Barraclough TG (2004) Environmental energy and evolutionary rates in flowering plants. *Proceedings of the Royal Society B: Biological Sciences*, **271**, 2195–2200.

Dowle EJ, Morgan-Richards M, Trewick SA (2013) Molecular evolution and the latitudinal biodiversity gradient. *Heredity*, **110**, 501–510.

Drummond AJ, Suchard MA (2010) Bayesian random local clocks, or one rate to rule them all. *BMC Biology*, **8**, 114.

Duchêne S, Duchêne D, Holmes EC, Ho SYW (2015) The performance of the date-randomization test in phylogenetic analyses of time-structured virus data. *Molecular Biology and Evolution*, **32**, 1895–906.

Duchêne S, Ho SYW (2015) Mammalian genome evolution is governed by multiple pacemakers. *Bioinformatics*, **31**, 2061–2065.

Eo SH, DeWoody JA (2010) Evolutionary rates of mitochondrial genomes correspond to diversification rates and to contemporary species richness in birds and reptiles. *Proceedings of the Royal Society B: Biological Sciences*, **277**, 3587–3592.

Gelman A, Meng X-L (1996) Model checking and model improvement. In: *Markov chain Monte Carlo in practice* (eds Gilks WR, Richardson S, Spiegelhalter DJ), pp. 189–201. Chapman and Hall, New York.

Gillman LN, Wright SD (2013) Species richness and evolutionary speed: the influence of temperature, water and area (R Ladle, Ed,). *Journal of Biogeography*, **41**, 39–51.

Goldie X, Lanfear R, Bromham L (2011) Diversification and the rate of molecular evolution: no evidence of a link in mammals. *BMC Evolutionary Biology*, **11**, 286.

Goldman N (1993a) Statistical tests of models of DNA substitution. *Journal of Molecular Evolution*, **36**, 182–198.

Goldman N (1993b) Simple diagnostic statistical tests of models for DNA substitution. *Journal of Molecular Evolution*, **37**, 650–661.

Heath TA, Holder MT, Huelsenbeck JP (2012) A dirichlet process prior for estimating lineage-specific substitution rates. *Molecular Biology and Evolution*, **29**, 939–955.

Ho SYW, Duchêne S, Duchêne D (2015) Simulating and detecting autocorrelation of molecular evolutionary rates among lineages. *Molecular Ecology Resources*, **15**, 688–696.

Hua X, Cowman P, Warren D, Bromham L (2015) Longevity is linked to mitochondrial mutation rates in rockfish: A test using Poisson regression. *Molecular Biology and Evolution*, **32**, 2633–2645.

Hugall AF, Lee MSY (2007) The likelihood node density effect and consequences for evolutionary studies of molecular rates. *Evolution*, **61**, 2293–2307.

Jarvis ED, Mirarab S, Aberer AJ *et al.* (2014) Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, **346**, 1320–1331.

Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO (2012) The global diversity of birds in space and time. *Nature*, **491**, 444–448.

Kerkhoff AJ, Moriarty PE, Weiser MD (2014) The latitudinal species richness gradient in New World woody angiosperms is consistent with the tropical conservatism hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 8125–8130.

King B, Lee MSY (2015) Ancestral state reconstruction, rate heterogeneity, and the evolution of reptile viviparity. *Systematic Biology*, **64**, 532–544.

Lanfear R, Ho SYW, Love D, Bromham L (2010) Mutation rate is linked to diversification in birds. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 20423–20428.

Lemmon AR, Moriarty EC (2004) The importance of proper model assumption in bayesian phylogenetics. *Systematic Biology*, **53**, 265–277.

Minin V, Abdo Z, Joyce P, Sullivan J (2003) Performance-based selection of likelihood models for phylogeny estimation. *Systematic Biology*, **52**, 674–683.

Mittelbach GG, Schemske DW, Cornell H V *et al.* (2007) Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecology Letters*, **10**, 315–331.

Murray GG, Wang F, Harrison EM *et al.* (2015) The effect of genetic structure on molecular dating and tests for temporal signal. *Methods in Ecology and Evolution*, DOI:10.1111/2041–210X.12466.

Pagel M, Venditti C, Meade A (2006) Large punctuational contribution of speciation to evolutionary divergence at the molecular level. *Science*, **314**, 119–121.

Penny D, Hendy MD, Steel MA (1992) Progress with methods for constructing evolutionary trees. *Trends in Ecology & Evolution*, **7**, 73–79.

Pigot AL, Owens IPF, Orme CDL (2010) The environmental limits to geographic range expansion in birds. *Ecology Letters*, **13**, 705–715.

Pigot AL, Tobias JA (2013) Species interactions constrain geographic range expansion over evolutionary time. *Ecology Letters*, **16**, 330–338.

Prum RO, Berv JS, Dornburg A *et al.* (2015) A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*, **526**, 569–573.

Pyron RA, Wiens JJ (2013) Large-scale phylogenetic analyses reveal the causes of high tropical amphibian diversity. *Proceedings of the Royal Society B: Biological Sciences*, **280**, 20131622.

Ramsden C, Melo FL, Figueiredo LM, Holmes EC, Zanotto PMA (2008) High rates of molecular evolution in hantaviruses. *Molecular biology and evolution*, **25**, 1488–92.

Reid NM, Hird SM, Brown JM *et al.* (2014) Poor fit to the multispecies coalescent is widely detectable in empirical data. *Systematic Biology*, **63**, 322–333.

Ripplinger J, Sullivan J (2008) Does choice in model selection affect maximum likelihood analysis? *Systematic Biology*, **57**, 76–85.

Spielman SJ, Wilke CO (2015) The relationship between dN/dS and scaled selection coefficients. *Molecular Biology and Evolution*, **32**, 1097–1108.

Springer MS, Gatesy J (2015) The gene tree delusion. *Molecular Phylogenetics and Evolution*, **94**, 1–33.

Steel M (2005) Should phylogenetic models be trying to “fit an elephant”? *Trends in Genetics*, **21**, 307–309.

Streicher JW, Schulte JA, Wiens JJ (2015) How should genes and taxa be sampled for phylogenomic analyses with missing data? An empirical study in Iguanian lizards. *Systematic Biology*, **65**, 128–145.

To T-H, Jung M, Lycett S, Gascuel O (2016) Fast dating using least-squares criteria and algorithms. *Systematic Biology*, **65**, 82–97.

Venditti C, Meade A, Pagel M (2006) Detecting the node-density artifact in phylogeny reconstruction. *Systematic Biology*, **55**, 637–643.

Venditti C, Pagel M (2010) Speciation as an active force in promoting genetic evolution. *Trends in Ecology & Evolution*, **25**, 14–20.

Webster AJ, Payne RJH, Pagel M (2003) Molecular phylogenies link rates of evolution and speciation. *Science*, **301**, 478.

Welch JJ, Bromham L (2005) Molecular dating when rates vary. *Trends in Ecology & Evolution*, **20**, 320–327.

Welch JJ, Waxman D (2008) Calculating independent contrasts for the comparative study of substitution rates. *Journal of Theoretical Biology*, **251**, 667–678.

Wertheim JO, Sanderson MJ, Worobey M, Bjork A (2010) Relaxed molecular clocks, the bias-variance trade-off, and the quality of phylogenetic inference. *Systematic Biology*, **59**, 1–8.

Yang Z, Rannala B (2006) Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Molecular Biology and Evolution*, **23**, 212–226.

Yang Z, Rannala B (2012) Molecular phylogenetics: principles and practice. *Nature Reviews Genetics*, **13**, 303–314.

Appendices

Appendix – Abbreviations

- ACLN. The Auto-Correlated Log Normal clock model describes the rate of substitutions across lineages as an inheritable process, such that adjacent branches in a phylogeny have similar rates. The rate in a given branch is drawn from a log normal distribution with mean equal to the rate of the ancestral branch.
- AIC. The Akaike Information Criterion is based on information theory and is the likelihood of a model penalised for the number of parameters it has, and is primarily used for model comparison and selection.
- ASFV. African Swine Fever Virus.
- BDM. Bateson-Dobzhansky-Muller incompatibilities are genetic changes among closely related populations that lead to hybrid incompatibility.
- CR. Constant Rates models, such as the Yule or birth-death processes assume that the rates of speciation and extinction are constant across lineages and through time.
- DEC. The Dispersal, Extinction, Cladogenesis model describes discrete geographic evolution, assuming that lineages can either remain in their ancestral range, split from part of their ancestral range, or evenly split their ancestral range in two. It does not allow migration to new ranges, which is allowed by the DEC+J model.
- DNA. Deoxyribonucleic Acid sequences for multiple taxa with homologous sites aligned are used as the data for phylogenetic inference.
- *dN*. Non-Synonymous Rates of substitution are those for the sites in the first and second codon positions, which change the amino-acid that is coded by the codon.

- *dS*. Synonymous Rates of substitution are those for the sites in the third codon positions, which do not have an impact in the amino-acid that is coded by the codon.
- GTR. General-Time Reversible models are a family that describes nucleotide substitution of which GTR is the most complex, allowing every kind of transition among nucleotides to have a free parameter.
- HKY. The Hasegawa-Kishino-Yano substitution model is in the GTR family, and is simpler than the GTR model. It distinguishes between transitions and transversions and the frequencies for each of the bases.
- LDG. The Latitudinal Diversity Gradient is the observation that is more biological diversity near the equator than near the poles.
- LI. The Latitude Index is a variable that describes how tropical or temperate is a given taxon, based on the proportion of its geographic range that overlaps with the tropics. It ranges from -1 for species that occur entirely south of the tropics, through 0 for species that are present only in the tropics, to 1 for species that occur entirely north of the tropics.
- MCMC. Markov Chain Monte Carlo are a class or algorithms for sampling probability distributions, such as those defined in many phylogenetic inference methods, in particular those implemented in a Bayesian framework.
- MRCA. The Most Recent Common Ancestor of a given group of taxa.
- MPD. Mean Pairwise Distance among the taxa in a given group or community. Usually measured in units of time.
- Mya. Millions of years ago.
- N_e. Effective population size is the number of individuals that contribute offspring to the following generation.

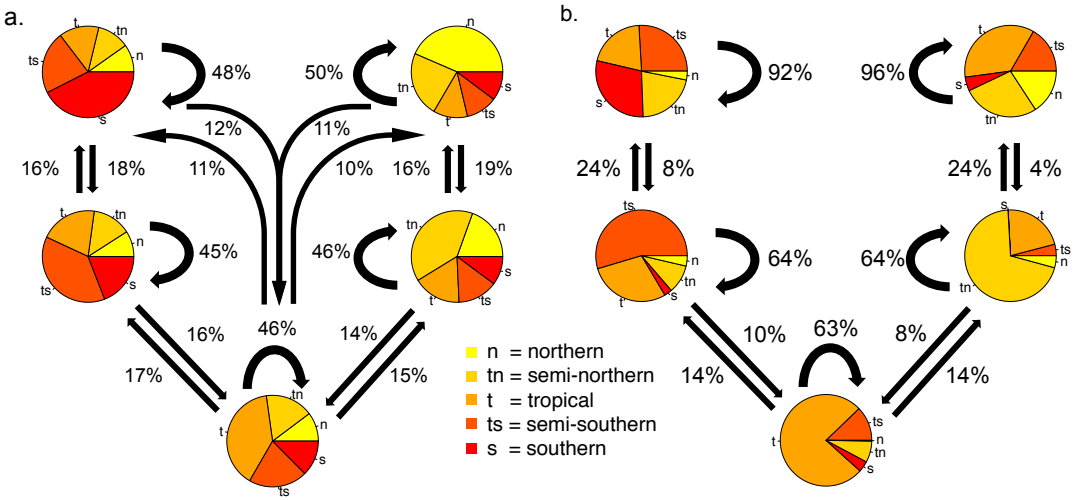
- NPRS. Non-Parametric Rate Smoothing is a method of divergence time estimation that assumes that rates across branches in a phylogeny are auto-correlated.
- NRI. The Net Relatedness Index is similar to the mean pairwise distance among a group or community of taxa. It is standardized using simulation such that it is comparable across groups or communities of different size.
- NT. Northern Temperate species are those with >75% of their geographic range north of the tropics.
- NST. Northern Sub-Tropical species are those with between 25% and 75% of their range north of the tropics.
- OTM. The Out Of The Tropics model states that diversity primarily emerges near the tropics and migrates to high latitudes.
- PD. Phylogenetic Diversity is a measure of biodiversity that takes into count the phylogenetic differences or relatedness among individuals in a community or group.
- SC. Strict Clock model assumes that the rate of substitutions is constant across lineages in a phylogeny.
- SIV. Simian Immunodeficiency Virus.
- SST. Southern Sub-Tropical species are those with between 25% and 75% of their range south of the tropics.
- ST. Southern Temperate species are those with >75% of their geographic range south of the tropics.
- TCH. The Tropical Conservatism Hypothesis states that the majority of current taxa have originated in tropical environments since the earth had was tropical until the mid-tertiary period. With the cooling of the earth, diversity contracted and has a tendency to inhabit the tropics, since it is what most taxa are adapted to.

- UCLN. Uncorrelated Lognormal clock allows molecular substitution rates across lineages to vary in an uncorrelated fashion. These rates are identically and independently drawn from a log normal distribution.
- UV. Ultra-Violet radiation.

Appendix – Chapter 2

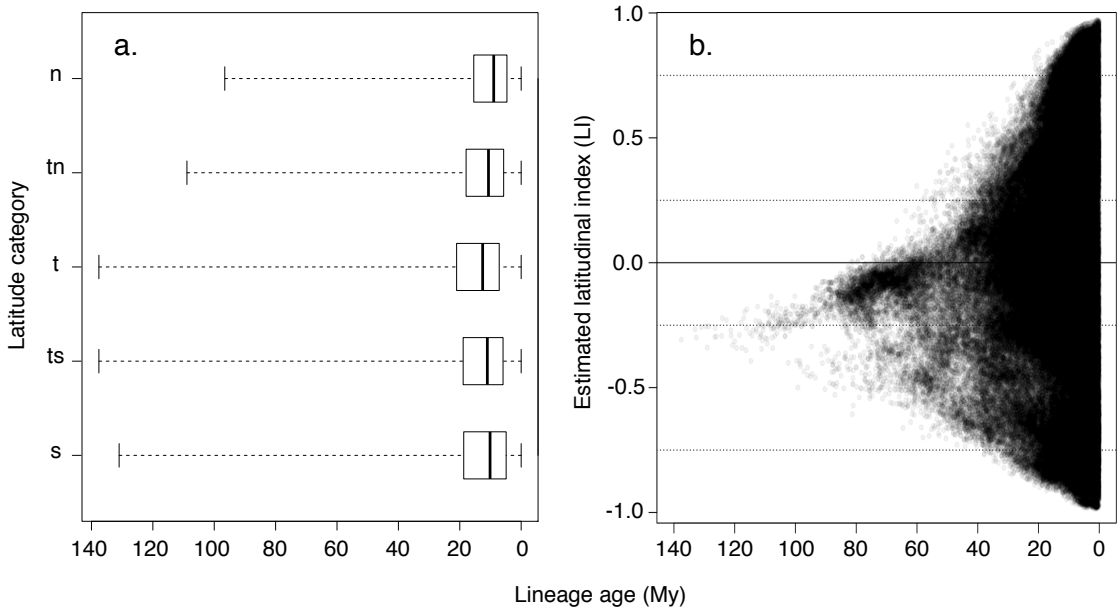
Appendix – Figure A2.1

Proportion of transitions into each category (i.e. ancestry) represented as circle sections from 100 posterior trees, using sub-sampled phylogenies to contain an equal number of species at each category. Arrows show the mean frequencies of evolutionary transitions between latitudinal zones. Values were inferred from reconstructed nodal values of LI using (a) geographic models of range evolution and (b) models of continuous-trait evolution. Transitions percentages below 1% are not shown.



Appendix – Figure A2.2

Ancestral estimates of LI for nodes across node age estimates (My) of 100 posterior trees, made using sub-sampled phylogenies to contain an equal number of species at each category. The diagrams show results for (a) geographic models of range evolution and (b) models of continuous-trait evolution.



Appendix – Table A2.1

Mean support for each of the models tested for 100 posterior trees, for (a) models of geographic range evolution, in which latitude was treated as a categorical variable, and (b) models of continuous-trait evolution.

(a)

Geographic range evolution models	Log likelihood	AICc	Proportion of times selected
DEC	-9379.71	18762.92	0
DEC+J	-6354.76	12715.77	1

(b)

Continuous trait evolution models	Log likelihood	AICc	Proportion of times selected
Ornstein-Uhlenbeck	-5160.78	10327.57	0
Pagel's λ-transformed random walk	-3432.03	6870.06	1
Trend	-9834.45	19674.91	0
Early burst	-10054.10	20114.20	0

Appendix – Chapter 3

Appendix – Table A3.1. Species and GenBank accession numbers used in the present study. Species names are present beside the accession number if sequences for more than one species were used for the same lineage.

Comparison	Clade	Sequence	*atpB*	*atpB-rbcL*	*matK*	*rbcL*	*trnLi*	*trnL-trnF*
1	Persoonia	*Persoonia spp.*			*P. linearis* EU169654	*P. falcata* EU676075		
	Garnieria + Acidonia	*Garnieria spathulaefolia*			EU169628	EU676113		
2	Symphionema	*Symphionema montanum*	AF060394	AF060733	EU169667	DQ875825		
	Agastachys	*Agastachys odorata*	AF060393	AF060717	EU169607	DQ875824		
3	Cenarrhenes + Dilobeia	*Cenarrhenes nitida*	AF060396	AF060746		DQ875827		
	Beaupreopsis	*Beaupreopsis paniculata*				EU642641	EU676044	EU676048
4	Conospermum + Synaphea	*Conospermum spp.*	*C. mitchelli* AF060398	*C. mitchelli* AF060728	*C. taxifolium* EU169617	*C. mitchelli* DQ875829		
	Stirlingia	*Stirlingia latifolia*	AF060397	AF060738	EU169666	DQ875828		
5	Protea	*Protea cynaroides*	DQ875866	AJ699186	EU169658	DQ875837	AJ698243	AJ698150
	Faurea	*Faurea spp.*		*F. macnaughtonii* AJ699262	*F. forficuliflora* EU169625	*F. saligna* EU676072	*F. macnaughtonii* AJ698320	*F. macnaughtonii* AJ698228
6	Petrophile	*Petrophile spp.*	*P. circinata* AF060401	*P. circinata* AF060735	*P. canescens* EU169655	*P. biloba* DQ875832		
	Aulax	*Aulax spp.*		*A. umbellata* AF060732	*A. umbellata* EU169610	*A. cancellata* DQ875863	*A. umbellata* EU676043	*A. umbellata* EU676047
7	Paranomus	*Paranomus spp.*			*P. bracteolaris* EU169653	*P. reflexus* EU676074		
	Vexatorella	*Vexatorella alpina*			EU169672	EU676079		
8	Leucospermum	*Leucospermum spp.*			*L. pedunculatu*	*L. bolusii* AM235083		

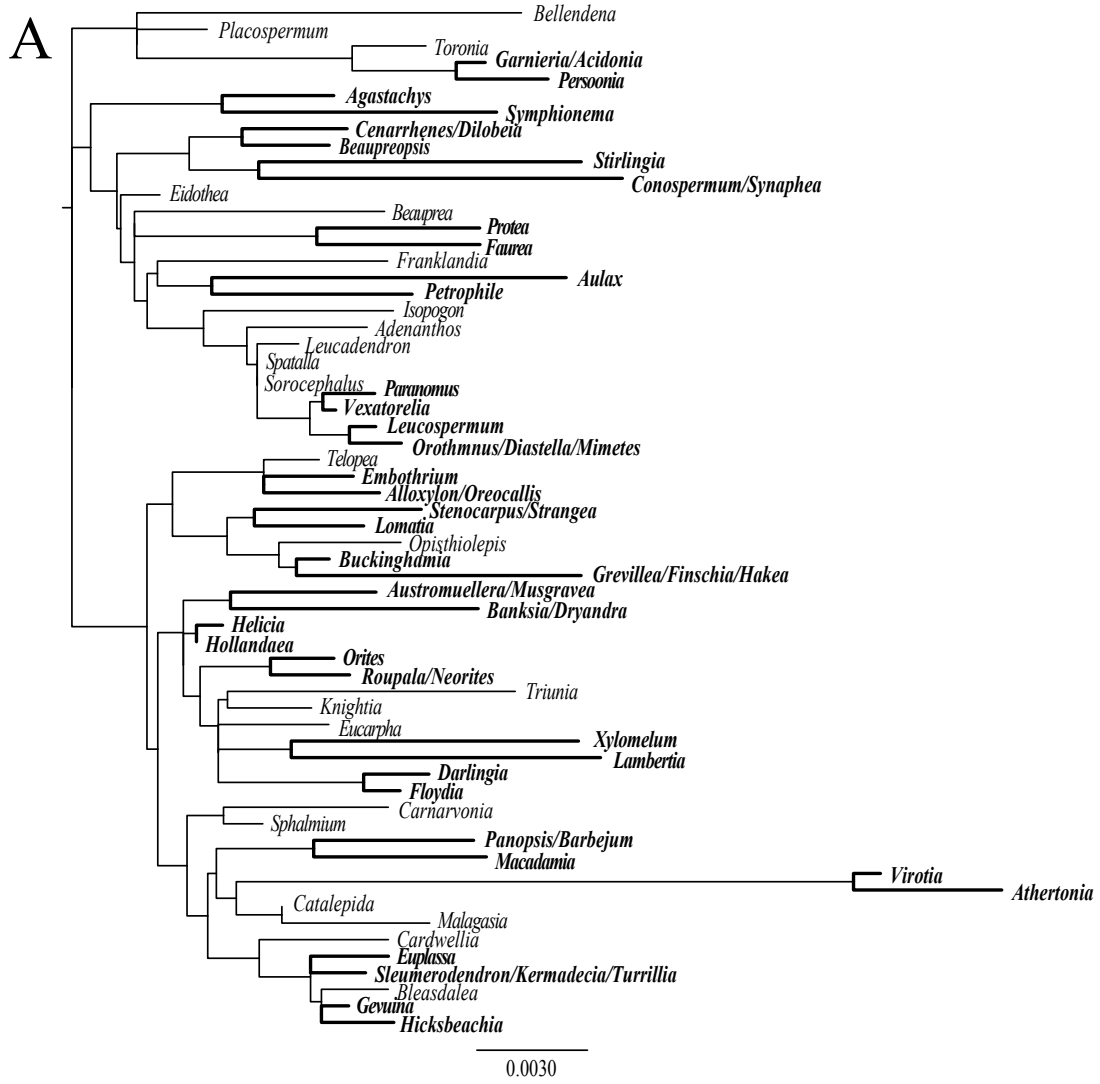
					m EU169642			
	Orothamnus + Diastella + Mimetes	*Mimetes spp.*			*M. hirtus* EU169647	*M. hottentoticus* EU676073		
9	Alloxylon + Oreocallis	*Alloxylon spp.*	*A.wickhamii* AF060428	*A.wickhamii* AF060752	*A. flammelum* EU169608	*A. flammelum* DQ875856		
	Embothrium	*Embothrium coccineum*	AF060429	AF060754	EU169622	DQ875857	AM397162	
10	Stenocarpus + Strangea	*Stenocarpus salignus*	AF060431	AF060743	EU169664	DQ875859	AF482149	AF482194
	Lomatia	*Lomatia spp.*	*L. myricoides* AF060430	*L. myricoides* AF060722	*L. silaifolia* EU169644	*L. silaifolia* U79171	*L. silaifolia* AF482143	*L. silaifolia* AF482188
11	Grevillea + Finschia + Hakea	*Grevillea spp.*	*G. baileyana* AF060434	*G. baileyana* AF060747	*G. robusta* EU169631	*G. robusta* AF197589	*G. banksii* AM397163	*G. banksii* AM397163
	Buckinghamia	*Buckinghamia spp.*	*B. celissima* AF060433	*B. celissima* AF060742	*B. ferruginiflora* EU169614	*B. ferruginiflora* DQ875861	*B. celissima* AF482145	*B. celissima* AF482190
12	Virotia	*Virotia leptophylla*				EU676122		
	Athertonia	*Athertonia diversifolia*			EU169609	EU676108		
13	Panopsis+ Brabejum	*Panopsis spp.*	*P. ferruginea* AF060421	*P. ferruginea* AF060756	*P. yolombo* EU169652	*P. cinnamomea* DQ875850		
	Macadamia	*Macadamia spp.*	*M. integrifolia* AY837827	*M. jansenii* AF060750	*M. integrifolia* AY823204	*M. ternifolia* U79172	*M. integriflora* AF482140	*M. integrifolia* AF482185
14	Hicksbeachia	*Hicksbeachia pinnatifolia*			EU169636	EU676115		
	Gevuina	*Gevuina avellana*				DQ875852		
15	Euplassa	*Euplassas occidentalis*				EU676051	EU676054	EU676059
	Sleumerodendron + Kermadecia + Turrillia	*Sleumerodendron austrocaledonicum + Kermadecia pronyensis*				EU676116	EU676057	EU676062
16	Banksia + Dryandra	*Banksia spp.*	*B. ericifolia* AY837809	*B. cuneata* AF060731	*B. ericifolia* AY823186	*B. ericifolia* DQ875843	*B. ericifolia* AF482126	*B. ericifolia* AF482171
	Austromuellera + Musgravea	*Austromuellera trinervia*	AY837825	AF060720	AY823202	DQ875865	AY823214	AY823219
17	Roupala + Neorites	*Roupala montana + Roupala monosperma + Neorites kevediana*	*N. kevediana* AF060411	*N. kevediana* AF060716	*R. montana* EU169661	*R. monosperma* EU676052	*R. montana* AF482144	*R. montana* AF482189
	Orites	*Orites spp.*	*O. lancifolia* AF060412	*O. lancifolia* AF060718	*O. excelsa* EU169650	*O. myrtoidea* DQ875842	*O. lancifolia* AF482142	*O. lancifolia* AF482187
18	Darlingia	*Darlingia darlingiana*			EU169618	EU676110		
	Floydia	*Floydia praealta*	AF060416	AF060713	EU169626	DQ875845	AF482147	AF482192

19	Lambertia	*Lambertia spp.*	*L. formosa* AF060417	*L. formosa* AF060737	*L. formosa* EU169639	*L. echinata* DQ875846
	Xylomelum	*Xylomelum spp.*	*X. scottianum* AF060418	*X. scottianum* AF060741	*X. angustifolium* EU169673	*X. pyriforme* DQ875847
20	Helicia	*Helicia spp.*	*H. australasica* AF060425	*H. australasica* AF060724		*H. sp*. DQ875853
	Hollandaea	*Hollandaea riparia*	AF060426	AF060751		DQ875854

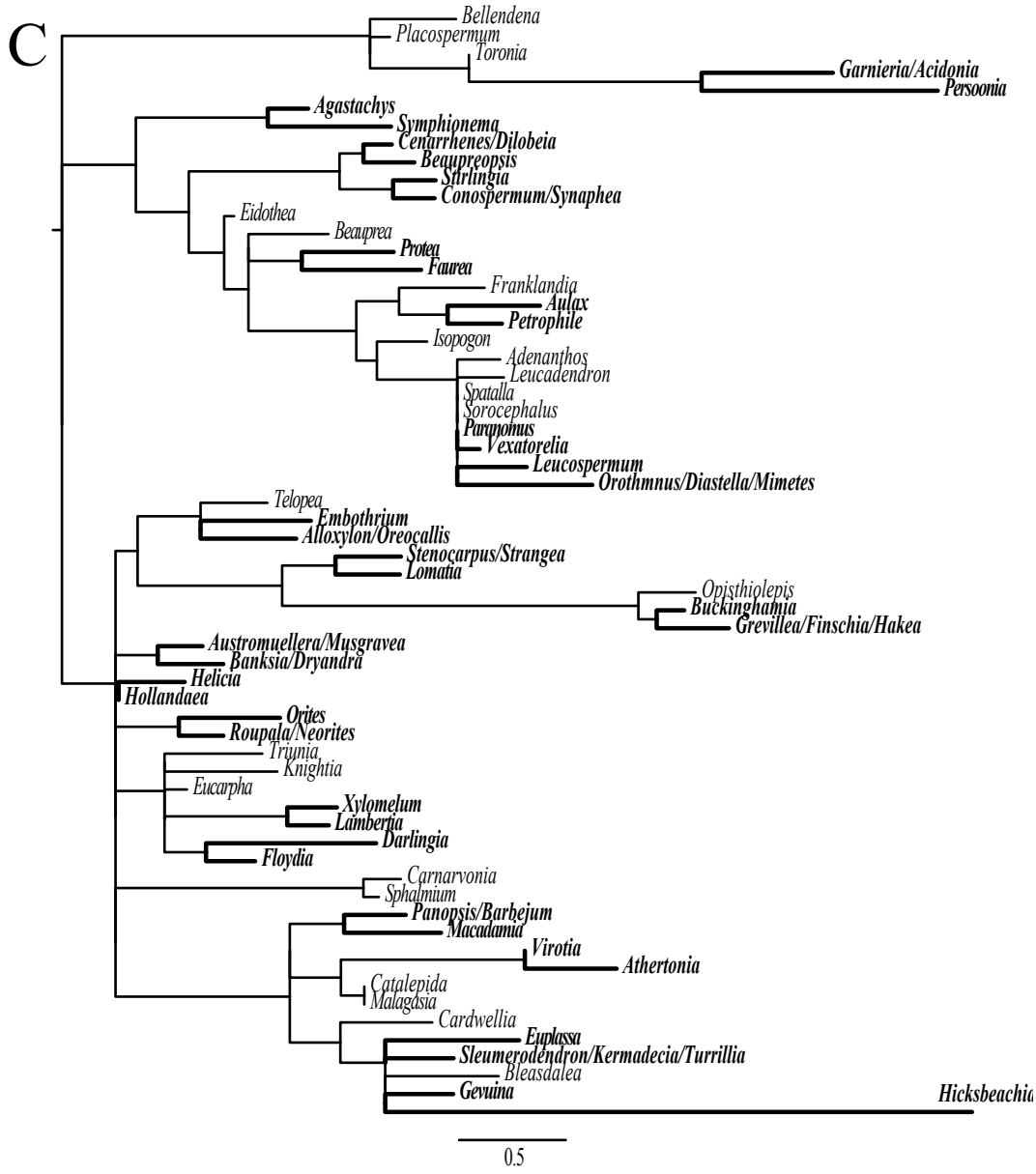
Appendix – Table A3.2. The chloroplast genes used in this study with the final alignment length and the substitution model selected.

Gene	atpB	atpB-rbcL	matK	rbcL	trnLi	trnL-trnF
Substitution model chosen	HKY+Γ	GTR+Γ	GTR+Γ	HKY+Γ+I	HKY+Γ	GTR+Γ
Alignment length	1497	860	1791	1406	554	407
Exon/Intron	Exon	Intron	Exon	Exon	Intron	Intron

Appendix – Figure A3.1. Molecular phylogenies of the family Proteaceae that highlight the 20 sister pair groups used for the present analyses. The branch lengths and scale bars are proportional to (a) the number of non-synonymous substitutions, (b) the number of synonymous substitutions, and (c) the dN/dS branch lengths calculated from the estimates of trees (a) and (b).




B
Bellendena
Placospermum
Toronia
Garnieria/Acidonia
Persoonia
Agastachys
Symphionema
Cenarrhenes/Dilobeia
Beaupreopsis
Stirlingia
Conospermum/Synaphea
Eidothea
Beauprea
Protea
Faurea
Franklandia
Aulax
Petrophile
Isopogon
Adenanthos
Leucadendron
Spatalla
Sorocephalus
Paranomus
Vexatorelia
Leucospermum
Orothamnus/Diastella/Mimetes
Telopea
Embothrium
Alloxylon/Oreocallis
Stenocarpus/Strangea
Lomatia
Opisthiolepis
Buckinghamia
Grevillea/Finschia/Hakea
Austromuellera/Musgravea
Banksia/Dryandra
Helicia
Hollandaea
Orites
Roupala/Neorites
Triunia
Knightia
Eucarpha
Xylomelum
Lambertia
Darlingia
Floydia
Carnarvonia
Sphalmium
Panopsis/Barbejum
Macadamia
Virotia
Athertonia
Catalepida
Malagasia
Cardwellia
Euplassa
Sleumerodendron/Kermadecia/Turrillia
Bleasdalea
Gevuina
Hicksbeachia
0.0090


C
Bellendena
Placospermum
Toronia
Garnieria/Acidonia
Persoonia
Agastachys
Symphionema
Cenarrhenes/Dilobeia
Beaupreopsis
Stirlingia
Conospermum/Synaphea
Eidothea
Beauprea
Protea
Faurea
Franklandia
Aulax
Petrophile
Isopogon
Adenanthos
Leucadendron
Spatalla
Sorocephalus
Paranomus
Vexatorelia
Leucospermum
Orothmnus/Diastella/Mimetes
Telopea
Embothrium
Alloxylon/Oreocallis
Stenocarpus/Strangea
Lomatia
Opisthiolepis
Buckinghamia
Grevillea/Finschia/Hakea
Austromuellera/Musgravea
Banksia/Dryandra
Helicia
Hollandaea
Orites
Roupala/Neorites
Triunia
Knightia
Eucarpha
Xylomelum
Lambertia
Darlingia
Floydia
Carnarvonia
Sphalmium
Panopsis/Barbejum
Macadamia
Virotia
Athertonia
Catalepida
Malagasia
Cardwellia
Euplassa
Sleumerodendron/Kermadecia/Turrillia
Bleasdalea
Gevuina
Hicksbeachia
0.5

Appendix – Chapter 4

Appendix – Supplementary information A4.1 – Literature search

To parameterize the simulations in this study, we focused on the routine practices for estimating phylogenies and diversification rate dynamics through time. We reviewed the literature of studies that estimated diversification rates or diversification dynamics based on molecular phylogenetic estimates. To access the relevant literature, we performed a search in the freely available scholarly literature database Google Scholar. We used the search term *phylogeny "diversification rates"* and extracted data from the first 100 articles published within the last ten years that used phylogenetic estimates to make estimates of diversification rates or diversification rate dynamics. Studies that only used simulated data or that provided only theoretical advances were not considered.

From each article, we recorded data about the phylogenetic estimate and about the method used to estimate diversification rates or diversification rate dynamics. The data recorded about the phylogenetic estimate included the number of sampled tips, the number of loci used for estimation, the number of sequence base pairs in the alignment, the method used for timescale estimation, and the number of time-calibrations used. The data recorded about the estimation of diversification rate or diversification rate dynamics included the method used and the estimates of net diversification rates if available.

A4.1.1 Characteristics of molecular data sets

The mean time of publication of the studies sampled was the second quarter of 2009. Twelve studies did not provide information about the number of loci used, either

because the data was not produced in that study or because they used multiple phylogenies sampled from the literature. In the remaining 88 studies, the number of loci ranged from 1 to 83, and had mean and median values of 5 and 4, respectively (Table A4.1). Twenty-one studies did not provide information about the number of sequence base pairs in the alignment. Alignment length in the remaining 79 studies ranged from 511 to $4.5e^6$ base pairs, with mean and median values of 71100 and 3773 base pairs, respectively (Table A4.1).

Eleven studies did not mention the number of tips sampled in phylogenies. These were studies that included several phylogenies sampled from the literature. The number of tips sampled in the remaining 89 studies ranged from 21 to 55470, with a mean and median of 989 and 133 tips respectively (Table A4.1).

A4.1.2 Methods for phylogenetic inference

At least 9 studies made phylogenetic timescale estimates using multiple methods. Fifty-two studies used non-parametric rate smoothing as implemented in the software package r8s (Sanderson 2003), while 38 studies used the uncorrelated lognormal distribution of rates as implemented in BEAST (Drummond & Rambaut 2007), and 6 used the model of rate autocorrelation implemented in the package MULTIDIVTIME (Yang, 2007; Figure A4.1). Eighteen studies used multiple phylogenies derived from the literature, or designed methods that were suited to their dataset (e.g. molecular rate estimates using local clocks, or no molecular rate estimates at all).

Twelve studies did not mention the number of time-calibrations used, either because the study did not estimate molecular rates (and inferred changes in diversification rates from topology), or because phylogenetic data was taken from several other studies. The

number of time-calibrations used in the remaining 88 studies ranged from 0 (for studies that used relative times) to 98, with a mean and median of 8 and 2 time-calibrations, respectively (Table A4.1).

A4.1.3 Methods to estimate diversification rates or dynamics

Fifty-five studies (55) used more than one method to estimate diversification rates or diversification rate dynamics. Thirty-two studies used the gamma statistic (Pybus & Harvey 2000) to infer changes in diversification rates through time, while 26 studies used the model-testing method with the same purpose from the software package LASER (Rabosky 2006). Twenty-three studies estimated net diversification rates using the methods by Magallon and Sanderson (Magallón & Sanderson 2001). Other methods were used with less frequency, like the topology-based method in the software SymmeTree (Chan & Moore 2005) or the methods MEDUSA (Alfaro *et al.* 2009) and TreePar (Stadler 2011), in 13, 12 and four studies, respectively (Figure A4.2). Nearly half of the studies sampled (46) used methods that were unique or which were used in less than 5 of the sample of studies. Less than half of the studies sampled (40) estimated net diversification rates. Thirteen studies used relative time to estimate diversification rate dynamics.

Fifty-three studies provided estimates of diversification rates. Some of these studies estimated net diversification rates while others provided estimates of diversification rates through time. All but two studies included confidence intervals for the estimates. To facilitate the visualization of the results for the sample of studies, we used the mid point between the maximum and minimum estimate in each study. The diversification rate estimates ranged from 0.018 to 3.8 species per million years, and had mean and median values of 0.46 and 0.14 species per million years, respectively (Table A4.1).

A4.1.4 Overall findings

We assumed that the sample of 100 studies provides a representative picture of the research in macroevolution in the past ten years. Median values provide a better summary than arithmetic means for the information recorded, because the data does not follow a normal distribution. Accordingly, a regular dataset used for phylogenetics and macroevolution has approximately four loci and 3800 base pairs. The sample size of a regular alignment is of approximately 130.

To estimate the chronograms that are used in analyses of macroevolution, two methods have been used with particularly high frequency compared to others. These methods are non-parametric rate smoothing using likelihood and the Bayesian implementation of uncorrelated and log-normally distributed rates. A regular study uses approximately two time-calibrations to estimate phylogenetic timescales.

There is a large diversity of methods used for macroevolutionary inference using phylogenies. This diversity means that any definition of the average study in this field is uncertain. Nevertheless, the most common methods in macroevolution have been the gamma statistic (Pybus & Harvey 2000), model-testing as implemented in the software LASER (Rabosky 2006), and estimates of net diversification rates as described by Magallon and Sanderson (Magallón & Sanderson 2001). The phylogeny in a regular study shows a diversification rate of approximately 0.14 species per million years.

Alfaro ME, Santini F, Brock C *et al.* (2009) Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 13410–13414.

Chan KMA, Moore BR (2005) SYMMETREE: whole-tree analysis of differential diversification rates. *Bioinformatics*, **21**, 1709–1710.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.

Magallón S, Sanderson MJ (2001) Absolute diversification rates in angiosperm clades. *Evolution*, **55**, 1762–1780.

Pybus OG, Harvey PH (2000) Testing macro-evolutionary models using incomplete molecular phylogenies. *Proceedings of the Royal Society B: Biological Sciences*, **267**, 2267–2272.

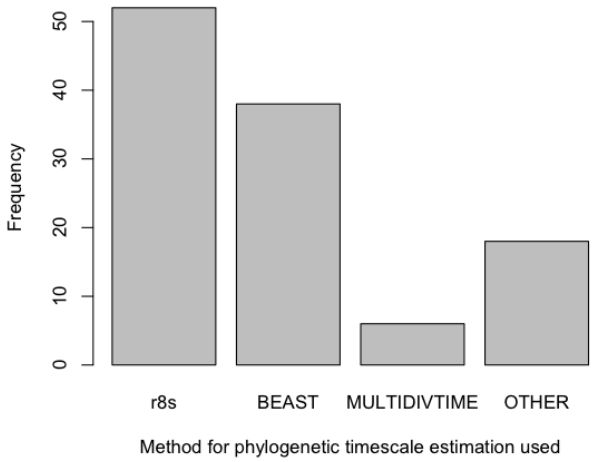
Rabosky DL (2006) LASER: a maximum likelihood toolkit for detecting temporal shifts in diversification rates from molecular phylogenies. *Evolutionary Bioinformatics Online*, **2**, 273–276.

Sanderson MJ (2003) r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics*, **19**, 301–302.

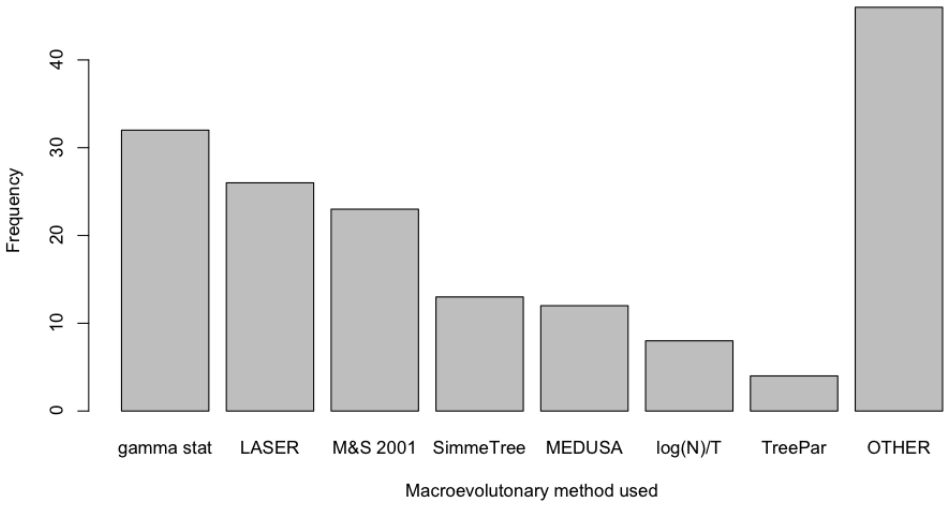
Stadler T (2011) Mammalian phylogeny reveals recent diversification rate shifts. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 6187–6192.

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**, 1586–1591.

Appendix – Figure A4.1. Frequency of usage of methods for phylogenetic timescale estimation in a sample of 100 studies of macroevolution from the last 10 years.



Appendix – Figure A4.2. Frequency of usage of macroevolutionary methods in a sample of 100 studies of from the last 10 years.



Appendix – Table A4.1. Summaries of information about the phylogenetic and macroevolutionary methods used in a sample of 100 studies from the last 10 years.

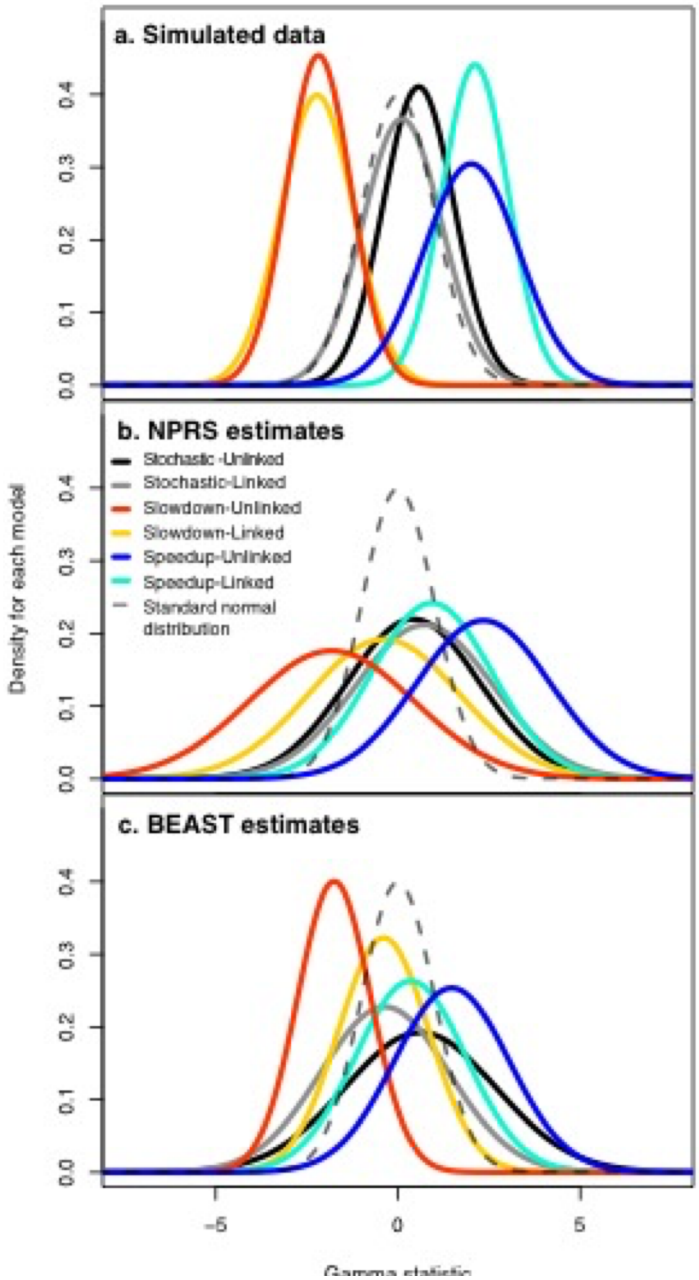
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Number of loci	1	2	4	5.443	5.25	83
Alignment length	511	2412	3773	71100	6431	4500000
Phylogenetic sample size	21	69	133	989.1	227	55470
Number of calibrations	0	1	3	8.261	7.25	98
Diversification rate estimates	0.01855	0.07	0.14	0.4645	0.42	3.8

Appendix – Table A4.2. Proportion of time each that each of the marcoevolutionary methods (gamma, laser, and TESS) were unable to reject the null model (expected for simulations under models A and B; Table 1), or rejected the null in favor of slowdown or speedup in speciation rates (expected for simulations under models C-F).

Macro-evolutionary inference method	Underlying model	Macro-evolutionary model favoured	Source of tree		
			True	NPRS	BEAST
Gamma statistic	A – Stochastic Unlinked	Unable to reject null	0.86 (0.068)	0.62 (0.095)	0.55 (0.098)
		Slowdown	0.01 (0.02)	0.13 (0.066)	0.14 (0.068)
		Speedup	0.13 (0.066)	0.25 (0.085)	0.31 (0.091)
	B – Stochastic Llinked	Unable to reject null	0.87 (0.066)	0.59 (0.096)	0.64 (0.094)
		Slowdown	0.05 (0.043)	0.1 (0.059)	0.23 (0.082)
		Speedup	0.08 (0.053)	0.31 (0.091)	0.12 (0.064)
	C – Slowdown Unlinked	Unable to reject null	0.27 (0.087)	0.41 (0.096)	0.46 (0.098)
		Slowdown	0.73 (0.087)	0.53 (0.098)	0.54 (0.098)
		Speedup	0 (0)	0.06 (0.047)	0 (0)
	D – Slowdown Linked	Unable to reject null	0.28 (0.088)	0.57 (0.097)	0.8 (0.078)
		Slowdown	0.72 (0.088)	0.27 (0.087)	0.14 (0.068)
		Speedup	0 (0)	0.16 (0.072)	0.05 (0.043)
	E – Speedup Unlinked	Unable to reject null	0.39 (0.096)	0.34 (0.093)	0.52 (0.098)
		Slowdown	0 (0)	0.01 (0.02)	0.02 (0.027)
		Speedup	0.61 (0.096)	0.65 (0.093)	0.46 (0.098)
	F – Speedup Linked	Unable to reject null	0.3 (0.09)	0.61 (0.096)	0.71 (0.089)
		Slowdown	0 (0)	0.06 (0.047)	0.09 (0.056)
		Speedup	0.7 (0.09)	0.33 (0.092)	0.19 (0.077)
Laser	A – Stochastic Unlinked	Unable to reject null	0.5 (0.098)	0.67 (0.092)	0.55 (0.098)
		Slowdown	0.35 (0.093)	0 (0)	0.44 (0.097)
		Speedup	0.15 (0.07)	0.3 (0.09)	0.07 (0.05)
	B – Stochastic Llinked	Unable to reject null		0.54 (0.098)	0.5 (0.098)
		Slowdown		0.19 (0.077)	0.34 (0.093)
		Speedup		0.27 (0.087)	0.14 (0.068)
	C – Slowdown Unlinked	Unable to reject null	0.13 (0.066)	0.06 (0.047)	0.3 (0.09)
		Slowdown	0.7 (0.09)	0.6 (0.096)	0.66 (0.093)
		Speedup	0.16 (0.072)	0.33 (0.092)	0.03 (0.033)
	D – Slowdown Linked	Unable to reject null		0.07 (0.05)	0.26 (0.086)
		Slowdown		0.5 (0.098)	0.3 (0.09)
		Speedup		0.43 (0.097)	0.43 (0.097)
	E – Speedup Unlinked	Unable to reject null	0.28 (0.088)	0.41 (0.096)	0.48 (0.098)
		Slowdown	0.07 (0.05)	0.03 (0.033)	0.24 (0.084)

		Speedup	0.64 (0.094)	0.55 (0.098)	0.28 (0.088)
	F – Speedup Linked	Unable to reject null		0.67 (0.092)	0.39 (0.096)
		Slowdown		0.03 (0.033)	0.46 (0.098)
		Speedup		0.29 (0.089)	0.14 (0.068)
TESS	A – Stochastic Unlinked	Unable to reject null	0.51 (0.098)	0.67 (0.092)	0.55 (0.098)
		Slowdown	0.29 (0.089)	0 (0)	0.44 (0.097)
		Speedup	0.18 (0.075)	0.3 (0.09)	0.07 (0.05)
	B – Stochastic Llinked	Unable to reject null	0.5 (0.098)	0.54 (0.098)	0.5 (0.098)
		Slowdown	0.35 (0.093)	0.19 (0.077)	0.34 (0.093)
		Speedup	0.15 (0.07)	0.27 (0.087)	0.14 (0.068)
	C – Slowdown Unlinked	Unable to reject null	0.06 (0.047)	0.06 (0.047)	0.3 (0.09)
		Slowdown	0.83 (0.074)	0.6 (0.096)	0.66 (0.093)
		Speedup	0.1 (0.059)	0.33 (0.092)	0.03 (0.033)
	D – Slowdown Linked	Unable to reject null	0.13 (0.066)	0.07 (0.05)	0.26 (0.086)
		Slowdown	0.7 (0.09)	0.5 (0.098)	0.3 (0.09)
		Speedup	0.16 (0.072)	0.43 (0.097)	0.43 (0.097)
	E – Speedup Unlinked	Unable to reject null	0.31 (0.091)	0.41 (0.096)	0.48 (0.098)
		Slowdown	0 (0)	0.03 (0.033)	0.24 (0.084)
		Speedup	0.69 (0.091)	0.55 (0.098)	0.28 (0.088)
	F – Speedup Linked	Unable to reject null	0.28 (0.088)	0.67 (0.092)	0.39 (0.096)
		Slowdown	0.07 (0.05)	0.03 (0.033)	0.46 (0.098)
		Speedup	0.64 (0.094)	0.29 (0.089)	0.14 (0.068)

Appendix – Figure A4.3. Gamma statistic of phylogenies simulated or reconstructed under each of six models: variable rates, linked rates, speciation slowdown, speciation speedup, linked slowdown, linked speedup (see section 4.2.1).



Appendix – Chapter 5

Appendix – Table A5.1. The mean difference between simulated and estimated node-ages in analyses of isochronous (left) and heterochronous (right) data, expressed as a proportion of node-age for each simulation scheme. Results are shown for analyses with simulations performed under each of the three clock models: strict clock (top), relaxed uncorrelated lognormal clock (middle), and relaxed autocorrelated lognormal clock (bottom). Darker shades of grey indicate wider 95% credible intervals, and colour gradients are relative to each data set type.

		Isochronous datasets			Heterochronous datasets		
Calibrations \ Balance		High	Medium	Low	High	Medium	Low
Strict clock	Deep	<0.01	0.04	<0.01	0.02	-0.05	-0.03
	Median	<0.01	<0.01	<0.01	-0.03	-0.08	-0.1
	Shallow	<0.01	-0.03	0.02	-0.04	-0.14	-0.21
Uncorrelated lognormal clock	Deep	0.02	-0.06	0.02	-0.03	-0.04	-0.07
	Median	0.09	-0.02	-0.08	-0.06	-0.09	-0.12
	Shallow	-0.01	-0.05	0.03	-0.09	-0.14	-0.3
Autocorrelated lognormal clock	Deep	-0.03	-0.02	-0.02	-0.02	-0.03	-0.07
	Median	0.02	-0.08	-0.08	-0.02	-0.11	-0.13
	Shallow	0.03	0.07	0.05	-0.07	-0.13	-0.24

Appendix – Supplementary information A5.1 – Data accessibility

GenBank DNA sequence accession numbers for the African swine fever virus dataset: FJ174401, KC990883, FJ174385, GQ477151, HM745325, HM745346, JX857503, HM745344, HM745350, FJ174397, FJ174412, HM745343, KC990875, GQ410767, JX857507, FJ174422, HM745354, HM745348, JX857497, KC610537, KF303302, HM745342, JQ771681, JQ771685, KC112569, HM745328, HM745336, FJ174415, KC990889, JX857505, JN582354, HM745347, JX857501, FJ174435, FJ174400, JX857506, HM745333, JN582355, KC990886, GQ477146, FJ174433, HM745339, FJ174417, KC112570, FJ174431, FJ174438, JQ771682, FJ174421, KC610535, FJ174405.

Genbank DNA sequence accession numbers for the primate dataset: AB371093, KC757396, AB371092, KC757407, AB371095, KC757405, AJ309867, KC757402, AB371094, KC757408, KC757397, AJ421451, KC757395, AB371087, AB371086, AM905040, AB371088, AB371089, KC757387, AB286049, KC757400, HM070254, KC757392, AB371085, AM905039, AF348159, AB371090, KC757409, KC757389, KC757388, KC757399, KC757385, FJ785421, KC757410, AJ309866, KC959987, FJ785425, KC757398, KC757386, FJ785422, KC757384, KC757393, KC959985, KC959986, FJ785423, EF597502, EF597500, KC757391, KC757394, FJ906803, AY612638, EU294187, AJ309865, KC757403, KC757390, KC757412, FJ85426, KC757406, Y18001, KC757401, JF293096, JF293094, JF293095, DQ355300, JF293093, DQ355299, AY863425, DQ355297, AY863427, DQ355301, JF293092, KC757411, X99256, AB504749, AB504748, KC757404, X97707, D38115, X93347, D38116, D38113, AM948965, X93334, AF217811, NC_005089, NC_001665, NC_000891.

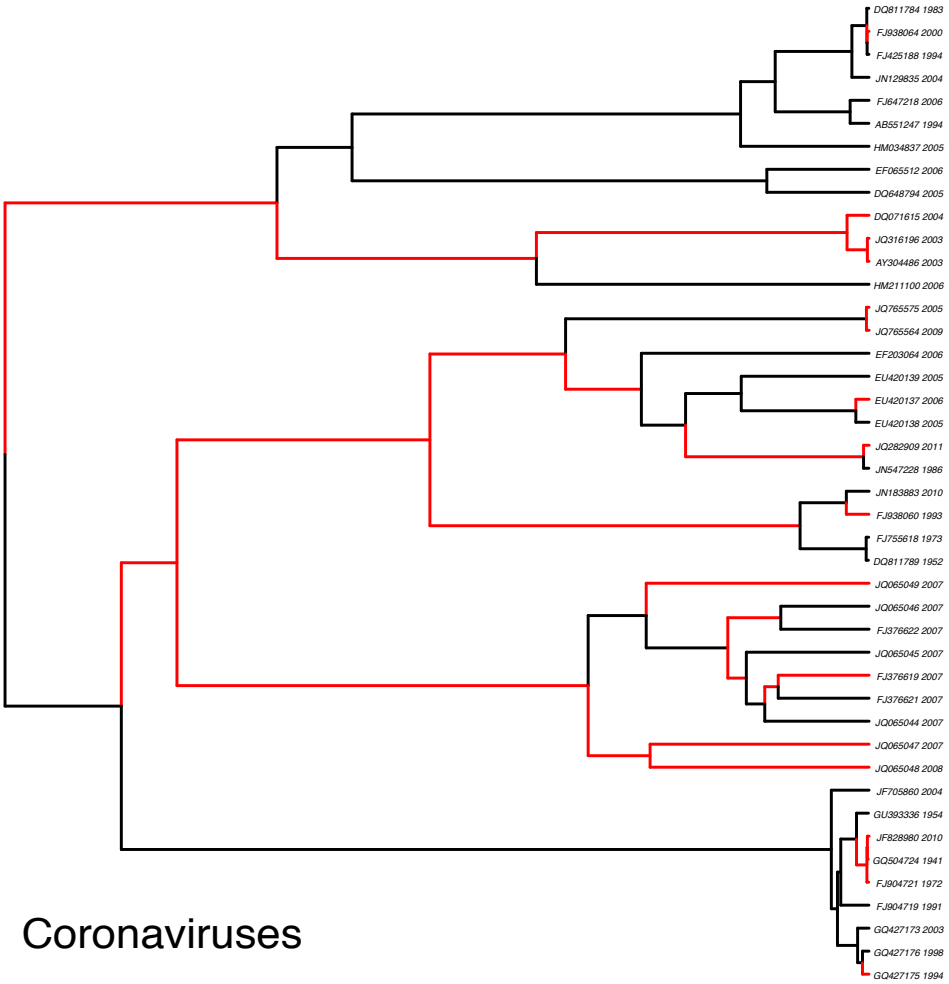
Empirical data alignments: DRYAD entry doi:10.5061/dryad.q4545

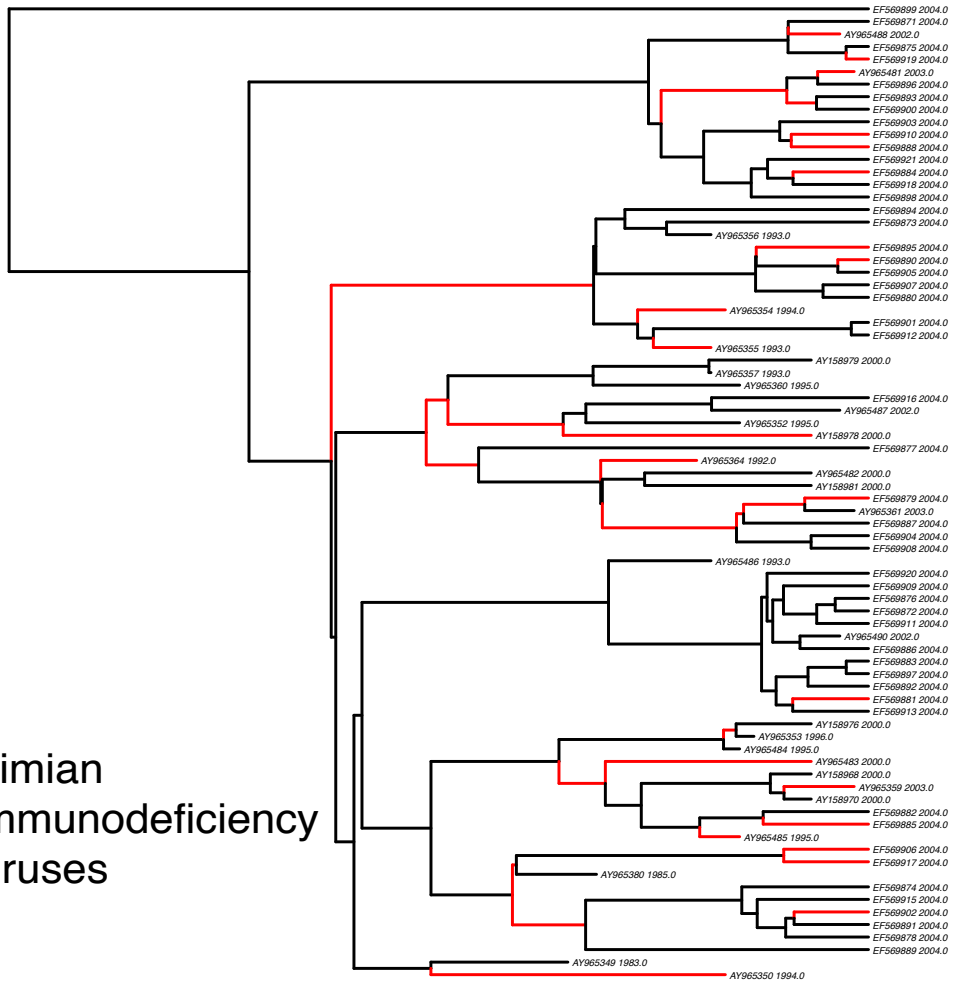
Code for simulations and example input/output files: GitHub repository:

https://github.com/duchene/imbaltime1

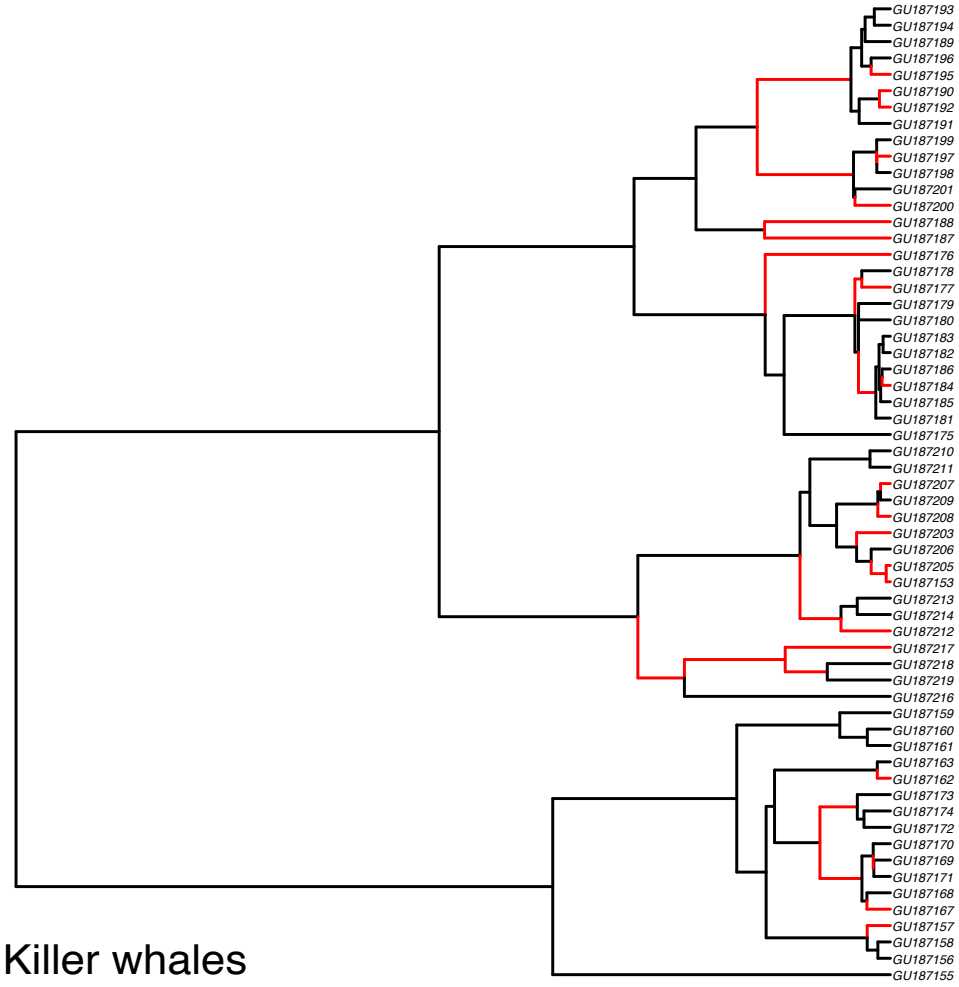
Appendix – Chapter 6

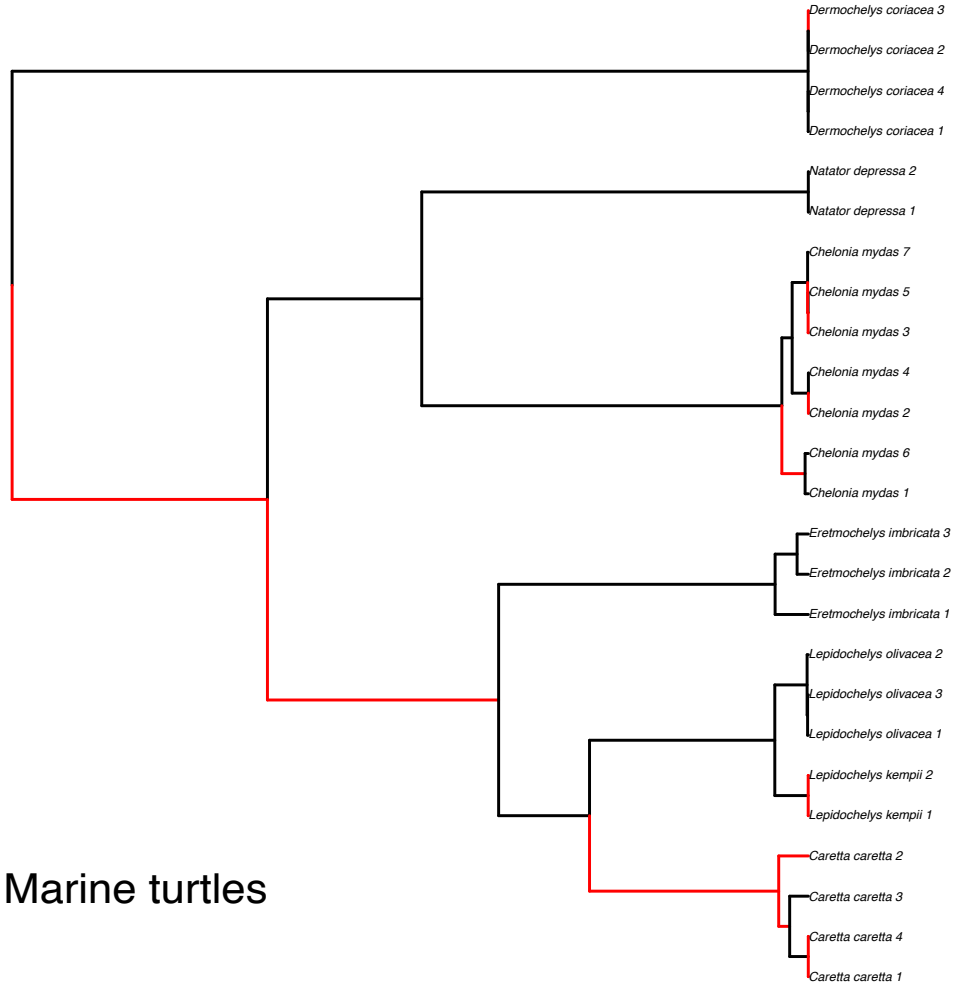
Appendix – Figure A6.1. Phylogenetic tree estimates for the four data sets used in the study, estimated using the best-fitting clock model. Branches in red were found to be inadequate under the clock model selected.





Simian immunodeficiency viruses


GU187193
GU187194
GU187189
GU187196
GU187195
GU187190
GU187192
GU187191
GU187199
GU187197
GU187198
GU187201
GU187200
GU187188
GU187187
GU187176
GU187178
GU187177
GU187179
GU187180
GU187183
GU187182
GU187186
GU187184
GU187185
GU187181
GU187175
GU187210
GU187211
GU187207
GU187209
GU187208
GU187203
GU187206
GU187205
GU187153
GU187213
GU187214
GU187212
GU187217
GU187218
GU187219
GU187216
GU187159
GU187160
GU187161
GU187163
GU187162
GU187173
GU187174
GU187172
GU187170
GU187169
GU187171
GU187168
GU187167
GU187157
GU187158
GU187156
GU187155
Killer whales


Dermochelys coriacea 3
Dermochelys coriacea 2
Dermochelys coriacea 4
Dermochelys coriacea 1
Natator depressa 2
Natator depressa 1
Chelonia mydas 7
Chelonia mydas 5
Chelonia mydas 3
Chelonia mydas 4
Chelonia mydas 2
Chelonia mydas 6
Chelonia mydas 1
Eretmochelys imbricata 3
Eretmochelys imbricata 2
Eretmochelys imbricata 1
Lepidochelys olivacea 2
Lepidochelys olivacea 3
Lepidochelys olivacea 1
Lepidochelys kempii 2
Lepidochelys kempii 1
Caretta caretta 2
Caretta caretta 3
Caretta caretta 4
Caretta caretta 1
Marine turtles

Appendix – Figure A6.2. Example of the simulated chronogram (top) and the median node ages from the prior distribution (bottom), after placing two calibrations with incorrect ages (A and B). Calibration A was placed in a node adjacent to a tip, and given an age of 0.9 times the age of the root. Calibration B was placed on a node directly descending from the root on the opposite side of the root from calibration A, and was given an age of 0.1 of the true node age. The root was calibrated in every case (calibration C).