

WHAT DRIVES BIOLOGICAL DIVERSIFICATION?  
DETECTING TRAITS UNDER  
SPECIES SELECTION

*by*

RICHARD GARETH FITZJOHN

B.Sc., Victoria University of Wellington, 2000  
M.Sc. (Hons), Victoria University of Wellington, 2003

A thesis submitted in partial fulfillment of  
the requirements for the degree of

DOCTOR OF PHILOSOPHY

*in*

The Faculty of Graduate Studies  
(Zoology)

The University of British Columbia  
(Vancouver)

AUGUST 2012

© Richard Gareth FitzJohn, 2012

## ABSTRACT

---

Species selection — heritable trait-dependent differences in rates of speciation or extinction — may be responsible for variation in both taxonomic and trait diversity among clades. While initially controversial, interest in species selection has been revived by the accumulation of evidence of widespread trait-dependent diversification. In my thesis, I developed and applied a number of new likelihood-based methods for investigating species selection by detecting the association between species traits and speciation or extinction rates. These methods are explicitly phylogenetic and incorporate simple, but commonly used, models of speciation, extinction, and trait evolution; I assume throughout that speciation and extinction can be modelled as a birth-death process where rates depend in some way on one or more traits, and that these traits evolve under a Markov process. In particular, I extended the BiSSE (Binary State Speciation and Extinction) method to allow use with incompletely resolved phylogenies, and developed analogous methods for multi-state discrete traits or combinations of binary traits (MuSSE; Multi-State Speciation and Extinction) and quantitative traits (QuaSSE; Quantitative State Speciation and Extinction). I tested the statistical performance of the methods using simulations, investigating their performance with variation in tree size, degree of resolution, number of traits, and departure from the true model. I used each method to consider a different biological question; I found that sexual dimorphism was shortlived but associated with elevated rates of speciation in shorebirds; that solitariness and monogamy are associated with decreased speciation rates in primates (showing that a previous analysis was robust to treating both traits simultaneously); and that body size was a poor predictor of speciation rates in primates. In chapter 5, I extended this analysis of body size to all mammals, and investigated if within-lineage increases in body size (Cope's rule) were balanced by species selection against large bodied species. I

found little support for this hypothesis, with clade-specific differences in the direction of species selection and idiosyncratic variation in speciation rates. Together, the methods I have developed allow testing of long-standing hypotheses about causes of variation in biological diversity.

## PREFACE

---

Several chapters from my thesis have been published elsewhere:

CHAPTER 2 has been previously published as:

FitzJohn R.G., Maddison W.P., and Otto S.P. 2009. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Systematic Biology* 58:595–611.

I derived most of the mathematical results, developed the software, carried out the analysis and wrote the manuscript. Wayne P. Maddison and Sarah P. Otto had the original idea for the “skeletal tree” and “unresolved clades” methods, respectively, and both also helped with writing, editing, and guidance.

CHAPTER 3 has been previously published as:

FitzJohn R.G. 2010. Quantitative traits and diversification. *Systematic Biology* 59:619–633.

Sarah P. Otto provided guidance and helped edit the manuscript.

CHAPTER 4 has been accepted for publication at *Methods in Ecology and Evolution*, pending minor revisions. I am the sole author of this manuscript. Sarah P. Otto provided guidance, advice, and helped edit the manuscript. The R package this paper is based on is my own work, but contains methods by Emma E. Goldberg (University of Illinois at Chicago), Karen Magnuson-Ford, and Sarah P. Otto.

CHAPTER 5 was coauthored with Nick D. Pyneson (Smithsonian Institution) and Sarah P. Otto. I developed the idea, ran the analyses and wrote the first version of the paper. Nick D. Pyneson provided data and advice about Cetacea. Sarah P. Otto provided guidance, advice, and editing.

## TABLE OF CONTENTS

---

ABSTRACT . . . . .	ii
PREFACE . . . . .	iv
TABLE OF CONTENTS . . . . .	v
LIST OF TABLES . . . . .	vii
LIST OF FIGURES . . . . .	viii
ACKNOWLEDGEMENTS . . . . .	x
DEDICATION . . . . .	xiii
1 INTRODUCTION . . . . .	1
1.1 Detecting species selection . . . . .	5
1.2 Speciation, extinction, & phylogenies . . . . .	6
1.3 Structure & contents of this thesis . . . . .	8
2 ESTIMATING TRAIT-DEPENDENT SPECIATION & EXTINCTION RATES FROM INCOMPLETELY RESOLVED PHYLOGENIES . . . . .	10
2.1 Summary . . . . .	10
2.2 Introduction . . . . .	11
2.3 BiSSE for complete phylogenies . . . . .	14
2.4 Likelihood calculations for incompletely resolved phylogenies . . . . .	16
2.5 Bayesian inference . . . . .	24
2.6 Numerical methods & application . . . . .	25
2.7 Results . . . . .	27
2.8 Application to shorebird data . . . . .	34
2.9 Discussion . . . . .	38
3 QUANTITATIVE TRAITS & DIVERSIFICATION . . . . .	41
3.1 Summary . . . . .	41
3.2 Introduction . . . . .	42
3.3 Character evolution & diversification . . . . .	44
3.4 Likelihood calculations . . . . .	46
3.5 Simulation results . . . . .	58
3.6 Application to primate body size data . . . . .	61

3.7	Discussion . . . . .	67
4	DIVERSITREE: COMPARATIVE PHYLOGENETIC ANALYSES OF DIVERSIFICATION IN R. . . . .	72
4.1	Summary . . . . .	72
4.2	Introduction . . . . .	72
4.3	The methods . . . . .	73
4.4	The approach . . . . .	75
4.5	The MuSSE model . . . . .	78
4.6	Simulation test assessing the power of MuSSE . . . . .	82
4.7	Closing comments . . . . .	90
5	SURVIVAL OF THE LITTLEST? SPECIES SELECTION, COPE'S RULE, & MAMMAL BODY SIZE. . . . .	93
5.1	Introduction . . . . .	93
5.2	Results & discussion . . . . .	95
5.3	Methods . . . . .	103
6	CONCLUSION . . . . .	111
6.1	Some issues with these methods . . . . .	113
6.2	Future directions . . . . .	121
6.3	Conclusion . . . . .	124
	BIBLIOGRAPHY . . . . .	125
	APPENDICES . . . . .	142
A	SUPPLEMENTARY INFORMATION TO CHAPTER 2. . . . .	142
A.1	Root-state calculations . . . . .	142
A.2	Character-independent model . . . . .	144
B	SUPPLEMENTARY INFORMATION TO CHAPTER 3. . . . .	147
B.1	Single character derivation . . . . .	147
B.2	Multivariate character derivation . . . . .	149
C	SUPPLEMENTARY INFORMATION TO CHAPTER 4. . . . .	155
C.1	Tuning diversitree . . . . .	155
C.2	A faster algorithm for BM & OU likelihood calculations . . . . .	162
C.3	MuSSE & multitrait diversification in primates . . . . .	169
D	SUPPLEMENTARY INFORMATION TO CHAPTER 5. . . . .	179
D.1	Partition compositions . . . . .	179

## LIST OF TABLES

---

TABLE 3.1	Summary of model fits for the association between body size and diversification for primates. . . . .	66
TABLE 4.1	Summary of model types available in diversitree. . . . .	76
TABLE 5.1	Properties of the 10 clades used . . . . .	96
TABLE 5.2	Properties of partitions recovered by MEDUSA . . . . .	108

## LIST OF FIGURES

---

FIGURE 1.1	Lineage-through-time plots for simulated trees . . . . .	7
FIGURE 2.1	Ways that phylogenetic information may be incomplete . . . . .	13
FIGURE 2.2	Schematic of the BiSSE method with and without full phylogenetic knowledge . . . . .	15
FIGURE 2.3	Posterior probability densities for BiSSE parameters . . . . .	28
FIGURE 2.4	Posterior probability densities for diversification rates . . . . .	30
FIGURE 2.5	Uncertainty around BiSSE parameter estimates as a function of phylogenetic knowledge . . . . .	32
FIGURE 2.6	Uncertainty around diversification rate estimates as a function of phylogenetic knowledge . . . . .	33
FIGURE 2.7	Phylogenetic tree of the 350 species of shorebirds (Charadriiformes) .	35
FIGURE 2.8	Association between sexual and diversification and character transition rates in shorebirds . . . . .	37
FIGURE 3.1	Possible ways a lineage extant at time $t + \Delta t$ might go extinct. . . .	48
FIGURE 3.2	Possible ways a lineage extant at time $t + \Delta t$ might lead to exactly the clade $N$ as observed. . . . .	50
FIGURE 3.3	Power to detect differential speciation and extinction with QuaSSE on simulated phylogenies. . . . .	59
FIGURE 3.4	Representative speciation and extinction function fits. . . . .	60
FIGURE 3.5	Power to detect trait-dependent speciation and directional character change . . . . .	62
FIGURE 3.6	Phylogenetic tree of the primates. . . . .	64
FIGURE 3.7	Primate speciation and extinction model fits . . . . .	68
FIGURE 4.1	Uncertainty around multitrait-MuSSE parameter estimates as a function of tree size and number of traits . . . . .	84
FIGURE 4.2	Power and error rates of multitrait MuSSE . . . . .	85
FIGURE 4.3	“Main effects” of monogamy and solitariness on speciation rate in primates . . . . .	91
FIGURE 5.1	Distribution of mammal body masses . . . . .	94
FIGURE 5.2	Mammal supertree showing the 10 focal clades . . . . .	97
FIGURE 5.3	Inferred body size/speciation relationships in 10 mammal clades . .	99



FIGURE 5.4	Slopes of speciation rate/body mass relationships within MEDUSA-derived partitions. . . . .	.101
FIGURE 5.5	Relationship between speciation rate and body size within mammals, across partitions with different diversification rates. . . . .	.102
FIGURE 5.6	Fits for the multi-slope analyses . . . . .	.109
FIGURE 5.7	Regression of body mass against length for cetacea. . . . .	.110
FIGURE C.1	Performance of Brownian motion likelihood algorithms . . . . .	.168
FIGURE C.2	Primate phylogeny, showing distribution of monogamy and solitariness among species . . . . .	.171

## ACKNOWLEDGEMENTS

---

First, a big thank you to my supervisor, Sally Otto. Sally has been a fantastic mentor, and provided me with enough freedom to let me focus on whatever I found interesting, while simultaneously providing enough guidance so I never felt too lost. Her ability to encourage, enlighten, and (at times) coax have been critical the completion of the work here.

I also thank my committee; Wayne Maddison, Dolph Schluter, and Jeannette Whitton for their helpful advice at a number of points in this thesis and for always being willing to stop and chat with no notice. Other faculty at UBC and SFU were always generous with their time and ideas; Arne Mooers and Michael Whitlock in particular. The SOWD/Delta-Tea group provided a great venue for airing and refining half-baked ideas.

The Otto lab has been a fantastic place to work over the last five years, thanks to Dilara Ally, Aleeza Gerstein, Philip Greenspoon, Kay Hodgins, Crispin Jordan, Liz Kleynhans, Nathan Kraft, Leithen M’Gonigle, Karen Magnusson-Ford, Itay Mayrose, Jasmine Ono, Kate Ostevik, Alirio Rosales, Michael Scott, and Thor Veen. Thanks also to the regular extras at lab meetings (especially over the last year): Florence Débarre, Kim Gilbert, Fred Guillaume, Katie Lotterhos, and Alana Schick. The diversity of ideas in the lab has helped me keep my interests broad.

This thesis is primarily theoretical, and I have benefited from data from other researchers. Gavin Thomas and Terje Lislevand, Jordi Figuerola, and Tamás Székely have made the data used in chapter 2 freely available. Arne Mooers, David Redding, and Rutger Vos generously shared data for the primate analysis in chapters 3 and 4, and Tyler Kuhn generated the trees with simulated branch lengths in chapters 3–5. Kate Jones

(and coauthors) provided the trait data used in chapter 5. More generally, I thank those involved in Open Data efforts, including Heather Piowar and Tim Vines at UBC.

In addition to those listed above, a number of people provided feedback on various chapters directly: David Bapst, Folmer Bokma, Will Cornwell, Luke Harmon, Dan Rabosky, and Graham Slater. Several people also contributed specific ideas; Rick Ree, initially suggested expanding BiSSE to account for trees containing exemplars, which inspired most of chapter 2. Wayne Maddison suggested the connection to species selection, which helped me clarify my thinking on my research. Graham Slater suggested I investigate the power of MuSSE, which turned out to be more interesting than I thought it would be. Matt Pennell suggested creating figure C.1, which made me feel a lot better about how painfully slow most of the models in diversitree are. I have also benefited tremendously from discussions with friends, collaborators, and colleagues from outside of UBC: Jeremy Beaulieu, Juan-López Cantalapiedra, Emma Goldberg, Gene Hunt, Boris Igić, Marc Johnson, Marcos Alexandrou, Lynsey McInnes, Brian O’Meara, Nick Pyensen, Dan Rabosky, Stacey Smith, and Amy Zanne.

I thank everyone who has helped by testing the perpetual beta that is “diversitree”. Special thanks to Emma Goldberg, Karen Magnusson-Ford, and Sally Otto for contributing code, to Wayne Maddison for developing the interface with Mesquite, and to Emmanuel Paradis for developing the “ape” package without which diversitree would never have been written.

This work was financially supported by a University Graduate Fellowship from the University of British Columbia, a Vanier Commonwealth Graduate Scholarship from NSERC, and a Capability Fund Grant from Manaaki Whenua/Landcare Research. Vast amounts of computing time was provided by the Zoology Computing Unit and Westgrid. Particular thanks to Alistair Blachford, Andy LeBlanc, and Richard Sullivan for running the Zoology Computing Unit, without which most of the research here would not have been possible.

UBC has been a fun and ridiculously interactive place to be, thanks in part to the people above, and Alistair Blachford, Dave Allen, Rose Andrews, Rowan Barrett, Ella Bowles, Alan Brelsford, Carla Crossman, Josh Chang Mell, Dan Ebert, Edd Hammill, Matt Herron, Will Iles, Andy LeBlanc, Robin LeCraw, Julie Lee-Yaw, Andrew MacDonald, Janet Maclean, Blake Matthews, Jon Mee, JS Moore, Jana Petermann, Jess Purcell, Seth Rudman, Kieran Samuk, Alana Schick, Laura Southcott, Travis Ingram, Kathryn Turner, Jabus Tyerman, and Sam Yeaman. Everyone I've played boardgames with, drank coffee or beer with, eaten a burger at the Fringe with, climbed with, hiked with. Copying Jon Mee, I thank the beer barons, reading group and seminar organisers, and retreat planners who make UBC Zoology/Biodiversity the place it is.

I thank my family for always being supportive of my research. Also, for not giving me too much of a hard time to leave a perfectly good job and go back to school.

Finally, thank you Andréa, for help in all sorts of ways.

*To my grandparents*  
Anthony & Pamela FitzJohn  
Dewi & Eruwen Davies

## CHAPTER 1

### INTRODUCTION

---

“At yet higher levels, the species and the community, natural selection obviously must occur. Species evolve to survive in a certain environmental range, and if the environment should suddenly change, some species will become extinct but others will survive.” Lewontin (1970), p. 15

The tree of life is remarkably uneven. The Passeriformes consist of over 5,000 species, but the deepest division in this clade is between two species of New Zealand wrens (the family Acanthisittidae) and the rest of the order. Similarly, a single species (*Amborella trichopoda*) is probably the sister to the remaining 250,000 species of angiosperms (Moore *et al.*, 2007). Such asymmetries in diversity are highly unlikely to have occurred by chance alone. One explanation is that species traits may cause differences in rates of speciation and extinction. By analogy to natural selection — trait-based differential survival and reproduction of individuals within population — trait-dependent speciation and extinction can be thought of as “species selection”. This idea has a long history, dating back as far as Lyell (1832), who was the first to argue that differential extinction among species may shape patterns of diversity (Van Valen, 1975). The concept is implicit in E.O. Wilson’s “taxon cycle” (Wilson, 1959, 1961) and was discussed (but dismissed) by Fisher (1958, p. 50) and Williams (1966, p. 115). Lewontin (1970) clarified the idea considerably by drawing parallels to other levels of selection, and the term “species selection” itself was coined by Stanley (1975b). While there has been little disagreement that species traits may affect their rates of speciation or extinction, species selection became controversial immediately after being proposed.

First, there has been disagreement about the nature of traits required to call trait-dependent speciation and extinction “species selection” (e.g., table 1 in Lieberman and Vrba, 2005). A “narrow-sense” view of species selection argues that the trait leading to differential speciation or extinction be apparent only as an emergent property of species (e.g., Vrba and Gould, 1986). Examples of such traits include geographic range, population structure, and variability of traits. This view is motivated by the idea that for selection to operate at the species level, traits must be apparent at this same level. Differential speciation and extinction due to traits apparent below the species level is referred to as “species sorting”. The alternative, “broad-sense”, view of species selection argues that any trait, whether it is the property of individuals or of the species as a whole, can be involved in species selection (e.g., Lloyd and Gould, 1993). What matters is that *fitness* is emergent at the level of species; species give birth (speciation) and die (extinction), and species selection operates on any trait that affects emergent fitness. Increasingly, the distinction between emergent and aggregate traits has been considered not sufficiently informative to warrant the recognition of two different processes (arguments in Damuth and Heisler, 1988; Grantham, 1995; Okasha, 2006; Jablonski, 2008b). As such, “emergent fitness” is sufficient for species selection to operate, regardless of the level at which the trait is apparent. That said, there is an interesting consequence of the emergent trait criterion; as individual selection cannot act on emergent traits, if species selection affects the distribution of traits among species then macroevolution may be fundamentally decoupled from microevolution (Erwin, 2010). If species selection on emergent traits is widespread then the distribution of any trait that “hitch-hikes” with these will be unpredictable from microevolutionary studies.

Second, species selection has been argued to be too slow, relative to individual-level selection, to be a significant force in shaping the distribution of traits (Fisher, 1958, p. 50; Williams, 1966, p. 115; Maynard Smith, 1983; Rice, 1995). Strong directional selection at the individual level should always be able to trump species selection due to the large differences in their relative speeds. That is, the number of selective deaths of individuals

(and the resulting opportunity for selection to act) will outweigh the number of selective deaths of species. However, this view eventually softened. Strong sustained directional selection may not be common, with changes in both strength and sign (e.g., Grant and Grant, 2002; Siepielski *et al.*, 2009), and stabilising selection may be pervasive (e.g., Estes and Arnold, 2009), leaving room for species selection to operate. Brown (1995) argued that if the rate of environmental change can outpace rates of individual adaptation, then species selection may play an important role during mass extinction; this frames the problem as a purely empirical question of how often such conditions are met. Maynard Smith (1989) argued that species selection has had a major effect on the distributions of different types of organisms, even if a trivial effect on the production of adaptations within organisms. Jablonski (2000) put it nicely when he said “species sorting may not construct a complex eye or a long neck, but it may determine how many species possess complex eyes or long necks over evolutionary timescales” (p. 38).

Finally, species selection has also been controversial due to its association with other contentious ideas; particularly group selection and punctuated equilibrium. Early arguments both for and against species selection often invoked traits that decreased the probability of extinction (e.g., Leigh, 1977). However, the extinction of a species requires the death of all individuals of that species (in contrast with speciation, which is entirely decoupled from the birth of individuals). Therefore, any trait that increases the probability of species extinction is perfectly correlated with decreased individual survival (at least in small populations), and individual selection would tend to oppose that trait, without having to invoke species selection. Similarly, the association with punctuated equilibrium and its perceived challenge to neo-Darwinian orthodoxy (Eldredge and Gould, 1972; Gould and Eldredge, 1977) did not initially endear the theory to many biologists (e.g., Lande, 1980; Charlesworth, 1981; Charlesworth *et al.*, 1982).

As a result of this opposition, species selection has not been a popular concept among evolutionary biologists. While debate about the nature of species selection remained active in philosophy journals, I was unable to find any primary research explicitly framed



as species selection between 1987 and 1999. In a rare primary paper that even mentioned species selection during this period, Herrera (1992) avoids the term not because it is an unlikely process, but because it is controversial (p. 423). This view persists; an anonymous reviewer of chapter 3 commented:

“Why start a paper that has very broad implications out on the narrow and controversial topic of species selection? [or at least cite Coyne and Orr’s book for a counterargument] Whatever evidence there is for it is limited, yet there is little doubt that traits can affect speciation rates and there are many implications of that connection irrespective of whether selection on clades has molded biodiversity in some major way.”

A survey of 15 biologists by Grimshaw (2001, p. 178) found that most regarded species selection as an unlikely phenomenon, too weak to be generally important for biology, although their objections were vague. Species selection seems to have become *theoria non grata* among most biologists, though often without a firm idea why.

Recently, species selection appears to have gained favour amongst evolutionary biologists. A series of influential reviews have recently been published, each making the case that trait-dependent differences in speciation and extinction rates can and should be considered species selection (Coyne and Orr, 2004<sup>1</sup>; Jablonski, 2008b; Rabosky and McCune, 2009). At least 27 primary papers framing their work as species selection have been published since 1999 (16 since 2010). The degree of embrace varies from mentioning species selection as one possible interpretation (e.g., Pillon *et al.*, 2010; Sallan *et al.*, 2011) to declaring it in the title of the paper (e.g., Duda and Palumbi, 1999; McGlone *et al.*, 2001; Goldberg *et al.*, 2010; Simpson, 2010; Eastman and Storfer, 2011). Despite claims of controversy, I found no articles by biologists arguing against species selection over the last decade, suggesting that even if it is not fully accepted, opposition has waned considerably since 20 years ago. My personal view, and the view taken in this thesis, is to

---

<sup>1</sup>As the quote above cites Coyne and Orr (2004) as opposing species selection. However, Coyne and Orr (2004) write “Thus, if we define species selection on a trait as *repeatable effects of that trait on the rate of diversification* of species possessing it, we avoid unproductive arguments about whether or not selection acts on emergent properties.” (p. 444, their emphasis) and “we regard most of the examples given in Table 12.2 as true cases of species selection” (p. 445).

define species selection as any trait-dependent differences in speciation and extinction where the trait is “heritable” (i.e., does not drastically change at speciation).

## 1.1 DETECTING SPECIES SELECTION

If species selection is a potentially important biological phenomenon, how do we detect it? More specifically, how do we detect whether a trait is associated with increased or decreased rates of speciation or extinction? The most direct route would be to know the state of a species immediately before speciation or extinction. However, this requires a more complete fossil record, phylogeny, and knowledge of ancestral states than we are ever likely to have. Instead, we must infer the effect of traits on speciation by looking for signatures in the patterns of diversity and trait distributions.

The first attempt to measure diversity-trait associations statistically (rather than descriptively) using extant species was Mitter *et al.* (1988), who introduced “sister-clade comparisons”. The idea is simple; by definition, sister clades have diversified over the same length of time. If each clade is characterised by a different state of a trait thought to affect diversification rates, then differences in diversity might be ascribed to that trait. If enough such diversity–trait comparisons show the same relationship, then there is statistical evidence for the nonrandom association of a state and increased rates of diversification. Such sister-clade comparisons have been used to detect trait-dependent differences in diversification in a wide range of species (e.g., table 12.2 in Coyne and Orr, 2004) and have been the dominant approach since their introduction.

However, sister-clade comparisons discard most of the potential information available from phylogenies. Clades consisting of hundreds of species are collapsed into a binary categorisation: are they larger or smaller than their sister clade? Any information about the relative timings of speciation events are discarded (e.g., Rosenzweig, 1996). If the trait varies among species within a clade, that clade cannot easily be used. Furthermore, the effects of differential speciation and differential extinction cannot be disentangled.

gled. One example that illustrates these issues is the long-standing hypothesis of species selection against asexual reproduction (Stanley, 1975a). It is unclear if asexual species tend to speciate less or go extinct more than sexual species (Schwander and Crespi, 2009), and sister-clade approaches cannot distinguish between these possibilities. Furthermore, as asexual species are often “tippy” (i.e., recently diverged species-poor groups scattered within a broader clade) most asexual species are sister to a single sexual species, which is uninformative in a sister-clade analyses. In such cases, methods that make better use of available phylogenetic information are needed to address questions about traits and diversification.

## 1.2 SPECIATION, EXTINCTION, & PHYLOGENIES

Beyond sister-species comparisons, the temporal position of nodes in a time-calibrated phylogeny can provide information about the timing of speciation; this information has been widely used to infer rates of speciation. The simplest model is the pure birth, or “Yule”, process (Yule, 1925), in which new species arise through speciation at rate  $\lambda$  and there is no extinction. This leads to exponential growth in species diversity at rate  $\lambda$ , which can be visualised by plotting the number of lineages over time on a semi-log plot, where the log-number of lineages is expected to increase linearly over time (figure 1.1). If a tree with  $N$  species has a total branch length of  $s$  since the root node (i.e., the length of all branches in the tree), the maximum likelihood speciation rate estimate is  $(N - 2)/s$  (Nee, 2001).

The pure birth process is a special case of the “birth-death” process, in which lineages give rise to new species at rate  $\lambda$  and go extinct at rate  $\mu$ . With extinction, we must make a distinction between lineages that survive to the present (and are therefore present in a phylogeny of all extant species: the “reconstructed phylogeny” of Nee *et al.*, 1994b) and those that go extinct. Under this process, for most of the history of a clade, lineages in the reconstructed phylogeny accumulate at the diversification rate,  $\lambda - \mu$ , with an

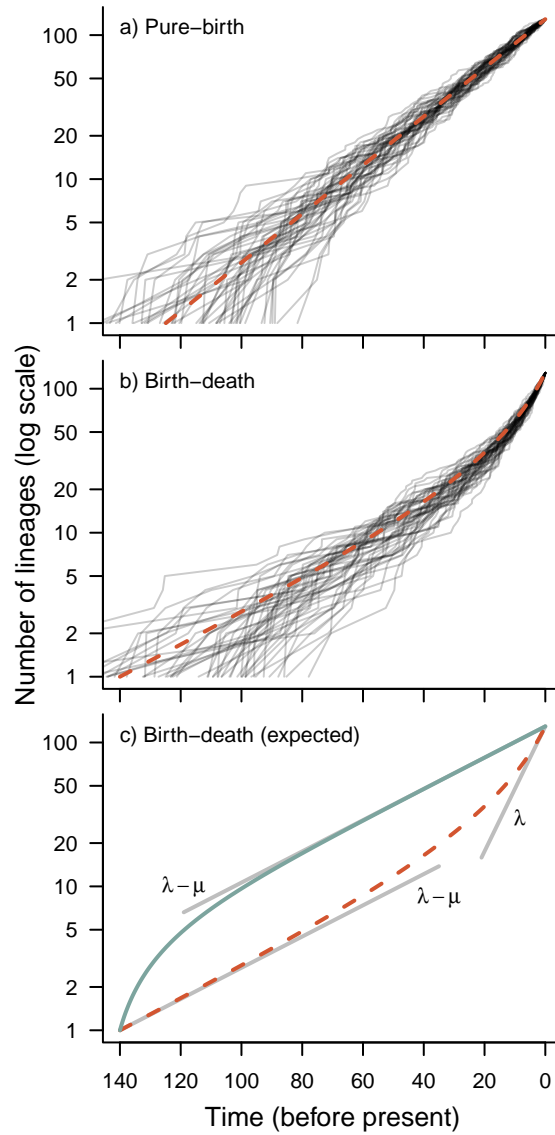


FIGURE 1.1: Lineage-through-time plots for pure-birth (a) and birth-death (b) processes, which count the lineages surviving to the present over time. Fifty simulated phylogenies are represented by the thin grey lines. The dashed red line is the expected number of lineages that will survive to the present. In the pure-birth process, the slope of the expected log number of surviving lineages over time is  $\lambda$  over the entire time course. In the birth-death process, this slope is  $\lambda - \mu$  initially, increasing towards the present to reach a slope of  $\lambda$  (grey lines panel c). The number of lineages at a given time (including those that subsequently go extinct) is larger than this (teal upper line in c). Figure adapted from Harvey *et al.* (1994).

upturn in the log-number of lineages over time close to the present (figure 1.1). Loosely, this occurs because recently diverged species have not yet had time to go extinct. This temporal difference in slope suggests that it may be possible to estimate both speciation and extinction rates from extant taxa only (Harvey *et al.*, 1994). Likelihood calculations under this model were derived by Nee *et al.* (1994a,b).

Even if we can estimate speciation and extinction rates from phylogenies, linking these rates to differences in traits is nontrivial. If we knew the character state at every point along the phylogeny we could construct lineage-through-time plots for each state and compare these. While these states are never directly known, some methods have attempted to combine reconstructed ancestral states with speciation rate estimation (Paradis, 2005; Ree, 2005; Freckleton *et al.*, 2008). However, these methods ignore the trait-dependence of speciation or extinction when estimating the rate at which the state changes, introducing biases in both the transition rates (Maddison, 2006), and the estimated ancestral states (Goldberg and Igić, 2008). Recently, the “BiSSE” (Binary State Speciation and Extinction) method of Maddison *et al.* (2007) was developed to account simultaneously for both the evolution of a trait and that trait’s effect on speciation and extinction. This is a likelihood method that computes the probability of a phylogenetic tree and the distribution of species traits without directly reconstructing ancestral states, avoiding the limitations above. BiSSE is the foundation for the techniques developed in this thesis and is described more fully in the next chapter.

### 1.3 STRUCTURE & CONTENTS OF THIS THESIS

In this thesis, I develop and apply a number of new methods for investigating species selection by detecting the association between species traits and speciation or extinction rates. In chapter 2, I derive a method for estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies, extending the BiSSE method and allowing it to be used with currently available phylogenies where an assumption of

complete sampling is often violated. In chapter 3, I develop a method (QuaSSE: Quantitative State Speciation and Extinction) for detecting an association between continuous traits (such as body size) and speciation or extinction rate. With continuous traits, the associations can take any form, and I explore the degree of detail that we can reasonably expect to recover under optimal circumstances. In chapter 4, I describe the R (R Development Core Team, 2012) package, “*diversitree*”, that I developed during my thesis and which implements BiSSE, QuaSSE, and other comparative phylogenetic methods. I also describe an extension of BiSSE to multiple states and multiple traits, and discuss ways in which models of discrete character evolution can be parametrised more intuitively. In chapter 5, I use QuaSSE to test the long-standing hypothesis that while body size tends to increase over time (Cope’s rule), this is opposed by species selection. If both of these processes are operating, then the distribution of mammalian body size might represent a trade-off between two levels of selection: individual selection for larger body size and species selection for smaller body size. Finally, in chapter 6, I close with a discussion of the limitations of the methods used in the thesis, and suggest some future directions.

## CHAPTER 2

# ESTIMATING TRAIT-DEPENDENT SPECIATION & EXTINCTION RATES FROM INCOMPLETELY RESOLVED PHYLOGENIES

---

### 2.1 SUMMARY

Species traits may influence rates of speciation and extinction, affecting both the patterns of diversification among lineages and the distribution of traits among species. Existing likelihood approaches for detecting differential diversification require complete phylogenies; i.e., every extant species must be present in a well-resolved phylogeny. We developed two likelihood methods that can be used to infer the effect of a trait on speciation and extinction without complete phylogenetic information, generalising the recent Binary State Speciation and Extinction (BiSSE) method. Our approaches can be used where a phylogeny can be reasonably assumed to be a random sample of extant species, or where all extant species are included but some are assigned only to terminal unresolved clades. We explored the effects of decreasing phylogenetic resolution on the ability of our approach to detect differential diversification within a Bayesian framework, using simulated phylogenies. Differential diversification caused by an asymmetry in speciation rates was nearly as well detected with only 50% of extant species phylogenetically resolved as with complete phylogenetic knowledge. We demonstrate our unresolved clade method with an analysis of sexual dimorphism and diversification in shorebirds (Charadriiformes). Our methods allow for the direct estimation of the effect of a trait on speciation and extinction rates using incompletely resolved phylogenies.

## 2.2 INTRODUCTION

Just as differences in traits may affect the relative survival and reproductive success of individuals, traits may affect the relative rate at which lineages go extinct or speciate (Stanley, 1975b; Coyne and Orr, 2004; Ricklefs, 2007). Sister-clade comparisons (Barraclough *et al.*, 1998) have been widely used to detect traits that are correlated with differential diversification. Using this method, traits that have been found to have a significant impact on diversification rates include diet in insects (Mitter *et al.*, 1988; Farrell, 1998), latitude in birds and butterflies (Cardillo, 1999), mating system in birds (Mittra *et al.*, 1996), and sex allocation in flowering plants (Heilbuth, 2000). These analyses have often been framed as tests of whether a character is a “key innovation”; i.e., has a particular character state lead to elevated rates of diversification? More recently, a variety of statistical approaches that directly estimate speciation rates have been developed that incorporate phylogenetic tree topology and the pattern of branching times (e.g., Pagel, 1997; Paradis, 2005; Ree, 2005; Maddison *et al.*, 2007). These approaches allow for greater statistical power than sister-clade comparisons, because they incorporate more information about the patterns of diversification. Among these is the BiSSE method (Binary State Speciation and Extinction; Maddison *et al.* 2007), a whole-tree likelihood method that can be used to detect the effect of a trait on diversification, where the trait can be classified into two states.

The BiSSE method as formulated by Maddison *et al.* (2007) assumes that the phylogenetic tree is complete and fully resolved; i.e., the tree must include every extant species. It also assumes that all character state information is known. These assumptions currently restrict its applicability, as few published phylogenies are both complete to the species level and large enough to detect differential diversification. Without appropriate correction, BiSSE will not produce valid likelihoods for incompletely resolved trees. Incomplete phylogenetic coverage decreases the apparent number of events over a phylogeny; there are fewer inferred speciation and character change events. Because of this, the BiSSE likelihood surface shifts to favour lower rates of diversification and



character change. Furthermore, inferred phylogenies that include only a fraction of extant species tend to have longer terminal branches (figure 2.1 b), and as a result the estimated extinction rates approach zero because there is a smaller increase in the number of lineages in time near the present (Nee *et al.*, 1994b).

Similar limitations have been overcome in likelihood approaches that estimate speciation and extinction rates when these rates do not depend on a character (character-independent diversification). The character-independent likelihood method of Nee *et al.* (1994b) includes corrections that assume that the species present in a phylogeny represent a random sample of extant species from a clade by incorporating the sampling process into the likelihood calculations. Recently, Bokma (2008a) developed a Bayesian approach for estimating character-independent diversification rates that treats the branching times for missing taxa as additional parameters to be estimated. In these studies, because speciation and extinction rates do not depend on a species' character state, only the branching times are required. However, if speciation and extinction rates depend on a character's state, then branching times are insufficient because the topology of the tree will depend on how the character evolves.

Here, we extend BiSSE to allow estimation of character-dependent speciation and extinction rates from incompletely resolved phylogenies. We develop likelihood calculations that compensate for incomplete phylogenetic knowledge in two cases: (1) where the species in a phylogeny represent a random sample of all extant species within a group (figure 2.1 b), and (2) where species not directly represented as tips in the phylogeny can be assigned to terminal unresolved clades (figure 2.1 c). We also develop methods to allow for incomplete character state knowledge, for both complete and incompletely resolved trees. We describe how these likelihoods can be used in Bayesian inference and apply our methods to simulated data sets. Finally, we demonstrate our method by applying it to the correlation between diversification and sexual dimorphism in shorebirds (Charadriiformes).

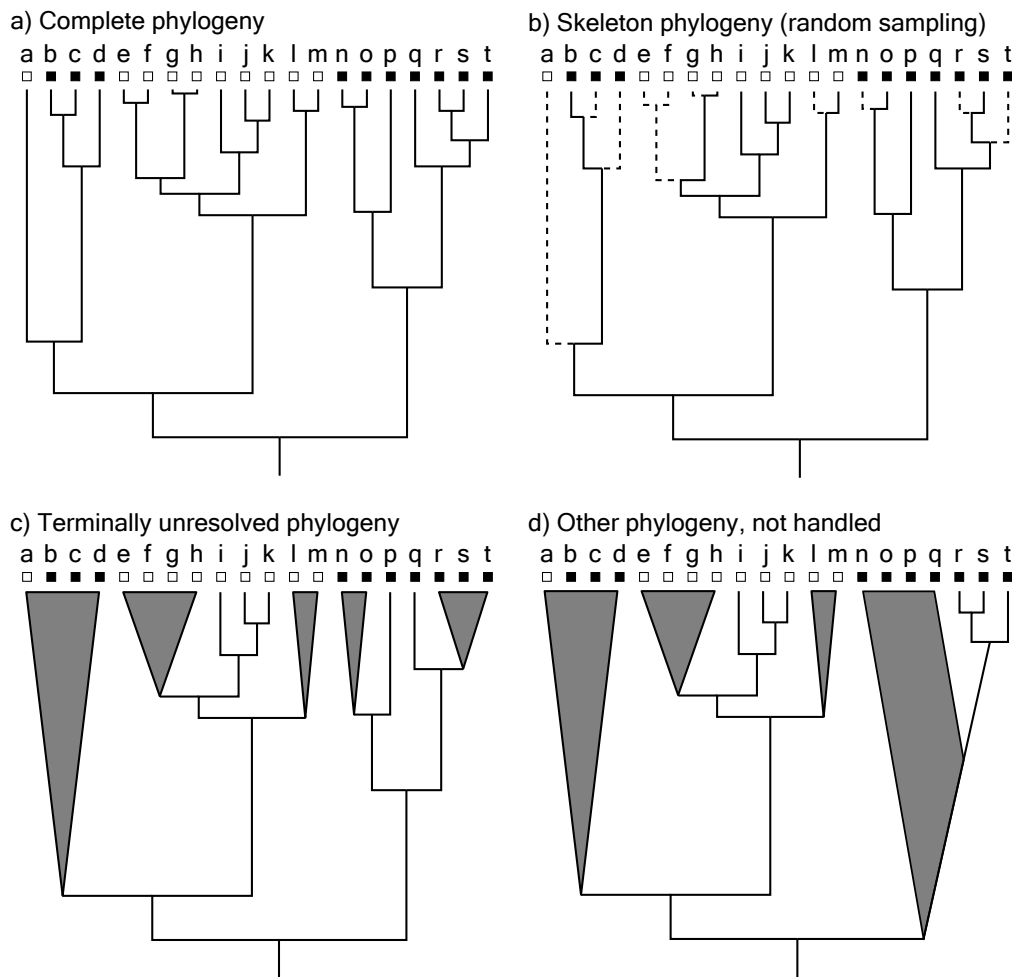


FIGURE 2.1: Different ways that phylogenetic information may be incomplete. Tree (a) is complete; every extant species is included and the tree is fully resolved. Black and white boxes above the tips refer to different character states. Tree (b) is a “skeletal tree”; species are included randomly from the full tree in (a). Sampled taxa are indicated by solid lines, and missing taxa are indicated by dashed lines. In general, nothing is known about the placement of these taxa. Tree (c) is a “terminally unresolved tree”; in this case the species not explicitly included as tips in the phylogeny are all known to belong to terminal unresolved clades. This tree is therefore “complete” in that it includes all extant taxa, but is incompletely resolved. This tree has the same branching structure as (b). Tree (d) contains a paraphyletic unresolved group and cannot be directly handled by either of the methods presented here. The relationships among species  $n - q$  are not resolved, and this group is known to be paraphyletic (see panel a). To convert this tree into a terminally unresolved tree, the known relationships within the  $r - t$  clade would have to be discarded to create an unresolved clade spanning species  $n - t$ .

### 2.3 BISSE FOR COMPLETE PHYLOGENIES

Because our aim is to generalise the BiSSE model of Maddison *et al.* (2007), we start with a brief description of this method. BiSSE computes the probability of a phylogenetic tree and the observed distribution of character states among extant species, given a model of character evolution, speciation, and extinction. The character states must be binary; we denote the possible character states as 0 or 1 (e.g., herbivorous or non-herbivorous insects). The likelihood calculation tracks two variables for each character state  $i$  along branches in a phylogeny:  $D_{Ni}(t)$  — the probability that a lineage in state  $i$  at time  $t$  would evolve into the extant clade  $N$  as observed, and  $E_i(t)$  — the probability that a lineage in state  $i$  at time  $t$  would go completely extinct by the present, leaving no extant members. (For compactness, we will often refer to the clade whose most recent common ancestor is node  $N$  as “clade  $N$ ”.) Time is measured backwards with the present at  $t = 0$ , and  $t > 0$  representing some time in the past. The changes in these quantities over time are described by a set of ordinary differential equations

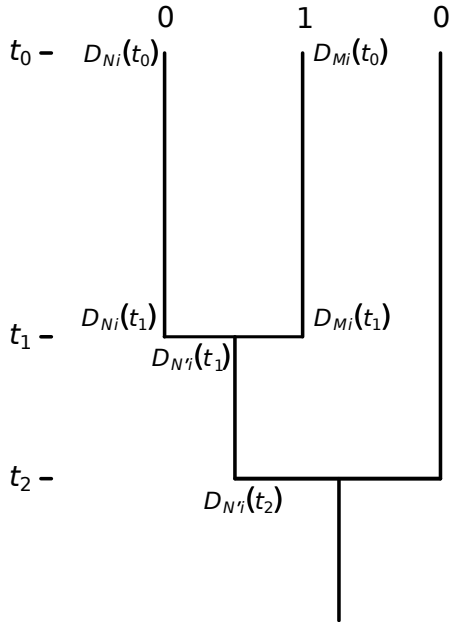
$$\frac{dD_{Ni}}{dt} = -(\lambda_i + \mu_i + q_{ij})D_{Ni}(t) + q_{ij}D_{Nj}(t) + 2\lambda_i E_i(t)D_{Ni}(t) \quad (2.1a)$$

$$\frac{dE_i}{dt} = \mu_i - (\lambda_i + \mu_i + q_{ij})E_i(t) + q_{ij}E_j(t) + \lambda_i E_i(t)^2 \quad (2.1b)$$

where  $\lambda_i$  is the speciation rate in state  $i$ ,  $\mu_i$  is the extinction rate in state  $i$ , and  $q_{ij}$  is the rate of transition from state  $i$  to  $j$  forward in time (Maddison *et al.*, 2007). These equations are solved numerically along each branch backwards in time to compute  $D_{Ni}(t)$  (figure 2.2).

On each branch, the character state at the tip provides the initial conditions for equations (2.1).  $D_{Ni}(0) = 1$  if the sampled tip  $N$  is in state  $i$  and 0 otherwise because the lineage must be in its observed state. Similarly,  $E_0(0) = E_1(0) = 0$  as a lineage cannot go extinct in zero time. At a node joining the lineages leading to clades  $N$  and  $M$ , the

a) Complete phylogeny



b) Terminally unresolved phylogeny

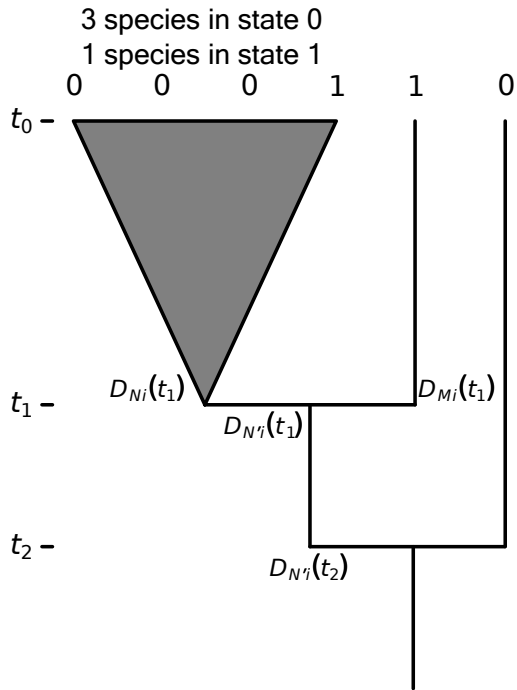


FIGURE 2.2: BiSSE with (a) and without (b) full phylogenetic knowledge. In panel (a), the values at the base of the nodes leading to the first tips ( $D_{Ni}(t_1)$  and  $D_{Mi}(t_1)$ ) are calculated backwards in time using equations (2.1) and then combined with equation (2.2) to become the initial condition  $D_{N'i}(t_1)$  for calculating  $D_{N'i}(t_2)$ . In panel (b), the four species on the left are unresolved but can be assigned to a tip that branches at time  $t_1$ .  $D_{Ni}(t)$  would be calculated forward in time using our new method, and  $D_{Mi}(t)$  with BiSSE, with these values combined as above.

probability of generating both daughter clades given that the node is in state  $i$  is

$$D_{N'i}(t) = D_{Ni}(t)D_{Mi}(t)\lambda_i \quad (2.2)$$

where  $N'$  represents the union of clades  $N$  and  $M$  (see figure 2.2). The likelihood calculation proceeds backwards in time down the tree from the tips until it reaches the root. At the root,  $R$ , we have the two probabilities  $D_{R0}$  and  $D_{R1}$ , corresponding to the possible character states at the root. The overall likelihood,  $D_R$ , must sum over the probabilities that the root was in each state (see Appendix A.1).

## 2.4 LIKELIHOOD CALCULATIONS FOR INCOMPLETELY RESOLVED PHYLOGENIES

Incompleteness in phylogenetic information can come in many forms. A species may be entirely unplaced phylogenetically, or placed into a clade but not into a precise relationship within the clade. Its character state may be known or unknown. We will derive methods for two situations: “skeletal trees,” where we have a fully resolved tree for a random sample of species whose states are fully known, and “terminally-unresolved trees,” where trees include all extant species and are fully resolved except for terminal clades that are completely unresolved phylogenetically and whose character states are known to varying degrees. Skeletal trees (figure 2.1 b) could arise when a biologist samples species simultaneously for their presence in a phylogenetic analysis and having data for the character of interest. For these trees, we assume that nothing is known about the phylogenetic placement of the missing taxa. Terminally-unresolved trees arise frequently when the species included in a molecular phylogeny are exemplars and where information on the non-included (unplaced) species is available (e.g., from previous systematic studies). If the unplaced species can be assigned to terminal clades containing the exemplar species, then our method can be used (figure 2.1 c). Here, we assume that every species can be assigned to an unresolved clade. Note that terminally unre-

solved trees are phylogenetically complete, in that they include all extant taxa, but are incompletely resolved, in that not all phylogenetic relationships are known. A broader class of incomplete phylogenies do not match either of these cases, and the methods we describe below cannot be used directly. This includes paraphyletic unresolved groups (figure 2.1 d).

#### 2.4.1 *Skeletal trees: unplaced missing taxa*

First, we consider skeletal trees, where a given phylogenetic tree represents a random sample of all extant species in a taxonomic group. To account for incomplete phylogenies, we model a “sampling” event at the present that corresponds to a biologist obtaining data for the species. This event occurs during an infinitesimally small time period, during which a species in state  $i$  has a probability  $f_i$  of being sampled for inclusion in a phylogeny. The  $f_i$  values should be determined from estimates of the numbers of species having each character state that are unsampled versus sampled. If the character states of unsampled species are unknown, then the  $f_i$  values could be set equal for all states and reflect the proportion of all extant species that have been sampled. With this sampling event,  $E_i(t)$  can be interpreted as the probability of a lineage not being present in the phylogeny, either by going extinct or not being sampled. The initial condition  $E_i(0)$  is therefore  $(1 - f_i)$ . Similarly, rather than representing the probability that a lineage in state  $i$  at time  $t$  would evolve into the full extant clade  $N$  at the present,  $D_{Ni}(t)$  includes the probability that the tip taxa present in the phylogeny are sampled. The initial conditions become  $D_{Ni}(0) = f_i$  if the sampled tip is in state  $i$  and 0 otherwise. After these modifications to the initial conditions, the calculations continue as described in Maddison *et al.* (2007). This is similar to the method used by Nee *et al.* (1994b) to correct likelihood calculations for inferring character-independent speciation and extinction rates from incomplete phylogenies. Indeed, in Appendix A.2.1 we show that when the speciation and extinction rates are independent of character state, the calculations are equivalent.

This approach assumes that the taxon sampling process is independent of the position in the phylogeny. However, it need not be independent of the character state as the  $f_i$  can differ between states 0 and 1. However, this approach also assumes that taxon sampling is even across the phylogeny, although in many cases it is not.

#### 2.4.2 *Terminally unresolved trees*

Terminally unresolved trees contain all species, but their relationships are not fully resolved, with some species grouped into unresolved clades (figure 2.1 c). We can envision this situation as comparable to the skeletal trees, but with the unplaced species not entirely unknown: we know to which terminal clades they belong, and we may know their character states. This extra information about placement and character states can be used to improve inference.

We will use the word “tip” to refer to a terminal unit in the tree, which may represent either a single extant species, or a terminal unresolved clade. Thus, the number of tips in the tree will be less than the number of species implied if there are unresolved terminal clades. We assume that the sampling of species is complete; i.e., every extant species is either present directly as a tip or can be assigned to a tip that represents an unresolved clade. We do not assume knowledge of the timing of the last common ancestor for a terminal unresolved clade, instead we assume that diversification happens at any point after splitting from its sister clade (figure 2.2). We also do not assume any particular topology for the unresolved clades, rather we sum over all possible phylogenetic histories, according to their probability. We initially assume complete knowledge of character states, but we relax this assumption in the next section.

If we can compute the probability of a terminal clade,  $D_{Ni}(t)$ , then we can combine this with the probability of their sister lineages using equation (2.2) and continue with BiSSE down the rest of the tree (figure 2.2 b). In contrast to the backward-time approach employed by BiSSE, we use a forward-time method to calculate  $D_{Ni}(t)$  for an unresolved clade. Because it has no phylogenetic structure, we cannot distinguish among the dif-

ferent possible evolutionary histories of an unresolved clade. Consequently we model clade evolution as a Markov process, tracking only the probability of different clade compositions over time. The possible clade compositions can be distinguished by the number of species in each state; let  $(n_0, n_1)$  represent a clade with  $n_0$  species in state 0 and  $n_1$  species in state 1. Even though the number of possible clade types is infinite, we truncate state space to a finite number of species. This process is similar to two birth-death processes (Nee, 2006), one for each character state, but includes transitions between the processes. We term this a “birth-death-transition process”.

If  $\mathbf{x}(t)$  is a column vector representing probabilities of different clade types at time  $t$  and  $\mathbf{Q}$  is a transition rate matrix describing the rates of changes between clade types, then the probability of generating any possible clade is given by:

$$\mathbf{x}(t_0) = \exp((t_1 - t_0) \mathbf{Q}) \cdot \mathbf{x}(t_1), \quad t_1 > t_0 \quad (2.3)$$

where  $t_1$  represents an earlier point in time to  $t_0$  and ‘exp’ represents the matrix exponential (Sidje, 1998). The values of  $\mathbf{x}(t_0)$  that correspond to the observed data can then be used as the probability of the clade evolving as observed,  $D_{Ni}(t_1)$ , for subsequent BiSSE calculations. For example, for a clade that begins in state 0 at time  $t_1$  and ends in clade type  $(3, 1)$  at the present ( $t_0$ , as in figure 2.2 b), we find  $\mathbf{x}(t_0)$  from equation (2.3) and pick out the probability of generating a  $(3, 1)$  clade from this vector, which is used as  $D_{N0}(t_1)$ . The probability of generating a clade with no species,  $(0, 0)$ , is the probability that the clade would have gone extinct,  $E_0(t_1)$ . This process is then repeated assuming that the lineage leading to the clade was initially in state 1, giving  $D_{N1}(t_1)$  and  $E_1(t_1)$ .

Before describing the transition rate matrix  $\mathbf{Q}$  we must first specify the structure of the state space  $\mathbf{x}(t)$ . Let the first element represent the probability of having zero species, the next two elements represent the two single-species clades with one species in state 0 or one species in state 1 (respectively), and so on. That is, probabilities are assigned to



the positions of  $\mathbf{x}(t)$  in the order:

$$\begin{array}{ccccccc}
 \text{zero species} & & \text{one species} & & \text{two species} & & \\
 \underbrace{(0, 0)} & , & \underbrace{(1, 0), (0, 1)} & , & \underbrace{(2, 0), (1, 1), (0, 2)} & , & \dots, \\
 & & & & \underbrace{(k, 0), (k-1, 1), (k-2, 2), \dots, (1, k-1), (0, k)}_{k \text{ species}} & , & \dots
 \end{array} \quad (2.4)$$

so that a clade with  $k$  species is represented by the  $k+1$  elements in positions  $k(k+1)/2+1$  to  $(k+1)(k+2)/2$ . To keep the state space finite, the final element of  $\mathbf{x}(t)$  is an absorbing state, representing the probability that a clade has at least  $n_{\max}$  species. By doing this, we assume that once a clade reaches  $n_{\max}$  species it is so large that there is a negligible probability of generating the observed number of species by time  $t_0$ . In practice,  $n_{\max}$  can be chosen to be large enough so that it does not significantly affect calculations (e.g., by monitoring the change in relevant values in  $\mathbf{x}(t_0)$  as  $n_{\max}$  is increased). At the base of the clade,  $t_1$ , there must have been a single ancestral lineage in either state 0 or 1. The state of the system at this time must have been a vector of zeros except for a 1 in either the second position (corresponding to  $(1, 0)$  to calculate  $D_{N0}(t)$ ) or third position (corresponding to  $(0, 1)$  to calculate  $D_{N1}(t)$ ).

To calculate  $\mathbf{Q}$ , we assume that each time step is small enough that only a single event may happen; a lineage currently in state  $(n_0, n_1)$  may have one species

- speciate, moving from state  $(n_0, n_1)$  to  $(n_0 + 1, n_1)$  or  $(n_0, n_1 + 1)$  at rate  $n_0\lambda_0$  or  $n_1\lambda_1$ , respectively,
- go extinct, moving to  $(n_0 - 1, n_1)$  or  $(n_0, n_1 - 1)$  at rate  $n_0\mu_0$  or  $n_1\mu_1$ ,
- change character state, moving to state  $(n_0 - 1, n_1 + 1)$  or  $(n_0 + 1, n_1 - 1)$  at rate  $n_0q_{01}$  or  $n_1q_{10}$ .

Using these rules, the transition rate matrix has a block structure, involving the blocks  $\mathbf{S}_k$ ,  $\mathbf{E}_k$  and  $\mathbf{C}_k$ . The block  $\mathbf{S}_k$  is a  $(k+2) \times (k+1)$  matrix describing speciation from  $k$  to

$k + 1$  species:

$$\mathbf{S}_k = \begin{pmatrix} k\lambda_0 & 0 & 0 & & & & \\ 0 & (k-1)\lambda_0 & 0 & & & & \\ 0 & \lambda_1 & (k-2)\lambda_0 & & & & \\ 0 & 0 & 2\lambda_1 & \ddots & & & \\ & & & \ddots & \lambda_0 & 0 & \\ & & & & (k-1)\lambda_1 & 0 & \\ & & & & 0 & k\lambda_1 & \end{pmatrix}.$$

$\mathbf{E}_k$  is a  $k \times (k + 1)$  matrix describing extinction from  $k$  to  $k - 1$  species:

$$\mathbf{E}_k = \begin{pmatrix} k\mu_0 & \mu_1 & 0 & & & & \\ 0 & (k-1)\mu_0 & 2\mu_1 & & & & \\ 0 & 0 & (k-2)\mu_0 & \ddots & & & \\ & & & \ddots & (k-1)\mu_1 & 0 & \\ & & & & \mu_0 & k\mu_1 & \end{pmatrix}.$$

$\mathbf{C}_k$  is a  $(k + 1) \times (k + 1)$  square matrix describing character state changes, leaving the number of species constant at  $k$ :

$$\mathbf{C}_k = \begin{pmatrix} \cdot & q_{10} & 0 & & & & \\ kq_{01} & \cdot & 2q_{10} & & & & \\ 0 & (k-1)q_{01} & \cdot & \ddots & & & \\ 0 & 0 & (k-2)q_{01} & \ddots & (k-1)q_{10} & 0 & \\ & & & \ddots & \cdot & kq_{10} & \\ & & & & q_{01} & \cdot & \end{pmatrix}$$

where the dotted elements along the diagonal of  $\mathbf{C}_k$  are chosen so that the columns of  $\mathbf{Q}$  sum to zero. Denoting matrices of zeros with  $\mathbf{0}$ , the transition rate matrix  $\mathbf{Q}$  is

$$\mathbf{Q} = \begin{pmatrix} \mathbf{C}_0 & \mathbf{E}_1 & \mathbf{0} & & & \\ \mathbf{0} & \mathbf{C}_1 & \mathbf{E}_2 & & & \\ \mathbf{0} & \mathbf{S}_1 & \mathbf{C}_2 & \ddots & & \\ \mathbf{0} & \mathbf{0} & \mathbf{S}_2 & \ddots & \mathbf{E}_{n_{\max}-1} & \mathbf{0} \\ & & & \ddots & \mathbf{C}_{n_{\max}-1} & \mathbf{0} \\ & & & & \mathbf{S}_{n_{\max}-1} & \mathbf{0} \end{pmatrix} \quad (2.5)$$

The final speciation block  $\mathbf{S}_{n_{\max}-1}$ , describing speciation into the absorbing state is an  $n_t$  element row vector:

$$((n_{\max} - 1)\lambda_0, (n_{\max} - 2)\lambda_0 + \lambda_1, \dots, (n_{\max} - 1)\lambda_1)$$

As a special case, this approach can be used to calculate likelihoods for terminally unresolved trees where speciation and extinction do not depend on a character's state, as described in Appendix A.2.2.

#### 2.4.3 Incomplete character state knowledge

Regardless of the level of phylogenetic completeness and resolution, character state information may be unknown for some species. Here, we describe corrections to the BiSSE likelihood for missing character state information for fully resolved phylogenies, skeletal trees, and terminally unresolved trees.

For fully resolved phylogenies, if no information on a character for a tip is available, then the “data” becomes the presence of the tip only. On the single branch leading to this tip, we can then interpret  $D_{Ni}(t)$  as the probability of giving rise to a single species, regardless of its character state. The initial conditions must therefore be  $D_{Ni}(0) = 1$  for both states, because with no time for extinction there is a 100% probability that

the branch will lead to the observed data. Using this logic, for skeletal trees, the initial conditions are  $D_{Ni}(t) = f_i$ .

For terminally unresolved trees, character state information may not be known for all members of an unresolved clade. In this case, we can calculate the joint probability that a clade evolved to a particular composition and that it was sampled as observed. Say that the unresolved clade of interest truly has  $x_0$  species in state 0 and  $x_1$  species in state 1 but that we know the state information only for a sample of these species so that  $s_i$  species are known to be in state  $i$ . If the probability of a species' state being known is independent of its state, then we can assume that the  $s_N = s_0 + s_1$  known species represent samples without replacement from a pool of  $x_N = x_0 + x_1$  species, and compute the sampling probability using the hypergeometric distribution. Of the  $\binom{x_N}{s_N}$  ways of sampling  $s_N$  species from this pool, there are  $\binom{x_i}{s_i}$  ways of sampling  $s_i$  species in state  $i$ . The sampling probability is therefore given by

$$\Pr(s_0, s_1 | x_0, x_1) = \frac{\binom{x_0}{s_0} \binom{x_1}{s_1}}{\binom{x_N}{s_N}}. \quad (2.6)$$

While we do not know the true number of species in each state, we can use equation (2.3) to compute the probability that the clade composition is  $(x_0, x_1)$  and then use equation (2.6) to give the probability of knowing that  $s_i$  species are in state  $i$ . To do this, we multiply the probability of generating the clade by the probability of sampling the clade as observed and sum over all possible clade compositions:

$$D_{Ni}(t) = \sum_{j=s_0}^{x_N-s_1} \Pr(j, x_N-j) \frac{\binom{j}{s_0} \binom{x_N-j}{s_1}}{\binom{x_N}{s_N}}. \quad (2.7)$$

Here,  $\Pr(x_0, x_1)$  is the probability of a clade with  $x_0$  species in state 0 and  $x_1$  species in state 1, calculated from equation (2.3). This calculation assumes that we know that there are  $x_N$  species in the clade, but this calculation can be generalised if  $x_N$  itself is not known exactly but can be described by a probability distribution. Where we have full state information (i.e.,  $s_0 + s_1 = x_N$ ) equation (2.7) reduces to  $\Pr(s_0, s_1)$ .

## 2.5 BAYESIAN INFERENCE

The above equations can be used to calculate from an incomplete phylogeny the likelihood, that is, the probability of the data given a model of speciation, extinction, and character evolution. This method can then be used to estimate rates using maximum likelihood and to compare models using likelihood ratio tests. Here we will discuss their application to Bayesian inference so that measures of parameter uncertainty can be simultaneously obtained. For a general introduction to Bayesian inference in phylogenetics, see Huelsenbeck *et al.* (2002). We will focus on the posterior probability distribution of the model parameters; that is, the probability of the parameters given the data.

To compute the posterior probability we need to specify the prior probability distribution for the parameters. We use an exponential prior for the six parameters (see Churchill, 2000). This choice reflects the philosophical preference for explanations requiring fewer events, all else being equal (Occam's razor). For example, if few species are present in state 0, then there is no information about the extinction rate for species in state 0 ( $\mu_0$ ). An exponential prior would then generate a posterior distribution with the same mean as the prior. Other common priors include a uniform prior and a uniform prior on the log of each parameter (e.g., on  $\ln(\mu_0)$ ). These are both "improper priors" because they do not integrate to a finite value over the possible range of the parameters  $(0, \infty)$ . Because of this, in the case where little signal is present in the data, the posterior will not integrate to a finite value, and cannot easily be interpreted (Gelman *et al.*, 1995). Because it is itself proper, the exponential prior always produces a proper posterior probability distribution, and it has the additional benefit that its influence on the posterior distribution can be easily detected by comparing the mean of prior and posterior distributions. The prior probability density associated with the parameter  $\theta_j$  is set to:

$$\Pr(\theta_j) = c_j e^{-c_j \theta_j} \quad (2.8)$$

where  $\theta_j$  is the value of the  $j$ th parameter, and  $c_j$  is a rate parameter. The posterior probability of the model given the data is proportional to

$$D_R(t_R) \prod_j c_j e^{-c_j \theta_j} \quad (2.9)$$

where the product is taken over the six model parameters. An exploration of the alternative priors indicated that the priors generally had negligible influence except where there were very few extant species in a given character state.

To choose values for  $c_j$ , we use a preliminary measure of the rate of diversification from the tree. Ignoring state changes and asymmetries in speciation or extinction rates, the expected number of species in a tree of length  $t_R$  is  $n = e^{(\lambda - \mu)t_R}$ , where  $\lambda$  and  $\mu$  are the character independent speciation and extinction rates (Nee *et al.*, 1994b). Rearranging, the diversification rate  $(\lambda - \mu)$  that would produce  $n$  species at time  $t_R$  is  $\ln(n)/t_R$ . We chose the prior rates so that the mean of the exponential distribution was twice this value (i.e.,  $c_j = t_R/2 \ln(n)$ ). The same prior was used for all model parameters.

## 2.6 NUMERICAL METHODS & APPLICATION

To test our method, we followed the same approach as Maddison *et al.* (2007) by simulating trees and character states using known rates and then attempting to infer those rates from the tree. We simulated trees containing 500 species with rates  $\lambda_0 = \lambda_1 = 0.1$ ,  $\mu_0 = \mu_1 = 0.03$ , and  $q_{01} = q_{10} = 0.01$  (equal rate trees) or with  $\lambda_1 = 0.2$  (unequal rate trees). These are the same rates as Maddison *et al.* (2007) for comparison, and the trees were simulated using their method.

We generated random incomplete phylogenies from these complete simulated phylogenies. To perform random taxonomic sampling to create skeletal trees (figure 2.1 b), we sampled a proportion of all tips independently of tip state. The per-state fraction of species in each state that were present in the final sample was calculated and used to specify  $f_0$  and  $f_1$  when calculating likelihoods. To simulate terminally unresolved trees,

a similar sampling routine can be used. Insofar as terminally unresolved trees can arise when character data are available for all species but detailed phylogenetic placement is available for only a sample of species, we can simulate this by choosing which species were sampled for detailed phylogenetic placement. The remaining unsampled species would be assigned to terminally unresolved clades represented by a single species that was sampled, the exemplar of the clade. However, this sampling requires some additional care, because every extant species must be either present in the phylogeny or assigned to an unresolved clade (compare figures 2.1 c and 2.1 d). Simply sampling species can leave orphaned species that fall below resolved clades and so cannot be placed into fully unresolved clades. For example, suppose that species  $j$  and  $k$  were chosen to have resolved placement from the phylogeny in figure 2.1 a, but species  $i$  left unresolved. The species  $i$  cannot be placed into an unresolved clade represented by a single sampled exemplar species and is thus “orphaned”. As a way of guaranteeing that there were no orphans in the final tree, we included a fraction of the orphan species in the sample and reassessed which species remained orphans, repeating until no orphan species were present. Note that this sampling approach does not generate a random sample of species, as assumed in our skeletal tree approach. For the results reported in this chapter, we assumed that the character states of all species were known.

### 2.6.1 *Implementation*

We implemented the above methods in the R package “diversitree” (available from <http://www.zoology.ubc.ca/prog/diversitree>). The “diversitree” package is also be accessible through recent versions of Mesquite (Maddison and Maddison, 2008). The matrix exponentiations were calculated numerically using the DMEXPV routine in Ex-pokit (Sidje, 1998). Because the transition rate matrix  $\mathbf{Q}$  is very sparse, it is practical to use this approach for unresolved clades containing up to several hundred species. The posterior probability distribution cannot be sampled from directly, so we use Markov Chain Monte Carlo (MCMC) to approximate the distribution, using slice sampling for the

parameter updates (MacKay, 2003; Neal, 2003). For each tree, we ran three independent MCMC chains for 10,000 steps from random starting locations, discarding the first 2,500 steps of each chain. While these chains are short compared to those used in tree inference, the sampler here is exploring a reasonably smooth continuous probability surface, rather than tree-space, with disjoint regions of high probability separated by areas of low probability (data not shown). Consequently, convergence of the MCMC chains was very rapid.

## 2.7 RESULTS

We briefly present the results of Bayesian inference using BiSSE with complete phylogenetic knowledge, then discuss how the statistical power is affected by incomplete phylogenetic knowledge.

### 2.7.1 *Bayesian inference with BiSSE*

Where speciation rates were equal for each character state ( $\lambda_0 = \lambda_1$ ; equal rate trees), the mean inferred speciation rates were close to the true values used to simulate the trees, and the posterior probability density was tightly distributed around this true value (figure 2.3, solid curves). Where speciation rates were unequal ( $\lambda_0 < \lambda_1$ ), species in state 0 were relatively rare (approximately 10% of extant species). Consequently, the rates for transitions in state 0 ( $\lambda_0$ ,  $\mu_0$  and  $q_{01}$ ) were less precisely estimated than for state 1, though still largely centred around their true values (figure 2.3, dashed curves). This pattern is consistent with that in Maddison *et al.* (2007), who found that the maximum likelihood estimates for the rare character state were more widely distributed than the estimates for the more common character state (their figure 4).

Some of the model parameter estimates were correlated; in particular the speciation and extinction rates for a particular character state were positively and linearly correlated (data not shown), indicating that a range of speciation/extinction rate combinations



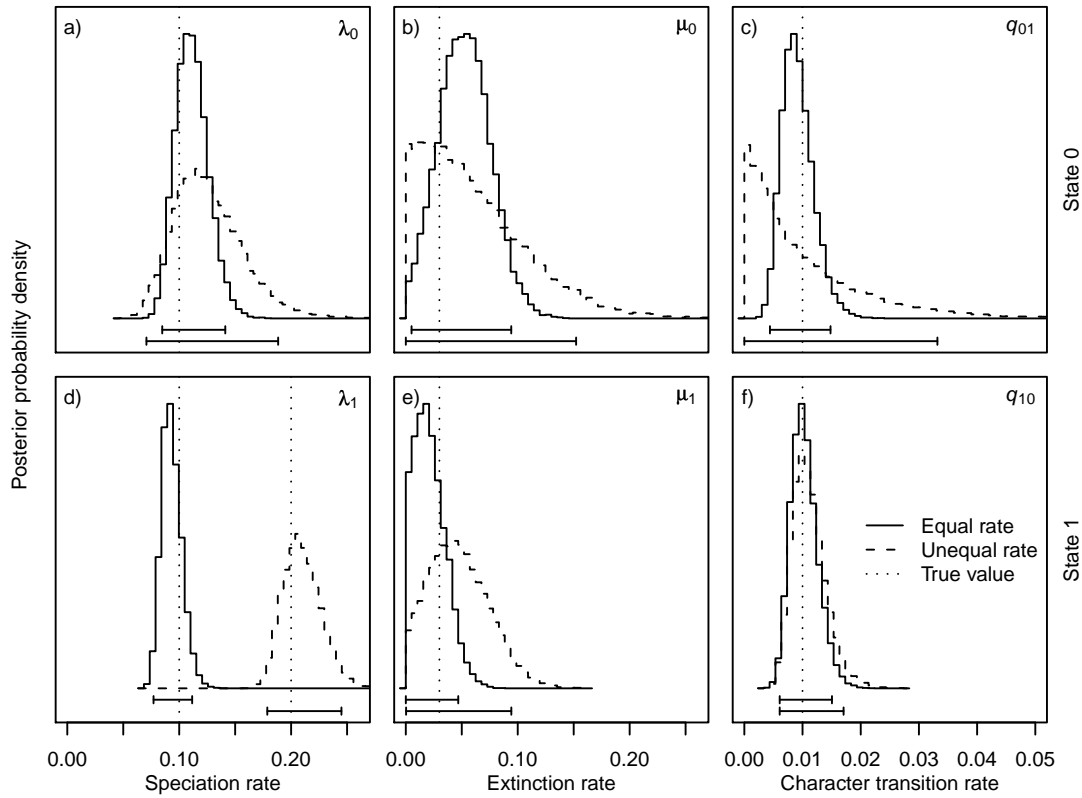


FIGURE 2.3: Posterior probability densities for the six BiSSE parameters on a fully-resolved phylogeny. Two trees were generated, each containing 500 species and with either all rates equal ( $\lambda_0 = \lambda_1$ ,  $\mu_0 = \mu_1$ ,  $q_{01} = q_{10}$ ; solid curves) or with unequal speciation rates ( $\lambda_0 < \lambda_1$ ; dashed curves). The histograms display posterior probabilities over the last 7,500 points from three independent MCMC chains, after discarding the first 2,500 points of each chain. The vertical lines indicate the true parameter values used in simulating the trees. The y axes differ between plots but are scaled so the area under each curve integrates to 1. The horizontal bars indicate the 95% credibility intervals for the equal rate (upper bar) and unequal rate (lower bar) tree.

had similar posterior probabilities. The diversification rate is the difference between the speciation and extinction rates ( $r_i = \lambda_i - \mu_i$ , Nee *et al.* 1994b). The uncertainty around the diversification rate estimate was similar to the uncertainty around the speciation rates, even where extinction rates were poorly estimated (figure 2.4). The difference between the diversification rates for the two character states (relative diversification rate;  $r_{\text{rel}} = r_1 - r_0$ ) gives a summary of the strength of differential diversification. The relative diversification rate for equal rate trees was well estimated; centred around the true value and with a narrow credibility interval. For unequal rate trees, the posterior probability distribution was flatter, but still centred around the true value. For the example shown in figure 2.4, the posterior probability of  $r_{\text{rel}} \leq 0$  for the unequal rate tree was 0.004, so we would correctly conclude that character state 1 increased diversification rate in this case.

### 2.7.2 *Effect of decreasing phylogenetic knowledge*

As fewer species were included in a phylogeny or as more species fell within unresolved clades, parameters were less accurately and precisely estimated (figure 2.5). Accuracy and precision were essentially unaffected for nearly completely resolved phylogenies (75%–100% complete), and for most parameters precision did not deteriorate substantially until trees contained fewer than  $\approx 50\%$  of the total possible tips. For all parameters, the mean parameter estimate increased when phylogenetic resolution became very low. This reflects the skew in the posterior probability distribution (figure 2.3), which increased with reduced phylogenetic resolution as the prior distribution increasingly dominates the posterior distribution. In addition, the prior means were higher than any of the simulated rates (approximately 0.35 for unequal rate trees and 0.16 for equal rate trees). Because medians are less sensitive to skew, the median parameter estimates were less affected by decreasing phylogenetic information than the mean, but they still tended to increase at very low phylogenetic resolution (not shown). The decrease in accuracy and precision was most pronounced for the rate parameters for the rare state

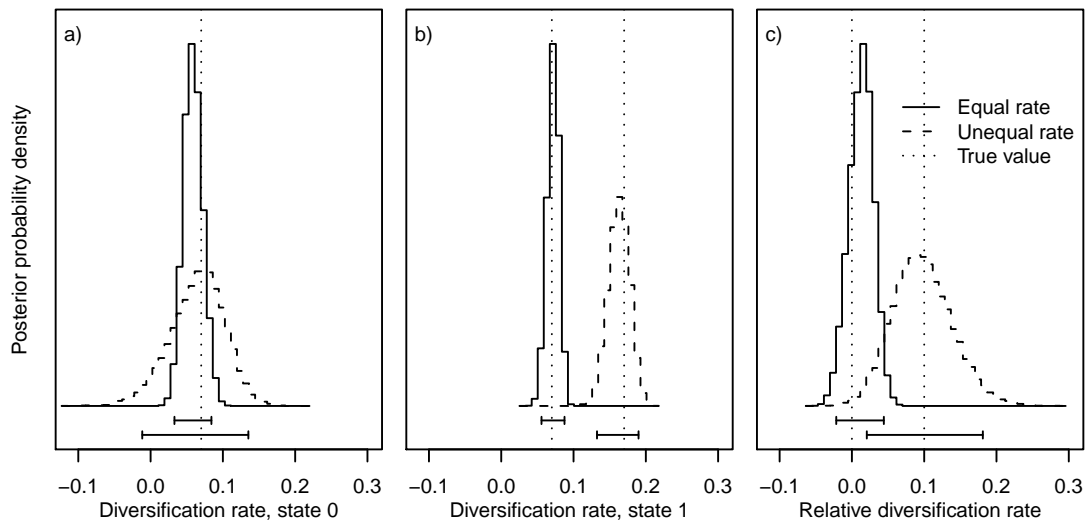


FIGURE 2.4: Posterior probability distribution for the diversification rates (a) in state 0 ( $r_0$ ) and (b) in state 1 ( $r_1$ ) and (c) the relative diversification rate ( $r_{\text{rel}} = r_1 - r_0$ ), for an equal rate tree ( $\lambda_0 = \lambda_1$ , solid curves) and an unequal rate tree ( $\lambda_0 < \lambda_1$ , dashed curves). The 95% credibility intervals are indicated by the horizontal bars, and the vertical lines indicate the true parameter values used in simulating the trees. Parameters are as indicated in the text.

on unequal rate trees ( $\lambda_0$ ,  $\mu_0$ , and  $q_{01}$ ). Decreasing phylogenetic resolution increased the uncertainty of the parameter estimates, with the widths of the credibility intervals growing as the proportion of tips sampled decreased (figure 2.5). On unequal rate trees at very low phylogenetic resolution, the posterior probability distribution for  $q_{01}$  became very similar to the prior distribution.

The decrease in accuracy and precision with decreasing phylogenetic knowledge was more pronounced for skeletal trees than for terminally unresolved trees. This is because of the additional information that the unresolved clades contain, in addition to the branching structure (i.e., the number of species and their states). In general, for a given number of tips present in a phylogeny (i.e., sampled species in skeletal trees, resolved species plus unresolved clades in terminally unresolved trees), terminally unresolved trees had lower bias in the mean parameter estimates and narrower credibility intervals than did skeletal trees. However, for well estimated parameters (e.g.,  $\lambda_1$  and  $\mu_1$  on the unequal rate tree; figure 2.5 b, e) the difference in uncertainty between these methods was small. The difference in precision was particularly pronounced for the character transition rates, which were typically well estimated on terminally unresolved trees, even with low phylogenetic resolution (figure 2.5 g–i).

Rates of net diversification were well estimated as phylogenetic resolution decreased, despite increasing uncertainty in extinction rates (figure 2.6). For equal rate trees, the net diversification rate for each trait ( $r_i = \lambda_i - \mu_i$ ) and the relative diversification rate  $r_{\text{rel}}$  were fairly insensitive to phylogenetic resolution, with no bias in the mean parameter estimates and little increase in the width of the credibility intervals, especially for terminally unresolved trees (figure 2.6 a, c, e). Where speciation rates differed (unequal rate trees), the estimated diversification rates were sensitive to decreasing phylogenetic resolution, but less so than for the individual parameters. Particularly for terminally unresolved trees, the net diversification rate was well estimated even with low phylogenetic resolution. With the parameters used here, differential diversification was detectable on unequal rate trees at the 5% significance level until fewer than 30% of taxa were

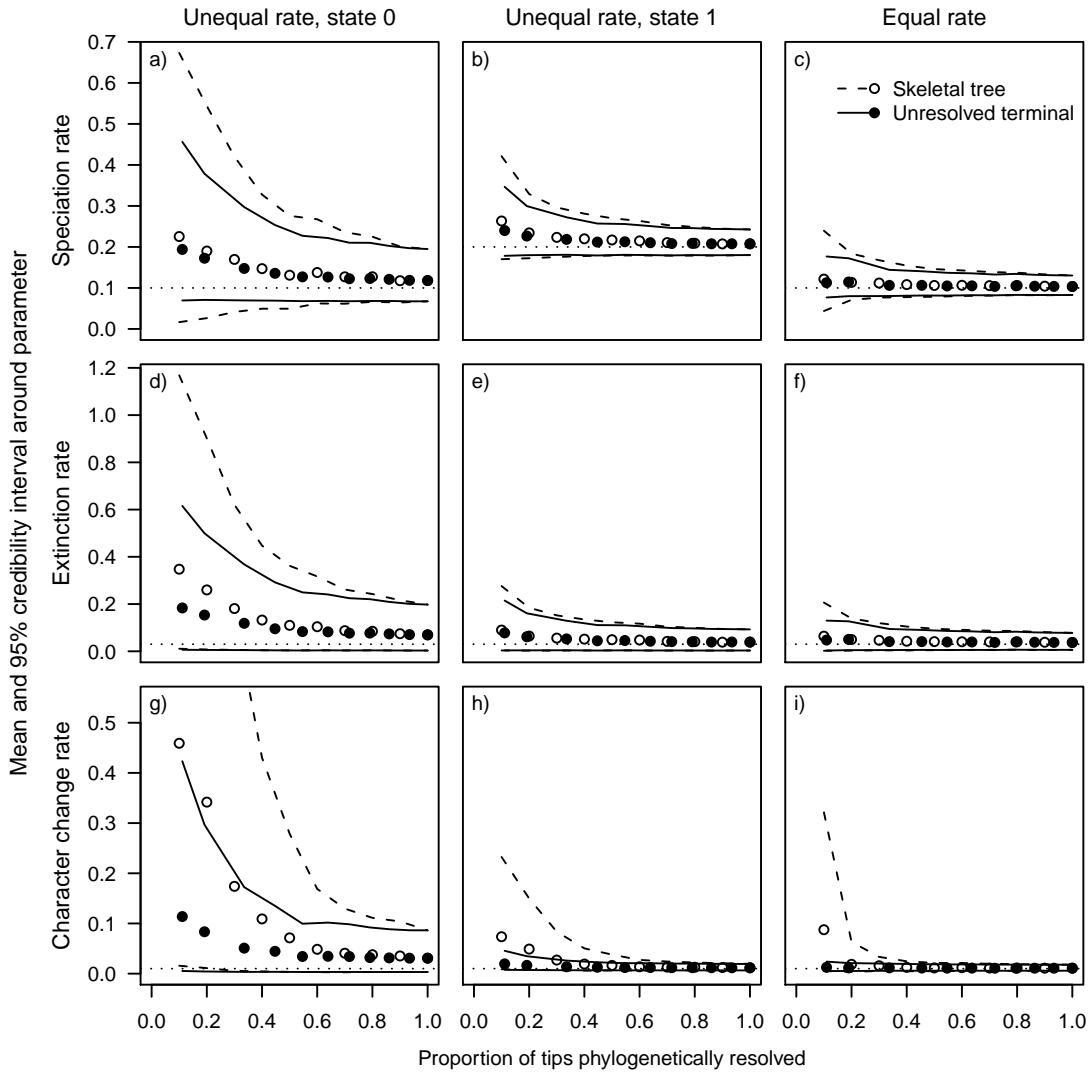


FIGURE 2.5: Uncertainty around BiSSE parameter estimates as a function of phylogenetic knowledge. Points represent the mean for the estimate of each parameters, and the curves above and below indicate the mean 95% credibility interval, averaged over 30 different phylogenies. Dashed curves/open circles represent skeletal trees and solid curves/filled circles represent terminally unresolved trees. The horizontal dotted line indicates the true rate from the simulations. For skeletal trees, the proportion of tips reflects phylogenetic completeness, while for terminally unresolved trees it represents the level of phylogenetic resolution. Trees were evolved with unequal speciation rates ( $\lambda_1 > \lambda_0$ , first two columns) or equal speciation rates (final column) and contained 500 species before sampling. Credibility intervals were calculated over the last 7,500 points of three independent MCMC chains per tree, discarding the first 2,500 points.

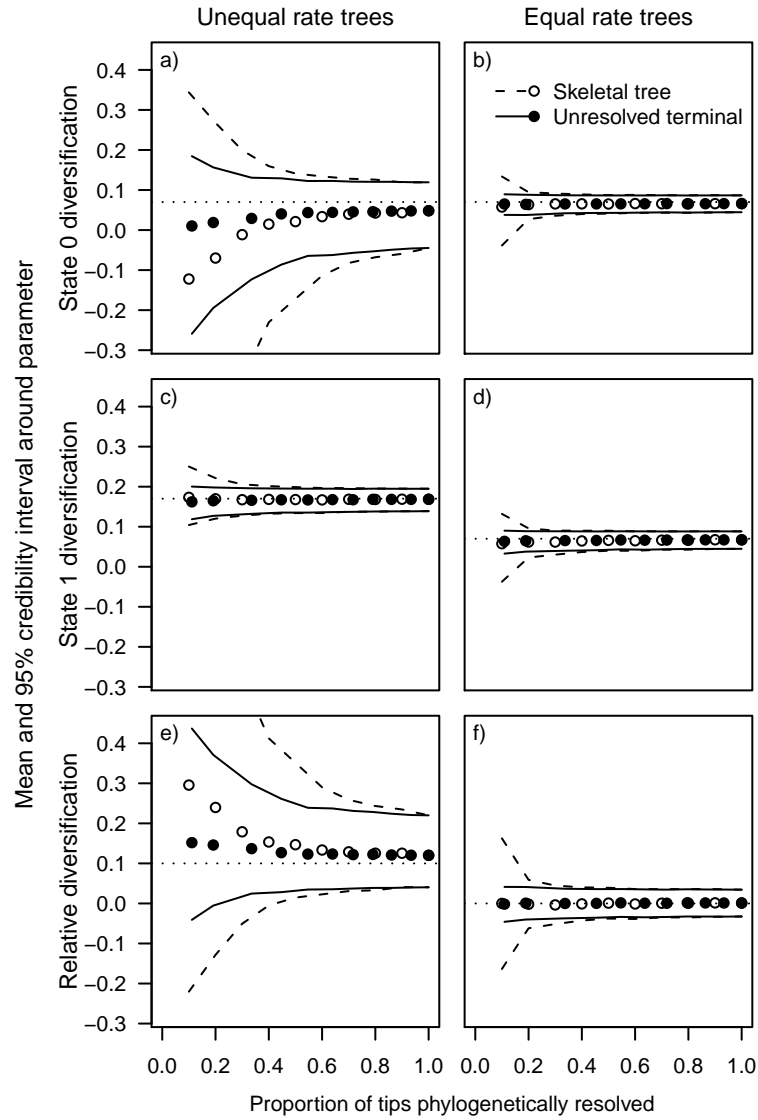


FIGURE 2.6: Uncertainty around diversification rate estimates as a function of phylogenetic knowledge. Panels (a) and (b) show the net diversification rate in state 0 ( $\lambda_0 - \mu_0$ ), c and d show the net diversification rate in state 1 ( $\lambda_1 - \mu_1$ ), and panels e and f show the relative diversification rate,  $r_{\text{rel}}$ . See figure 2.5 for details.

explicitly included in terminally unresolved trees and until 50% of taxa were included using skeletal trees (figure 2.6 f).

## 2.8 APPLICATION TO SHOREBIRD DATA

Sexual dimorphism in body size or other traits in birds is thought to be driven by sexual selection (Darwin, 1871). Larger males might be favoured by females or fare better in intra-sexual conflict over mates, while small males (reversed sexual dimorphism) might be favoured when sexual displays are acrobatic (Figueroa, 1999). Sexual differences in any trait may indicate different optima for the two sexes, and therefore inter-sexual conflict, which may increase speciation rates (Parker and Partridge, 1998; Gavrilovs, 2000; Jablonski, 2008b). Comparative evidence linking sexual dimorphism with speciation is mixed. Several studies have found that sexual dimorphism in plumage or other display traits might promote increased diversification (Barraclough *et al.*, 1995; Parker and Partridge, 1998; Owens *et al.*, 1999), while other studies failed to find correlations between measures of sexual selection and diversification rates (e.g., Gage *et al.*, 2002; Morrow and Pitcher, 2003; Morrow *et al.*, 2003).

Here, we use a recent supertree of shorebirds (Charadriiformes) to investigate the correlation between speciation rate and sexual dimorphism. Thomas *et al.* (2004) constructed a complete supertree of all 350 shorebird species. While complete, this tree lacks resolution among many of the terminal clades, with large polytomies including up to 50 species. For each polytomy, we collapsed all species descended from any lineage within the polytomy into a terminal unresolved clade (figure 2.7). The resulting tree had 134 tips (with the 215 unresolved species included in 14 unresolved terminal clades). Many of the branch lengths in this tree are not strictly proportional to time, which reduces the information about extinction rates available in the tree.

We used a database of bird traits with separate measurements for males and females of body mass, wing length, tarsus length, bill length and tail length (Lislevand *et al.*,

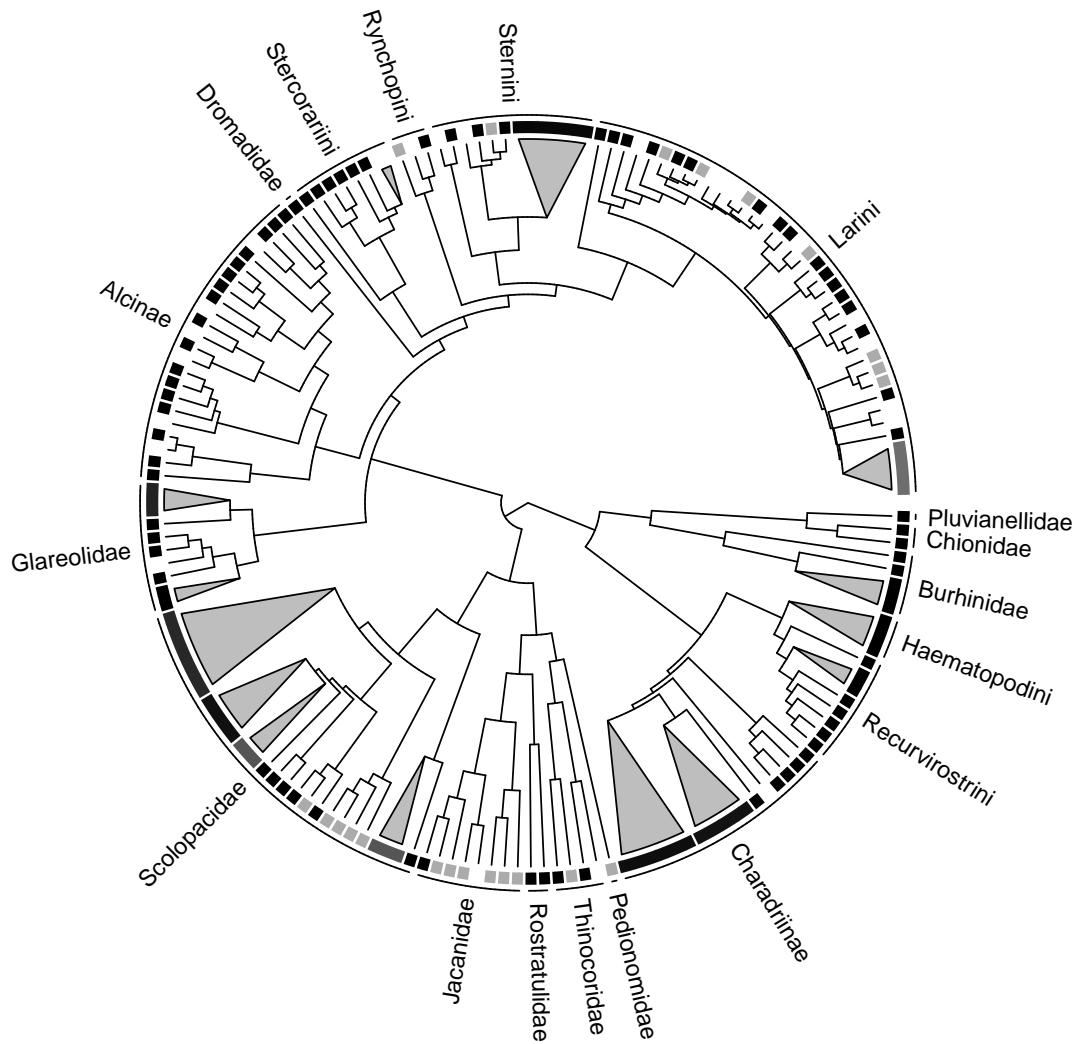


FIGURE 2.7: Phylogenetic tree of the 350 species of shorebirds (Charadriiformes) and measures of sexual dimorphism, based on Thomas *et al.* (2004). Gray triangles indicate unresolved clades, with the height of the triangle being proportional to the square root of the number of species. Character states at the 15% threshold level are indicated at the tips; grey indicates sexually dimorphic, black indicates sexually monomorphic, and white indicates no data. For unresolved clades, the degree of shading indicates the proportion of species in each state. For clarity, only family or subfamily names are shown.



2007). For each trait, we computed a standardised measure of dimorphism as  $(x_m - x_f) / \bar{x}$ , where  $x_m$  and  $x_f$  are the trait values in the males and females and  $\bar{x}$  is the mean of the male and female values. We regarded species as dimorphic if the absolute value of this dimorphism measure was greater than some threshold value for at least one of the five traits. This data set did not include state information for 77 species (22%). These were treated using the methods described in section 2.4.3.

While the general form of the marginal posterior distributions was well characterised after 10,000 steps of the MCMC algorithm, it was difficult to characterise some of the peaks in the multimodal posterior distributions. To improve resolution we ran eight independent MCMC chains for 100,000 iterations. The precise credibility intervals changed slightly, but not our general conclusions.

The relationship between sexual dimorphism and speciation and diversification rates depended on the threshold difference in body size used. For low to medium thresholds of sexual dimorphism ( $\leq 15\%$ ), the maximum likelihood and mean posterior probability speciation and diversification rates were higher for sexually dimorphic lineages than monomorphic lineages (figure 2.8). In contrast, for very high thresholds (20%) the diversification rate for dimorphic species was lower than that of monomorphic species. The difference was supported by high posterior probability values only for the 15% threshold. The maximum likelihood extinction rate and the mode of the posterior probability distribution was generally zero for both character states across all threshold values examined. We found that character transition rates from sexual dimorphism to monomorphism were higher than the reverse across most thresholds used (figure 2.8), with this difference being most pronounced at 15% (significant at the 5% level for the 10% and 15% thresholds). It is perhaps not surprising that the choice of threshold has such an effect, as the dimorphic state becomes rare as the threshold is raised, and the rarer a state the more likely its diversification rate would be biased downward.

As with previous studies, our results suggest that evidence for a correlation between sexual dimorphism and diversification rates is mixed at best (Barracough *et al.*, 1995;

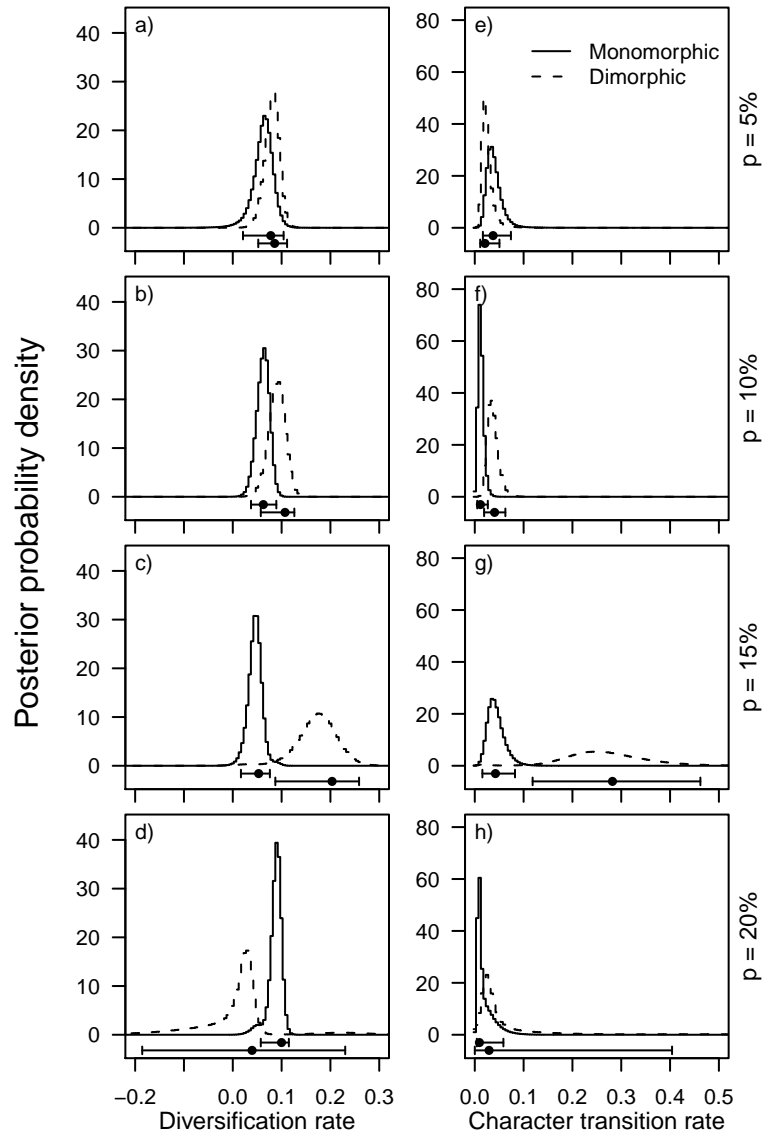


FIGURE 2.8: Marginal posterior probability distributions for the sexual dimorphism-dependent diversification rates (a–d) and character transition rates (e–h) inferred from a supertree of shorebirds (Thomas *et al.*, 2004). Panels in different rows use a different threshold level of sexual dimorphism to classify species as monomorphic and dimorphic. Solid curves show the distribution for sexually monomorphic species, and dashed curves for dimorphic species. The horizontal bar and point indicate the 95% credibility interval and maximum likelihood estimate.

Parker and Partridge, 1998; Owens *et al.*, 1999; Gage *et al.*, 2002; Morrow and Pitcher, 2003; Morrow *et al.*, 2003). However, rather than dividing groups arbitrarily into clades that have just one character state (e.g., Barraclough *et al.*, 1995), our approach allowed us to make use of all of the available phylogenetic and character state information.

## 2.9 DISCUSSION

In this chapter, we have developed two methods for estimating the effect of a trait on speciation and extinction rates from incomplete and incompletely resolved phylogenies. Testing these methods with simulations, it was possible to estimate diversification rates from even poorly sampled phylogenies (figure 2.6). Where trees were simulated with equal speciation rates, there was little increase in uncertainty in the estimates of differential diversification with decreasing phylogenetic information, even when as few as 20% of species were phylogenetically placed. This is surprising, because terminally unresolved trees lack much of the fine branching structure present at the tips of a completely resolved phylogeny (figure 2.1). However, the power to detect differences in individual parameters depended more strongly on phylogenetic structure.

Because the terminally unresolved clade method uses the branching structure available to the skeletal method (for a given number of tips in a tree), differences between these two methods are due to the additional information about the placement of the missing taxa in the terminal unresolved clades. In cases where a given species sample can be reasonably assumed to be a random draw from all extant species, the skeletal tree method provides a simple way of estimating speciation and extinction. In particular, the phylogenetic relationships and character states of non-sampled species do not need to be incorporated. Where phylogenies are almost complete, the loss of power using this method is fairly low. For poorly sampled phylogenies (fewer than 25% species included in our 500 species phylogenies), the uncertainty around parameters became very large to the point where inference was not possible (figures 2.5–2.6).

The terminal unresolved clade approach can avoid most of this loss of power, provided all species not included in the phylogeny can be grouped into terminally unresolved clades. The effect of including the terminally unresolved clades was strongest for the character transition rates ( $q_{01}$  and  $q_{10}$ , figure 2.5), and it allowed detection of differential diversification on poorly sampled trees (figure 2.6). However, the terminally unresolved tree method can only be used where every species can be assigned to a terminal unresolved clade. Deeper phylogenetic uncertainties such as unresolved paraphyletic groups have not yet been incorporated, and some known phylogenetic information may need to be discarded to use the current methods by including only terminal unresolved clades (see figure 2.1 d). This method is also substantially more computationally demanding than the skeletal tree approach and is limited at present to unresolved clades that contain fewer than approximately 200 species. With 200 species, there are over 20,000 possible clade compositions (numbers of species in each state), and even with modern matrix exponentiation techniques the calculations become both very slow and prone to numerical underflow (Sidje, 1998).

Missing data and incompleteness are generally unavoidable in comparative macroevolutionary analyses. Frequently, phylogenetic trees will contain species that are less related than expected by chance in order to maximise coverage over the true phylogeny (e.g., Moyle *et al.*, 2009). In these cases, the terminally unresolved tree method will be appropriate. If a phylogeny is almost complete, missing only a few taxa, but for which the placement is uncertain (e.g., the cases considered by Bokma, 2008a), the skeletal tree method should be satisfactory (figure 2.5). It may not always be possible to know with complete certainty where taxa that are not included in a tree should be placed within a terminally unresolved tree. In this case, one could run an analysis over possible placements of missing taxa, integrating over this uncertainty (Lutzoni *et al.*, 2001). It is probably not possible to know in general for cases such as figure 2.1 d whether collapsing known structure into an terminal unresolved clade, or treating the tree as a skeleton tree will suffer the least reduction in power, as these approaches both lose data.

A final caution about the pattern of missing taxa: outgroups are generally poorly sampled relative to the ingroup and should be removed from the tree prior to calculation of likelihoods with BiSSE. Poorly-sampled outgroups would certainly violate assumptions of random taxon sampling.

Maddison *et al.* (2007) tested whether simpler models may fit the data better by comparing models where rates were character dependent to simpler models where rates were character independent (e.g., a model where  $\lambda_0 \neq \lambda_1$  to a model where  $\lambda_0 = \lambda_1$ ) using likelihood ratio tests. Model selection between the full model and reduced models could also be done in a Bayesian framework using reversible jump Markov Chain Monte Carlo (RJMCMC; Green 1995). RJMCMC alters the Markov chain to propose different models at some steps, for example changing from the full six parameter model to one of the three simpler five parameter models. The posterior probability distribution of different models can then be directly compared. An RJMCMC approach would provide a natural way of removing parameters from the analysis where there is little to no phylogenetic signal (e.g.,  $q_{01}$  in the poorly-resolved unequal rate tree, figure 2.5 g). This approach has been used successfully elsewhere in phylogenetic inference (e.g., Pagel and Meade, 2006).

Using either of the sampling methods explored here, BiSSE likelihoods may be computed for partially complete phylogenies. Future work is needed to handle much larger unresolved clades (> 200 species) and to handle orphan taxa, whose phylogenetic position is deep in the tree and uncertain. Such extensions are needed to analyse “higher-level” phylogenies that are complete at a taxonomic level above species but contain large numbers of species in their unresolved clades (e.g., Hackett *et al.* 2008; Davies *et al.* 2004).

## CHAPTER 3

### QUANTITATIVE TRAITS & DIVERSIFICATION

---

#### 3.1 SUMMARY

Quantitative traits have long been hypothesised to affect speciation and extinction rates. For example, smaller body size or increased specialisation may be associated with increased rates of diversification. Here, I present a phylogenetic likelihood-based method (QuaSSE; Quantitative State Speciation and Extinction) that can be used to test such hypotheses using extant character distributions. This approach assumes that diversification follows a birth-death process where speciation and extinction rates may vary with one or more traits that evolve under a diffusion model. Speciation and extinction rates may be arbitrary functions of the character state, allowing much flexibility in testing models of trait-dependent diversification. I test the approach using simulated phylogenies and show that a known relationship between speciation and a quantitative character could be recovered in up to 80% of the cases on large trees (500 species). Consistent with other approaches, detecting shifts in diversification due to differences in extinction rates was harder than when due to differences in speciation rates. Finally, I demonstrate the application of QuaSSE to investigate the correlation between body size and diversification in primates, concluding that clade-specific differences in diversification may be more important than size-dependent diversification in shaping the patterns of diversity within this group.

### 3.2 INTRODUCTION

Species selection may be responsible for much of the variation in diversity among clades. Species fulfil the Lewontin (1970) definition of units of selection; different species have different traits, differential “fitness” (rates of speciation and extinction) that may be attributable to these traits, and the traits and their fitness differences are heritable. While the concept of species selection has been controversial since its inception (e.g., Stanley, 1975b; Vrba and Gould, 1986), there now seems to be reasonable agreement that the process may operate (recently discussed in Okasha, 2006; Jablonski, 2008b; Rabosky and McCune, 2009). Many traits have been proposed to affect rates of speciation and extinction, such as body size (Gittleman and Purvis, 1998), sexual system (Heilbuth, 2000), and dispersal ability (Phillimore *et al.*, 2006). In addition, hypotheses that invoke “key innovations” (e.g., floral nectar spurs: Hodges and Arnold, 1995) or evolutionary “dead ends” (e.g., asexuality: Schwander and Crespi, 2009) generally invoke trait-dependent variation in rates of diversification.

Phylogenies contain information about the timings of speciation events and patterns of diversification (Nee *et al.*, 1994b) and have been used extensively in comparative analyses to attempt to identify correlates of elevated speciation or extinction rates. Sister-clade analyses have been widely used for detecting correlates of diversification for binary traits (e.g., Mitter *et al.*, 1988; Heilbuth, 2000; Vamosi and Vamosi, 2005). These require that clades are characterised by a single character state and assume that all lineages within the clade have taken this value for the majority of its evolutionary history. More recently, likelihood approaches such as BiSSE (Binary State Speciation and Extinction; Maddison *et al.* 2007) have allowed this assumption to be relaxed, allowing any distribution of characters among extant species by using the entire pattern of branching in a phylogeny. For example, BiSSE has recently been used to demonstrate a correlation between live-bearing (vs. egg-laying) snake species and elevated speciation rates (Lynch, 2009).

Sister-clade analyses will not generally be appropriate for detecting correlation between diversification rates and continuous traits because a clade does not have a single

body size, geographic range, or latitude (but see Gittleman and Purvis, 1998, for an approach that uses mean clade traits). Several recent methods have been developed explicitly for continuous traits. Clauset and Erwin (2008) used a diffusion model to calculate equilibrium trait frequency distributions under a model where species selection is opposed by individual-level selection. However, this approach cannot incorporate phylogenetic information. The methods developed by Paradis (2005) and Freckleton *et al.* (2008) are explicitly phylogenetic, but they assume that ancestral character states can be estimated without accounting for the effect of the character on speciation and extinction (see below). None of these methods can distinguish between differential speciation and differential extinction. However, speciation and extinction rates may be correlated (high extinction rates may accompany high speciation rates; e.g., Gilinsky 1994; Coyne and Orr 2004; Liow *et al.* 2008), and many traits are thought to change diversification rates through their effect on extinction, rather than speciation (e.g., Harcourt *et al.*, 2002; Cardillo *et al.*, 2005). Approaches that allow differential speciation to be distinguished from differential extinction will therefore allow testing of a broader array of evolutionary hypotheses.

Inferring the states of ancestral nodes is problematic when the character affects speciation or extinction (Maddison, 2006; Paradis, 2008). For example, if the processes of character evolution and speciation/extinction were treated separately, then trait values might be inferred at ancestral nodes that are unlikely because they would lead to high rates of extinction. More subtly, the trait values estimated in this way will be incorrect if the evolution of the trait has a directional tendency, as it is not possible to detect directional shifts in a character under a Brownian motion model of character evolution on an ultrametric tree (Schluter *et al.*, 1997). However, where speciation, extinction, and character evolution are treated simultaneously, and where speciation rates are state-dependent, these directional changes can potentially be inferred. This would allow detection of cases where species selection and individual-level selection oppose one another (i.e., directional character change towards trait values that are disfavoured by species



selection). For example, it has been suggested that large-bodied individuals tend to have higher fitness (leading to an increase in mean size over time), while populations of large-bodied individuals are more prone to extinction (Clauset and Erwin, 2008).

Here, I describe a new comparative phylogenetic method, QuaSSE (Quantitative State Speciation and Extinction), for inferring the effect of quantitative traits on speciation and extinction rates. I first derive likelihood equations that can be used to calculate the probability of a phylogenetic tree and distribution of character states among species under a general model of cladogenesis and character evolution. I then investigate the power of this method to detect differential diversification by applying it to simulated trees. Finally, I demonstrate the method and illustrate some potential pitfalls by investigating the correlation between body size and diversification in primates.

### 3.3 CHARACTER EVOLUTION & DIVERSIFICATION

I model speciation and extinction as a birth-death process (similar to Nee *et al.*, 1994b), allowing the rates of speciation and extinction to vary with a simultaneously evolving character. Assume that a species can be characterised by its mean value of some character trait,  $x$ , which varies on the interval  $(-\infty, \infty)$ , and that this character affects diversification through its effect on the rate of speciation or extinction (or both). Let the rate of speciation for a lineage in state  $x$  be  $\lambda(x)$  and the rate of extinction be  $\mu(x)$ . These may be arbitrary non-negative functions of  $x$ , and I do not assume anything about their form. In the most general case, these can also be functions of time, so that the speciation rate for a lineage in state  $x$  at time  $t$  is  $\lambda(x, t)$ , but for notational brevity I will omit this time dependence. Incorporating time-dependence allows modelling of clade-wide changes in diversification rates (e.g., Rabosky, 2006).

It is convenient to model character evolution along lineages using a diffusion process (Allen, 2003). Diffusion processes are attractive for modelling character evolution because they allow for stochasticity while being mathematically tractable. I will measure

time backwards, with the present at time 0, and  $t > 0$  representing some time in the past. Let  $g(z, t|x, t + \Delta t)$  be the transition probability density function for the diffusion process; the probability density that a character state changes from  $x$  at time  $t + \Delta t$  to state  $z$  at time  $t$ , where  $t$  is closer to the present than  $t + \Delta t$  ( $0 < t < t + \Delta t$ ). The diffusion assumptions state that

$$\phi(x, t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} (z - x) g(z, t|x, t + \Delta t) dz \quad (3.1a)$$

$$\sigma^2(x, t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} (z - x)^2 g(z, t|x, t + \Delta t) dz \quad (3.1b)$$

$$0 = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} (z - x)^k g(z, t|x, t + \Delta t) dz, \quad k > 2, \quad (3.1c)$$

where the integral is taken over all possible character transitions (Allen, 2003). I will refer to  $\phi(x, t)$  as the “directional” term, which captures the deterministic or directional component of character evolution; this is the expected rate of change of the character over time and may be due to selection or other any other within-lineage process that has a directional tendency. This term is typically referred to as the “drift” term (e.g., Allen, 2003), but I avoid this terminology to prevent confusion with genetic drift. The term  $\sigma^2(x, t)$  is the “diffusion” term, and is the expected squared rate of change; this captures the stochastic elements of character evolution. The condition (3.1c) formally captures the assumption that large changes are unlikely by asserting that character evolution is described entirely by the first two moments of the transition probability density function. Note that both  $\phi(x, t)$  and  $\sigma^2(x, t)$  may be functions of both the character state and time. I assume that the character state is perfectly inherited by both daughter species during speciation (e.g., speciation does not lead to character displacement).

The above diffusion process generalises other models of character evolution. Brownian motion can be modelled by setting the functions  $\phi(x, t)$  and  $\sigma^2(x, t)$  to the constants  $\phi$  and  $\sigma^2$ . Where  $\phi = 0$ , this is standard Brownian motion, and where  $\phi$  is nonzero this is Brownian motion with a directional tendency (Felsenstein, 1988). The Ornstein-Uhlenbeck process captures stabilising selection that pulls the character towards a long-

term mean  $\hat{x}$  (Hansen and Martins, 1996). This can be modelled by setting the directional term,  $\phi(x, t)$ , to the linear function  $\alpha(\hat{x} - x)$  where  $\alpha$  is the strength of this stabilising force and setting the diffusion term,  $\sigma^2(x, t)$ , to the constant  $\sigma^2$ . Note that when the model of character evolution is Brownian motion (with no directional tendency), this birth-death-diffusion process is essentially that described by Paradis (2005) and Freckleton *et al.* (2008), and by Slatkin (1981) and Clauset and Erwin (2008), but disallowing character changes at nodes.

Before describing the approach, it is worth emphasising some limitations that follow from the above assumptions. The birth-death process leads to an exponential growth in the number of species (at rate  $\lambda - \mu$  when these rates are character-independent; Nee *et al.* 1994b), which is clearly not sustainable indefinitely. In addition, no interaction is possible between any of the lineages in the phylogeny; the rates of speciation, extinction and character evolution cannot depend on the number of extant species or the character states of those species. This prevents modelling of density-dependent diversification (Phillimore and Price, 2008) or frequency-dependent character evolution.

### 3.4 LIKELIHOOD CALCULATIONS

In this section, I will derive equations to compute the probability of a phylogenetic tree and character state distribution among extant species under the above model of character evolution and character-dependent speciation and extinction. I will assume that the calculations are carried out on a single ultrametric phylogenetic tree that has branch lengths proportional to time. It is straightforward to extend this analysis to integrate over a family of trees (e.g., bootstrapped trees or samples from a Bayesian analysis). I also assume that the tree is complete and fully resolved; i.e., that it includes every extant species above a common ancestor and contains no polytomies. Later, I will relax this assumption slightly to allow for partial taxon sampling.

The calculations follow the same general structure as those of BiSSE (Maddison *et al.*, 2007). Following the notation of BiSSE, let  $E(x, t)$  be the probability that a lineage in state  $x$  at time  $t$  goes completely extinct, leaving no descendants by the present (time 0). This is a continuous function in both trait-space and time, in contrast to the analogous quantities in BiSSE that were continuous only in time. Similarly, let  $D_N(x, t)$  be the probability that a lineage in state  $x$  at time  $t$  would evolve into the extant clade  $N$  as observed, including branch lengths and present-day character states. The subscript  $N$  denotes that this function applies to a particular lineage  $N$ .

### 3.4.1 Probability of extinction

Assume that we know the function  $E(x, t)$  at some time  $t$  in the past. If we can express  $E$  at a time immediately prior to this,  $t + \Delta t$ , in terms of its values at time  $t$ , then we can continue to do this until reaching the origin of a branch (Felsenstein, 1981). To do this, consider all the events that could occur over a very short period of time,  $\Delta t$ , and write  $E(x, t + \Delta t)$  by multiplying the probability of each event happening by the probability of extinction given that a particular event happened (figure 3.1). I assume that in this small period of time at most one speciation or extinction event may occur (specifically, I assume that the probability of two or more events occurring is of order  $(\Delta t)^2$  and therefore negligible with sufficiently small  $\Delta t$ ). Over this period of time, there are three possibilities; the lineage (1) goes extinct with probability  $\mu(x)\Delta t$ , (2) speciates with probability  $\lambda(x)\Delta t(1 - \mu(x)\Delta t)$ , or (3) neither speciates nor goes extinct with probability  $(1 - \lambda(x)\Delta t)(1 - \mu(x)\Delta t)$  (see Maddison *et al.*, 2007). If extinction does not occur in this time interval, character change may have occurred along the branch, and we must account for all possible character transitions that might have occurred. Where speciation occurred, this character change occurs independently on both lineages and both lineages must be extinct by the present (figure 3.1). Summing over these possibilities

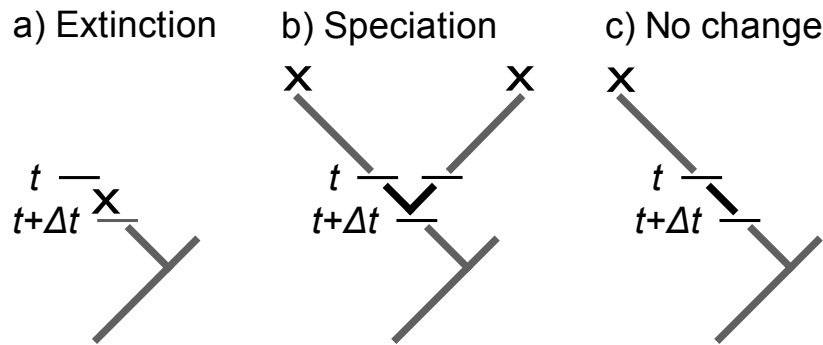


FIGURE 3.1: Possible ways a lineage extant at time  $t + \Delta t$  might go extinct. If at most a single lineage-changing event occurs, then a) extinction happens with probability  $\mu(x)\Delta t$ , leading to total extinction with probability 1, b) a speciation event happens with probability  $\lambda(x)(1 - \mu(x))\Delta t$ , leading to total extinction with probability  $E(x, t)^2$ , or c) no speciation or extinction happens with probability  $(1 - \mu(x)\Delta t - \lambda(x)\Delta t)$ , leading to total extinction with probability  $E(x, t)$ . We must integrate over the character change that might occur during this period of time; lineages in which the character may change are indicated in black.

gives

$$\begin{aligned}
E(x, t + \Delta t) = & \mu(x)\Delta t \times 1 \\
& + (1 - \mu(x)\Delta t)\lambda(x)\Delta t \left[ \int_{-\infty}^{\infty} g(z, t|x, t + \Delta t)E(z, t)dz \right]^2 \\
& + (1 - \lambda(x)\Delta t)(1 - \mu(x)\Delta t) \int_{-\infty}^{\infty} g(z, t|x, t + \Delta t)E(z, t)dz \\
& + O(\Delta t^2)
\end{aligned} \tag{3.2}$$

where  $O(\Delta t^2)$  includes terms of order  $(\Delta t)^2$  or higher (see figure 3.1). Subtracting  $E(x, t)$  from both sides, dividing by  $\Delta t$ , taking the limit  $\Delta t \rightarrow 0$ , and using the diffusion conditions above, the following partial differential equation can be derived (see Appendix B.1 for details):

$$\begin{aligned}
\frac{\partial E(x, t)}{\partial t} = & \mu(x) + \lambda(x)E(x, t)^2 - (\lambda(x) + \mu(x))E(x, t) \\
& + \phi(x, t)\frac{\partial E(x, t)}{\partial x} + \frac{\sigma^2(x, t)}{2}\frac{\partial^2 E(x, t)}{\partial x^2}. \tag{3.3}
\end{aligned}$$

### 3.4.2 Probability of the data

Next, consider the probability of a lineage including topology, branch lengths, and character states amongst its extant descendants in clade  $N$ ,  $D_N(x, t)$ . Because the calculations here assume we are not at a node, only a single lineage can be present in the reconstructed phylogeny. The three possible events that could occur over the period of time  $\Delta t$  that are consistent with this are (1) extinction, with no chance to explain the data (the extant clade  $N$ ), (2) speciation, requiring the eventual extinction of either of the resulting lineages (with the other becoming clade  $N$ ), and (3) no speciation or extinction, leaving a single lineage to become clade  $N$  (figure 3.2). Incorporating the

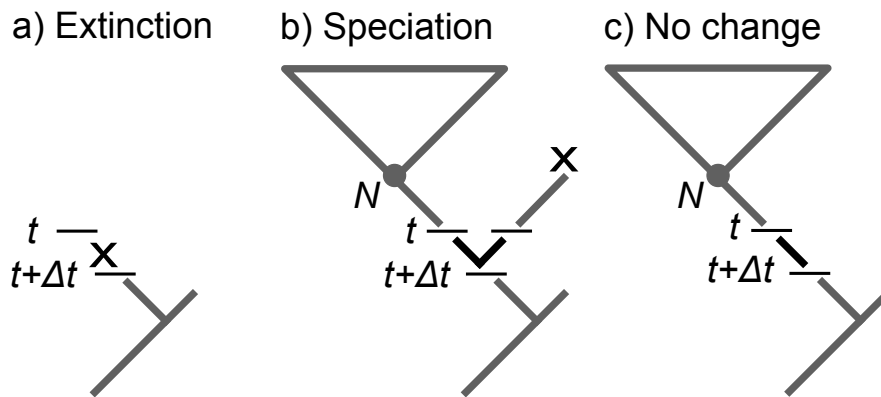


FIGURE 3.2: Possible ways a lineage extant at time  $t + \Delta t$  might lead to exactly the clade  $N$  as observed. If at most a single lineage-changing event occurs, then a) extinction happens with no chance of explaining the data, b) a speciation event requiring the extinction of either lineage (probability  $2D_N(x, t)E(x, t)$  of explaining the data), or c) no speciation or extinction, with probability  $D_N(x, t)$  of explaining the data. See figure 3.1 for other details.

possible character transitions in all non-extinct lineages as for  $E(x, t)$  gives

$$\begin{aligned}
D_N(x, t + \Delta t) = & \mu(x)\Delta t \times 0 \\
& + 2\lambda(x)(1 - \mu(x)\Delta t)\Delta t \left[ \int_{-\infty}^{\infty} g(z, t|x, t + \Delta t)D_N(z, t)dz \right] \times \\
& \left[ \int_{-\infty}^{\infty} g(z, t|x, t + \Delta t)E(z, t)dz \right] \\
& + (1 - \lambda(x)\Delta t)(1 - \mu(x)\Delta t) \int_{-\infty}^{\infty} g(z, t|x, t + \Delta t)D_N(z, t)dz \\
& + O(\Delta t^2).
\end{aligned} \tag{3.4}$$

The 2 in equation (3.4) appears because either of the two lineages that are extant after a speciation event could be consistent with the observed data, provided the other goes extinct (Maddison *et al.*, 2007). Using the same logic as was used to derive  $E(x, t)$  gives the partial differential equation

$$\begin{aligned}
\frac{\partial D_N(x, t)}{\partial t} = & 2\lambda(x)D_N(x, t)E(x, t) - (\lambda(x) + \mu(x))D_N(x, t) \\
& + \phi(x, t)\frac{\partial D_N(x, t)}{\partial x} + \frac{\sigma^2(x, t)}{2}\frac{\partial^2 D_N(x, t)}{\partial x^2}.
\end{aligned} \tag{3.5}$$

Equations (3.3) and (3.5) form the core of QuaSSE.

### 3.4.3 Initial & boundary conditions

Equations (3.3) and (3.5) do not have known analytic solutions. However, given appropriate initial and boundary conditions, they may be integrated numerically along a branch towards the root of the tree. For the initial condition for  $E$ , note that a lineage cannot go extinct in zero time, so  $E(x, 0) = 0$  for all  $x$ . The initial condition for  $D_N(x, t)$  must be a probability distribution function; i.e., it must integrate to 1 over all  $x$  because at time 0 a lineage does exist. If we knew with absolute certainty that an extant species had state



$x_{\text{obs}}$ , we could use a Dirac delta function,

$$D_N(x, 0) = \delta(x - x_{\text{obs}}),$$

which concentrates the probability distribution on the observed character state  $x_{\text{obs}}$  and integrates to 1. However, in contrast with discrete data, a species' state is never known without error, due to both within-species variation and measurement error. In the examples below, I will use a normal distribution centred on  $x_{\text{obs}}$ , with standard deviation  $\sigma_{\text{obs}}$ , but any probability distribution could be used.

To integrate these equations numerically, a finite domain and boundary conditions need to be specified. Suppose that the range  $(x_l, x_r)$  is modelled; I assume that at these boundaries the derivative of  $E(x, t)$  and  $D_N(x, t)$  with respect to  $x$  is zero (i.e., Neumann boundary conditions). This requires that the derivative of  $\lambda(x)$  and  $\mu(x)$  with respect to  $x$  is approximately zero at the boundaries and that the region is sufficiently wide that  $D_N(x, t)$  is very close to zero at the boundaries so that the probability of explaining the data from states beyond these boundaries is negligible.

#### 3.4.4 *Calculations at the nodes*

Given the initial and boundary conditions above, equations (3.3) and (3.5) can be integrated along a branch to give distributions at the base of nodes. At the node  $N'$  that joins the branches leading from nodes  $N$  and  $M$ , the initial condition is

$$D_{N'}(x, t) = D_N(x, t)D_M(x, t)\lambda(x). \quad (3.6)$$

This is the probability of the lineage at the node being in state  $x$  at time  $t$  speciating, then giving rise to both the  $N$  and  $M$  clades. This value is then used as the initial condition for the integration along the branch leading down from this node.

### 3.4.5 Calculations at the root

At the base of the tree, we have a function  $D_R(x, t_R)$ , where  $t_R$  is the time at the root. To get a single likelihood value,  $D_R$ , we must integrate over all possible character states  $x$ . This has been discussed elsewhere for the binary case (Goldberg and Igić, 2008; FitzJohn *et al.*, 2009). The simplest approach is to integrate over the possible states:

$$D_R = \int_{x_l}^{x_r} D_R(x, t_R) dx. \quad (3.7)$$

This is equivalent to assigning a flat prior to the character state at the root (e.g., Pagel, 1994). However, the tree and model contain some information about the likely state at the root, and we can use this by weighting the state  $x$  by its relative probability of yielding the observed data

$$D_R = \int_{x_l}^{x_r} D_R(x, t_R) \frac{D_R(x, t_R)}{\int_{x_l}^{x_r} D_R(y, t_R) dy} dx. \quad (3.8)$$

The latter approach is used in the calculations in this paper.

### 3.4.6 Extensions

**INCOMPLETE PHYLOGENY** — Where a phylogenetic tree does not include all extant relatives the above calculations may not be used, as they will produce incorrect likelihoods (see FitzJohn *et al.*, 2009). However, if the species included in a phylogeny represent a random sample from the extant taxa a simple modification to the calculations above allows the tree to be used, following Nee *et al.* (1994b) and FitzJohn *et al.* (2009). Suppose that a species in state  $x$  has probability  $f(x)$  of being sampled for inclusion in the tree. We can model this sampling event similar to a mass extinction at the present (Nee *et al.*, 1994b). If the probability of being included in a phylogenetic tree is thought to be independent of the character of interest, then we may set  $f(x) = f$ . Above,  $E(x, t)$  was defined as the probability that a lineage in state  $x$  at time  $t$  would have no extant descendants. For incomplete trees, we can interpret this as the probability of failing to

appear in the phylogenetic tree, either by extinction or by not being sampled. The initial conditions then become  $E(x, 0) = 1 - f(x)$ . Likewise, we can interpret  $D_N(x, t)$  as the probability that the lineage evolves into a clade  $N$  and is sampled. The initial condition  $D(x, 0)$  is then the product of a distribution describing uncertainty in the extant species state and  $f(x)$ .

**MULTIPLE CHARACTERS** — Multiple traits may affect diversification rate, and these may not evolve independently. For example, body size and latitude are likely to be correlated and are thought to both have effects on speciation and/or extinction rates (Jablonski, 2008b). Suppose that we are tracking  $k$  traits. Let  $\mathbf{x}$  be a vector of character states of length  $k$ , and let  $x_i$  be the  $i$ th trait ( $i = 1, 2, \dots, k$ ). The speciation and extinction functions become  $\lambda(\mathbf{x})$  and  $\mu(\mathbf{x})$ . As above, we retain just the first two moments of character evolution (which now include covariances) so that  $\phi_i(x, t)$  is the rate of directional evolution of the  $i$ th trait,  $\sigma_{i,i}(x, t)$  is the rate of diffusion of the  $i$ th trait, and  $\sigma_{i,j}(x, t)$  is the instantaneous covariance between the  $i$ th and  $j$ th traits ( $i \neq j$ ). In Appendix B.2, I derive the multivariate analogues to equations (3.3) and (3.5):

$$\begin{aligned} \frac{\partial E(\mathbf{x}, t)}{\partial t} = & \mu(\mathbf{x}) + \lambda(\mathbf{x})E(\mathbf{x}, t)^2 - (\lambda(\mathbf{x}) + \mu(\mathbf{x}))E(\mathbf{x}, t) \\ & + \sum_{i=1}^k \phi_i(\mathbf{x}, t) \frac{\partial E(\mathbf{x}, t)}{\partial x_i} + \sum_{i=1}^k \sum_{j=1}^k \frac{\sigma_{i,j}(\mathbf{x}, t)}{2} \frac{\partial^2 E(\mathbf{x}, t)}{\partial x_i \partial x_j} \quad (3.9) \end{aligned}$$

and

$$\begin{aligned} \frac{\partial D_N(\mathbf{x}, t)}{\partial t} = & 2\lambda(\mathbf{x})D_N(\mathbf{x}, t)E(\mathbf{x}, t) - (\lambda(\mathbf{x}) + \mu(\mathbf{x}))D_N(\mathbf{x}, t) \\ & + \sum_{i=1}^k \phi_i(\mathbf{x}, t) \frac{\partial D_N(\mathbf{x}, t)}{\partial x_i} + \sum_{i=1}^k \sum_{j=1}^k \frac{\sigma_{i,j}(\mathbf{x}, t)}{2} \frac{\partial^2 D_N(\mathbf{x}, t)}{\partial x_i \partial x_j}, \quad (3.10) \end{aligned}$$

where the single sums are taken over the directional parameters, and the double sums are taken over the diffusion parameters (when  $i = j$ ) and the covariances between characters

( $i \neq j$ ). A similar approach can be used where the second character is a binary state (see Appendix B.2).

### 3.4.7 Implementation & technical details

To integrate equations (3.3) and (3.5) numerically, I used an implicit integration scheme, where the propagation of the values in character space are performed using future values of  $E$  and  $D_N$ . To do this, I discretised both the character space and time. In each time step of size  $\Delta t$ , the changes in  $E$  and  $D_N$  through time are initially set to the character-independent solutions to equations (3.3) and (3.5),

$$E(x, t + \Delta t) = \frac{\mu(x) - \lambda(x)E(x, t) + e^{(\lambda(x) - \mu(x))\Delta t}(1 + E(x, t) - \mu(x))}{\mu(x) - \lambda(x)E(x, t) + e^{(\lambda(x) - \mu(x))\Delta t}(1 + E(x, t) - \lambda(x))} \quad (3.11a)$$

$$D_N(x, t + \Delta t) = D_N(x, t) \left( \frac{e^{(\lambda(x) - \mu(x))\Delta t}(\lambda(x) - \mu(x))}{e^{(\lambda(x) - \mu(x))\Delta t}\lambda(x)(1 - E(x, t)) - \mu(x) + \lambda(x)E(x, t)} \right)^2. \quad (3.11b)$$

(While these are character-independent, these equations need to be evaluated for each of the discretised  $x$  positions.) Following this, for each step I integrate over the character evolution that may have occurred during this period of time. For constant directional and diffusion terms (i.e., when the character evolves under Brownian motion), this can be done efficiently by convolving the functions  $E(x, t)$  and  $D_N(x, t)$  with a normal distribution with mean  $\phi\Delta t$  and variance  $\sigma^2\Delta t$ , that is the solution to the diffusion process described by the partial derivatives on the right hand side of equations (3.3) and (3.5). I implemented these calculations in R (R Development Core Team, 2012), using the Fast Fourier Transform routines in the package `FFTW` (Frigo and Johnson, 2005) to perform the convolutions. I focus on maximum likelihood estimation in the analyses in this paper, using the subplex algorithm in R to maximise the likelihood function with respect to the parameters of  $\lambda(x)$ ,  $\mu(x)$ , and the directional and diffusion coefficients. However, the likelihoods computed could be used in Bayesian calculations (e.g., FitzJohn *et al.*, 2009), although the choice of appropriate priors may not be trivial and the number

of calculations required to draw samples from the posterior using Markov Chain Monte Carlo will make this fairly slow in practice. This implementation (currently allowing only one character) is available in the R package “diversitree” (available from <http://www.zoology.ubc.ca/prog/diversitree>).

### 3.4.8 *Tree simulation*

I tested the performance of QuaSSE on simulated trees. To simulate a tree, I started with a single lineage in state  $x_0$ . Each time step, I allowed at most one lineage to speciate or go extinct, and then updated the character state of every lineage stochastically following a Brownian motion model of character evolution. I scaled time so that on average 500 time steps would occur between lineage-changing events by setting the time step equal to  $1/(500(\sum_i \lambda(x_i) + \mu(x_i)))$ , where  $x_i$  is the character state of the  $i$ th lineage and the sum is taken over all extant lineages in the tree. This ensured that the character had adequate time to evolve between speciation events to approximate continuous character evolution.

I simulated trees where there was no effect of a character on speciation or extinction (i.e.,  $\lambda(x) = \lambda$ ,  $\mu(x) = \mu$ ), and where speciation or extinction were sigmoidal functions of the character state,

$$y_0 + \frac{(y_1 - y_0)}{1 + \exp(r(x_{\text{mid}} - x))},$$

where  $y_0$  and  $y_1$  are the asymptotic values at low and high  $x$ ,  $r$  describes the steepness of the sigmoid and  $x_{\text{mid}}$  is the inflection point. I chose a sigmoidal function as this captures a directional effect of a character on speciation or extinction, while preventing negative or extremely large speciation or extinction rates. When constant, the speciation rate was 0.1 and the extinction rate was 0.03. For the differential speciation simulations the speciation rates varied with  $x$  from 0.1–0.15 (low difference), 0.1–0.2 (medium), or 0.1–0.3 (high). For differential extinction, the rates varied from 0.03–0.045 (low difference), 0.03–0.06 (medium), or 0.03–0.09 (high). For all simulations, I set  $x_{\text{mid}} = 0$  and  $r = 2.5$ . I also used two rates of character diffusion; low ( $\sigma^2 = 0.01$ ) or high ( $\sigma^2 = 0.025$ ). For these

simulations, there was no directional tendency ( $\phi = 0$ ). Note that the scale used for  $x$  is arbitrary (making the choice of  $x_{\text{mid}}$  arbitrary), and changing  $r$  is equivalent to changing  $\sigma^2$  when only one of  $\lambda(x)$  and  $\mu(x)$  varies with  $x$  and  $\phi = 0$ . For the parameter values used, most of the variation in speciation rate with respect to the character occurred over the region  $[-2, 2]$ ; I started simulated trees in a character state chosen randomly from a uniform distribution on this range.

Sigmoidal functions require four parameters (plus parameters for extinction and character evolution), and there may not be sufficient signal in the data to be able to fit such complicated models (see results). Therefore, I fit models where speciation and extinction were constant, linear, or sigmoidal functions of the trait. To make the linear models satisfy the boundary condition that  $\partial E(x, t)/\partial x$  is effectively zero at extreme values of  $x$ , I set the slope to zero for  $\lambda(x)$  and  $\mu(x)$  once the character was 20 times the character-independent maximum likelihood diffusion coefficient away from the extant character distribution. I also set the functions to zero if they became negative.

To test if speciation or extinction functions that vary with character state fit better than constant functions, I used likelihood ratio tests where model comparisons were nested. Even with large trees, the appropriate cut-off value for the  $\chi^2$  statistic may not be the expected 3.84 (e.g., Maddison *et al.*, 2007). I used simulated trees where speciation and extinction rates were independent of any character states to estimate the false positive rate. For all tree sizes the false positive rate for tests of differential speciation was close to 5% at the 5% level (Kolmogorov–Smirnov test of observed distribution vs.  $\chi^2$  distribution with 1 d.f.:  $p > 0.45$ ). However, for differential extinction, the false positive rate was higher (12% significant results at the 5% level), and the distribution of likelihood ratios significantly deviated from a  $\chi^2_1$  distribution ( $p < 0.008$ ). I therefore used empirically determined cutoff values for the three tree sizes of 6.50 (125 species), 5.98 (250 species), and 5.72 (500 species) based on these simulations for the power calculations reported below.

### 3.5 SIMULATION RESULTS

On simulated trees, there was often power to detect differential speciation, while differential extinction was always difficult, but not impossible, to detect (figure 3.3). Because a linear function was significant for almost all cases where a sigmoidal function was significant, I interpret a significant fit of the linear function as detection of differential speciation or extinction. As tree size increased, the power to detect differential speciation grew from 10–40% on 125 species trees to up to 70% on 500 species trees. As the difference between minimum and maximum speciation or extinction rates increased, differential speciation and extinction became easier to detect. However, there was essentially no power to detect differential extinction for the simulated trees unless the character had a large effect on rates of extinction (figure 3.3 f–h). Differential speciation was easier to detect in simulations with higher diffusion parameters. Increasing the diffusion parameter increases the sampling of the character space where  $\lambda(x)$  changes most with respect to  $x$ , effectively increasing the sampling of the relevant character states. Power to detect trait-dependent speciation also depended strongly on the starting position of the simulation (data not shown). The power to detect differential speciation was highest on trees where the simulation started slightly below the mean diversification rate, again reflecting the amount of sampling of the informative region of differential speciation.

Even though the trees were simulated using a sigmoidal effect of character on speciation or extinction, linear models were often preferred to the full sigmoidal model when fits were compared using the Akaike information criterion (figure 3.4). Sigmoidal models were rarely preferred for the extinction function. As tree size increased, the sigmoidal speciation models were more often preferred over linear models (data not shown). However, even for 500 species trees, sigmoidal functions were preferred in less than 40% of significant cases. This is probably due to the fact that for most smaller trees, most species occupy the roughly linear part of the sigmoidal function (figure 3.4). In addition, the sigmoidal model that was fit was often a step function rather than a

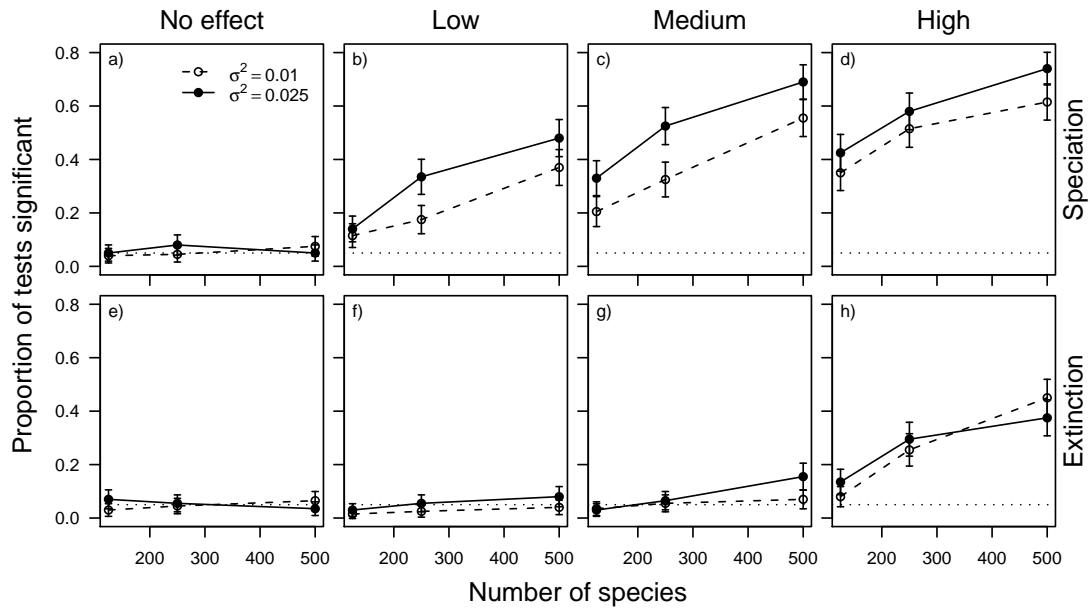


FIGURE 3.3: Power to detect differential speciation (a–d) and differential extinction (e–h) with QuaSSE on simulated phylogenies. Trees were evolved with 125, 250 or 500 species, and no, low, medium, or high effect of a character state on speciation or extinction. The character state evolved under Brownian motion at a low or high rate  $\sigma^2$ , as indicated by dashed or solid lines. Error bars are 95% binomial confidence intervals over the 200 replicate simulated trees. Horizontal dotted lines indicate the 5% significance level.



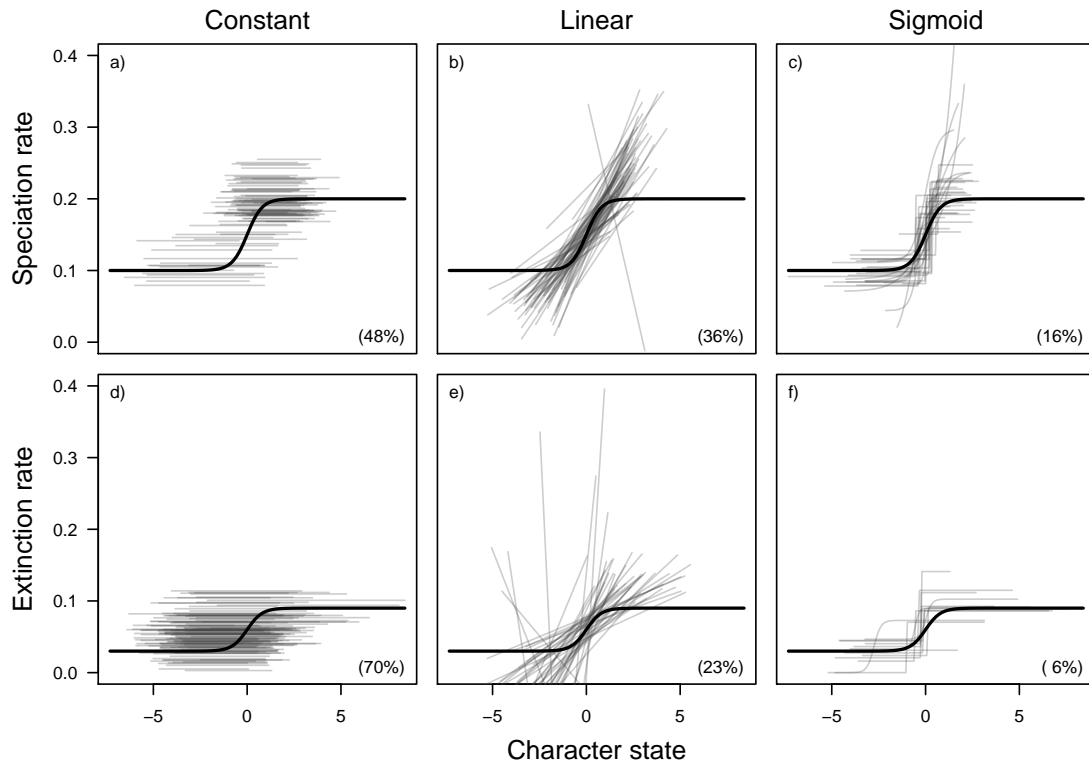


FIGURE 3.4: Representative speciation (a–c) and extinction (d–f) function fits. Functions were fit to 200 trees containing 250 taxa with a high rate of character evolution ( $\sigma^2 = 0.025$ ) and a medium effect of speciation ( $\lambda(x)$  ranged from 0.1 to 0.2) or a high effect of extinction ( $\mu(x)$  ranged from 0.03 to 0.09). Thick black lines show the true values used in the simulations. Gray lines show fits for individual trees, and span the range of observed character data for each tree. Only the best fit chosen by likelihood ratio test or AIC is displayed (see text for details).

smooth sigmoid (figure 3.4), showing that while differences in extreme speciation rates are detectable, the exact pattern may not be.

To investigate if there is power to detect directional changes in character states, I ran some simulations that included a nonzero directional parameter,  $\phi$ . I used only the set of parameters with the highest power in the absence of directional evolution (500 species, high rate of character evolution, large effect of the trait on speciation). I ran simulations where this directional tendency was negative and opposed species selection (i.e., the trait tended to decrease along a lineage, while species with larger trait values had higher rates of speciation) and where the tendency was positive and reinforced species selection. When rates of this directional tendency were very high in either direction, the power to detect differential diversification was reduced as character states tended to evolve into flatter regions of the speciation function (figure 3.5). When the directional tendency opposed species selection, there was some power to detect the trend, but this power was never high for the parameter values explored (figure 3.5). There was essentially no power to detect the presence of the directional tendency where it reinforced species selection, as both processes moved most character states into regions where speciation was constant with respect to the character state.

### 3.6 APPLICATION TO PRIMATE BODY SIZE DATA

Several studies have suggested that speciation and/or extinction are correlated with animal body size. Typically, smaller bodied species have been hypothesised to have higher speciation rates or lower extinction rates than larger bodied species (e.g., Cardillo *et al.*, 2005; Clauset and Erwin, 2008). In some groups there is also palaeontological evidence for increases in body size over evolutionary time, with species tending to be larger than their ancestors (Cope's rule; Jablonski 1997; Alroy 1998).

I investigated body size evolution in primates, testing whether body size is a correlate of speciation or extinction. I used a recent primate supertree (Vos and Mooers, 2006).

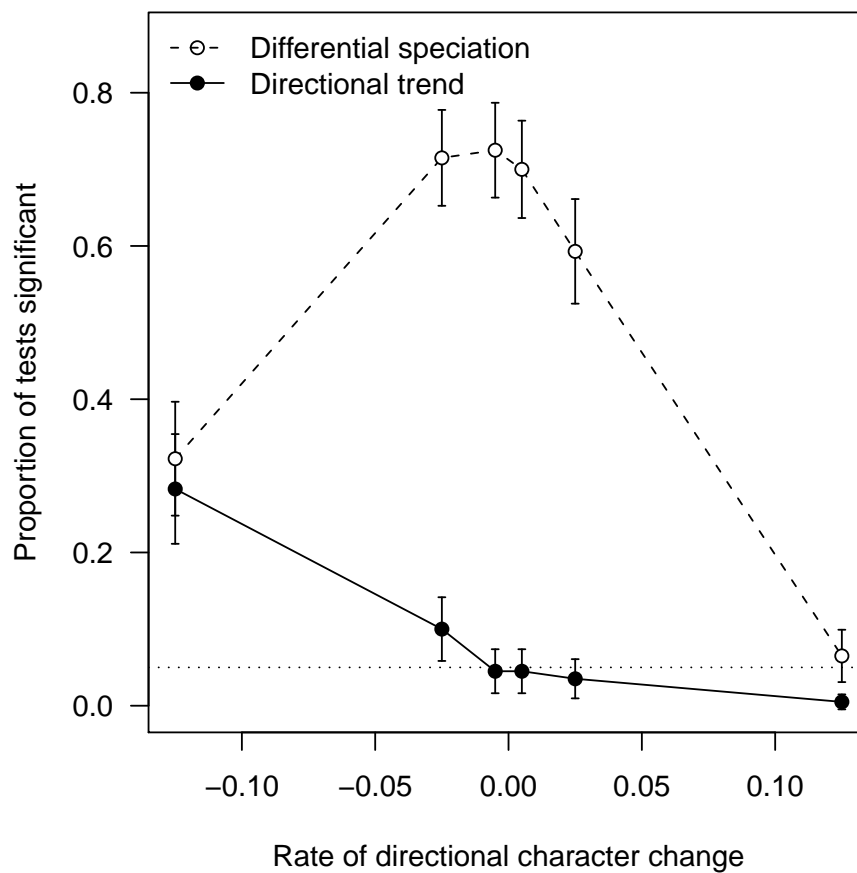


FIGURE 3.5: Power to detect trait-dependent speciation (dashed lines/open circles) and directional character change (solid lines/filled circles) at different rates of the directional tendency on simulated 500-species phylogenies.

This tree contains several polytomies that need to be resolved before running the analysis (213 of 232 internal nodes are resolved; figure 3.6). The polytomies reflect phylogenetic uncertainty, but the likelihood calculations above would misinterpret them as bursts of speciation followed by relatively low rates of speciation. It is not sufficient to randomly resolve the nodes and leave branch lengths as effectively zero, as branch lengths need to be specified in some way to prevent this misinterpretation. To do this, I used a bifurcating tree generated by T. Kuhn, in which he randomly resolved the topology of the polytomies and then used BEAST (Drummond and Rambaut, 2007) to simulate unknown branch lengths under a constant-rates birth-death model (Kuhn *et al.*, 2011). I used log-transformed female body mass from a recent collection of primate trait data as a measure of size (Redding *et al.*, 2010).

I fit several functions: constant rates for both speciation and extinction, and models where the speciation or extinction function was linear, sigmoidal, or modal. For the modal function, I used a vertically offset Gaussian

$$y_0 + (y_1 - y_0) \exp\left(-\frac{(x - x_{\text{mid}})^2}{2\omega_x^2}\right)$$

where  $y_0$  is the rate at low and high  $x$ ,  $y_1$  is the rate at the mid point  $x_{\text{mid}}$ , and  $\omega_x^2$  is the width (variance) of the Gaussian kernel. To test for the presence of directional body size evolution (e.g., Cope's rule) I also ran models where there was a nonzero, but constant, directional term ( $\phi$ ).

Among models with a directional relationship between log body size and speciation, there was strong support for a positive linear relationship between log body size and speciation rates (likelihood ratio test (LRT) against the constant rate model:  $\chi_1^2 = 10.8$ ,  $p = 0.001$ ), with a step-shaped sigmoidal curve preferred ( $\Delta \text{AIC} = 2.8$ , figure 3.7). Contrary to the predictions above, speciation rates were inferred to increase with increasing body size. However, the best fit model was a modal-speciation model (table 3.1), where species with body masses around 2.5–13.4 kg had elevated speciation rates (figure 3.7).

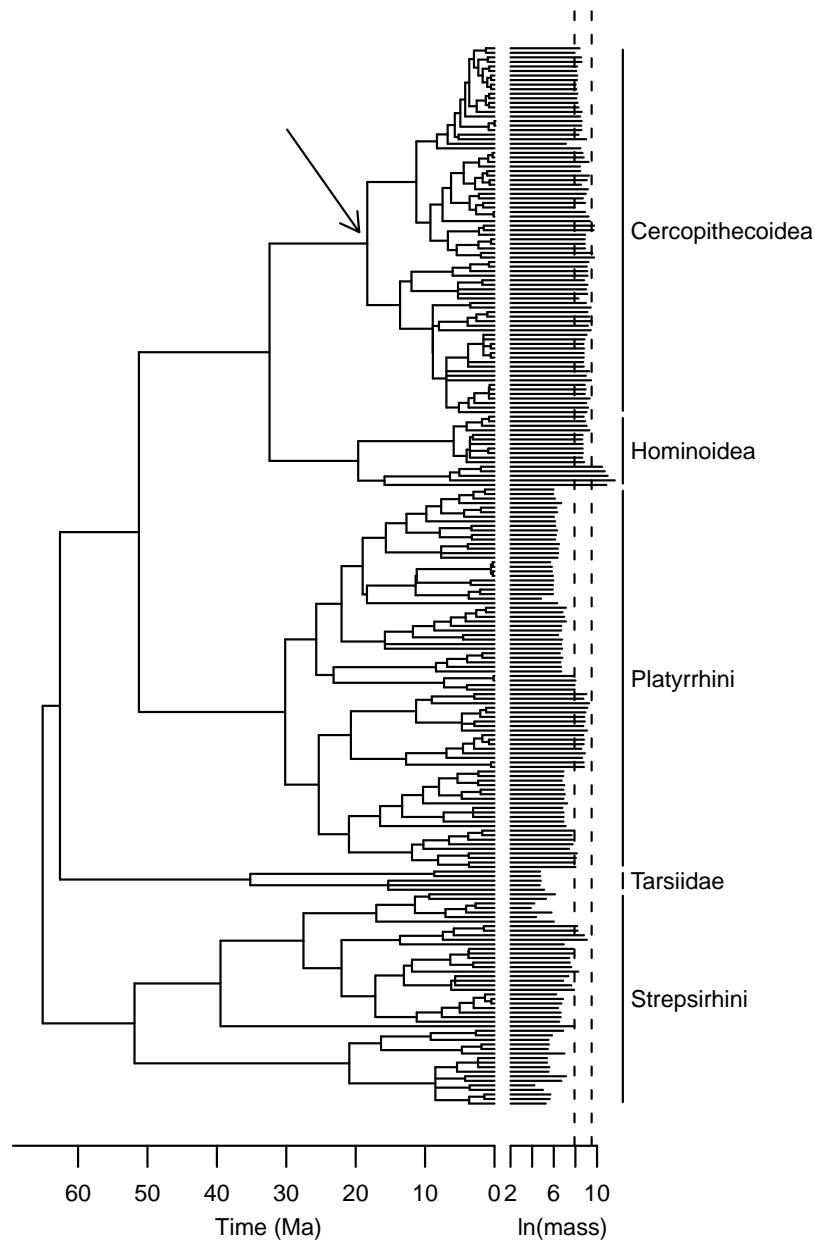


FIGURE 3.6: Phylogenetic tree of the primates, from Vos and Mooers (2006). Log body size (in grams) is shown by the horizontal bar for each species. The vertical dashed lines indicate the approximate range of body masses for which QuaSSE inferred elevated speciation rates under the “modal” speciation model (the lines indicate masses that are 10% above the base speciation rate). The arrow indicates where MEDUSA inferred a shift in speciation and extinction rates compared with the rest of the tree.

Including a positive directional term, consistent with increasing average body size along lineages, improved model fit significantly (table 3.1).

The body size with elevated speciation rates inferred using the modal-speciation function is concentrated in the Cercopithecoidea and Hominoidea (old world monkeys and apes), and this section of the tree does appear to have undergone a recent burst of relatively rapid diversification compared with the rest of the tree (figure 3.6). It is possible that any character that is concentrated in this clade could lead to a significant correlation with diversification, so this body size result could be spurious. To explore this further, I used MEDUSA (Modelling Evolutionary Diversification Under Stepwise AIC; Alfaro *et al.* 2009) to test for clade-specific differences in diversification across the tree. Using the suggested AIC difference of 4, there was support for a single partition that separated the tree into the old world monkey clade (superfamily Cercopithecoidea; figure 3.6), and the rest of the tree (LRT  $\chi^2_3 = 24.9$ ,  $p < 0.0001$ ).

I modified QuaSSE to allow this partition. I allowed a “background” group (all clades except for the old world monkeys) to have one set of speciation and extinction functions and a “foreground” group (the old world monkeys) to have another. The two groups shared a common diffusion coefficient and I set the directional term to zero. A model with constant speciation and extinction functions that could differ between the partitions had a lower (better) AIC and fewer parameters than the unpartitioned modal-speciation model (table 3.1). I found no support for any relationship between body size and either speciation or extinction for the “background” group (table 3.1). However, there was support for a model where speciation was a decreasing linear function of log-body size among the old world monkeys (LRT vs. the constant-rate partitioned model:  $\chi^2_1 = 5.1$ ,  $p = 0.024$ ) or where extinction was an increasing function of log-body size within this group ( $\chi^2_1 = 9.8$ ,  $p = 0.002$ ), suggesting decreasing diversification with increased body size in this group. However, the latter fit suggested an extremely high rate of extinction among large old world monkeys (figure 3.7).

TABLE 3.1: Summary of model fits for the association between body size and diversification for primates.

Model type	$\ln L$	$n$	AIC	$\Delta\text{AIC}$
Constant	-834.8	3	1675.7	29.6
Linear $\lambda$	-829.4	4	1666.9	20.8
Sigmoidal $\lambda$	-826.0	6	1664.1	18.0
Modal $\lambda$	-822.4	6	1656.8	10.7
<i>With directional tendency:</i>				
Linear $\lambda$	-826.0	5	1661.9	15.8
Sigmoidal $\lambda$	-823.8	7	1661.7	15.6
Modal $\lambda$	-818.8	7	1651.7	5.6
<i>Partitioned tree (no directional tendency):</i>				
Constant	-822.0	5	1654.0	7.8
Linear $\lambda$ (fg)	-819.4	6	1650.9	4.7
Linear $\lambda$ (bg)	-821.6	6	1655.2	9.0
Linear $\lambda$ (both)	-819.1	7	1652.2	6.1
Linear $\mu$ (fg)	-817.1	6	1646.1	0.0
Linear $\mu$ (bg)	-821.7	6	1655.4	9.3
Linear $\mu$ (both)	-816.8	7	1647.7	1.6

Notes: “Fg” and “bg” refer to the Cercopithecoidea clade (old world monkeys) and the rest of the tree, respectively. “Both” is where the functions were fit to both groups separately.  $\ln L$  is the log likelihood of the maximum likelihood fit,  $n$  is the number of parameters, and  $\Delta\text{AIC}$  is the AIC difference relative to the best model (linear  $\mu$  (fg)).

The results presented here are in broad accord with the analysis of body size evolution in primates by Paradis (2005), Gittleman and Purvis (1998) and Freckleton *et al.* (2008), despite using different methods, phylogenetic trees, and data sets. These studies all initially inferred a relationship of increasing diversification rates with increasing body size, though this was not significant in Gittleman and Purvis (1998). Paradis (2005) also looked at a partitioned data set and found support for decreasing diversification with increasing body size after allowing clade-specific diversification rates.

### 3.7 DISCUSSION

How much signal is there within a phylogeny about the evolutionary processes that generated it? On the simulated trees used here, it was generally possible to infer the correct trend in the character dependence of speciation, but difficult to infer the exact functional form of the trend. For instance, both the linear and sigmoidal functions capture the tendency of speciation to increase or decrease with increasing character state and the inferred linear speciation function was often a rough characterisation of the true function (figure 3.4). It is often difficult to infer ancestral states with confidence (which are needed to identify a speciation-trait correlation, even though this is only done implicitly here), as the information provided by the tips attenuates deeper into the past. Here, adding more species improved the ability to recover the more specific model, but this may be through the larger number of shallow nodes, rather than through more accurate information about deep ancestral states (Mossel and Steel, 2005).

It is possible that extinction is not possible to reliably detect on real (non-simulated) molecular phylogenies. Accurate detection of extinction requires that we determine the rate at which species fail to appear in our phylogeny, which is a difficult task. Maximum likelihood estimates of the extinction rates are frequently zero, despite fossil evidence of nonzero extinction (e.g., Nee, 2006; Purvis, 2008). However, even when ML estimates are zero, the confidence intervals around extinction rate estimates may be large, allowing



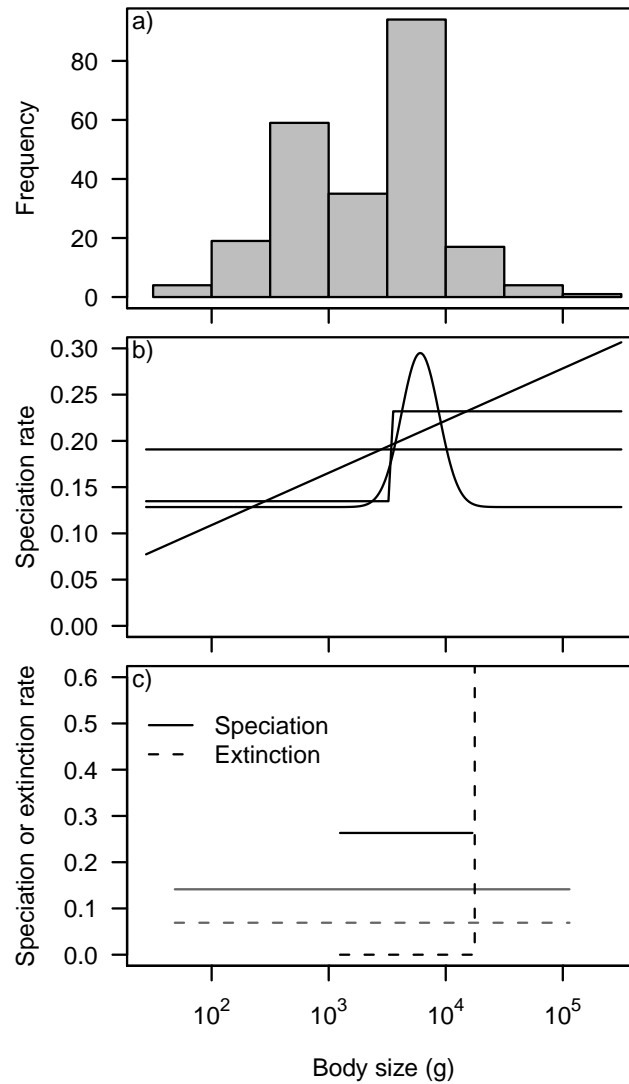


FIGURE 3.7: Primate speciation and extinction rate model fits. (a) Primate body mass distribution. (b) Maximum likelihood speciation rate model fits for the complete tree, showing constant, linear, sigmoidal and modal functions. The modal model provided the best fit to the data (table 3.1). (c) Maximum likelihood speciation and extinction functions on the partitioned tree: grey lines are fits for the “background group”, and the black lines are fits for the Cercopithecoidea. The reverse-L shape of the Cercopithecoidea extinction rate function (black dashed line in panel c) indicates a zero extinction rate for all observed body sizes and an extremely high extinction rate for species with body masses slightly greater than the largest observed mass in the Cercopithecoidea.

potentially high levels of extinction to be consistent with the observed data. Where we have strong independent evidence of high extinction rates, perhaps our analyses would be improved by including these rates directly, either through a prior distribution on extinction rates in a Bayesian analysis, or by using this estimated rate and not attempting to directly estimate it from the phylogeny. The likelihood calculations proposed here would hold in either case.

Many phylogenies appear to show some sort of slowdown in lineage accumulation towards the present, which will generate low extinction rate estimates. The response to this has generally been to alter the model of diversification. Most commonly, slowdowns have been interpreted as evidence that speciation rates may be density dependent (e.g., McPeck, 2008; Phillimore and Price, 2008), and various alternative models of cladogenesis have been proposed and tested based on this pattern (e.g., McPeck, 2008; Rabosky, 2009b). Because of its use of the birth-death model, which does not allow interaction among lineages, it would not be straightforward to incorporate these types of dynamics directly into QuaSSE, though it is possible that they may be approximated (Rabosky and Lovette 2008, but see Bokma 2009). Care should be taken to interpret results from QuaSSE and other birth-death based models (e.g., Nee *et al.*, 1994b; Paradis, 2005; Rabosky, 2006; Maddison *et al.*, 2007; Freckleton *et al.*, 2008; Alfaro *et al.*, 2009) in light of these limitations.

An alternative explanation for the observed “slowdown”, and consequent problems for estimating speciation and extinction rates, is that our methods of tree construction and ultrametricisation creates trees that are incongruous with the model. Extinction rate estimates will always be sensitive to the precise lengths of terminal branches, and any consistent bias towards lengthening the terminal branches will cause problems (Purvis, 2008). Furthermore, our delineation of species is generally retrospective, with lineages counted as species once both morphological changes and reproductive isolation have occurred. However, many isolated lineages may be considered “species” in that they will never again exchange genes. Some of these would eventually become recognised

species, but most will go extinct. However, simple birth-death models do not include this sort of process; incorporating such lags in species recognition into tree construction or diversification models, along with information from the fossil record where available, may help with efforts to infer meaningful speciation and extinction rates.

The likelihood equations derived here provide exact solutions to the forward-time dynamics described by Paradis (2005) and Freckleton *et al.* (2008), and also to the early model of Slatkin (1981), but ignoring character evolution at nodes. The key advance of this work is that it treats character evolution and cladogenesis simultaneously. Though the equations cannot be solved directly, likelihoods computed using this approach will correspond exactly to those under this model of character evolution and cladogenesis. Because the likelihood method here uses all of the available phylogenetic and character data, it should have higher statistical power than methods based on approximations, such as first inferring ancestral states and ignoring the character-dependent diversification process when doing so. Run on the same trees, the model of Freckleton *et al.* (2008) had approximately 26% of the power of QuaSSE at detecting differential speciation (data not shown). However, the factors that affect power were the same as identified by Freckleton *et al.* (2008); increased rates of character evolution, stronger effects of a character on speciation, and larger trees all increased power (figure 3.3). QuaSSE does retain some ability to detect differential extinction in contrast to the method of Freckleton *et al.* (2008), but this power appears to be limited and parameter dependent (figure 3.3). QuaSSE was also robust to the levels of background extinction used here (c.f. Paradis, 2005).

Despite their assumptions, diffusion models of character evolution and birth-death models of cladogenesis have given us insights over the last few decades into correlated character evolution (Felsenstein, 1985), evolutionary constraints (Hansen and Martins, 1996) and patterns of diversification (Alfaro *et al.*, 2009). While the combination of the birth-death and diffusion methods used in QuaSSE may inherit the limitations of both methods, it presents a tractable and powerful method that will help to answer long

standing questions about the correlates of diversification from phylogenetic data and current character distributions. As Freckleton *et al.* (2008) noted, we have no general expectation of what the relationship between speciation or extinction and character states might look like. Because QuaSSE can use arbitrary speciation and extinction functions, it allows investigation of alternative functions. However, we should not generally expect to extract more than general trends from the data, especially where variation in extinction is important in affecting patterns of diversification.

## CHAPTER 4

# DIVERSITREE: COMPARATIVE PHYLOGENETIC ANALYSES OF DIVERSIFICATION IN R

---

### 4.1 SUMMARY

The R package “diversitree” contains a number of classical and contemporary comparative phylogenetic methods. Key included methods are BiSSE (Binary State Speciation and Extinction), MuSSE (a multi-state extension of BiSSE), and QuaSSE (Quantitative State Speciation and Extinction). Diversitree also includes methods for analysing trait evolution and estimating speciation/extinction rates independently. In this chapter, I describe the features and demonstrate use of the package, using a new method, MuSSE (Multi State Speciation and Extinction), to examine the joint effects of two traits on speciation. Diversitree is open source and available on CRAN (the Comprehensive R Archive Network). A tutorial and sample data sets can be downloaded from <http://www.zoology.ubc.ca/prog/diversitree>.

### 4.2 INTRODUCTION

The tree of life is remarkably uneven in both taxonomic and trait diversity; describing this unevenness and revealing its underlying causes are major focuses of evolutionary biology. Comparative phylogenetic methods have been widely used to study patterns and rates of both trait evolution (Felsenstein, 1985; Pagel, 1994) and diversification (Nee *et al.*, 1994b). A recently developed set of models unites both trait evolution and species diversification, avoiding biases that occur when the two are treated separately (Maddi-

son, 2006). This includes the “BiSSE” method (Binary State Speciation and Extinction; Maddison *et al.*, 2007), as well as similar methods that generalise the approach to non-anagenetic trait evolution and to quantitative traits (see below).

In this chapter, I describe the “diversitree” package for R (R Development Core Team, 2012). Diversitree implements several recently developed methods for analysing trait evolution, speciation, extinction, and their interactions. Below, I describe the general approach of the package and the method that it contains. I introduce a generalisation of the BiSSE method to multi-state characters or to combinations of binary traits (MuSSE: Multi State Speciation and Extinction). Finally, I demonstrate the package, and MuSSE, with an example of social trait evolution in primates.

### 4.3 THE METHODS

The diversitree package implements a series of methods for detecting associations between traits and rates of speciation and/or extinction given a phylogeny and trait data, including the BiSSE method (Maddison *et al.*, 2007). Under BiSSE, speciation and extinction follow a birth-death process, where the rate of speciation and extinction may vary with a binary trait, itself evolving following a continuous-time Markov process. BiSSE has been used to look at the associations between many different traits and speciation or extinction, including migration in warblers (Winger *et al.*, 2012), fruiting body morphology in fungi (Wilson *et al.*, 2011), and recombination in plants (Johnson *et al.*, 2011).

In its original formulation, BiSSE assumes that character change occurs only along branches (anagenetic change), using the same model of character evolution as used in the “discrete” (Pagel, 1994) or “Mk” models (Lewis, 2001). This may not always be a reasonable assumption, and we might expect some characters to show considerable change during speciation (cladogenetic change). One such example is geographic range; while geographic ranges are expected to change anagenetically, allopatric speciation should also alter range sizes. The GeoSSE (Geographic SSE; Goldberg *et al.*, 2011) method allows

speciation rates to vary depending on a species' presence in two different geographic regions, allowing within- and between-region speciation. This has been used to examine diversification in plants endemic to serpentine regions (Anacker *et al.*, 2010). More recently, the BiSSE-ness (BiSSE-Node Enhanced State Shift; Magnuson-Ford and Otto, 2012) and ClaSSE (Cladogenetic SSE; Goldberg and Igić, in press) models have been developed to allow both anagenetic and cladogenetic character evolution, such as that expected for traits involved in ecological speciation (Schluter, 2009). Importantly, with extinction or incomplete taxonomic sampling not all speciation events will appear as nodes in a phylogeny; these missing nodes must be modelled to accurately estimate the rate of cladogenetic trait change (see Nee *et al.*, 1994b and Bokma, 2008b, and note that the placement of these missing nodes is nonlinear in time).

Diversitree also includes methods for non-binary traits. QuaSSE (Quantitative SSE; FitzJohn, 2010) allows speciation and extinction rates to be modelled as any function of a continuously varying trait, which itself evolves under Brownian motion. This has been used to test for associations between diversification rates and body size in snakes (Burbrink *et al.*, 2012) and dispersal ability in birds (Claramunt *et al.*, 2012). Finally, MuSSE extends BiSSE to multi-state traits or combinations of binary traits.

Diversitree includes variants that relax some of the original assumptions of the included methods. Birth-death based speciation/extinction models will give biased parameter estimates unless all extant taxa in the focal clade are present in a phylogeny. For cases where not all extant species are included in a phylogeny, diversitree includes methods for where species are included randomly or where all species are represented in “unresolved clades” (FitzJohn *et al.*, 2009). Rates of speciation, extinction or character change can be set to vary as any user-supplied function of time. Similar approaches have been used elsewhere to model slowdowns in speciation or diversification over time (e.g., Rabosky and Glor, 2010).

Rates of speciation, extinction, and character change may also be allowed to vary in different regions of a tree. This is similar to MEDUSA (Modelling Evolutionary Diversifi-

cation Under Stepwise AIC: Alfaro *et al.*, 2009) for diversification and AUTEUR (Eastman *et al.*, 2011) for continuous character evolution. Such methods can be used to test whether membership of a clade that has undergone a shift in diversification rates is misleading BiSSE or other methods. For example, if particular trait values are concentrated in a highly diverse clade BiSSE may detect an association when none exists (see applications in Johnson *et al.*, 2011 and FitzJohn, 2010, the diversitree tutorial for a worked example, and further discussion in Read and Nee, 1995).

In the above models, if speciation and extinction do not vary with character state, the models converge on classical models of character evolution (e.g., Pagel, 1994) and state-independent speciation and extinction (Nee *et al.*, 1994b). For completeness, these models are also included. However, when comparing models to determine if traits are associated with speciation or extinction using likelihood ratio tests, comparisons must involve only nested models to be valid. For example, BiSSE and Mk2 are not directly comparable, but BiSSE can be compared to a constrained version of BiSSE that disallows state-dependent diversification. See table 4.1 for a summary of included methods.

In addition to the likelihood calculations, tree simulation routines are implemented for birth-death, BiSSE, MuSSE, and QuaSSE. Simulating character evolution on a given tree is possible for discrete (binary or multi-state) characters and continuous characters under Brownian motion and Ornstein-Uhlenbeck processes. Likelihood-based ancestral state reconstruction (Schluter *et al.*, 1997) and stochastic character mapping (Bollback, 2006) are implemented for discrete characters.

#### 4.4 THE APPROACH

In diversitree the inference process is decoupled from the likelihood calculations, allowing users to take advantage of the programmatic flexibility of R. Analyses therefore require at least two steps. First, the user creates a likelihood function from their tree and data, using a `make.xxx` function (where xxx is one of the model types available). For



TABLE 4.1: Summary of model types available in diversitree (as of version 1.0).

Name	Trait <sup>a</sup>	Missing taxa <sup>b</sup>	Extensions <sup>c</sup>	Description and reference
bd	—	Sk, Un	Sp, Tv	Constant-rate birth-death (Nee <i>et al.</i> , 1994b)
mk2, mkn	B,M	—	Sp, Tv	Markov discrete character evolution (Pagel, 1994; Lewis, 2001)
bisse	B	Sk, Un	Sp, Tv	Binary State Speciation and Extinction (Maddison <i>et al.</i> , 2007; FitzJohn <i>et al.</i> , 2009)
bisness	B	Sk, Un	—	BiSSE-ness (Magnuson-Ford and Otto, 2012)
geosse	T	Sk	—	Geographic State Speciation and Extinction (Goldberg <i>et al.</i> , 2011)
musse	M	Sk, Un	Sp, Tv	Multi State Speciation and Extinction
classe	M	Sk	—	Clade State Speciation and Extinction (Goldberg and Igić, in press)
bm	Q	—	—	Brownian motion
ou	Q	—	—	Ornstein-Uhlenbeck
quasse	Q	Sk	Sp	Quantitative State Speciation and Extinction (FitzJohn 2011)

a: Trait type key: B = Binary (0/1), T = Ternary (three combinations of presence/absence in two regions), M = Multi-state (1, 2, 3, ...), Q = Quantitative (real-valued).

b: Missing taxa support: Sk = “Skeleton tree” (random sampling) correction, Un = “Unresolved clade”.

c: Extensions: Sp = “Split tree” (allows MEDUSA-style different rate classes in different partitions of the tree), Tv = Time-varying rates.

example, to look at character evolution under a two-state Markov model (Lewis, 2001), we would enter:

```
lik <- make.mk2(tree, states)
```

Secondly, we can find the maximum likelihood (ML) parameter vector:

```
fit <- find.mle(lik, starting.pars)
```

or use it in a Bayesian analysis by running an MCMC (Markov chain Monte Carlo) chain (with an appropriate prior):

```
samples <- mcmc(lik, starting.pars, nsteps, proposal.widths, prior)
```

or in some other use (for example, integrating the function numerically to compute the “integrated likelihood” for Bayes factors, e.g., Kass and Raftery, 1995).

Between these steps, the likelihood function can be constrained arbitrarily. *Diversitree*’s *constrain* function allows several natural constraints, such as setting one parameter equal to another, or to a specific numerical value. For example, to constrain the forward and backward transition rates to be equal (reducing the Mk2 model to the Jukes-Cantor model):

```
lik.jc <- constrain(lik, q01 ~ q10)
```

We could then find the ML parameter by entering

```
fit.jc <- find.mle(lik.jc, starting.parameters)
```

These nested models could then be compared using a likelihood ratio test.

Most of the methods included in *diversitree* are computationally challenging, but there are a number of options for controlling how the calculations are performed. Among these, the user can use different ODE solvers, and the accuracy of the calculations can be traded off against speed for most methods. Algorithms that have proven to be reasonably robust (in my experience) are used by default. For some models, such as Mk2, Brownian motion, and Ornstein-Uhlenbeck, *diversitree* provides alternative algorithms

that perform better with large numbers of states or large trees. The possible options and algorithms are discussed in Appendix C.1.

Diversitree builds on much existing software: *ape* (Paradis *et al.*, 2004) is used for tree loading and manipulation, the *deSolve* package (Soetaert *et al.*, 2010) and *sundials* library (Hindmarsh *et al.*, 2005) are used for solving the systems of differential equations for the discrete trait models, and *FFTW* (Frigo and Johnson, 2005) is used to solve the partial differential equations in QuaSSE. In addition to the R interface, Wayne Maddison has developed a wrapper around some of diversitree’s functionality to allow use from within Mesquite (Maddison and Maddison, 2008), using a user-friendly point-and-click interface.

## 4.5 THE MUSSE MODEL

MuSSE is a straightforward extension of BiSSE to discrete traits with more than two states. Some characters are not naturally binary (e.g., mating systems, diets, or count data), and MuSSE allows these to be treated naturally. This method has been used to examine the effect of diet (faunivore, folivore, frugivore) in primates (Gómez and Verdú, 2012). Alternatively, MuSSE can be used to disentangle the relative importance of two or more traits to diversification (see below).

Suppose that we have a trait that takes values  $1, 2, \dots, k$  that might influence speciation and/or extinction. Using the notation and approach of Maddison *et al.* (2007), let lineages in state  $i$  speciate at rate  $\lambda_i$ , go extinct at rate  $\mu_i$ , and transition to state  $j \neq i$  at rate  $q_{ij}$ . For  $k$  states, there are  $k$  speciation rates,  $k$  extinction rates, and  $k(k - 1)$  transition rates.

### 4.5.1 Derivation

Let  $D_{N,i}(t)$  be the probability of a lineage in state  $i$  at time  $t$  before the present ( $t = 0$ ) evolving into its descendant clade as observed, and let  $E_i(t)$  be the probability that a

lineage in state  $i$  at time  $t$ , and all of its descendants, goes extinct by the present. Under the same assumptions as Maddison *et al.* (2007) and using the same approach, it is possible to derive a set of ordinary differential equations that describe the evolution of the  $D$  and  $E$  variables over time:

$$\frac{dE_i(t)}{dt} = \mu_i - \left( \lambda_i + \mu_i + \sum_{j \neq i} q_{ij} \right) E_i(t) + \lambda_i E_i(t)^2 + \sum_{j \neq i} q_{ij} E_j(t) \quad (4.1a)$$

$$\frac{dD_{N,i}(t)}{dt} = - \left( \lambda_i + \mu_i + \sum_{j \neq i} q_{ij} \right) D_{N,i}(t) + 2\lambda_i E_i(t) D_{N,i}(t) + \sum_{j \neq i} q_{ij} D_{N,j}(t). \quad (4.1b)$$

For  $k$  states, there are  $2k$  equations.

We can solve this system of equations numerically from the tip to base of a branch. As with BiSSE the initial conditions for the  $D$  variables are 1 when the trait combination is consistent with the data, and 0 otherwise, while the initial conditions for all  $E$  variables is zero. Missing trait data is allowed by setting all  $D$  values to 1 (any state is consistent with the observed data). For the multi-trait case, if state information is available for some traits and not the others, the initial conditions are modified to allow any trait combination consistent with the observed data. For example, if trait  $A$  is in state 0 and the state of trait  $B$  is unknown, the  $D$  variables will be 1 for the combinations  $(0, 0)$  and  $(0, 1)$  and zero for combinations  $(1, 0)$  and  $(1, 1)$ . When the phylogeny is incomplete, the initial conditions can be modified by assuming random sampling (see FitzJohn *et al.*, 2009).

At the node  $N'$  that joins lineages  $N$  and  $M$ , we multiply the probabilities of both daughter lineages together with the rate of speciation

$$D_{N',i}(t) = D_{N,i}(t) D_{M,i}(t) \lambda_i. \quad (4.2)$$

The equations here assume no cladogenetic change, but this can be added following the approach in Magnuson-Ford and Otto (2012) or Goldberg and Igić (in press).

As the number of parameters in MuSSE grows quadratically with the number of states, care will often be required to prevent over-fitting and pathological behaviour associated with estimation of rate parameters involving states that are rarely observed. In particular, if some state  $i$  is not observed, then the the likelihood surface never has a negative slope with increasing  $q_{ij}$  ( $j \neq i$ ) and  $\mu_i$ , causing ML values for these parameters to tend to infinity, in turn causing problems for both the maximisation and likelihood calculation routines. For ordinal data, constraining the transition rates so that  $q_{ij} = 0$  for  $|i - j| > 1$  may be useful.

#### 4.5.2 *Analysing multiple traits simultaneously*

Alternatively, this method can be generalised to combinations of binary traits, following Pagel (1994); in this scheme, a discrete state would represent the combination of different binary traits; for  $n$  binary *traits* there are  $2^n$  possible *states*. For example, for a pair of binary traits there are four possible state combinations: (0, 0), (0, 1), (1, 0), (1, 1). We can denote these (1, 2, 3, 4), and use MuSSE directly. However, in this “multi-trait” model, parameters may be unintuitive to interpret, particularly as the number of traits increases. Moreover, with multiple traits we may be explicitly interested in asking if combinations of traits affect speciation or extinction non-additively, and this is difficult to determine with this parametrisation.

In diversitree, an alternative parametrisation is available to facilitate interpretation and model testing. Let  $\lambda_{i,j}$  be the speciation rate of a species with states  $A = i$ ,  $B = j$ , for two binary traits  $A$  and  $B$ . We can use a linear modelling approach and write

$$\lambda_{i,j} = \lambda_0 + \lambda_A X_A + \lambda_B X_B + \lambda_{AB} X_A X_B, \quad (4.3)$$

where  $X_A$  and  $X_B$  are indicator variables that are 1 when trait  $A$  and  $B$  are in the “1” state (respectively),  $\lambda_0$  is the “intercept” speciation rate (if all traits are in state 0),  $\lambda_A$  and  $\lambda_B$  are the “main effects” of traits  $A$  and  $B$ , and  $\lambda_{AB}$  is the interaction between these. If a

combination of  $A$  and  $B$  drives speciation, then a model with  $\lambda_{AB}$  will fit better than a model with just the main effects. Similarly, for the extinction rate, we write

$$\mu_{i,j} = \mu_0 + \mu_A X_A + \mu_B X_B + \mu_{AB} X_A X_B. \quad (4.4)$$

The same approach can be used for the character transition rates. If we follow Pagel (1994) and allow change in only a single trait during a single point in time, then for  $n$  traits there are only  $2n$  possible “types” of transitions (i.e., a  $0 \rightarrow 1$  or  $1 \rightarrow 0$  transition in one of the  $n$  traits). However, the rate at which these transitions happen may vary depending on the state of the other traits. For example, with two traits, we can write the rate of transition in trait  $A$  from 0 to 1, given that trait  $B$  is in state  $j$ , as

$$q_{A01,j} = q_{A01,0} + q_{A01,B} X_B. \quad (4.5)$$

where  $q_{A01,0}$  is the intercept term and  $q_{A01,B}$  is the main effect of trait  $B$ . In this scheme, if a model with  $q_{A01,B}$  fits better than a model without, then the rates of  $0 \rightarrow 1$  transition of trait  $A$  depends on the state of trait  $B$ .

Similar schemes can be derived for more traits; for more than two states, interaction terms will appear in the equations. For example, with three traits ( $A$ ,  $B$ , and  $C$ )

$$q_{A01,j,k} = q_{A01,0} + q_{A01,B} X_B + q_{A01,C} X_C + q_{A01,BC} X_B X_C \quad (4.6)$$

where  $q_{A01,C}$  is the main effect of trait  $C$  on the rate of character change of trait  $A$  from 0 to 1, and  $q_{A01,BC}$  is an interaction effect that specifies the level of non-additivity of the traits  $B$  and  $C$  on character change of trait  $A$ . Of course, this parametrisation of transition rates is valid for studying character evolution in multiple binary traits without modelling its effect on diversification (as in Pagel, 1994), and this can be done with the `make.mkn.multitrait` function.

## 4.6 SIMULATION TEST ASSESSING THE POWER OF MUSSE

There are a large number of distinct ways of modelling diversification with MuSSE, and I expect that the power of the model will depend strongly on the model specification. For example, one might have an ordinal multi-state trait, where transitions can only occur between adjacent states, and be interested asking whether large or small values of that trait are associated with elevated rates of diversification. For a given number of states ( $> 2$ ), such a model will have far fewer parameters (and greater power) than a model where the trait is purely categorical, such as diet, if all transitions are possible. The power of MuSSE will strongly depend on the number of estimated parameters (especially the character transition parameters), and I expect that for any more than four states, careful consideration of constraints in the transition parameters will be needed.

Here, I focus on a simple multi-trait case where there is some number of uncorrelated binary traits that evolve at the same rate, one of which influences the rate of speciation. I investigate the ability of MuSSE to correctly identify the trait associated with elevated speciation and to rule out the association with other traits, as a function of clade size and number of possible traits.

To simulate trees, I set the intercept speciation and extinction rates ( $\lambda_0$  and  $\mu_0$ ) to 0.1 and 0.03 respectively, and character transition rates ( $q_{X01}$ ,  $q_{X10}$ , for traits  $X = A, B, \dots$ ) to 0.01. I set  $\lambda_A = 0.1$  so that when trait  $A$  is in state 1, the speciation rate is 0.2. When only a single trait is considered, these are the same parameters used by Maddison *et al.* (2007) in their “asymmetric speciation” case. I simulated phylogenies and character state transitions under the multitrait MuSSE model, starting at the root in one of the “low” speciation states (with  $A$  in state 0), sampling randomly for the other traits. Trees were simulated to contain 50, 100, 200, or 400 species, with 1, 2, 3, or 4 traits, and with 100 replicate trees for each of the 16 combinations.

For each tree, I ran a Markov chain Monte Carlo (MCMC) analysis on a model where all speciation main effects were free to vary (but excluded interactions), fitting only intercepts for extinction and character change. For example, with two traits this meant

that the free parameters were  $\lambda_0$ ,  $\lambda_A$ ,  $\lambda_B$ ,  $\mu_0$ ,  $q_{A01,0}$ ,  $q_{A10,0}$ ,  $q_{B01,0}$ , and  $q_{B10,0}$ . This model is very close to the true model, but allows for uncertainty in which trait is responsible for increased speciation (trait *A* or *B*). I used an exponential prior with a mean of twice the state-independent diversification rate for all the underlying rate parameters (see Appendix C.3). I ran each chain for 10,000 steps, and discarded the first 500 steps as “burn-in”. Because the “dummy” traits *B*, *C*, and *D* are equivalent where present, I report results primarily for trait *A* (which increases speciation rates when in state 1) and trait *B* (which does not affect speciation rates).

As the size of the tree increased, the credibility intervals around the main effects on speciation decreased, and the mean estimated effect converged on the true values (see figure 4.1). The uncertainty around the dummy trait, *B*, was not strongly affected by the number of dummy traits that were included, and decreased slightly as more traits were included. For small trees ( $\leq 100$  species), MuSSE underestimated the effect of trait *A* on speciation rates, especially as the number of traits increased.

Significance showed similar patterns. As tree size increased, power to correctly identify *A* as the trait associated with increased speciation increased (figure 4.2, blue lines), but for trees with 100 species or more this varied only weakly with the number of included traits. The dummy trait *B* was significant approximately 5% of the time (based on 95% credibility intervals): the rate expected due to type I error (figure 4.2, solid green lines).

To test how model misspecification would affect the results, I also reran the analyses with trait *A* omitted so that none of the analysed traits were truly associated with state-dependent diversification. The dummy trait *B* was incorrectly associated with increased speciation in up to 27% of trees (figure 4.2). While this effect was strongest when there were fewer dummy traits, the possibility of *any* trait being falsely associated with diversification increased. Indeed, where three dummy traits are included, the probability of associating any trait with increased speciation increased to 59% for the 400 species tree (figure 4.2, dotted orange lines).



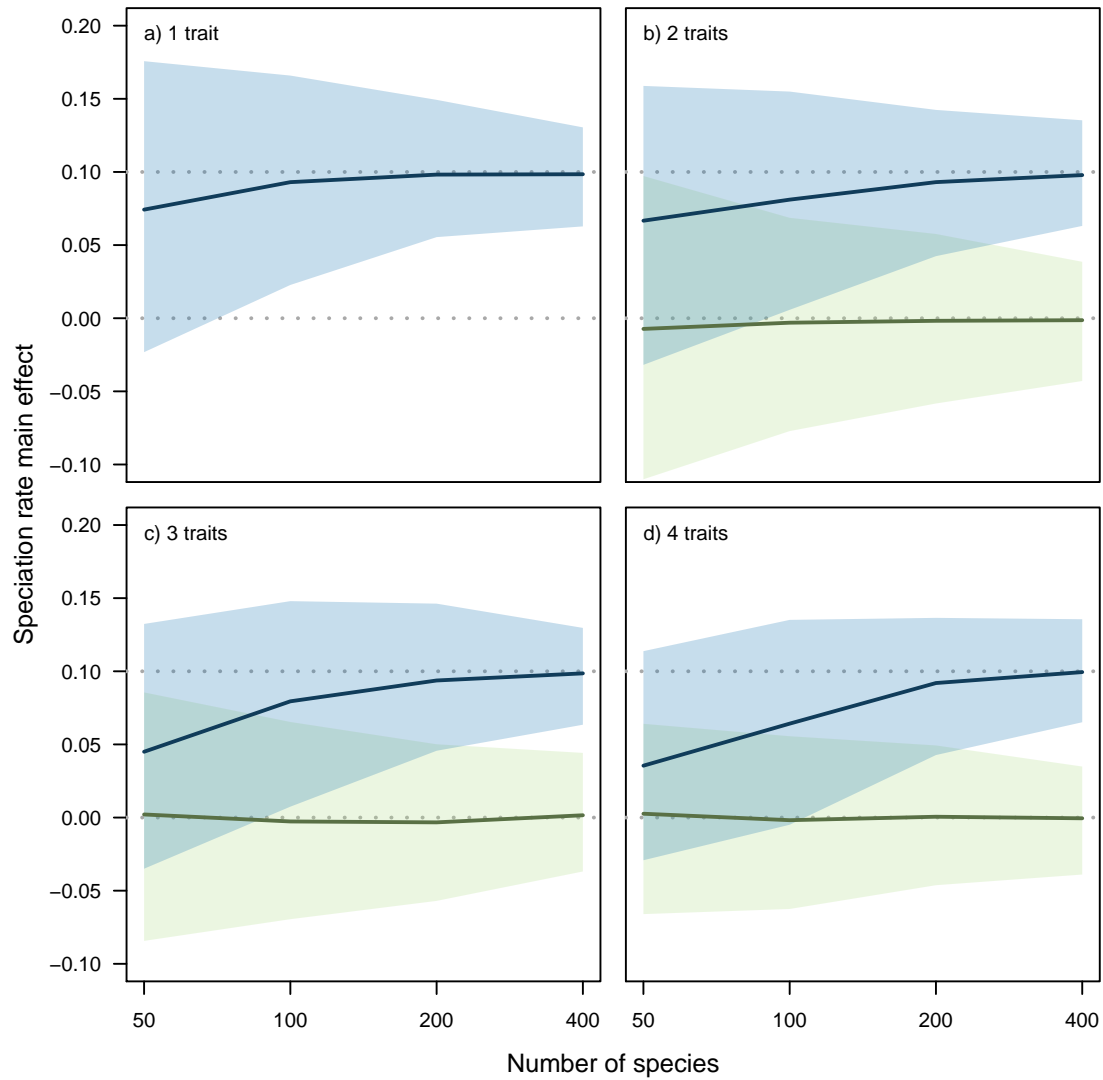


FIGURE 4.1: Uncertainty around multitrait MuSSE parameter estimates as a function of tree size and number of traits. The solid blue line and blue region represent the mean and 95% credibility interval (CI) over 100 trees for the estimated speciation rate main effect of trait A, which increases speciation rates (true value is 0.1, indicated by the grey dotted line). The solid green line and region represent the mean and 95% CI for the speciation rate main effect for trait B, which has no effect on speciation rates (true value of zero indicated by dotted grey line). Panel (a), with one trait, is equivalent to BiSSE.

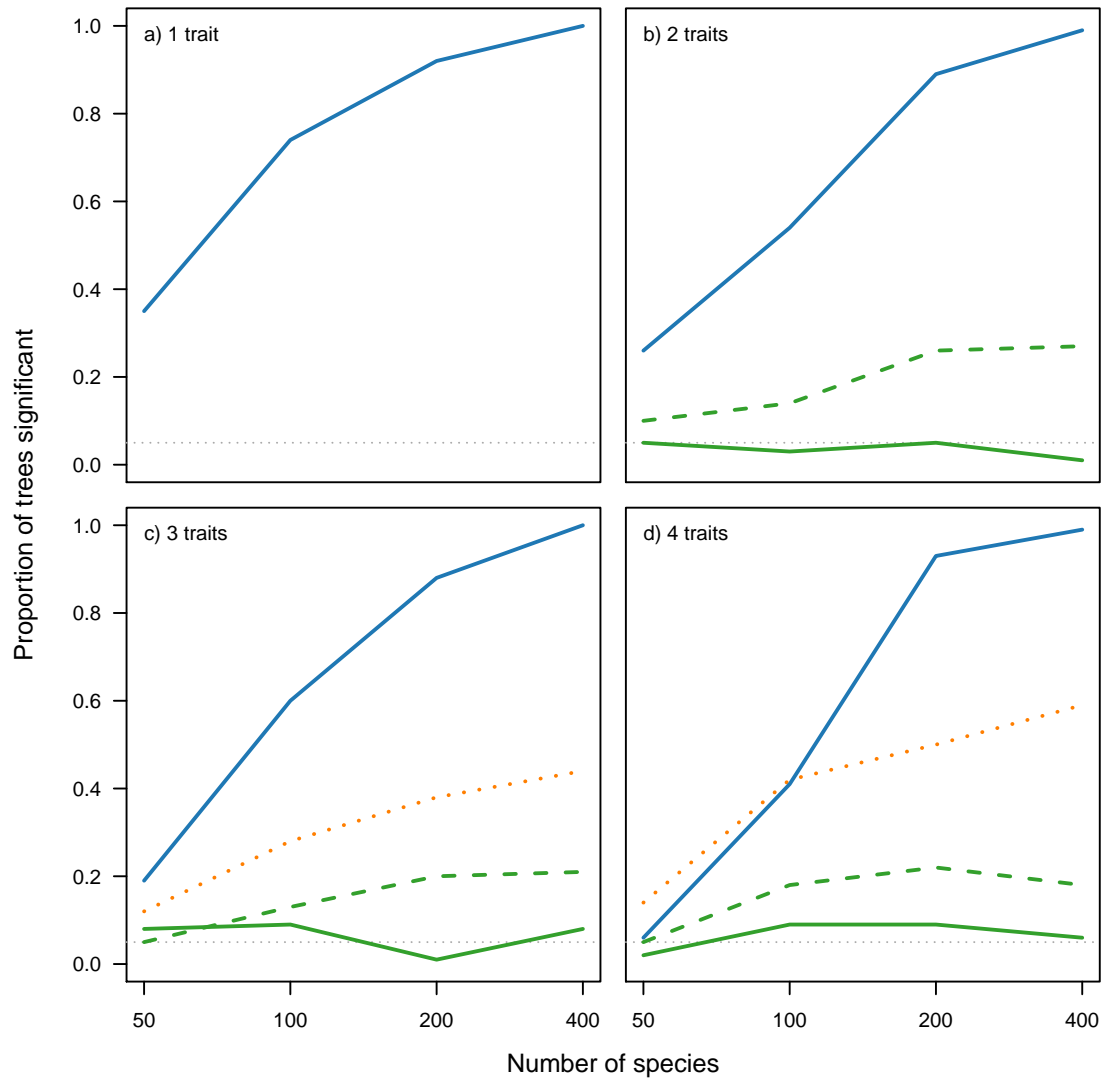


FIGURE 4.2: Power and error rates of multitrait MuSSE, as a function of tree size. The lines are the proportion of 100 simulated trees that have 95% credibility intervals of speciation main effects that do not include zero (indicating significant state-dependent speciation). The blue line represents trait A, which increases speciation rates when in state 1. The solid green line represents a trait B with no effect on speciation. The dashed green line indicates the same trait B, but when trait A is omitted from the analysis. The dotted orange line in panels (c) and (d) is the probability of finding *any* of the dummy traits (B, C, or, where present D) significant in an analysis that omits trait A. The 5% expected type I error rate is indicated by the dotted grey line. Panel (a), with one trait, and the dashed green line in panel (b) are equivalent to BiSSE.

These results are simultaneously encouraging and sobering. When a trait that is affects speciation is included in the model, it is easily detected, and this is robust to the number of additional traits included. However, if no traits do affect speciation, as we add additional traits we risk false positives at an alarming rate. However, the rates of false positives are perhaps not surprising. The trees used do not conform well to the expectations of a constant rate birth-death tree (there is strong phylogenetically structured variation in speciation rates) and the model is using the only parameters it has to explain this deviation. I expect that similar problems will affect other comparative analyses such as detecting correlated trait evolution with the Mk/discrete models.

The code for this analysis is available on the diversitree github site<sup>2</sup>.

#### 4.6.1 *Social evolution and speciation in primates*

Here I give a worked example, using the trait data compiled by Redding *et al.* (2010) to look at social evolution in primates. Previously, Magnuson-Ford and Otto (2012) found that both monogamy and solitary behaviour in primates reduced speciation rates, though this was only marginally significant for solitariness. However, if these characters are correlated, then it is possible that the decreased speciation rates could be truly associated with just one trait. That is, the effect of one character might bias the estimated effects of the other when these are treated independently. Alternatively, it could be that an elevated (or decreased) speciation rate occurs only with some combination of trait states (e.g., only social, polygamous taxa speciate more rapidly).

Here, I illustrate the method with R input in italics (preceded by “>”), while output is upright. The full version of this analysis is presented in Appendix C.3. The phylogeny is stored in NEXUS format (Maddison *et al.*, 1997), and loaded using the `read.nexus` function in `ape` as the object “tree”. For multi-trait MuSSE, the data must be stored in a data frame with species names as row labels. The two traits are “M” (TRUE for monogamous, FALSE otherwise) and “S” (TRUE for solitary, FALSE otherwise).

---

<sup>2</sup><http://github.com/richfitz/diversitree/tree/pub/simulations>

```
> head(dat)
```

	M	S
Allenopithecus_nigroviridis	NA	FALSE
Allocebus_trichotis	TRUE	TRUE
Alouatta_belzebul	NA	FALSE
Alouatta_caraya	NA	FALSE
Alouatta_coibensis	FALSE	FALSE
Alouatta_fusca	NA	FALSE

Note that some of the species lack state information (i.e., have NA values). These are accommodated using the method described above.

The first step is to make a likelihood function with `make.musse.multitrait`. The “depth” argument controls the number of terms to include from equations (4.3), (4.4), and (4.5); 0 includes only intercepts, 1 also includes main effects, 2 includes interactions between two parameters and so on. If specified as a 3-element vector, the elements apply to the  $\lambda$ ,  $\mu$ , and  $q$  parameters; if a scalar is given, the same depth is used for all three parameter types. To make a model with intercepts only:

```
> lik.0 <- make.musse.multitrait(tree, dat, depth=0)
```

This likelihood function takes a vector of parameters as its first argument. To get the vector of names for the parameters, use the `argnames` function:

```
> argnames(lik.0)
[1] "lambda0" "mu0"      "qM01.0"  "qM10.0"  "qS01.0"  "qS10.0"
```

This shows the six parameters: the speciation rate (`lambda0`), extinction rate (`mu0`) and four transition rates (e.g., `qM01.0` is the rate of transition of the breeding system from non-monogamous to monogamous, and this rate does not depend on the social state *S*).

To find the maximum likelihood (ML) point, a sensible starting point must be supplied (discussed in Appendix C.3); with such a point, `p.0`, we can find the ML parameters using the `find.mle` function:

```
> fit.0 <- find.mle(lik.0, p.0)
```

This returns an object (`fit.0`) that contains estimated parameters, likelihood values, and other information about the fit (see the help page `?find.mle` for more information).

```
> round(coef(fit.0), 4)
lambda0      mu0  qM01.0  qM10.0  qS01.0  qS10.0
  0.1912  0.1110  0.0251  0.0259  0.0009  0.0163
> fit.0$lnLik
[1] -786.3427
```

By default “subplex” (Rowan, 1990) is used for the optimisation. However, different optimisation algorithms can be selected through the “method” argument to `find.mle`.

To include state-dependent diversification, we construct a likelihood function that includes “main effects” of the two traits on speciation and extinction. To allow this while retaining the independent model of character evolution, we change the `depth` argument:

```
> lik.1 <- make.musse.multitrait(tree, dat, depth=c(1, 1, 0))
> argnames(lik.1)
[1] "lambda0" "lambdaM" "lambdaS" "mu0"      "muM"      "muS"
[7] "qM01.0"  "qM10.0"  "qS01.0"  "qS10.0"
```

Running an ML search from a suitable point `p.1`:

```
> fit.1 <- find.mle(lik.1, p.1)
```

These models can be compared using a likelihood ratio tests using the `anova` function; the model with state-dependent speciation and extinction fits much better than the state-independent version ( $\chi^2_4 = 24.7$ ,  $p < 0.001$ ).

```
> anova(fit.1, noSDD=fit.0)
      Df  lnLik   AIC  ChiSq Pr(>|Chi|)
full  10 -773.97 1568.0
noSDD  6 -786.34 1584.7 24.739  5.677e-05
```

(The use of `anova` for general model comparison is a fairly widespread convention in R packages, and does not imply that an ANOVA was performed!)

We can expand the model further to allow interactions between the two traits in speciation and extinction; is a combination of mating system and sociality associated with elevated speciation or extinction? Specifying `depth=c(2, 2, 0)` introduces the terms “`lambda.MS`” and “`mu.MS`” (see equations 4.3 and 4.4) to model non-additive effects of these traits on speciation and extinction, and again leaves character transitions to occur independently for the two traits.

```
> lik.2 <- make.musse.multitrait(tree, dat, depth=c(2, 2, 0))
> fit.2 <- find.mle(lik.2, p.2)
> anova(fit.2, noInteraction=fit.1)
```

	Df	lnLik	AIC	ChiSq	Pr(> Chi )
full	12	-773.73	1571.5		
noInteraction	10	-773.97	1568.0	0.49143	0.7821

This time the improvement is not significant, implying that there is no evidence for an interaction between these traits on speciation and extinction rates.

To test the significance of the each trait (solitariness and monogamy) in a maximum likelihood framework, we could fit models where the main effect of each trait was set to zero and compare these against the model `fit.1` using a likelihood ratio test. This approach is explored in Appendix C.3. Alternatively, we might run an MCMC and examine the posterior distributions of the `lambdaM` and `lambdaS` values:

```
> samples <- mcmc(lik.1, p.1, nsteps=10000, w=0.5, prior=prior)
```

The prior distribution used here is exponential with respect to the underlying rates in the model (e.g.,  $\lambda_{i,j}$ , not  $\lambda_{AB}$ : see equation (4.3) and Appendix C.3), but any prior function may be specified by the user (see the main `diversitree` tutorial). The “slice sampling” MCMC algorithm (Neal, 2003) is used by default and is fairly insensitive to tuning parameters. In particular, specifying a too large or too small value for the width of the proposal step (`w`) just increases the mean number of function evaluations per step, rather than the rate of mixing of the chain.

The marginal distributions of both the monogamy and sociality main effects on speciation rates are negative over the bulk of their distribution (figure 4.3). However, in

contrast with treating the traits separately using BiSSE (figure 4.3 a), we find that the 95% credibility intervals for both traits do not include zero (figure 4.3 b). Therefore these results support the conclusions of Magnuson-Ford and Otto (2012) that both monogamy and sociality are associated with decreased speciation rates in primates. Surprisingly, simultaneously accounting for both traits increased our confidence levels, suggesting that incorporating additional traits can reduce noise caused by shifts in diversification due to other traits.

More comprehensive examples are included in a tutorial document on the diversitree website, <http://www.zoology.ubc.ca/prog/diversitree>, as well as within the on-line help for the package.

## 4.7 CLOSING COMMENTS

The diversitree package implements several methods for jointly modelling character evolution and speciation. The package is open source and designed to be fairly straightforward to extend. In particular, any model that can be expressed by moving down a tree (post-order traversal, or “pruning”; Felsenstein, 1981) can be implemented using only a modest number of lines of R code. To facilitate the development of related methods, there is a “Writing diversitree extensions” manual available from the diversitree website. Stable versions of diversitree are available on CRAN (the Comprehensive R Archive Network) and from the website above. Development can be followed or joined on github (<http://github.com/richfitz/diversitree>).

I hope that the package will enable users to test a wide variety of macroevolutionary questions. However, I will close with a caution. All included methods are correlative only (Maddison *et al.*, 2007; Losos, 2011); they can merely show a statistical association between traits and speciation or extinction rates and cannot prove that the trait does affect speciation or extinction. Any unconsidered trait that is correlated with the target trait could be causal (Maddison *et al.*, 2007, and figure 4.2). Alternatively, the associations

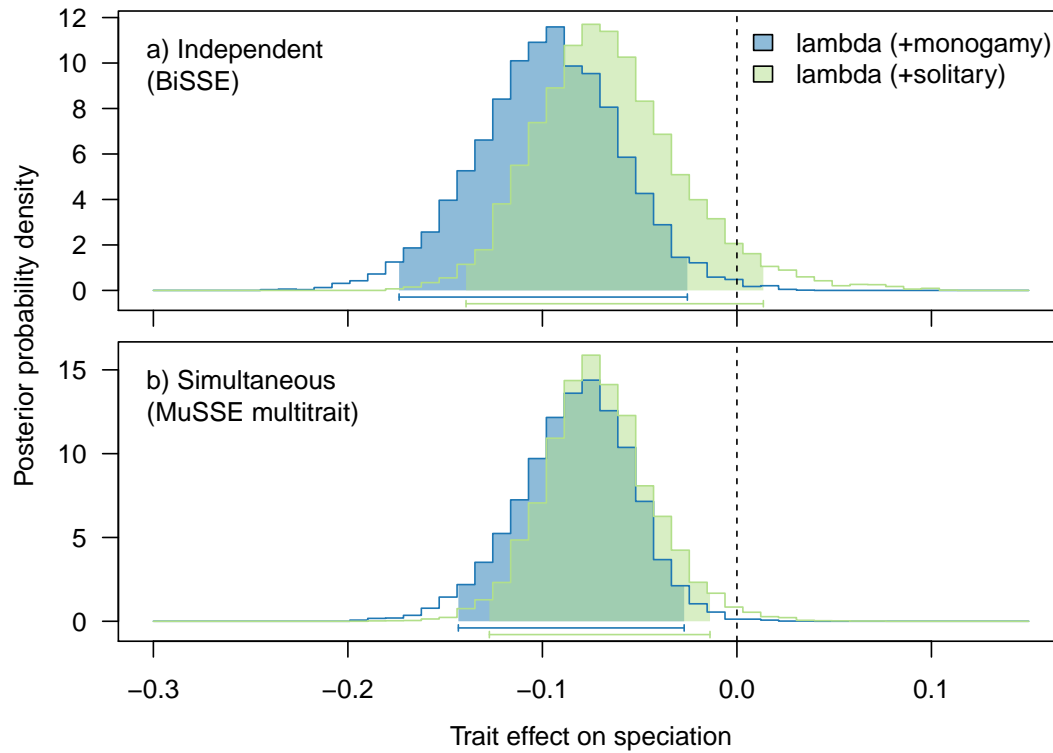


FIGURE 4.3: Posterior probability distributions for the effects of monogamy (dark grey) and solitariness (light grey) on speciation rate. Shaded areas and bars indicate the 95% credibility intervals for each parameter. In the top panel, BiSSE was run on each character independently. In the bottom panel, the `musse.multitrait` fit the effects of both traits simultaneously. In both cases, the MCMC chain was run for 10,000 steps, and the first 500 points were dropped as burn-in.



may be spurious, perhaps driven by departures from the assumed model of cladogenesis or character evolution. There is currently no way of testing absolute goodness-of-fit with any method, and all conclusions should be recognised as being conditional on a particular model, and on that model being appropriate.

## CHAPTER 5

# SURVIVAL OF THE LITTLEST? SPECIES SELECTION, COPE’S RULE, & MAMMAL BODY SIZE

---

### 5.1 INTRODUCTION

Cope’s rule states that body size tends to increase over evolutionary time (Cope, 1887, 1896)<sup>3</sup>. While palaeontological evidence appears to support this pattern in mammals (Alroy, 1998; Valkenburgh *et al.*, 2004; Raia *et al.*, 2012), most species are relatively small; the modal body size of mammals is approximately 100 g and the distribution is skewed to the right (figure 5.1, Jones *et al.*, 2009). This relative overabundance of small species in mammals (and other groups) has invited explanations including an optimal body size (Brown and Maurer, 1989), the effect of the perceived size of the environment (Morse *et al.*, 1985), and decreased rates of diversification of larger species (Stanley, 1973; Brown, 1995). This last hypothesis is the focus of this chapter: that species selection for small body size opposes a general tendency for phyletic size increase, a combination of processes that we will refer to as the “Cope–Stanley hypothesis”.

As phylogenies contain information about the timing and pattern of both morphological and taxonomic diversification, they have been widely used to test for size-dependent speciation. However, analyses of individual mammal clades have generally found mixed results; while some have shown the expected negative relationship between body size and diversification rates, others recovered no pattern or even a positive relationship (Gardezi and da Silva, 1999; Gittleman and Purvis, 1998; Isaac *et al.*, 2003; Paradis, 2005;

---

<sup>3</sup>Cope never coined the phrase “Cope’s rule”, but these two papers are commonly cited. Some argue that the rule is implicit in Cope’s writing (e.g., Stanley, 1973; Benton, 2002), though others dispute even this (Polly, 1998). The phrase was apparently coined by Rensch (1948), and has been widely used since.

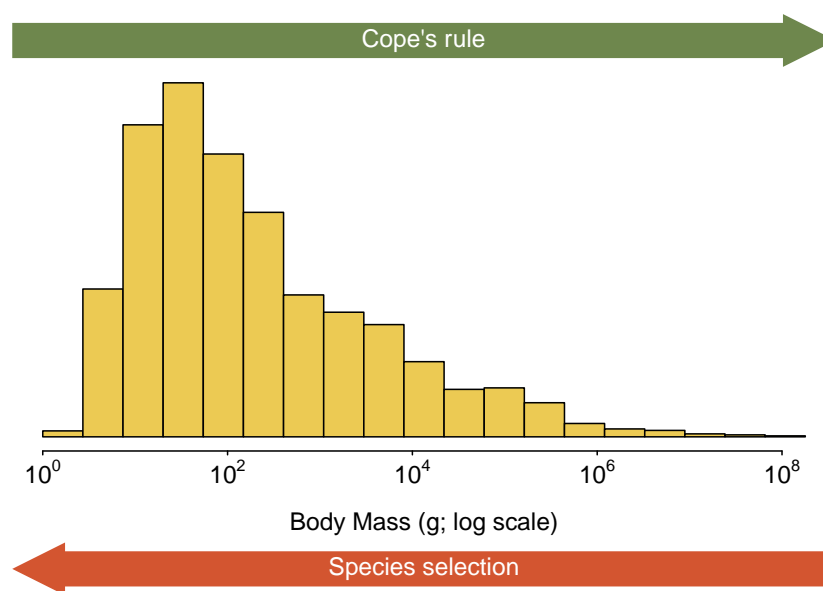


FIGURE 5.1: The distribution of mammal body masses. The Cope–Stanley hypothesis states that species mass tends to increase over time (Cope’s rule), balanced by species selection against large species. Data from PanTHERIA (Jones *et al.*, 2009).

FitzJohn, 2010). Different groups have been studied with different methods, making generalisation to all species difficult. Recently, Clauset and Erwin (2008) argued that the distribution of body sizes across all mammals is consistent with the Cope–Stanley hypothesis, and Etienne *et al.* (2012a) argued that such a relationship holds across multiple animal phyla.

Here, we test the Cope–Stanley hypothesis across all mammal species using an explicitly phylogenetic Bayesian approach. We modelled speciation rates as linear functions of log female body mass, which we assumed evolves according to a Brownian motion process, using QuaSSE (Quantitative State Speciation and Extinction; FitzJohn, 2010). The possibility that body size tends to increase (or decrease) within lineages was incorporated by allowing for directional mass change, in addition to the diffusion present in the Brownian motion model. To avoid over-parameterisation, we held extinction rates constant across body sizes. This combination of processes allows us to simultaneously test for an among-lineage disadvantage of large size with respect to diversification, and for a within-lineage tendency for body size to increase (Cope’s rule). We selected 10 major clades that include 98% of known extant mammal species (figure 5.2, table 5.1) from a recent supertree of mammals (Bininda-Emonds *et al.*, 2007; Fritz *et al.*, 2009) or from clade-specific phylogenies (Hernández Fernández and Vrba, 2005; Vos and Mooers, 2006; Steeman *et al.*, 2009), and obtained mammal body mass estimates from the PanTHERIA database (Jones *et al.*, 2009). Markov chain Monte Carlo (MCMC) was used to sample from the posterior distribution of the model (see section 5.3).

## 5.2 RESULTS & DISCUSSION

Across the clades, we found no consistent relationship between speciation rates and body size. In four clades (Afrotheria, Cetacea, Eulipotyphla, and Rodentia) we found that speciation rates were negatively correlated with body size, consistent with species selection against large body size. However, four clades showed the opposite pattern

TABLE 5.1: The 10 mammal clades used to test the Cope-Stanley hypothesis. Species were excluded either because of a lack of data on mass or because they were not present in the phylogeny. However, they are accounted for in the analysis by assuming that such species were randomly omitted.

Name	Taxonomic level	Species	Included	Excluded
Afrotheria	Superorder	79	63	16
Carnivora	Order	286	250	36
Cetacea	Order	89	87	2
Chiroptera	Order	1116	924	192
Eulipotyphla	Order	452	206	246
Lagomorpha	Order	92	74	18
Marsupials	Infraclass	331	270	61
Primates	Order	233	233	0
Rodentia	Order	2277	1442	835
Ruminantia	Sub order	197	196	1

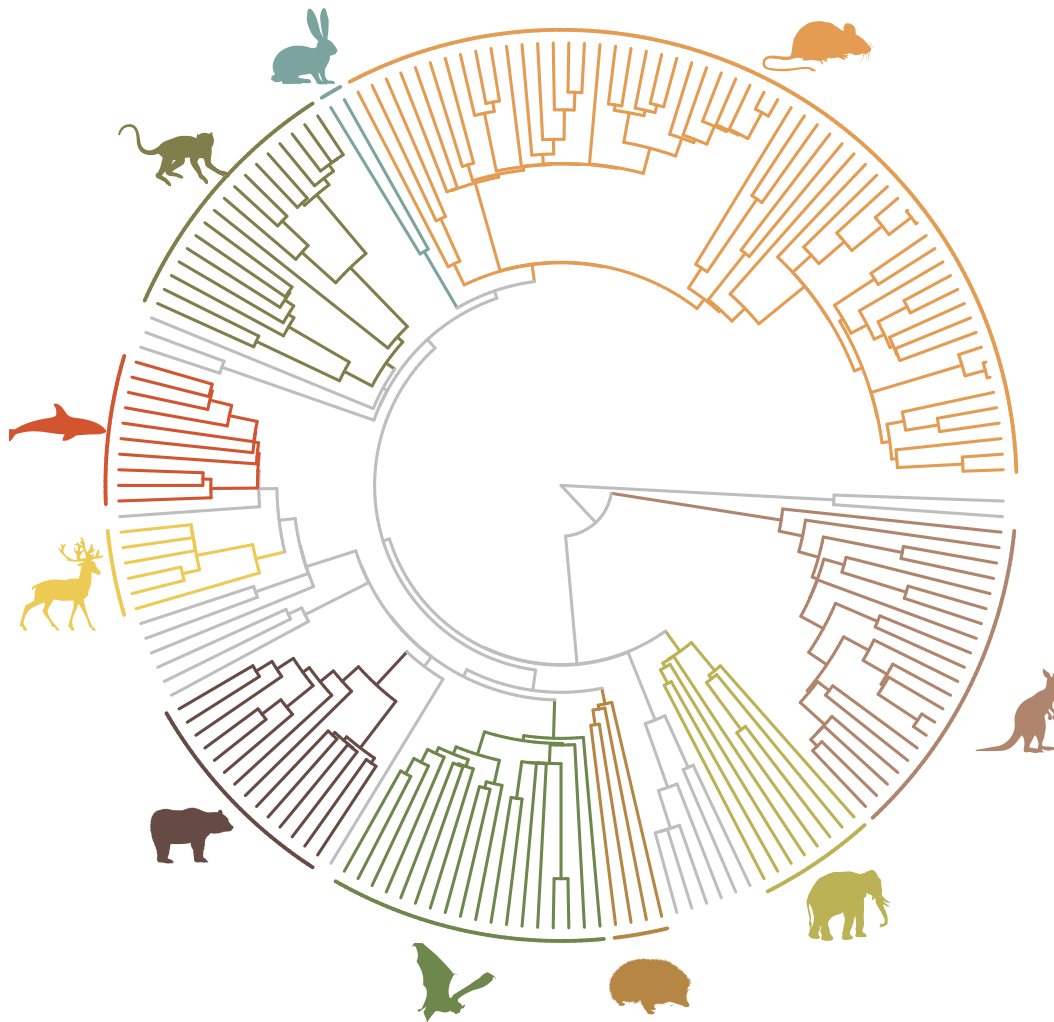


FIGURE 5.2: Mammal supertree showing the 10 focal clades. Each terminal branch represents a family. The grey branches indicate groups that were not included in the analysis. Clockwise from the right, included clades are Marsupials, Afrotheria, Eulipotyphla, Chiroptera, Carnivora, Ruminantia, Cetacea, Primates, Lagomorpha, Rodentia.

(Chiroptera, Lagomorpha, Marsupials, and Primates), and two clades (Carnivora and Ruminantia) showed no significant pattern (figure 5.3). Consistent with Cope's rule, there was a general tendency for body size to increase within lineages (7/10 clades had a positive mean trend parameter). Nevertheless, this positive trend was only significant in Cetacea, and Lagomorpha had a significant negative trend (figure 5.3). These results argue against the generality of the Cope–Stanley hypothesis and suggest that body size is not a consistent decelerator of speciation rates. Indeed, across clades, being larger was just as likely to accelerate diversification as decelerate it.

That a model can be fit to data does not prove that the model is correct. There is considerable variation among clades in both the slope and mean of the speciation rate/body size relationship, which argues against a single common model (figure 5.3). However, similar rate heterogeneity may exist *within* each clade. In particular, it may be that evolutionary transitions in some unknown trait along certain branches of the tree of life have altered diversification rates and led to a spurious correlation between body size and diversification. For example, the dolphin family (Delphinidae) is a relatively recent but rapid radiation containing 40% of cetacean species (34 of 84 species), including 70% of cetacean species shorter than 4 m in length. Consequently, we might falsely attribute variation in diversification rate within the Cetacea to body size when these differences are in fact due to some other shared feature of the Delphinidae.

To determine if the observed associations between body size and diversification have been driven by major shifts in diversification caused by other factors, we divided each clade into regions that appeared to have different diversification rates using MEDUSA (Alfaro *et al.*, 2009); we then allowed speciation rate/body mass relationships to vary among these these partitions. When analysed in this way, only two partitions showed the expected negative relationship between body size and speciation rate (paraphyletic basal groups in Afrotheria and Cetacea; figure 5.4), while three clades showed a significant positive relationship (Dasyuididae and Macropodidae within Marsupials, and a clade that included Nataloidea and Noctilionoidea within Chiroptera). The evidence

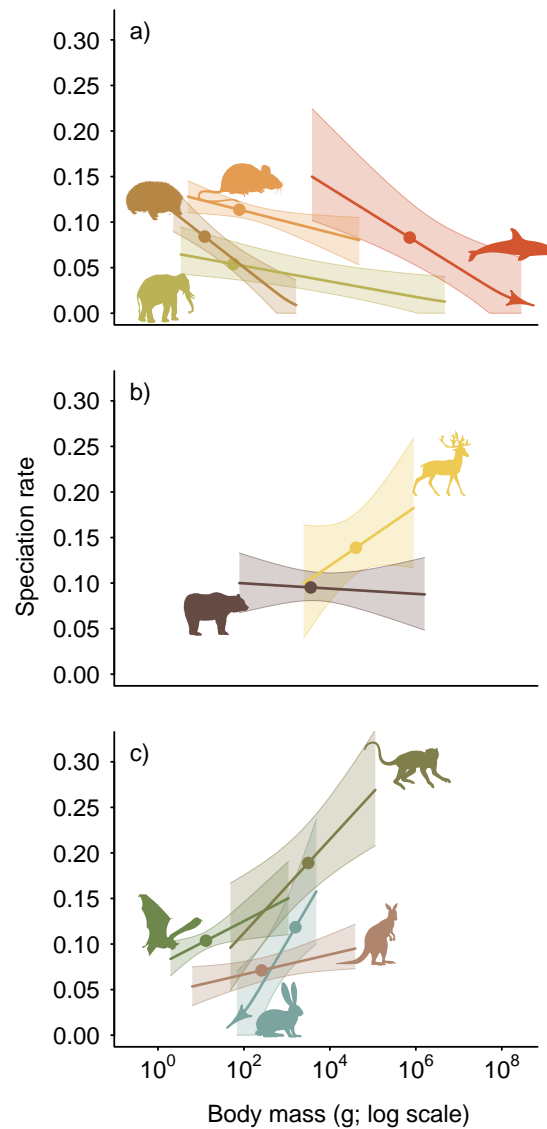


FIGURE 5.3: The inferred relationship between body size and speciation rate for 10 mammal clades. The thick lines represent the mean relationship, and the lighter envelopes the 95% credibility interval around these, estimated using MCMC. Arrows indicate the mean direction of the trend body size evolution when significantly different to zero (right arrows indicate an increase, left arrows indicate decrease). Points indicate the median body size within each clade and the mean estimated speciation rate at that body size. The three panels separate cases where the slope was a) significantly negative (Afrotheria, Cetacea, Eulipotyphla, and Rodentia), b) not significantly different from zero (Carnivora and Ruminantia), and c) significantly positive (Chiroptera, Lagomorpha, Marsupials, and Primates).



for directional change toward larger body size remained significant in Cetacea, and there was evidence for a negative body size trend within Marsupials.

Our results suggest that there is no single overarching pattern of size-dependent speciation in mammals. We found significant relationships between body size and speciation rates in most clades (figure 5.3); the problem is not that we cannot detect associations, but that the nature of the associations that we detected differs among groups. In contrast to detecting differences in speciation rates, detecting directional trends in quantitative characters is notoriously difficult (it is impossible under a simple Brownian motion model without fossil data), and the power of QuaSSE to detect such trends is low (FitzJohn, 2010). Thus, the lack of positive trends here does not preclude Cope's law operating, especially given the considerable palaeontological support (e.g., Alroy, 1998; Valkenburgh *et al.*, 2004; Raia *et al.*, 2012).

Our results support the idea that clade-specific differences in speciation rate are more important than body-size related differences in determining mammal diversity. It is possible that major shifts in body size have occurred along the branches separating the clades in our analyses, so that analysing the clades separately reduces the power to detect any association with diversification. That is, the clades that are speciating more rapidly also tend to contain smaller species (such as Delphinidae within Cetacea). This would be consistent with purely taxonomic analyses showing that the most diverse clades (e.g., Chiroptera and Rodentia) tend to be relatively smaller in size (though perhaps not the smallest; Dial and Marzluff, 1988; Gardezi and da Silva, 1999). However, we found no clear relationship between speciation rate at the median body size and median body size, either across our 10 major clades (figure 5.3), or across the partitions identified by MEDUSA (figure 5.5 and 5.6). For example, the group with the highest speciation rate within the Chiroptera, *Rhinolophus*, has a median body size that is similar to the median over the whole order.

Most of the phylogenetic information used by QuaSSE comes from the recent branches, as about half the branches in a tree lead to single extant species. Significant relation-

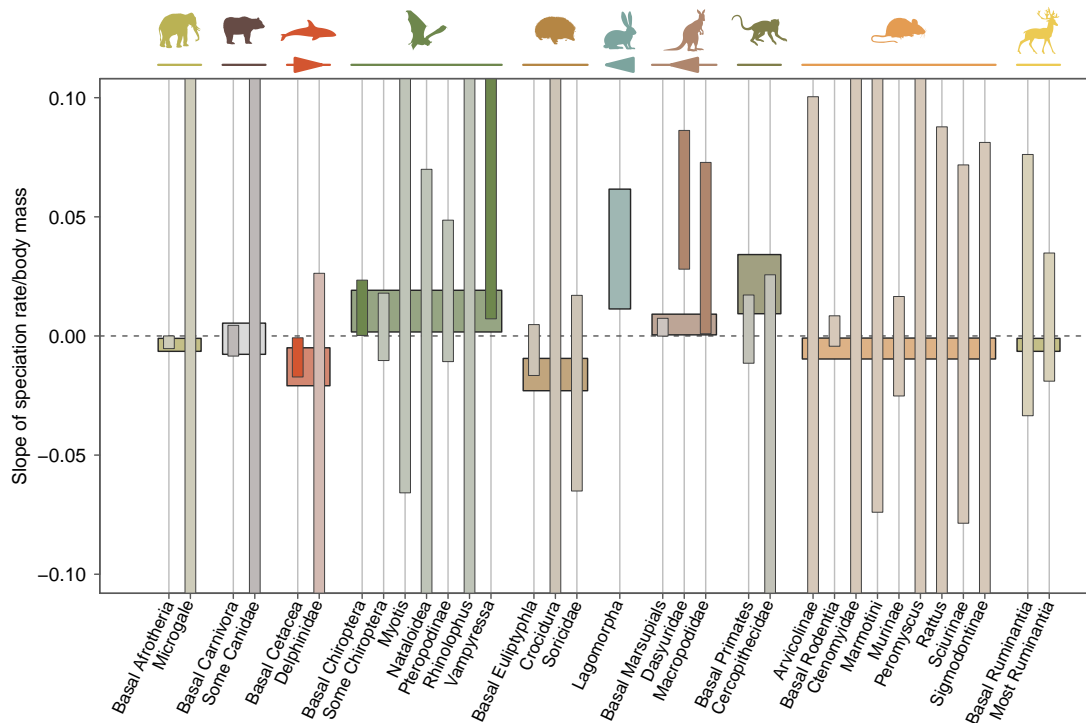


FIGURE 5.4: Slopes of speciation rate/body mass relationships within MEDUSA-derived partitions (see table 5.2 and Appendix D for definitions). Each bar represents the 95% credibility interval for the slope of the speciation rate/body size relationship within a partition, and arrows above the bars indicate the direction of the trend parameter when significantly different from zero. For contrast, the single-slope credibility interval is displayed behind the bars. Dark shading represents cases where the slope was significantly different from zero; that most bars have light shading and are as likely to be negative as positive indicates little support for the Cope–Stanley hypothesis. For Chiroptera and Rodentia, only partitions that were recovered in at least 10% of trees are displayed.

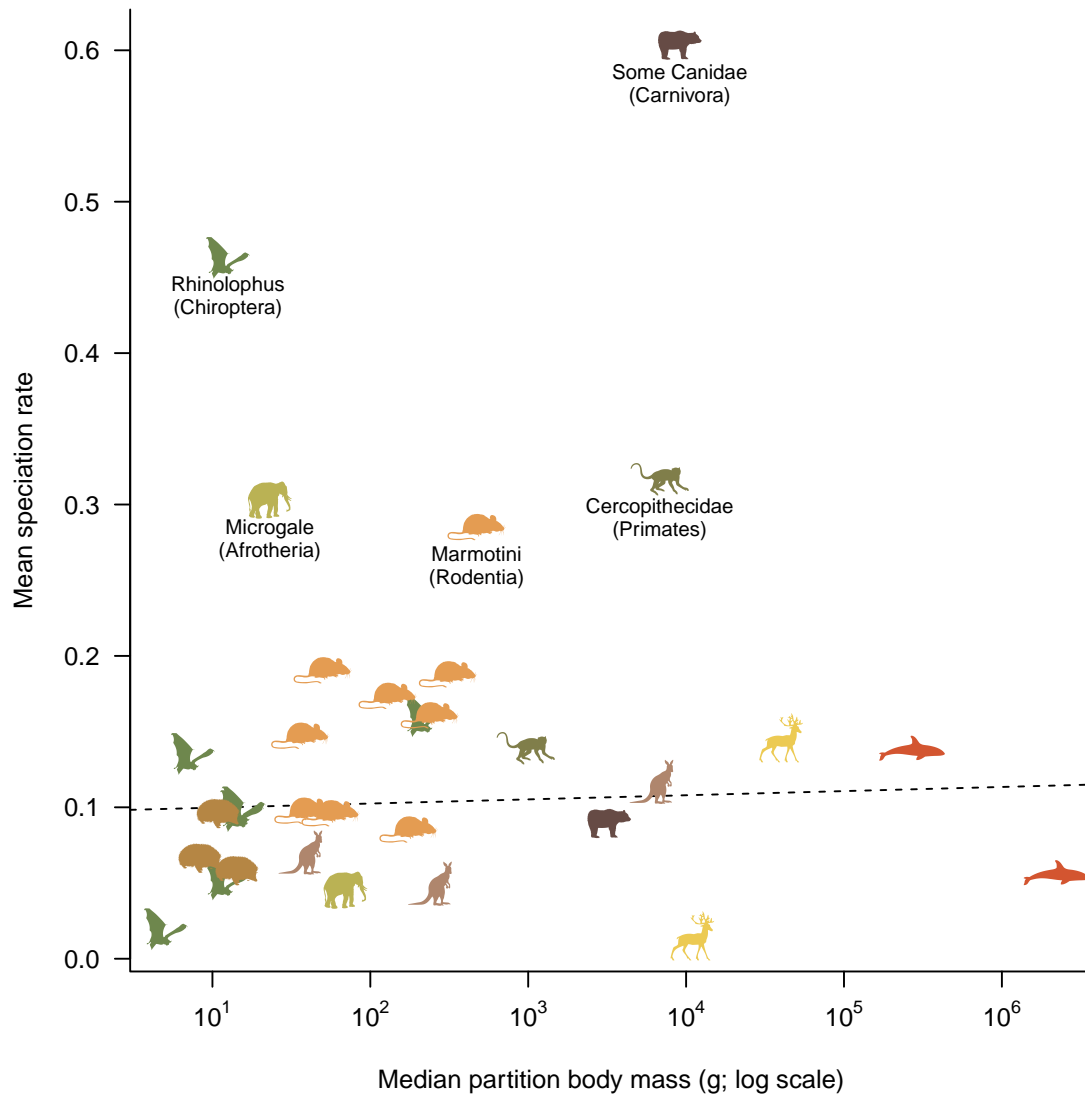


FIGURE 5.5: Relationship between speciation rate and body size within mammals, across partitions with different diversification rates identified by MEDUSA. Each point is the estimated speciation rate at the median body size, integrated over the posterior distribution of the model (see figure 5.3). The dashed line indicates a linear regression through these points, excluding the five partitions with the highest mean speciation rate ( $r^2 = 0.006$ ,  $p = 0.7$ ); the regression was also not significant with these points included. For Chiroptera and Rodentia, only partitions that were recovered in at least 10% of trees are displayed.

ships between speciation rate and body size can be driven by recent nodes that lead to species concentrated at one end of the size distribution. As such, even if a fundamentally different model of cladogenesis is closer to the true model (e.g., the niche-filling model of Rabosky, 2009a), our approach would interpret size-specific differences in turnover as a speciation-body size relationship. The lack of a consistent pattern here indicates that a single general rule relating diversification and body size in mammals is unlikely to hold.

Explaining the great disparity in clade diversity remains a major thrust of evolutionary investigation. Therian mammals outnumber Monotremes by 1000 to 1, and over 40% of mammals are rodents; it is natural to suspect that traits within these clades account for some of these differences. Here, we have tested the long-standing hypothesis that variation in speciation rate driven by body size has led to the excess abundance of small species relative to large species. We found little support for this hypothesis, but instead infer a more idiosyncratic relationship between body size and diversification, with some clades exhibiting a positive and some a negative relationship.

### 5.3 METHODS

To investigate the relationship between body size and speciation rate we used the QuaSSE (Quantitative State Speciation and Extinction) method (FitzJohn, 2010) as implemented in the R package diversitree (Chapter 4) to fit size dependent birth-death models of speciation and extinction. Specifically, we modelled speciation rate as a linear function of body size, disallowing negative speciation rates (these were truncated to zero). A negative slope of the speciation rate–body size relationship captures the hypothesised long-term disadvantage of large body size on speciation rates, while requiring only two parameters. We modelled change in body size over time using Brownian motion with both a diffusion and a trend term, where the latter captures directional changes within a lineage (e.g., due to selection or mutational pressure). Thus, this method allows for the

possibility of detecting conflicts between individual selection (e.g., favouring a trend toward larger body size) and species selection (e.g., favouring higher diversification among small mammals).

For body size estimates, we used the PanTHERIA database (Jones *et al.*, 2009), expressed as log female mass in grams (for brevity, we will refer to this simply as “mass”). Direct estimates were available for 3,542 species, with another 392 determined by linear regression from body length estimates (see Jones *et al.*, 2009, for details). These were supplemented by estimates of cetacean length (unpubl. dat., Nick Pyensen). We converted cetacean log-lengths to log-weight estimates by predicting from linear regression among cetacean species only (figure 5.7).

We focused on 10 major clades of mammals (figure 5.2, table 5.1). This omitted 111 species (2% of all species) in 9 additional clades that we felt were too small to be statistically informative, the largest of which were the superorder Xenarthra (31 species) and the order Scandentia (20 species). For most clades, the tree was pruned from a recent mammal supertree (Bininda-Emonds *et al.*, 2007; Fritz *et al.*, 2009). For three clades we used trees constructed specifically for that clade: Cetacea (Steeman *et al.*, 2009), Primates (Vos and Mooers, 2006), and Ruminantia (Hernández Fernández and Vrba, 2005).

The mammal, Primate, and Ruminantia supertrees contained polytomies; before use with QuaSSE these were randomly resolved, and replaced with branch lengths simulated under a birth-death model (see Kuhn *et al.*, 2011 for details). This produced a family of plausible trees, from which we sampled 100 trees to account for uncertainty in polytomy resolution (note that this includes only a fraction of possible phylogenetic uncertainty). The polytomy resolution was performed without reference to body size, which sometimes created unrealistically short branches separating sister species that were very different in size, causing numerical instability in the integrations and preventing likelihood calculation. To avoid this, we dropped one species of any sister species pair that were “impossibly disparate” given their estimated divergence date. We arbitrarily defined this

as a log probability of less than  $-10$  (approximately 0.000045) under a Brownian motion model of character evolution with the maximum likelihood rate estimate (estimated ignoring speciation and extinction). In practice, this removed 10–30 of the 3,934 included species across the 10 clades, the exact taxonomic composition of which varied among the trees. For Cetacea, we sampled 100 trees from a distribution of time-calibrated trees obtained from a Bayesian tree search (Steeman *et al.*, 2009), which more fully accounts for phylogenetic uncertainty in all branches.

We used two different approaches to analyse the data: “single-slope” fits (one speciation rate/body size relationship per clade), and “multiple-slope” fits where different regions of each clade were allowed to have different relationships. For the single-slope analyses, we fit QuaSSE models to 10 major clades of mammals (table 5.1). Each model has five parameters: the slope and intercept of the speciation rate–body mass relationship, extinction rate, diffusion parameter, and trend parameter. Not all species were represented on the phylogeny, and some species lacked body mass data (table 5.1). We corrected for these missing data in the likelihood analysis by assuming that species missing from a clade were randomly excluded with respect to both phylogeny and mass (FitzJohn, 2010). This approach treats all the clades as fully independent data points, and no information was shared among them during the inference process.

To account for major shifts in diversification across the tree of mammals, we fit “multiple-slope” models where the speciation rate/body mass relationship was allowed to vary among partitions in a clade. We first used MEDUSA (Alfaro *et al.*, 2009) to identify regions within the 10 major clades exhibiting significant shifts in diversification rates. To avoid over-fitting, and because extinction rate estimates are sensitive to the exact distribution of branch lengths, we constrained the MEDUSA algorithm to use a single extinction rate over the entire clade, set to the maximum likelihood estimate from a constant rate birth-death model for that clade while using MEDUSA. Each partition adds two parameters: a location and a new speciation rate. To compensate for the very large number of comparisons made when fitting MEDUSA models (considering potential shifts

at every branch in each phylogeny), we used Bonferroni corrections. This correction increased the required log-likelihood improvement from 3.0 (assuming a  $\chi^2$  distribution with two degrees of freedom) to between 7.1 (Afrotheria, 62 nodes) and 10.0 (Rodentia, 1,425 nodes).

The smallest clades were rarely (Afrotheria) or never (Lagomorpha) partitioned (table 5.2). Most other clades were usually split into two or more partitions. The largest orders, Chiroptera and Rodentia, were split into many partitions that (in contrast to other orders) varied widely in presence among individual trees. This appears to reflect taxonomic uncertainty in the backbones of these clades. Many of the regions that MEDUSA fit as having diversification rates that differed from the rest of the tree were fairly recent, with most species included in a basal paraphyletic partition.

We used Markov chain Monte Carlo (MCMC) to sample from the posterior distribution for our models, using slice sampling (Neal, 2003) for the parameter updates. For the single-slope analysis, this model has five parameters: slope and intercept of the speciation rate function, extinction rate, Brownian motion trait diffusion coefficient, and a Brownian motion drift coefficient describing the rate of directional trait change (“trend”). For the multiple-slope analysis with  $n$  partitions, there are  $n$  slope and intercept parameters, in addition to a common extinction rate, diffusion coefficient, and trend coefficient. We had no strong prior beliefs about most parameters, so we used unbounded uniform priors for all model parameters. However, we used a uniform prior on the root state that spanned the range of known mammal sizes: 1.3 g (*Batodonoides vanhouteni*; Bloch *et al.* 1988) to 190 tonnes (*Balaenoptera musculus*). All chains were started from the maximum likelihood point, estimated using the subplex algorithm (Rowan, 1990). We ran the chains for 3,000 steps (single-slope analysis) or 6,000 steps (multiple-slope analysis), discarding the first 500 steps as burn-in. While this is a relatively small number of steps, the estimated effective sample size was over 1,000 for most parameters (Plummer *et al.*, 2006), and the total computational time for the analyses here exceeded 50 CPU years. We then pooled these sampled parameters over the 100 trees to produce posterior

distributions representing 250,000 or 550,000 MCMC steps that include both parameter and polytomy uncertainty (phylogenetic uncertainty in the case of the Cetacea).



TABLE 5.2: Properties of partitions recovered by MEDUSA, indicating the number of trees for which that partition was significant, and the median log-likelihood improvement ( $\Delta$  likelihood). The exact species composition varied among trees, and the name is indicative of composition only; see Appendix D.

Clade/partition name	Number of trees	$\Delta$ likelihood (range)
Afrotheria		
Not partitioned	97	
Microgale	3	7.9 (7.4, 9.3)
Carnivora		
Some Canidae	100	16.3 (14.1, 19.7)
Cetacea		
Not partitioned	64	
Delphinidae	36	8.7 (7.5, 15.2)
Chiroptera		
Some Chiroptera	99	30.1 (25.1, 35.1)
Myotis	83	13.0 (10.0, 31.3)
Nataloidea	41	13.6 (10.0, 15.0)
Pteropodinae	13	10.5 (10.0, 18.3)
Rhinolophus	100	48.3 (41.4, 60.5)
Vampyressa	10	17.5 (12.7, 18.9)
Other	2	
Eulipotyphla		
Crocridura	100	42.6 (36.6, 48.4)
Soricidae	21	9.6 (9.0, 11.4)
Lagomorpha		
Not partitioned	100	
Marsupials		
Dasyuridae	100	10.3 (9.5, 11.5)
Macropodidae	100	11.9 (10.6, 13.5)
Primates		
Cercopithecidae	100	11.6 (10.8, 12.4)
Rodentia		
Arvicolinae	28	13.2 (10.8, 16.7)
Ctenomyidae	26	11.9 (10.8, 14.3)
Marmotini	100	22.0 (20.2, 24.3)
Murinae	85	15.1 (10.8, 23.5)
Peromyscus	17	12.4 (10.9, 15.5)
Rattus	11	15.6 (12.2, 21.0)
Sciurinae	99	16.4 (11.1, 20.4)
Sigmodontinae	100	46.1 (39.0, 56.3)
Other	9	
Ruminantia		
Not partitioned	4	
Most Ruminantia	96	12.9 (10.7, 14.3)

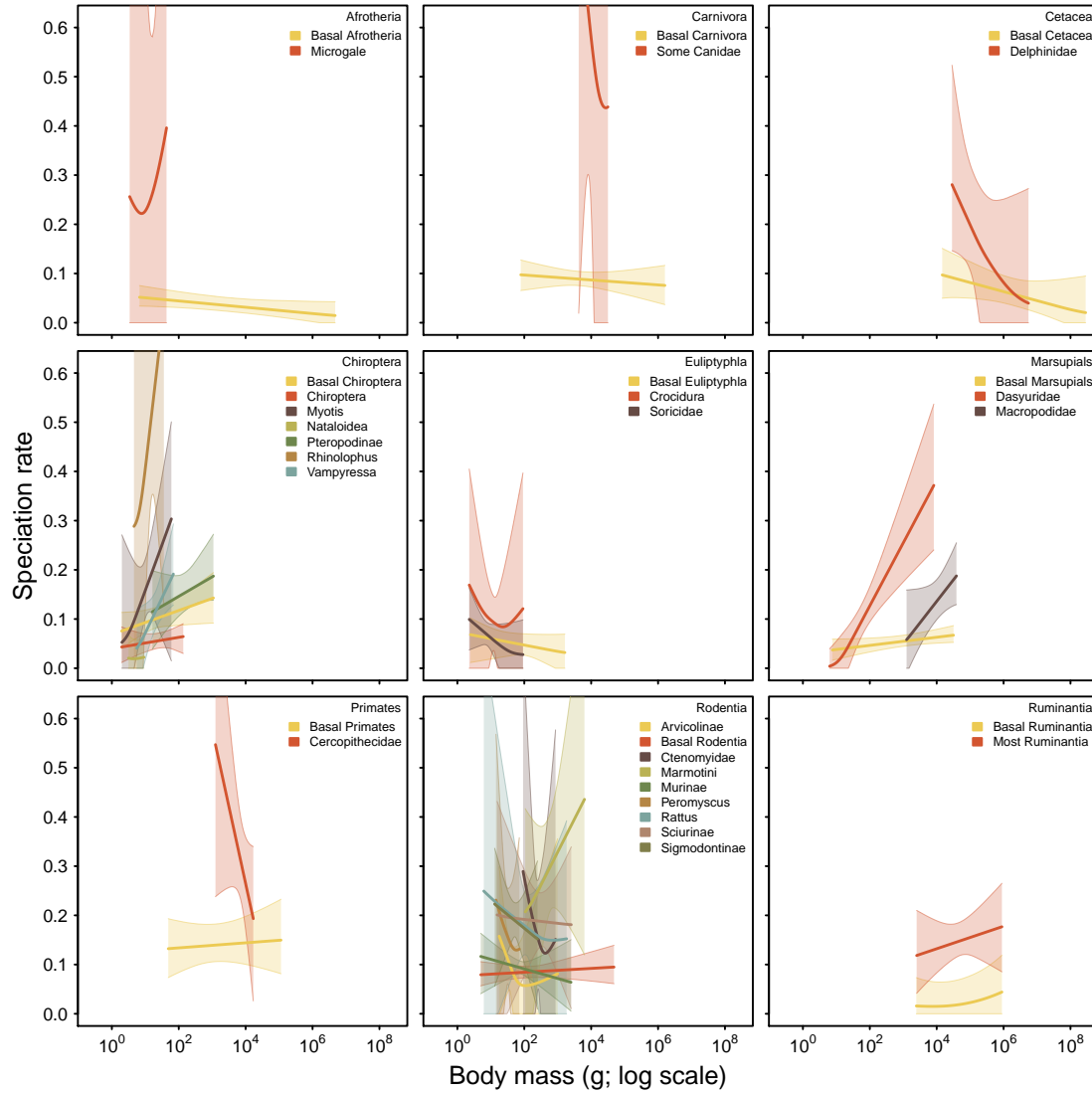


FIGURE 5.6: Speciation rate inferences for partitions. In each major group, QuaSSE was performed separately for the partitions that exhibited significantly different diversification rates, according to MEDUSA. No such partitions were inferred for Lagomorpha, so only the remaining 9 clades are shown. Each envelope represents the 95% credibility interval for the speciation rate/body size relationship within a partition. Thick lines indicate the mean relationship. For Chiroptera and Rodentia, only partitions that were recovered in at least 10% of trees are displayed.

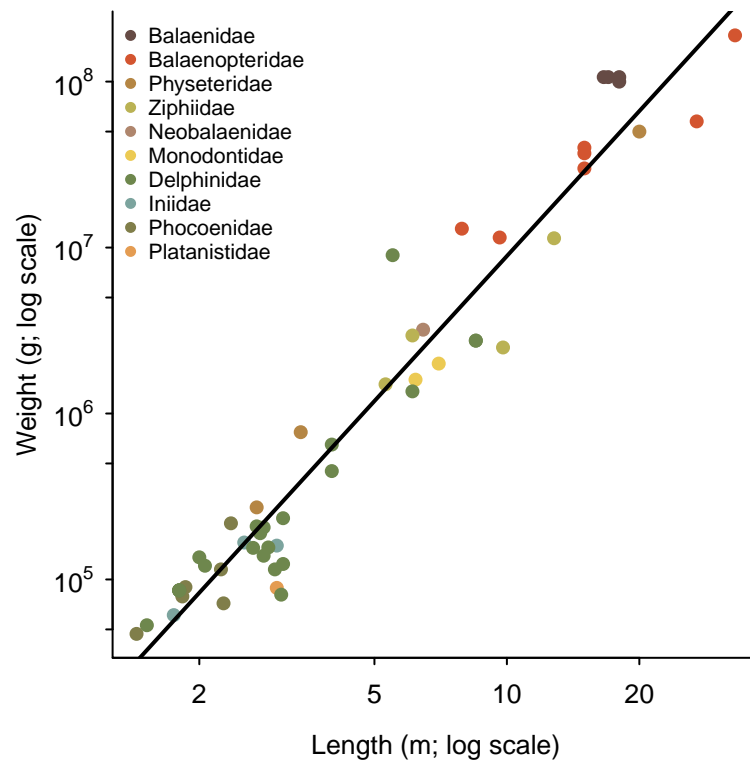


FIGURE 5.7: Regression of body mass against length for Cetacea ( $r^2 = 0.94$ ). Different colours indicate different families (sorted by decreasing mean mass in the legend). Because we perform a linear transformation between log-length and log-mass, the rank ordering of size estimates is unchanged. This transform rescales the parameter estimates for more direct comparison with the other clades.

## CHAPTER 6

### CONCLUSION

---

I believe that one of the reasons that the concept of species selection has gained favour among biologists is the evidence for trait-based differences in diversification that have accumulated over the last 25 years (Coyne and Orr, 2004; Jablonski, 2008b; Rabosky and McCune, 2009). In my thesis, I have developed and applied a number of methods for detecting species selection using phylogenies. The methods presented here improve on earlier approaches by allowing exploration of both the magnitude of species selection and the within-lineage processes that might oppose it.

In chapter 2, I developed a method for detecting the association between a binary trait and rate of speciation or extinction when a phylogeny is incompletely resolved. This built upon the BiSSE (Binary State Speciation and Extinction) method of Maddison *et al.* (2007). I then used this method to investigate the association between speciation rate and sexual dimorphism in shorebirds, finding that strong sexual dimorphism was associated with elevated rates of speciation, but may quickly revert back to monomorphism.

In chapter 3, I developed a method for inferring the association between a continuous trait and speciation or extinction rates, using a diffusion model for character evolution (QuaSSE: Quantitative State Speciation and Extinction). State-dependent diversification under such a model has been long discussed (e.g., Slatkin, 1981), but fitting this model to data has been only approximate prior to the development of this method. I used QuaSSE to investigate the association between body size and speciation and extinction in Primates, finding mixed evidence for any size-specific speciation or extinction.

In chapter 4, I described an R package, “diversitree”, that I developed to facilitate phylogenetic analyses. This package contains and documents the methods above, and

also methods that ignore trait evolution when studying diversification (as in Nee *et al.*, 1994b) or model character evolution independently of speciation and extinction (e.g., Pagel, 1994). While this is a small chapter, developing diversitree has been a large component of my thesis. The package has been widely adopted by the comparative phylogenetic community; it has been used in 34 published studies since May 2009 and has over 100 users (based on email contact). In this chapter I also introduced “MuSSE” — a multi-state extension of BiSSE — and discuss how this can be applied to combinations of traits to simultaneously account for the effects of multiple traits on speciation or extinction using an approach akin to a general linear model regression. Importantly, this allows for quantifying the association between one trait and diversification while accounting for other possible traits. I reanalysed an investigation of social and mating structure in Primates originally discussed by Magnuson-Ford and Otto (2012). Surprisingly, when taking into account sociality, the effect of mating system (monogamy vs. polygamy) on speciation became more apparent.

In chapter 5, I carried out an analysis of body size evolution in mammals, aiming to address the long-standing question of why there are so many small mammal species (and so few large) by testing the hypothesis that small mammals tend to diversify more rapidly than large species. I found that there was no consistent relationship between body size and speciation rate, with some clades having a negative relationship and others having a positive relationship. I found limited evidence for Cope’s law (the tendency for species to get larger over time), which was only well supported in the Cetacea. Furthermore, I found that within clades there was often large variation in rates of speciation that was not directly related to body size (figure 5.6).

Here, I conclude with some general comments about the approaches used in my thesis and their limitations. I will consider some lines of evidence challenging the sort of methods described in my thesis, focusing on inconsistencies between inference from phylogenies and inference from the fossil record. I consider modifications to phyloge-

netic methods that have been suggested to address their shortcomings. Finally, I finish with some thoughts on the future directions of the field.

## 6.1 SOME ISSUES WITH THESE METHODS

Model-based inference is only accurate to the extent that the model reasonably captures the key features of the (unknown) true process. Some simplification of reality is needed, as models cannot capture every aspect of a process while remaining tractable. The dominant paradigm in comparative phylogenetics is “null model testing”; for analyses of speciation and extinction the constant rate birth-death process is most commonly used, as a null model. We prefer the more complicated model over the null model (perhaps adding time- or state-dependence) if the complex model fits significantly better. However, this approach gives only a relative measure of support; even if the more complex model fits better, we have no strong evidence that it is actually a good fit to the data, rather than merely “less bad” than the simple model.

How badly do trait-dependent diversification models fit biological data? The models that I have considered in my thesis combine the birth-death model of speciation and extinction with simple Markov models of character evolution. There are a number of highly improbable assumptions and consequences of these models, including (i) exponential growth (or decline) in the number of species, (ii) no interaction among species during diversification or character evolution, (iii) independence of character change and speciation events from the time of the last such event, and (iv) homogeneity of rates over all species in the tree, over all time. While both the character evolution and diversification models have weaknesses, most attention has focused on the shortcomings of the birth-death model.

### 6.1.1 *The birth-death model, extinction, & the fossil record*

The most commonly cited line of evidence that the birth-death model is a poor fit to biological phylogenies is that the maximum-likelihood estimate of extinction rate is often zero. With extinction, there should be an upturn in the log number of lineages in a reconstructed phylogeny over time (figure 1.1), which would lead to positive values of the “gamma statistic” of Pybus and Harvey (2000). However in many molecular phylogenies, the gamma statistic is negative, indicating a decreasing slope in the log number of lineages over time, leading to extinction rate estimates of zero (reviews in McPeck, 2008; Phillimore and Price, 2008). This absence of extinction conflicts with even the most cursory comparison with the fossil record and with the motivation for the initial development of birth-death methods over the pure birth methods that preceded them (Nee *et al.*, 1994a; Kubo and Iwasa, 1995). This problem has motivated extensive work into alternatives to the constant rate birth-death process for modelling diversification. These alternatives fall into two broad (but overlapping) classes.

One class posits that rates of speciation or diversification have decreased over time. Consequently, methods for fitting time-varying speciation and extinction rates have proliferated recently (e.g., Rabosky, 2006; Rabosky and Lovette, 2008; Morlon *et al.*, 2010; Paradis, 2010; Rabosky and Glor, 2010; Stadler, 2011). However, there are a number of undesirable consequences of this interpretation. For trees where the lineage through time plot appears to be saturating (negative gamma), not only will extinction rates still be estimated to be zero at the present, but contemporary speciation rates will also be estimated to be very low (e.g., Paradis, 2010; Rabosky and Glor, 2010). This is difficult to reconcile with the view that both speciation and extinction are ongoing and can be rapid (e.g., Seehausen, 2006; Taylor *et al.*, 2006; Hendry *et al.*, 2007; Grant and Grant, 2009). Also, given the general prevalence of apparent slowdowns, this interpretation implies that diversification is slowing down simultaneously in *almost every clade* (Rabosky, 2009b). Perhaps we are at a peculiar moment in history where diversification is simultaneously slowing down in all groups, but this seems unlikely (though there

is palaeontological evidence for declining origination rates over time over broad taxonomic scales: e.g., Gilinsky and Bambach, 1987; Sepkoski, 1998).

The second major class of explanations postulates that species occupy niches, which may become filled over time (e.g., McPeck, 2008; Phillimore and Price, 2008; Rabosky, 2009a,b; Morlon *et al.*, 2010, 2011; Etienne *et al.*, 2012b). After an initial period of diversification, speciation and extinction rates balance to maintain the size of a clade at a characteristic carrying capacity of diversity. In its simplest form, this would imply logistic growth, rather than exponential growth for the constant rate birth-death process. While approaches for detecting diversity-dependent diversification from phylogenies are new, the idea itself has a long history in ecological and palaeontological thought (e.g., MacArthur, 1969; Raup *et al.*, 1973; Levinton, 1979; Walker and Valentine, 1984).

Proponents claim that modelling this process brings inferences from phylogeny closer to the fossil record (Rabosky, 2009a; Etienne *et al.*, 2012b). Based on fossil evidence, clades have been often been thought to reach a stable level of diversity early on in their evolutionary history, followed by periods with no consistent trend in diversity over time (e.g., Gould *et al.*, 1977; Alroy, 2008; Alroy *et al.*, 2008). For example, the fossil record supports equal origination and extinction rates at the family level in marine organisms (Gilinsky, 1994). Similarly, while diversity in tropical plants has generally increased over the last 65 million years, this increase has not been exponential, and estimated speciation rates are similar to extinction rates (Jaramillo *et al.*, 2006). Consistent with the predictions of diversity-dependent diversification, speciation rates have been inferred to increase following mass extinctions (Sepkoski, 1997) and to decrease when diversity is high (Alroy, 1996).

However, not all groups have fossil records that show strong signs of diversity dependence. While speciation rates may decline with diversity, this effect has been shown to be weaker in terrestrial systems than marine (Benton, 1997; Eble, 1999). In fact, insect groups show signs of a steady increase in diversity over time (Labandeira and Sepkoski, 1993; Mayhew, 2007). At the broadest scales, some have argued that fossil data are consis-



tent with a steady increase in diversity over hundreds of millions of years (Benton, 1995; Jablonski, 2008a; Kalmar and Currie, 2010). Furthermore, whether or not a saturating pattern of diversity is seen depends on the taxonomic level used; orders have stronger evidence for a carrying capacity than lower taxonomic units (Benton, 1997). Any density-dependent effects on speciation or extinction should most strongly affect species that are most closely related, most similar, and with overlapping ranges. Contrary to this, Alroy (1996) found that while overall diversity of North American mammals has not strongly increased over the last 25 Myr, speciation and extinction rates within genera are not tightly coupled.

It is also difficult to see that the major interactions that constrain diversity should necessarily occur within a clade. Particularly for clades that are not very diverse, why should the presence of a species in one continent affect the propensity of a species in another continent to speciate (Wiens, 2011)? Alternatively, speciation in one clade may be affected by distantly related clades, such as cospeciation in parasites of insects (Forbes *et al.*, 2009), mites of gophers (Huelsenbeck *et al.*, 2000), and wasp pollinators of figs (Weiblen and Bush, 2002). At a broader scale, the hypothesised codiversification of angiosperms and insects (e.g., Pellmyr, 1992; Farrell, 1998; Grimaldi, 1999) suggests that a single-clade view is rather too narrow. In a phylogenetic context, extensive periods of turnover with constant diversity (i.e., balanced speciation and extinction rates) should result in a proliferation of branches near the present, as in a coalescent tree, causing an upturn in a lineage through time plot — the lack of which was part of the original motivation for creating these sorts of models (Etienne *et al.*, 2012a)!

If it turns out that the first class of explanation (time varying rates) is a reasonable approximation of the true process, then it is straightforward to adapt the preceding chapters of my thesis to allow for this. Speciation and extinction rates would vary both in time and according to the value of a trait. Indeed, this has been done already (Rabosky and Glor, 2010, Chapter 4, J. L. Cantalapiedra *et al.*, in prep.). However, including the parameters required to model temporal variation in diversification may leave little power

to identify trait-dependent diversification. In contrast, if diversity-dependence really is a better model of diversification than the birth-death model, this will be difficult to combine with the approaches I have developed in my thesis, especially as there appears to be multiple distinct ways of modelling such dynamics. This view also raises deeper questions about what state-dependent diversification or species selection means in this context, as the interactions necessarily become frequency dependent. Most postulated species-selection scenarios correspond to simple directional (or possibly stabilising) selection; with diversity dependent diversification, different traits may confer different carrying capacities, rates of increase, or rates of character evolution, all of which may vary with the number and identity of other species.

#### 6.1.2 *Alternative explanations for poor fit*

Returning to the apparent slowdowns in diversification observed in phylogenies, there are additional explanations that have received less attention. In a birth-death model, “birth” can happen at any instant before the present. Indeed, we expect a large number of births very close to the present, especially with nonzero extinction rates (figure 1.1). However, we are reluctant to call two species separate until they are morphologically, ecologically, or reproductively distinct (Coyne and Orr, 2004). Recognising species therefore typically requires the passage of significant amounts of time since sharing a common ancestor, so that for many groups distinct species are at least hundreds of thousands or millions of years old. In essence, there may be a lag period between speciation (i.e., the point at which two populations share only trivial amounts of gene flow) and our recognition of species. This taxonomic lag will hide recent speciation events, essentially collapsing the last million years into unresolved clades that we recognise as a single species (Weir and Schluter, 2007; Purvis, 2008; Purvis *et al.*, 2009; Rosenblum *et al.*, 2012). Biased sampling of taxa towards including more distinct species (i.e., deeper nodes) would also lead to apparent slowdowns in diversification in pure birth trees, even when most taxa are included (Cusimano and Renner, 2010; Brock *et al.*, 2011).

The number of missing species may be substantial; cryptic diversity may more than double the number of known species (e.g., Hebert *et al.*, 2004a; Pfenniger and Schwenk, 2007; Funk *et al.*, 2012). Even DNA divergence-based estimates likely underestimate the true number of cryptic species, given that DNA divergence of the level often considered still requires substantial time to develop (e.g., 2% in Hebert *et al.*, 2004b). Low mutation rates, small sample sizes, and incomplete lineage sorting will all hamper identification of lineages that will never again exchange genes. In addition to cryptic species, new species are being discovered even in well-studied groups. It seems far more likely that we are missing species that are closely related to known species than missing entire deep splits, at least among vertebrates. For example, while 341 mammal species were discovered between 1992 and 2006, most were closely related to known species with only 18 new genera created (Reeder *et al.*, 2007). This pattern of missing taxa will generate apparent slowdowns in diversification.

An additional area of disconnect between the models and the data is caused by the quality of the phylogenies that we use. In most comparative phylogenetic approaches, we take the tree as given or, at best, compute a statistic over a distribution of trees. Compared to the extensive work on developing and comparing alternative models of cladogenesis, there has been relatively little work considering the impact of tree estimation error on estimates from any of these models (but see Revell *et al.*, 2005; Cusimano and Renner, 2010). Ages of recent splits may be systematically overestimated, which would create apparent slowdowns (Pulquério and Nichols, 2007; Purvis, 2008), as would a discordance between gene-trees and species-trees (Burbrink and Pyron, 2011). Estimates of speciation and extinction rates are sensitive to branch length, particularly near the tips, and differences between fossil date estimates, calibration, and methods of reconstruction can have pronounced effects on node ages (Pulquério and Nichols, 2007; Warnock *et al.*, 2012). While many methods for creating time-calibrated phylogenies attempt to take into account some biologically realistic patterns of deviations from a molecular clock, we are still understanding the ways that the clock can be violated (e.g., Smith and Donoghue,

2008; Schwartz and Mueller, 2010; Mayrose and Otto, 2011). Together, these issues raise the question of whether our tree inferences are good enough to accurately detect the true nature of deviations from a birth-death process.

### 6.1.3 *What about character evolution?*

The criticisms above focus on the birth-death model, but the models of character evolution used in my thesis are just as simple, and likely just as flawed. However, there does not seem to be the same groundswell of dislike against our current models of character evolution, to the point where we do not even know how badly they fit. We do know that trait-dependent diversification can bias estimates of character transition rates (Maddison, 2006). In addition, there is a general mistrust for ancestral state reconstruction methods (e.g., Omland, 1999; Losos, 2011), especially for continuous traits (Oakley and Cunningham, 2000; Webster and Purvis, 2002). However, the weaknesses in these models have not stimulated the same proliferation of methods as has occurred for birth-death models. For example, for binary data on small trees there is often insufficient power to distinguish between a model where rates of  $0 \rightarrow 1$  and  $1 \rightarrow 0$  transitions differ, and where these rates are equal (the Jukes–Cantor model). This led Mooers and Schluter (1999) to suggest that a two rate model was rarely justified. In contrast, estimation of zero extinction rates (which mean that the pure birth model is statistically indistinguishable from the birth-death model) has led to development of alternative models. I can offer no explanation for why two such similar issues have led to such dramatically different degrees of scepticism.

### 6.1.4 *Model adequacy*

Are simple models useful in any way? Can we learn anything from phylogenies about speciation, extinction, or character evolution that we didn't already know? Our models trade off realism in order to remain tractable:

“Brownian motion is a poor model, and so is Ornstein-Uhlenbeck, but just as democracy is the worst method of organising a society ‘except for all the others’, so these two models are all we’ve really got that is tractable. Critics will be admitted to the event, but only if they carry with them another tractable model.” Joe Felsenstein (R-SIG-phylo mailing list, 7 April 2008)

All models are incorrect. All we hope is that they capture enough “truth” to tell us something about a complex system. While I recognise that our models of cladogenesis and character evolution are a gross simplification of reality, it is not clear to me that there currently is a much better alternative to the birth-death model, with all options lacking in one way or other.

The main focus in this thesis is relating species traits to patterns of species diversity in order to detect the signature of species selection. It is quite possible that even if the model of cladogenesis is totally incorrect, and the rates are meaningless in an absolute sense, there may still be valuable information in the estimated parameters. For example if the “true model” involves some form of niche-filling, and if species with some particular trait have a higher rate of turnover or a higher carrying capacity, then it is likely that a method like BiSSE or QuaSSE would still correctly associate the trait with increased “speciation” (see also Wiens, 2011). How often this will happen, and how often this fortuitous association will be missed will depend on the true model of cladogenesis and deserves future study.

Actual patterns of diversification in both species and traits are far more complicated than any of the models discussed in my thesis or elsewhere. Patterns of fossil diversity in groups have risen and fallen, with complex temporal patterns of past diversity. For example large North American carnivores have experienced several bouts of wholesale replacement (Valkenburgh *et al.*, 2004), rather than an increase or steady pattern of diversity. Fossil data suggest Cetacean diversity was up to five times higher than today just 10 million years ago (Quental and Marshall, 2010). This seems equally at odds with both models positing a steady increase in diversity or a carrying capacity at a low

taxonomic level. Clearly, more work is needed to both improve our methods, and to better characterise how well they fit our data.

## 6.2 FUTURE DIRECTIONS

Here, I briefly outline some areas for possible future development and synthesis that would improve our ability to infer about possible past processes using phylogenies.

### 6.2.1 *Better diversification models*

Recently, there have been attempts to reconcile the models of cladogenesis with more realistic models of speciation and species concepts. Incorporating “protracted speciation”, where the probability of speciation of a lineage depends on the time since the previous speciation event, may capture biological aspects of the speciation process that the birth-death model ignores (Rosindell *et al.*, 2010; Etienne and Rosindell, 2012). Accounting for biased sampling can avoid misinference of speciation and extinction rates (Hönha *et al.*, 2011; Cusimano *et al.*, 2012). Better still, *detecting* such biased sampling and correcting for it automatically would increase the robustness of our inferences. Similarly, different species concepts imply different interpretations of “birth” and “death”, which in turn can lead to very different inferences about patterns of diversification from phylogenies (Ezard *et al.*, 2012). Very little research has been done along these lines, but this should bring our models considerably closer to matching biology. However, determining what the “correct” concept is will be challenging (and most likely, contentious). All these developments are recent, and it is not yet clear how they will affect choice among competing models of diversification.

### 6.2.2 *Better trait evolution models*

While development of alternative models of trait evolution appear to lag models of diversification, development of new approaches has recently accelerated. Methods to identify

topological shifts in the rate of trait evolution are now available (O’Meara *et al.*, 2006; Eastman *et al.*, 2011; Slater *et al.*, 2012), though only for continuous traits<sup>4</sup>. Felsenstein (2012) recently developed an method that allows for a “threshold” model of discrete trait evolution, in which an observed binary trait represents the sign of an unobserved continuous trait. This relaxes the assumption of constant rates of trait change by allowing traits to change more rapidly if they have recently changed, and might be expected whenever traits are parts of co-adapted trait complexes. Detecting the association between such a trait and speciation and extinction would be challenging (both statistically and computationally) but could represent a more realistic model than those I have presented. Models that allow for speciation shifts in character states (e.g., Bokma, 2008b; Magnuson-Ford and Otto, 2012; Goldberg and Igić, in press) are still in early stages of development but hold promise for testing hypotheses about the role of ecological speciation and punctuated equilibrium (but see Rabosky, 2012). It is possible to extend QuaSSE to allow for punctuated trait changes, but doing so will be computationally challenging. Similarly, methods that allow for more realistic modelling of species ranges, constraints, or simultaneous character evolution in multiple groups would significantly broaden the fairly meagre toolbox of trait models currently available. Probably the most useful development with respect to improving models of trait evolution would be the creation of goodness-of-fit statistics. The gamma statistic (Pybus *et al.*, 2002, see above) has been instrumental in motivating research into alternative diversification models, and a similar statistic for traits could do the same for trait evolution.

### 6.2.3 *Combining phylogenetics & the fossil record*

The fossil record is already the main source of comparison for phylogenetic approaches to studying speciation, extinction, and trait evolution. However, most studies treat fossils and phylogenies as two separate sources of information: one generates a hypothesis, and

---

<sup>4</sup>This is straightforward to implement for discrete traits for given shift positions in diversitree, though such a method’s statistical properties are unknown. Implementing search over possible shifts, as in Eastman *et al.* (2011), is less trivial.

the other tests it. Closer integration of phylogenetic and fossil analyses would allow better inferences than either alone. For example, incorporating fossil data can improve ancestral state estimation (Finarelli and Flynn, 2006). Recent efforts in a diversification context include Simpson *et al.* (2011), who used extinction data from the fossil record when inferring rates of speciation and extinction from a phylogeny of reef corals, and “Fossil MEDUSA” (J. Brown *et al.*, in prep.), which uses records of past diversity to relax assumptions of both temporal and topological homogeneity in rates.

While there is scope for improving inference by combining phylogenetic approaches with fossil data, this integration will introduce challenges that will have to be solved. Using fossils to date nodes in a phylogeny is probably the most common way that fossils are used to complement purely phylogenetic analyses, and this application illustrates some potential problems. Fossils belong to branches of unknown length that connect to unknown places within a phylogeny and adding them directly to phylogenies is not trivial (Felsenstein, 2002; Forey *et al.*, 2004). Even under ideal circumstances, with a complete fossil record and clock-like molecular evolution, discrepancies are expected between molecular and fossil dates, with fossil-based node ages biased towards the present as fossils indicate only minimum clade age (Brown *et al.*, 2008; Lukoschek *et al.*, 2012). The species concepts used by palaeontologists are necessarily morphological, and analyses are often performed at the level of morphological fossil “genera” or “families” (e.g., Benton, 1995; Mayhew, 2007; Jablonski, 2008a; Liow *et al.*, 2008). More than a semantic issue, different species concepts can entirely change interpretations about patterns of diversity both from the fossil record (Benton, 1997; Alroy, 2000; Forey *et al.*, 2004) and from phylogenies (Ezard *et al.*, 2012). Taxonomic misidentification can cause “pseudo-extinction”, where taxa disappear from the fossil record due to morphological change, rather than extinction (Alroy, 2009). Such pseudo-extinctions are paired with pseudo-speciation events, leading to inflation of both estimates of speciation and extinction rates, and the apparent correlation between these rates. Multiple specimens from a single species may be recorded as different species or even different genera (Forey *et al.*, 2004).



However, many of these issues can probably be addressed through modelling, and I believe there is scope to improve our analyses considerably when using all the available evidence simultaneously.

### 6.3 CONCLUSION

Much of the value in comparative phylogenetics is in formalising hypotheses about the processes that might have shaped diversity. Simple approaches such as sister-clade comparisons allow us to formally test hypotheses about the associations of traits with higher levels of diversity in a coarse manner (e.g., Mitter *et al.*, 1988; Barraclough *et al.*, 1998; Vamosi and Vamosi, 2005). I feel that the strength of the methods presented in my thesis is that they allow a more fine-grained test of associations and can lead to a more pluralistic view of how species have diversified and how traits may have become distributed (e.g., Goldberg *et al.*, 2005; Maddison, 2006; Clauset and Erwin, 2008; Schwander and Crespi, 2009). I hope that in combination with other types of data, the approaches discussed in my thesis will be a useful source of future hypotheses.

## BIBLIOGRAPHY

---

- Alfaro M.E., Santini F., Brock C., Alamillo H., Dornburg A., Rabosky D.L., Carnevale G., and Harmon L.J. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences, USA* 106:13410–13414.
- Allen L.J.S. 2003. *An Introduction to Stochastic Processes with Applications to Biology*. Pearson Prentice Hall, Upper Saddle River, N.J.
- Alroy J. 1996. Constant extinction, constrained diversification, and uncoordinated stasis in North American mammals. *Palaeogeography, Palaeoclimatology and Palaeoecology* 127:285–311.
- Alroy J. 1998. Cope's rule and the dynamics of body mass evolution in North American fossil mammals. *Science* 280:731–734.
- Alroy J. 2000. Successive approximations of diversity curves: ten more years in the library. *Geology* 28:1023–1026.
- Alroy J. 2008. Dynamics of origination and extinction in the marine fossil record. *Proceedings of the National Academy of Sciences, USA* 105:11536–11542.
- Alroy J. 2009. Speciation and extinction in the fossil record of North American mammals. *In* *Speciation and Patterns of Diversity* (R. Butlin, J. Bridle, and D. Schluter, eds.), pages 301–323, Cambridge University Press.
- Alroy J., Aberhan M., Bottjer D.J., Foote M., Fürsich F.T., Harries P.J., Hendy A.J.W., Holland S.M., Ivany L.C., Kiessling W., Kosnik M.A., Marshall C.R., McGowan A.J., Miller A.I., Olszewski T.D., Patzkowsky M.E., Peters S.E., Villier L., Wagner P.J., Bonuso N., Borkow P.S., Brenneis B., Clapham M.E., Fall L.M., Ferguson C.A., Hanson V.L., Krug A.Z., Layou K.M., Leckey E.H., Nürnberg S., Powers C.M., Sessa J.A., Simpson C., Tomasovych A., and Visaggi C.C. 2008. Phanerozoic trends in the global diversity of marine invertebrates. *Science* 321:97–100.
- Anacker B.L., Whittall J.B., Goldberg E.E., and Harrison S.P. 2010. Origins and consequences of serpentine endemism in the California flora. *Evolution* 65:365–376.
- Barracough T.G., Harvey P.H., and Nee S. 1995. Sexual selection and taxonomic diversity in passerine birds. *Proceedings of the Royal Society of London Series B* 259:211–215.
- Barracough T.G., Nee S., and Harvey P.H. 1998. Sister-group analysis in identifying correlates of diversification. *Evolutionary Ecology* 12:751–754.
- Benton M.J. 1995. Diversification and extinction in the fossil record. *Science* 268:52–58.

- Benton M.J. 1997. Models for the diversification of life. *Trends in Ecology and Evolution* 12:490–495.
- Benton M.J. 2002. Cope's rule. *In* *Encyclopedia of Evolution* (M. Pagel, ed.), pages 185–186, Oxford University Press, Oxford.
- Bininda-Emonds O.R.P., Cardillo M., Jones K.E., MacPhee R.D.E., Beck R.M.D., Grenyer R., Price S.A., Vos R.A., Gittleman J.L., and Purvis A. 2007. The delayed rise of present-day mammals. *Nature* 446:507–512.
- Bloch J.I., Rose K.D., and Gingerich P.D. 1988. New species of *Batodonoides* (Lipotyphla, Geolabididae) from the early Eocene of Wyoming: smallest known mammal? *Journal of Mammalogy* 79:804–827.
- Bokma F. 2008a. Bayesian estimation of speciation and extinction probabilities from (in)complete phylogenies. *Evolution* 69:2441–2445.
- Bokma F. 2008b. Detection of “punctuated equilibrium” by Bayesian estimation of speciation and extinction rates, ancestral character states, and rates of anagenetic and cladogenetic evolution on a molecular phylogeny. *Evolution* 62:2718–2726.
- Bokma F. 2009. Problems detecting density-dependent diversification on phylogenies. *Proceedings of the Royal Society of London Series B* 276:993–994.
- Bollback J.P. 2006. SIMMAP: Stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics* 7:88.
- Brock C.D., Harmon L.J., and Alfaro M.E. 2011. Testing for temporal variation in diversification rates when sampling is incomplete and nonrandom. *Systematic Biology* 60:410–419.
- Brown J.H. 1995. *Macroecology*. University of Chicago Press, Chicago.
- Brown J.H. and Maurer B.A. 1989. Macroecology: the division of food and space among species on continents. *Science* 243:1145–1150.
- Brown J.W., Rest J.S., García-moreno J., Sorenson M.D., and Mindell D.P. 2008. Strong mitochondrial DNA support for a Cretaceous origin of modern avian lineages. *BMC Biology* 6:6.
- Burbrink F.T. and Pyron R.A. 2011. The impact of gene-tree/species-tree discordance on diversification-rate estimation. *Evolution* 65:1851–1861.
- Burbrink F.T., Ruane S., and Pyron R.A. 2012. When are adaptive radiations replicated in areas? ecological opportunity and unexceptional diversification in West Indian dipsadine snakes (Colubridae: Alsophiini). *Journal of Biogeography* 39:465–475.
- Cardillo M. 1999. Latitude and rates of diversification in birds and butterflies. *Proceedings of the Royal Society of London Series B* 266:1221–1225.

- Cardillo M., Mace G.M., Jones K.E., Bielby J., Bininda-Emonds O.R.P., Sechrest W., Orme C.D.L., and Purvis A. 2005. Multiple causes of high extinction risk in large mammals. *Science* 309:1239–1241.
- Charlesworth B. 1981. Macroevolution: pattern and process (book review). *Biological Journal of the Linnean Society* 16:169–172.
- Charlesworth B., Lande R., and Slatkin M. 1982. A neo-Darwinian commentary on macroevolution. *Evolution* 36:474–498.
- Churchill G.A. 2000. Inferring ancestral character states, vol. 32 of *Evolutionary Biology*, chap. 6, pages 117–134. Kluwer Academic, New York.
- Claramunt S., Derryberry E.P., Remsen Jr J.V., and Brumfield R.T. 2012. High dispersal ability inhibits speciation in a continental radiation of passerine birds. *Proceedings of the Royal Society of London, Series B Biological Sciences* 279:1567–1574.
- Clauset A. and Erwin D.H. 2008. The evolution and distribution of species body size. *Science* 321:399–401.
- Cohen S.D. and Hindemarsch A.C. 1996. CVODE, a stiff/nonstiff ODE solver in C. *Computers in Physics* 10:138–143.
- Cope E. 1887. *The origin of the fittest*. Appleton, New York.
- Cope E.D. 1896. *The primary factors of organic evolution*. Open Court Publishing Company, Chicago.
- Coyne J.A. and Orr H.A. 2004. *Speciation*. Sinauer Associates, Massachusetts.
- Cusimano N. and Renner S.S. 2010. Slowdowns in diversification rates from real phylogenies may not be real. *Systematic Biology* 59:458–464.
- Cusimano N., Stadler T., and Renner S.S. 2012. A new method for handling missing species in diversification analysis applicable to randomly or non-randomly sampled phylogenies. *Systematic Biology* in press.
- Damuth J. and Heisler I.L. 1988. Alternative formulations of multilevel selection. *Biology and Philosophy* 3:407–430.
- Darwin C. 1871. *The Descent of Man, and Selection in Relation to Sex*. Murray, London.
- Davies T.J., Barraclough T.G., Chase M.W., Soltis P.S., and Soltis D.E. 2004. Darwin's abominable mystery: insights from a supertree of the angiosperms. *Proceedings of the National Academy of Sciences, USA* 101:1904–1909.
- Dial K.P. and Marzluff J.M. 1988. Are the smallest organisms the most diverse? *Ecology* 69:1620–1624.
- Drummond A.J. and Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7:214.

- Duda T F J. and Palumbi S.R. 1999. Developmental shifts and species selection in gastropods. *Proceedings of the National Academy of Sciences, USA* 96:10272–10277.
- Eastman J.M., Alfaro M.E., Joyce P., Hipp A.L., and Harmon L.J. 2011. A novel comparative method for identifying shifts in the rate of character evolution on trees. *Evolution* 65:3578–3589.
- Eastman J.M. and Storfer A. 2011. Correlations of life-history and distributional-range variation with salamander diversification rates: Evidence for species selection. *Systematic Biology* 60:503–518.
- Eble G.J. 1999. Originations: land and sea compared. *Geobios* 2.
- Eldredge N. and Gould S.J. 1972. Punctuated equilibria: an alternative to phyletic gradualism. *In* *Models in Paleobiology* (T.J.M. Schopf, ed.), pages 82–115, Freeman, Cooper and Company, San Francisco.
- Erwin D.H. 2010. Macroevolution is more than repeated rounds of microevolution. *Evolution and Development* 2:78–84.
- Estes S. and Arnold S.J. 2009. Resolving the paradox of stasis: Models with stabilizing selection explain evolutionary divergence on all timescales. *The American Naturalist* 169:227–244.
- Etienne R.S., de Visser S.N., Janzen T., Olsen J.L., Olff H., and Rosindell J. 2012a. Can clade age alone explain the relationship between body size and diversity? *Interface Focus* 2:170–179.
- Etienne R.S., Haegeman B., Stadler T., Aze T., Pearson P., Purvis A., and Phillimore A.B. 2012b. Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. *Proceedings of the Royal Society of London Series B* in press.
- Etienne R.S. and 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.
- Ezard T., Pearson P., Aze T., and Purvis A. 2012. The meaning of birth and death (in macroevolutionary birth-death models). *Biology Letters* 8:139–142.
- Farrell B.D. 1998. “Inordinate fondness” explained: why are there so many beetles? *Science* 281:555–559.
- Felsenstein J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. *American Journal of Human Genetics* 25:471–492.
- Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* 17:368–376.
- Felsenstein J. 1985. Phylogenies and the comparative method. *The American Naturalist* 125:1–15.

- Felsenstein J. 1988. Phylogenies and quantitative characters. *Annual Review of Ecology and Systematics* 19:445–471.
- Felsenstein J. 2002. Quantitative characters, phylogenies, and morphometrics. *In* *Morphology, Shape, and Phylogenetics* (N. MacLeod, ed.), Systematics Association Special Volume Series 64, Taylor and Francis, London.
- Felsenstein J. 2012. A comparative method for both discrete and continuous characters using the threshold model. *The American Naturalist* 179:145–156.
- Figuerola J. 1999. A comparative study on the evolution of reversed size dimorphism in monogamous waders. *Biological Journal of the Linnean Society* 67:1–18.
- Finarelli J.A. and Flynn J.J. 2006. Ancestral state reconstruction of body size in the Caniformia (Carnivora, Mammalia): The effects of incorporating data from the fossil record. *Systematic Biology* 55:301–313.
- Fisher R.A. 1958. *The genetical theory of natural selection*. Dover, New York.
- FitzJohn R.G. 2010. Quantitative traits and diversification. *Systematic Biology* 59:619–633.
- FitzJohn R.G., Maddison W.P., and Otto S.P. 2009. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Systematic Biology* 58:595–611.
- Forbes A.A., Powell T.H.Q., Stelinski L.L., Smith J.J., and Feder J.L. 2009. Sequential sympatric speciation across trophic levels. *Science* 323:776–779.
- Forey P.L., Fortey R.A., Kenrick P., and Smith A.B. 2004. Taxonomy and fossils: a critical appraisal. *Philosophical Transactions of the Royal Society B: Biological Sciences* 359:369–653.
- Freckleton R.P., Phillimore A.B., and Pagel M. 2008. Relating traits to diversification: a simple test. *The American Naturalist* 172:102–115.
- Frigo M. and Johnson S.G. 2005. The design and implementation of FFTW3. *Proceedings of the IEEE* 93:216–231.
- Fritz S.A., Bininda-Emonds O.R.P., and Purvis A. 2009. Geographical variation in predictors of mammalian extinction risk: big is bad, but only in the tropics. *Ecology Letters* 12:538–549.
- Funk W.C., Caminer M., and Ron S.R. 2012. High levels of cryptic species diversity uncovered in Amazonian frogs. *Proceedings of the Royal Society of London Series B* in press.
- Gage M.J.G., Parker G.A., Nylin S., and Wiklund C. 2002. Sexual selection and speciation in mammals, butterflies and spiders. *Proceedings of the Royal Society of London Series B* 269:2309–2316.

- Gardezi T. and da Silva J. 1999. Diversity in relation to body size in mammals: a comparative study. *The American Naturalist* 153:110–123.
- Gavrilets S. 2000. Rapid evolution of reproductive barriers driven by sexual conflict. *Nature* 403:886–889.
- Gelman A., Carlin J.B., Stern H.S., and Rubin D.B. 1995. *Bayesian Data Analysis*. Chapman & Hall, London.
- Gilinsky N.L. 1994. Volatility and the phanerozoic decline of background extinction intensity. *Paleobiology* 20:445–458.
- Gilinsky N.L. and Bambach R.K. 1987. Asymmetrical patterns of origination and extinction in higher taxa. *Paleobiology* 13:427–445.
- Gittleman J. and Purvis A. 1998. Body size and species-richness in carnivores and primates. *Proceedings of the Royal Society of London Series B* 265:113–119.
- Goldberg E.E. and Igić B. 2008. On phylogenetic tests of irreversible evolution. *Evolution* 62:2727–2741.
- Goldberg E.E. and Igić B. in press. Tempo and mode in plant breeding system evolution. *Evolution*.
- Goldberg E.E., Kohn J.R., Lande R., Robertson K.A., Smith S.A., and Igić B. 2010. Species selection maintains self-incompatibility. *Science* 330:493–495.
- Goldberg E.E., Lancaster L.T., and Ree R.H. 2011. Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. *Systematic Biology* 60:451–465.
- Goldberg E.E., Roy K., Lande R., and Jablonski D. 2005. Diversity, endemism, and age distributions in macroevolutionary sources and sinks. *The American Naturalist* 165:623–633.
- Gómez J.M. and Verdú M. 2012. Mutualism with plants drives primate diversification. *Systematic Biology* in press.
- Gould S.J. and Eldredge N. 1977. Punctuated equilibrium: the tempo and mode of evolution reconsidered. *Paleobiology* 3:115–151.
- Gould S.J., Raup D.M., Sepkoski Jr. J.J., Schopf T.J.M., and Simberloff D.S. 1977. The shape of evolution: A comparison of real and random clades. *Paleobiology* 3:23–40.
- Grant P.R. and Grant B.R. 2002. Unpredictable evolution in a 30-year study of darwin's finches. *Science* 296:707–711.
- Grant P.R. and Grant B.R. 2009. The secondary contact phase of allopatric speciation in Darwin's finches. *Proceedings of the National Academy of Sciences, USA* 106:20141–20148.

- Grantham T.A. 1995. Hierarchical approaches to macroevolution: recent work on species selection and the “effect hypothesis”. *Annual Review of Ecology and Systematics* 26:301–321.
- Green P.J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82:711–732.
- Grimaldi D. 1999. The co-radiations of pollinating insects and angiosperms in the cretaceous. *Annals of the Missouri Botanical Garden* 86:373–406.
- Grimshaw A.J. 2001. The Adventures of Punctuated Equilibria. Ph.D. thesis, Deakin University.
- Hackett S.J., Kimball R.T., Reddy S., Bowie R.C.K., Braun E.L., Braun M.J., Chojnowski J.L., Cox W.A., Han K.L., Harshman J., Huddleston C.J., Marks B.D., Miglia K.J., Moore W.S., Sheldon F.H., Steadman D.W., Witt C.C., and Yuri T. 2008. A phylogenomic study of birds reveals their evolutionary history. *Science* 320:1763–1768.
- Hansen T.F. and Martins E.P. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution* 50:1404–1414.
- Harcourt A.H., Copperto S.A., and Parks S.A. 2002. Rarity, specialization and extinction in primates. *Journal of Biogeography* 29:445–456.
- Harmon L.J., Weir J.T., Brock C.D., Glor R.E., and Challenger W. 2008. GEIGER: investigating evolutionary radiations. *Bioinformatics* 24:129–131.
- Harvey P.H., May R.M., and Nee S. 1994. Phylogenies without fossils. *Evolution* 48:523–529.
- Hebert P.D.N., Penton E.H., Burns J.M., Janzen D.H., and Hallwachs W. 2004a. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astrartes fulgerator*. *Proceedings of the National Academy of Sciences, USA* 101:14812–14817.
- Hebert P.D.N., Stoeckle M.Y., Zemplak T.S., and Frances C.M. 2004b. Identification of birds through DNA barcodes. *PLoS Biology* 2:e312.
- Heilbuth J.C. 2000. Lower species richness in dioecious clades. *The American Naturalist* 156:221–241.
- Hendry A.P., Nosil P., and Rieseberg L.H. 2007. The speed of ecological speciation. *Functional Ecology* 21:455–464.
- Hernández Fernández M. and Vrba E.S. 2005. A complete estimate of the phylogenetic relationships in Ruminantia: a dated species-level supertree of the extant ruminants. *Biological Reviews* 80:269–302.



- Herrera C.M. 1992. Historical effects and sorting processes as explanations for contemporary ecological patterns: Character syndromes in mediterranean woody plants. *The American Naturalist* 140:421–446.
- Herron M.D., Castoe T.A., and Parkinson C.L. 2004. Sciurid phylogeny and the paraphyly of holarctic ground squirrels (*Spermophilus*). *Molecular Phylogenetics and Evolution* 31:1015–1030.
- Hindmarsh A.C., Brown P.N., Grant K.E., Lee S.L., Serban R., Shumaker D.E., and Woodward C.S. 2005. SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. *ACM Transactions on Mathematical Software* 31:363–396.
- Hodges S.A. and Arnold M.L. 1995. Spurring plant diversification: are floral nectar spurs a key innovation? *Proceedings of the Royal Society of London, Series B* 262:343–348.
- Hönha S., Stadler T., Ronquist F., and Britton R. 2011. Inferring speciation and extinction rates under different sampling schemes. *Molecular Biology and Evolution* 28:2577–2589.
- Huelsenbeck J.P., Larget B., Miller R.E., and Ronquist F. 2002. Potential applications and pitfalls of Bayesian inference of phylogeny. *Systematic Biology* 51:673–688.
- Huelsenbeck J.P., Rannala B., and Larget B. 2000. A Bayesian framework for the analysis of cospeciation. *Evolution* 54:352–364.
- Isaac N.J.B., Agapow P.M., Harvey P.H., and Purvis A. 2003. Phylogenetically nested comparisons for testing correlates of species richness: a simulation study of continuous variables. *Evolution* 57:18–26.
- Jablonski D. 1997. Body-size evolution in Cretaceous molluscs and the status of Cope's rule. *Nature* 385:250–252.
- Jablonski D. 2000. Micro- and macroevolution: scale and hierarchy in evolutionary biology and paleobiology. *Paleobiology* 26:15–52.
- Jablonski D. 2008a. Biotic interactions and macroevolution: extensions and mismatches across scales and levels. *Evolution* 62:715–739.
- Jablonski D. 2008b. Species selection: theory and data. *Annual Review of Ecology, Evolution, and Systematics* 39:501–524.
- Jaramillo C., Rueda M.J., and Mora G. 2006. Cenozoic plant diversity in the Neotropics. *Science* 311:1893–1896.
- Johnson M.T.J., FitzJohn R.G., Smith S.D., Rausher M.D., and Otto S.P. 2011. Loss of sexual recombination and segregation is associated with increased diversification in evening primroses. *Evolution* 65:3230–3240.

- Jones K.E., Bielby J., Cardillo M., Fritz S.A., O'Dell J., Orme C.D.L., Safi K., Sechrest W., Boaks E.H., Carbone C., Connolly C., Cutts M.J., Foster J.K., Grenyer R., Habib M., Plaster C.A., Price S.A., Rigby E.A., Rist J., Teacher A., Bininda-Emonds O.R.P., Gittleman J.L., Mace G.M., and Purvis A. 2009. PanTHERIA: A species-level database of life-history, ecology and geography of extant and recently extinct mammals. *Ecology* 90:2648.
- Kalmar A. and Currie D.J. 2010. The completeness of the continental fossil record and its impact on patterns of diversification. *Paleobiology* 36:51–60.
- Karlin S. and Taylor H.M. 1981. *A Second Course in Stochastic Processes*. Academic Press, London.
- Kass R.E. and Raftery A.E. 1995. Bayes factors. *Journal of the American Statistical Association* 90:773–795.
- Kubo T. and Iwasa Y. 1995. Inferring the rates of branching and extinction from molecular phylogenies. *Evolution* 49:694–704.
- Kuhn T.S., Mooers A.O., and Thomas G.H. 2011. A simple polytomy resolver for dated phylogenies. *Methods in Ecology and Evolution* 2:427–436.
- Labandeira C.C. and Sepkoski Jr. J.J. 1993. Insect diversity in the fossil record. *Science* 261:310–315.
- Lande R. 1980. Microevolution in relation to macroevolution. *Paleobiology* 6:233–238.
- Leigh E G J. 1977. How does selection reconcile individual advantage with the good of the group? *Proceedings of the National Academy of Sciences, USA* 74:4542–4546.
- Leisch F. 2002. Sweave: Dynamic generation of statistical reports using literate data analysis. *In* *Compstat 2002 — Proceedings in Computational Statistics* (W. Härdle and B. Rönz, eds.), pages 575–580, Physica Verlag, Heidelberg, ISBN 3-7908-1517-9.
- Levinton J.S. 1979. A theory of diversity equilibrium and morphological evolution. *Science* 204:335–336.
- Lewis P.O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* 50:913–925.
- Lewontin R. 1970. The units of selection. *Annual Review of Ecology and Systematics* 1:1–18.
- Lieberman B.S. and Vrba E.S. 2005. Stephen Jay Gould on species selection: 30 years of insight. *Paleobiology* 31:113–121.
- Liow L.H., Fortelius M., Bingham E., Lintulaakso K., Mannila H., Flynn L., and Stenseth N.C. 2008. Higher origination and extinction rates in larger mammals. *Proceedings of the National Academy of Sciences, USA* 105:6097–6102.
- Lislevand T., Figuerola J., and Székely T. 2007. Avian body sizes in relation to fecundity, mating system, display behavior, and resource sharing. *Ecology* 88:1605.

- Lloyd E.A. and Gould S.J. 1993. Species selection on variability. *Proceedings of the National Academy of Sciences, USA* 90:595–599.
- Losos J. 2011. Seeing the forest for the trees: The limitations of phylogenies in comparative biology. *The American Naturalist* 177:709–727.
- Lukoschek V., Keogh J.S., and Avise J.C. 2012. Evaluating fossil calibrations for dating phylogenies in light of rates of molecular evolution: A comparison of three approaches. *Systematic Biology* 61:22–43.
- Lutzoni F., Pagel M., and Reeb V. 2001. Major fungal lineages are derived from lichen symbiotic ancestors. *Nature* 411:937–940.
- Lyell C. 1832. *Principles of Geology*, vol. 2. John Murray, London.
- Lynch V.J. 2009. Live-birth in vipers (Viperidae) is a key innovation and adaptation to global cooling during the Cenozoic. *Evolution* 63:2457–2465.
- MacArthur R.H. 1969. Patterns of communities in the tropics. *Biological Journal of the Linnean Society* 1:19–30.
- MacKay D.J.C. 2003. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, New York.
- Maddison D.R., Swofford D.L., and Maddison W.P. 1997. NEXUS: an extensible file format for systematic information. *Systematic Biology* 46:590–621.
- Maddison W.P. 2006. Confounding asymmetries in evolutionary diversification and character change. *Evolution* 60:1743–1746.
- Maddison W.P. and Maddison D.R. 2006. Mesquite: a modular system for evolutionary analysis. Version 1.1 <http://www.mesquiteproject.org>.
- Maddison W.P. and Maddison D.R. 2008. Mesquite: A modular system for evolutionary analysis. Version 2.5. <http://www.mesquiteproject.org>.
- Maddison W.P., Midford P.E., and Otto S.P. 2007. Estimating a binary character's effect on speciation and extinction. *Systematic Biology* 56:701–710.
- Magnuson-Ford K. and Otto S.P. 2012. Linking the investigations of character evolution and species diversification. *The American Naturalist* .
- Mayhew P.J. 2007. Why are there so many insect species? Perspectives from fossils and phylogenies. *Biological Review* 82:425–454.
- Maynard Smith J. 1983. Models of evolution. *Proceedings of the Royal Society of London Series B* 219:315–325.
- Maynard Smith J. 1989. The causes of extinction. *Philosophical Transactions of the Royal Society B: Biological Sciences* 325:241–252.
- Mayrose I. and Otto S.P. 2011. A likelihood method for detecting trait-dependent shifts in the rate of molecular evolution. *Molecular Biology and Evolution* 28:759–770.

- McGlone M.S., Duncan R.P., and Heenan P.B. 2001. Endemism, species selection and the origin and distribution of the vascular plant flora of new zealand. *Journal of Biogeography* 28:199–216.
- McPeck M.A. 2008. The ecological dynamics of clade diversification and community assembly. *The American Naturalist* 172:E270–E284.
- Mitra S., Landel H., and Pruett-Jones S. 1996. Species richness covaries with mating system in birds. *The Auk* 113:544–551.
- Mitter C.B., Farrell B., and Wiegmann B. 1988. The phylogenetic study of adaptive zones: has phytophagy promoted insect diversification? *The American Naturalist* 132:107–128.
- Moler C. and Loan C.V. 2003. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Review* 45:3–49.
- Mooers A.Ø. and Schluter D. 1999. Reconstructing ancestor states with maximum likelihood: support for one- and two-rate models. *Systematic Biology* 48:623–633.
- Moore M.J., Bell C.D., Soltis P.S., and Soltis D.E. 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proceedings of the National Academy of Sciences, USA* 104:19363–19368.
- Morlon H., Parsons T., and Plotkin J. 2011. Reconciling molecular phylogenies with the fossil record. *Proceedings of the National Academy of Sciences, USA* 108:16327–16332.
- Morlon H., Potts M.D., and Plotkin J. 2010. Inferring the dynamics of diversification: a coalescent approach. *PLoS Biology* 8:e1000493.
- Morrow E.H. and Pitcher T.E. 2003. Sexual selection and the risk of extinction in birds. *Proceedings of the Royal Society of London Series B* 270:1793–1799.
- Morrow E.H., Pitcher T.E., and Arnqvist G. 2003. No evidence that sexual selection is an ‘engine of speciation’ in birds. *Ecology Letters* 6:228–234.
- Morse D.R., Lawton J.H., Dodson M.M., and Williamson M.H. 1985. Fractal dimension of vegetation and the distribution of arthropod body lengths. *Nature* 314:731–733.
- Mossel E. and Steel M. 2005. How much can evolved characters tell us about the tree that generated them? *In* *Mathematics of Evolution and Phylogeny* (O. Gascuel, ed.), pages 384–412, Oxford University Press.
- Moyle R.G., Filardi C.E., Smith C.E., and Diamond J. 2009. Explosive Pleistocene diversification and hemispheric expansion of a “great speciator”. *Proceedings of the National Academy of Sciences, USA* 106:1863–1868.
- Neal R.M. 2003. Slice sampling. *Annals of Statistics* 31:705–767.
- Nee S. 2001. Inferring speciation rates from phylogenies. *Evolution* 55:661–668.

- Nee S. 2006. Birth-death models in macroevolution. *Annual Review of Ecology and Systematics* 37:1–17.
- Nee S., Holmes E.C., May R.M., and Harvey P.H. 1994a. Extinction rates can be estimated from molecular phylogenies. *Philosophical Transactions of the Royal Society B: Biological Sciences* 344:77–82.
- Nee S., May R.M., and Harvey P.H. 1994b. The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society B: Biological Sciences* 344:305–311.
- Oakley T.H. and Cunningham C.W. 2000. Independent contrasts succeed where ancestor reconstruction fails in a known bacteriophage phylogeny. *Evolution* 54.
- Okasha S. 2006. *Evolution and the Levels of Selection*. Oxford University Press, New York.
- O'Meara B.C., Anè C., Sanderson M.J., and Wainwright P.C. 2006. Testing for different rates of continuous trait evolution using likelihood. *Evolution* 60:922–933.
- Omeland K.E. 1999. The assumptions and challenges of ancestral state reconstructions. *Systematic Biology* 48:604–611.
- Owens I.P.F., Bennett P.M., and Harvey P.H. 1999. Species richness among birds: body size, life history, sexual selection or ecology. *Proceedings of the Royal Society of London Series B* 266:933–939.
- Pagel M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society of London Series B* 255:37–45.
- Pagel M. 1997. Inferring evolutionary processes from phylogenies. *Zoologica Scripta* 26:331–348.
- Pagel M. and Meade A. 2006. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *The American Naturalist* 167:808–825.
- Paradis E. 2005. Statistical analysis of diversification with species traits. *Evolution* 59:1–12.
- Paradis E. 2008. Asymmetries in phylogenetic diversification and character change can be untangled. *Evolution* 61:241–247.
- Paradis E. 2010. Time-dependent speciation and extinction from phylogenies: a least squares approach. *Evolution* 65:661–672.
- Paradis E., Claude J., and Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290.
- Parker G.A. and Partridge L. 1998. Sexual conflict and speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences* 353:261–274.

- Pellmyr O. 1992. Evolution of insect pollination and Angiosperm diversification. *Trends in Ecology and Evolution* 7:46–49.
- Pfenniger M. and Schwenk K. 2007. Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. *BMC Evolutionary Biology* 7:121.
- Phillimore A.B. and Price T.D. 2008. Density-dependent cladogenesis in birds. *PLoS Biology* 6:483–489.
- Phillimore A.B.,reckleton R.P., Orme C.D.L., and Owens I.P.F. 2006. Ecology predicts large-scale patterns of phylogenetic diversification in birds. *The American Naturalist* 168:220–229.
- Pillon Y., Munzinger J., Amir H., and Lebrun M. 2010. Ultramafic soils and species sorting in the flora of New Caledonia. *Journal of Ecology* 98:1108–1116.
- Plummer M., Best N., Cowles K., and Vines K. 2006. CODA: Convergence diagnosis and output analysis for MCMC. *R News* 6:7–11.
- Polly P.D. 1998. Cope's rule. *Science* 282:50–51.
- Pulquério M.J.F. and Nichols R.A. 2007. Dates from the molecular clock: how wrong can we be? *Trends in Ecology and Evolution* 180–184.
- Purvis A. 2008. Phylogenetic approaches to the study of extinction. *Annual Review of Ecology and Systematics* 39:301–319.
- Purvis A., Orme C.D.L., Toomey N.H., and Pearson P.N. 2009. Temporal patterns in diversification rates. *In* *Speciation and Patterns of Diversity* (R. Butlin, J. Bridle, and D. Schluter, eds.), pages 278–300, Cambridge University Press.
- Pybus O.G. and Harvey P.H. 2000. Testing macro-evolutionary models using incomplete molecular phylogenies. *Proceedings of the Royal Society of London Series B* 267:2267–2272.
- Pybus O.G., Rambaut A., Holmes E.C., and Harvey P.H. 2002. New inferences from tree shape: numbers of missing taxa and population growth curves. *Systematic Biology* 51:881–888.
- Quental T.B. and Marshall C.R. 2010. Diversity dynamics: Molecular phylogenies need the fossil record. *Trends in Ecology and Evolution* 25:434–441.
- R Development Core Team. 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
- Rabosky D.L. 2006. Likelihood methods for detecting temporal shifts in diversification rates. *Evolution* 60:1152–1164.
- Rabosky D.L. 2009a. Ecological limits and diversification rate: alternative paradigms to explain the variation in species richness among clades and regions. *Ecology Letters* 12:735–743.

- Rabosky D.L. 2009b. Heritability of extinction rates links diversification patterns in molecular phylogenies and fossils. *Systematic Biology* 58:629–640.
- Rabosky D.L. 2012. Positive correlation between diversification rates and phenotypic evolvability can mimic punctuated equilibrium on molecular phylogenies. *Evolution in press*.
- Rabosky D.L., Donnellan S.C., Talaba A.L., and Lovette I.J. 2007. Exceptional among-lineage variation in diversification rates during the radiation of Australia's most diverse vertebrate clade. *Proceedings of the Royal Society of London Series B* 274:2915–2923.
- Rabosky D.L. and Glor R.E. 2010. Equilibrium speciation dynamics in a model adaptive radiation of island lizards. *Proceedings of the National Academy of Sciences, USA* 51:22178–22183.
- Rabosky D.L. and Lovette I.J. 2008. Explosive evolutionary radiations: decreasing speciation or increasing extinction through time? *Evolution* 62:1866–1875.
- Rabosky D.L. and McCune A.R. 2009. Reinventing species selection with molecular phylogenies. *Trends in Ecology and Evolution* 25:68–74.
- Raia P., Carotenuto F., Passaro F., Fulgione D., and Fortelius M. 2012. Ecological specialization in fossil mammals explains cope's rule. *The American Naturalist* 179:328–337.
- Raup D.M., Gould S.J., Schopf R.J.M., and Simberloff D.S. 1973. Stochastic models of phylogeny and the evolution of diversity. *Journal of Geology* 81:525–542.
- Read A.F. and Nee S. 1995. Inference from binary comparative data. *Journal of Theoretical Biology* 173:99–108.
- Redding D.W., DeWolff C., and Mooers A.Ø. 2010. Evolutionary distinctiveness, threat status and ecological oddity in primates. *Conservation Biology* 24:1052–1058.
- Ree R.H. 2005. Detecting the historical signature of key innovations using stochastic models of character evolution and cladogenesis. *Evolution* 59:257–265.
- Reeder D.M., Helgen K.M., and Wilson D.E. 2007. Global trends and biases in new mammal species discoveries. *Occasional Papers, Museum of Texas Tech University* 269:1–35.
- Rensch B. 1948. Histological changes correlated with evolutionary changes of body size. *Evolution* 2:218–230.
- Revell L.J., Harmon L.J., and Glor R.E. 2005. Underparameterized model of sequence evolution leads to bias in the estimation of diversification rates from molecular phylogenies. *Systematic Biology* 54:973–983.
- Rice S.H. 1995. A genetical theory of species selection. *Journal of Theoretical Biology* 177:237–245.

- Ricklefs R.E. 2007. Estimating diversification rates from phylogenetic information. *Trends in Ecology and Evolution* 22:601–610.
- Rosenblum E.B., Sarver B.A.J., Brown J.W., Des Roches S., Hardwick K.M., Hether T.D., Eastman J.M., Pennell M.W., and Harmon L.J. 2012. Goldilocks meets Santa Rosalia: An ephemeral speciation model explains patterns of diversification across time scales. *Evolutionary Biology* in press.
- Rosenzweig M.L. 1996. Colonial birds probably do speciate faster. *Evolutionary Ecology* 681–683.
- Rosindell J., Cornell S.J., Hubbell S., and Etienne R.S. 2010. Protracted speciation revitalizes the neutral theory of biodiversity. *Ecology Letters* 13:716–727.
- Rowan T. 1990. Functional Stability Analysis of Numerical Algorithms. Ph.D. thesis, University of Texas at Austin.
- Sallan L.C., Kammer T.W., Ausich W.I., and Cook L.A. 2011. Persistent predator–prey dynamics revealed by mass extinction. *Proceedings of the National Academy of Sciences, USA* 108:8335–8338.
- Schluter D. 2009. Evidence for ecological speciation and its alternative. *Science* 323:737–741.
- Schluter D., Price T., Mooers A.Ø., and Ludwig D. 1997. Likelihood of ancestor states in adaptive radiation. *Evolution* 51:1699–1711.
- Schwander T. and Crespi B. 2009. Twigs on the tree of life? neutral and selective models for integrating macroevolutionary patterns with microevolutionary processes in the analysis of asexuality. *Molecular Ecology* 18:28–42.
- Schwartz R.S. and Mueller R.L. 2010. Branch length estimation and divergence dating: estimates of error in bayesian and maximum likelihood frameworks. *BMC Evolutionary Biology* 10:5.
- Seehausen O. 2006. Conservation: Losing biodiversity by reverse speciation. *Current Biology* 16:R334–R337.
- Sepkoski Jr. J.J. 1997. Biodiversity: past, present, and future. *Journal of Paleontology* 71:533–539.
- Sepkoski Jr. J.J. 1998. Rates of speciation in the fossil record. *Philosophical Transactions of the Royal Society B: Biological Sciences* 353:315–326.
- Sidje R.B. 1998. Expokit: A software package for computing matrix exponentials. *ACM Transactions on Mathematical Software* 24:130–156.
- Siepielski A.M., DiBattista J.D., and Carlson S.M. 2009. It's about time: the temporal dynamics of phenotypic selection in the wild. *Ecology Letters* 12:1261–1276.
- Simpson C. 2010. Species selection and driven mechanisms jointly generate a large-scale morphological trend in monobathrid crinoids. *Paleobiology* 36:481–496.



- Simpson C., Kiessling W., Mewis H., Baron-Szabo R.C., and Müller J. 2011. Evolutionary diversification of reef corals: a comparison of the molecular and fossil records. *Evolution* 65:3274–3284.
- Slater G.J., Harmon L.J., Wegmann D., Joyce P., Revell L.J., and Alfaro M.E. 2012. Fitting models of continuous trait evolution to incompletely sampled comparative data using Approximate Bayesian Computation. *Evolution* 66:752–762.
- Slatkin M. 1981. A diffusion model of species selection. *Paleobiology* 7:421–425.
- Smith S. and Donoghue M. 2008. Rates of molecular evolution are linked to life history in flowering plants. *Science* 322:86–89.
- Soetaert K., Petzoldt T., and Setzer R.W. 2010. Solving differential equations in r: Package desolve. *Journal of Statistical Software* 33:1–25.
- Stadler T. 2011. Mammalian phylogeny reveals recent diversification rate shifts. *Proceedings of the National Academy of Sciences, USA* 108:6187–6192.
- Stanley S.M. 1973. An explanation for Cope's rule. *Evolution* 27:1–26.
- Stanley S.M. 1975a. Clades versus clones in evolution: why we have sex. *Science* 190:382–383.
- Stanley S.M. 1975b. A theory of evolution above the species level. *Proceedings of the National Academy of Sciences, USA* 72:646–650.
- Steeman M.E., Hebsgaard M.B., Fordyce R.E., Ho S.Y.W., Rabosky D.L., Nielsen R., Rahbek C., Glenner H., Sørensen M.V., and Willerslev E. 2009. Radiation of extant cetaceans driven by restructuring of the oceans. *Systematic Biology* 58:573–585.
- Taylor E.B., Boughman J.W., Groenboom M., Sniatynski M., Schluter D., and Gow J.L. 2006. Speciation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Molecular Ecology* 15:343–355.
- Thomas G.H., Wills M.A., and Székely T. 2004. A supertree approach to shorebird phylogeny. *BMC Evolutionary Biology* 4:28.
- Valkenburgh B.V., Wang X., and Damuth J. 2004. Cope's rule, hypercarnivory, and extinction in north american canids. *Science* 306:101–104.
- Vamosi S.M. and Vamosi J.C. 2005. Endless tests: guidelines for analysing non-nested sister-group comparisons. *Evolutionary Ecology Research* 7:567–579.
- Van Valen L.M. 1975. Group selection, sex, and fossils. *Evolution* 29:87–94.
- Vos R.A. and Mooers A.Ø. 2006. A new dated supertree of the Primates. chapter 5. *In* Inferring large phylogenies: the big tree problem. (R.A. Vos) PhD. thesis, Simon Fraser University.

- Vrba E.S. and Gould J.S. 1986. The hierarchical expansion of sorting and selection: sorting and selection cannot be equated. *Paleobiology* 12:217–228.
- Walker T.D. and Valentine J.W. 1984. Equilibrium models of evolutionary species diversity and the number of empty niches. *Evolution* 124:887–899.
- Warnock R.C.M., Yang Z., and Donoghue P.C.J. 2012. Exploring uncertainty in the calibration of the molecular clock. *Biology Letters* 8:156–159.
- Webster A.J. and Purvis A. 2002. Testing the accuracy of methods for reconstructing continuous characters. *Proceedings of the Royal Society of London Series B* 269:143–149.
- Weiblen G.D. and Bush G.L. 2002. Speciation in fig pollinators and parasites. *Molecular Ecology* 11:1573–1578.
- Weir J.T. and Schluter D. 2007. The latitudinal gradient in recent speciation and extinction rates of birds and mammals. *Science* 315:1574–1576.
- Wiens J.J. 2011. The causes of species richness patterns across space, time, and clades and the role of “ecological limits”. *Quarterly Review of Biology* 86:75–96.
- Williams G.C. 1966. *Adaptation and natural selection*. Princeton University Press, New Jersey.
- Wilson A.W., Binder M., and Hibbett D.S. 2011. Effects of gasteroid fruiting body morphology on diversification rates in three independent clades of fungi estimated using binary state speciation and extinction analysis. *Evolution* 65.
- Wilson E.O. 1959. Adaptive shift and dispersal in a tropical ant fauna. *Evolution* 13:112–144.
- Wilson E.O. 1961. The nature of the taxon cycle in the Melanesian ant fauna. *The American Naturalist* 95:169–193.
- Winger B.M., Lovette I.J., and Winker D.W. 2012. Ancestry and evolution of seasonal migration in the Parulidae. *Proceedings of the Royal Society of London Series B* 279:610–618.
- Yule G.U. 1925. A mathematical theory of evolution, based on the conclusions of Dr. J.C. Willis, FRS. *Philosophical Transactions of the Royal Society B: Biological Sciences* 21:21–87.

## APPENDIX A

### SUPPLEMENTARY INFORMATION TO CHAPTER 2

---

#### A.1 ROOT-STATE CALCULATIONS

Under BiSSE, at the root of a tree,  $R$ , we have the two probabilities  $D_{R0}$  and  $D_{R1}$ , corresponding to the possible character states at the root. The overall likelihood must sum over the probabilities that the root was in each state. In Schluter *et al.* (1997), the  $D$ s at the root were weighted evenly, which assumes that the lineage arose out of a group with 50% of taxa in each state, therefore potentially being out of equilibrium with the inferred model of evolution. The model parameters provide some knowledge, however. For example, if transitions away from character state 0 are much more frequent than the reverse ( $q_{01} > q_{10}$ ) and if speciation/extinction rates do not depend strongly on the character state, then we would expect that the system is more likely to be in state 1 than in state 0 at any point in time, including at the root. In Maddison *et al.* (2007), the information provided by the model was used to weight the  $D$ s at the root by the equilibrium frequencies for the character states given by the model (following Maddison and Maddison 2006). This implicitly assumes that a sufficient amount of time has passed prior to the root, so that the root state can be assumed to be a random draw from an equilibrium distribution. However, this assumption does not account for cases where the traits are novel or have yet to reach equilibrium.

Here, we treat the root state as a nuisance parameter (Gelman *et al.*, 1995) and use an alternative root assignment that weights each root state according to its probability of giving rise to the extant data, given the model parameters and the tree. This probability is given by the likelihood given that the root is in state  $i$  divided by the sum of the

likelihoods over both root states,  $D_{Ri}/(D_{R0} + D_{R1})$ . The overall likelihood is then:

$$D_R = D_{R0} \frac{D_{R0}}{D_{R0} + D_{R1}} + D_{R1} \frac{D_{R1}}{D_{R0} + D_{R1}} \quad (\text{A.1})$$

As a test case, consider a tree that consists of a single branch with an infinitesimally short branch length where the single extant taxon is in state 0. Assuming that none of the transition parameters is very large, then  $D_{R0}$  is nearly one and  $D_{R1}$  is nearly zero. Assigning the root state according to the probability of the root leading to the data, as we do in equation (A.1), we infer that there is nearly a 100% probability that the root was in state 0, and the overall probability of the data given the model is nearly one. Assigning the root state uniformly, we would be assuming that there is a 50% probability that the root state was 1, even though we know this to be impossible (more precisely, it has an infinitesimally small probability of being true given the infinitesimally short branch and the fact that the single extant taxon is in state 0). Similarly, assigning the root state according to the equilibrium distribution would give some non-zero probability to the root being in state 1, unless  $q_{01} = 0$ .

Furthermore, if there is directional change in the character, assigning the root to the equilibrium distribution incorrectly forces the character to take on the value that it will have in the long-term future, not what it is likely to have been in the past. For example, if all organisms started in state 0 and are evolving into state 1 ( $q_{01} \gg q_{10}$ , with all else equal), the equilibrium distribution method will incorrectly assign the root to state 1, even if most extant species are still in state 0. By contrast, assigning the root states according to their relative likelihoods of explaining the data will assign a high probability on the root having been in state 0 when most species are in state 0. Equation (A.1) has the further advantage that the only quantities needed for its calculation are  $D_{R0}$  and  $D_{R1}$ , which are already known once BiSSE has traversed the tree. Goldberg and Igić (2008) have recently explored the effect of root state on BiSSE calculations and found that they can have a strong effect on conclusions, especially where character change is unidirectional. Our approach (equation (A.1)) can approximate the ancestral root state and result in

reasonable character change rate estimates for the situations described above. In practice, sensitivity to the root state is easily detected by comparing  $D_{R0}$  and  $D_{R1}$ .

## A.2 CHARACTER-INDEPENDENT MODEL

When the speciation and extinction rates do not depend on a character, our likelihood calculations reduce to existing models of character-independent evolution. Analytical solutions for these models are known, removing the need to use numerical approaches to calculating the likelihoods.

### A.2.1 *Skeletal trees*

With character-independent speciation rates, the skeletal-tree likelihoods reduce to the method of Nee *et al.* (1994b). The character-independent analogues of equation (2.1) are

$$\frac{dD_N}{dt} = -(\lambda + \mu)D_N(t) + 2\lambda E(t)D_N(t) \quad (\text{A.2a})$$

$$\frac{dE}{dt} = \mu - (\lambda + \mu)E(t) + \lambda E(t)^2 \quad (\text{A.2b})$$

(Maddison *et al.*, 2007), where  $\lambda$  and  $\mu$  are the character-independent speciation and extinction rates. If a fraction,  $f$ , of all species are sampled in the phylogeny, then the initial conditions are  $E(0) = 1 - f$  and  $D(0) = f$ . It is possible to derive an analytical solution to equations (A.2) describing changes along a single branch. Using the initial condition  $E(0) = 1 - f$ , the solution to equation (A.2b) for the extinction rate is

$$E(t) = 1 - \frac{f(\lambda - \mu)}{f\lambda - e^{(\lambda - \mu)t}(\mu - \lambda(1 - f))} \quad \text{if } \lambda \neq \mu \quad (\text{A.3a})$$

$$E(t) = \frac{1 - f + f\lambda t}{1 - f\lambda t} \quad \text{if } \lambda = \mu \quad (\text{A.3b})$$

Substituting equation (A.3) into equation (A.2a) and solving for  $D_N(t)$  gives

$$D_N(t) = e^{-(\lambda-\mu)(t-t_N)} \frac{(f\lambda - e^{-(\lambda-\mu)t_N}(\mu - \lambda(1-f)))^2}{(f\lambda - e^{-(\lambda-\mu)t}(\mu - \lambda(1-f)))^2} D_N(t_N) \quad \text{if } \lambda \neq \mu \quad (\text{A.4a})$$

$$D_N(t) = \frac{(1 + f\lambda t_N)^2}{(1 + f\lambda t)^2} D_N(t_N) \quad \text{if } \lambda = \mu, \quad (\text{A.4b})$$

where  $t_N$  represents the time depth (since the present) of node  $N$ . Equations (A.3) and (A.4) reduce to equations (9) and (10) in Maddison *et al.* (2007) if sampling is complete ( $f = 1$ ).

These equations can be used as in Maddison *et al.* (2007) to compute the likelihood for the entire phylogeny:

$$D_R(t_R) = \left[ \prod_{k=1}^{2n} e^{-(\lambda-\mu)(t_{k,b}-t_{k,t})} \left( \frac{f\lambda - e^{-(\lambda-\mu)t_{k,t}}(\mu - \lambda(1-f))}{f\lambda - e^{-(\lambda-\mu)t_{k,b}}(\mu - \lambda(1-f))} \right)^2 \right] \lambda^n \quad \text{if } \lambda \neq \mu \quad (\text{A.5a})$$

$$D_R(t_R) = \left[ \prod_{k=1}^{2n} \frac{(1 + f\lambda t_{k,t})^2}{(1 + f\lambda t_{k,b})^2} \right] \lambda^n \quad \text{if } \lambda = \mu \quad (\text{A.5b})$$

where  $t_{k,b}$  and  $t_{k,t}$  are the times at the base and tip of the  $k$ th branch, respectively, and the product is taken over all  $2n$  branches for a tree containing  $n$  nodes.

Equation (A.5) is consistent with the results from Nee *et al.* (1994b) for the character-independent case. Nee *et al.* (1994b) does not explicitly give equations for the probability of a sampled phylogeny, given a speciation rate, extinction rate and sampling probability, so we state them here. Using the notation of Nee *et al.* (1994b), the probability of the data is

$$L = (N-1)! f^{N-2} \lambda^{N-2} \left( \prod_{i=3}^N P_s(t_i, T) \right) (1 - u_s(x_2))^2 \prod_{i=3}^N (1 - u_s(x_i)) \quad (\text{A.6})$$

where  $N$  is the number of tips in the phylogeny,  $x_i$  is the time between the present and the node that splits the phylogeny into  $i$  branches (so that  $x_2$  is the distance to the root),  $P_s(t_i, T)$  is the probability that a lineage originating at time  $t_i$  leaves at least one descendant at the present, time  $T$  (given by Nee *et al.*'s (1994) equation (34)), and  $u_s(t)$

is

$$u_s(t) = 1 - \frac{1 - a}{f e^{rx_i} - a + 1 - f} \quad (\text{A.7})$$

where  $a = \mu/\lambda$  and  $r = \lambda - \mu$ . Equation (A.7) can be derived from equations (29) and (33) in Nee *et al.* (1994b). After some algebra, equation (A.5a) can be shown to be equal to equation (A.6), after conditioning on the existence of a root node and two surviving lineages (see Maddison *et al.* 2007 for similar calculations with  $f = 1$ ).

### A.2.2 Terminally unresolved trees

To calculate the likelihood for terminally unresolved trees, we must first calculate the probability of unresolved clades given the speciation and extinction rates. We could re-derive  $\mathbf{Q}$  and  $\mathbf{x}$ , ignoring character state changes, which now do not affect diversification, and use equation (2.3) to compute the probability of the clade. However, without transitions, this can be viewed as a single birth-death process, for which an analytical solution is available. The probability of  $k$  lineages arising and surviving to the present from a single ancestor over a period of time  $t$  is given by (Nee *et al.*, 1994b):

$$P(k, t) = \frac{\lambda - \mu}{\lambda - \mu e^{-(\lambda - \mu)t}} (1 - u(t)) u(t)^{k-1} \quad k > 0, \lambda \neq \mu \quad (\text{A.8a})$$

$$P(k, t) = \frac{(t\lambda)^{i-1}}{(1 + t\lambda)^{i+1}} \quad k > 0, \lambda = \mu \quad (\text{A.8b})$$

where  $u(t)$  is  $u_s(t)$  from equation (A.7) with  $f = 1$ . If an unresolved clade has  $n$  species and originated at time  $t_N$ , then  $P(n, t_N)$  can be used for  $D_N(t_N)$  in equation (9) of Maddison *et al.* (2007). This approach has been used by Rabosky *et al.* (2007) to estimate speciation and extinction rates from a terminally unresolved lizard phylogeny.

## APPENDIX B

### SUPPLEMENTARY INFORMATION TO CHAPTER 3

---

#### B.1 SINGLE CHARACTER DERIVATION

Here, I derive equation (3.3): the partial differential equation that describes changes in the probability of extinction of a lineage in state  $x$  over time  $t$ ,  $E(x, t)$ , under QuaSSE. This derivation parallels both that of BiSSE (Maddison *et al.*, 2007) and the Kolmogorov backward differential equation (Allen, 2003). Starting with equation (3.2), subtracting  $E(x, t)$  from both sides and dropping remaining terms that are of order  $(\Delta t)^2$  gives

$$\begin{aligned} E(x, t + \Delta t) - E(x, t) = & \mu(x)\Delta t + \lambda(x)\Delta t \left[ \int_{-\infty}^{\infty} g(z, t|x, t + \Delta t)E(z, t)dz \right]^2 \\ & - (\lambda(x) + \mu(x))\Delta t \int_{-\infty}^{\infty} g(z, t|x, t + \Delta t)E(z, t)dz \\ & + \int_{-\infty}^{\infty} g(z, t|x, t + \Delta t)E(z, t)dz - E(x, t) + O(\Delta t^2). \end{aligned} \quad (\text{B.1})$$

Next, note that because  $g$  is a probability distribution function it must integrate to 1 over all possible future character states,  $\int_{-\infty}^{\infty} g(z, t|x, t + \Delta t)dz = 1$ , so that

$$E(x, t) = \int_{-\infty}^{\infty} g(z, t|x, t + \Delta t)E(x, t)dz. \quad (\text{B.2})$$



Replacing the final  $E(x, t)$  term in equation (B.1) with equation (B.2) and dividing both sides by  $\Delta t$  gives

$$\begin{aligned} \frac{E(x, t + \Delta t) - E(x, t)}{\Delta t} = & \mu(x) + \lambda(x) \left[ \int_{-\infty}^{\infty} g(z, t|x, t + \Delta t) E(z, t) dz \right]^2 \\ & - (\lambda(x) + \mu(x)) \int_{-\infty}^{\infty} g(z, t|x, t + \Delta t) E(z, t) dz \\ & + \frac{1}{\Delta t} \int_{-\infty}^{\infty} g(z, t|x, t + \Delta t) (E(z, t) - E(x, t)) dz + O(\Delta t). \end{aligned} \quad (\text{B.3})$$

We then take the limit  $\Delta t \rightarrow 0$ . In this limit, the left hand side becomes the partial derivative  $\partial E(x, t)/\partial t$ . Because wild jumps are not allowed and we are considering an infinitesimally small time period, the term  $E(z, t)$  can be expanded as a Taylor series in  $z$  around the point  $z = x$ :

$$E(z, t) = E(x, t) + (z - x) \frac{\partial E(x, t)}{\partial x} + \frac{(z - x)^2}{2} \frac{\partial^2 E(x, t)}{\partial x^2} + O((z - x)^3) \quad (\text{B.4})$$

Using this expansion, the third term on the right hand side of equation (B.3) can then be written

$$\begin{aligned} \lim_{\Delta t \rightarrow 0} (\lambda(x) + \mu(x)) \int_{-\infty}^{\infty} g(z, t|x, t + \Delta t) \times \\ \left( E(x, t) + (z - x) \frac{\partial E(x, t)}{\partial x} + \frac{(z - x)^2}{2} \frac{\partial^2 E(x, t)}{\partial x^2} + O((z - x)^3) \right) dz. \end{aligned}$$

In the limit  $\Delta t \rightarrow 0$ , transitions any distance away from  $x$  become increasingly unlikely. That is,

$$\lim_{\Delta t \rightarrow 0} g(z, t|x, t + \Delta t) = \delta(z - x),$$

where  $\delta(x)$  is the Dirac delta function, concentrating all probability density on the point  $z = x$ . This means that

$$\lim_{\Delta t \rightarrow 0} \int_{-\infty}^{\infty} (z - x)^k g(z, t|x, t + \Delta t) dz = 0, \quad k > 0$$

and the third term of equation (B.3) can be rewritten

$$\lim_{\Delta t \rightarrow 0} (\mu(x) + \lambda(x)) \int_{-\infty}^{\infty} g(z, t|x, t + \Delta t) E(z, t) dz = (\mu(x) + \lambda(x)) E(x, t). \quad (\text{B.5})$$

The same logic applied to the second term of equation (B.3) gives

$$\lim_{\Delta t \rightarrow 0} \lambda(x) \left[ \int_{-\infty}^{\infty} g(z, t|x, t + \Delta t) E(z, t) dz \right]^2 = \lambda(x) E(x, t)^2. \quad (\text{B.6})$$

After substituting in the Taylor expansion from equation (B.4), the fourth term of equation (B.3) becomes

$$\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} g(z, t|x, t + \Delta t) \left( (z - x) \frac{\partial E(x, t)}{\partial x} + \frac{(z - x)^2}{2} \frac{\partial^2 E(x, t)}{\partial x^2} + O((z - x)^3) \right) dz.$$

This can be simplified using the diffusion conditions from equation (3.1) to give

$$\phi(x, t) \frac{\partial E(x, t)}{\partial x} + \frac{\sigma^2(x, t)}{2} \frac{\partial^2 E(x, t)}{\partial x^2}. \quad (\text{B.7})$$

Substituting equations (B.5), (B.6), and (B.7) into equation (B.3) gives the partial differential equation (3.3).

## B.2 MULTIVARIATE CHARACTER DERIVATION

I start with the two-character case, from which the extension to an arbitrary number of characters immediately follows. Suppose we have two characters,  $x$  and  $y$ . Let  $g(a, b, t|x, y, t + \Delta t)$  be the probability density that the character changes from  $x, y$  at time  $t + \Delta t$  to state  $a, b$  at time  $t$ , where  $t$  is closer to the present than  $t + \Delta t$  ( $0 < t < t + \Delta t$ ). The functions  $E$  and  $D_N$  are now functions of both character variables;  $E(x, y, t)$  and  $D_N(x, y, t)$ .

Continuing as for the single character case,  $E(x, y, t + \Delta t)$  can be written

$$\begin{aligned}
E(x, y, t + \Delta t) = & \mu(x, y)\Delta t \\
& + (1 - \mu(x, y)\Delta t)\lambda(x, y)\Delta t \times \\
& \left[ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(a, b, t|x, y, t + \Delta t)E(a, b, t)dad b \right]^2 \\
& + (1 - \mu(x, y)\Delta t)(1 - \lambda(x, y)\Delta t) \times \\
& \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(a, b, t|x, y, t + \Delta t)E(a, b, t)dad b \\
& + O(\Delta t^2).
\end{aligned} \tag{B.8}$$

Dropping terms of  $O(\Delta t^2)$ , subtracting  $E(x, y, t + \Delta t)$  from both sides, dividing by  $\Delta t$ , and rearranging using the fact that

$$E(x, y, t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(a, b, t|x, y, t + \Delta t)E(x, y, t)dad b,$$

we have

$$\begin{aligned}
\frac{E(x, y, t + \Delta t) - E(x, y, t)}{\Delta t} = & \mu(x, y) \\
& + \lambda(x, y) \left[ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(a, b, t|x, y, t + \Delta t)E(a, b, t)dad b \right]^2 \\
& - (\mu(x, y) + \lambda(x, y)) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(a, b, t|x, y, t + \Delta t)E(a, b, t)dad b \\
& + \frac{1}{\Delta t} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(a, b, t|x, y, t + \Delta t)(E(a, b, t) - E(x, y, t))dad b \\
& + O(\Delta t). \tag{B.9}
\end{aligned}$$

To simplify equation (B.9), we need an expression for  $E(a, b, t)$  in terms of the original location  $(x, y)$ . This can be expanded as a Taylor series in two variables around the

point  $(a = x, b = y)$ :

$$\begin{aligned}
 E(a, b, t) = & E(x, y, t) \\
 & + (a - x) \frac{\partial E}{\partial x} + \frac{(a - x)^2}{2} \frac{\partial^2 E}{\partial x^2} \\
 & + (b - y) \frac{\partial E}{\partial y} + \frac{(b - y)^2}{2} \frac{\partial^2 E}{\partial y^2} \\
 & + (a - x)(b - y) \frac{\partial^2 E}{\partial x \partial y} + O((\Delta x, \Delta y)^3)
 \end{aligned} \tag{B.10}$$

where  $O((\Delta x, \Delta y)^3)$  includes the terms  $O((a - x)^3)$ ,  $O((b - y)^3)$ ,  $O((a - x)^2(b - y))$ , and  $O((a - x)(b - y)^2)$ .

We take the limit  $\Delta t \rightarrow 0$  for equation (B.9) and consider each term in sequence. Using the same logic as the one character case,

$$\lim_{\Delta t \rightarrow 0} g(a, b, t | x, y, t + \Delta t) = \delta(a - x) \delta(b - y),$$

so

$$\lim_{\Delta t \rightarrow 0} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (a - x)^{k_x} (b - y)^{k_y} g(a, b, t | x, y, t + \Delta t) da db = 0, \quad \text{for } k_x, k_y > 0.$$

With this, the second and third terms of equation (B.9) can be written

$$\lim_{\Delta t \rightarrow 0} \lambda(x, y) \left[ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(a, b, t | x, y, t + \Delta t) E(x, y, t) da db \right]^2 = \lambda(x, y) E(x, y, t)^2. \tag{B.11}$$

and

$$\begin{aligned}
 \lim_{\Delta t \rightarrow 0} (\mu(x, y) + \lambda(x, y)) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(a, b, t | x, y, t + \Delta t) E(x, y, t) da db = \\
 (\mu(x, y) + \lambda(x, y)) E(x, y, t) \tag{B.12}
 \end{aligned}$$

After substituting in the Taylor expansion (B.10) into the fourth term of equation (B.9) and taking the limit  $\Delta t \rightarrow 0$ , we have

$$\begin{aligned} \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(a, b, t|x, y, t + \Delta t) \times \\ \left( (a-x) \frac{\partial E}{\partial x} + \frac{(a-x)^2}{2} \frac{\partial^2 E}{\partial x^2} + (b-y) \frac{\partial E}{\partial y} + \frac{(b-y)^2}{2} \frac{\partial^2 E}{\partial y^2} \right. \\ \left. + (a-x)(b-y) \frac{\partial^2 E}{\partial x \partial y} + O((\Delta x, \Delta y)^3) \right) da db. \end{aligned}$$

A similar set of assumptions to those in equation (3.1) can be made. Consider first the derivatives involving just one character. Rearranging the first of these gives:

$$\begin{aligned} \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(a, b, t|x, y, t + \Delta t) (a-x) \frac{\partial E}{\partial x} da db = \\ \frac{\partial E}{\partial x} \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} (a-x) \int_{-\infty}^{\infty} g(a, b, t|x, y, t + \Delta t) db da. \end{aligned}$$

Defining

$$\int_{-\infty}^{\infty} g(a, b, t|x, y, t + \Delta t) db = g(a, t|x, t + \Delta t),$$

so that integrating over all transitions in  $y$  gives the transition probability density function for  $x$ . The equation above then becomes

$$\frac{\partial E}{\partial x} \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} (a-x) g(a, t|x, t + \Delta t) da$$

where the term within the limit is identical to equation (3.1a). With similar manipulation for the other terms, the diffusion conditions become

$$\phi_x(x, y, t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (a - x) g(a, b, t | x, y, t + \Delta t) da db \quad (\text{B.13a})$$

$$\phi_y(x, y, t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (b - y) g(a, b, t | x, y, t + \Delta t) da db \quad (\text{B.13b})$$

$$\sigma_{x,x}(x, y, t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (a - x)^2 g(a, b, t | x, y, t + \Delta t) da db \quad (\text{B.13c})$$

$$\sigma_{y,y}(x, y, t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (b - y)^2 g(a, b, t | x, y, t + \Delta t) da db \quad (\text{B.13d})$$

$$\sigma_{x,y}(x, y, t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (a - x)(b - y) g(a, b, t | x, y, t + \Delta t) da db \quad (\text{B.13e})$$

$$0 = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (a - x)^{k_x} (b - y)^{k_y} g(a, b, t | x, y, t + \Delta t) da db, \quad (\text{B.13f})$$

where condition (B.13f) holds for  $k_x + k_y > 2$ . The term  $\sigma_{x,y}(x, y, t)$  represents the instantaneous covariance between  $x$  and  $y$ .

Substituting equations (B.11), (B.12), and (B.13) into equation (B.9) and applying the limit to the left-hand side gives

$$\begin{aligned} \frac{\partial E(x, y, t)}{\partial t} = & \mu(x, y) + \lambda(x, y)E(x, y, t)^2 - (\lambda(x, y) + \mu(x, y))E(x, y, t) \\ & + \phi_x(x, y, t) \frac{\partial E(x, y, t)}{\partial x} + \phi_y(x, y, t) \frac{\partial E(x, y, t)}{\partial y} \\ & + \frac{\sigma_{x,x}(x, y, t)}{2} \frac{\partial^2 E(x, y, t)}{\partial x^2} + \frac{\sigma_{y,y}(x, y, t)}{2} \frac{\partial^2 E(x, y, t)}{\partial y^2} \\ & + \sigma_{x,y}(x, y, t) \frac{\partial^2 E(x, y, t)}{\partial x \partial y} \quad (\text{B.14}) \end{aligned}$$

and a similar process leads to the partial differential equation

$$\begin{aligned} \frac{\partial D_N(x, y, t)}{\partial t} = & 2\lambda(x, y)D_N(x, y, t)E(x, y, t) - (\lambda(x, y) + \mu(x, y))D_N(x, y, t) \\ & + \phi_x(x, y, t)\frac{\partial D_N(x, y, t)}{\partial x} + \phi_y(x, y, t)\frac{\partial D_N(x, y, t)}{\partial y} \\ & + \frac{\sigma_{x,x}(x, y, t)}{2}\frac{\partial^2 D_N(x, y, t)}{\partial x^2} + \frac{\sigma_{y,y}(x, y, t)}{2}\frac{\partial^2 D_N(x, y, t)}{\partial y^2} \\ & + \sigma_{x,y}(x, y, t)\frac{\partial^2 D_N(x, y, t)}{\partial x \partial y} \quad (\text{B.15}) \end{aligned}$$

This can be extended to an arbitrary number of characters to give equations (3.9) and (3.10).

Similar logic can also be used to derive equations when speciation, extinction and character transition functions also depend on the state of an additional binary character; let  $\lambda_i(x)$  and  $\mu_i(x)$  denote the speciation and extinction function in a continuous character state  $x$  while in the binary state  $i$ , where  $i = 0$  or  $1$ ,  $\phi_i(x, t)$  and  $\sigma_i^2(x, t)$  be the directional and diffusion functions, while in state  $i$ , and  $q_{ij}(x)$  be the rate of transition from binary state  $i$  to  $j$ , which may depend on the continuous trait. The variables become  $E_i(x, t)$  and  $D_{iN}(x, t)$ , which are the probability of extinction or of the lineage leading to node  $N$  (respectively) for the lineage in binary state  $i$  and continuous state  $x$  at time  $t$ . Following logic similar to above and in Maddison *et al.* (2007) yields the equations

$$\begin{aligned} \frac{\partial E_i(x, t)}{\partial t} = & \mu_i(x) + \lambda_i(x)E_i(x, t)^2 - (\lambda_i(x) + \mu_i(x) + q_{ij}(x))E_i(x, t) \\ & + q_{ij}(x)E_j(x, t) + \phi_i(x, t)\frac{\partial E_i(x, t)}{\partial x} + \frac{\sigma_i^2(x, t)}{2}\frac{\partial^2 E_i(x, t)}{\partial x^2} \quad (\text{B.16a}) \end{aligned}$$

$$\begin{aligned} \frac{\partial D_{iN}(x, t)}{\partial t} = & 2\lambda_i(x)D_{iN}(x, t)E_i(x, t) - (\lambda_i(x) + \mu_i(x) + q_{ij}(x))D_{iN}(x, t) \\ & + q_{ij}(x)D_{jN}(x, t) + \phi_i(x, t)\frac{\partial D_{iN}(x, t)}{\partial x} + \frac{\sigma_i^2(x, t)}{2}\frac{\partial^2 D_{iN}(x, t)}{\partial x^2}. \quad (\text{B.16b}) \end{aligned}$$

## APPENDIX C

### SUPPLEMENTARY INFORMATION TO CHAPTER 4

---

#### C.1 TUNING DIVERSITREE

Many of the models included in `diversitree` are computationally expensive. For example, all the state-dependent speciation and extinction models (abbreviated `xxSSE`) involve numerically solving systems of nonlinear differential equations for every branch in a tree (of which there are  $2n - 2$  for a tree with  $n$  species). This can lead to long computation times, but it is possible to tune the performance of `diversitree` by changing how these calculations are performed

To control the way calculations are carried out, every “make” function takes a `control` argument. When specified, this is a list of one or more tag/value pairs, such as

```
control=list(tag1=value1, tag2=value2, ...)
```

Below, for each class of models I list the possible tags and the values that they may take. As `diversitree` handles discrete and continuous traits in quite different ways, I describe the tuning options separately.

##### C.1.1 *Discrete traits*

`backend` This switches between different algorithms for solving the system of differential equations, and takes values `"deSolve"` (the default), `"cvcodes"`, and `"CVODES"` (quotes are required). For example, passing the argument `control=list(backend="CVODES")`



to a make function will use the "CVODES" backend. Keys never require quotes, while values do unless logical or numeric.

The "deSolve" backend uses the LSODA algorithm from the R package deSolve (Soetaert *et al.*, 2010) to solve the system of differential equations. This is a great general purpose ODE solver and is available on all R platforms. However, while the calculations for each branch are done entirely using compiled code, the calculations at nodes and substantial amounts of book-keeping are done in R code, which can be a bottleneck. See `safe` below for the other disadvantage of this backend.

The "cvodes" backend uses the CVODES algorithm (Cohen and Hindmarsh, 1996) from the sundials library of solvers (Hindmarsh *et al.*, 2005)<sup>5</sup>. Like the deSolve backend, only the integration is done in compiled code, with node calculations and book-keeping done in R. The "cvodes" backend is about 20% slower than "deSolve" for BiSSE, and is not available on Windows.

The "CVODES" backend also uses the CVODES algorithm. However, *all* calculations are carried out in compiled code. This option is not available for all model types. In particular, ClaSSE, BiSSE-ness, split models, and time-dependent models are not yet implemented. This backend is about 5 times faster than "deSolve" for BiSSE, but is also not available on Windows.

`safe` When using the "deSolve" backend, diversitree uses non-exported compiled functions within the deSolve package<sup>6</sup>. deSolve's function definitions may change between package versions, and if called can cause R to crash. To avoid this, if the installed deSolve version is not known to work, diversitree will fall back on a "safe" version, in which only exported deSolve functions are used.

---

<sup>5</sup>This actually uses the CVODE algorithm, not CVODES, as sensitivity calculations are not yet supported.

<sup>6</sup>Specifically, we use the compiled LSODA wrapper function `call_lsoda` directly, rather than the R function `lsoda`.

This is about 5 times slower than `safe=FALSE` for BiSSE<sup>7</sup>. This approach can be forced by specifying `safe=TRUE`, though there are few cases where this is desired. This option has no effect for the "cvcodes" and "CVCODES" backends.

`unsafe` This is the opposite of `safe`. When `TRUE`, even if the deSolve version is not known to work, the calculations will proceed anyway. This can crash R, or potentially produce incorrect results (though the latter is unlikely and can be checked by confirming with `safe=TRUE`). However, this option may be useful if the diversitree version lags behind the installed deSolve version. This is especially true for Windows users, for whom compilation of diversitree is often tricky. This option has no effect for the "cvcodes" and "CVCODES" backends.

`tol` This controls the degree of accuracy of the integration of each branch. By default a value of  $1e-8$  is used ( $10^{-8}$ ). Decreasing this value increases the accuracy of the calculations, but increases the required running time. Informally, the LSODA algorithm estimates errors,  $e$ , for each variable  $y$  and attempts to keep these errors smaller than  $tol \times y + tol$ . The CVCODES algorithm uses a similar error target,  $\sum_{i=1}^n \frac{1}{n} (e_i / (tol \times y_i + tol))^2$ , where  $e_i$  is the estimated error for the  $i$ th variable. If the requested accuracy is not possible, the calculations will fail with an error. Values below `sqrt(.Machine$double.eps)`, which is usually around  $1e-8$ , are optimistic. Note that because of error propagation, the error for the entire likelihood calculation will be substantially higher<sup>8</sup>.

`eps` At the end of each branch, diversitree checks the value of all “data” variables (in BiSSE-type models, these are the  $D$  variables). If any of these are smaller than “eps” diversitree splits the branch in two and runs the calculations on each half, repeating this until the desired accuracy is reached. This is useful on very

---

<sup>7</sup>The reason for the slowdown is due to looking up the memory address of the derivative function every time it is used. In diversitree, we create a wrapper that remembers the address as it will not change during an R session.

<sup>8</sup>The form  $tol \times y + tol$  in the above expressions arises because the algorithm *actually* uses  $rtol \times y + atol$  where  $rtol$  is a relative tolerance and  $atol$  is an absolute tolerance. However, these are not separately tunable in the current diversitree version.

long branches where speciation and extinction rates are similar, as negative  $D$  values can be produced. The default is  $\text{eps}=0$ , which enforces positive values. Specifying  $\text{eps}=-\text{Inf}$  will disable this check. In theory, small positive numbers may help in some difficult-to-fit models (again, with speciation rates close to extinction rates). However, in models with rapid character evolution, variables may never take values much above zero, which means that this criterion may never be satisfied for a given  $\text{eps} > 0$ , and the calculations will fail.

As an example, using a BiSSE likelihood function,

```
make.bisse(tree, states,
           control=list(backend="CVODES", tol=1e-5, eps=-Inf))
```

specifies that the "CVODES" backend will be used, error tolerances have increased to 0.00001, and positive  $D$  values are no longer being enforced. This will run faster than the default options, but the answers will be less accurate.

The elements of the `control` argument are checked to make sure that the values make sense, but misspellings of element names are not checked. For example, specifying `control=list(back="CVODES")` will still use the the default deSolve algorithm LSODA solver for the calculations, with no warning given.

**Mk2 & Mkn** — By default, likelihoods under these models are computed differently to the xxSSE models. Because there are no  $E$  variables to compute, it is feasible to generate a matrix  $P_{ij}(\Delta t)$  that describes the probability of a transition from state  $i$  to state  $j$  over time  $\Delta t$  simultaneously for all branches. This is not possible in BiSSE and other xxSSE models as the starting and ending times matter (rather than simply the elapsed time) because of the influence of  $E(t)$ . By default, diversitree computes this matrix for all branch lengths in the tree and then uses this matrix to quickly compute the likelihood. Almost all of these calculations are in compiled code and should be fairly quick.

For `mk2`, the  $P_{ij}(\Delta t)$  matrix can be computed exactly, and all of the above `control` options are ignored. For `mkn`, the calculation of  $P_{ij}(\Delta t)$  still requires numerical inte-

gration, and the options `backend`, `safe`, `unsafe`, and `tol` in the previous section are available. However, neither `backend="CVODES"` or `eps` make sense in this case, and will cause an error (`backend="CVODES"`) or be silently ignored (`eps`).

As the number of states increases, computing the transition probability matrix becomes expensive. This is particularly the case for `make.mkn.multitrait` where the number of possible states grows exponentially in the number of traits;  $n$  binary traits require  $2^n$  possible states, which is  $2^{2^n}$  elements in  $P_{ij}(\Delta t)$ . In such cases, there is a control option that changes the algorithm:

`method` Specifying `method="ode"` uses a branch-by-branch approach that avoids computing the transition probability matrix. When this is specified, the algorithm used is exactly the same as for BiSSE-type models, and all the control parameters in the previous section are available. The default is `method="exp"` (the name here derives from the approach used by “ape” to carry out the same calculations through matrix exponentiation — see Moler and Loan, 2003).

### C.1.2 *Continuous traits*

**QUASSE** — The options for tuning QuaSSE are not shared with any other model. Unfortunately, no matter what control parameters are specified, QuaSSE will be fairly slow compared with other models in `diversitree`.

`method` One of `"fftC"` or `"fftR"` to switch between C (relatively fast) and R (extremely slow) back-ends for the integration. Both use non-adaptive fast Fourier transform (FFT) based convolutions to solve the partial differential equations (see FitzJohn, 2010). Specifying `"fftC"` uses the “Fastest Fourier Transforms in the West” library (Frigo and Johnson, 2005), while specifying `"fftR"` method uses R’s built-in `fft` function. Specifying `"mol"` will use an experimental method-of-lines approach.

- `dt.max` Maximum time step to use for the integration. By default, this will be set to 1/1000 of the tree depth. Smaller values will slow down calculations, but improve accuracy.
- `nx` The number of bins into which the character space is divided (the default is 1024). Larger values will be slower and more accurate. For the "fftC" integration method, this must be an integer power of 2 (512, 2048, etc).
- `r` Scaling factor that multiplies `nx` for a "high resolution" section at the tips of the tree (the default is 4, giving a high resolution character space divided into 4,096 bins). This helps improve accuracy and makes it possible to allow for narrow initial probability distributions, which flatten out as time progresses towards the root. Larger values will be slower and more accurate. For the "fftC" integration method, this must be a small power of 2 (e.g., 2, 4, 8), so that `nx*r` is also a power of 2.
- `tc` This specifies when in the tree to switch the resolution from the high resolution section specified by `r` to the lower resolution. Zero corresponds to the present, larger numbers moving towards the root. By default, this happens at 10% of the tree depth (so the default is `0.1 * max(branching.times(tree))`). Smaller values will be faster, but less accurate, and must be specified in time units.

BROWNIAN MOTION AND ORNSTEIN-UHLENBECK — These models take two possible control parameters:

- `method` This switches between two different algorithms for computing likelihoods. The default, `method="vcv"`, uses a variance-covariance matrix approach, as implemented in the `geiger` and `ape` packages (Paradis *et al.*, 2004; Harmon *et al.*, 2008). This is very fast for small trees.
- For large trees (hundreds of species), the matrix calculations make computing the likelihood very expensive. In this case, specifying `method="pruning"`

uses an algorithm more like that used in BiSSE and QuaSSE likelihood calculations, where each branch is treated separately. This algorithm is similar to that described in Felsenstein (1973); however, I have not seen this version described elsewhere and so describe it in section C.2.

`backend` When `method="pruning"` is selected, this controls which of two methods of calculation is used; the default, `backend="R"`, computes the likelihood entirely within R code, which should be fairly robust. Specifying `backend="C"` uses the same algorithm, but entirely implemented in C, which will be faster, but is less extensively tested.

### C.1.3 “Split” models

All models that allow for different regions of the tree to have different rates (such as `make.bisse.split` and `make.quasse.split`) have one additional control parameter:

`caching.branches` When `TRUE`, this will try to minimise recalculating the likelihood for regions of the tree where parameters have not changed. Every function evaluation, the values at the beginning and end of each branch, plus the parameters, are stored. If in the next evaluation both the parameters and initial conditions are unchanged, the previous base values are returned. Otherwise the branch is calculated as usual. This is useful during MCMC updates, or with many ML search algorithms, where only a fraction of parameters are changed at once. In particular, the default MCMC algorithm (slice sampling; Neal, 2003) updates each parameter separately, so without this set most calculation time is wasted. For computationally intensive models, such as QuaSSE, this can make MCMC analyses with split models take the same time to take one step (updating all parameters) as non-split models, despite the increase in the dimensionality of parameter space. For less intensive models, such as BiSSE, the overhead of the caching process can erase some of the time savings.

## C.2 A FASTER ALGORITHM FOR BM & OU LIKELIHOOD CALCULATIONS

Here, I describe an algorithm for computing likelihoods under Brownian motion and Ornstein-Uhlenbeck that can be considerably faster than conventional approaches. The algorithm here uses the pruning algorithm of Felsenstein (1981), and is closely related to the algorithm presented in Felsenstein (1973). It differs mostly in presentation, using the same approach as the xxSSE models. The conventional algorithm, as implemented in `fitContinuous` in the `geiger` package (Harmon *et al.*, 2008) and the REML algorithm in `ace` in the `ape` package (Paradis *et al.*, 2004) uses the phylogenetic variance-covariance matrix to compute the likelihood of a rate of diffusion by estimating the probability of the observed trait data under a multivariate-normal distribution. In what follows, I will refer to this as the “vcv” algorithm due the central role of the phylogenetic variance-covariance matrix.

### C.2.1 *Brownian motion*

Using the notation of QuaSSE (FitzJohn, 2010), let  $D_N(x, t)$  be the probability of the observed trait distribution for some lineage  $N$ , given that it is in state  $x$  at time  $t$ . Unlike QuaSSE, this does not account for the distribution of branching times, which are assumed to be unaffected by the trait state. Assuming that  $D$  will always be Gaussian, it can be characterised by a vector of three elements: a mean  $\mu_N$ , variance  $V_N$ , and a normalising factor  $1/z_N$ , such that

$$D_N(x, t) = \frac{1}{z_N} \frac{1}{\sqrt{2\pi V_N}} \exp\left(-\frac{(x - \mu_N)^2}{2V_N}\right). \quad (\text{C.1})$$

For example, the initial distribution at a tip,  $D_N(x, 0)$ , will be have  $\mu_N$  as the mean extant trait value. If the trait value is known without error, then  $V_N = 0$  and  $D_N(x, 0)$  is a delta function at  $\mu_N$ . Alternatively,  $V_N$  can be set greater than zero to capture measurement

error or uncertainty in the mean  $\mu_N$ . The normalising factor,  $1/z_N$ , will be 1 so that  $\int D_N(x, 0)dx = 1$ .

Consider a branch that has a tip at time  $t$  and a base at time  $t + \Delta t$  (further back in time than  $t$ ), so that the branch has length  $\Delta t$ . Brownian motion has a normal transition probability density function; given a rate of diffusion of  $\sigma^2$ , the probability density of state  $y$  at a branch tip (time  $t$ ) given a state of  $x$  at the branch base (time  $t + \Delta t$ ) is:

$$g(y, t|x, t + \Delta t) = \frac{1}{\sqrt{2\pi\Delta t\sigma^2}} \exp\left(-\frac{(x - y)^2}{2\Delta t\sigma^2}\right). \quad (\text{C.2})$$

Given  $D_N(x, t)$  at the tip of the branch, we can compute  $D_N(x, t + \Delta t)$  at the base of that branch as

$$D_N(x, t + \Delta t) = \int_{-\infty}^{\infty} g(y, t|x, t + \Delta t) D_N(y, t) dy, \quad (\text{C.3})$$

which is the probability that we move from  $x$  to  $y$  multiplied by the probability of the data at  $y$ , integrated over possible values of  $y$  (this is the convolution of  $D_N$  and  $g$ ). This has the solution

$$D_N(x, t + \Delta t) = \frac{1}{z_N} \frac{1}{\sqrt{2\pi(V_N + \sigma^2\Delta t)}} \exp\left(-\frac{(x - \mu_N)^2}{2(V_N + \sigma^2\Delta t)}\right), \quad (\text{C.4})$$

which is a Gaussian with a mean  $\mu_N$ , variance  $V_N + \sigma^2\Delta t$ , and normalising factor  $1/z_N$ . Note that only the variance is changed by this calculation, and it is always increased.

At the node  $N'$  that is the parent of the lineages leading to nodes  $N$  and  $M$ , we have two Gaussian functions with means  $\mu_N$  and  $\mu_M$ , variances  $V_N$  and  $V_M$ , and normalising factors  $1/z_N$  and  $1/z_M$ . Both daughter lineages share the same state as their parent immediately following speciation. Therefore, at the node the probability density of the data given that we are in some state  $x$  is

$$D_{N'}(x, t) = D_N(x, t) D_M(x, t), \quad (\text{C.5})$$



or

$$D_{N'}(x, t) = \frac{1}{z_N z_M} \frac{\exp\left(-\frac{(\mu_N - \mu_M)^2}{2(V_N + V_M)}\right)}{\sqrt{2\pi(V_N + V_M)}} \frac{1}{\sqrt{2\pi \frac{V_N V_M}{V_N + V_M}}} \exp\left(-\frac{(x - \frac{\mu_N V_M + \mu_M V_N}{V_N + V_M})^2}{2 \frac{V_N V_M}{V_N + V_M}}\right), \quad (C.6)$$

which is Gaussian with mean  $\frac{\mu_N V_M + \mu_M V_N}{V_N + V_M}$ , variance  $\frac{V_N V_M}{V_N + V_M}$ , and normalising factor  $\frac{1}{z_N z_M} \times \frac{\exp\left(-\frac{(\mu_N - \mu_M)^2}{2(V_N + V_M)}\right)}{\sqrt{2\pi(V_N + V_M)}}$ . See also equations (7)–(15) in Felsenstein (1973), which involve similar terms.

With these two operations (moving down a branch and combining at nodes) we can move to the root of the tree, where we have a distribution  $D_R(x, t_R)$ , which is Gaussian with mean  $\mu_R$ , variance  $V_R$ , and normalising factor  $z_R$ . Diversitree implements four possible root treatments. First, following Pagel (1994) we can specify a flat prior on the root state, computing the likelihood as

$$\int_{-\infty}^{\infty} D_R(x, t_R) dx = \frac{1}{z_R}. \quad (C.7)$$

Second, following FitzJohn (2010), we can weight the  $D$  function by the probability of observing the data:

$$\int_{-\infty}^{\infty} D_R(x, t_R) \frac{D_R(x, t_R)}{\int_{-\infty}^{\infty} D_R(y, t_R) dy} dx = \frac{1}{z_R} \frac{1}{2\sqrt{\pi V_R}}, \quad (C.8)$$

where the second fraction on the right hand side comes from integrating the product of two Gaussians. Third, we can evaluate  $D_R(x, t)$  at its ML value. This is  $x = \mu_R$ , giving

$$\max_x [D_R(x, t_R)] = \frac{1}{z_R} \frac{1}{\sqrt{2\pi V_R}}. \quad (C.9)$$

Finally, the user can supply a character state at the root  $x_R$  at which  $D_R(x, t_R)$  will be evaluated. At present this is assumed to be known without error, but in principle a distribution could be used here.

When the ML approach is used (equation c.9), this algorithm produces identical likelihoods to the vcv algorithm. I have proven this statement for a three-species tree and confirmed it numerically for some larger trees.

### c.2.2 Ornstein-Uhlenbeck

For an Ornstein-Uhlenbeck process (OU), the above algorithm can again be used after altering the along-branch calculations (i.e., equations c.2 and c.4). In addition to the diffusion parameter,  $\sigma^2$ , let  $\theta$  be the “optimum” character state, towards which traits are pulled and let  $\alpha$  be the strength of restoring force bringing traits back to  $\theta$ . The transition probability density function for OU (probability density of state  $y$  at time  $t$  given we were in state  $x$  at time  $t + \Delta t$  is normal with mean  $e^{-\alpha t}(x - \theta) + \theta$ , variance  $\frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha t})$ , such that

$$g(y, t|x, t + \Delta t) = \frac{1}{\sqrt{\pi \frac{\sigma^2}{\alpha}(1 - e^{-2\alpha \Delta t})}} \exp\left(-\frac{e^{-\alpha \Delta t}(x - \theta) + \theta - y}{\frac{\sigma^2}{\alpha}(1 - e^{-2\alpha \Delta t})}\right) \quad (\text{C.10})$$

(Karlin and Taylor, 1981). The solution to the convolution (c.3) using equation (C.10) as the transition probability density function gives a Gaussian with mean  $e^{\alpha \Delta t}(\mu_N - \theta) + \theta$ , variance  $\frac{(e^{2\alpha \Delta t} - 1)\sigma^2}{2\alpha} + e^{2\alpha \Delta t} V_N$ , and normalising factor  $e^{\Delta t \alpha} / z_N$ .

The same treatment at nodes applies (equation c.6), and the root calculations in equations (c.7)–(c.9) can be used. However, the “flat prior” root treatment (equation c.7) appears to give likelihoods that are not directly comparable to BM in a likelihood ratio test, with a hugely inflated type I error.

### c.2.3 Performance

The above calculations can often be much faster than the vcv-based calculations. I generated random Yule trees using diversitree’s `tree.yule` function with a speciation rate of 0.1 up to 16, 32, 64, ..., 4096 species. For each tree, I simulated traits under Brownian motion using diversitree’s `sim.character` function with a diffusion parame-

ter,  $\sigma^2$ , of 0.1. Note that both the rate of speciation and diffusion rate of character state evolution are arbitrary, as rescaling the edge lengths or trait distributions is equivalent to changing the parameters. To measure the time taken to compute likelihoods, for each tree I then computed the likelihood of the true diffusion parameter repeatedly until the total evaluation time exceeded 0.5 s (this ranged from a single evaluation to several thousand evaluations). I repeated this on 10 different simulated trees and traits.

I tested three algorithms; first the vcv algorithm that is based on the code in the *geiger* package. I modified this algorithm slightly to avoid inverting the phylogenetic covariance matrix and other matrix calculations for every likelihood calculation in the case where “measurement error” is assumed to be zero. This is the default algorithm used by `make.bm`, corresponding to passing in `control=list(method="vcv")`. Second, I used the pruning algorithm above, entirely coded in R. This is the algorithm selected by passing `control=list(method="pruning", backend="R")` to `make.bm`. Third, I used the pruning algorithm, coded in C. This is selected by specifying `backend="C"` rather than “R” in the control list above. This third algorithm should be most directly comparable to the vcv algorithm in terms of speed, as both involve primarily compiled code. To avoid the optimisation I made in the vcv algorithm, and to measure performance when measurement error is included, I repeated the above, but included very small measurement errors ( $10^{-7}$ ). Timings were carried out on a 2008 Mac Pro (2.8 GHz Intel Xeon processor).

Where measurement error is zero, the vcv algorithm outperformed the pruning/R algorithm for all but the largest trees (4,096 species). However, the required computational time of the vcv grew very quickly at this point (figure C.1, solid lines). Including “measurement errors” (initial  $V_N > 0$ ), the vcv algorithm requires significant extra time to run. As a result, pruning/R (which was unaffected by the addition of errors), outperformed the vcv beyond around 128 species (figure C.1, dashed lines). The pruning/C algorithm was the fastest in all cases, and was only marginally affected by the addition of measurement errors. For very large trees (4,096) species, the difference in running

times was very large; comparing vcv with pruning/C, there was a 700-fold difference without measurement errors and over 150,000-fold difference including measurement errors.

The required running time in the number of species  $n$ , for the three algorithms at 4,096 species grows as approximately  $O(n^3)$  and  $O(n^{2.3})$  for vcv with and without measurement errors respectively,  $O(n)$  for pruning/R, and  $O(n^{0.9})$  for pruning/C. The sub-linear growth of the pruning/C algorithm is due to the impact of fixed computational costs<sup>9</sup>, and should approach  $O(n)$  for larger trees. These performance timings do not include the time to compute and invert the phylogenetic variance-covariance matrix for the vcv algorithm, which would currently represent a large fraction of the time in carrying out an ML search. For example, on a 4,096 species tree creating a likelihood function (including creating and inverting the phylogenetic covariance matrix, and computing the determinant of the inverse) takes 124 s with the vcv algorithm compared with 0.23 s with the pruning algorithm.

The tree sizes used here are large, and the performance differences on modest sized tree should be fairly small (figure C.1). However, trees larger than the largest size used here have increasingly become available (e.g., Bininda-Emonds *et al.*, 2007; Smith and Donoghue, 2008), and which exceed the ability of the vcv algorithm to run. Extending this pruning algorithm approach to relax the model may also be easier than within the conventional vcv framework.

The code for the timing tests is available on the diversitree github site<sup>10</sup>.

---

<sup>9</sup>Simply checking that the diffusion parameter is non-negative takes 12% of the calculation time for a 256 species tree.

<sup>10</sup><http://github.com/richfitz/diversitree/tree/pub/bm>

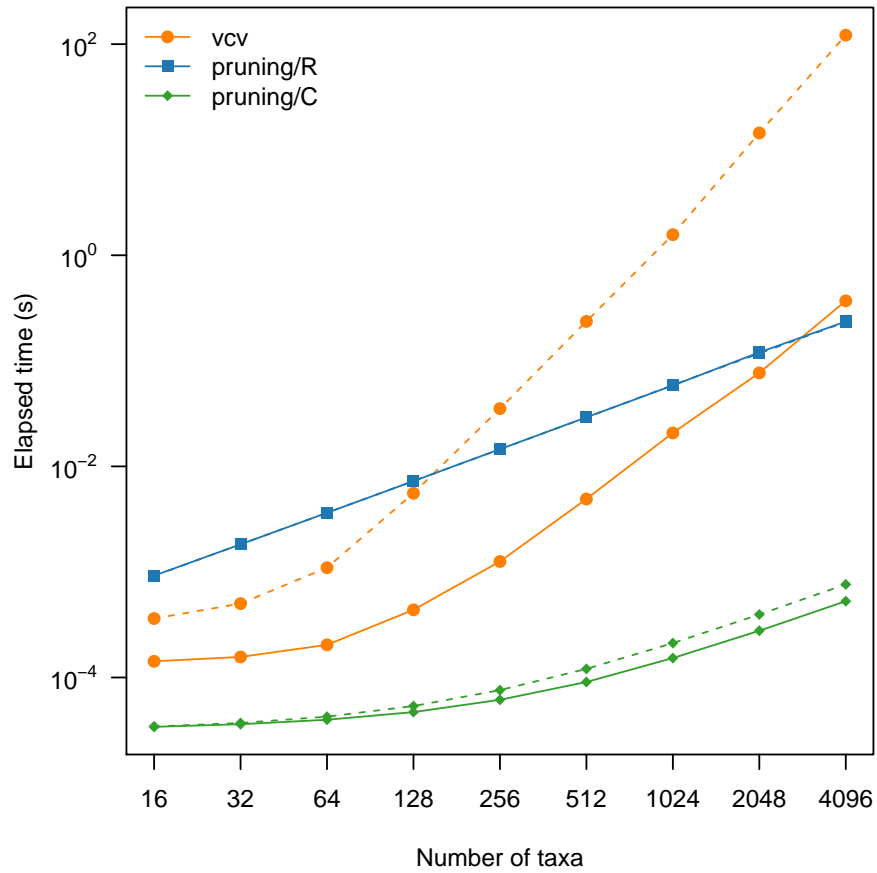


FIGURE C.1: Mean running times for each likelihood function evaluation under the three algorithms. Solid lines assume that species traits are known without error, while dashed lines include “measurement error”. Note the logarithmic scale of both axes.

### C.3 MUSSE & MULTITRAIT DIVERSIFICATION IN PRIMATES

This is a full version of the analysis of social structure and mating system in primates. However, it is not a reference and the reader is directed to the diversitree help pages for further information (in particular the help page `?make.musse.multitrait`).

This section is a “Sweave” document: a mix of R and  $\text{\LaTeX}$  code (see Leisch, 2002). It is possible to recompile the document, which runs the R code and regenerates this output. For instructions on how to do this, please see the instructions at the top of the `primates.Rnw` source file, available on the diversitree github site<sup>11</sup>. To run the examples, or recompile this file, you will need two data files: “`primates-100.nex`” containing the phylogenetic trees and “`primates-social.csv`” containing the trait data. These are also available on the diversitree github site<sup>12</sup>. I assume that these files are in the directory “`data`”.

```
> library(diversitree)
```

First, load the distribution of trees. These were generated by Tyler Kuhn, using the approach in Kuhn *et al.* (2011). Here, we will use just the first tree to demonstrate the methods.

```
> trees <- read.nexus("data/primates-100.nex")
> tree <- trees[[1]]
```

Next, load the species trait data; the first line below creates a data.frame with the species names as row names. The two data columns are “M”, and “S”, which are TRUE when the species is monogamous and social, respectively. Alternatively, these could be integer values with “0” and “1” being equivalent to FALSE and TRUE, respectively (I will refer to states 0 and 1 below).

```
> dat <- read.csv("data/primates-social.csv", row.names=1)
> head(dat)
```

<sup>11</sup><https://github.com/richfitz/diversitree/tree/master/pub/example/>

<sup>12</sup><https://github.com/richfitz/diversitree/tree/master/pub/example/data>, or clone the diversitree github repository and run the analysis from the `diversitree/pub/example` directory.

	M	S
<i>Allenopithecus nigroviridis</i>	NA	FALSE
<i>Allocebus trichotis</i>	TRUE	TRUE
<i>Alouatta belzebul</i>	NA	FALSE
<i>Alouatta caraya</i>	NA	FALSE
<i>Alouatta coibensis</i>	FALSE	FALSE
<i>Alouatta fusca</i>	NA	FALSE

Note that some of the species lack state information. The distribution of traits on the phylogeny is shown in figure C.2.

Start by creating a simple model, in which the speciation and extinction rates do not depend on the character state, and the two traits have forward ( $0 \rightarrow 1$ ) and backward ( $1 \rightarrow 0$ ) transition rates that do not depend on the state of the other trait.

```
> lik.0 <- make.musse.multitrait(tree, dat, depth=0)
```

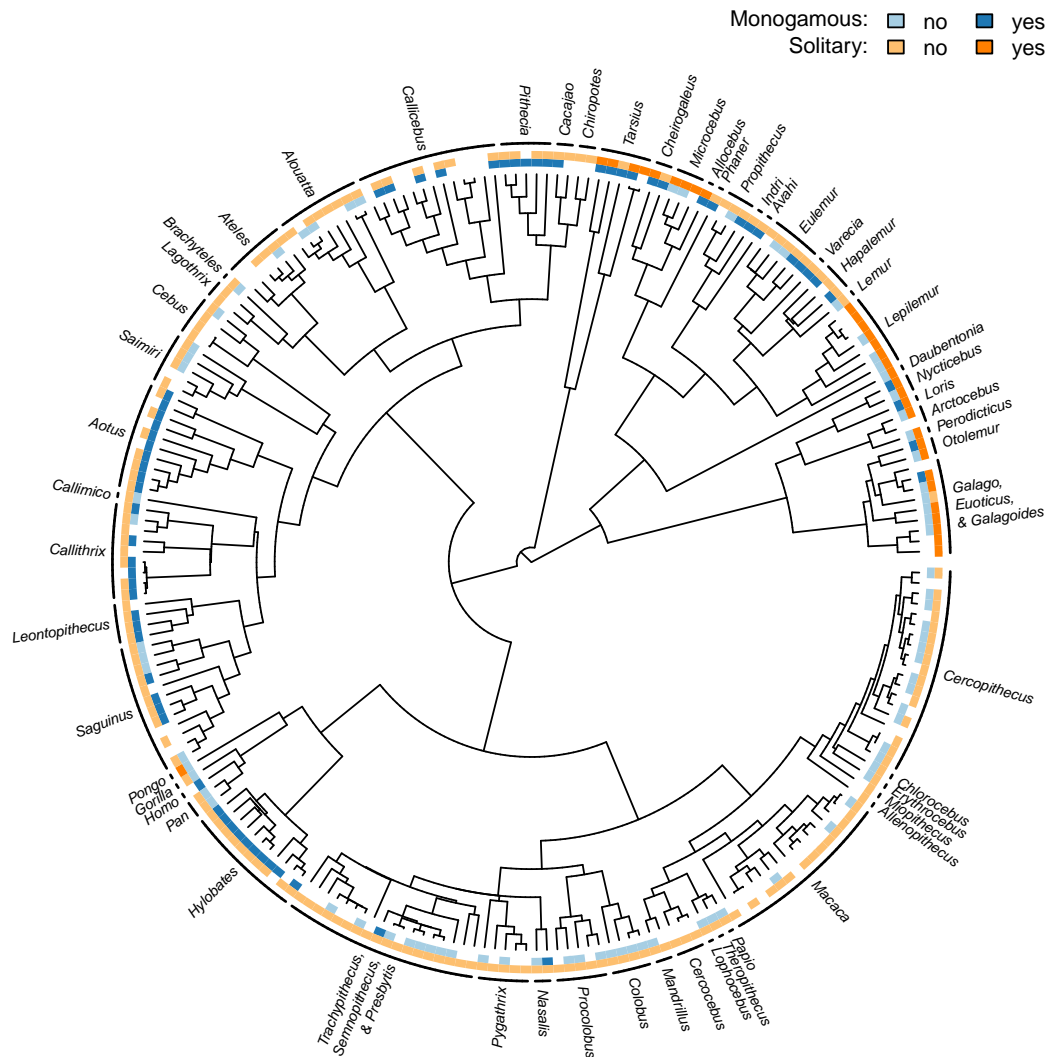
All diversitree likelihood functions take as their first argument a vector of parameters.

To get the vector of names for the parameters, use the `argnames` function:

```
> argnames(lik.0)
[1] "lambda0" "mu0"      "qM01.0"  "qM10.0"  "qS01.0"  "qS10.0"
```

This shows the six parameters: the speciation rate (`lambda0`), extinction rate (`mu0`) and six transition rates (e.g., `qM01.0` is the rate of transition of the breeding system from non-monogamous to monogamous). The “0” after all parameter names indicates that these are intercepts.

To perform a maximum likelihood (ML) analysis, we search for the parameter vector with the highest likelihood. To do this, diversitree uses a numerical optimisation routine (by default, subplex; Rowan, 1990); most algorithms start at some point in parameter space and work “uphill” finding parameters that improve the likelihood until no further improvement is possible. To start this search, we therefore need a starting parameter vector from which the ML point is reachable. It is not possible in general to prove that a point will lead to the ML point, or that the best point found really is the ML point. However, the state-independent speciation and extinction, combined with reasonable



```
> cols <- list(M=c("#a6cee3", "#1f78b4"), S=c("#fdbf6f", "#ff7f00"))
> genus <- sub("_.$", "", tree$tip.label) # extract genus name
> trait.plot(tree, dat, cols, lab=c("Monogamous", "Solitary"),
  str=c("no", "yes"), genus)
```

FIGURE C.2: Primate phylogeny, showing states of the two traits considered here: monogamy (blue, inner circle) and solitariness (orange, outer circle). Species are grouped by genus (or groups of genera in the case of polyphyletic groupings).



guesses for the character state transition rates appears to converge with reasonable success:

```
> p.0 <- c(starting.point.bd(tree), rep(.1, 4))
> names(p.0) <- argnames(lik.0)
```

It is good practice to confirm that this point has finite likelihood:

```
> lik.0(p.0)
[1] -863.5594
```

We can then carry out the ML search, with `find.mle`:

```
> fit.0 <- find.mle(lik.0, p.0)
```

This returns an object of class “`fit.mle`”, which has a `lnLik` element with the log-likelihood at the ML point, and a `par` element with the ML parameter vector. This object is described in more detail in the help page `?find.mle`. As with other model fits, the `coef` function can access parameters:

```
> fit.0$lnLik
[1] -786.3427
> round(coef(fit.0), 4)
lambda0      mu0  qM01.0  qM10.0  qS01.0  qS10.0
0.1912  0.1110  0.0251  0.0259  0.0009  0.0163
```

Now, we expand this model to allow state-dependent diversification. To make a likelihood function that includes “main effects” of the two traits for speciation and extinction, but leaves the character state change parameters the same, we use `depth=c(1, 1, 0)`.

```
> lik.1 <- make.musse.multitrait(tree, dat, depth=c(1, 1, 0))
> argnames(lik.1)
[1] "lambda0" "lambdaM" "lambdaS" "mu0"      "muM"      "muS"
[7] "qM01.0"  "qM10.0"  "qS01.0"  "qS10.0"
```

(Specifying `depth=c(1, 1, 1)`, or equivalently `depth=1`, would also introduce main effects for the transition rates, which would then mean we would have to determine why

`lik.1` might fit better than `lik.0` — the additional degrees of freedom from state dependent diversification or from correlated character evolution.) To find the ML point for this model, we again need a starting parameter vector. The model `lik.0` is a special case of model `lik.1`, so it makes sense to start the ML point found in `fit.0`. To do this, expand the parameter vector of the model in `lik.0` to include main effects, but set them equal to zero:

```
> p.1 <- rep(0, length(argnames(lik.1)))
> names(p.1) <- argnames(lik.1)
> p.1[names(coef(fit.0))] <- coef(fit.0)
> round(p.1, 4)
lambda0 lambdaM lambdaS      mu0      muM      muS qM01.0 qM10.0
0.1912  0.0000  0.0000  0.1110  0.0000  0.0000  0.0251  0.0259
qS01.0 qS10.0
0.0009  0.0163
```

(compare this parameter vector to `coef(fit.0)`, above). The parameter vector `p.1` must have the same likelihood as the previous ML fit:

```
> lik.0(coef(fit.0))
[1] -786.3427
> lik.1(p.1)
[1] -786.3427
```

Find the ML point for the model that includes main effects:

```
> fit.1 <- find.mle(lik.1, p.1)
```

This model fits substantially better than the state-independent model, with a likelihood improvement of 12.4:

```
> fit.1$lnLik - fit.0$lnLik
[1] 12.36941
```

With a difference of 4 parameters, we can compare twice this value to a  $\chi^2_4$  and see that the improvement is statistically significant.

```
> 1 - pchisq(2*(fit.1$lnLik - fit.0$lnLik), 4)
[1] 5.677339e-05
```

The `anova` function does these likelihood ratio tests automatically, also reporting AIC values:

```
> anova(fit.1, noSDD=fit.0)
      Df   lnLik    AIC  ChiSq Pr(>|Chi|)
full   10 -773.97 1568.0
noSDD   6 -786.34 1584.7 24.739  5.677e-05
```

We can expand the model further to include interaction terms; is a combination of mating system and sociality associated with elevated speciation or extinction?

```
> lik.2 <- make.musse.multitrait(tree, dat, depth=c(2, 2, 0))
> argnames(lik.2)
[1] "lambda0"  "lambdaM"  "lambdaS"  "lambdaMS" "mu0"      "muM"
[7] "muS"      "muMS"     "qM01.0"   "qM10.0"   "qS01.0"   "qS10.0"
```

Now fit the model, starting from the ML point found in `fit.1`, expanded to include intercept terms for both speciation and extinction rates:

```
> p.2 <- rep(0, length(argnames(lik.2)))
> names(p.2) <- argnames(lik.2)
> p.2[names(coef(fit.1))] <- coef(fit.1)
> fit.2 <- find.mle(lik.2, p.2)
```

Comparing this against the no-interaction fit, there is no significant improvement in model fit:

```
> anova(fit.2, noInteraction=fit.1)
      Df   lnLik    AIC  ChiSq Pr(>|Chi|)
full      12 -773.73 1571.5
noInteraction 10 -773.97 1568.0 0.49143  0.7821
```

The improvement in fit between `fit.1` and `fit.0` could be due to either of the monogamy or sociality traits (or both). We can determine the contribution of each by constructing models that omit main effects of each trait. The `constrain` function can

be used to simplify models by removing parameters. Parameters can either be set to be equal to a different parameter or to a constant. For example, to remove the main effect of the breeding system trait on speciation from the `lik.1` model, we can enter:

```
> lik.1M <- constrain(lik.1, lambdaM ~ 0)
```

Notice that `lambdaM` is no longer present in the `argnames` of this likelihood function:

```
> argnames(lik.1M)
[1] "lambda0" "lambdaS" "mu0"      "muM"      "muS"      "qM01.0"
[7] "qM10.0" "qS01.0" "qS10.0"
```

Similarly, for sociality:

```
> lik.1S <- constrain(lik.1, lambdaS ~ 0)
```

These models can then be fit as before (adjusting the starting point to account for the reduction in the number of parameters)

```
> fit.1M <- find.mle(lik.1M, p.1[argnames(lik.1M)])
> fit.1S <- find.mle(lik.1S, p.1[argnames(lik.1S)])
```

There is a significant drop in likelihood when the breeding system (monogamy/non-monogamy) speciation main effect is dropped from the model, but the social/nonsocial trait association is marginally non-significant (both comparisons here are made against `fit.1`):

```
> anova(fit.1, noM=fit.1M, noS=fit.1S)
      Df  lnLik   AIC  ChiSq Pr(>|Chi|)
full 10 -773.97 1568.0
noM   9 -778.24 1574.5 8.5236   0.003506
noS   9 -775.52 1569.0 3.1031   0.078145
```

Alternatively, we might run use Markov chain Monte Carlo (MCMC) to sample from the posterior distribution of the `lambdaM` and `lambdaS` values. This procedure requires a prior probability distribution for the parameters. Here, I use a prior on the actual rates in the model, not the multitrait parametrisation (partly because it is easier to constrain

the former to being positive; negative speciation and extinction rates are not allowed, and partly because I do not have a strong prior belief about the form of the main effect parameters). Given a multi-trait parametrisation, the underlying rate parameters can be found by passing in the argument `pars.only=TRUE` to the likelihood function, which returns the underlying parameters without computing the likelihood:

```
> round(coef(fit.1), 3)
lambda0 lambdaM lambdaS      mu0      muM      muS qM01.0 qM10.0
  0.224  -0.095  -0.076   0.061  -0.061   0.000   0.020   0.030
qS01.0 qS10.0
  0.000   0.017

> round(lik.1(coef(fit.1), pars.only=TRUE), 3)
lambda00 lambda10 lambda01 lambda11      mu00      mu10      mu01
  0.224   0.128   0.148   0.052   0.061   0.000   0.061
  mu11   q00.10   q00.01   q00.11   q10.00   q10.01   q10.11
  0.000   0.020   0.000   0.000   0.030   0.000   0.000
q01.00   q01.10   q01.11   q11.00   q11.10   q11.01
  0.017   0.000   0.020   0.000   0.017   0.030
```

To specify an exponential prior with a mean set to twice the state-independent diversification rate:

```
> r <- p.0[[1]] - p.0[[2]]
> prior1 <- make.prior.exponential(1/(2*r))
```

Using the translation above, we can make a function that takes as arguments the multi-trait parametrisation and returns the prior probability density:

```
> prior <- function(pars)
  prior1(lik.1(pars, pars.only=TRUE))
```

Running the MCMC for 10,000 steps (this takes several hours, and by default will print the parameters visited on each step and their posterior probability).

```
> samples <- mcmc(lik.1, p.1, nsteps=10000, w=0.5, prior=prior)
```

The posterior distributions for the main effects of the speciation rates ( $\lambda_M$  and  $\lambda_S$ ) are concentrated below zero, with the 95% credibility interval below zero (figure 4.3). Therefore, we can conclude both traits have a negative effect on speciation.

For completeness, I will show how a similar analysis would proceed if we treat each trait separately with BiSSE. First, we will need two named state vectors — one for each trait:

```
> st.m <- dat$M
> names(st.m) <- rownames(dat)
> st.s <- dat$S
> names(st.s) <- rownames(dat)
```

Then we build likelihood functions (using `make.bisse`).

```
> lik.m <- make.bisse(tree, st.m)
> lik.s <- make.bisse(tree, st.s)
```

We can then run MCMC chains from a “sensible” starting point (see the help page `?starting.point.bisse`):

```
> p <- starting.point.bisse(tree)
```

Running the chains:

```
> samples.m <- mcmc(lik.m, p, nsteps=10000, w=.5, prior=prior1)
> samples.s <- mcmc(lik.s, p, nsteps=10000, w=.5, prior=prior1)
```

For a single trait, the difference in the speciation rates (i.e.,  $\lambda_1 - \lambda_0$ ) is mathematically equivalent to the main effect of that trait. Monogamy is significantly associated with decreased speciation rates (the 95% credibility interval of  $\lambda_M$  is below zero; figure 4.3). However, the effect of solitariness is no longer significant.

Similarly, we can run an ML analysis. First, fit the full six parameter model for each trait:

```
> fit.m <- find.mle(lik.m, p)
> fit.s <- find.mle(lik.s, p)
```

Then fit constrained models where  $\lambda_1$  is set equal to  $\lambda_0$  for both traits:

```
> lik.m.eqL <- constrain(lik.m, lambda1 ~ lambda0)
> fit.m.eqL <- find.mle(lik.m.eqL, p[argnames(lik.m.eqL)])
> lik.s.eqL <- constrain(lik.s, lambda1 ~ lambda0)
> fit.s.eqL <- find.mle(lik.s.eqL, p[argnames(lik.s.eqL)])
```

Comparing the “with state-dependent speciation” model against the simpler model that lacks state-dependent speciation with a likelihood ratio test for the monogamy trait:

```
> anova(fit.m, equal.lambda=fit.m.eqL)

          Df   lnLik    AIC  ChiSq Pr(>|Chi|)
full          6 -755.62 1523.2
equal.lambda  5 -758.74 1527.5 6.2242      0.0126
```

and the solitariness trait

```
> anova(fit.s, equal.lambda=fit.s.eqL)

          Df   lnLik    AIC  ChiSq Pr(>|Chi|)
full          6 -710.90 1433.8
equal.lambda  5 -711.59 1433.2 1.3821      0.2397
```

confirms the previous BiSSE result: there is evidence of state-dependent speciation for monogamy, but not for solitariness.

Note that it is not possible to directly compare the multitrait MuSSE model with the BiSSE model, because these models explain different data; MuSSE accounts for the observed distribution of multiple traits, BiSSE accounts only for one trait. Likelihood ratio tests are only valid where models are nested and where the data are identical between both models. Attempting a non-nested comparison will generate an error. Instead, to perform such comparisons, the user must generate constrained models using MuSSE, which accounts for all of the trait data but allows only one of the traits to affect diversification (as performed using `lik.1M` and `lik.1S` above).

## APPENDIX D

### SUPPLEMENTARY INFORMATION TO CHAPTER 5

---

#### D.1 PARTITION COMPOSITIONS

Species composition of different tree partitions identified with MEDUSA. Full species lists of each partition for each tree will be made available on DRYAD.

##### *Afrotheria*

Afrotheria was only partitioned three times, into the same two groups:

BASAL AFROTHERIA — containing most of the super order

MICROGALE — a genus of tenrecs. Only 9 or 10 of the species in this genus were included in the group.

##### *Carnivora*

Carnivora was always split into two groups

BASAL CARNIVORA — most species in the order.

SOME CANIDAE — a group of 4–8 recently diverged genera within Canidae, always including *Canis* and *Lycalopex*. Other genera sometimes included are *Atelocynus*, *Cerdocyon*, *Chrysocyon*, *Cuon*, *Lycaon*, and *Speothos*.

##### *Cetacea*

Cetacea was partitioned into two groups in 36 trees:



BASAL CETACEA — a group containing the baleen whales (suborder Mysticeti, with families Balaenidae, Neobalaenidae, and Balaenopteridae), and most families of toothed whales.

DELPHINIDAE — most of the Delphinidae (dolphins). This group never included *Orcinus orca* and in three cases also did not include *Feresa*, *Globicephala*, *Grampus*, *Orcaella*, *Peponocephala*, and *Pseudorca* (trees 61, 81, 86).

### *Chiroptera*

Partitioned in all trees into a variable number of groups:

BASAL CHIROPTERA (100 trees) — basal group that includes most species.

RHINOLOPHUS (100 trees) — horseshoe bats. This group includes only some species in this genus.

SOME CHIROPTERA (99 trees) — this is a complicated group of fairly small families. Always includes the families Craseonycteridae, Emballonuridae, Hipposideridae, Megadermatidae, Nycteridae, Rhinolophidae, and Rhinopomatidae.

MYOTIS (83 trees) — mouse-eared bat and relatives. Up to 36 genera may be included in this group, but usually only two (*Cistugo* is always present in addition to *Myotis*).

NATALOIDEA (41 trees) — funnel-eared bats and relatives. This superfamily includes the families Furipteridae, Myzopodidae, Natalidae, and Thyropteridae.

PTEROPODINAE (13 trees) — this clade approximately corresponds to Pteropodinae, Rousettinae (missing *Eidolon*), and Epomophorinae.

VAMPYRESSA (10 trees) — in one tree this is a group of 18 genera, corresponding approximately to Stenodermatinae. For the other trees, this group is extended out to include most genera in Phyllostomidae.

### *Eulipotyphla*

Always splits into two or three groups

**BASAL EULIPOTYPHLA** — A basal group that always includes the families Erinaceidae (Erinaceomorpha) and Solenodontidae and Talpidae (Soricomorpha).

**CROCIDURA** — white-toothed shrews. A group within the Soricidae that always includes *Crocidura*, but up to 9 others mostly in the Crocidurinae.

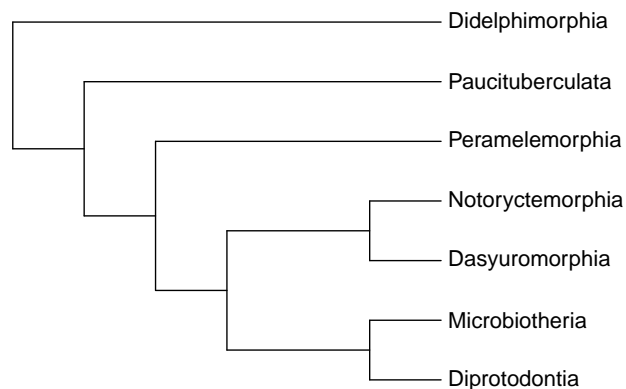
**SORICIDAE** (21 trees) — shrews. Where present, this group includes all Soricidae not included in CROCIRUDA.

### *Lagomorpha*

Never partitioned by MEDUSA.

### *Marsupials*

The Marsupials include 7 orders:



This always splits into three groups

**BASAL MARSUPIALS** — most of the Marsupials: everything except for the two families below.

**DASYURIDAE** — “marsupial mice”, within Dasyuromorphia. The entire family except for *Lagostrophus* (Banded Hare-wallaby) is always included.

**MACROPODIDAE** — kangaroos, wallabies, etc, within Diprotodontia. The entire family is always included.

## *Primates*

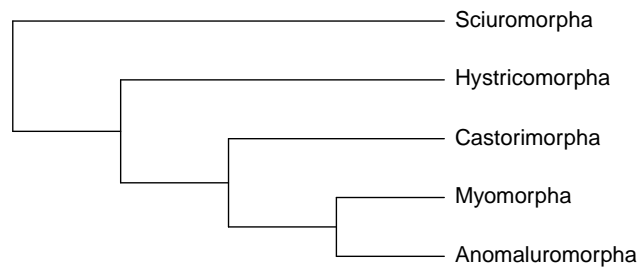
Always splits into two groups

BASAL PRIMATES — most of the primates: everything except the family below.

CERCOPITHECIDAE — “Old World monkeys”. The entire family is always included.

## *Rodentia*

Rodentia can be divided into several suborders:



Most of the diversity is in Myomorpha (which contains Muridae and Cricetidae), so we treat this separately from the other suborders for simplicity. The only group that spans both the background group of suborders (Anomaluromorpha, Castorimorpha, Hystricomorpha and Sciuromorpha) and Myomorpha is

BASAL RODENTIA (100 trees) — This collects everything that is not in one of the groups below and spans the root of the tree. It primarily contains non-Myomorpha species, with suborders Anomaluromorpha (Anomaluridae and Pedetidae) and Castorimorpha (Castoridae, Geomyidae, and Heteromyidae) always entirely included. Other families are always in this group, such as Dipodidae (Myomorpha), most families in Hystricomorpha, and Aplodontiidae (Sciuromorpha).

Non-Myomorpha groups:

MARMOTINI (100 trees) — most ground squirrels, within Sciuridae. This is a group of three genera of ground squirrels *Cynomys*, *Marmota*, *Spermophilus*. This

broadly matches Herron *et al.* (2004)'s, with the exclusion of *Ammospermophilus*, which is the sister to the three species here.

SCIURINAE (99 trees) — flying squirrels and tree squirrels, within Sciuridae. With the exception of one tree, the same set of 16 genera are included: (*Aeromys*, *Eoglaucmys*, *Eupetaurus*, *Glaucmys*, *Hylopetes*, *Iomys*, *Microsciurus*, *Petaurillus*, *Petaurista*, *Petinomys*, *Pteromys*, *Pteromyscus*, *Rheithrosciurus*, *Sciurus*, *Syntheosciurus*, *Tamiasciurus*). Tree 72 also includes *Ratufa* and *Sundasciurus*, which are the sister groups to this subfamily.

CTENOMYIDAE (26 trees) — tucos, within Ctenomyidae (Hystriocomorpha). This is the only genus in the family, and only part of the genus is included in this group.

#### Myomorpha groups

SIGMODONTINAE (100 trees) — New World rats and mice, within Cricetidae. This either includes the entire subfamily (64 genera: 15 trees) or most of the family (as few as 42 genera, only 11 trees have fewer than 57 genera).

MURINAE (85 trees) — this is a large group that varies in composition across different trees. It varies from approximately the Murinae (within family Muridae, only 5 trees have this small a span), to the entire Myomorpha group except for Dipodidae.

ARVICOLINAE (28 trees) — voles, lemmings, and muskrats, within Cricetidae. The entire subfamily is not always present, with the group ranging from 10–21 species, but usually 21 (see tree 35 for the smallest collection of genera). Most commonly (10 trees) the tribe Lemmini (lemmings: *Synaptomys*, *Lemmus*, *Myopus*) is missing.

PEROMYSCUS (17 trees) — deer mouse, within Cricetidae/Sigmodontinae. This sometimes did not cover the full genus, but always get most of it.

RATTUS (11 trees) — rats and related genera, within Muridae/Murinae. This group varies from 2–14 genera in a fairly continuous manner.

*Ruminantia*

Ruminantia was partitioned in all but four trees. When partitioned, this was always into two groups:

BASAL RUMINANTIA — Antilocapridae, Giraffidae, and Tragulidae. These three small families span the root of the tree.

MOST RUMINANTIA — Moschidae, Cervidae, and Bovidae. These families make up the vast majority of the diversity in Ruminantia.