Developments for the Next Generation of Evolutionary Paleobiology

by

Caroline Parins-Fukuchi

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Ecology and Evolutionary Biology)
in the University of Michigan
2019

Doctoral Committee:

Professor Christopher W. Dick, Co-Chair
Professor Daniel C. Fisher, Co-Chair
Associate Professor Matt Friedman
Professor Edward Ionides
Associate Professor Daniel L. Rabosky

Caroline Parins-Fukuchi

cfukuchi@umich.edu

ORCID iD: 0000-0003-0084-2323

For E, T, and S

Acknowledgements

I owe many people gratitude. First, I would like to thank my advisors, **Dr. Christopher W. Dick** and **Dr. Daniel C. Fisher**.

Chris is, in many ways, an ideal mentor for my idiosyncratic and headstrong style. Although my research interests quickly drifted past his immediate area of specialization, Chris showed remarkable intellectual flexibility along with both enthusiasm and astute criticism with each new direction upon which I embarked. While I was immersed in the depths of statistical methods development, Chris consistently provided very healthy nudges to continue watering my roots as a biologist. I am indebted to Chris for his guidance in shaping my approach as a scholar.

It is difficult for me to overstate the centrality of Dan's influence in my growth as a paleontologist and evolutionary biologist. My first interaction with Dan was as an undergraduate taking his paleontology seminar course. Entering as a budding paleoanthropologist and computational biologist, Dan's practice of digging deeply into the foundations of paleontological inference was a defining influence as I shaped the approach that lead to the content of my doctoral work. The directions that I forged while creating this dissertation were shaped in large part by my inheritance of Dan's predilection toward getting his hands dirty to rebuild ideas from the ground up. My work has grown quite distinct from Dan's, but I feel it important to recognize that my position as a descendant within the same lineage has imbued in my approach many deep homologies inherited from him. I am humbled to contribute as student to the legacy of such a thought-provoking and uncompromising scientist.

I am also indebted to my committee members, **Dr. Matt Friedman**, **Dr. Edward Ionides**, and **Dr. Daniel Rabosky**. My interactions with each of these phenomenal scientists have strengthened my work considerably.

There are many additional colleagues and collaborators who have been instrumental as I shaped my research program. Although I didn't get to spend as much time on hominoid functional morphology during my dissertation as I'd hoped, my collaborations and guidance from **Dr. Laura MacLatchy** have been critical in helping me keep some grounding in the paleoanthropological world. I hope that we will have many more interactions as I rebuild my connection to the hominoids over the next few years. **Dr. Joseph Brown** was an excellent collaborator and early mentor. His phylogenetics nerdiness and enthusiasm for Electric Wizard were always charming and made for highly entertaining and worthwhile collegial interactions. **Dr. Greg Stull** made an honest effort to make a paleobotanist out of me. It didn't quite take, but our many (continuing) interactions have been very enjoyable and fruitful (sickening pun intended). **James Saulsbury** is a newer colleague. I have benefited greatly from encountering another continuous-trait-enthusiastic paleobiologist. **Nat Walker-Hale's** ideas were frequent highlights in our 'PRUSSIA' group meetings. His intellect and enthusiasm will serve him well as he grows further into a phenomenal dialectical biologist. **Dr. Joseph Walker** has been like a brother since our shared infancy in the EEB PhD program. It was an honor to share our bantering, bickering, and mutual intellectual growth. I will always view my trajectory as tied in some way to his.

Finally, I express my deep gratitude to **Dr. Diana Gannett**. Although I left my potential as a bassist unfulfilled, Diana's willingness to fight to accept into her studio a scruffy teenager

was a life-shaping and necessary condition for the creation of this dissertation. As an avid student of historical contingency, I recognize deeply the importance of this event in allowing my trajectory as a scientist. Her many nudges to ‘think ahead’ have tamed my speaking in many class lectures and academic talks. I will hold this, along with many more of Diana's teachings, throughout my life.

TABLE OF CONTENTS

DEDICATION . . . . . **ii**

ACKNOWLEDGEMENTS . . . . . **iii**

LIST OF FIGURES . . . . . **ix**

ABSTRACT . . . . . **xii**

CHAPTER

I. Introduction 1

1.1 Paleontology as evolutionary biology 1
1.2 The emergence of paleontology as a biological science 2
1.3 The molecular revolution in evolutionary biology 5
1.4 A burgeoning revolution in evolutionary paleobiology? 7
1.5 Chapter summaries 10

II. Continuous Trait Phylogenetics 14

2.1 Abstract 14
2.2 Introduction 15
2.3 Methods 20
2.3.1 Simulations 20
2.3.2 Estimation of phylogenies and reconstruction accuracy 23
2.4 Results 24
2.4.1 Unconstrained and independently evolving continuous traits 24
2.4.2 Continuous traits evolving under selective constraint 26
2.4.3 Covarying continuous characters 26
2.5 Discussion 27
2.5.1 Can using continuous characters benefit morphological phylogenetics? . 30
2.5.2 Is Mk a reasonable model for discrete character evolution? 34

III. Bayesian Placement of Fossils on Phylogenies Using Quantitative Morphometric Data . 40

3.1 Abstract 40
3.2 Introduction 42
3.3 Methods and Materials 47
3.3.1 Software 47
3.3.2 Brownian motion model 48
3.3.3 Computation of the likelihood 49

3.3.4 Priors . . . 50
3.3.5 Markov-chain Monte Carlo . . . 51
3.3.6 Filtering for concordant sites . . . 53
3.3.7 Simulations . . . 55
3.3.8 Empirical analyses . . . 58
3.4 Results and Discussion . . . 60
3.4.1 Simulations . . . 60
3.4.2 Vitaceae . . . 63
3.4.3 Carnivorans . . . 64
3.4.4 Continuous vs discrete morphological characters . . . 66
3.4.5 Comparison to other approaches . . . 68
3.4.6 Conclusions . . . 69

IV. Phylogeny, Ancestors, and Anagenesis in the Hominin Fossil Record . . . 74

4.1 Abstract . . . 74
4.2 Introduction . . . 75
4.3 Methods and materials . . . 80
4.3.1 Use of terms . . . 80
4.3.2 Inference of ML topology . . . 81
4.3.3 Identification of a fully-bifurcating ML tree . . . 81
4.3.4 Stratigraphic and morphologic datasets . . . 86
4.4 Results and Discussion . . . 87
4.4.1 Anagenesis in the hominin fossil record . . . 87
4.4.2 Ancestors, anagenesis, and evolutionary processes . . . 92
4.4.3 Some practical methodological considerations . . . 95
4.4.4 How can we proceed? . . . 97

V. Detecting Mosaic Evolutionary Patterns in Phenotypic Disparity . . . 101

5.1 Abstract . . . 101
5.2 Introduction . . . 102
5.3 Methods and Materials . . . 105
5.3.1 Code and data availability . . . 105
5.3.2 Partitioning traits into mosaic suites . . . 105
5.3.3 Tree model . . . 106
5.3.4 Search procedure . . . 110
5.3.5 Stochastically varying mosaic patterns . . . 112
5.3.6 Power analysis and detection of relative rate shifts . . . 112
5.3.7 Distinguishing rate shifts among covarying traits . . . 113
5.3.8 Evaluation of reconstruction accuracy . . . 114
5.3.9 Empirical analysis . . . 115
5.4 Results and Discussion . . . 115
5.4.1 Simulated stochastically varying rates . . . 115
5.4.2 Relative rate shifts in independently evolving traits . . . 116
5.4.3 Relative rate shifts in covarying traits . . . 117
5.4.4 Type 1 error . . . 117
5.4.5 Empirical analysis . . . 118
5.4.6 Utility in transcriptomic studies . . . 120
5.4.7 Utility to phylogenetic comparative methods . . . 120
5.4.8 Utility for phylogenetic inference and divergence time estimation . . . 122
5.4.9 Overfitting, covariance, and the true number of clusters . . . 123
5.4.10 Morphological integration, mosaic evolution, and geometric morphometrics . . . 124

5.4.11 Scale and rate . . . 125
5.4.12 Choice of information criteria . . . 126
5.5 Conclusions . . . 127

VI. Conclusion . . . 133

6.1 Paths forward . . . 133
6.2 Concluding remarks . . . 134

BIBLIOGRAPHY . . . **136**

LIST OF FIGURES

Figure

2.1 A) Exemplar true simulated tree. B) Tree inferred from 20 discrete characters simulated under Mk from true tree. C) Tree inferred from 20 continuous characters simulated under Brownian motion. Blue dots denote incorrect bipartitions. 21

2.2 Topological error calculated as the proportion of maximum symmetric distance across trees estimated from independently evolving continuous characters. A) Error averaged across all rates except for the highest rate category, which resulted in the highest error when inferring under Mk. B) Error across all matrix sizes and rates. 25

2.3 Topological error achieved after reconstructing trees from discrete traits simulated under Mk at rate 0.5, and single rate Ornstein Uhlenbeck at rate 0.5 with no directional drift and constraint set equal to the discrete characters. 36

2.4 A) Topological error, calculated as proportion of maximum symmetric distance across trees estimated from covarying continuous characters. B) Branch length distance (BLD) across trees estimated from covarying continuous characters. Dimensions refers to the number of traits within covarying blocks. Covariance strength refers to the strength of the correlation between covarying characters, with a value of 0 describing to complete independence and 1 describing perfect correlation. 37

2.5 Discrete and continuous characters simulated A) at slow evolutionary rate and B) fast evolutionary rate. Y axis represents continuous phenotype. Changes in colour represent changes in discrete character state. Note how continuous characters retain phylogenetic signal at fast rates, while discrete characters saturate. 37

2.6 A) True tree. B) Tree estimated from 50 discrete and 50 continuous characters C) Tree estimated from 100 continuous characters simulated at rate 1.0 D) Tree estimated from 100 discrete characters simulated at rate 1.0. Blue dots signify incorrectly estimated bipartitions. The tree in panel b. was generated by randomly subsampling the matrices used to generate trees C and D, and combining into a single matrix. This matrix was analysed in RevBayes. An example script is provided in the supplement. 38

2.7 Branch length distance (BLD) across trees estimated from independently evolving continuous characters. A) BLD averaged across all rates except for the highest rate category, which resulted in the highest error when inferring under Mk. B) BLD across all matrix sizes and rates. 39

3.1 Random fossil prune and regraft procedure. 71

3.2 Vitaceae fossil placements inferred under the compound Dirichlet branch length prior. Fossil taxa and branches are highlighted in red. Values following fossil tip labels indicate posterior support for placement. Topology is summarized from the posterior using the set of maximally credible clades (MCC). Figure displays only the clade containing all 6 fossils. . 72

3.3 Fossil placements inferred from the carnivoran dataset using the compound Dirichlet prior. Placements are displayed as the maximum clade credibility summary of the posterior distribution of trees. Branch lengths represent morphological disparity. Values trailing fossil tip names display posterior support. 73

4.1 Comparison between anagenetic and cladogenetic trees. A) Likelihoods are calculated on bifurcating arrangements using the standard 'pruning algorithm'. B) Direct ancestor likelihood calculation. 85

4.2 Phylogenetic relationships between hominin species. A) Reconstruction with the best AIC score when temporal data are considered alongside morphological data and anagenesis is accommodated. B, C) Areas of the tree inferred from morphology alone that differ from panel A. Both regions display misleading results when anagenesis is not considered. This is reflected in the improvement in AIC observed in the preferred anagenetic topology. Silhouettes obtained from phylopic.org. 88

4.3 AIC scores calculated for each possible arrangement between *H. sapiens*, *H. neanderthalensis*, and *H. heidelbergensis*. 89

4.4 A) Speciation mode interpreted as budding when sampling is complete. B) Incomplete stratigraphic sampling may create an inability to distinguish between anagenesis and budding when sampling is sparse. 100

5.1 Search procedure to identify mosaic suites. 106

5.2 Fitting increased disparity observed in taxa T2 and T3. A) Single rate BM fit to a chronogram assumes clock-like evolution, ignoring both positive jumps in disparity and stasis. B) Expressing branch lengths in units of disparity under the phylogenetic Brownian parameterization used here captures both increased disparity and stasis. C) Fitting multiple Brownian rates or shifts in adaptive optima or D) assuming the presence of a single jump in optimum trait value may explain the same pattern with more model parameters. Note that panels B, C, and D all display similar information, while panel A assumes clock-like evolution. 109

5.3 A) Adjusted Rand indices across reconstructions of simulated datasets. B) Number of clusters resulting from analyses of simulated data. Barplots are stacked to represent the proportion of replicates that resulted in k character suites. All violin and barplots are separated by true number of simulated character suites. 116

5.4 A) Adjusted Rand indices from reconstruction of datasets comprised of mosaic suites with one simulated under equal rates across branches, and the other displaying a randomly placed rate shift at magnitudes of 8, 16, 32, and 64. B) Stacked barplots to show the frequency that each of k clusters are inferred from the simulated datasets. C) and D) show the same metrics as above, summarized from datasets containing simulated covarying modules of characters. 129

5.5 Type 1 error. Reconstructions show no type 1 error when traits are evolved independently, or display weak covariance. Type 1 error increases at moderate and high levels of covariance, as the method overfits biased patterns to place covarying modules in superfluous mosaic suites. 130

5.6 A-D) Four best configurations with AIC scores. E) Weighted graph calculated by summing the AIC weights associated with the each model to form edges and edge weights. All graphs were drawn using the "lgl" format implemented in igraph. 131

5.7 Branch lengths reconstructed from traits contained within: A) suite 0 and B) suite 1 display distinct patterns in disparity. 132

5.8 Expected patterns between traits with and without integration. A) Two possible patterns in disparity represented by differences in disparity-scaled branch lengths. B) Values of two traits generated according to the pattern implied by tree 1 in panel A, but evolved independently. C) Two traits generated according to tree 1, but with strongly covarying evolutionary paths. D) Two traits generated independently and distinctly according to tree 1 and 2. Note that the covariance in panel C implies a shared pattern in disparity, while the examples in panels B and D do not. 132

ABSTRACT

During the modern synthesis, researchers merged insights from natural history, evolutionary genetics, and paleontology to develop a cohesive theoretical foundation for evolutionary theory. Since then, the rapid emergence of genomic resources has revolutionized our understanding of evolutionary processes. Despite neontological successes, paleobiology has lagged behind, due in part to perceived challenges in collecting and analyzing morphological data. As a result, the earlier synthetic evolutionary view developed between neo- and paleontology has not kept pace with the current data-centric landscape. To address these issues, I aim to integrate morphological data representing fossil and living taxa into the modern evolutionary framework through the development of novel statistical approaches that leverage sources of data previously thought to be unconventional. These developments follow two main threads: 1) development of a statistical framework through which to infer phylogeny among fossil taxa by merging increasingly large and high-throughput quantitative morphological datasets with stratigraphic information, and 2) developing empirical applications of new approaches to comprehensively examine long-hypothesized but under-studied patterns in evolutionary rate throughout time, and mosaic change by integrating morphological, stratigraphic, and developmental data.

CHAPTER I

Introduction

1.1 Paleontology as evolutionary biology

Paleontology has long had a challenging relationship with evolutionary biology. In *On the Origin of Species*, Darwin emphasized both the importance and unreliability of the fossil record in providing a basis of empirical evidence for evolutionary theory. Darwin argued that, while records of extinction and 'transitional' taxa can only be obtained from fossil evidence, the fragmentary preservation of biological material often renders the rock record inadequate in documenting detailed evolutionary change (Darwin 1859, Chapter 9). The tension between evolutionary theory and paleontology reflected in Darwin's early pessimism regarding the usefulness of the fossil record in understanding evolutionary patterns and processes was perhaps indicative of a broader tendency among prominent 19th century paleontologists, including Georges Cuvier and Richard Owen, to reject both pre- and post-Darwinian evolutionary explanations. While the fields of evolutionary biology and paleontology continued their growth after the publication of *Origin*, they followed fairly isolated tracks. The Darwinian heritage in the growing field of evolutionary biology was reflected by skepticism over the ability of fossils to address many evolutionary questions. At the same time, paleontology grew into a geologic discipline that was primarily developed as a tool to define, delimit, and associate major epochs within stratigraphic

sequences.

By the 1920's, evolutionary theory entered a phase of substantial maturation referred to as the 'modern synthesis'. This period was defined by the development of a mechanistic basis for evolution that explained and expanded Darwinian concepts by merging them with an understanding of Mendelian genetics at the population level. Researchers such as Thomas Hunt Morgan, Sewall Wright, J.B.S Haldane, and R.A. Fisher incorporated mathematical models and experimental approaches to explain the emergence of evolutionary patterns from fluctuations in allele frequencies within and between populations of organisms. The modern synthesis movement culminated in the 1940's, diffusing into more historical and organismally-focused subdisciplines of biology, such as zoology and botany, through the work of the researchers Ernst Mayr and G. L. Stebbins. These researchers were more similar to naturalists such as Darwin, and sought to apply the concepts and models from the population geneticists to understand broad patterns in the evolution of taxa outside of a strictly theoretical or experimental context.

1.2 The emergence of paleontology as a biological science

It was during the modern synthesis that paleontology first [1] featured prominently as a full participant in the formulation of evolutionary theory with the publication by George Gaylord Simpson of his book *Tempo and Mode in Evolution* (Simpson 1944). Simpson was a vertebrate paleontologist who closely followed the development of the population genetic work of the modern synthesis throughout the 1920's and 1930's and strongly advocated for the formalization of a quantitative basis for paleontology. In *Tempo and Mode*, Simpson analyzed fossil datasets to explore the ways in which evolutionary patterns take shape over long timescales, demonstrating that rates of morphological evolution vary

[1] Although, for completeness, I should add that there was a movement toward an evolutionary paleobiology on the European continent, shaped most notably by Louis Dollo and Othenio Abel.

throughout evolutionary time from near-stasis to rapid 'quantum' bursts. Simpson proposed explanations for these fluctuations using population processes, developing a model that explained the emergence of novel morphologies over paleontologic timescales as alternating episodes of rapid selection and drift along an 'adaptive landscape', a metaphor describing the fitness attributes of combinations of phenotypes borrowed from the population geneticist Sewall Wright. By convincingly applying this model to the evolution of fossil horses, Simpson provided a long-needed demonstration of both the ability for paleontological data to bear on major evolutionary questions and the value of evolutionary models in understanding patterns displayed by the fossil record.

Simpson's explorations provided an important initial step toward demonstrating the important place of paleontology in modern evolutionary theory. However, his seminal work participated alongside that of several contemporaries who were increasingly interested in analyzing the fossil record in an evolutionary context. Although paleontology spent most of its history before the 1920's as a discipline distinct from evolutionary biology, the growing consensus toward accepting Darwinian evolution and strong theoretical underpinnings provided by the early contributors to the modern synthesis marked an increased interest among paleontologists in harnessing the fossil record to evaluate and propose new evolutionary hypotheses. Norman Newell explained large scale patterns observed in the invertebrate fossil record, such as directional trends in body size evolution (Newell 1949), and periodic bursts in the rate of evolution (Newell 1952) using the evolutionary concepts developed during the modern synthesis.

The work of evolutionary paleontologists such as Simpson and Newell from the 1930's through the 1950's provided an initial injection of hope by giving early demonstrations of the viability of paleontology as a subdiscipline of evolutionary biology. However, their goals came to fruition more fully in the 1960's and 1970's with the maturation of their

students and others following in their work. This generation of paleobiologists fully realized the potential of the fossil record as a laboratory through which to both apply modern evolutionary concepts derived through population genetics and develop new theoretical contributions to evolutionary theory that were uniquely informed by the vast timescales and empirical context available to paleontology. Many of the most influential and enlightening concepts in post-synthesis evolutionary biology were developed by paleobiologists during this time period, such as Eldredge and Gould's 'punctuated equilibria' and Van Valen's 'Red Queen' hypothesis (Van Valen 1973; Gould and Eldredge 1977).

The ideas developed by researchers during this period simultaneously validated, extended, and challenged the views of evolutionary theory developed during the modern synthesis. Encompassing and extending beyond Simpson and Newell's application of synthesis-era concepts to understanding the fossil record, new ideas emerging from paleontologists suggested a much more complex picture of evolutionary theory (Gould 1980b,a). These contributions collectively provided a full realization of the ability for paleontology, now referred to as 'evolutionary paleobiology' to reflect the renewed evolutionary emphasis, to yield unique contributions to evolutionary theory. The transformed scope of evolutionary biology through the renewed synthesis with the fossil record was validated in 1984 with the prominent population geneticist and game theoretician John Maynard Smith's article in *Nature*, titled 'Palaeontology at the High Table' (Maynard Smith 1984). In the article, Maynard Smith chastised previous generations of paleontologists, while celebrating the contributions of Gould and his contemporaries as heralding the emergence of a new evolutionary synthesis wherein paleontology would be fully integrated into the fold of modern evolutionary theory. Although the article may have given a slightly unfair treatment of past work, it marked a significant shift in the growth and acceptance of paleobiology as an evolutionary discipline.

1.3 The molecular revolution in evolutionary biology

Beginning the 1960's, the field of evolutionary biology was reshaped by the emergence and proliferation of molecular data that enabled the direct study of the evolution of biological molecules. Early studies, first using protein data and later incorporating DNA, began to glean the complexity of evolutionary processes at the genomic level. While many synthesis-era geneticists envisioned a fairly simple, atomized view of evolution where single alleles become fixed as a consequence of natural selection or genetic drift, the emergence of molecular data revealed more complicated patterns. For instance, molecular data suggested many evolutionary changes at the molecular level were both selectively neutral, and may even be 'clock-like' in the sense of maintaining a consistent substitution rate (Zuckerkandl and Pauling 1965; Kimura *et al.* 1968). Protein data also showed genetic variation in populations of both *Drosophila* and humans to be substantially higher than previously suggested by Mendelian population genetics alone (Hubby and Lewontin 1966; Lewontin 1967). This early promise for molecular data to contribute to a more nuanced view of evolutionary theory grew rapidly with the emergence of DNA sequencing. Molecular data revolutionized existing disciplines such as systematics and population genetics, and encouraged the development of new approaches, such as coalescent theory.

The diverse disciplines that were reshaped by the proliferation of molecular sequence data were quickly united by their shared methodological foundation that was rooted from the beginning in the emerging tools of statistical phylogenetics. While phylogenies were first used as tools for the taxonomic classification of organisms (Sneath and Sokal 1962; Hennig 1965), their usefulness in understanding evolutionary dynamics at both the population (Felsenstein 1973a; Thompson 1975; Felsenstein 1981b) and interspecific (Cracraft 1974; Gingerich 1979a; Brooks 1981; Cracraft 1982) levels became quickly apparent. As

the burgeoning field of molecular evolution developed, statistical phylogenetic methods, based on parametric models of evolutionary change, rapidly developed and became heavily preferred over previous 'phenetic' and cladistic approaches (Cavalli-Sforza and Edwards 1967; Felsenstein 1978, 1981a). Some of the key benefits to the use of parametric models in phylogenetic inference included the ability to weigh alternative hypotheses using statistical criteria such as likelihood and to characterize evolutionary patterns and processes using inferred model parameters, such as substitution rates. Many important questions in molecular evolution were addressed by combining sequence datasets with increasingly sophisticated parametric phylogenetic approaches.

The field of molecular evolution underwent another major shift that began in the early 2000's with the emergence of next-generation sequencing (NGS). Instead of being confined to the single genomic regions accessible through Sanger sequencing, NGS approaches enabled researchers to examine evolutionary and phylogenetic patterns across entire genomes. Combining NGS datasets with phylogenetic methods has facilitated many examinations of evolutionary processes with previously unimaginable detail. For example, NGS data have both revealed the extent and enabled the disentangling of discordance in the phylogenetic signal displayed by different genes that occurs due to population processes (Maddison and Knowles 2006; Fontaine *et al.* 2015; Pease *et al.* 2016). NGS data have also contributed to an understanding of the importance of gene and whole genome duplication in shaping patterns in phylogeny and molecular evolution among non-model organisms (Dehal and Boore 2005; De Bodt *et al.* 2005; Cui *et al.* 2006; Smith *et al.* 2015, 2018a).

Although a comprehensive review of all of the insights and breakthroughs in evolutionary biology that have been facilitated by the synergistic emergence and growth of molecular data and parametric phylogenetic approaches falls outside the scope of relevance for this work, a key lesson from molecular evolution is the immense potential for advances in

approaches to both the collection and analysis of data to drive revolutionary innovations in our understanding of evolutionary processes.

1.4 A burgeoning revolution in evolutionary paleobiology?

The culmination of the theoretical contributions from evolutionary paleobiologists coincided with the rise of molecular evolution during the 1980's. On the whole, the molecular revolution has spurred massive changes in evolutionary biology since the publication of John Maynard Smith's welcoming of paleontology to the 'high table'. Although paleontology has remained a source of important concepts, the field has been limited in its ability to empirically evaluate hypotheses in comparison to molecular evolution. While there have been many important developments in both data collection and model-based analysis, the potential for evolutionary paleobiology as a theoretical and empirical discipline, akin to molecular evolution, has yet to be fully realized. However, recent advances in the collection and public accessibility of morphologic datasets (Boyer *et al.* 2015, 2016; Pomidor *et al.* 2016) hint at the potential for a data revolution in paleobiology.

As the availability of morphological datasets spanning increasingly large clades of fossil and living taxa increases, methods for their analysis are quickly becoming a limiting factor. The construction of molecular evolution over a methodological foundation derived from statistical phylogenetics can provide an indication of the necessity moving forward for a revitalized set of approaches for phylogenetic inference among fossil taxa. Although the important place for phylogenetics in understanding the evolutionary dynamics that have shaped the fossil records has long been recognized (Fisher 1991), paleontology has been limited by antiquated cladistic approaches until very recently.

The strong sway held over paleontology by cladistic dogmatism has resulted at least partially from the limitations in collecting morphologic data. Data matrices of manually-

coded qualitative character states have long been the dominant medium of information in paleontology. The emergence of geometric morphometrics in the 1980's appeared promising (Rohlf and Marcus 1993). Although there was substantial optimism during this time for the development of a fully statistical, quantitative framework for morphological phylogenetics (Felsenstein 1988), the following years witnessed substantial controversy over the usability of quantitative traits generated through geometric approaches in phylogenetics (Zelditch *et al.* 1995; Adams and Rosenberg 1998; Rohlf 1998; Felsenstein 2002; MacLeod 2002; Rohlf 2002). Since this time, new approaches to morphometrics have reignited the potential for a data-rich future in morphology (Boyer *et al.* 2015, 2016; Pomidor *et al.* 2016). However, new statistical phylogenetic approaches will be needed to facilitate the growth of this new generation of morphological phylogenetics.

The fragmentary nature of the fossil record presents additional challenges for the development of a new framework for paleobiology. The incomplete anatomical sampling of many fossil taxa can create challenges in confidently reconstructing evolutionary relationships from morphology alone. Although the past and emerging methodological advances in morphology discussed above are likely to open many new possibilities in paleontological research, incompleteness in the sampling of the fossil record will remain a substantial challenge. One approach that has improved resolution of phylogeny among fossil taxa in the past has been the addition of stratigraphic information when evaluating phylogenetic hypotheses (Gingerich 1974, 1979a; Fisher *et al.* 1994; Fisher 2008). These previous approaches, referred to as stratophenetics and stratocladistics, evaluated support for phylogenetic trees by combining evidence from both morphologic and stratigraphic data, ultimately selecting the tree that provided the best fit across both data types. The incorporation of these stratigraphic approaches, reimagined through a statistical lens, will be important in developing a new data-rich era in evolutionary paleobiology.

In this dissertation, I have aimed to develop a rebuilt statistical methodological foundation for the next generation of evolutionary paleobiology. The new methods that I have developed in the course of this work embrace two data sources that have been historically unconventional to paleontology: quantitative traits and stratigraphic range data. Through a combination of simulation-based pilot studies, development of new computational algorithms and statistical approaches, and empirical case studies, I have attempted to provide the groundwork for a new era of evolutionary paleobiology. Like the molecular revolution of the 1980's and 1990's, this foundation is based on a new set of statistical phylogenetic approaches. The first three chapters provide new computational phylogenetic methods for quantitative traits and stratigraphic data. In the fourth and last chapter, I build upon these developments to provide an example of how they may be harnessed to evaluate major evolutionary questions throughout the fossil record.

While many previous contributions to analytical paleobiology have focused on the marine invertebrate record (Foote and Raup 1996; Wagner 1996; Solow and Smith 1997, for example), the methods developed in the course of this dissertation are geared primarily toward analysis of the terrestrial vertebrate fossil record. This is motivated in part by my interests as a vertebrate paleontologist, but is also intended to provide a somewhat limiting case, where the methods are structured to accommodate the poor sampling in the vertebrate fossil record relative to the marine invertebrate record. Extending beyond this extreme, I even provide a paleobotanical case study, by applying one of my methods to fossil taxa from the grape family (Vitaceae). Nevertheless, although they show utility in the poorly sampled records of vertebrates and plants, the methods that I have developed here will also be applicable and useful in more densely sampled taxa.

1.5 Chapter summaries

In **chapter 2**, I provide a pilot study using simulated data to evaluate the feasibility of employing quantitative trait data in parametric phylogenetic inference. I use the trials on simulated data as a launching point to advocate for a shift among the paleobiological community from emphasizing traditionally scored qualitative traits to continuous traits quantified using traditional or geometric morphometric methods. Such a shift would facilitate the assembly of comprehensive datasets that reduce bias and subjectivity in manual character coding and provide a more faithful and resolute representation of interspecific morphological variation. In some ways, the ability for quantitative traits to inform phylogeny is self-evident. Their use was also commonplace in the pre-cladistics era of paleontology, which developed a long tradition of reconstructing phylogeny by plotting (strato)phenetic similarity in quantitative traits such as body size or molar width (Simpson 1944; Gingerich 1979b). However, my goal in this study was to 1) provide concrete documentation for their utility in phylogenetic inference when placed in an evaluative, parametric framework and 2) use this demonstration as a basis to explore issues relating to the fidelity of phylogenetic information content and to construct a vision for a future of paleontology that is more rigorous and efficient in its treatment of morphologic data.

In **chapter 3**, I present an implementation of a method for inferring fossil placements and phylogeny from continuous traits. I also provide two empirical demonstrations of the method, where I explore the ability of the method to infer the phylogenetic placements of fossil taxa along an extant scaffolding. These empirical trials explore the performance of the method on both geometric and traditional morphometric data representing the mammalian order Carnivora and the flowering plant family Vitaceae (grapes), respectively. In addition to validating the implementation and evaluating the performance of different data

types, I explore the utility of weighing the contribution of each trait based on its assessed reliability. This weighing procedure is intended to separate phylogenetically reliable from unreliable traits through the construction of a Bayesian prior, where traits that are not consistent with the extant scaffolding are de-emphasized when inferring the positions of fossil taxa. This procedure is designed to bolster the accuracy and confidence in the phylogenetic placement of fossil taxa. Another major goal in this work was to provide a demonstration for the advantages of gathering comprehensive quantitative morphologic datasets. Using the procedures that I introduce in the chapter, the reliability of signal can be assessed objectively using an explicit reference point. This offers a procedural and epistemological improvement over the construction and analysis of traditional cladistic datasets, where sampling of traits is often biased (relative to a comprehensive sampling of all accessible morphologic features) according to subjective judgments of their reliability.

In **chapter 4**, I designed and implemented a parametric reformulation of the stratocladistic approach to phylogenetic inference (Fisher *et al.* 1994; Fisher 2008). This approach was intended to faithfully reproduce the logic and goals of the original parsimony-based stratocladistics within a maximum-likelihood (ML) statistical framework. The defining feature of my approach, like stratocladistics, is the merging of temporal and morphologic evidence to select a best-supported tree that includes both collateral and ancestor-descendant relationships. I applied this framework to develop a stronger understanding of the evolutionary relationships among taxa within the hominin lineage. While paleoanthropological hypotheses often consider direct ancestry between hominin species, the phylogenetic approaches that have been applied to hominins in the recent past have only considered collateral relationships. As a result, hominin relationships have remained highly contentious, owing to the lack of statistical support for any of the detailed qualitative hypotheses proposed by paleoanthropologists and confusing patterns supported by previous

phylogenetic analyses. The ML stratocladistic approach that I introduced provided strong support for several ancestor-descendant hypotheses that align well with common qualitative interpretations.

In **chapter 5**, I introduced a new procedure that reconstructs the mosaic patterns in evolutionary disparity displayed by suites of continuous traits. I introduce and describe a new method through which to interrogate large datasets of phenotypic traits to reveal the shared and divergent patterns across anatomical regions that have driven their disparification through time by identifying modules that display similar patterns in rate and disparity across phylogenetic trees. By identifying this structure, the approach identifies specific patterns in mosaic evolution that describe change in phenotypes throughout time. Using both simulated and empirical datasets, I demonstrate the capability of the approach to identify mosaic evolution and consider the potential of the method to contribute to a more detailed view of the diversity of patterns in phenotypic evolution across lineages and throughout evolutionary time.

In this dissertation, I have aimed to produce a meaningful contribution toward the construction of a renewed, computational, evolutionary paleobiology. I recognize that the approaches that I have developed in the course of this work represent a skeletal fulfillment toward this goal. However, I hope that the contributions presented in this dissertation will contribute to a broader dialectic in their attempt to reconsider the ways in which we reconstruct macroevolutionary and phylogenetic patterns from the fossil record. Although the generational task of developing and incorporating a new set of approaches and applications into the paleobiological mainstream is daunting, I feel that it will be necessary in developing a renewed conceptual and theoretical synthesis between neontological and paleontological evolutionary biology. Moving forward, many creative developments in methods and empirical applications will surely unleash unprecedented insight into the dis-

parate and complex patterns and processes that have shaped the tree of life.

CHAPTER II

Continuous Trait Phylogenetics

Preamble: The contents of this chapter have been published in *Systematic Biology*. The published version appears as: Parins-Fukuchi, Caroline. Use of continuous traits can improve morphological phylogenetics. *Systematic Biology* 67.2 (2018): 328-339.

2.1 Abstract

The recent surge in enthusiasm for simultaneously inferring relationships between extinct and extant species has reinvigorated interest in statistical approaches for modeling morphological evolution. Current statistical methods use the Mk model to describe substitutions between discrete character states. Although representing a significant step forward, the Mk model presents challenges in biological interpretation, and its adequacy in modeling morphological evolution has not been well explored. Another major hurdle in morphological phylogenetics concerns the process of coding discrete characters. The often subjective nature of discrete character coding can generate discordant results that are rooted in individual researchers' subjective interpretations. Employing continuous measurements to infer phylogenies may alleviate some of these issues. Although not widely used in the inference of topology, models describing the evolution of continuous characters have been well examined, and their statistical behavior is well understood. Also, continuous measurements avoid the substantial ambiguity often associated with the assignment

of discrete states to characters. I present a set of simulations to determine whether use of continuous characters is a feasible alternative or supplement to discrete characters for inferring phylogeny. I compare relative reconstruction accuracy by inferring phylogenies from simulated continuous and discrete characters. These tests demonstrate significant promise for continuous traits by demonstrating their higher overall accuracy as compared to reconstruction from discrete characters under Mk when simulated under unbounded Brownian motion, and equal performance when simulated under an Ornstein-Uhlenbeck model. Continuous characters also perform reasonably well in the presence of covariance between sites. I argue that inferring phylogenies directly from continuous traits may benefit efforts to maximize phylogenetic information in morphological datasets by preserving larger variation in state space compared to many discretization schemes. I also suggest that the use of continuous trait models in phylogenetic reconstruction may alleviate potential concerns of discrete character model adequacy, while identifying areas that require further study in this area. This study provides an initial controlled demonstration of the efficacy of continuous characters in phylogenetic inference.

2.2 Introduction

The development and widespread adoption of statistical phylogenetic methods has revolutionized disparate disciplines in evolutionary biology, epidemiology, and systematics. Studies utilizing maximum-likelihood (ML) and Bayesian approaches have become the preferred means to analyze molecular data, largely eclipsing parsimony and distance methods. Despite this, approaches which draw inference from morphological data have remained comparatively underdeveloped (but see relevant discussion and citations below). As a result, non-probabilistic tree inference methods have continued to be employed for the phylogenetic analysis of morphological characters. Nonetheless, several landmark ad-

vances in the development of statistical morphological phylogenetic methods have demonstrated the benefits of further developing this framework. This will be particularly important in the near future as burgeoning approaches enabling the rapid collection of morphological data may begin to outstrip methods through which to analyze them (Chang and Alfaro 2015b,a). This may significantly alter and enhance our view of the tree of life, especially considering that the majority of macro-organisms, represented by fossil taxa, can only be analyzsed from their morphology.

A foundational contribution in morphological phylogenetics has been the Mk model of discrete trait evolution (Lewis 2001). This is a version of the Jukes-Cantor model of nucleotide substitution generalized to accommodate varying numbers of character states (Jukes and Cantor 1969). Extensions to this model accommodate biased sampling of parsimony informative characters (Lewis 2001), rate heterogeneity between sites (Wagner 2012), and asymmetric transition rates (Ronquist and Huelsenbeck 2003; Wright *et al.* 2015). The deployment of this model has demonstrated the utility of statistical approaches to morphological phylogenetics. Such approaches improve estimates of uncertainty over non-probabilistic approaches, enable a clearer statement of modeling assumptions, and enable branch length estimation. This has enabled a better understanding of much of the fossil tree of life (Dávalos *et al.* 2014; Pattinson *et al.* 2014; Dembo *et al.* 2015). These approaches have also enabled the application of tip dating methods to the combined analysis of extinct taxa represented by morphological data with extant taxa (Nylander *et al.* 2004; Ronquist *et al.* 2012). These total evidence tip dating methods have been widely used since their introduction, and are implemented in the BEAST (Bouckaert *et al.* 2014) and MrBayes (Ronquist and Huelsenbeck 2003) packages. These have more clearly resolved the timing of species divergences and relationships between fossil and living taxa (Wiens *et al.* 2010; Wood *et al.* 2012; Lee *et al.* 2013, 2014, but see Arcila *et al.* 2015). Overall,

probabilistic approaches to morphological phylogenetics appear to represent an improvement in accuracy compared to cladistic methods, and are indispensable in their distinct ability to allow the estimation of branch lengths and evolutionary rate. The benefits of a statistical total-evidence framework as applied to fossil taxa will only become clearer as more data become available and improved methods are developed (Pennell and Harmon 2013; Lee and Palci 2015).

Despite the these strides, discrete character models represent an imperfect solution in their current usage. Although Bayesian inference under Mk appears to outperform parsimony under certain conditions, error increases at high evolutionary rates (Wright and Hillis 2014). Also, under many circumstances, phylogenetic inference under the Mk model includes imprecision and uncertainty, both in simulations (O'Reilly *et al.* 2016) and empirical studies (Lee and Worthy 2012; Dembo *et al.* 2015). Previous researchers have also expressed concerns over the efficacy of model-based approaches in the presence of missing data (Livezey and Zusi 2007; O'leary *et al.* 2013). However, these have been assuaged and any issues arising from missing data are likely not specific to probabilistic approaches (Wright and Hillis 2014; Guillerme and Cooper 2016). Another potential issue is the lack of clarity in interpreting the Mk model biologically. Although transition rates have a strong theoretical and empirical basis in population genetics, their significance beyond serving as nuisance parameters is less straightforward when applied to morphological data. Discrete morphological characters may not undergo change in a manner analogous to nucleotides, which are well understood to alternate between states repeatedly. Conversely, many characters used for phylogenetic inference consist of single, parsimony informative directional changes between taxa (Klopfstein *et al.* 2015). It is unclear how adequately discrete Markov models describe such variation. The Mk model itself does not accommodate directional evolution, and previous researchers have questioned the adequacy of

existing discrete character models (Ronquist *et al.* 2016). This is particularly important when considering the importance of branch lengths in total evidence tip dating methods discussed above, but may also be expected to mislead inference of topology.

Aside from the modeling concerns discussed above, discrete morphological characters present a non-trivial set of challenges to phylogenetics that are distinct from those presented by molecular data. Perhaps foremost among these is disagreement between researchers in the categorisation, ordering, and weighing of discrete character states (Farris 1990; Hauser and Presch 1991; Pleijel 1995; Wilkinson 1995). Despite extensive discussion among comparative biologists, the interpretive nature of the process of character coding has continued to leave major palaenotological questions unresolved (Upchurch 1995; Wilson and Sereno 1998; Bloch and Boyer 2002; Kirk *et al.* 2003).

Although continuous traits share with discrete traits a reliance on a pre-specified character concept, and so retain several the conceptual challenges in morphological analyses, they may help to improve some of the most egregious challenges discussed above. They can be collected more objectively than qualitative observations and do not require ordering of states. Their use in phylogenetic inference has been discussed among the earliest advancements in statistical phylogenetics (Cavalli-Sforza and Edwards 1967; Felsenstein 1973a), and their phylogenetic informativeness has been demonstrated empirically (Goloboff *et al.* 2006; Smith and Hendricks 2013). Still, the use of continuous characters for the inference of phylogenetic topology has remained uncommon, with methods for their use in phylogenetics remaining relatively poorly examined beyond the foundational works referenced above. Although many palaeontological studies incorporate continuous measurements, they are binned into categories and analysed as discrete. However, since fossil data are often scarce, it may be beneficial to maximise the amount of information gleaned from available specimens by representing such variation in its entirety.

Another potential benefit to inferring phylogeny from continuous characters is the wealth of models developed in phylogenetic comparative methods to describe their evolution. Most comparative models of continuous trait evolution belong to the Gaussian class, which are also well utilized in disparate fields such as physics, economics, and engineering. In comparative biology, they are used to describe stochastic Markovian movement through continuous trait space along continuous time. This class of models includes Brownian motion (BM) (Felsenstein 1973a, 1985; Gingerich 1993), Ornstein-Uhlenbeck (OU) (Hansen 1997a; Butler and King 2004; Beaulieu *et al.* 2012), and Lévy processes (Landis *et al.* 2013). Under BM, evolution is described as a random walk, with phenotypic change being normally distributed with a mean displacement of zero, and variance σ^2. OU models expand upon this by introducing terms producing a stabilizing force which stabilizes movement around an optimal trait value, while Lévy processes contain terms producing saltational jumps in character space, interspersed either by BM diffusion or stasis. Two major benefits to Gaussian models in phylogenetics are their relatively straightforward interpretability and the relative ease of deriving mathematical extensions to describe a range of biological processes.

Given the existence of well understood and clearly interpretable models describing their evolution, the use of continuous traits may offer several advantages over discrete characters in phylogenetic inference. However, their behaviour is not well understood when applied to the inference of phylogenetic topology, and so further investigation is needed. In addition, there are potential hurdles to their efficacy. Possibly foremost among these is the widespread covariance between continuous measurements that is expected through both genetic and morphometric perspectives (Lynch *et al.* 1998; Uyeda *et al.* 2015; Adams and Felice 2014). Nevertheless, the expected magnitude in covariance among continuous morphological measurements and the robustness of phylogenetic methods to this violation

is not known. Furthermore, it is also generally reasonable to expect evolutionary covariance between nucleotide sites, and phylogenetic methods that do not accommodate for this are routinely applied to molecular data.

In this study, I carry out simulations to compare the relative performance of binary discrete and continuous characters at reconstructing phylogenetic relationships. Simulations of continuous characters were designed to reflect a range of scenarios that may influence accuracy including overall evolutionary rate and matrix sizes. I also conduct inference on continuous traits that have undergone correlated evolution, an important violation to single-rate BM thought to be widespread in continuous character evolution.

2.3 Methods

2.3.1 Simulations

I generated a set of 100 pure birth trees using the Phytools package (Revell 2012a) in R (R Core Team 2016), each containing ten taxa. All trees were ultrametric and generated with a total length of 1.0 units for consistency in parameter scaling for trait simulations (Fig. 2-1). These trees were used to simulate continuous characters evolving along an unbounded BM process, again using Phytools. This is a Markovian process in continuous time where the variance of the process can increase infinitely through time. This differs from the BM σ^2 parameter, which gives the variance in the amount of character displacement at each draw, effectively describing the magnitude of the random BM walk or a rate of character displacement. To assess performance across several biological scenarios, traits were simulated at σ^2 parameterizations of 0.05, 0.5, 1.0, 1.5, and 3. Since the process under which traits were simulated is unbounded, phylogenetic signal is expected to remain consistent across rates (Revell *et al.* 2008), but different rates were chosen to illustrate this consistency and to provide even comparison to discrete trait simulations.

Discrete characters were simulated in the Phytools package (Revell 2012a) under an Mk model with homogeneous transition probabilities. Traits were generated at transition rates 0.05, 0.5, 1.0, 1.5, and 3. All character matrices were generated without rate heterogeneity, and include invariable sites (i.e. no acquisition bias).

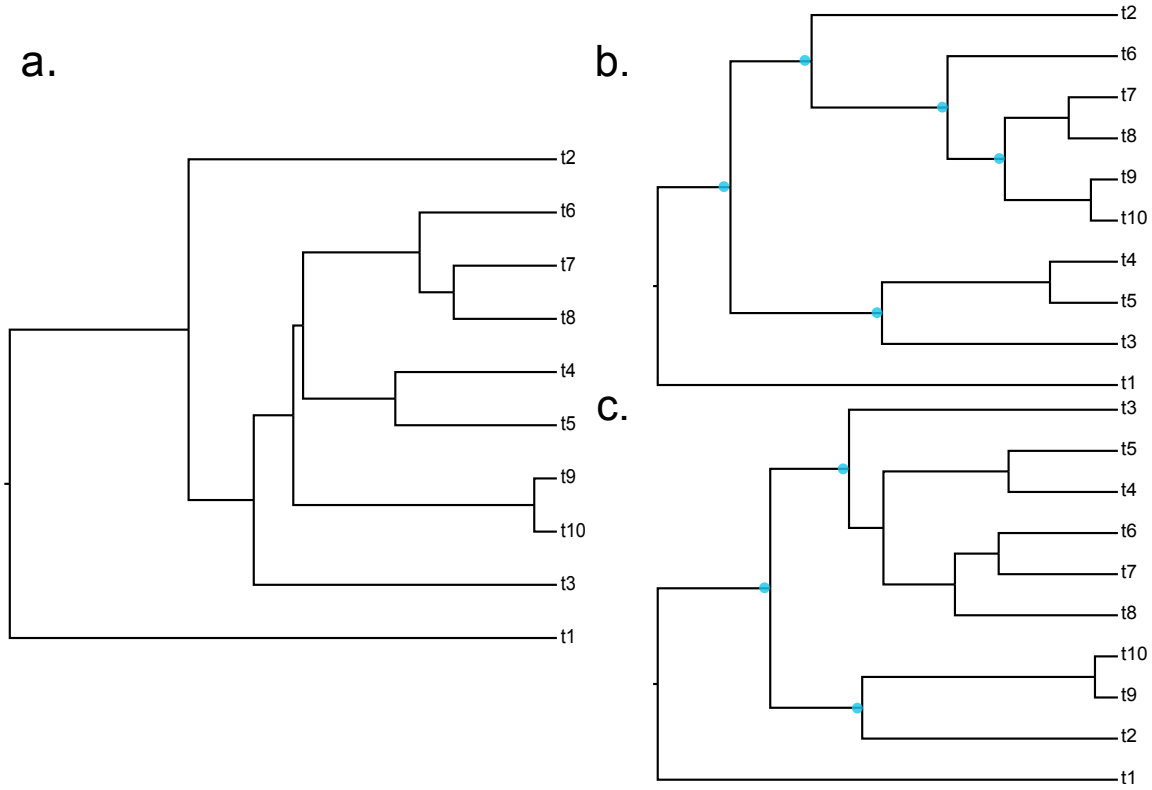

Figure 2.1: A) Exemplar true simulated tree. B) Tree inferred from 20 discrete characters simulated under Mk from true tree. C) Tree inferred from 20 continuous characters simulated under Brownian motion. Blue dots denote incorrect bipartitions.

Matrices containing 500 traits were generated and randomly subsampled to create smaller sets of 20 and 100 characters to reflect a range of sampling depths. These were chosen because many published morphological matrices fall within this range. The subsampled matrix sizes were chosen to represent reasonably sized palaeontological datasets, while the 500 trait matrices were tested to assess performance when data are abundant. While such large datasets are uncommon in morphology, several studies have produced character matrices of this size, and for continuous characters, it may be feasible to gener-

ate such large datasets from morphometric data.

I also simulated continuous characters under an OU model parameterised without directional drift ($\theta = 0$), and with the stabilizing (α) parameter set to yield the same phylogenetic half-life present in the binary Mk model used for comparison. For OU continuous characters, phylogenetic half-life is defined by:

$$\frac{\log(2)}{\alpha} \tag{2.1}$$

and for binary discrete characters as:

$$\frac{\log(2)}{(q01 + q10)} \tag{2.2}$$

With $q01$ and $q10$ corresponding to the respective transition rates between binary character states.

When phylogenetic half-life is set to be equal, phylogenetic constraint should be the same between both sets of characters in the sense that they reach saturation over the same timescale. This comparison examines whether either data source performs inherently better when phylogenetic signal is held constant. These data were generated in matrices of 100 traits at an evolutionary rate of 0.5. Because the phylogenetic information content of both sets of constrained traits should be the same, both sets are expected to perform similarly. Nevertheless, this comparison provides a control by assessing whether unknown differences in the behaviour of each model (or other properties of each method) themselves lead to any differences in reconstruction accuracy.

Data were also generated under a correlated BM process to mimic inference in the presence of multidimensionality. These datasets were constructed at covariance strengths of 0.1, 0.5, and 0.9 and covarying dimensions of 5 and 25 traits. These were chosen to represent situations where traits range from being loosely to tightly correlated to each another,

and where the number of correlated dimensions is large to small. Although differing, these values were chosen to loosely follow the scheme of Adams and Felice (2014).

2.3.2 Estimation of phylogenies and reconstruction accuracy

I estimated Bayesian phylogenetic trees under a single rate BM model for all sets of continuous characters using RevBayes (Höhna *et al.* 2016). Trait likelihoods were computed after Felsenstein (1973a, 1985). MCMC simulations were run for 150,000-1,000,000 generations and checked manually for convergence using Tracer v1.6. Runs were accepted when the effective sample size (ESS) for logged parameters exceeded 200. Trees were inferred from discrete data in MrBayes version 3.2.6 (Ronquist and Huelsenbeck 2003), simulating for 1,000,000 generations. Different programs were used because, while MrBayes remains the standard in the field for Bayesian phylogenetic inference, its current version does not implement likelihood functions for continuous character models. So the continuous character approach needed to be developed in RevBayes, however, I preferred to remain with the standard and proven implementation where possible. For both continuous and discrete characters, I incorporated a birth-death prior on node heights. This was done to enable an even comparison of branch lengths obtained through both methods that are scaled to time. Example configuration files for RevBayes and MrBayes analyses are provided as supplementary data. Tree distributions were summarized using TreeAnnotator version 2.4.2 (Rambaut and Drummond 2013) to yield maximum clade credibility (MCC) topologies. MCC trees maximize the posterior probability of each individual clade, summarizing across all trees sampled during MCMC simulation. Once summarised, all trees were rescaled to match inferred tree lengths to the true trees using Phyx (https://github.com/FePhyFoFum/phyx).

I assessed topological accuracy from simulated trait data using the symmetric (Robinson-Foulds) distance measure (Robinson and Foulds 1981), giving the topological distance

between true trees and inferred trees. Symmetric distance is calculated as a count of the number of shared and unshared partitions between compared trees. As such, the maximum symmetric distance between two unrooted trees can be calculated as *2(N-3)*. These values were then scaled to the total possible symmetric distance for interpretability. Additionally, I measured error in branch length reconstruction using the branch length distance (BLD) (Kuhner and Felsenstein 1994). This is calculated as the sum of the vector representing the individual differences between the branch lengths of all shared bipartitions. The scale of this value depends on the lengths of the trees under comparison. If trees of different lengths are compared, BLD can be very high. However, in this study, all trees are scaled to a root height of 1 to allow comparison of topological and internal branch length reconstruction error. All distances were calculated using the DendroPy Python package (Sukumaran and Holder 2010). Summary barplots were constructed using ggplot2 (Wickham 2016).

2.4 Results

2.4.1 Unconstrained and independently evolving continuous traits

Topological reconstruction error is lower overall for trees estimated from continuous characters than from binary discrete (Fig. 2-2a, Fig 2-7a). For discrete characters, symmetric distance increases significantly at high evolutionary rates, likely due to saturation and loss of phylogenetic signal. Distance also increases in discrete characters when rate is very slow, due to lack of time for phylogenetic signal to develop. This pattern is similar to that recovered by (Wright and Hillis 2014) in their test of Bayesian inference of Mk, which revealed highest topological error at very low and high rates. As expected, continuous characters perform consistently across rates because saturation cannot occur, even at very fast rates. Because of the differing sensitivities of each data type to evolutionary rate, topological error should also be compared using the most favourable rate class for

discrete characters, 0.5 substitutions per million years (Fig. 2-2b, Supp. Fig. 2-1b). Even at this rate, continuous reconstruction performs more consistently than discrete, with error more tightly distributed around a slightly lower mean. A likely explanation is that discrete characters retain less information that continuous characters. The small state space of the binary character model likely causes phylogenetic signal to become saturated more quickly at fast rates, and develop too slowly at slow rates than multi-state characters. BM and Mk appear to perform fairly similarly in reconstructing branch lengths (Fig. 2-2). The pattern across rates and matrix sizes is very similar between BLD and symmetric distances, with the fastest rates producing the most error. This likely results from increased saturation at fast rates, causing underestimation of hidden character changes.

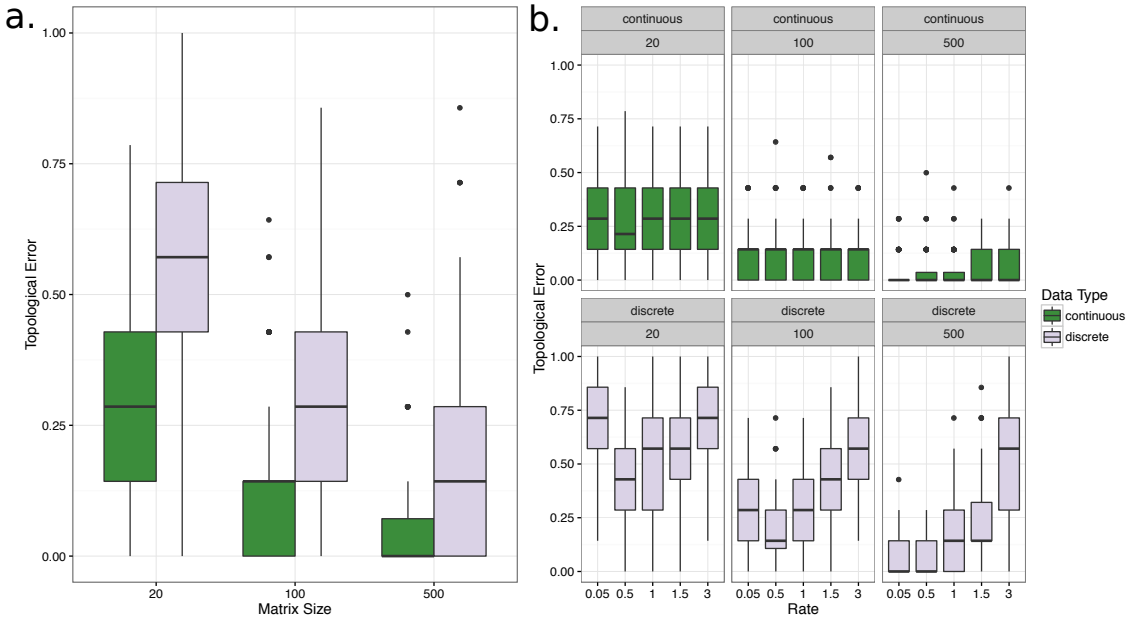

Figure 2.2: Topological error calculated as the proportion of maximum symmetric distance across trees estimated from independently evolving continuous characters. A) Error averaged across all rates except for the highest rate category, which resulted in the highest error when inferring under Mk. B) Error across all matrix sizes and rates.

Matrix size has a major impact on tree reconstruction accuracy. Estimations from both discrete and continuous traits improve substantially at each increasing matrix size (Fig. 2-2). Estimates from 20-character matrices possess fairly high error in both data types,

with approximately 1 in 5 bipartitions being incorrectly estimated from continuous characters, and 2 in 5 being incorrectly estimated from discrete data. Increasing matrix size to 100 traits improves accuracy significantly, with both data types estimating approximately 1 in 10 bipartitions incorrectly. Although at several rates, mean symmetric distance compared between data types is close, continuous characters tend to be less widely distributed, and thus appear to reconstruct trees with more consistent accuracy. When matrix size is increased to 500 characters, both continuous and discrete characters are able to recover phylogeny with very high accuracy, except for at very fast rates, where discrete characters estimate approximately half of all bipartitions incorrectly on average.

2.4.2 Continuous traits evolving under selective constraint

Phylogenies inferred from continuous traits simulated under an OU model achieve virtually identical performance to binary discrete characters simulated under the same phylogenetic constraint (Fig. 2-3). Both sets of characters display a very similar range of error, with approximately 15% of bipartitions estimated incorrectly on average. This result demonstrates that any performance increases observed for continuous traits over discrete traits result from differences in realised phylogenetic information.

2.4.3 Covarying continuous characters

Tree inference under BM appears relatively robust to the violation of co-evolving continuous characters. Although error is recognisably greater with strong covariance and many trait dimensions, symmetric distance remains close to values from uncorrelated traits at lower covariance strengths and/or fewer trait dimensions (Fig. 2-4). When correlated traits are of low dimensionality and covariance strength, reconstruction appears to be nearly as accurate as uncorrelated traits, with all bipartitions estimated correctly on average. As covariance strength and dimensionality are increased to intermediate val-

ues, topological error increases such that between 0 and 17% of bipartitions are estimated incorrectly, with a wider distribution than is present at the lowest values. Accuracy is most diminished when covariance is strongest and dimensionality is largest, with most reconstructions estimating between 17-29% of bipartitions incorrectly. Although statistical significance cannot be estimated for BLD and symmetric distance, estimation under low to intermediate trait covariance appears at least qualitatively similar, albeit slightly worse, to uncorrelated continuous and binary discrete characters. The decreases in accuracy observed can likely be attributed to the decrease in total information content caused by covariance. This reduces the effective amount of data from which to draw inference. This is reflected in the results, with higher covariances and dimensionalities reconstructing trees with a similar magnitude of error as is shown for the 100 character datasets.

2.5 Discussion

The results demonstrate that phylogenetic reconstruction from continuous trait data can provide a reasonable supplement or alternative to inference from discrete characters. Continuous characters that are unconstrained and unbounded in their evolution outperform discrete characters, and perform equally well when constrained by selection. The unconstrained traits' resilience to high evolutionary rate is expected, because continuous characters evolving under an unbounded and unconstrained BM process will continue to increase in variance through time. Therefore, such characters are able to retain phylogenetic information at high evolutionary rates that may cause rampant saturation in discrete characters (Fig. 2-4). Further work is needed in this area to investigate the extent to which continuous characters are bounded and constrained in their evolution relative to discrete characters. This will be especially important moving forward, as temporal variation in evolutionary regimes and model parameters can interact in complex ways, sometimes

extending the maintenance of phylogenetic signal through time (Revell *et al.* 2008). Although continuous characters in empirical are undoubtedly constrained in their evolution, the added information contained in continuous character datasets may lessen the extent of saturation relative to discrete characters in practice.

The demonstration that performance becomes equal when the amount of phylogenetic constraint is held constant between both data sources identifies the major source of the performance increase observed in unconstrained BM traits compared to discrete traits. The average amount of phylogenetic constraint exhibited by discrete and continuous traits, however, is not well understood in empirical datasets. Conversely, the susceptibility of discrete traits to the loss of phylogenetic signal at high evolutionary rates and deep timescales has long been recognised (Hillis and Huelsenbeck 1992; Yang 1998). Although this effect is understood to affect molecular data, discrete morphological datasets may possess increased susceptibility to this effect because of the frequent use of binary character coding schemes. Discrete characters constrained to fewer states increases signal loss at high evolutionary rates due to increased levels of homoplasy, saturation, and lower information content overall (Donoghue and Ree 2000). The extent to which continuous traits are constrained in their evolution on average is not well understood. However, the results here suggest that researchers would benefit in treating continuous traits as such and inferring phylogenies under continuous trait models in order to maximise usable information contained in datasets.

My results demonstrate that the fundamental issues in comparing continuous and discrete traits are state space, selective constraint, and evolutionary boundedness. When selective constraint in continuous characters occurs at levels which restrict phylogenetic signal with the same strength as binary characters, reconstruction accuracy is predictably equal. Nevertheless, it is unclear the extent to which phylogenetic half-life in continu-

ous and discrete traits tends to differ in empirical datasets. Continuous characters may be expected to commonly evolve under some manifestation of selective constraint, but it is unclear whether such effects typically mask phylogenetic signal to the same extent as rapidly saturating binary traits.

Discrete traits with more than two states possess a significantly longer phylogenetic half-life than binary characters, but could be supplanted by continuous characters in many cases. Although empirical morphological datasets typically incorporate discrete characters with more than two states, these are typically fewer in number than binary coded characters. Multi-state characters are also typically discretized codings of continuous measurements. Such "discrete" traits would be susceptible to the same selective forces as their continuous counterparts, and so treatment of the multi-state partitions of morphological matrices as continuous can only increase the amount of phylogenetic information contained within datasets. The tendency of morphological matrices to be predominantly composed of binary characters should encourage further consideration of continuous traits in future empirical and theoretical studies.

Error in branch length estimation was fairly high with the 20-trait matrices but decreased substantially when matrix size was increased to 100 traits. Although BM and Mk achieve similar accuracy in estimating branch lengths in this study, careful thought should continue to be applied when relying upon Mk branch length estimates in the future. Branch length error may be higher when inferring under Mk from empirical datasets, since many discrete morphological matrices are constructed to include only parsimony informative characters. In these cases, characters are expected to have undergone only single synapomorphic changes. Although the lack of invariable sites in datasets tailored to parsimony is addressed through the ascertainment bias correction developed by (Lewis 2001), it is unclear how meaningfully the directional single character changes often ob-

served in these datasets can inform evolutionary rates. This mode of change, which may characterise much of discrete character evolution, differs from the population dynamics of nucleotide substitution.

Although continuous traits may often follow covarying evolutionary trajectories in nature, this appears to have a relatively minor impact on reconstruction. Accuracy was only greatly lowered in the simultaneous presence of very high dimensionality and covariance strength. Offering further support to the ability of continuous characters to reconstruct phylogeny despite evolutionary covariance, Adams and Felice (2014) also report the presence of phylogenetic information in multidimensional characters, even when the number of dimensions is greater than the number of taxa. Despite these generally positive findings, it should be noted that inference may be misled if sampling is significantly biased to include relatively small numbers of strongly correlated measurements. In these cases, it would be beneficial to examine the correlation structure and information content of the dataset to assess the amount of biased redundancy in signal.

2.5.1 Can using continuous characters benefit morphological phylogenetics?

Use of continuous traits has the benefit of reducing subjectivity in the construction of data matrices in many cases. Categorizing qualitative characters often requires subjective interpretation. However, quantitative measurements can be taken without this source of human error. This increased objectivity in the measurement of quantitative characters would expand biologists' capacity to assess statistical uncertainty. Although the likelihood approaches to morphological phylogenetics enabled by the Mk model represent a major step in this direction, discordance in tree estimates can still be attributed to differences in qualitative categorization of variation by researchers. Translation of morphological observations into data that can be analysed can present serious complications in discrete characters. Steps such as the determination of whether or not to order states, the total number of

states chosen to describe characters, and the assignment of character states can vary greatly and often yield widely different results (Hauser and Presch 1991; Pleijel 1995; Wilkinson 1995; Hawkins *et al.* 1997; Scotland and Pennington 2000; Scotland *et al.* 2003; Brazeau 2011; Simões *et al.* 2017). Continuous measurements avoid many of these issues because they can be measured, by definition, objectively and quantitatively. In addition, they may better describe variation than discrete characters. Several workers have suggested that the majority of biological variation is fundamentally continuous (Thiele 1993; Rae 1998; Wiens 2001). Although continuous characters have long been employed in phylogenetic analysis, they are generally artificially discretised, either by applying thresholds to interspecific measurements or through gross categorisations such as "large" and "small". The major disadvantage to this approach is the loss of valuable biological information. Several researchers have condemned the use of continuous characters in phylogenetics, arguing that intraspecific variation may be too great for clear phylogenetic signal to exist (Pimentcl and Riggins 1987; Chappill 1989). However, these arguments have been largely undermined by studies demonstrating the phylogenetic informativeness of continuous measurements (Goloboff *et al.* 2006; Smith and Hendricks 2013).

The expectation of correlated evolution between continuous characters has been a major argument against their use in phylogenetic reconstruction in the past (Felsenstein 1985). However, evolutionary covariance between sites is not a phenomenon that is restricted to continuous morphological characters. Population genetic theory predicts tight covariance between nucleotide sites under many conditions (e.g. Hill and Robertson 1968; Reich *et al.* 2001; Palaisa *et al.* 2004; Schlenke and Begun 2004; McVean 2007). Such covariance has also been demonstrated among discrete characters (Pagel 1994), and so this concern is not unique to continuous measurements but is shared by all phylogenetic approaches. While it is difficult to assess the relative magnitude of sitewise covariance between contin-

uous, discrete, and molecular data, examination of the correlation structure of traits may be more straightforward in continuous characters using standard regressional techniques. This would ease the identification of biased and positively misleading signal among continuous characters, enabling correction through common transformation approaches such as principal components analyses or by weighting likelihood calculations by the amount of overall variance contributed by covarying sets of characters.

The fundamentally continuous nature of many biological traits is supported by differential gene expression and quantitative trait loci mapping studies, which demonstrate their quantitative genetic basis (Andersson *et al.* 1994; Hunt *et al.* 1998; Frary *et al.* 2000; Valdar *et al.* 2006). Nevertheless, there remain well known instances where traits are truly discrete. Studies in evolutionary developmental biology have shown that many traits can be switched on or off in response to single genes controlling genetic cascades (e.g. Wilkinson *et al.* 1989; Burke *et al.* 1995; Cohn and Tickle 1999). Characters used in phylogenetic analysis are also frequently truly discrete, representing qualitative categories (eg., presence/absence). These traits may be incorporated as separate partitions into integrated analyses along with continuous measurements (Fig. 2-6). Such combined analyses can be performed in RevBayes by adding a discrete trait model, such as Mk, and discrete character data. In practice, this may improve inference from discrete characters alone, and would represent a conceptual advance in its ability to treat all available data as faithfully as is possible. Doing so may improve upon existing paradigms, which group continuous variation into multi-state discrete characters, potentially preserving more phylogenetic information. An added benefit would be the greater flexibility in modelling the evolution of such traits by making available all existing continuous trait models. An example RevBayes script for a phylogenetic analysis combining continuous and discrete characters is available in the supplement. Characters under the control of developmental expression pathways may also

exhibit very deep phylogenetic signal (De Rosa *et al.* 1999; Cook *et al.* 2001). Thus, such integrated analyses may enable the construction of large phylogenies from morphology by use of datasets containing phylogenetic signal at multiple taxonomic levels.

Depending on the extent to which individual morphometric datasets are bounded and constrained in their evolution, analysis of continuous characters may help to increase phylogenetic information. Collecting morphometric measurements in many dimensions may enable the assembly of datasets that are large in size compared to those comprised of discrete characters alone. Although large collections of morphometric measurements may be strongly covarying, analysis of the correlation structure of such datasets, as mentioned above, would enable correction for biased signal and may reveal additional phylogenetic information. This would signify a more data-scientific approach to morphological phylogenetics by enabling researchers to dissect signal present in large morphometric datasets rather than reconstruct relationships using carefully curated data matrices. Such a paradigm shift would bring morphological phylogenetics closer in spirit to phylogenomic studies and enable deeper biological inferences through co-estimation of species relationships and dynamics in trait evolution. This would provide a firm phylogenetic backing to morphometric studies, and potentially reinvigorate the field in a similar way to the previous merging of phylogenetics and genomics. Improved ability to infer phylogeny among fossil taxa would also benefit molecular phylogenetics because the incorporation of fossils into total evidence matrices can improve both inference of molecular dates and alleviate long branch attraction (Huelsenbeck 1991; Wiens 2005; Ronquist *et al.* 2012). Though further study is needed to measure the expected phylogenetic information content of both continuous and discrete traits, all of the points discussed above should urge palaeontologists to give greater consideration to continuous traits in phylogenetic analysis of evolutionary patterns and relationships. This may improve efficiency in the use of hard-won

palaeontological data by maximizing the amount of information gleaned from specimens and transform the field by facilitating new lines of questioning in palaeobiology.

And despite this optimistic tone, it should be noted that major work is still needed to provide deeper understanding of the behaviour of continuous trait models when used to infer phylogeny. It will be also important to gain a better understanding of expected empirical properties of continuous and discrete characters. As is shown here, discrete and continuous characters perform equally well when phylogenetic constraint is held constant, but there still lacks a clear characterisation of the relative expected constraint found in empirical datasets. As such, further work will be necessary to develop knowledge of the relative phylogenetic information content expressed across data types.

Moving forward, several extensions to existing Gaussian trait models should be explored. For example, further work is needed to determine the extent and distribution of rate heterogeneity between sites in continuous alignments. Since its presence has been well documented in molecular and discrete morphological data, it is likely that such rate heterogeneity is present in continuous measurements, and should be accommodated in empirical studies. Since traits can evolve under a broad range of processes, the fit of alternative models of continuous character evolution to empirical data and their adequacy in describing variation among them should also be examined.

2.5.2 Is Mk a reasonable model for discrete character evolution?

Although likelihood approaches making use of the Mk model have been increasingly adopted in morphological phylogenetics, it is unclear whether it provides a reasonable approximation of the evolutionary process. Although there are explicit theoretical links between Markov substitution models and population genetic processes (Jukes and Cantor 1969), such theory does not exist in morphology. It should also be noted that molecu-

lar data are rarely modelled using the single parameter Jukes-Cantor model, with more complex generalisations typically preferred (Felsenstein 1981a; Tavaré 1986). More sophisticated Markov processes can in principle be applied to morphological data, though this is rarely done. Nonetheless, MrBayes and RAxML implement HKY and General Time Reversible models, respectively, that can be applied to data with varying numbers of states (Ronquist and Huelsenbeck 2003; Stamatakis 2006). More work is needed to examine the adequacy of the Mk model in describing discrete character evolution. Such work will guide dataset assembly and the development of new model extensions. This is especially important in total-evidence tip dating methods employing Mk, as poor branch length estimates may weaken the ability to infer branching times. Although presenting a unique set of challenges, the use of continuous characters may alleviate some of issues concerning model misspecification. Models describing their change have been demonstrated to provide a reasonable description of character change resulting from several different microevolutionary processes (Hansen and Martins 1996). Further work is needed to address the relative adequacy of discrete and continuous trait models in describing the evolution of phenotypic data. In light of the results presented here, I suggest that continuous trait models be favoured in phylogenetic analysis in cases where morphological variation can be described quantitatively. Moving forward, deeper insight concerning the behaviour and adequacy of both discrete and continuous character models will enable increasingly powerful inferences to be drawn from morphological data. These issues will be of critical importance as advances in data collection and fossil evidence usher in an age of unprecedented discovery in morphological phylogenetics.

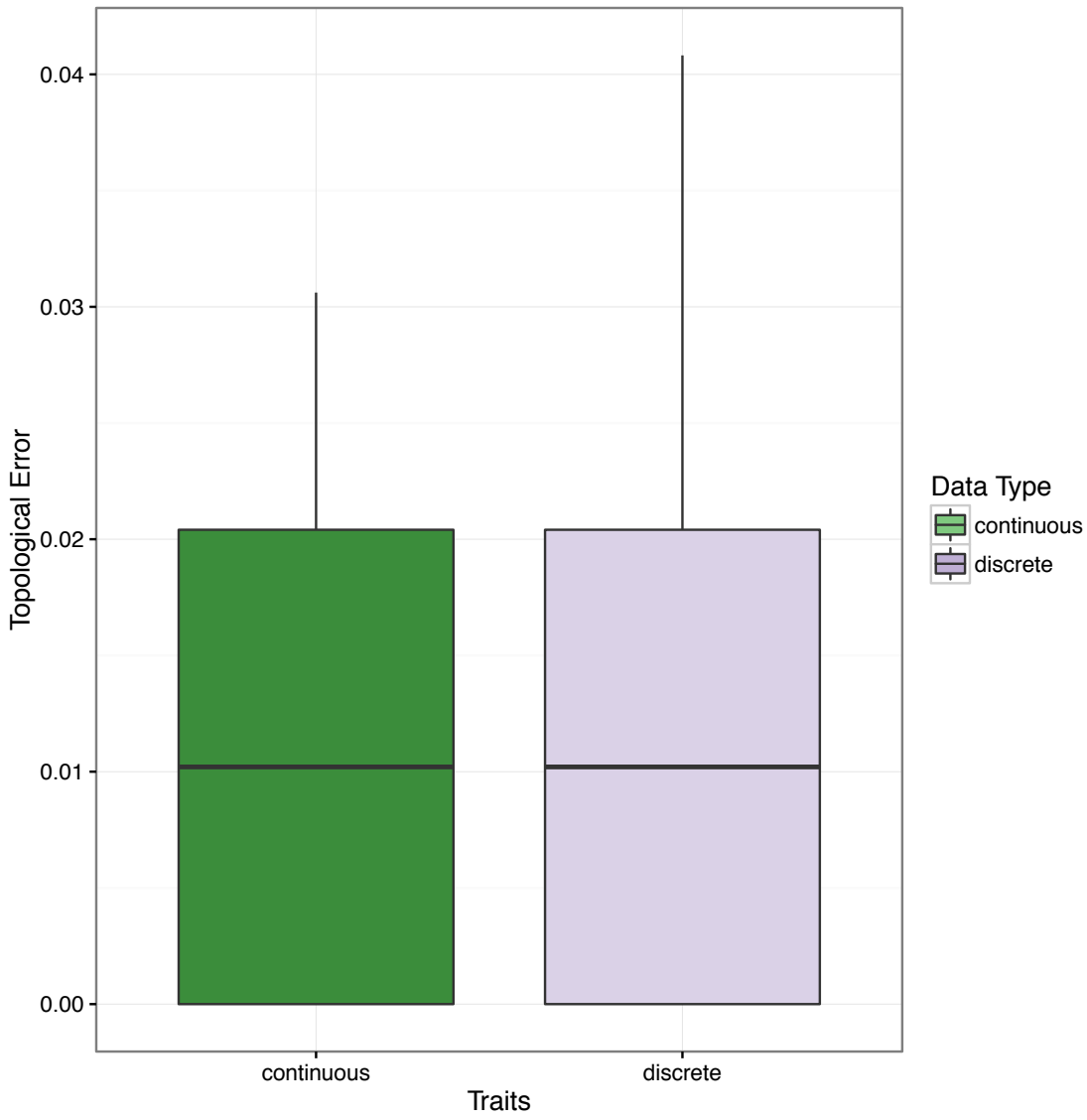

Figure 2.3: Topological error achieved after reconstructing trees from discrete traits simulated under Mk at rate 0.5, and single rate Ornstein Uhlenbeck at rate 0.5 with no directional drift and constraint set equal to the discrete characters.

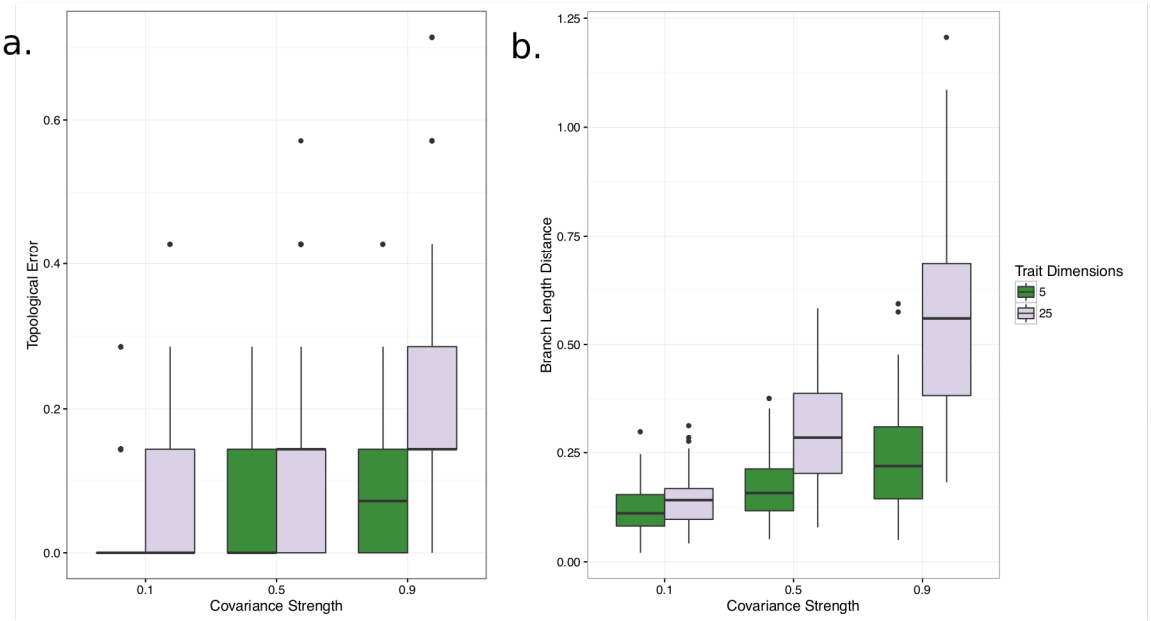

Figure 2.4: A) Topological error, calculated as proportion of maximum symmetric distance across trees estimated from covarying continuous characters. B) Branch length distance (BLD) across trees estimated from covarying continuous characters. Dimensions refers to the number of traits within covarying blocks. Covariance strength refers to the strength of the correlation between covarying characters, with a value of 0 describing to complete independence and 1 describing perfect correlation.

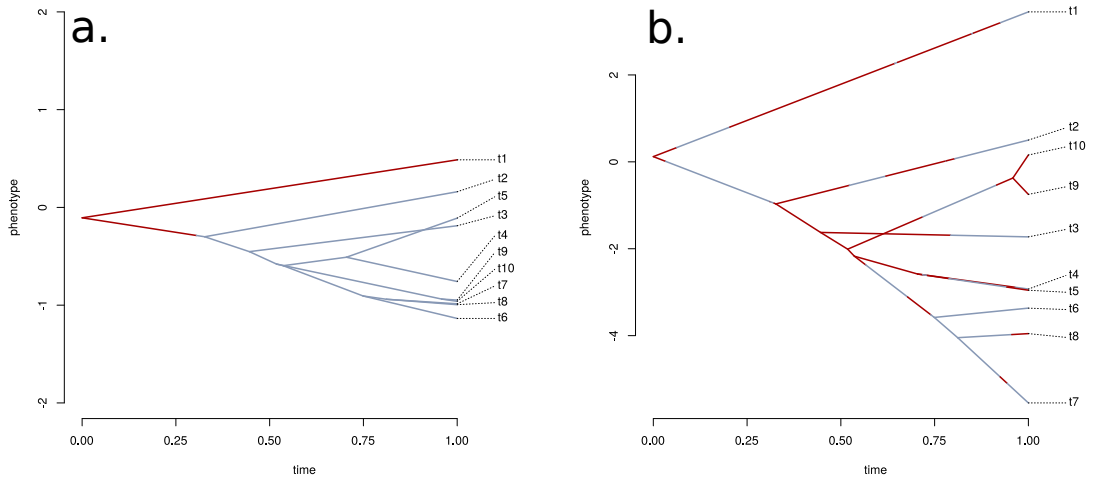

Figure 2.5: Discrete and continuous characters simulated A) at slow evolutionary rate and B) fast evolutionary rate. Y axis represents continuous phenotype. Changes in colour represent changes in discrete character state. Note how continuous characters retain phylogenetic signal at fast rates, while discrete characters saturate.

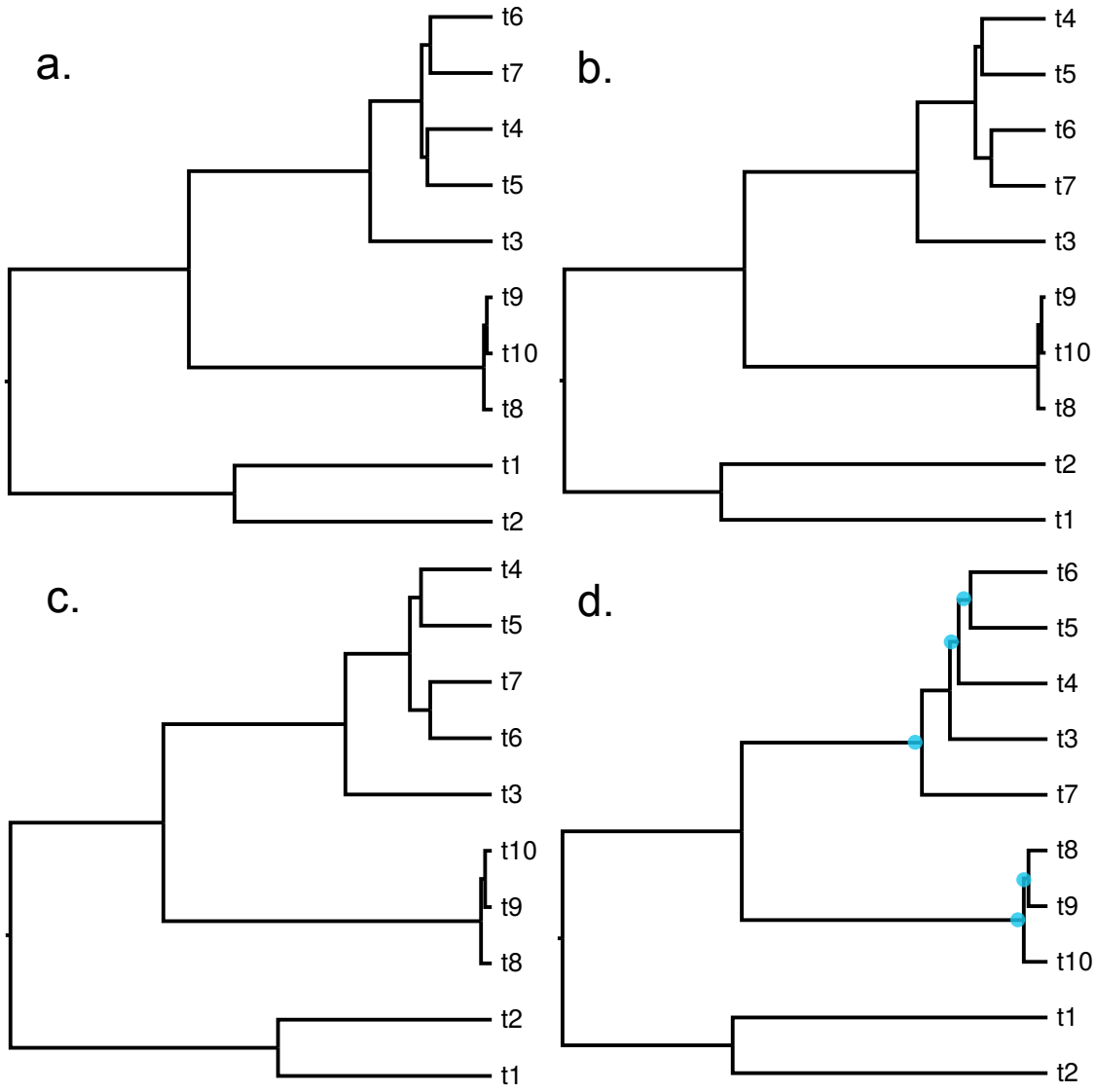

Figure 2.6: A) True tree. B) Tree estimated from 50 discrete and 50 continuous characters C) Tree estimated from 100 continuous characters simulated at rate 1.0 D) Tree estimated from 100 discrete characters simulated at rate 1.0. Blue dots signify incorrectly estimated bipartitions. The tree in panel b. was generated by randomly subsampling the matrices used to generate trees C and D, and combining into a single matrix. This matrix was analysed in RevBayes. An example script is provided in the supplement.

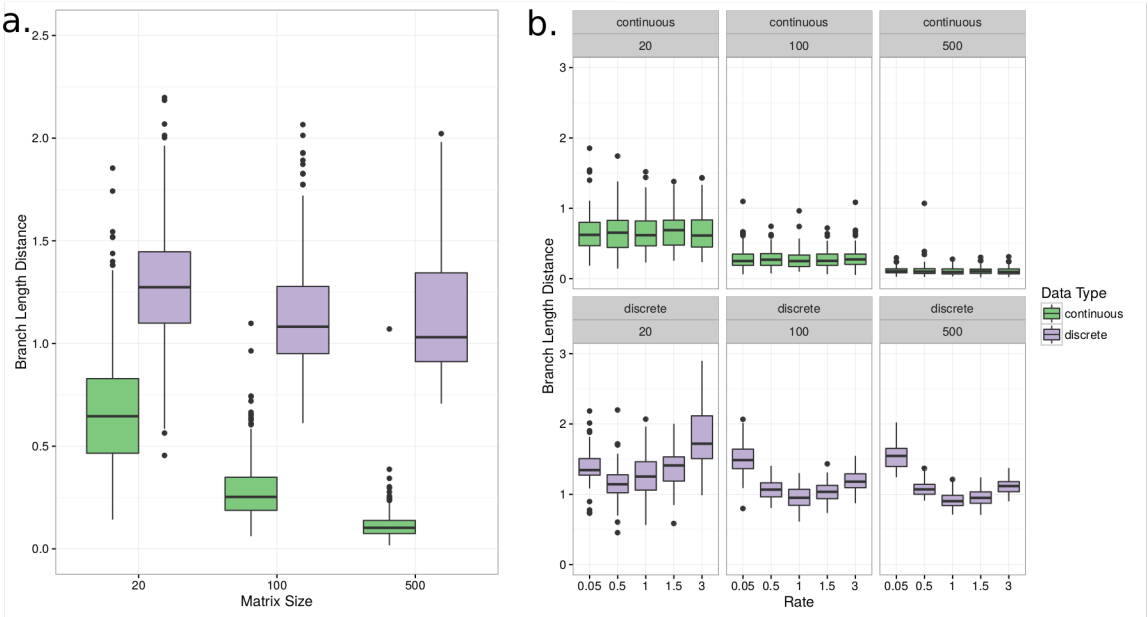

Figure 2.7: Branch length distance (BLD) across trees estimated from independently evolving continuous characters. A) BLD averaged across all rates except for the highest rate category, which resulted in the highest error when inferring under Mk. B) BLD across all matrix sizes and rates.

CHAPTER III

Bayesian Placement of Fossils on Phylogenies Using Quantitative Morphometric Data

Preamble: The contents of this chapter have been published in *Evolution*. The published version appears as: Parins–Fukuchi, Caroline. Bayesian placement of fossils on phylogenies using quantitative morphometric data. *Evolution* 72.9 (2018): 1801-1814.

3.1 Abstract

Jointly developing a comprehensive tree of life from living and fossil taxa has long been a fundamental goal in evolutionary biology. One major challenge has stemmed from difficulties in merging evidence from extant and extinct organisms. While these efforts have resulted in varying stages of synthesis, they have been hindered by their dependence on qualitative descriptions of morphology. Though rarely applied to phylogenetic inference, traditional and geometric morphometric data can improve these issues by generating more rigorous ways to quantify variation in morphological structures. They may also facilitate the rapid and objective aggregation of large morphological datasets. I describe a new Bayesian method that leverages quantitative trait data to reconstruct the positions of fossil taxa on fixed reference trees composed of extant taxa. Unlike most formulations of phylogenetic Brownian motion models, this method expresses branch lengths in units of morphological disparity, suggesting a new framework through which to construct Bayesian

node calibration priors for molecular dating and explore comparative patterns in morphological disparity. I am hopeful that the approach described here will help to facilitate a deeper integration of neo- and paleontological data to move morphological phylogenetics further into the genomic era.

3.2 Introduction

The role of fossil data in reconstructing phylogeny among living organisms has long been a central, yet contentious, topic in evolutionary biology. This has manifested over the past decade in the rapid proliferation of 'total-evidence' methods that seek to simultaneously reconstruct the relationships and divergence times between living and fossil taxa using cladistic morphological matrices. These approaches, based upon probabilistic models of molecular and morphological character evolution, have increased understanding of evolutionary tempo across large clades, and provide compelling evidence in favor of incorporating fossils in phylogenetic analyses (Pyron 2011; Ronquist *et al.* 2012). This can benefit both paleo- and neontological studies by improving the accuracy and treatment of uncertainty in estimation of divergence times and comparative dynamics (Slater *et al.* 2012; Guindon 2018).

A constant source of difficulty when jointly estimating phylogeny between living and extinct organisms is the unavailability of molecular data in nearly all fossil taxa. As a result, there has been a need to explore the compatibility of molecular with morphological data to better understand the capability of fossil and extant species to reciprocally inform reconstruction of phylogeny and divergence times. Previous work has sought to determine whether the inclusion of molecular data representing extant species can improve the reconstruction of relationships among fossils represented by morphology alone (Wiens 2009; Wiens *et al.* 2010). The results of these studies suggest that the inclusion of morphological characters comprising living and fossil species does not have a tendency to decrease the accuracy of phylogenetic reconstructions, and can improve estimation of fossil placements in well-behaved datasets. Expanding upon these observations, Berger and Stamatakis (2010) have shown that methods placing fossils on fixed molecular phy-

logenies can yield accurate results. Their study also shows that a scaffolding approach can further improve fossil reconstructions by offering a straightforward means of filtering through noise in morphological datasets by leveraging information from the molecular reference topology.

Morphological data present other unique challenges important to phylogenetic analysis. For example, morphological data are frequently susceptible to displaying biased or misleading signal. Although discordance in morphological datasets may sometimes reflect biological processes such as convergent evolution, there is also frequently substantial noise stemming from systematic error and poor preservation of fossil taxa. Systematic sources of discordance often stem from the general practice of assigning discrete character states to taxa through qualitative assessment. The subjective nature of this process can cause major irreconcilable disagreement between results achieved from different researchers (Hauser and Presch 1991; Pleijel 1995; Wilkinson 1995; Hawkins *et al.* 1997; Scotland and Pennington 2000; Scotland *et al.* 2003; Brazeau 2011; Simões *et al.* 2017). As an added source of potential bias, these matrices are also frequently filtered to exclude characters that researchers suspect to be homoplasious. However, since these judgments are typically made subjectively, it may be of benefit to introduce a quantitative framework to evaluate the reliability of morphological traits.

As another challenge, the discrete character matrices most commonly employed in phylogenentics can often be difficult to adequately model. At present, researchers employing probabilistic methods generally use the so-called 'Mk' model (Lewis 2001). This is a generalization of the Jukes-Cantor model of nucleotide substitution that accommodates k possible character states. Although previous work based upon simulated data has suggested that Mk-based approaches outperform parsimony (Wright and Hillis 2014), the extent and conditions under which this is the case in empirical datasets are unclear (Goloboff

et al. 2017). Empirical datasets are also likely to depart significantly from the assumptions of the Mk model. This poor match between model assumptions and data can lead to erratic results and high uncertainty in posterior estimates of divergence times (Ronquist *et al.* 2016). Although recent studies have proposed more sophisticated models (Wright *et al.* 2016), the standard symmetric Mk model remains in frequent use, and the sensitivity of topological reconstruction to this frequent mismatch is fairly unclear at present.

For all of these reasons, continuous traits have been suggested as a potential alternative (Felsenstein 1973a, 1988; MacLeod 2002). Nevertheless, their use has remained relatively unexplored. In a previous study (Parins-Fukuchi 2018b), I explored through simulations the relative performance of continuous and discrete traits in phylogenetic inference. I found that continuous characters perform similarly to discrete characters when phylogenetic half-life is set to be equal, while exploring the possibility that continuous traits may extend phylogenetic informativeness over some discretized character codings.

Traditional linear morphometric measurements have long been employed in morphological phylogenetics, but are typically discretized to more easily analyze them alongside present-absence data. Several approaches have been proposed for the discretization of quantitative morphological data (Thiele 1993; Wiens 2001). However, these can yield inconsistent or misleading results (Rae 1998; Goloboff *et al.* 2006), and may in principle reduce the amount of information in continuous datasets by binning fine-scaled variation into shared discrete categories. As a result, it may often be preferable to analyze continuous traits directly.

Tools that quantify morphological size and shape have the capacity to alleviate many of the concerns relating to bias and subjectivity that occur with discrete characters. Approaches such as geometric morphometrics offer the potential to holistically incorporate all dimensions of shape to inform phylogeny. The continuous state space of morphometric

data might also increase the amount of information that can be extracted from morphological datasets, which may be beneficial when analyzing poorly-sampled fossil data. Continuous traits in general may engender benefits on two levels when available by 1) reducing subjective bias often encountered when constructing discrete character matrices, and 2) potentially preserving hard-won phylogenetic information over discretized character codings by representing the full range of observed interspecific variation. Although I explored point 2 previously (Parins-Fukuchi 2018b), future studies will be needed to quantify the extent to which this is the case in diverse empirical datasets.

As another source of continuous traits, geometric morphometric data have shown utility in several previous phylogenetic studies using parsimony-based methods (González-José *et al.* 2008a; Catalano *et al.* 2010; Smith and Hendricks 2013), but have not gained substantial traction. This may be in part due to the lack of available tools to analyze continuous trait data in a probabilistic framework. In addition, previous authors have raised concerns about the use of morphometric data in phylogenetic analysis, based primarily upon potential error stemming from covariance across characters and difficulties in parsing out homologous interspecific variation from variation resulting from rotations in morphospace (Felsenstein 2002). However, these concerns have been partially alleviated by the success of other workers in reconstructing phylogeny from landmark coordinates that are derived from truly homologous regions that have been properly aligned using Procrustes transposition (MacLeod 2001, 2002; Catalano *et al.* 2010; Goloboff and Catalano 2016).

The earliest studies investigating probabilistic methods of phylogenetic inference were developed using continuous characters modeled under Brownian motion (BM) (Cavalli-Sforza and Edwards 1967; Felsenstein 1973a). Due in part to the abundant discrete character data that became available with the emergence of DNA sequencing, these approaches were quickly overshadowed in popularity by discrete trait approaches based upon Markov

nucleotide substitution models. Continuous trait models have since gained significant popularity in phylogenetic comparative methods, but still are rarely used for phylogenetic inference. As a result, few implementations exist, with only ContML in the PHYLIP package and RevBayes providing such functionality (Höhna *et al.* 2016). However, the PHYLIP implementation uses a very simple tree searching procedure. RevBayes is very flexible, however, it is perhaps best suited to total-evidence analyses, where extant and fossil taxa are estimated simultaneously. An alternative procedure involves fixing extant relationships using the results of a molecular analysis, and estimating the positions of fossil taxa along this scaffolding. Previously, Revell *et al.* (2015) described a method that places individual taxa on phylogenies using quantitative data.The authors found that the approach performed well, but the implementation developed for the study was restricted to the placement of only extant and recently extinct taxa. In addition, the authors explored only the placement of a single taxon at a time.

Although, like the Mk model, BM is fairly simplistic, it may offer a degree of flexibility that improves its fit to empirical data in comparison to Mk. For instance, the Mk model assumes that stationary frequencies of character states are equal, whereas BM assumes that traits at the tip of a phylogeny are distributed according to a multivariate Gaussian distribution, with a set of covariances defined by the topology and branch lengths. While the Mk equilibrium assumption is violated in most empirical datasets, the BM assumption of normality can often be justified by the central limit theorem. This suggests that, even in cases where character state changes may better conform to a non-Gaussian distribution over short timescales, these collapse into a Gaussian-like distribution over longer timespans with many repeated draws. The standard phylogenetic BM model may still be violated by patterns such as directional change, but the effect is not well understood. Quantitative trait evolution might also proceed according to stasis and sudden jumps (Lan-

dis *et al.* 2013), but the identifiablility between BM and more complicated models across a tree when branch lengths are expressed in unit variance are not clear.

In this paper, I describe a new approach that places multiple fossils on molecular trees using quantitative characters modeled under BM. Departing from Revell *et al.* (2015), the phylogenetic BM model used here treats branch lengths in terms of morphological divergence rather than time. This simplifies the estimation procedure, and allows morphological disparity across taxa to be easily visualized across the resulting tree, similarly to molecular phylograms. The approach here seeks to tackle some of the most pressing obstacles associated with the use of traditional and geometric morphometric data in phylogenetic inference. Using simulated data, I validate and explore the behavior of the implementation. I also analyze empirical datasets representing the Vitaceae family of flowering plants (Chen 2009) and carnivoran mammals (Jones *et al.* 2015) comprised of traditional and geometric morphometric measurements, respectively. The method uses Markov chain Monte Carlo (MCMC) to infer the evolutionary placements of fossils and branch lengths.

3.3 Methods and Materials

3.3.1 Software

All fossil placement analyses were performed using the new software package *cophymaru* written in the Go language. The source code is publicly available as free software at https://github.com/carolinetomo/cophymaru. This package estimates the positions of fossil taxa on a user-specified reference tree of extant species using continuous traits contained within a PHYLIP-formatted data file where each trait is separated by tabs. Examples can be gleaned from the simulated and empirical data generated from this study, available online.

3.3.2 Brownian motion model

The approaches that I describe in this paper all rely upon the familiar BM model of evolution (Butler and King 2004; O'Meara *et al.* 2006) . Under BM, traits are assumed to be multivariate distributed, with variances between taxa defined by the product of their evolutionary distance measured in absolute time and the instantaneous rate parameter (σ):

$$dX(t) = \sigma dB(t) \tag{3.1}$$

where *dX(t)* is the time derivative of the change in trait *X* and *dB(t)* corresponding to normally distributed random variables with mean 0 and variance *dt*. This leads to the expectation that over time *t*,

$$E(X_t) = X_0 \tag{3.2}$$

with

$$Var(X_t) = \sigma^2 t \tag{3.3}$$

where X_0 gives the trait value at t_0.

The methods that I describe use a slightly different parameterization and likelihood calculation than most conventional implementations used in modern phylogenetic comparative methods (PCMs). These generally construct a variance-covariance (VCV) matrix from a dated, ultrametric phylogeny to calculate the likelihood of the data, assuming a multivariate normal distribution (Butler and King 2004; O'Meara *et al.* 2006). Since these methods treat the topology and branching times as known, the goal is typically to obtain

the maximum likelihood estimate (MLE) of the rate parameter (σ^2) to examine evolutionary rate across clades.

In typical usage, researchers employ phylogenetic BM models where branch lengths are scaled to absolute time, and a rate parameter is estimated. Although it is possible to simultaneously estimate divergence times and topology while analyzing continuous traits, this requires the specification of a tree prior that can accommodate non-ultrametric trees that include fossils. In addition, this approach would effectively perform morphological dating using continuous traits. The behavior and feasibility of such a procedure is not understood, and falls outside the scope of this article. Perhaps more importantly, this would also create circularity when using the method to place fossils used as calibrations in molecular dating. To overcome the need for simultaneously estimating divergence times and fossil placements, the method estimates the product $\sigma^2 t$ together. As a result, rate and absolute time are confounded in the trait and tree models. Branch lengths, which reflect the morphological disparity between taxa, are thus measured in units of morphological standard deviations per site. This interpretation could be thought roughly of as a continuous analogue to the branch lengths obtained from discrete substitution models. Similarly to the discrete case, long branch lengths could reflect either a rapid rate of evolution or a long period of divergence (in absolute time) along that lineage.

3.3.3 Computation of the likelihood

Rather than use the computationally expensive VCV likelihood calculation, I use the reduced maximum likelihood (REML) calculation described by Felsenstein (1973). Full derivations of the likelihood and algorithm are also given by Felsenstein (1981b) and Freckleton (2012), and summarized briefly below. The tree likelihood is computed from the phylogenetic independent contrasts (PICs) using a 'pruning' algorithm. In this procedure, each internal node is visited in a postorder traversal, and the log- likelihood, L_{node} is

calculated as multivariate normal, with a mean equal to the contrast between the character states, x_1 and x_2 at each subtending edge and variance calculated as the sum of each child edge, v_1 and v_2:

$$L_{node} = \frac{1}{2} * \frac{\log(2\pi) + \log(v_1 + v_2) + (x_1 - x_2)^2}{v_1 + v_2} \tag{3.4}$$

The PIC, $x_{internal}$, is calculated at each internal node and used as the character state representing the internal node during the likelihood computation at the parent node. The edge length of the internal node, $v_{internal}$ is also extended by averaging the lengths of the child nodes to allow the variance from the tips to propagate from the tips to the root:

$$x_{internal} = \frac{(x_1 * v_2) + (x_2 * v_1)}{v_1 + v_2} \tag{3.5}$$

$$v_{internal} = v_{internal} + \frac{(v_1 * v_2)}{(v_1 + v_2)} \tag{3.6}$$

The total log-likelihood of the tree, L_{tree} is calculated by summing the log-likelihoods calculated at each of the n internal nodes.

$$L_{tree} = \sum_{node=1}^{n} L_{node} \tag{3.7}$$

3.3.4 Priors

Since the estimation of branch lengths from continuous traits is relatively uncharted territory in phylogenetics, I implemented and tested three different branch length priors derived from the molecular canon: 1) flat (uniform), 2) exponential, and 3) a compound Dirichlet prior after (Rannala *et al.* 2011). The compound Dirichlet prior also offers the

option to set the scale of the expected tree length using the initial rough estimate of branch lengths.

3.3.5 Markov-chain Monte Carlo

This method uses a Metropolis-Hastings (MH) algorithm (Hastings 1970) to simulate the posterior distribution of fossil insertion points and branch lengths. Rearrangements of the topological positions of fossil taxa are performed by randomly pruning and reinserting a fossil taxon to generate a proposal. This is a specific case of the standard subtree pruning and regrafting (SPR) move for unrooted tees (Fig. 3-1). In this procedure, the two edge lengths that link the fossil to the rest of the tree are merged when the fossil tip is pruned, while the edge upon which the tip is inserted is split into two. The move is described in detail, along with a full derivation of the appropriate MH proposal ratio in Yang (2014, p. 287). Branch lengths are updated both individually and by randomly applying a multiplier to subclades of the tree. MH proposal ratios for branch length updates follow the derivations given for the the 'multiplier' or 'proportional scaling' move described by Yang (2014, p. 225).

I re-implemented the approach used in the ContML program to generate an approximate ML starting tree. These initial placements are achieved using stepwise addition. Unlike ContML, this step successively adds fossils to the molecular guide tree, and so only the fossil positions are estimated. Each fossil is individually inserted along all existing branches of the tree, with the insertion point that yields the highest likelihood retained. At each step, MLEs of the branch lengths are computed using the iterative procedure introduced by (Felsenstein 1981a). In this procedure, the tree is rerooted along each node. PICs are calculated to each of the three edges subtending the new root, and are treated as 'traits' at the tips of a three-taxon tree. The MLE of each edge length of the pruned

three-taxon tree (v_i) is computed analytically using the expressions::

$$\hat{v_{1j}} = \frac{\sum_{j=1}^{n} (x_{1j} - x_{2j})(x_{1j} - x_{3j})}{n} \tag{3.8}$$

$$\hat{v_{2j}} = \frac{\sum_{j=1}^{n} (x_{2j} - x_{1j})(x_{2j} - x_{3j})}{n} \tag{3.9}$$

$$\hat{v_{3j}} = \frac{\sum_{j=1}^{n} (x_{3j} - x_{1j})(x_{3j} - x_{2j})}{n} \tag{3.10}$$

This process is iterated by successively rerooting on each node of the tree and calculating the branch lengths until their values and the likelihoods converge. Felsenstein (1981) gives a more detailed explanation of the algorithm, along with a complete derivation of the MLE branch length calculations.

Once an initial placement has been assigned for all of the fossils, the branch lengths are optimized on the complete tree. These starting lengths can be used to inform branch length priors used during MCMC simulation. One problem with interpreting the results of the ML approach on their own is that it has a strong propensity to becoming trapped in local optima. As a result, it should be interpreted cautiously, and not used without further MCMC searching. In the applications here, the topologies achieved from this procedure are used only to construct starting trees, while the branch lengths inform the specification of branch length priors. This procedure allows straightforward construction of non-random starting trees for the MCMC and priors that reflect the the dataset under analysis.

3.3.6 Filtering for concordant sites

One major hurdle involved in the use of morphological data is their frequent tendency to display noisy and discordant signal. This problem might be expected to manifest even more intrusively in morphometric datasets than in discrete datasets, since traits are much less likely to be excluded *a priori* on the basis of perceived unreliability. As a result, there is a need to filter through noisy signal to favor more reliable sites. I developed a procedure adapted from Berger and Stamatakis (2010) for this purpose. This computes a set of weights based upon the concordance of each site with the reference tree. In this procedure, the likelihood (L_{ref}) of each site is calculated on the reference tree (excluding fossil taxa). Next, the likelihood (Ln) of each site is calculated along each n of 100 phylogenies generated randomly by successively grafting nodes in a stepwise manner until a full tree is formed. Branch lengths are then assigned using uniform random draws. If the likelihood of the site is higher along the reference tree than the current random tree, the weight of the site is incremented by one. Thus, site j receives the integer weight:

$$\overrightarrow{W}_j^{int} = \sum_{n=1}^{100} \delta_{nj} \tag{3.11}$$

where $\delta_{nj} = 1$ if:

$$L_{ref} > L_n \tag{3.12}$$

and $\delta_{nj} = 0$ if:

$$L_{ref} < L_n \tag{3.13}$$

This yields a weight vector that is the same length as the character matrix, with each site possessing a weight between 0 and 100. The sites are then weighted using one of three schemes: 1) whole integer values, where the weight equals the value obtained from equation 11, 2) a floating point value between 0 and 1, where the value generated from the random comparison is divided by 100, and 3) a binary value where the weight is equal to 1 if the site displayed a higher likelihood in the reference tree than 95 or more of the random trees, and 0 if less than 95:

(3.14) $$\overrightarrow{W}_j^{binary} = 1$$

if

(3.15) $$\overrightarrow{W}_j^{int} > 95$$

and

(3.16) $$\overrightarrow{W}_j^{binary} = 0$$

if

(3.17) $$\overrightarrow{W}_j^{int} < 95$$

After the weights are computed using the input guide tree, they are stored, and used in all subsequent likelihood computations during MCMC simulations.

In application, I found that integer weighting caused poor MCMC mixing, and so the floating and binary schemes are probably most practical in most cases. The poor mixing achieved by the integer scheme is likely due to the large increase in the scale of the

log-likelihoods. This causes nearly all proposals to be rejected, substantially reducing the efficiency of the algorithm. In effect, the MCMC algorithm becomes a very inefficient hill-climbing ML search, since only proposals that increase the likelihood are accepted. Since it filters out discordant sites completely, the binary scheme enforces a harsher penalty than the floating and integer schemes, and so might be of greatest use in particularly noisy datasets. As an additional note, although these procedures share similar terminology to the site weights calculated during parsimony analysis of multi-state characters, they differ in their purpose. Parsimony site weights are intended to normalize the contribution of characters with differing state spaces to the overall tree length. In contrast, the site weighting approach deployed here is designed to decrease the contribution of sites that disagree with the reference topology to the overall tree likelihood, instead highlighting signal taken to be more reliable. As a result, the guide tree is used to identify sites that are most likely to reliably inform fossil placements.

Although this procedure was originally implemented in an ML context, the application here functions as a prior. By assuming that the molecular guide tree provides an accurate view of extant species relationships, characters that appear to show significant error, homoplasy, or reflect other processes yielding discordant signal, are filtered out or de-emphasized. This procedure has the effect of increasing posterior support in datasets possessing many discordant characters.The Bayesian framework offers a straightforward means to interpret the resulting support values as standard posterior credibility estimates. Nevertheless, the filtering approach, as any prior, should be applied thoughtfully, and compared to results when the prior is not used.

3.3.7 Simulations

To explore the behavior of these approaches under different settings and validate the implementation, I performed a set of simulations. From a single simulated tree, I pruned

five “fossil” taxa and estimated their positions along the tree using 100 datasets of 50 characters simulated under BM. The tree was simulated under a birth-death model, with a birth parameter of 1.0 and a death parameter of 0.5. The resulting tree conained 41 taxa, leaving a 36-taxon reference tree when the five fossils were pruned. To explore the effect of conflicting and noisy signal, I also generated alignments consisting of 50 “clean” traits simulated along the true tree, and combined with sets “dirty” traits in intervals of 10, 25, and 50 traits generated along random trees. All trait (clean and dirty) simulations were performed using the “fastBM” function in the phytools package (Revell 2012a). All traits were simulated using a rate parameter of 1.0. Random trees were generated by collapsing the true tree into a star topology using the ”di2multi” function, which was randomly resolved using the ”multi2di” function. Branch lengths were then assigned randomly by drawing from an exponential distribution with mean set to 1. The simulated data sets, Newick trees, and all scripts used to generate them are available at https://github.com/carolinetomo/fossil_placement_tests.

I restricted the simulations to a fairly small number of traits because this reflected a similar size as the two empirical datasets. This level of sampling is fairly common among existing continuous datasets, which are often compiled from only one or two anatomical regions (eg., “cranium”, “pelvis”, “leaf”). In the future, methods such as that described here may encourage the assembly of more comprehensive quantitative morphometric datasets, but at present, it seemed most sensible to examine the level of sampling expected from existing datasets. Each simulated trait was evolved independently (ie. displaying no covariance with other sites). This is because 1) I showed in a previous study (Parins-Fukuchi 2018b) that sitewise covariance does not in and of itself significantly handicap reconstructions from continuous traits, and 2) because in this study I was primarily interested in examining the effect of inducing random noise without the potentially confounding effect of

covariance. Although covariance has been expressed as a major concern in morphometric phylogenetics (Felsenstein 1988, 2002), there is no reason to expect greater covariance between continuous traits than discrete traits, which, ideally, should describe similar aspects of morphology. Nevertheless, a fairly common source of error in molecular phylogenetic studies can occur when many sites exhibit shared misleading signal due to some legitimate biological process. A similar effect may in principle occur in studies using continuous morphological characters. And so, although continuous trait matrices may not necessarily carry greater inherent risk toward being mislead by covariance across sites than studies based on molecular and discrete morphological characters, careful analysis is important to properly dissect the distribution of signal across character matrices to properly identify biological and systematic sources of conflict and error.

These simulated datasets were then used to reconstruct the placements of the five fossils. To explore the relative performance of weighting schemes, I performed reconstructions using both the binary and floating approaches. These were supplemented by analyses of the noisy datasets without applying site weights. MCMC simulations were run for 1,000,000 generations and checked to ensure that the effective sample sizes (ESS) exceeded 200. The exponential branch length prior was employed for the simulated data with a mean of 1.0. To evaluate the accuracy of the placement method, I then calculated the distances between the true and reconstructed fossil placements. This was calculated by counting the number of nodes separating the true insertion branch from the reconstructed insertion branch. These distances were divided by the largest possible distance between two tips in the simulated tree to yield a measure of placement error falling between 0 and 1. Placement accuracy was evaluated using the *maximum a posteriori* (MAP) summaries of tree distributions. MAP trees represent the single most sampled tree during the MCMC run. Tree summary and placement distances were calculated using custom Python scripts.

3.3.8 Empirical analyses

To assess the utility of the new approach in analyzing continuous morphological data, I performed analyses on empirical datasets comprised of 1) linear measurements and proportions, and 2) geometric morphometric data composed of 3-dimensional landmark coordinates. These are two common sources of continuous trait data, and so were chosen to test the method across different possible data types. In the *cophymaru* implementation of the method, these characters are input as character matrices similar to those used to store discrete traits, with homologous measurements arranged in columns, corresponding to rows of taxa. In the case of the geometric morphometric data, each landmark coordinate represents a column, similarly to previous phylogenetic approaches that explicitly use geometric morphometric data (Catalano *et al.* 2010). Empirical character matrices, trace files, and reference trees are all available online at https://github.com/carolinetomo/fossil_placement_tests.

I estimated the phylogenetic positions of fossils using a morphological matrix comprised of 51 continuous measurements gathered from pollen and seed specimens sampled across 147 extant and 8 fossil Vitaceae taxa. These data were acquired from Chen (2009). I constructed a guide tree for the extant taxa from 8 nuclear and chloroplast genes gathered from Genbank using the PHLAWD system (Soltis *et al.* 2011). The sequence alignment used to construct the guide tree is available in the online data supplement. Using this scaffolding, I analyzed the morphological data to estimate the positions of the fossil taxa. Individual runs were performed under all three branch length priors to assess stability across models. All analyses were run for 30,000,000 generations and visually checked for convergence. Analyses were performed with binary weights applied to the sites and compared to an unweighted analysis. To ensure that MCMC runs were not trapped in local optima, several redundant runs were performed under each combination of settings. For each, the analysis with the highest mean likelihood was retained.

To explicitly test the informativeness of geometric morphometric data in fossil placement, I also performed analyses on a dataset of 33 3D landmark coordinates representing 46 extant and 5 extinct fossil carnivoran crania (Jones *et al.* 2015). A reference tree composed of the 46 extant taxa was obtained from the data supplement of the original study. These coordinates were subjected to Procrustes transposition using MorphoJ (Klingenberg 2011). This yielded a matrix where each character represented the aligned X, Y, or Z position of one landmark. These characters are 'aligned' such that each column contains the coordinates in one dimension of a single landmark occupied by each taxon. Although the details surround the analytical approaches differ, this use of morphometric data is similar to that used in the method described by Catalano *et al.* (2010). The resulting traits displayed phylogenetic signal, but the transposed coordinates showed very low dispersion (variance) on an absolute scale. Low variance can result in narrower peaks in the MCMC surface, which causes difficulties in achieving MCMC convergence. To remedy this, I scaled all of the traits to increase the absolute variance evenly across taxa evenly at each site while maintaining the original pattern of relative variances across taxa using the scale() function in R (R Core Team 2016). This procedure preserved the signal present in the original dataset, since the relative distances between taxa remained the same. Final analyses were performed on this transformed set of measurements. As with the Vitaceae dataset, I analyzed the canid data under all three branch length priors, and performed several runs, retaining the one with the highest mean likelihood. MCMC simulations were run for 20,000,000 generations, and visually examined using Tracer v1.6 to assess convergence. Both empirical datasets achieved large ESS values (over 1000) under all settings.

For both datasets, I used starting trees and branch lengths generated from the rough ML method described above. Sites were weighted using the binary for the final analyses. Intermediate analyses using unweighted and float-weighted sites were also performed, and are

presented in the data supplement. Dirichlet priors were assigned alpha parameters of 1.0 and beta parameters specified as the total tree length of the ML starting tree. Exponential branch length priors were assigned mean values of 1.0.

Since the empirical datasets were more complex than the simulated data, I summarized the tree distributions as maximum clade credibility (MCC) summaries. These summaries maximize the support of each clade. These were compared to the MAP estimates, however, and yielded generally concordant placements (supplementary material). MCC summaries were obtained using the SumTrees script that is bundled with the DendroPy package (Sukumaran and Holder 2010). Branch lengths were summarized as the mean across all sampled trees.

3.4 Results and Discussion

3.4.1 Simulations

Reconstructions of fossil placements from the simulated datasets showed that the method is generally accurate in placing fossil taxa (Table 3-1). In the absence of noisy traits, reconstruction is nearly always correct, displaying 0.9% error on average. In the presence of random noise, the reconstructions are fairly accurate under the binary scheme, except when noise becomes severe. And although the procedure reconstructs fossil positions that are quite distant in the worst case (50% error under the exponential prior with no weighting scheme), application of the weighting procedures reduces placement error by over half, even though the signal-to-noise ratio is quite high.

In the *cophymaru* implementation, the compound Dirichlet prior outperforms the exponential branch length prior on the simulated datasets. Placement error lower under the compound Dirichlet in all but one of the comparisons. The improvement exhibited under the compound Dirichlet is greatest when using the binary weighting scheme, resulting in

a 6% reduction in error compared to exponential prior on the noisiest dataset. The improvement also increases with the noisiness of the simulated dataset, with the 50 clean+50 dirty dataset displaying the largest increase in placement accuracy when using the binary weighting scheme. This result suggests that the compound Dirichlet branch prior combined with binary weighting scheme may be the ideal mode through which to analyze particularly noisy datasets.

Across both branch length priors, binary weighting shows improved accuracy over float and unweighted analyses. However, despite the apparent advantage of binary weighting, it is possible that the float weighting scheme could remain beneficial in cases where the distribution of noise varies between different regions of trees. This is because the float weighting scheme limits the contribution of noisy sites to the likelihood rather than entirely excluding them. This possibility was not examined in this set of simulations, since the dirty traits were generated to reflect completely random noise. However, in reality, noise may be structured to display discordance in only certain taxa. In these cases, continuous traits may display misleading signal among some subset of taxa, but correctly informative signal among other subsets. Further work will be needed to determine the extent to which weights calculated under the float weighting scheme vary when conflict is localized to particular regions of the reference tree.

Overall, the simulations demonstrate the efficacy of the method for the phylogenetic placement of fossils and provide a validation of the computational implementation. The analysis of clean datasets shows that the method performs well, estimating fossil placements with very low error when signal is clear. The adaptation of Berger and Stamatakis' (2010) site weight calibration approach also appears to effectively filter through noisy datasets to improve estimation. The binary weight calibrations appear particularly effective at dealing with rampant misleading random noise, improving accuracy by 2 to 20

times depending on the relative proportion of signal and noise compared to unweighted analyses. These results show promise toward the prospect of applying the method developed in this work to the analysis of large-scale morphometric datasets, where significant noise might be expected. Although introducing noise decreases reconstruction accuracy, the method performs predictably, and still manages to place fossils on average within the correct neighborhood. However, when weighting schemes are applied, the performance improves drastically, highlighting the promise of this method for the analysis of empirical datasets.

dataset	prior	binary_weights	float_weights	unweighted
50 clean	Exp	0.009	0.012	0.009
50 clean	Dir	0.009	0.009	0.012
50 clean + 10 dirty	Exp	0.024	0.195	0.396
50 clean + 10 dirty	Dir	0.021	0.189	0.390
50 clean + 25 dirty	Exp	0.153	0.420	0.489
50 clean + 25 dirty	Dir	0.120	0.402	0.501
50 clean + 50 dirty	Exp	0.291	0.485	0.528

dataset	prior	binary_weights	float_weights unweighted
50 clean + 50 dirty	Dir	0.234	0.468
0.522			

Table 3-1. Mean error when placing simulated fossils under the exponential and Dirichlet branch length priors. Error is measured as the average number of nodes separating reconstructed placements from their true positions across all 100 replicates of each dataset divided by the maximum possible path length between nodes.

3.4.2 Vitaceae

Application of the fossil placement method to the Vitaceae dataset showed generally positive results (Fig. 3-2). The weight calibration procedure revealed substantial noise in the dataset, with 10-12 of 51 sites failing to favor the molecular reference tree over the random trees at least 95% of the time across all runs. Despite this noise, the binary weighting scheme appeared to adequately filter through this noise to generate biologically reasonable results. *Vitis tiffneyi*, *Parthenocissus clarnensis*, and *Ampelopsis rooseae* all share clades with the extant members of their respective genera. *Palaeovitis paradoxa*, and *Cissocarpus jackesiae*, which represent genera with no extant species, both group with separate, non-monophyletic groups of crown *Cissus*. *Ampelocissus wildei* placed within crown *Cissus*, separated by only a node from *Palaeovitis paradoxa*. All six of these taxa are stable in their placements, grouping within the same clades across runs, and when both the exponential and empirical compound Dirichlet priors are applied.

The remaining two fossils are unstable in their placements across branch length priors. *Ampelocissus parvisemina* alternately occupies clades shared by crown *Vitis* or *Nekemias* in the exponential and Dirichlet prior runs, respectively. This taxon shows poor support under the exponential prior, and achieves higher posterior support under the compound Dirichlet prior. Under the exponential prior, the *Ampelocissus parvisemina* placement shows a 0.2 posterior probability, and increases to 0.62 under the Dirichlet prior (Fig. 3-2). Similarly, *Vitis magnisperma* alternately resolves into clades shared by crown *Cissus* and *Ampelocissus* under the exponential and Dirichlet priors, with posterior support values of 0.23 and 0.54, respectively.

The simulations show that the compound Dirichlet prior achieves higher accuracy than the exponential prior, especially when combined with the binary scheme and applied to noisy datasets. If this observation can be extended to the empirical results, it is reasonable to prefer the placements inferred for these two taxa under the compound Dirichlet prior. This interpretation is supported by the greater stability and higher posterior support observed under the compound Dirichlet branch length prior.

3.4.3 Carnivorans

Analysis of the carnivoran dataset also yielded generally reasonable results (Fig. 3-3). The placements of *Piscophoca pacifica*, *Acrophoca longirostris*, *Enaliarctos emlongii*, and *Allodesmus* agree with previous results (Amson and de Muizon 2014; Jones *et al.* 2015). The placement of *Piscophoca pacifica* and *Acrophoca longirostris* differs slightly from the topology generated by Jones et al., placing the two taxa in a more nested position. However, this placement is consistent with the results of Amson and Muison. *Enaliarctos emlongii* and *Allodesmus* resolve in positions identical to the topology used by Jones and colleagues (2015). *Pontolis magnus* is more erratic in its placement, alternating between placement at the center of the unrooted topology, or grouping with *Vulpes* and

Otocyon. The latter placement is unlikely to be correct, because it places *Pontolis magnus* within the Canidae family, while is canonically known as the only extant member of family Odobenidae. Nevertheless, like the problem taxa in the Vitaceae example above, the placement of *Pontolis* displays reassuringly weak support, both in terms of its posterior density and in its tendency to group at the center of the tree. Interestingly, although the placements of *Enaliarctos emlongii* and *Allodesmus* remain stable across runs, both display weak support.

In both datasets, placement under the exponential branch length prior yields conservative estimates of uncertainty in the fossil placements, displaying generally low posterior support, except when placements are exceptionally stable such as with *Ampelocissus wildei*. This is especially important in 'rogue' taxa such as *Vitis magnisperma*. Branch support under the compound Dirichlet prior is higher across several fossils in the Vitaceae dataset. The positions of the six taxa with stable behavior (listed above) do not change significantly under the compound Dirichlet compared to the exponential prior. Closer examination is needed to better determine the significance of this apparent sensitivity of posterior support measures to prior choice observed in Vitaceae. The carnivoran dataset does not exhibit the same behavior, with both branch support and fossil placements similar across priors.

3.4.4 Continuous vs discrete morphological characters

Previous work investigating the degradation of phylogenetic signal over time has implied that continuous traits can benefit over discrete traits under certain circumstances (Revell *et al.* 2008). In principle, methods that analyze continuous traits directly are preferable over those that bin continuous variation into discrete categories (Goloboff *et al.* 2006), due to their avoidance of error stemming from discretization schemes (Rae 1998), and potential to better preserve information that can be gleaned from morphological datasets (Parins-Fukuchi 2018b). Nevertheless, depending on the type of continuous data that are used, the incorporation of features that can be uniquely described qualitatively, such as the loss and gain of structures, may be helpful. It would be straightforward to combine such discrete information into the morphometric framework described here. As progress in this area develops, it will be important to better understand the behavior of different sources of morphological data at different timescales, and the most appropriate ways to combine, model, and gather such datasets.

Although the performance of this new approach on simulated and empirical data appears generally promising, there are several caveats to consider in its use. When applying this method to geometric morphometric data, authors should be cautious to properly align landmark coordinates using Procrustes transformation to remove the effects of rotation in 3D space as a source of variation. In addition, as is shown by the simulations, when the signal-to-noise ratio becomes high, the weighting procedure performs significantly less accurately than when the amount of noisy/misleading signal is lower. Further work will be needed to assess the source of this discrepancy, and the possibility of additional steps that fortifies the approach when noise becomes high. The weighting procedure also may become more complicated in cases where a reliable scaffolding tree cannot be estimated due to genealogical discordance, or where signal displayed by the quantitative traits is shaped by such discordance (Mendes *et al.* 2018). This could in principle be accommodated in future extensions to the method by relaxing the number of topologies accepted as scaffolding trees, or by extending the model to accommodate such discordance.

Despite the potential utility in phylogenetics, there may be cases where useful phylogenetically-informative characters cannot be extracted from geometric morphometric data. This may be the case when any of the concerns stated by Felsenstein (2002) cannot be overcome, or when geometrically-defined characters exhibit inconsistent or weak signal, such as was found by Smith and Hendricks (2013) when using a semi-landmark geometric method to capture morphological variation in *Conus* snails. In these cases, it may be necessary to resort to using traditional linear measurements and proportions, or qualitative characters. Finally, there are cases where fossils may simply present weak information due to shortcomings in geologic and taxonomic sampling. When this is occurs, it is unlikely that any greater certainty in their placement can be achieved except by adding data.

3.4.5 Comparison to other approaches

The method that I describe here differs substantially from existing approaches to the phylogenetic placement of fossil taxa. Although it is most similar to the fossil placement method developed by Revell *et al.* (2015), it extends their approach in several important ways. For instance, my approach does not require that branch lengths be scaled to time, simplifying the estimation procedure. In addition, the implementation here allows for the estimation of long extinct fossil taxa. Finally, the adaptation of Berger and Stamatakis' approach to filtering character matrices can improve upon the accuracy achieved from existing methods. The method described here also differs from recent 'total-evidence' methods that seek to simultaneously estimate both extinct and extant relationships. Although total-evidence methods are useful tools in the phylogenetic canon, splitting the estimation process into stages may be beneficial in certain datasets, and better suited to certain questions. For instance, the approach here may be used to generate priors for the placement of fossil calibrations in node dating. A new method has been developed that accommodates uncertainty in the phylogenetic placement of node calibrations in Bayesian molecular dating (Guindon 2018), which could, in principle, be combined with my fossil placement approach, by using posterior support of fossil calibrations as the prior probabilities in the dating analysis.

It is also worth noting that the method that I describe here would be straightforward to implement in existing phylogenetics packages, such as RevBayes, and adapted to a total-evidence framework. Although RevBayes does not feature a native implementation of the model that I describe, including the data-filtering approach, adapting the present procedure to this framework may be useful in addressing certain biological questions. This may include an exploration of the feasibility of incorporating continuous data into total-evidence morphological clock analyses (Zhang *et al.* 2016) .

Moving forward, it will be important to explore the behavior of this method when applied to morphometric data collected under a variety of approaches and sampling schemes. The success of the weight calibrations on the simulated and empirical datasets suggests the possibility of applying the method to very large morphometric datasets by providing a means to filter through the noise that may occur when sampling densely across taxa and organs. Such a framework would facilitate the development of a more data-centric approach to morphological phylogenetics that reduces common sources of bias in morphological datasets by filtering data matrices statistically rather than through subjective judgement. This would encourage an exploration of conflict and concordance in signal through quantitative data analysis rather than by attempting to filter subjectively at the stage of data collection. One major gap in the approach presented here concerns the assumption that all continuous traits under analysis evolve under a shared rate. In the empirical analyses performed above, I rescaled the traits at each site so that the variance is set to be equal. However, it will be important to explore model extensions that accommodate rate heterogeneity across characters. This has been done in continuous characters to positive effect by Schraiber *et al.* (2013) using a Gamma site-rate model, and adapting this or alternative approaches to modeling rate heterogeneity (Huelsenbeck and Suchard 2007) will be a key priority in future extensions the method.

3.4.6 Conclusions

The method described here provides a new means for biologists to reliably and confidently place fossils in the tree of life. Although the simulated and empirical analyses show several imperfections and a need for further refinement of these methods, the overall accuracy and conservative assessment of uncertainty displayed in the examples appear encouraging. As molecular phylogenetics advances in its use of genomic data to answer fundamental questions across the tree of life, it will be important for morphological phylo-

genetics and paleontology to keep pace. Analysis of morphometric data using the approach shown here will help to improve issues surrounding subjectivity in character collection, and will help morphological datasets to scale better in the genomic era. New advances in the collection of morphometric data, combined with refinements to the approach developed here will better equip morphology to speak to major outstanding questions across the tree of life.

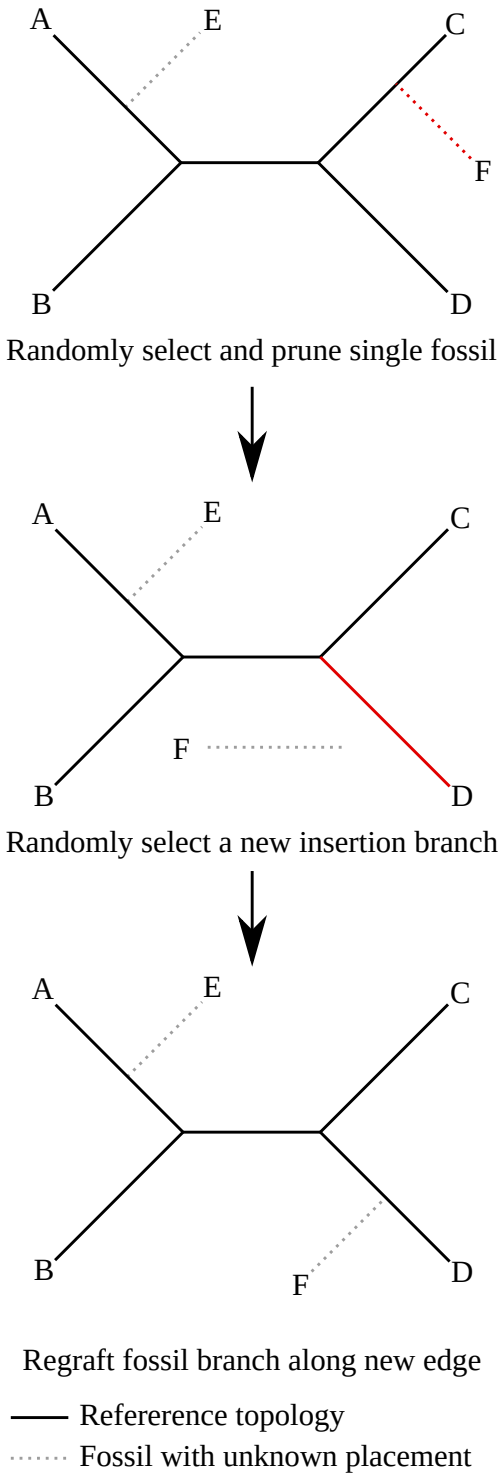

Figure 3.1: Random fossil prune and regraft procedure.

Figure 3.2: Vitaceae fossil placements inferred under the compound Dirichlet branch length prior. Fossil taxa and branches are highlighted in red. Values following fossil tip labels indicate posterior support for placement. Topology is summarized from the posterior using the set of maximally credible clades (MCC). Figure displays only the clade containing all 6 fossils.

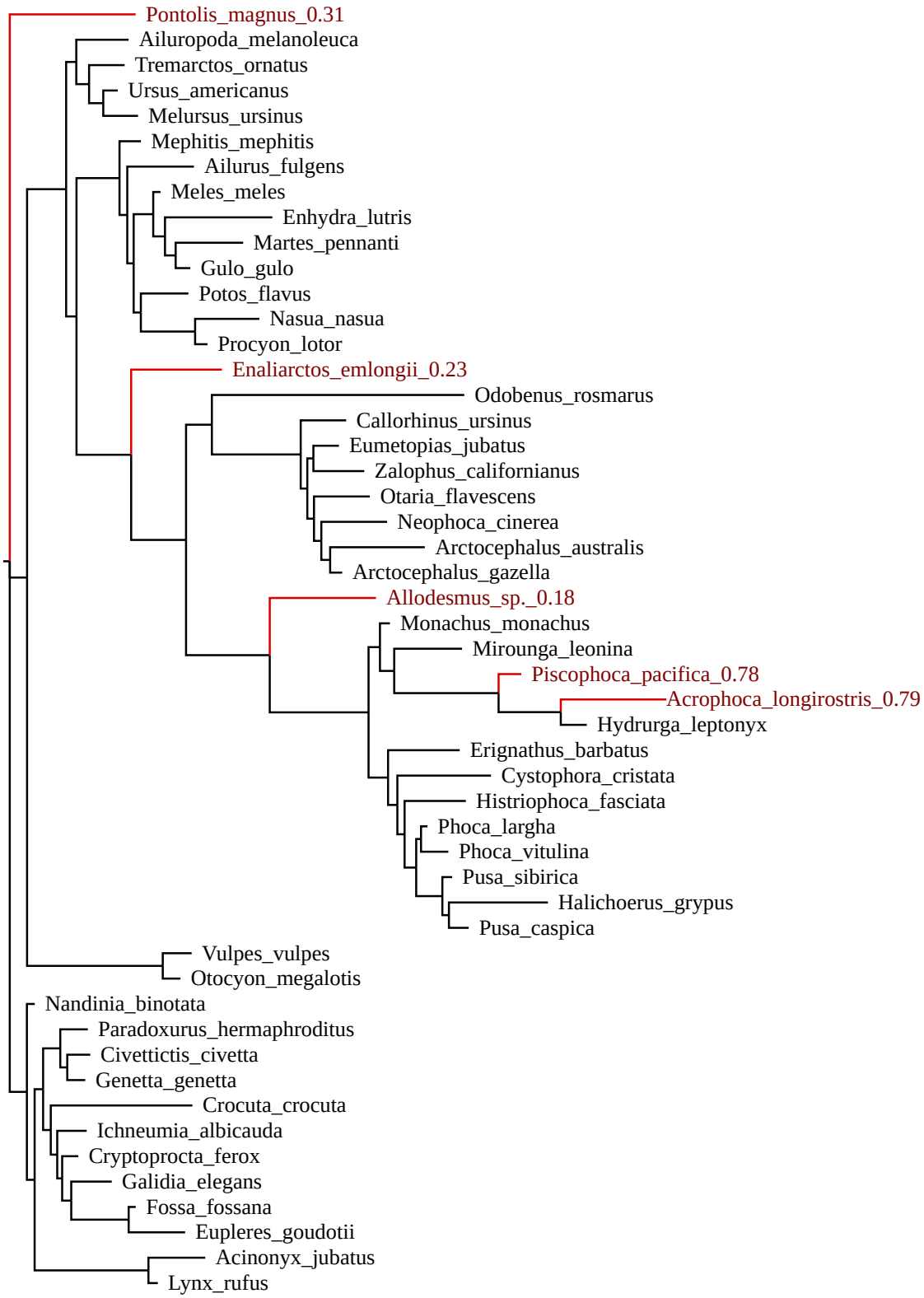

Figure 3.3: Fossil placements inferred from the carnivoran dataset using the compound Dirichlet prior. Placements are displayed as the maximum clade credibility summary of the posterior distribution of trees. Branch lengths represent morphological disparity. Values trailing fossil tip names display posterior support.

CHAPTER IV

Phylogeny, Ancestors, and Anagenesis in the Hominin Fossil Record

Preamble: This chapter has been accepted for publication in an upcoming issue of *Paleobiology*. The manuscript will appear as: Parins–Fukuchi, Caroline, Elliot Greiner, Laura M. MacLatchy, and Daniel C. Fisher. Phylogeny, ancestors, and anagenesis in the hominin fossil record. (2018)

4.1 Abstract

Probabilistic approaches to phylogenetic inference have recently gained traction in paleontological studies. Because they directly model processes of evolutionary change, probabilistic methods facilitate a deeper assessment of variability in evolutionary pattern by weighing evidence for competing models. Although phylogenetic methods used in paleontological studies have generally assumed that evolution proceeds by splitting cladogenesis, extensions to previous models help explore the potential for morphological and temporal data to provide differential support for contrasting modes of evolutionary divergence. Recent methodological developments have integrated ancestral relationships into probabilistic phylogenetic methods. These new approaches rely on parameter-rich models and sophisticated inferential methods, potentially obscuring the respective contributions of data and models. In this study, we describe a simple likelihoodist approach that combines probabilistic models of morphological evolution and fossil preservation to reconstruct both

cladogenetic and anagenetic relationships. By applying this approach to a dataset of fossil hominins, we demonstrate the capability of existing models to unveil evidence for anagenesis presented by morphological and temporal data. This evidence was previously recognized by qualitative assessments, but largely ignored by quantitative phylogenetic analyses. For example, we find support for directly ancestral relationships in multiple lineages: Sahelanthropus is ancestral to later hominins; Australopithecus anamensis is ancestral to Au. afarensis; Au. garhi is ancestral to Homo; H. antecessor is ancestral to H. heidelbergensis, which in turn is ancestral to both H. sapiens and H. neanderthalensis. These results show a benefit of accommodating direct ancestry in phylogenetics. By so doing, quantitative results align more closely with previous qualitative expectations. Keywords: anagenesis, phylogenetics, morphology, paleontology, hominin

4.2 Introduction

Phylogenetic methods that employ parametric evolutionary models, such as the Lewis Mk model (Lewis 2001) have recently begun to emerge as important tools for addressing paleobiological issues at macroevolutionary timescales. Probabilistic approaches appear promising in their performance relative to earlier cladistic methods and offer several important advantages. These include their tendency to increase the clarity with which models are specified and their ability to weigh evidence under competing evolutionary models using modern inferential machinery. While probabilistic approaches have long been dominant in molecular phylogenetics, their application to morphological datasets remains immature. One topic that has only recently been addressed in parametric approaches to phylogenetic inference has been the incorporation of paleobiological evolutionary patterns, such as budding cladogenesis and anagenesis.

Traditional approaches usually depict evolutionary relationships in terms of branching,

or in paleobiological terms, splitting cladogenesis. However, when analyzing the fossil record, it is often desirable to entertain hypotheses where taxa occurring earlier in time are directly ancestral to taxa that occur later in time. The presence of such directly ancestral taxa is expected in the fossil record (Foote and Raup 1996), and is central to identification of paleobiological patterns such as budding cladogenesis and anagenesis. Characterization of evolutionary mode by distinguishing splitting cladogenesis from budding and anagenesis in the fossil record is fundamental to many paleobiological studies (Simpson 1944; Stanley 1979; Levinton 2001). Although the significance of budding, anagenesis, and direct ancestry has been fiercely debated (Gould and Eldredge 1977; Gould 1980a; Levinton and Chris 1980; Gingerich 1985), these patterns have all been observed in the fossil record and can greatly impact inferences of evolutionary processes (MacLeod 1991; Soul and Friedman 2017). This may be particularly important when constructing and testing hypotheses of phylogenetic relationships, where accommodation of direct ancestry, whether as budding cladogenesis or anagenesis, may improve accuracy and yield insights otherwise unattainable (Gingerich 1979a; Fox *et al.* 1999; Aze *et al.* 2011; Strotz and Allen 2013; Aze *et al.* 2013).

These issues are particularly relevant in taxa such as hominins (humans and all other taxa more closely related to humans than to chimpanzees), where hypotheses of direct ancestry are often constructed and entertained qualitatively. Discerning human evolutionary patterns is a focus traceable to Darwin (1871), and ancestry is almost certainly more frequently proposed (even if informally) for clusters of hominin fossils than for those from other taxonomic groups. Nonetheless, there are few formal routes for recognizing this status based on statistical evaluations of morphological traits.

Despite the importance of directly ancestral relationships in the fossil record their role in phylogenetic reconstruction has remained under-explored. Stratocladistic methods were

developed in part to explore use of temporal occurrence data to test hypotheses of direct ancestor-descendant relationship (Fisher 2008). As a non-parametric approach, stratocladistics uses the criterion of maximum parsimony (MP) to minimize both the number of homoplasious character changes and unsampled stratigraphic intervals implied by topologies that include both bifurcating and serially linked segments of phyletic lineages. Several authors have expressed objections both to the use of temporal data in phylogenetic inference and the capability of available data to adequately test ancestral and collateral relationships between species (Smith 2000). Nevertheless, direct ancestors are expected to occur in the fossil record (Smith 2000), and the integration of temporal data has been shown to improve reconstruction accuracy over that of morphological analyses alone (Fox *et al.* 1999). In addition, those earlier criticisms, which have been frequently raised by proponents of cladistic methodologies, are also less relevant when placed in the context of modern probabilistic approaches. While cladistics operates at the level of cladograms, both stratocladistics and recent probabilistic approaches reconstruct phylogenetic trees (Fisher *et al.* 1994; Fisher 2008). This renders earlier criticism of temporal data largely irrelevant given the current landscape of phylogenetic methodology. As such, it stands to reason that probabilistic approaches are remiss when they fail to accommodate the possibility of ancestor-descendant relationships between taxa occurring at different times.

Two early attempts to explicitly extend the intent and logic of stratocladistics into parametric methods using maximum-likelihood (ML) (Huelsenbeck and Rannala 1997; Wagner 1998), showed strong potential, but did not fulfill all of the goals of stratocladistics, such as identifying direct ancestors. A more recent set of methods combine morphological, fossil preservation, and lineage diversification models to infer relationships and lineage divergence times in a Bayesian context (Pyron 2011). This framework has been extended to explicitly accommodate ancestor-descendant relationships by modeling lin-

eage diversification and fossil preservation processes (Stadler 2010; Bapst and Hopkins 2017; Gavryushkina *et al.* 2017), in particular through use of the 'Fossilized Birth-Death' (FBD) process (Heath *et al.* 2014). The FBD process models lineage diversification using speciation, extinction, and fossil preservation parameters and so is very similar to previous 'birth-death-sampling' (BDS) models (Foote 1997, for example), but is usually estimated in a Bayesian context as an informative prior.

Although these approaches have been shown to be useful, several outstanding questions remain regarding their application to the fossil record. These methods have been largely developed for use in epidemiological systems, where they are used to model molecular sequence evolution along single lineages. Although these systems are useful models of patterns in the fossil record, their sampling is often incomplete, and so they may sometimes call for different approaches in practice. Several a posteriori time-scaling (APT) approaches, such as cal3, that accommodate ancestor-descendant relationships also exist (Bapst 2013). These apply divergence times to unscaled cladograms that have been inferred from character data alone using a model similar to the FBD process. Although both Bayesian and APT approaches have been shown effective when applied to fossil taxa, there has been limited discussion of the statistical properties and identifiability of model parameters such as speciation and extinction rates given incomplete fossil sampling. One more straightforward question regards the extent to which morphological data alone can provide evidence for direct ancestor-descendant relationships without modeling abstract lineage diversification parameters such as speciation or extinction rates. This may be especially important in clades with gap-prone fossil records, such as many lineages of terrestrial vertebrates. These taxa may possess less information from which to infer speciation and extinction parameters, and so a characterization of the evidence for hypotheses of direct ancestry provided by morphology alone will be important in the continued development

of parametric approaches for phylogenetic inference in fossil species.

In this study, we describe an approach to phylogenetic inference that combines models of stratigraphic preservation and morphological evolution to reconstruct time-scaled phylogenies and distinguish between anagenesis and cladogenesis using ML and the Akaike Information criterion (AIC). This approach seeks to simplify existing methods, such as the APT and FBD approaches described above, to clearly identify the signal for directly ancestral relationships presented by morphological data alone. Unlike more complex approaches, ours seeks only to identify information in morphological and temporal data that establishes differential support for ancestor-descendant and bifurcating relationships. As a result, our approach explicitly relies on models of morphological evolution and stratigraphic preservation rather than on models of lineage diversification.

We demonstrate the utility of our approach using the hominin fossil record. Hominins are a particularly compelling taxon for a case study. Although cladistic methods have been applied to hominin evolution (Delson *et al.* 1977; Chamberlain and Wood 1987; Strait *et al.* 1997; Irish *et al.* 2013), prior authors' exclusive focus on bifurcating relationships has precluded the study of direct ancestry except in cases where character polarity and specimen sampling have been carefully considered (Kimbel *et al.* 2006). This limitation has been a major detriment to paleoanthropological studies, where direct ancestry has often been considered to be especially important. The hominin fossil record has more recently been examined in a parametric context (Dembo *et al.* 2015, 2016). However, these results still contain ambiguities and do not consider the possibility of direct ancestry. As a result, phylogenetic results have often conflicted with qualitative interpretations of hominin relationships. For instance, this divide occurs in previous treatments of Homo heidelbergensis and Homo antecessor. Previous studies have alternatively suggested that either Homo heidelbergensis or Homo antecessor are directly ancestral to Homo sapiens and Homo ne-

anderthalensis (Mounier *et al.* 2009), while others have disagreed that Homo antecessor is ancestral to Homo heidelbergensis (Stringer 2012). Others have suggested that Homo heidelbergensis is either a chronospecies directly ancestral to H. neanderthalensis, or directly ancestral to both H. sapiens and H. neanderthalensis (Rightmire 1998; Rosas and De Castro 1998; Stringer 2012). This uncertainty is underscored by the suggestion that cladistic methods and data are unreliable in their ability to describe hominin evolutionary patterns and history (Collard and Wood 2000).

For our empirical exploration, we borrowed a morphological supermatrix from the literature (Dembo *et al.* 2015, 2016). Our analysis identified several areas of direct ancestry in hominins, demonstrating cases where temporal data both corroborate and refute results achieved using morphology alone. We recognize that increasingly thorough compilation of morphological data and new fossil discoveries is likely to continue to revise and refine current understanding of hominin evolution. Nevertheless, our approach sheds light on the capabilities of existing models and data to accommodate anagenesis and budding cladogenesis without relying on abstract diversification parameters. Moving forward, application of our method in hominins provides general demonstration of the importance of accommodating directly ancestral relationships in phylogenetic methods to generate deeper understanding of evolutionary patterns in the fossil record.

4.3 Methods and materials

4.3.1 Use of terms

We use the term ‘ancestor’ modelled loosely after (Gingerich 1979b). Ancestors identified through our method represent taxa that display character states that are not sufficiently differentiated from those of a taxon occurring later in time to warrant assignment to a separate lineage. Operationally, we consider anagenesis as any evolutionary change occurring

along these serially linked phyletic lineage segments. Strictly speaking, these instances might alternatively represent budding cladogenesis if new fossil evidence shows greater temporal overlap between ancestral and descendant taxa than is currently known. However, the gap-prone nature of much of the hominin fossil record complicates the ability to develop a precise mechanistic view of speciation patterns. Since candidate direct ancestors typically do not overlap in their temporal ranges with close relatives, it was simpler to consider all direct ancestors as connected to their descendants through 'anagenesis' (again, given the lack of evidence for budding patterns in the stratigraphic record). Nevertheless, it would be straightforward to apply our approach to clades where putative ancestors overlap with putative descendants.

4.3.2 Inference of ML topology

Our approach evaluates the likelihood of candidate topologies using probabilistic models of fossil preservation and morphological evolution (Huelsenbeck and Rannala 1997; Lewis 2001). We perform a semi-automated tree search by calculating the likelihoods of these models on a set of candidate topologies. This approach tests hypotheses of direct ancestry by combining branches with non-overlapping ranges and comparing cladogenetic and anagenetic models using the AIC. All code developed for these analyses is publicly available and implemented in the mandos package (www.github.com/carolinetomo/mandos).

4.3.3 Identification of a fully-bifurcating ML tree

Our approach has yet to be implemented with a fully automated tree searching algorithm, so we combine semi-automated rearrangements with manual perturbations to search for the ML topology. This is done by exploring tree-space surrounding a starting tree estimated from morphological characters alone. In this study, we obtained a ML starting tree using RAxML, version 8.2.11, using the Mk model of morphological evolution (Sta-

matakis 2014). The morphological data were separated into partitions according to the number of possible states (i.e., binary, trinary, etc.) and analyzed under separate models. This partitioning scheme was maintained for all subsequent morphological likelihood calculations, including those used below in the AIC comparisons. We then performed a series of nearest neighbor interchange (NNI) and subtree pruning and regrafting (SPR) operations. These yielded a set of 1700 candidate topologies from which we identified the fully-bifurcating topology best supported by both morphologic and stratigraphic data. This tree provided a starting point drawn from both morphologic and stratigraphic lines of evidence. From here, we explored ancestor-descendant relationships using the model-testing approach described below.

Modeling stratigraphic preservation. Stratigraphic likelihoods were calculated under a homogeneous Poisson process of geologic preservation. When applied to phylogenetics, inference under this model has been shown to accurately recover simulated phylogenetic relationships (Huelsenbeck and Rannala 1997). The likelihood function is derived from a Poisson process as the probability of observing the first occurrence (o_f), last occurrence (o_l) and number (n_o) of occurrences in the stratigraphic record given some origination and extinction time (t_f and t_l, respectively) and preservation rate (λ). These likelihoods are calculated independently for each (i) of b total lineages and multiplied to yield the overall tree likelihood:

$$\hat{L} = \prod_{i=1}^{b} \frac{(o_f^i - o_l^i)^{n_o^i - 2} \lambda^{n_o^i} e^{-\lambda (t_f^i - t_l^i)}}{(n_o^i - 2)!} \tag{4.1}$$

This equation reaches its maximum as branching and extinction times approach the first and last occurrences in the fossil record, and so the likelihood is maximized across the tree when the total amount of unsampled time implied by the topology is minimized. Although

this approach differs from stratocladistic parsimony in its treatment of occurrence data as continuous rather than discrete, this property causes the preservation model to behave as a statistical formalization of the stratigraphic parsimony debt calculations undertaken in parsimony-based stratocladistic analyses. We estimate lineage origination and termination times along with preservation rate using multivariate numerical optimization routines implemented in SciPy (Jones *et al.* 2016). When combined with morphological data and models of character evolution, this approach represents a comprehensive extension of traditional stratocladistics based entirely on probabilistic models.

Identification of anagenesis. Using the bifurcating topology with the highest likelihood under both stratigraphic and morphologic models, we manually identified a set of potential ancestor-descendant relationships. To explore a more comprehensive range of both anagenetic and cladogenetic arrangements, we also extensively perturbed results manually and compared likelihoods. Although a fully-automated tree searching approach will ultimately be desirable in future versions of our method, our approach to tree-searching is similar to those used in previous stratocladistic studies. Starting with the fully bifurcating tree, we identified putative ancestor-descendant arrangements by collapsing each branch with a temporal range beginning earlier than the range represented by its sister lineage. We isolated each putative episode of anagenesis and compared the morphologic and stratigraphic likelihoods of anagenetic and cladogenetic arrangements using AIC scores. This was required because cladogenetic nodes assume one more parameter than anagenetic nodes (i.e., the branch length or node height connecting the new lineage), and so bifurcating trees contain more parameters than anagenetic trees. Comparison of AIC scores enables a comparison of the relative quality of models with different numbers of parameters. AIC score is calculated from the number of model parameters (k), and the likelihood (L):

$$AIC = 2k - 2log(L) \tag{4.2}$$

More complex models almost always have higher likelihoods than simpler models because added parameters allow a better fit to slight deviations in the data. Since they represent a more parameter-rich phylogenetic model, likelihoods from fully bifurcating trees are not directly comparable to ancestor-descendant trees. AIC scores represent the amount of information lost by a model when representing data, with lower scores being indicative of models that preserve a greater amount of information. AIC accommodates differences in parameter count between models by penalizing the addition of new parameters, seeking to optimize the trade-off between improvements in model fit associated with added parameters and the loss of statistical power that results from over-parameterization. In our use, AIC facilitates comparison of phylogenetic models of differing dimensions by penalizing the addition of branches that are better explained through an anagenetic pattern. The parameter count, k, for each phylogeny is calculated by summing the number of estimated branch lengths for each tree with the number of parameters used in the partitioned Mk substitution model.

Under the Poisson preservation model, stratigraphic likelihood predictably improves when unsampled time implied by the phylogeny is reduced, and so the acceptance of ancestor-descendant arrangements also requires the support of morphology. We accommodated direct ancestry using a novel calculation of morphological likelihood where the probability of transitioning from an observed, rather than an uncertain, parental character state to a single or multiple descendant character states is calculated under the Mk model (Fig. 4-1). This calculation differs from that used on multi-furcating nodes. Since the sequences at internal nodes representing unobserved taxa are unknown, the likelihood of

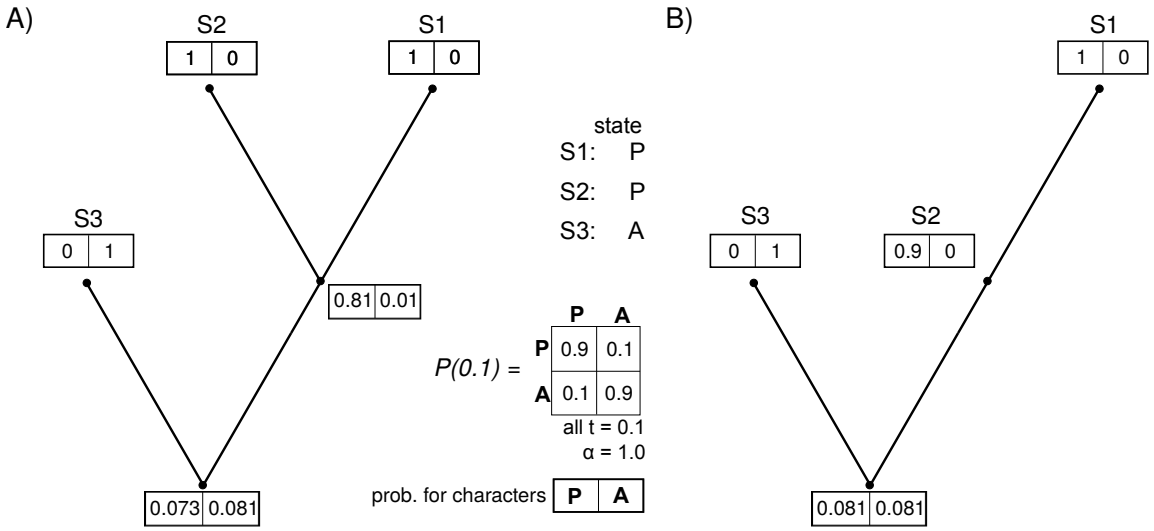

Figure 4.1: Comparison between anagenetic and cladogenetic trees. A) Likelihoods are calculated on bifurcating arrangements using the standard 'pruning algorithm'. B) Direct ancestor likelihood calculation.

character data at the tips is typically calculated by summing over all possible states at each unobserved internal node (Felsenstein 1981a). However, when dealing with direct ancestry, the sets of character states at some internal nodes are known and the marginal likelihoods for states absent in the ancestor are necessarily 0. In these cases, the likelihood is calculated as the conditional probability of observing the set of traits at the tips given the set of traits possessed by the putative ancestor. Ancestor-descendant arrangements are only accepted when there has been a sufficiently small amount of character change. This procedure resembles model testing procedures used to reconstruct ancestral DNA sequences, which compare conditional likelihoods of different permutations of character states at ancestral nodes (Yang *et al.* 1995). A similar effect could also be achieved by fixing the branch length leading to the putative ancestor to zero and performing the standard pruning algorithm. Although the two calculations are equivalent, our approach treats direct ancestry explicitly in the likelihood calculation and tree representation. Like stratocladistics, this use of probabilistic models of character change enables morphology to occupy a central role in identifying direct ancestors. As a result, this approach can, in principle, be ap-

plied without temporal data. A demonstration and test of this method using simulated data is provided in the data supplement (https://github.com/carolinetomo/hominin_anagenesis). These preliminary validations of the implementation show that the approach identifies direct ancestors correctly, even when rates of change are very uneven across taxa. However, the approach can incorrectly prefer arrangements where a taxon with a low rate of morphologic change relative to its sister taxon is collapsed into a directly ancestral position. This result demonstrates the importance of surveying broadly across the anatomy when constructing character matrices and the key role played by temporal data in constraining a set of possible ancestor-descendant relationships.

4.3.4 Stratigraphic and morphologic datasets

We performed our analysis on a supermatrix of 391 discrete craniodental characters compiled by Dembo and colleagues (Dembo *et al.* 2015, 2016). We removed all ambiguous character states, as researchers did not identify whether these were truly ambiguous or polymorphic. While ambiguous character codings are unlikely to provide significant phylogenetic information, existing Markov models of discrete character evolution do not accommodate polymorphism. We excluded the taxa Kenyanthropus platyops and Homo naledi from the present analysis. The features that are diagnostic of K. playtyops have been suggested to result from taphonomic distortion resulting from matrix expansion, rather than from true biological differences (White 2003). Thus, we omitted this taxon in hopes of shedding greater light on the remaining, more widely accepted hominin taxa. Homo naledi was omitted because the data provided in the original study yielded an ML topology placing H. naledi as sister to H. sapiens. Although the phylogenetic affinity of H. naledi is a major outstanding question in paleoanthropology, the confusing signal presented by the H. naledi data, which are relatively recently acquired and therefore represent less well-studied fossils overall, reduced our confidence in the ability of this dataset to resolve its

placement. Therefore, to avoid any confounding effects from the potential unreliability of the H. naledi data, we performed our analyses on the remaining subset of the data after H. naledi was removed. This enabled us to explore the phylogenetic relationships between better-known hominins.

We surveyed the literature to obtain the observed temporal range of each taxon in continuous time. Reported radiometric dates for the oldest and youngest fossils were taken as the first and last observations. Some specimens are ambiguous in their taxonomic assignment; these were excluded from the analysis. We also gathered the number of total occurrences as the number of localities where each taxon has been identified as listed by MacLatchy *et al.* (2010), and supplemented these with additional localities identified in the literature. Cases where multiple specimens belonging to the same taxon have been identified at a single locality were treated as single occurrences. Although we recognize the potential ambiguity in delineating between sites, localities, and occurrences, we attempted to coarsely characterize the total number of occurrences using the number of sites at which each taxon occurs. This approach is more likely to underestimate the number of occurrences than overestimate them, which we expect to yield more conservative statistical support for competing topologies under the preservation model. A comprehensive list of the sites used to define temporal ranges for all taxa is provided in the supplement.

4.4 Results and Discussion

4.4.1 Anagenesis in the hominin fossil record

Our analysis yielded evidence for several instances of anagenesis in the hominin fossil record (Fig. 4-2). Our analysis reconstructed Australopithecus anamensis as directly ancestral to Au. afarensis. This result agrees with broad acceptance of Au. anamensis and Au. afarensis as phyletically linked chronospecies (Leakey *et al.* 1995; Ward *et al.*

2001; Kimbel *et al.* 2006). Although our analysis recovered Ar. ramidus as sister to the anagenetic Au. anamesis-Au. afarensis branch, the morphological data did not support the collapse of Ar. ramidus. Nevertheless, Ar. ramidus possessed poor character sampling in the matrix, and so its placement should be regarded as tentative. Sahelanthropus tchadensis is recovered as a direct ancestor to the rest of the hominin clade. This result should also be treated cautiously due to the small number of characters recovered for analysis of this portion of the phylogeny, but it is in line with with Sahelanthropus' status as the oldest recognized hominin (Brunet *et al.* 2002; Guy *et al.* 2005; Zollikofer *et al.* 2005).

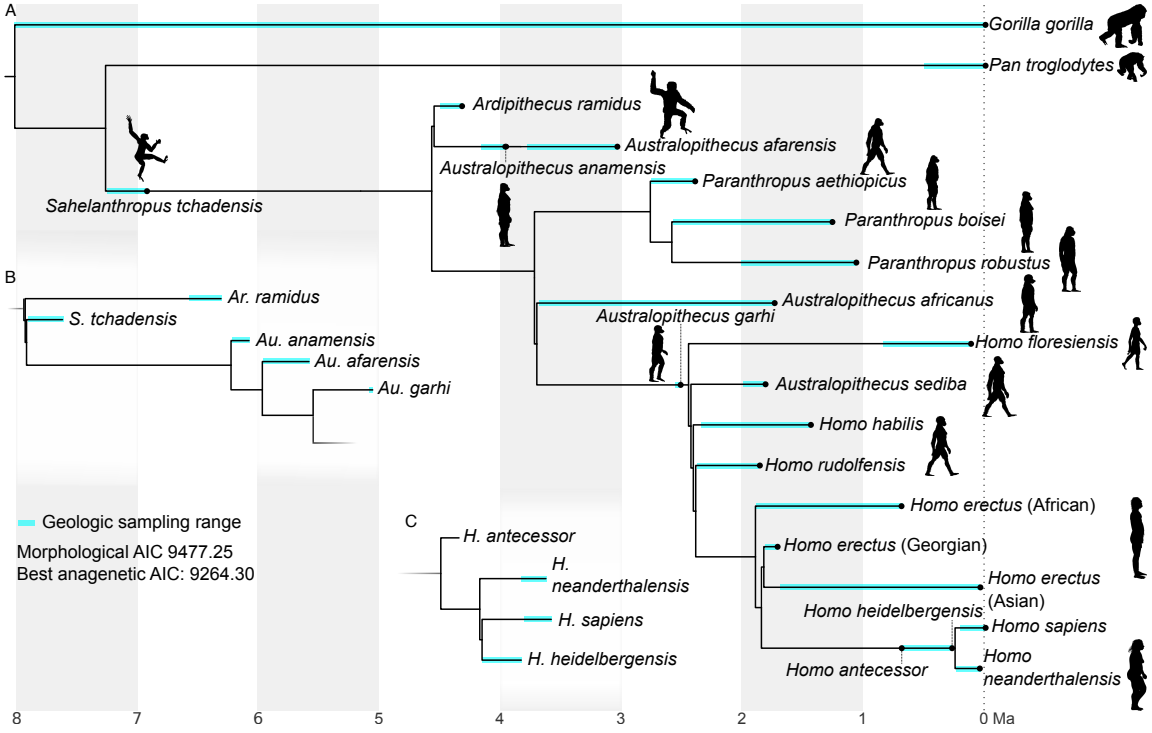

Figure 4.2: Phylogenetic relationships between hominin species. A) Reconstruction with the best AIC score when temporal data are considered alongside morphological data and anagenesis is accommodated. B, C) Areas of the tree inferred from morphology alone that differ from panel A. Both regions display misleading results when anagenesis is not considered. This is reflected in the improvement in AIC observed in the preferred anagenetic topology. Silhouettes obtained from phylopic.org.

Consistent with an early appraisal (Asfaw *et al.* 1999), our final analysis inferred Au. garhi to be directly ancestral to the Homo clade (Fig. 4-2a). This conflicts with cladistic analyses that placed Au. garhi as outgroup to Au. africanus, Paranthropus, and Homo

(Strait and Grine 1999). However, when anagenesis is not considered and phylogeny is inferred from morphology alone, we recover the same placement for Au. garhi as the cladistic result (Fig. 4-2b). Like the example above, this may reflect the constraint that strictly bifurcating methods impose on phylogenetic reconstructions among fossil taxa. However, preference for Au. garhi as ancestral to Homo is weak, with an AIC score (9264.3) that is only slightly better than that of the Au. garhi outgroup hypothesis (9271.11).

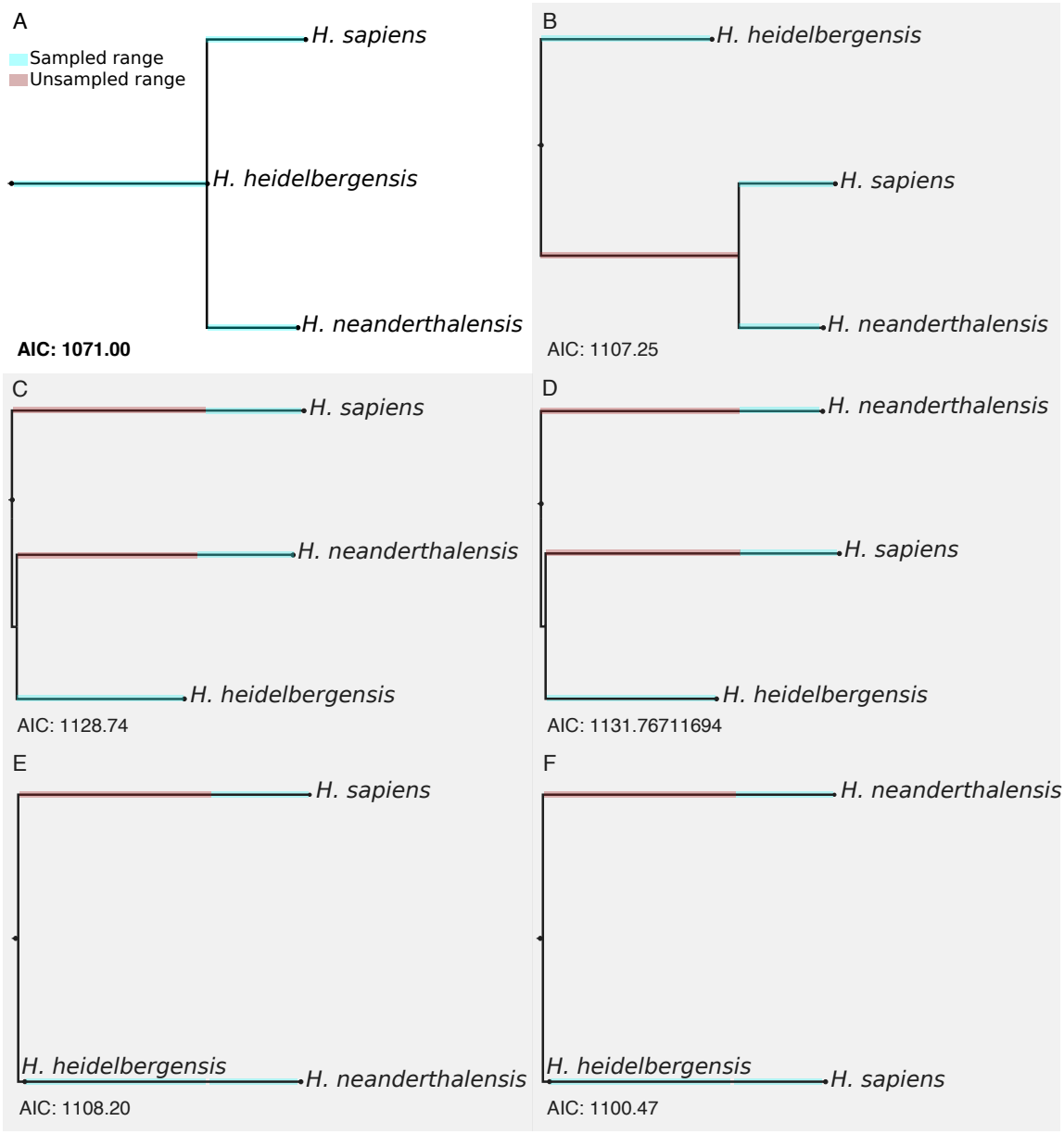

Figure 4.3: AIC scores calculated for each possible arrangement between *H. sapiens*, *H. neanderthalensis*, and *H. heidelbergensis*.

Our results also better reconcile quantitative and qualitative interpretations of the evolutionary relationships (involving potentially some combination of direct ancestry and collateral connections) among the Homo species leading to modern humans and Neanderthals. Paleoanthropologists have variously interpreted Homo heidelbergensis as either: 1) a direct ancestor to both Neanderthals and modern humans (Rightmire 1998; Mounier *et al.* 2009; Stringer 2012; Buck and Stringer 2014), or 2) a chronospecies leading to Neanderthals (Rosas and De Castro 1998). However, quantitative phylogenetic analysis has placed H. heidelbergensis as sister to Neanderthals (Dembo *et al.* 2015). Our ML result based on morphological characters alone places H. heidelbergensis as sister to H. sapiens. However, AIC support improves substantially when H. heidelbergensis is collapsed to represent a direct ancestor that preceded the split between modern humans and Neanderthals (Fig. 4-3). We also uncover one other instance of anagenesis in this clade. Opinions are divided as to whether H. antecessor represents a direct ancestor of later hominin species or is an evolutionary dead end (De Castro *et al.* 1997; Stringer 2012; Dembo *et al.* 2015), but our results provide support for combining H. antecessor and H. heidelbergensis into a single lineage. This suggests a long episode of anagenetic evolution immediately prior to the divergence between modern humans and Neanderthals.

Overall, our reconstruction of relationships among later species of Homo immediately preceding and encompassing modern humans and Neanderthals supports the hypothesis that H. heidelbergensis is directly ancestral to both modern humans and Neanderthals. This result differs from the hypothesis supported by the analysis performed by Dembo and colleagues (Dembo *et al.* 2015), and is instead more consistent with an earlier exploratory statistical analysis (Mounier *et al.* 2009), and with the position frequently suggested by paleoanthropologists (Rightmire 1998; Mounier *et al.* 2009; Stringer 2012; Buck and Stringer 2014). While previous phylogenetic analyses have yielded results that equiv-

ocate or disagree with the common interpretation of H. heidelbergensis as the last common ancestor of modern humans and Neanderthals, our analysis shows that the consideration of direct ancestry can generate statistical support for phylogenetic results that conform more closely to positions generated through researchers' subjective interpretations and exploratory data analyses. This finding supports a general argument against the use of cladistic and phylogenetic methods that are restricted to bifurcating relationships in fossil taxa, where the possibility of variability in evolutionary mode (i.e., occurrence of both anagenesis and cladogenesis) is at odds with an assumption that evolution proceeds by cladogenesis alone (Fig. 4-2).

Our results differ markedly from previous phylogenetic studies seeking to reconstruct hominin phylogeny using probabilistic and cladistic methods. In key regions of the tree, results achieved under our method reveal support for hypotheses more consistent with many qualitative interpretations of hominin relationships, demonstrating the importance of explicitly accommodating anagenesis in the phylogenetic reconstruction of fossil taxa. This may explain some of the historical difficulty in reconciling paleontological interpretations of hominin relationships with cladistic results. For instance, Dembo and colleagues' results are inconsistent with earlier suggestions that Au. anamensis and Au. afarensis are chronospecies differentiated through anagenesis. However, by considering ancestor-descendant relationships and incorporating temporal data, our analysis reveals that a linked Au. anamensis-Au. afarensis lineage is the arrangement most strongly supported by the data. Generally, this suggests that cladograms and strictly bifurcating phylogenies may be inadequate when describing evolutionary relationships between fossil taxa, and so ancestor-descendant relationships should be considered during topological inference. Further, we argue that explicit testing of ancestor-descendant relationships is important even in cases where bifurcating trees are not wholly misleading, as their omission precludes

us from considering the full range of possible evolutionary scenarios. Although previous studies addressing hominin phylogeny using probabilistic methods represent a significant step forward in weighing alternative evolutionary hypotheses, we suggest that their phylogenetic reconstructions have suffered from methodological limitations that were not generally perceived. We suggest that overcoming these limitations can provide a substantial step forward in closing the gap between the paleobiologists' interpretations and previous cladistic and phylogenetic results. In particular, we show through our analyses that the apparent discordance between quantitative and qualitative assessments of evolutionary relationships can be reconciled by extending phylogenetic models to explicitly accommodate anagenesis.

4.4.2 Ancestors, anagenesis, and evolutionary processes

Our method does not seek to distinguish between speciation modes at a mechanistic level. As noted by Fisher (2008), overlap between the extinction and origination times of taxon pairs does not necessarily preclude an ancestor-descendant relationship. Instead, it is possible for ancestral and descendant lineages to coexist. This reality complicates the identification and evolutionary interpretation of ancestor-descendant relationships from temporal data alone. Doing so requires diversification and preservation models that contain additional parameters that quantify completeness of the fossil record. Such models are currently implemented in the cal3 time-scaling method, which seeks to distinguish between splitting, budding, and anagenesis from a cladogram with a fixed topology (Bapst and Hopkins 2017). Our approach is distinct both theoretically and operationally from cal3, instead relying most heavily on morphological evidence to weigh the likelihood of ancestor-descendant relationships without considering the completeness of fossil sampling. At the population level, one might expect some temporal overlap between ancestral and descendant taxa undergoing anagenetic change, especially if the participants

are widely distributed geographically (for additional remarks on the relevance of biogeographic data see Fisher 1994, pp. 158-164, and Fisher 2008, p. 380). As a result, even if the temporal ranges corresponding to taxa identified as ancestor-descendant pairs are discovered to underestimate slightly the time of coexistence, the inference may simply highlight the fuzziness in the taxonomic placement of fossils belonging to lineages undergoing continuous transformation and in discerning between anagenesis and evolutionary budding given incomplete sampling (Fig. 4-4). This interpretation is consistent with previous authors' treatment of temporal ranges when identifying anagenesis between taxa, which has allowed a period of overlap between putative ancestor-descendant pairs (Aze *et al.* 2011; Strotz and Allen 2013; Aze *et al.* 2013).

Under our method, ancestor-descendant relationships might be interpreted either as true anagenesis (i.e., a single population undergoing gradual transformation), or as some form of budding cladogenesis. Previous researchers have argued that true anagenesis is rare compared to budding when analyzing the fossil records of densely-sampled marine invertebrate lineages using more complex preservation models (Bapst and Hopkins 2017). Nevertheless, we suggest that distinction between these two modes may often be impossible in terrestrial vertebrate lineages with large sampling gaps. For example, our results among early hominin species include multiple inferred direct ancestors, but the large gaps in stratigraphic sampling throughout this region of the tree hamper the ability to determine whether these relationships represent true anagenesis or budding that has been obfuscated by poor sampling.

As employed here, our method makes no attempt to distinguish between budding and anagenesis by extending lineage durations using BDS models. Although temporal data occupy an important place in our approach, they are largely used as a guide to constrain the set of possible ancestors and descendants and to provide additional insight when mor-

phological data are equivocal. Coarsely speaking, our approach focuses on patterns in morphological differentiation and lineage disparification, while approaches such as cal3 model lineage diversification. FBD models employed in Bayesian analyses weigh both morphological data and diversification parameter estimates and so combine elements of both approaches. Further empirical and simulation-based work is needed to determine the differences in behavior between methods that test ancestor-descendant relationships in a BDS/FBD framework, and ours. We speculate that their relative accuracy may depend largely on the completeness of sampling in the rock record and the correlation strength between morphological change and lineage diversification, although other factors may also be important.

The scales at which phylogenetic data are sampled may further complicate mechanistic evolutionary interpretations. Morphological character matrices often lack samples across the entire stratigraphic range of each taxon, so in the absence of evidence to the contrary, analysts often assume morphological stasis within lineages. It is therefore often impossible to observe gradual morphological change within and between taxa. These considerations might cause anagenetic relationships identified here to represent either true anagenesis or some form of 'pseudo-anagenesis', where stratigraphic and morphological data appear consistent with anagenesis but the persistence of the ancestor has not been sampled. The ancestor-descendant relationships identified by our method may be interpreted in several ways. As sampled, these results may be roughly conceived as anagenesis in the sense that the mode of evolutionary change between taxa is indistinguishable from evolution occurring along a single lineage, depending upon the completeness of sampling and the degree to which morphological disparity correlates with true biological species diversity. This interpretation is consistent with historical usage by paleobiologists (Gingerich 1979b; Levinton 2001). Thus, our approach seeks to reveal the extent to which existing cladistic

and temporal data can provide evidence for non-branching evolutionary modes and does not seek to resolve conceptual issues that may stem from incomplete sampling, lineage diversification, or population-level evolutionary change. Regardless of the fine-scale evolutionary interpretation, failure to accommodate phyletic change and ancestor-descendant relationships when inferring phylogenetic relationships can generate views of evolutionary history that are positively misleading in the sense that inaccurate results do not improve with added data.

4.4.3 Some practical methodological considerations

Concerns regarding the accuracy of probabilistic approaches have been raised, stemming from the reliance of these methods on the overly simplistic Lewis Mk model of morphological evolution (Goloboff *et al.* 2018b). These critics advocate the use of cladistic methods, arguing that Markov models inadequately capture the complexities of morphological evolution. Although we agree that existing substitution models oversimplify these processes, our results suggest that the accommodation of ancestor-descendant hypotheses in probabilistic methods can improve the fidelity of phylogenetic reconstructions, even when Lewis Mk is used as the underlying model of morphological change. As a result, concerns regarding the adequacy of existing morphological substitution models may be partially alleviated by considering hypotheses of direct ancestry. This is supported by simulation work showing that stratocladisics outperforms cladistics in topology reconstruction (Fox *et al.* 1999). Further exploration is needed to demonstrate more thoroughly the limitations of our new approach, which builds upon stratocladistics by incorporating the benefits of probabilistic analyses, including 1) more explicit statements of the assumptions involved, and 2) the ability to weigh competing models using modern inferential criteria.

The method we describe seeks to enhance understanding of the fossil record by explicitly testing support for existing hypotheses of direct ancestry while attempting to make

simpler assumptions than stratocladistics or recently developed Bayesian methods. Although Bayesian implementations that employ the FBD model also seek to identify direct ancestry (Zhang *et al.* 2016), they often struggle to parse through complex signals to choose between competing ancestor-descendant and collateral hypotheses (Luo *et al.* 2018). Although future extensions to those approaches may improve their response in this context, we view our approach as a foundational, minimally complex framework for exploring the behavior of probabilistic models when evaluating support for direct ancestry in temporal and morphological data. As such, our method should be viewed as a complement to, rather than a simplification of, existing Bayesian approaches. Our method encourages examination of the informativeness of the data without the increased complexity of assessing prior probabilities conditioned upon models of linage diversification. Thus, our method differs from both existing Bayesian and parametric APT approaches by explicitly omitting diversification parameters and instead placing morphological data in a central role when evaluating hypotheses of direct ancestry. In doing so, temporal data help to delineate the set of possible ancestors and play an important role in measuring the fit of candidate trees to the observed stratigraphic record. Our method does not seek to reconstruct diversification processes, and instead focuses on identifying hypotheses that best describe only the information contained within morphological and temporal datasets. Assessment of information contained within datasets and tests of hypotheses can also be achieved using Bayesian approaches (Lewis *et al.* 2016), but likelihoodist approaches such as ours streamline these procedures by reducing complications presented by prior probabilities.

Although Bayesian methods can be beneficial in certain circumstances, our method simplifies identification of anagenetic hypotheses using evolutionary and stratigraphic models. We observe that the likelihood surface surrounding certain nodes may possess low peaks, which likely results from sparse sampling and relatively low information in

the morphological characters. Since Bayesian approaches often average results across this surface, it is possible that they may fail to capture those relationships best supported by the data by including information from weakly supported hypotheses. This is of greater concern in paleontological than neontological data because the increased abundance of molecular data is often likely to result in more clearly defined peaks in the likelihood surface. In cases where information is sparse, likelihoodist approaches such as ours offer the benefit of filtering through noisy and equivocal signal to reveal the hypothesis best supported by the data. Although these benefits are also achievable through Bayesian approaches, additional caution must be taken to select priors that do not dominate weakly informative data. In addition, careful thought should be given when summarizing the posterior/likelihood surface. Averaging across a relatively flat surface might yield poor results (Yang and Zhu 2018), while the comparison of individual point estimates, as is done here, may more clearly shed light on best supported models while clearly contrasting this support relative to competing models. Moving forward, further extensions to our method that more clearly evaluate uncertainty across the likelihood surface will be useful.

4.4.4 How can we proceed?

Anagenesis and splitting cladogenesis were most pertinent to our analysis of hominin evolution, and the straightforward dichotomy between these simplified the assumptions and interpretation of our tests. Our approach may need to explicitly accommodate budding before being applied to groups with very dense fossil records. However, the assumptions required, which may often include morphological stasis within lineages, may make application to some phylogenetic datasets impractical. This is especially true in cases such as hominins, for which the fossil record implies large sampling gaps, and characters representative across stratigraphic ranges are often sampled from only a single individual. Expansions of our approach through implementation of new models will further test the

implications of existing paleontological datasets for reconstructing complex evolutionary and geologic processes over deep timescales. We hope that the example provided here will encourage integration of more diverse evolutionary modes into phylogenetic methods yielding better explanations of temporal patterns in critical parts of life's history.

As we emphasize above, the approach described here should be viewed as an attempt to explore the capability of phylogenetic methods to identify the signature of direct ancestry using existing models to interrogate morphological and stratigraphic data. In doing so, we acknowledge that there are many complicated biological and geological factors that could be incorporated into this framework. For instance, previous researchers have accommodated heterogeneity in fossil preservation rates across time and among lineages (Foote 1997, 2001). There have also been several concerns raised in the literature regarding the adequacy of existing models of discrete trait evolution to inform complex evolutionary scenarios (Goloboff *et al.* 2017; Brown *et al.* 2017). Alternative models that use continuous characters may help to improve some of these issues (Parins-Fukuchi 2018b). Moving forward, elaborations making use of new data sources and models will only continue to improve resolution of evolutionary patterns in the fossil record.

Finally, we acknowledge that our empirical results beg qualification. In particular, we expect that future studies will generate a more comprehensive and authoritative view of hominin evolution as improved data continue to become available. For instance, although we are currently cautious about making strong statements concerning the ancestral position of Sahelanthropus using this dataset, additional information may resolve this issue. This may be the case for several other areas of the hominin tree, which may be better resolved as temporal and taxonomic gaps in sampling are better filled by new discoveries. In addition, we concede the possibility that more comprehensive automated tree searching routines may reveal support for hypotheses that we failed to consider under our semi-automated

approach. Therefore, instead of providing an authoritative view of hominin evolution, our study provides a springboard for future studies by showing that the accommodation of anagenesis can improve our view of the processes and relationships underpinning the evolution of fossil taxa. Nevertheless, due to the improved support and congruence of hypotheses that explicitly consider ancestor-descendant relationships, we recommend that future phylogenetic studies in hominins avoid methods that only consider bifurcating relationships. Future studies that build upon existing work in other taxa will also be important to better characterize the extent to which this suggestion can be generalized across the tree of life. Although the accommodation of directly ancestral relationships is especially relevant in hominin taxa, for which hypotheses of anagenesis have been long entertained through qualitative anatomical assessment, these results may also be important in other taxa. Further empirical work will be needed to develop a better understanding of the extent to which the consideration of direct ancestors can improve resolution of evolutionary patterns throughout the fossil record.

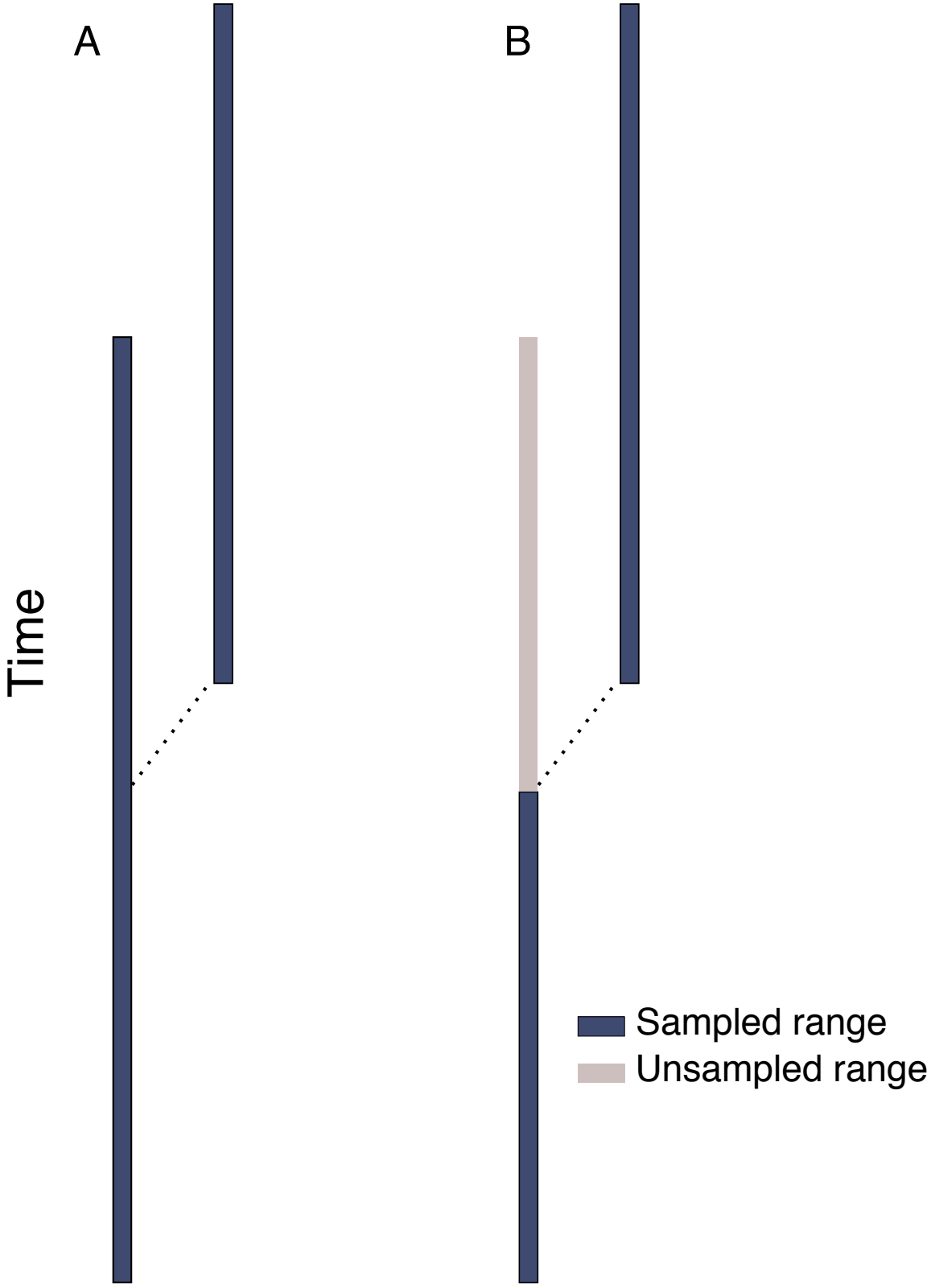

Figure 4.4: A) Speciation mode interpreted as budding when sampling is complete. B) Incomplete stratigraphic sampling may create an inability to distinguish between anagenesis and budding when sampling is sparse.

CHAPTER V

Detecting Mosaic Evolutionary Patterns in Phenotypic Disparity

Preamble: The contents of this chapter are currently in review. It appears in preprint as: Parins-Fukuchi, Caroline. Detecting mosaic patterns in phenotypic disparity. *BioRxiv* DOI: 423228

5.1 Abstract

Understanding the complex patterns underlying phenotypic diversification across the tree of life has long been a fundamental aim in evolutionary biology. The modern evolutionary synthesis was characterized by the integration of disparate biological fields to better understand the diverse processes that drive phenotypic change. The centrality of the 'mosaic' evolutionary patterns that emerge from these diverse processes is a key feature of post-synthesis thought. Despite this fundamental importance, researchers have been limited in their ability to explore such complex patterns in nature analytically. However, recent advances in data collection provide new potential for investigating broad patterns shaping variation across the tree of life. These possibilities emphasize the need for new comparative approaches that will facilitate the examination of mosaic change. In this study, I introduce a novel comparative framework that accommodates the mosaic evolutionary processes that shape whole organisms. The approach recasts the comparative method to harness patterns in phenotypic disparity to identify macroevolutionary suites of

continuous traits. Through demonstrations on simulated and empirical data, I demonstrate the utility of this new framework. This framework offers a first step toward evaluating mosaic variability comprehensively across whole organisms—an integrating goal of the modern synthesis.

5.2 Introduction

Characterizing the ways in which phenotypic disparity and evolutionary rates differ across lineages and throughout time has long been a central goal in evolutionary biology. Shifts in the rate of phenotypic change often coincide with the rapid diversification and ecological innovation of diverse lineages (Rabosky and Adams 2012; Rabosky *et al.* 2013). In addition, studying fluctuations in the tempo of evolutionary change can shed light on our knowledge of important evolution processes, such as adaptation and evolutionary constraint. A more recent body of work has focused on the development and application of phylogenetic comparative methods (PCMs) that identify patterns of phenotypic change using phylogenies (Harvey and Pagel 1991; Hansen 1997a; Butler and King 2004; O'Meara *et al.* 2006; Eastman *et al.* 2011; Beaulieu *et al.* 2012). Modern PCMs enable researchers to infer shifts in evolutionary rate, constraint, and disparity throughout time by applying stochastic models of trait evolution to comparative phenotypic data using phylogenetic trees. These approaches have helped to answer important questions concerning the tempo and mode of phenotypic evolution across deep timescales (Harmon *et al.* 2003; Scales *et al.* 2009; Harmon *et al.* 2010; Rohlfs *et al.* 2013; Slater 2013).

Paleobiological and comparative studies have typically examined univariate evolutionary patterns. This has perhaps grown from the tendency of early foundational work to document evolutionary patterns using single key traits, such as molar shape (Simpson 1944; Gingerich 1974). Although often limiting the ability to ask detailed evolutionary

questions at broad scales, this framework has been effectively leveraged to alternately explore specific case studies (Scales *et al.* 2009), or general questions using single gross morphological traits, such as body mass, as an approximation for overall phenotypic variation (Harmon *et al.* 2003; Beaulieu *et al.* 2007; Harmon *et al.* 2010; Burbrink and Pyron 2010; Rabosky *et al.* 2013; Zanne *et al.* 2014; Bokma *et al.* 2015; Landis and Schraiber 2017).

Despite these past successes, a central goal in post-synthesis evolutionary biology has been to evaluate patterns across anatomical regions and at different phenotypic levels. Such broad investigations require approaches that can accommodate multiple characters. Recent contributions have explored multivariate approaches that enable the analysis of multiple traits simultaneously (Adams 2014). These enable inference of trait models commonly employed in univariate PCMs among high-dimension data. Another set of methods statistically evaluates correlated evolutionary change between multiple traits along a phylogenetic tree (Revell and Harmon 2008; Caetano and Harmon 2018). These methods will be critical extensions to PCMs moving forward, given the current influx of large, publicly available databases of morphology (Boyer *et al.* 2016), and emerging approaches for the rapid, algorithmic quantification of variation from digital specimen images (Boyer *et al.* 2015; Pomidor *et al.* 2016).

One aspect that has been under-explored in a comparative analytical context has been the tendency for distinct subsets of traits to display unique patterns in evolutionary rate and relative disparity. The historical focus on univariate analysis in PCMs and paleobiology may have contributed to a general lack of recognition of the diversity of evolutionary patterns that can combine to organismal body plans. Nevertheless, mosaic evolution has been a fundamental concept in evolutionary biology since the modern synthesis in its acknowledgement of the reality that anatomical regions are often exposed to natural selection at

differing magnitudes and directions at different times (Stebbins 1983). The mosaic concept dates to Dollo (Gould 1970), and has frequently been invoked qualitatively as a key factor driving divergent morphological adaptation in different anatomical regions across lineages (De Beer 1954; Cracraft 1970; Mayr 1970; Gould 1977b; Stanley 1979). Although examination in a quantitative context has been limited overall, several key studies have shown that mosaic patterns explain the emergence of important traits in diverse case studies, such as structural variation in the brain across mammals (Barton and Harvey 2000), phenotypic and genomic diversity across angiosperms (Stebbins 1984), and the defining suite of morphological characters displayed by the hominin lineage (McHenry 1975; Gould 1977a; Holloway and Post 1982).

In addition to being of fundamental biological interest, mosaic patterns have long been argued to present unique challenges when inferring phylogeny from morphological characters (Farris 1971). This concern has recently been reasserted by Goloboff and colleagues (2018), who suggest that heterogeneity in relative disparity, as measured by phylogenetic branch lengths, displayed across separate suites of morphological characters can confound phylogenetic inference from morphological traits using Bayesian and maximum likelihood approaches. These concerns parallel the conclusions of important recent studies that validate the prevalence of mosaic patterns in paleobiological and comparative data by manually testing the variability in evolutionary mode that can occur across large morphological datasets (Hopkins and Lidgard 2012; Felice and Goswami 2018). These studies demonstrate the urgent need for overdue extensions to existing uni- and multivariate PCMs that accommodate mosaic patterns in large phenotypic datasets. Methods that separate traits according to overall rate have long been available for phylogenetic inference (Yang 1996; Schraiber *et al.* 2013). However, these approaches do not adequately address mosaic patterns, which focus more on heterogeneity in relative disparity across lineages rather than

absolute rate. Despite the importance, there has not yet been a computational approach that algorithmically partitions traits according to their patterns in disparity.

In this paper, I present a novel method that identifies suites of continuous traits displaying shared patterns in disparity to reconstruct the mosaic trends that have shaped organismal phenotypes. The mosaic character suites identified by the approach are defined by partitions of traits that are best explained by shared phylogenetic branch lengths, measured in units of average disparity, along a fixed topology. The phylogenetic models underlying the construction of mosaic character suites, by representing the accummulation of disparity across lineages, thus provide information on relative rates of evolution. After introducing the method, I evaluate its performance using simulated data. I also present an analysis of an empirical dataset of developmental ossification times complied by Rose (2003). This dataset has previously been leveraged to explore mosaic heterogeneity in evolutionary pattern (Germain and Laurin 2009; Laurin 2014) and so is well suited as an empirical test of the method that I introduce here.

5.3 Methods and Materials

5.3.1 Code and data availability

The approach described below is implemented in a program called *greedo*. It is available freely on Github at (links are available from the journal office). All analyses on simulated and empirical data were performed using this program. Scripts and data used for the simulated and empirical analyses are also available on Github (links available from journal office).

5.3.2 Partitioning traits into mosaic suites

The method described here combines several unsupervised learning strategies to partition traits into separate mosaic character suites, with each possessing its own set of phy-

logenetic branch lengths expressed in units of disparity. These strategies are applied in sequence (Fig. 5-1), with the goal of identifying the configuration that yields the lowest AIC score.

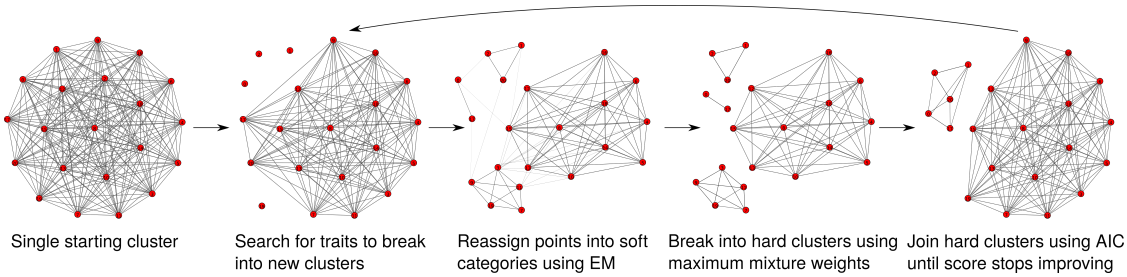

Figure 5.1: Search procedure to identify mosaic suites.

Each mosaic character suite that defines the classification model contributes to the likelihood independently. The log-likelihoods of each of the traits belonging to character suite j are calculated under the associated branch lengths and added to yield an independent log-likelihood. The log-likelihood of the trait matrix, $\mathrm{LL}_{\mathrm{classification}}$, is calculated by summing the log-likelihoods of all k character suites:

$$LL_{classification} = \sum_{j=1}^{k} LL_j \tag{5.1}$$

5.3.3 Tree model

Each mosaic character suite of the classification model is defined by a phylogeny where the topology is fixed, but its branch lengths are free to vary and calculated from the constituent traits. Branch lengths are expressed in units of disparity and are calculated using a Brownian model of evolution. The distribution underlying the traits belonging to each partition are assumed to be multivariate Gaussian, with variances between taxa defined by the product of their evolutionary distance measured in absolute time and the instantaneous rate parameter (σ^2). The phylogenetic comparative methods literature often estimates σ^2

alone by assuming a fixed timescale given by branch lengths that have been scaled to absolute time using a molecular clock model. However, here the absolute times are assumed to be unknown, and the rate and time parameters are confounded with one another. Thus, branch lengths are expressed in units of average morphological disparity, or variance per trait.

This parameterization differs from the typical use of stochastic continuous trait models in comparative studies. Traditional approaches generally start by applying single rate (clock-like) Brownian models to chronograms, comparing the fit to more complex models. For instance, σ^2 is sometimes allowed to vary locally in distinct subclades to yield 'multi-rate' BM models (O'Meara *et al.* 2006; Eastman *et al.* 2011; Thomas and Freckleton 2012). Expanding upon this, more parameter rich models have also been developed, such as Ornstein-Uhlenbeck (OU) (Hansen 1997b) and jump-diffusion (JD), or Lévy, processes (Landis *et al.* 2012). OU models expand upon standard BM by introducing a term, α, that generates a stabilizing force that constrains movement around an optimal trait value, while JD processes contain terms describing the frequency of jumps in character spaces. OU models are often used similarly to multi-rate BM, by allowing optimal trait values to vary across the tree (Butler and King 2004; Beaulieu *et al.* 2012). JD models relax single-rate BM by allowing sudden jumps in mean trait values (Eastman *et al.* 2013; Landis and Schraiber 2017).

The parameterization of the tree and branch lengths used here is continuous equivalent to that encountered in phylograms reconstructed from molecular and discrete morphological data. These express branch lengths in units of substitutions per site by similarly confounding the rate and time parameters. As used here, this approach has several benefits over the more common comparative approaches described above. Although it may be possible in principle to develop a similar approach that assigns traits to suites associated

with separate OU models fit to chronograms with different adaptive regimes, or best explained by distinct BM and JD processes, the tree model used here captures much of the same information (Fig. 5-2), while drastically simplifying inference from both a statistical and computational standpoint. While multi-rate BM, OU, and JD models are often aimed to describe explicit evolutionary processes, such as environmental adaptation or quantum evolution, they may simply improve model fit over simpler models by accommodating heterogeneity in disparity accumulated across the tree. Several studies have demonstrated substantial limitations in the ability to identify the extended OU model parameters using typical comparative datasets (Ho and Ané 2014; Cressler *et al.* 2015; Cooper *et al.* 2016), complicating the use and evolutionary interpretability of these models in empirical studies.

The likelihood is calculated in a recursion from the tips to the root after Felsenstein (1973). Full derivations of the likelihood and algorithm are also given by Felsenstein (1981b) and Freckleton (2012), and summarized briefly here. The tree likelihood is computed from the phylogenetic independent contrasts (PICs) using a 'pruning' algorithm. Each internal node is visited in a postorder traversal, and the log-likelihood, L_{node} is calculated as univariate Gaussian, with a mean equal to the contrast between the character states, x_1 and x_2 at each subtending edge and variance calculated as the sum of each child edge, v_1 and v_2:

$$L_{node} = \frac{1}{2} * \frac{\log(2\pi) + \log(v_1 + v_2) + (x_1 - x_2)^2}{v_1 + v_2} \tag{5.2}$$

The PIC, x_{internal}, is calculated at each internal node and used as the character representing the internal node during the likelihood computation at the parent node. The edge length of the internal node, v_{internal} is also extended by averaging the lengths of the child nodes.

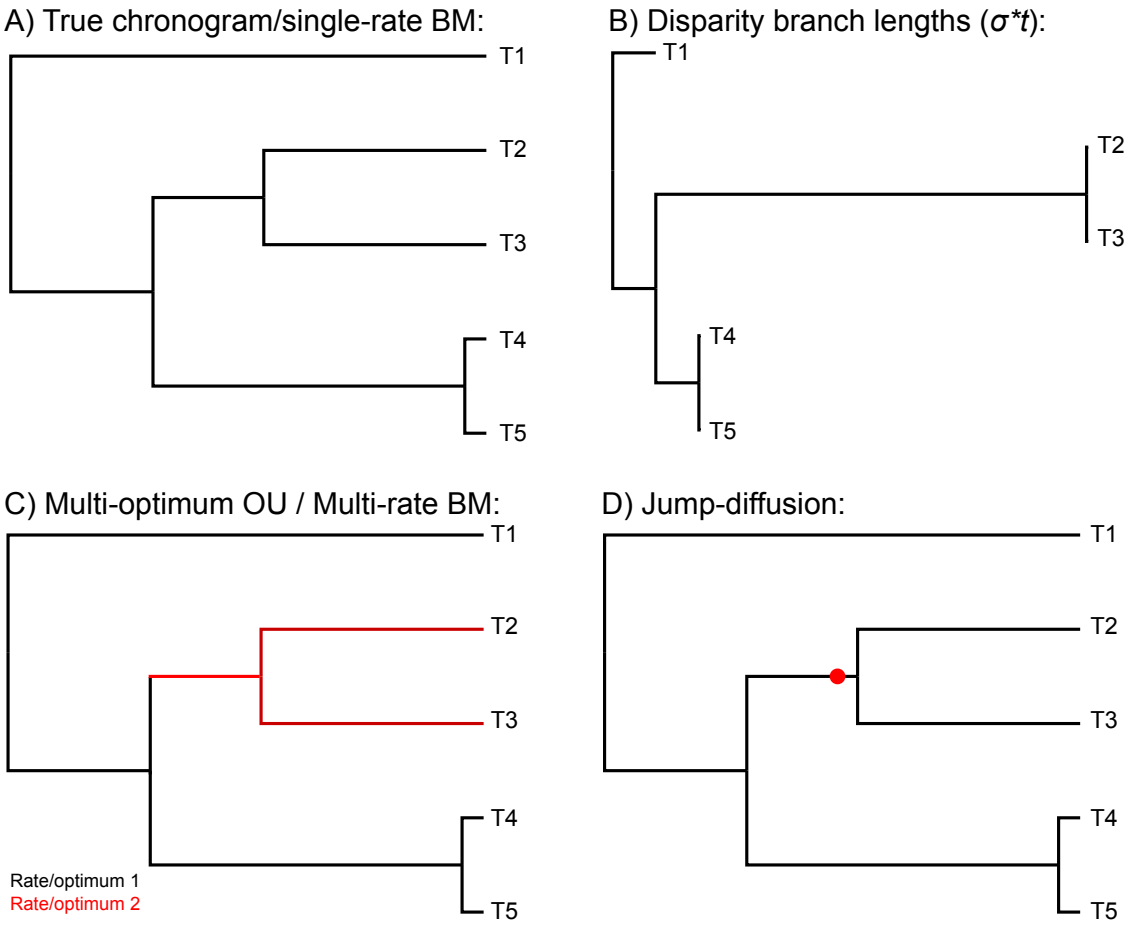

Figure 5.2: Fitting increased disparity observed in taxa T2 and T3. A) Single rate BM fit to a chronogram assumes clock-like evolution, ignoring both positive jumps in disparity and stasis. B) Expressing branch lengths in units of disparity under the phylogenetic Brownian parameterization used here captures both increased disparity and stasis. C) Fitting multiple Brownian rates or shifts in adaptive optima or D) assuming the presence of a single jump in optimum trait value may explain the same pattern with more model parameters. Note that panels B, C, and D all display similar information, while panel A assumes clock-like evolution.

$$x_{internal} = \frac{(x_1 * v_2) + (x_2 * v_1)}{v_1 + v_2} \tag{5.3}$$

$$v_{internal} = v_{internal} + \frac{(v_1 * v_2)}{(v_1 + v_2)} \tag{5.4}$$

The total log-likelihood of the tree, L_{tree} is calculated by summing the log-likelihoods calculated at each of the n internal nodes.

$$L_{tree} = \sum_{node=1}^{n} L_{node} \tag{5.5}$$

Branch lengths are estimated by iteratively solving the analytical solution to the maximum likelihood (ML) branch lengths for a 3-taxon star topology. In this procedure, the tree is treated as unrooted. Picking a single internal node, PICs are calculated to each of the three connected branches. These are treated as 'traits' at the tips of a three-taxon tree. The edge lengths of the pruned tree (v_i) are then computed analytically using the MLE solutions for a three taxon tree (Felsenstein 1981). This procedure is performed on all of the internal nodes. This process is iterated until the branch lengths and the likelihoods converge, yielding a local optimum of the likelihood function. The algorithm and derivation of the 3-taxon ML solutions are given a detailed explanation by Felsenstein (1981).

5.3.4 Search procedure

All traits start in a single shared partition. From here, traits that exhibit an improved likelihood in their own partition compared to their current placement are split into their own cluster. A penalty is imposed that is proportional to the difference in size between the existing partitions, as measured in the number of constituent traits. This functions similarly to a Dirichlet process prior by forcing traits to prefer belonging to partitions with a larger number of traits. This constraint is designed to prioritize traits with strongly divergent signal and to discourage overfitting of the clustering model. As a result, only traits with a strong preference for the new component over the existing component are selected. This step is repeated either until the number of occupied categories reaches a user-specified maximum threshold, or there are no more traits left to separate.

From here, the problem is temporarily recast as a finite mixture model, with the number of components corresponding to the user-specified value. First, membership weights are

calculated for each trait-component pair as the probability of the trait (x_i) belonging to each *j* of *K* components. This value is calculated for each component as the proportion of the likelihood of x_i (L_{ij}) under the corresponding set of branch lengths relative to the summed likelihoods of x_i under all *K* components.

$$P(x_i|K_j) = \frac{L_{ij}}{\sum_{k=1}^{K} L_{ik}} \tag{5.6}$$

Expectation-maximization (EM) (Dempster *et al.* 1977) is performed to update the mixture weights and the branch length parameters. The branch lengths associated with each component are updated as part of the mixture model, with each site in the matrix contributing to the branch lengths in each component according to the weights defined above. During this step, the model could be thought of as a variation of a typical multivariate Gaussian mixture model, where the covariance matrix is constrained to reflect the structure of a phylogenetic tree, since the phylogenetic Brownian model yields a multivariate Gaussian likelihood function.

Once the mixture model has been updated for several iterations, the components are broken into hard clusters, with the assignment for each site chosen to be the component with the maximum mixture weight. This arrangement is then reduced in an agglomerative manner. At each step of this procedure, the pair of components that results in the greatest improvement in AIC, calculated using the classification likelihood defined above, is merged. This hierarchical merging continues until either the AIC score cannot be further improved, or only a single component is left. The entire procedure is then repeated from this reduced configuration for a user-specified number of iterations.

5.3.5 Stochastically varying mosaic patterns

To examine the strengths and shortcomings of the method in detecting multiple suites of traits that vary stochastically, I performed tests using simulated datasets. A single topology of 20 taxa was simulated under a pure-birth model. For each partition of continuous traits, a new set of branch lengths was generated by drawing randomly from either a gamma or exponential distribution, then simulated under Brownian motion. The rate parameter of the Brownian process was set to 1 across the entire tree so that the matrices reflected the scale and heterogeneity of rates resulting from the altered branch lengths. Each trait matrix contained a single partition simulated under the original ultrametric branch lengths. The randomly drawn branch lengths were intended to mimic the differing rates of evolution that can be experienced by different lineages during evolutionary divergence, with the ultrametric branch lengths reflecting clock-like evolution. This procedure resulted in highly complex simulated datasets that tested the ability of the method to detect mosaic structure of high dimensionality. All trees and traits were simulated using the phytools package in R (Revell 2012b).

Using this procedure, datasets comprised of 2, 3, and 4 partitions of 50 continuous traits each were generated. All traits were rescaled to a variance of 1. I ran *greedo* on these datasets to attempt to reconstruct these partitions. The maximum number of clusters for these runs was set to half the number of traits in each matrix.

5.3.6 Power analysis and detection of relative rate shifts

I performed another simulation experiment to evaluate the statistical power and limits of the method under more controlled conditions. To evaluate the performance of the method computationally, I expanded the size of the simulated trees to 100 taxa. In this trial, I simulated datasets using a procedure modeled after Eastman *et al.* (2011). First,

I simulated 100 unique pure-birth trees. I then altered the branch lengths to all be equal, with the trees scaled to a total height of 1. This yielded a set of 100 'equal rates' trees that displayed uniform disparity across all branches. To simulate heterogeneity in evolutionary pattern, I then randomly selected once clade in each of the 100 trees, and multipled the branch lengths by factors of 8, 16, 32, and 64. Like in the tests performed by Eastman *et al.* (2011), these rates were inherited across the entire clade. Clades randomly chosen for shifts were constrained to those containing at least 10, and no more than 90, terminal taxa. This resulted in a set of 400 'rate shift' trees of differing magnitude. I then compiled datasets by simulating 50 traits on the trees with equal branch lengths and the trees displaying rate shifts, respectively, with the goal to ascertain the ability of the method to separate traits displaying a shift in rate from those simulated under uniform rates. The 400 resulting datasets contained 100 traits, half of which were simulated along an equal rates tree, and the other half along a rate shift tree of one of the four shift magnitudes. To examine the prevalence of type 1 error, I also simulated datasets of 100 traits along each of the equal branch length trees to test whether the method correctly identified only one suite of traits. Like above, all traits were scaled to a variance of 1 to ensure that the method was correctly identifying heterogeneous patterns in relative disparity, rather than finding heterogeneity reflected in empirical variance, which would unfairly bias results in its favor. This step was also performed in the test described below.

5.3.7 Distinguishing rate shifts among covarying traits

In nature, many continuous traits are expected to covary. Such covariance has been suggested as a challenge to phylogenetic analyses (Felsenstein 1988) by increasing complexity and bias in morphological datasets. To test the behavior of my approach when detecting mosaic patterns among traits that form covarying modules, I performed an additional experiment. I simulated covarying datasets along the 16x and 64x rate shift trees

generated above. Again, 50 traits each were simulated along the equal lengths and rate shift trees, respectively, but with the traits separated into two covarying sets of 25 traits. These were compiled into datasets of 100 traits, displaying two distinct evolutionary patterns (equal and shifted rates), and four separate covarying modules. To examine the effect of covariance on type 1 error, I also simulated datasets of 100 traits along the equal rates trees, as above, but separated into four covarying modules. The strength of covariance is likely to affect results, and so data were simulated at correlational intervals of 0.1 (weak), 0.5 (moderate), and 0.9 (strong). Since the approx‘ach introduced here does not explicitly model evolutionary or phenotypic integration, this test examined the extent to which these common patterns confound the identification of mosaic patterns in relative disparity. All simulated datasets were generated using the ‘fastBM()’ function in phytools (Revell 2012a).

5.3.8 Evaluation of reconstruction accuracy

I used the adjusted Rand index (ARI) to evaluate the accuracy of the inferred partitionings (Hubert and Arabie 1985) against the true partitionings. The RI measures congruence by counting the pairs of elements that either occupy the same or different clusters in both of the two clusterings, and calculating the proportion of this value relative to all of the possible permutations of elements. As a result, the RI can range from 0, indicating total disagreement, and 1, indicating total agreement. The ARI corrects for the propensity for elements to be randomly placed within the same cluster, with a value of 0 indicating a result indistinguishable from a random assignment of elements, and 1 indicating complete congruence. The metric takes negative values when a clustering is worse than random.

5.3.9 Empirical analysis

To examine the performance of the method on empirical data, I analyzed a dataset comprised of the ossification sequences of 21 cranial bones sampled across 21 taxa obtained from Laurin (2014), initially assembled by Rose (2003). Laurin identified suites of traits displaying distinct phylogenetic patterns using an 'evolutionary' principal component analysis (PCA) and also performed a distance-based hierarchical clustering of the data, making these data well-suited to a test of the method introduced here. I was interested in evaluating the ability of my new approach to detect patterns in relative, rather than absolute, rates across lineages, and so I standardized the variance between the traits to 1. The tree presented in the original study was obtained from the authors and used to perform the comparative analyses here.

5.4 Results and Discussion

5.4.1 Simulated stochastically varying rates

The method is generally able to recover the structure of the simulated datasets. The number of inferred mosaic suites is usually correct, and ARI values are typically well above random. The two-partition analyses are very accurate, with high ARI values, and nearly always correctly identifying the correct number of clusters. The three- and four-partition analyses were less accurate, but still yield results much higher than random, and typically recovering the correct number of character suites. ARI indices achieved for the three- and four- partition analyses are similar to results from simulated data using general Gaussian clustering approaches, such as Gibbs sampling under a Dirichlet process (Dahl 2006).

Despite the generally encouraging results from the simulated data, the trend toward decreasing accuracy when components are added suggests: 1) a limitation of the method in

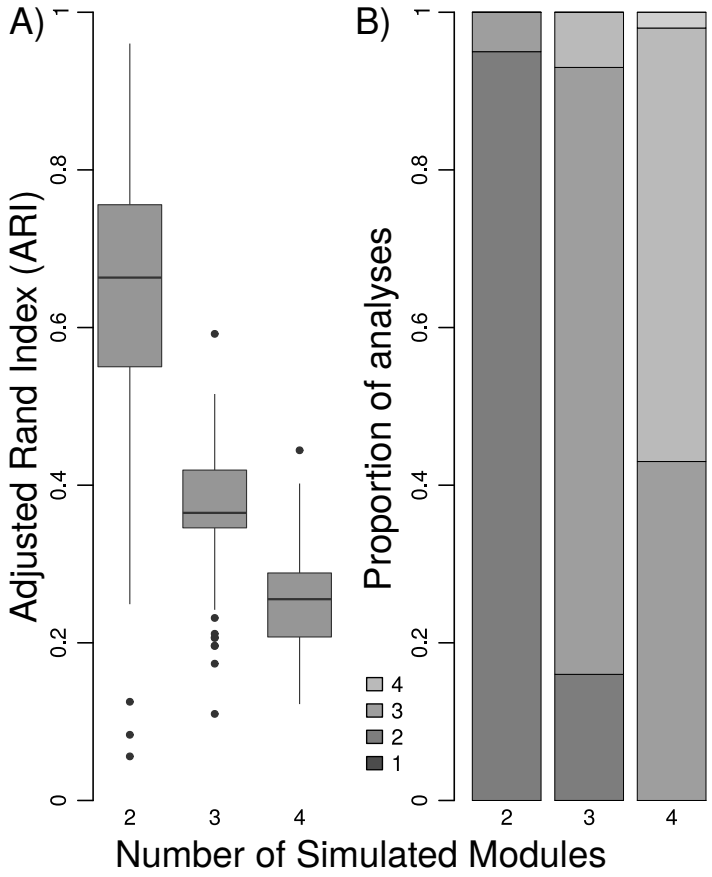

Figure 5.3: A) Adjusted Rand indices across reconstructions of simulated datasets. B) Number of clusters resulting from analyses of simulated data. Barplots are stacked to represent the proportion of replicates that resulted in *k* character suites. All violin and barplots are separated by true number of simulated character suites.

adequately exhausting the search space of component assignments or 2) a limitation in the power of the approach to detect subtle, stochastically varying differences in branch-wise evolutionary disparity. It may be helpful to initialize the analysis using results obtained from a less intensive approach, such as the evolutionary PCA developed by Laurin (2014).

5.4.2 Relative rate shifts in independently evolving traits

The method performs well in distinguishing traits evolved under a single rate shift from those under equal rates (Figs. 4a and 4b). The most subtle rate shift (8x faster) is detectable in most of the datasets, although mean accuracy is slightly lower than the 2-suite datasets tested above. This difference is due to decreased power in detecting this

more subtle heterogeneity. The method underfits the model in 8x rate shift datasets 30% of the time, by assigning all of the traits to a single shared character suite (Fig. 5-4b). Power to detect mosaic structure increases with the magnitude of the shift, with the 16x datasets correctly binning the traits into two character suites 95% of the time. The 32x and 64x datasets always correctly identify the number of character suites. Accuracy also increases with the strength of the rate shift. While ARI values in the 8x dataset are all acceptably high, reconstruction error approaches zero as the magnitude of the rate shifts are increased to 64x.

5.4.3 Relative rate shifts in covarying traits

Evolutionary covariance across traits has a detectable, but not overwhelmingly deleterious effect on the capability of the approach in categorizing the number of character suites present in the data (Figs. 4c and 4d). When the simulated modules weakly covary (0.1), the method performs comparably to the independently evolving examples. However, accuracy decreases with the strength of the covariance, with reconstruction being notably less accurate in the strongly covarying datasets. This decrease in accuracy at high levels of covariance can be explained by the tendency for the approach to overfit the datasets by inferring the presence of more than two character suites in 59% of cases (Fig. 5-4d).

5.4.4 Type 1 error

As with the multi-rate covarying examples above, the method has a tendency to overfit datasets that display high levels of covariance. Nevertheless, the method fits the data correctly in the presence of no and weak covariance (Fig. 5-6). As with above, results generated from datasets suspected to display particularly high levels of covariance should be inspected to ensure that mosaic suites are not overly similar in their resultant branch lengths. As a more immediately tenable alternative to full multivariate estimation, future

extensions would allow the ability for users to specify covarying modules of traits using a conventional approach (Goswami and Finarelli 2016), to ensure that strongly covarying sets of traits are placed together. And nevertheless, even in the absence of these tools, it should be noted that the extremely high and rampant levels of covariance that were needed to induce overfitting in both the type 1 and type 2 tests may be unlikely in certain types of empirical datasets.

5.4.5 Empirical analysis

Four separate runs each yielded different partitionings into two mosaic suites. All arrangements overlap in their assignments, and the AIC scores are all close to one another. To visualize the overall support for the categorization of each trait across partitionings, I calculated the AIC weight of each model (Burnham and Anderson 2002). The AIC weight of model i, w_i can be interpreted as its probability of being the best model among a set of K candidates.

$$w_i = \frac{L_i^{rel}}{\sum_{k=1}^{K} L_k^{rel}} \tag{5.7}$$

where L_i^{rel} is the relative likelihood of model i:

$$L_i^{rel} = \exp(-0.5(AIC_i - AIC_{min})) \tag{5.8}$$

These weights were used to visualize the the strengths of the support connecting traits across all the four best partitionings in a graph (Fig. 5-2b). An edge was drawn between traits i and j if they occurred in the same component in any of the four results, with a weight given by the summed AIC weights of all of the configurations where i and j occur in the same character suite. The maximum weight possible is 1.0, when traits i and j share a suite

in all of the configurations. The resulting graph suggests that traits 0, 1, 2, 17, and 20 all form a suite, with the rest of the traits sharing a separate suite. This result is very close to the pattern reconstructed by Laurin (2014) using an 'evolutionary PCA' approach, differing only in the assignment of the stapes. This is likely because the approach developed by Laurin seeks to join traits that display similar values in phylogenetic independent contrasts (PICs), and so focuses on patterns similar to the phylogenetic Brownian likelihood model used here.

In his original study, Laurin (2014) also performed an exploratory hierarchical clustering of the ossification times and found substantial differences in structure as compared to that revealed by his evolutionary PCA approach. This discordance occurs because the the PCA considers covariance between PIC values, while the hierarchical clustering only reflects shared similarity in absolute value. Like the original study, the results here differ substantially from the pattern resulting from the exploratory hierarchical clustering performed by Laurin, instead aligning very closely to the PCA approach. This is reassuring for the performance of my method, as Laurin considered the evolutionary PCA to yield the correct answer, and the hierarchical clustering to demonstrate the inadequacy of raw similarity in delimiting meaningful patterns (Laurin, pers. comm.). Although they differ in the specific criteria used to identify character suites, the similarity in results between my and Laurin's method are not surprising. Laurin's PCA method identifies structure from patterns in covariance that have been corrected for phylogenetic non-independence, whereas my method identifies a minimally complex set of models, defined by phylogenies with non-negative branch lengths. The trees describing each suite in my method may be thought of as representing disparity between taxa as patristic distances. And so, although they differ in formulation and statistical paradigm, both approaches share overlap in their treatment of phenotypic variation.

5.4.6 Utility in transcriptomic studies

In addition to morphological phenotypes, the method described here may be useful in identifying mosaic suites among molecular phenotypic traits, such as normalized gene expression levels. Expression data have been increasingly examined in a comparative, phylogenetic context, but previous studies have not had a meaningful way in which to partition sets of genes. As a result, researchers typically fall back on methods such as binning all genes expressed in the transcriptome together into a single analysis (Chaix *et al.* 2008), binning genes based upon functional pathways (Schraiber *et al.* 2013), and using non-phylogenetic clustering approaches (Brawand *et al.* 2011). The method described here may benefit such studies by identifying the major axes of variation in evolutionary pattern across transcriptomic datasets.

5.4.7 Utility to phylogenetic comparative methods

By identifying suites of characters that display similar patterns in disparity across lineages, my approach seeks to integrate existing work that takes a broad view of the tempo and mode of phenotypic evolution with under-examined patterns in mosaic evolution. Although the tendency for different traits to evolve according to different patterns is expected and well documented (Stanley 1979; Stebbins 1984), there has not yet been an approach that explicitly incorporates phylogeny to reveal the complex mosaic of patterns in divergence underlying the evolution of phenotypes. The analyses of simulated and empirical data showed the capability of my new method to identify biologically meaningful suites of continuous traits that reflect differences in their patterns in disparity across taxa.

Although Brownian motion is often interpreted in comparative analyses as a neutral process of phenotypic change reflecting genetic drift (e.g., Butler and King 2004) occurring under a single rate, the parameterization used in my approach is more general. As in

previous approaches (Felsenstein 1981), rate and time are confounded with one another. As a result, a long branch representing high disparity to adjacent lineages could reflect either a fast rate, or a long time of divergence. As a result, a tree with heterogeneity in branch lengths could express variation in evolutionary rates across lineages, or tips that were sampled at different points in time. Since phenotypic disparity can be generated by a broad range of processes at the population level, the phylogenetic Brownian model used here does not assume that the traits are selectively neutral. By representing branch lengths in terms of phenotypic disparity, my method is able to capture much of the same information from the data that is sought by more parameter-rich PCMs (Fig. 5-2). As a result, the approach could be form an analytical basis for delimiting suites of traits evolving under distinct evolutionary regimes, such as those found manually by Hopkins and Lidgard (2012). The suites recovered using this method could then be placed into a context more similar to existing PCMs by mapping the disparity branch lengths to a time-scaled phylogeny. This would enable the exploration of absolute rates throughout time, facilitating the exploration of patterns in a framework similar to existing PCMs.

Previous studies have shown that morphological (Lynch 1990) and gene expression phenotypes (Yang *et al.* 2017) often display patterns in rate that are not easily distinguishable from conservative evolutionary forces such as genetic drift and stabilizing selection. Nevertheless, comparative analysis of key traits used in classic studies (Simpson 1944; Gingerich 1983, 1993) have shown that certain features can show substantial variation in rate across lineages that can provide crucial evolutionary insights. As phenotypic datasets increase in scale, it may become increasingly likely for the signature underlying interesting patterns in tempo and mode to become swamped by large numbers of conservatively evolving traits. By separating traits according to their implied patterns in disparity, the approach introduced will help to identify less conservatively evolving traits that may be of

particular interest for more detaileds comparative analyses.

5.4.8 Utility for phylogenetic inference and divergence time estimation

In addition to being comparatively interesting in their own right, the mosaic suites identified by my approach will be useful in methods for inferring phylogeny and divergence times. Several recent articles have demonstrated the strong potential for continuous traits as an alternative to discrete traits when reconstructing phylogeny (Parins-Fukuchi 2018b,a) and divergence times (Alvarez-Carretero *et al.* 2018). Despite these successes, one persistent issue in morphological phylogenetics more generally has been the difficulty in accommodating the expected heterogeneity in branch lengths across traits. This has been argued to be a foundational limitation to model-based approaches by cladists when demonstrating the prevalence of these patterns across empirical datasets (Goloboff *et al.* 2018a). Error stemming from the effect of mosaic branch lengths may be expected to exhibit an even greater effect on divergence time estimation, which depends on calculating rates from branch-wise patterns in disparity.

Although long postulated to be a problem in phylogenetic inference, there has not yet been a computational approach that delimits suites of traits displaying similar patterns in branch lengths. The approach introduced here, if implemented in existing approaches for phylogenetic reconstruction and divergence time estimation, would help to alleviate this source of error. Identifying suites of traits based on patterns in relative evolutionary rate may also provide a framework through which to separate clock-like characters from those displaying more erratic patters. Filtering character data based on conformity to clock-like patterns improves divergence time estimation in molecular data (Smith *et al.* 2018b), and so would also be beneficial in morphological data.

5.4.9 Overfitting, covariance, and the true number of clusters

The approach that I introduce here does not explicitly model covariance among characters. However, when data are simulated as sets of strongly covarying modules, the method infers multiple character suites, despite having been evolved along the same set of branch lengths. This result should not be taken to suggest that the method is well-suited to identifying covarying modules of traits in addition to the mosaic suites that it seeks to discover. Instead, users should be aware that particularly high levels of covariance may lead to slight overfitting, by erroneously splitting up mosaic suites into multiple categories. Ideally, future extensions to this method will explicitly take covariance into account by incorporating older or recent approaches to identifying evolutionary covariance across multiple continuous traits (Felsenstein 1973b). And despite the error encountered at high levels of covariance, the method performs well in the presence of this common pattern by correctly inferring the co-occurrences of traits in the same cluster, despite the fact that suites may be broken up in more extreme cases. As a result, this behavior is biased in a direction that is preferable when attempting to infer the number of clusters in a dataset, which is generally either unknown, or ambiguous. While it is simple to manually combine clusters that have been overfit due to bias, underfitting causes users to miss potentially valuable structure in the data.

The 'overfitting' effect observed in the simulation experiment is predictable. The method that I introduce here is fundamentally a clustering method in design and implementation. Identification of the 'correct' number of clusters can often become a philosophical problem when analyzing complex datasets, such as the covarying modules that comprise mosaic suites above (Hennig 2015). Although multiple modules may be evolved according the same pattern in branch lengths, stochastic variation between separate instantiations of random walks may create real differences in the resulting evolutionary signal

between them. This stochastic effect, compounded by the large number of highly correlated traits within each covarying module, (Fig. 5-8c) generates strongly informative signal favoring precise differences from the generating set of branch lengths. Put simply for the user, larger and more strongly covarying modules of traits will increase the probability of encountering this effect. However, this behavior might be viewed as a benefit, by demonstrating the method's ability to detect small differences from covarying data that result from stochastic variation in macroevolutionary processes. Researchers who are interested in both macroevolutionary patterns and morphological integration should employ my method in conjunction with an approach explicitly designed to recover integration patterns. This is because the simulations demonstrate that, although the method can identify strongly covarying modules, it will miss weak and moderate covariance among traits. This property is implied within the aims of the method, and so researchers should be thoughtful in parsing through these issues.

5.4.10 Morphological integration, mosaic evolution, and geometric morphometrics

The method that I have introduced here is designed to identify shared mosaic patterns in evolutionary disparity, or relative rates across lineages. This aim is related to, but distinct from, an existing body of literature that aims to identify correlational patterns in morphological integration and modularity (Olson and Miller 1958; Cheverud 1982; Goswami 2006; Goswami *et al.* 2009). The mosaic suites identified by this study are a superset of the modules identified in typical morphological integration studies (Fig. 5-8). Strong covariance between traits implies the presence of a shared underlying pattern of macroevolutionary disparity (Fig. 5-8c) (Felice *et al.* 2018). However, as is demonstrated by the simulation experiment performed here, independent modules of covarying traits may share macroevolutionary patterns. This has also been demonstrated empirically by Felice and Goswami (2018) in avian cranial modules, by revealing both mosaic and shared patterns

between integrated modules. Although the method that I present here is capable of 'accidentally' discovering covarying modules of traits by overfitting biased patterns that result from strong and rampant covariance, future extensions of the method should more rigorously explore this interaction. One avenue would be to incorporate extensions to the likelihood model that explicitly account for covariance among traits (Adams 2014; Alvarez-Carretero *et al.* 2018). Alternatively, it would be more immediately managable to apply the approach to data that have already been explored for integration. This would enable the algorithmic exploration of mosaic patterns among a large number of pre-identified covarying modules of traits, as has been done previously to infer phylogeny and mosaic evolution from geometric data (González-José *et al.* 2008b; Felice and Goswami 2018). This step will be important when applying the method to geometric morphometric data, which are often collected to exhibit high levels of integration and covariance, but may be less necessary when exploring patterins in data that represent higher-level morphological structures, such as the Laurin ossification dataset explored here, or large sets of linear measurements and shape ratios. Emerging methods that algorithmically segment CT scans of whole organisms (Dunmore *et al.* 2018) and quantify shape differences in specimen images often provide a small number of independent metrics for each recovered segment (Pomidor *et al.* 2016). The method introduced here would be useful in harnessing such data to identify mosaic patterns across the entire organisms. s

5.4.11 Scale and rate

The approach described here seeks to identify suites of traits sharing similar patterns in relative evolutionary disparity across lineages. Continuous traits displaying greater empirical variances will display higher absolute rates of evolution when modeled under Brownian motion. As a result, I normalized the variances of the datasets used in the simulation

experiments and empirical case study. Nevertheless, alteration of the scale of continuous traits may often change the interpretation of results, and so should be performed thoughtfully. In cases where phenotypes are quantified using a single, shared set of units, such as with geometric morphometric data, standardization of the variances across traits erases information characterizing absolute evolutionary rate. In such carefully constructed datasets, including the matrix of developmental sequences used in the empirical example above, researchers may wish to quantify differences in absolute evolutionary rate across characters. For instance, using the same dataset, Germain and Laurin (2009) demonstrated substantial variability in absolute rate across traits. Study of absolute and relative rates can each yield unique insights into evolutionary processes, and so the scaling of traits should be considered carefully. I did not explore inference of heterogeneity in absolute rates here because it is an easier clustering problem, and thus would have been an overly favorable test of the method that I introduced. In addition, approaches that identify heterogeneity in absolute rate have been well explored in the molecular phylogenetics literature (Yang 1996), and successfully ported to continuous traits (Schraiber *et al.* 2013).

5.4.12 Choice of information criteria

In the analyses performed here, I exclusively used the AIC, in lieu of the corrected version, AICc, and the Bayesian Information Criterion (BIC). Previous authors have suggested that the AICc should be generally preferred to the uncorrected version (Burnham and Anderson 2002). My preference for the AIC was driven by several factors. The number of clusters is generally completely unknown prior to the analysis, and perhaps more importantly, there is generally no single 'true' clustering underlying the mosaic evolutionary patterns sought by the method. As a result, it might generally be preferable in the context of addressing comparative questions to identify a small number of spurious com-

ponents in the final configuration than to ignore important biological variation that could be missed due to the steeper penalty imposed by the AICc. The analyses here support this justification. The simulated analyses show that, when AIC is used, overestimating the number of components is not a major problem (Fig. 5-2). In addition, the results of the empirical analysis suggest that more coherent patterns emerge when several well-supported configurations are averaged. If spurious partitions are encountered in some arrangements, averaging over the results should generally reveal reasonably strong connections between points occupying overfit components.

Although BIC has been used successfully to select the number of components in mixture models (Fraley and Raftery 1998), I preferred the behavior and basis of AIC for these analyses. BIC assumes that the true model is within the set of candidate models, and so can be sensitive to model-misspecification (Wagenmakers and Farrell 2004). This assumption is incompatible with the goals of my method, which does not seek to identify a single 'true' configuration, but instead characterize the major axes of heterogeneity in disparity across lineages. This goal is more consistent with AIC, which simply seeks to identify the model that yields the lowest amount of information loss relative to the dataset. Despite my preference for AIC in the analyses presented here, AICc or BIC may be more appropriate in other situations. As such, researchers should be thoughtful in their choice of information criterion when performing the approach introduced here.

5.5 Conclusions

While there have been fundamental and seminal works sketching out major historical trends in morphological disparification, most offer only a glimpse into the diversity of patterns that shape separate anatomical regions and operate at different phenotypic levels that extend from the gene to the environment. Although a foundational concept in post-

synthesis evolutionary biology, researchers have been limited in their ability to explore mosaic evolution by both the difficulty in assembling comprehensive phenotypic datasets, and the lack of computational methods to handle comparative problems of significant breadth. Emerging approaches that can quantify morphology across entire organisms along with the increasing availability of large-scale transcriptomic and environmental datasets offer the unprecedented opportunity to develop a synthetic evolutionary picture that includes complex mosaic patterns operating at separate phenotypic levels and timescales. Methods like the one introduced here will be critical to these developments by facilitating the reconstruction of diverse mosaic patterns that have shaped evolutionary variation at different phenotypic levels across the tree of life, fulfilling the promises of the modern synthesis.

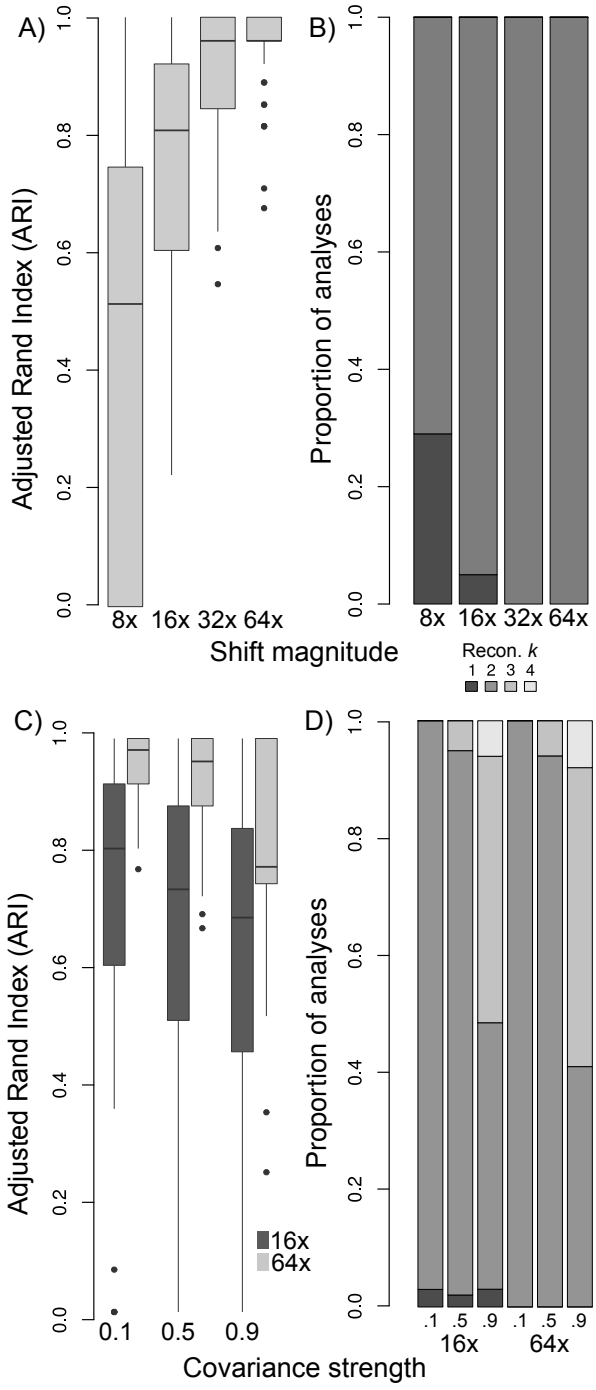

Figure 5.4: A) Adjusted Rand indices from reconstruction of datasets comprised of mosaic suites with one simulated under equal rates across branches, and the other displaying a randomly placed rate shift at magnitudes of 8, 16, 32, and 64. B) Stacked barplots to show the frequency that each of k clusters are inferred from the simulated datasets. C) and D) show the same metrics as above, summarized from datasets containing simulated covarying modules of characters.

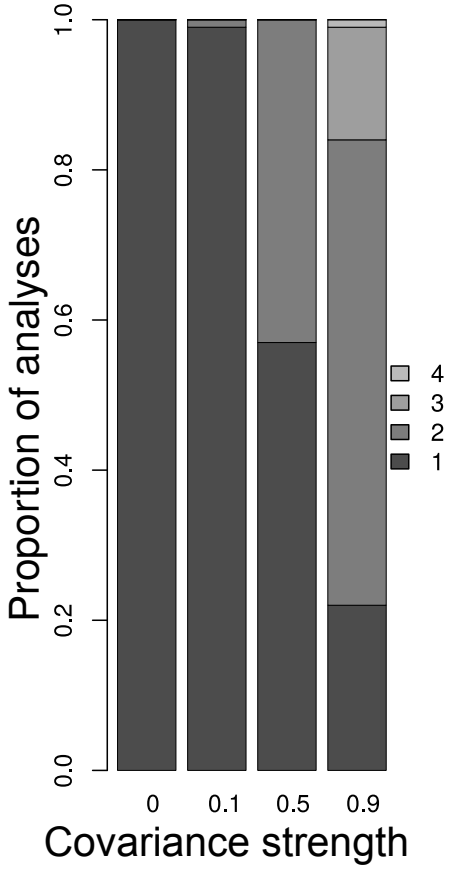

Figure 5.5: Type 1 error. Reconstructions show no type 1 error when traits are evolved independently, or display weak covariance. Type 1 error increases at moderate and high levels of covariance, as the method overfits biased patterns to place covarying modules in superfluous mosaic suites.

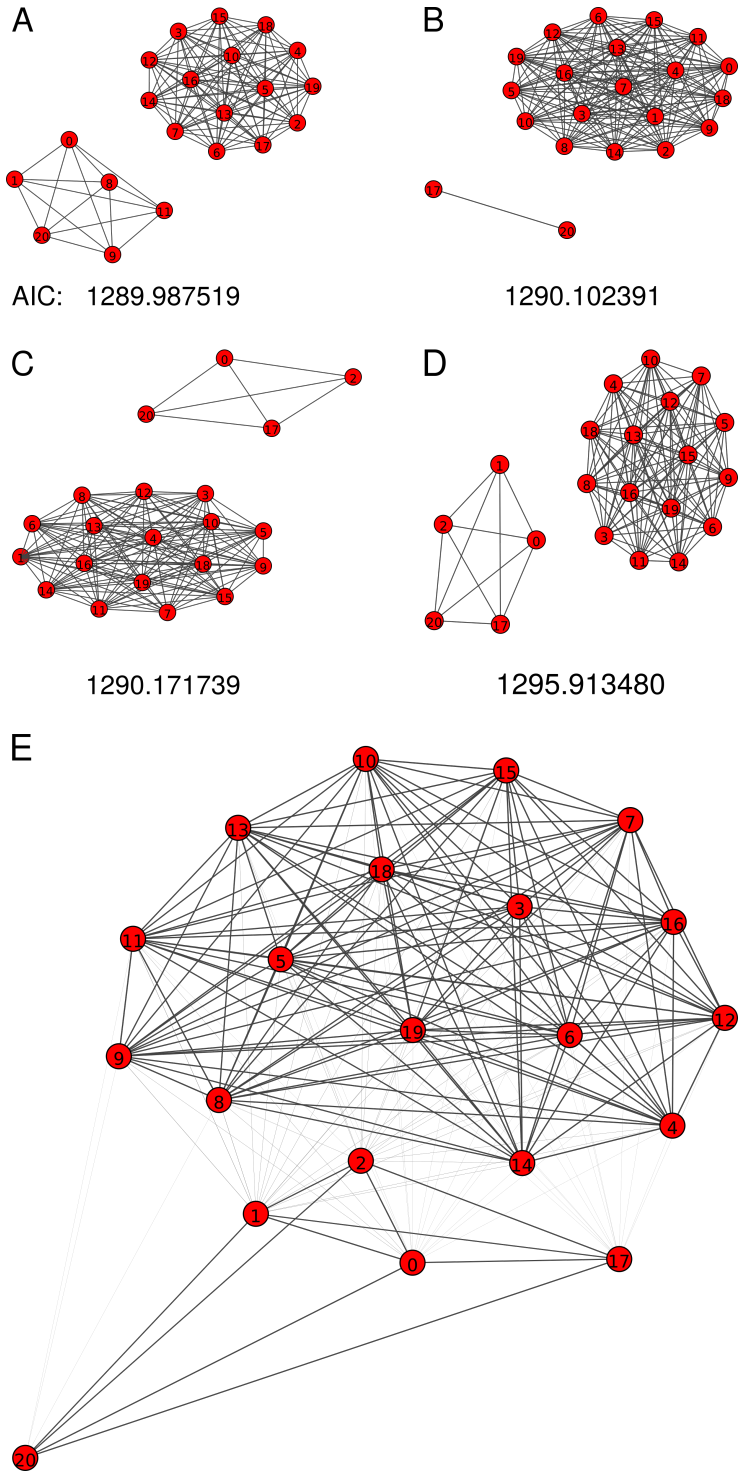

Figure 5.6: A-D) Four best configurations with AIC scores. E) Weighted graph calculated by summing the AIC weights associated with the each model to form edges and edge weights. All graphs were drawn using the ”lgl” format implemented in igraph.

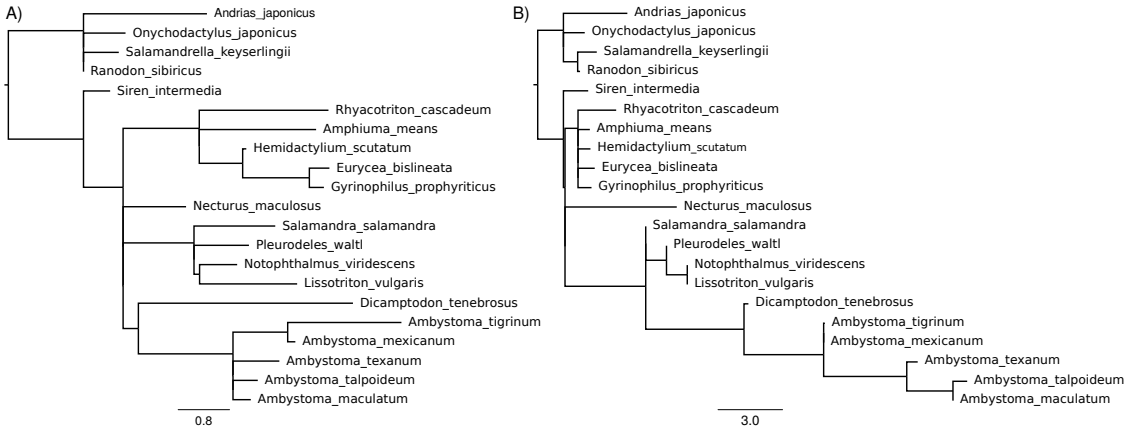

Figure 5.7: Branch lengths reconstructed from traits contained within: A) suite 0 and B) suite 1 display distinct patterns in disparity.

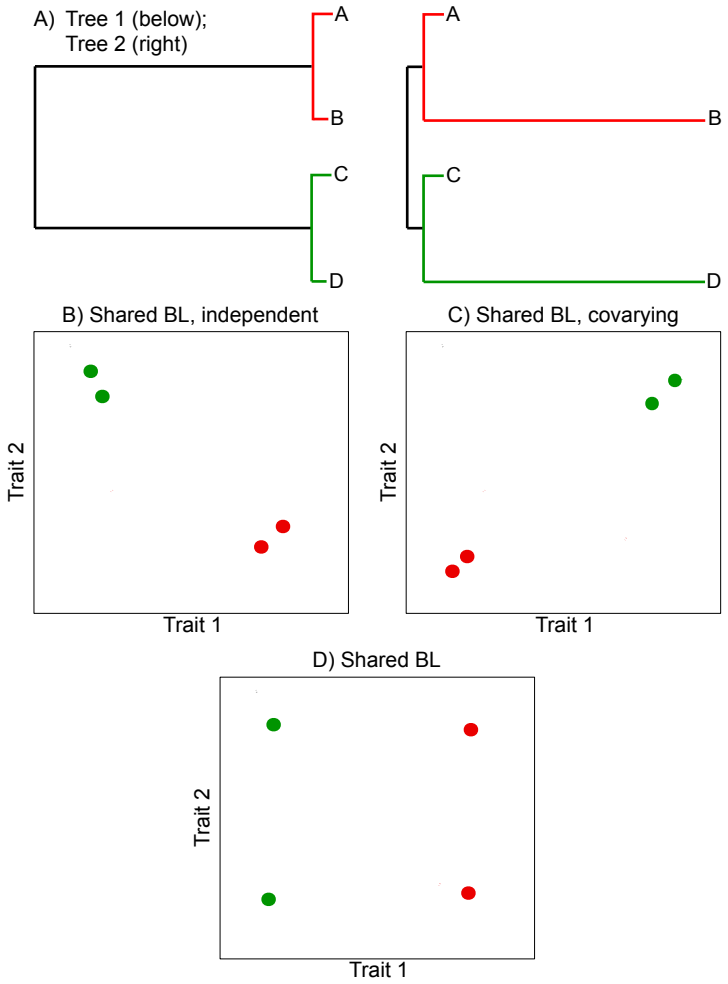

Figure 5.8: Expected patterns between traits with and without integration. A) Two possible patterns in disparity represented by differences in disparity-scaled branch lengths. B) Values of two traits generated according to the pattern implied by tree 1 in panel A, but evolved independently. C) Two traits generated according to tree 1, but with strongly covarying evolutionary paths. D) Two traits generated independently and distinctly according to tree 1 and 2. Note that the covariance in panel C implies a shared pattern in disparity, while the examples in panels B and D do not.

CHAPTER VI

Conclusion

The work contained within this dissertation is intended as a series of sketches outlining some potential methodological and conceptual directions for an evolutionary paleobiology for the genomic era. Far from being complete, my aim in the preceding chapters was to provide some fundamental methological and conceptual tools that can help to facilitate the development of a modern computational evolutionary paleobiology. Although retaining many rough edges in implementation, I feel that the work here can provide a useful methodological groundwork for a paleontological data revolution that synthesizes high-throughput quantitative morphologic data, stratigraphy, and development.

6.1 Paths forward

Methods for the algorithmic quantification of variation in morphology have experienced substantial recent advances (Boyer *et al.* 2015; Pomidor *et al.* 2016). However, the field has yet to reach a critical mass in the usage of these new approaches. In addition to highlighting the obvious benefits of facilitating the rapid, objective construction of large morphologic datasets, increased adoption of these approaches will undoubtedly reveal substantial challenges and gaps. For instance, while public repositories for 3D specimen data are experiencing increased use (Boyer *et al.* 2016), public quantitative morphologic datasets that maintain clear ontology of traits remain scarse. In addition, as these

new sources of data increase, they will surely demand many new methodological advances.

Although the development of analytical approaches that accommodate emerging sources of 'high-throughput' morphologic data is a recurrent motivating theme throughout this dissertation, traditional morphologic and morphometric techniques will remain highly valuable. For example, in chapter 3, while the Carnivoran geometric morphometric landmarks adequately informed the placement of fossil taxa, the linear ratios measured provided higher information and confidence. If this pattern proves to be common, linear measurement will remain an important tool when assembling continuous datasets for phylogenetic analyses. Despite these likely challenges moving ahead, I remain convinced that the further development of a quantitative, statistical phylogenetic framework will yield substantial growth, facilitating a renewed synthesis of paleontological and neontological theory fit for the 21st century.

6.2 Concluding remarks

When designing and developing the work contained in this dissertation, my goal was to provide a somewhat comprehensive conceptual and methodological foundation for a more data-rich future in paleontology. While the completion of this task is surely a generational goal that will demand the input of many researchers, this dissertation is aimed to provide a set of initial, minimally complex approaches intended as a nudge in the 'right direction'. While the years in which I have developed these works have witnessed the proliferation of new, increasingly complex parametric approaches to phylogenetic inference in analytical paleobiology, my goals were somewhat distinct from most of this emerging body of work. Instead of emphasizing increased model complexity, I sought to encourage a re-evaluation of the foundational sources of information that we use to assess phylogeny and more complicated evolutionary dynamics. In my role as a parametric data scientist, I also sought

to investigate the information presented by these unconventional data sources when evaluated as nakedly as is possible. From that starting point, I attempted, such as in chapter 5, to extend the models and computational approaches in ways that accommodate the most pressing sources of heterogeneity in comparative pattern.

The approaches introduced in the course of this dissertation were designed to further the goals toward a 'next-generation' of evolutionary paleobiology. Since John Maynard Smith's welcoming of the field to the high table, the molecular revolution in evolutionary biology has left paleontology comparatively limited in its ability to generate sweeping empirical insights. The continuing construction of a new methodological foundation that accommodates emerging sources of data unconventional to traditional paleontology will be essential as our field climbs its way to the high table of the genomic era.

Bibliography

Adams, D. C. 2014. A method for assessing phylogenetic least squares models for shape and other high-dimensional multivariate data. *Evolution*, 68(9): 2675–2688.

Adams, D. C. and Felice, R. N. 2014. Assessing trait covariation and morphological integration on phylogenies using evolutionary covariance matrices. *PloS One*, 9(4): e94335.

Adams, D. C. and Rosenberg, M. S. 1998. Partial warps, phylogeny, and ontogeny: a comment on fink and zelditch (1995). *Systematic Biology*, 47(1): 168–173.

Alvarez-Carretero, S., Goswami, A., Yang, Z., and dos Reis, M. 2018. Bayesian estimation of species divergence times using correlated quantitative characters. *bioRxiv*, page 441105.

Amson, E. and de Muizon, C. 2014. A new durophagous phocid (mammalia: Carnivora) from the late neogene of peru and considerations on monachine seals phylogeny. *Journal of Systematic Palaeontology*, 12(5): 523–548.

Andersson, L., Haley, C. S., Ellegren, H., Knott, S. A., Johansson, M., Andersson, K., Andersson-Eklund, L., Edfors-Lilja, I., Fredholm, M., Hansson, I., *et al.* 1994. Genetic mapping of quantitative trait loci for growth and fatness in pigs. *Science*, 263(5154): 1771–1775.

Arcila, D., Pyron, R. A., Tyler, J. C., Ortí, G., and Betancur-R, R. 2015. An evaluation of

fossil tip-dating versus node-age calibrations in tetraodontiform fishes (teleostei: Percomorphaceae). *Molecular Phylogenetics and Evolutiion*, 82: 131–145.

Asfaw, B., White, T., Lovejoy, O., Latimer, B., Simpson, S., and Suwa, G. 1999. Australopithecus garhi: a new species of early hominid from ethiopia. *Science*, 284(5414): 629–635.

Aze, T., Ezard, T. H., Purvis, A., Coxall, H. K., Stewart, D. R., Wade, B. S., and Pearson, P. N. 2011. A phylogeny of cenozoic macroperforate planktonic foraminifera from fossil data. *Biological Reviews*, 86(4): 900–927.

Aze, T., Ezard, T. H., Purvis, A., Coxall, H. K., Stewart, D. R., Wade, B. S., and Pearson, P. N. 2013. Identifying anagenesis and cladogenesis in the fossil record. *Proceedings of the National Academy of Sciences*, 110(32): E2946–E2946.

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

Bapst, D. W. and Hopkins, M. J. 2017. Comparing cal3 and other a posteriori time-scaling approaches in a case study with the pterocephaliid trilobites. *Paleobiology*, 43(1): 49–67.

Barton, R. A. and Harvey, P. H. 2000. Mosaic evolution of brain structure in mammals. *Nature*, 405(6790): 1055.

Beaulieu, J. M., Moles, A. T., Leitch, I. J., Bennett, M. D., Dickie, J. B., and Knight, C. A. 2007. Correlated evolution of genome size and seed mass. *New Phytologist*, 173(2): 422–437.

Beaulieu, J. M., Jhwueng, D.-C., Boettiger, C., and O'Meara, B. C. 2012. Modeling

stabilizing selection: expanding the ornstein–uhlenbeck model of adaptive evolution. *Evolution*, 66(8): 2369–2383.

Berger, S. A. and Stamatakis, A. 2010. Accuracy of morphology-based phylogenetic fossil placement under maximum likelihood. pages 1–9.

Bloch, J. I. and Boyer, D. M. 2002. Grasping primate origins. *Science*, 298(5598): 1606–1610.

Bokma, F., Godinot, M., Maridet, O., Ladevèze, S., Costeur, L., Solé, F., Gheerbrant, E., Peigné, S., Jacques, F., and Laurin, M. 2015. Testing for depéret's rule (body size increase) in mammals using combined extinct and extant data. *Systematic biology*, 65(1): 98–108.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M. A., Rambaut, A., and Drummond, A. J. 2014. Beast 2: a software platform for bayesian evolutionary analysis. *PLoS Comput. Biol.*, 10(4): e1003537.

Boyer, D. M., Puente, J., Gladman, J. T., Glynn, C., Mukherjee, S., Yapuncich, G. S., and Daubechies, I. 2015. A new fully automated approach for aligning and comparing shapes. *The Anatomical Record*, 298(1): 249–276.

Boyer, D. M., Gunnell, G. F., Kaufman, S., and McGeary, T. M. 2016. Morphosource: Archiving and sharing 3-d digital specimen data. *The Paleontological Society Papers*, 22: 157–181.

Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csárdi, G., Harrigan, P., Weier, M., Liechti, A., Aximu-Petri, A., Kircher, M., *et al.* 2011. The evolution of gene expression levels in mammalian organs. *Nature*, 478(7369): 343.

Brazeau, M. D. 2011. Problematic character coding methods in morphology and their effects. *Biological Journal of the Linnean Society*, 104(3): 489–498.

Brooks, D. R. 1981. Hennig's parasitological method: A proposed solution. *Systematic Biology*, 30(3): 229–249.

Brown, J. W., Parins-Fukuchi, C., Stull, G. W., Vargas, O. M., and Smith, S. A. 2017. Bayesian and likelihood phylogenetic reconstructions of morphological traits are not discordant when taking uncertainty into consideration: a comment on puttick et al. *Proceedings of the Royal Society B: Biological Sciences*, 284(1864): 20170986.

Brunet, M., Guy, F., Pilbeam, D., Mackaye, H. T., Likius, A., Ahounta, D., Beauvilain, A., Blondel, C., Bocherens, H., Boisserie, J.-R., *et al.* 2002. A new hominid from the upper miocene of chad, central africa. *Nature*, 418(6894): 145.

Buck, L. T. and Stringer, C. B. 2014. Homo heidelbergensis. *Current Biology*, 24(6): R214–R215.

Burbrink, F. T. and Pyron, R. A. 2010. How does ecological opportunity influence rates of speciation, extinction, and morphological diversification in new world ratsnakes (tribe lampropeltini)? *Evolution: International Journal of Organic Evolution*, 64(4): 934–943.

Burke, A. C., Nelson, C. E., Morgan, B. A., and Tabin, C. 1995. Hox genes and the evolution of vertebrate axial morphology. *Development*, 121(2): 333–346.

Burnham, K. P. and Anderson, D. R. 2002. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer.

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

Caetano, D. S. and Harmon, L. J. 2018. Estimating correlated rates of trait evolution with uncertainty. *Systematic Biology*, syy067.

Catalano, S. A., Goloboff, P. A., and Giannini, N. P. 2010. Phylogenetic morphometrics (i): the use of landmark data in a phylogenetic framework. *Cladistics*, 26(5): 539–549.

Cavalli-Sforza, L. L. and Edwards, A. W. 1967. Phylogenetic analysis. models and estimation procedures. *American Journal of Human Genetics*, 19(3): 233.

Chaix, R., Somel, M., Kreil, D. P., Khaitovich, P., and Lunter, G. 2008. Evolution of primate gene expression: drift and corrective sweeps? *Genetics*.

Chamberlain, A. and Wood, B. A. 1987. Early hominid phylogeny. *Journal of Human Evolution*, 16(1): 119–133.

Chang, J. and Alfaro, M. 2015a. Crowdsourced morphometric data are as accurate as traditionally collected data in 7 ray-finned fish families. *Integr. Comp. Biol.*, 55: E28–E28.

Chang, J. and Alfaro, M. E. 2015b. Crowdsourced geometric morphometrics enable rapid large-scale collection and analysis of phenotypic data. *Methods in Ecology and Evolution*.

Chappill, J. A. 1989. Quantitative characters in phylogenetic analysis. *Cladistics*, 5(3): 217–234.

Chen, I. 2009. *History of Vitaceae inferred from morphology-based phylogeny and the fossil record of seeds*. University of Florida.

Cheverud, J. M. 1982. Phenotypic, genetic, and environmental morphological integration in the cranium. *Evolution*, 36(3): 499–516.

Cohn, M. J. and Tickle, C. 1999. Developmental basis of limblessness and axial patterning in snakes. *Nature*, 399(6735): 474.

Collard, M. and Wood, B. 2000. How reliable are human phylogenetic hypotheses? *Proceedings of the National Academy of Sciences*, 97(9): 5003–5006.

Cook, C. E., Smith, M. L., Telford, M. J., Bastianello, A., and Akam, M. 2001. Hox genes and the phylogeny of the arthropods. *Curr. Biol.*, 11(10): 759–763.

Cooper, N., Thomas, G. H., Venditti, C., Meade, A., and Freckleton, R. P. 2016. A cautionary note on the use of ornstein uhlenbeck models in macroevolutionary studies. *Biological Journal of the Linnean Society*, 118(1): 64–77.

Cracraft, J. 1970. Mandible of archaeopteryx provides an example of mosaic evolution. *Nature*, 226(5252): 1268.

Cracraft, J. 1974. Phylogeny and evolution of the ratite birds. *Ibis*, 116(4): 494–521.

Cracraft, J. 1982. Geographic differentiation, cladistics, and vicariance biogeography: reconstructing the tempo and mode of evolution. *American Zoologist*, 22(2): 411–424.

Cressler, C. E., Butler, M. A., and King, A. A. 2015. Detecting adaptive evolution in phylogenetic comparative analysis using the ornstein–uhlenbeck model. *Systematic biology*, 64(6): 953–968.

Cui, L., Wall, P. K., Leebens-Mack, J. H., Lindsay, B. G., Soltis, D. E., Doyle, J. J., Soltis, P. S., Carlson, J. E., Arumuganathan, K., Barakat, A., *et al.* 2006. Widespread genome duplications throughout the history of flowering plants. *Genome research*, 16(6): 738–749.

Dahl, D. B. 2006. Model-based clustering for expression data via a dirichlet process mixture model. *Bayesian inference for gene expression and proteomics*, 201: 218.

Darwin, C. 1859. *On the Origin of Species*.

Darwin, C. 1871. The descent of man.

Dávalos, L. M., Velazco, P. M., Warsi, O. M., Smits, P. D., and Simmons, N. B. 2014. Integrating incomplete fossils by isolating conflicting signal in saturated and non-independent morphological characters. *Systematic Biology*, 63(4): 582–600.

De Beer, G. R. 1954. *Archaeopteryx lithographica: A study based upon the British Museum specimen*. Trustees of the British Museum.

De Bodt, S., Maere, S., and Van de Peer, Y. 2005. Genome duplication and the origin of angiosperms. *Trends in ecology & evolution*, 20(11): 591–597.

De Castro, J. B., Arsuaga, J. L., Carbonell, E., Rosas, A., Martınez, I., and Mosquera, M. 1997. A hominid from the lower pleistocene of atapuerca, spain: possible ancestor to neandertals and modern humans. *Science*, 276(5317): 1392–1395.

De Rosa, R., Grenier, J. K., Andreeva, T., Cook, C. E., Adoutte, A., Akam, M., Carroll, S. B., and Balavoine, G. 1999. Hox genes in brachiopods and priapulids and protostome evolution. *Nature*, 399(6738): 772–776.

Dehal, P. and Boore, J. L. 2005. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS biology*, 3(10): e314.

Delson, E., Eldredge, N., and Tattersall, I. 1977. Reconstruction of hominid phylogeny: a testable framework based on cladistic analysis. *Journal of Human Evolution*, 6(3): 263–278.

Dembo, M., Matzke, N. J., Mooers, A. Ø., and Collard, M. 2015. Bayesian analysis of a morphological supermatrix sheds light on controversial fossil hominin relationships. *Proceedings of the Royal Society B*, 282(1812): 20150943.

Dembo, M., Radovčić, D., Garvin, H. M., Laird, M. F., Schroeder, L., Scott, J. E., Brophy, J., Ackermann, R. R., Musiba, C. M., de Ruiter, D. J., *et al.* 2016. The evolutionary relationships and age of homo naledi: An assessment using dated bayesian phylogenetic methods. *Journal of Human Evolution*, 97: 17–26.

Dempster, A. P., Laird, N. M., and Rubin, D. B. 1977. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1): 1–38.

Donoghue, M. J. and Ree, R. H. 2000. Homoplasy and developmental constraint: A model and an example from plants 1. *Amer. Zool.*, 40(5): 759–769.

Dunmore, C. J., Wollny, G., and Skinner, M. M. 2018. Mia-clustering: a novel method for segmentation of paleontological material. *PeerJ*, 6: e4374.

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

Eastman, J. M., Wegmann, D., Leuenberger, C., and Harmon, L. J. 2013. Simpsonian 'evolution by jumps' in an adaptive radiation of anolis lizards. *arXiv preprint arXiv:1305.4216*.

Farris, J. S. 1971. The hypothesis of nonspecificity and taxonomic congruence. *Annual Review of Ecology and Systematics*, 2(1): 277–302.

Farris, J. S. 1990. Phenetics in camouflage. *Cladistics*, 6(1): 91–100.

Felice, R. N. and Goswami, A. 2018. Developmental origins of mosaic evolution in the avian cranium. *Proceedings of the National Academy of Sciences*, 115(3): 555–560.

Felice, R. N., Randau, M., and Goswami, A. 2018. A fly in a tube: Macroevolutionary expectations for integrated phenotypes. *Evolution*.

Felsenstein, J. 1973a. Maximum-likelihood estimation of evolutionary trees from continuous characters. *American Journal of Human Genetics*, 25(5): 471.

Felsenstein, J. 1973b. Maximum-likelihood estimation of evolutionary trees from continuous characters. *American Journal of Human Genetics*, 25(5): 471.

Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology*, 27(4): 401–410.

Felsenstein, J. 1981a. Evolutionary trees from dna sequences: a maximum likelihood approach. *Journal of Molecular Evolution*, 17(6): 368–376.

Felsenstein, J. 1981b. Evolutionary trees from gene frequencies and quantitative characters: finding maximum likelihood estimates. *Evolution*, 35(6): 1229–1242.

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

Felsenstein, J. 1988. Phylogenies and quantitative characters. *Annual Review of Ecology and Systematics*, 19(1): 445–471.

Felsenstein, J. 2002. Quantitative characters, phylogenies, and morphometrics. *Morphology, shape and phylogeny*, pages 27–44.

Fisher, D. C. 1991. Phylogenetic analysis and its application in evolutionary paleobiology. *Short Courses in Paleontology*, 4: 103–122.

Fisher, D. C. 2008. Stratocladistics: integrating temporal data and character data in phylogenetic inference. *Annual Review of Ecology, Evolution, and Systematics*, 39: 365–385.

Fisher, D. C., Grande, L., and Rieppel, O. 1994. Stratocladistics: morphological and temporal patterns and their relation to phylogenetic process. *Interpreting the Hierarchy of Nature*, pages 133–171.

Fontaine, M. C., Pease, J. B., Steele, A., Waterhouse, R. M., Neafsey, D. E., Sharakhov, I. V., Jiang, X., Hall, A. B., Catteruccia, F., Kakani, E., *et al.* 2015. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science*, 347(6217): 1258524.

Foote, M. 1997. Estimating taxonomic durations and preservation probability. *Paleobiology*, 23(3): 278–300.

Foote, M. 2001. Inferring temporal patterns of preservation, origination, and extinction from taxonomic survivorship analysis. *Paleobiology*, 27(4): 602–630.

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

Fox, D. L., Fisher, D. C., and Leighton, L. R. 1999. Reconstructing phylogeny with and without temporal data. *Science*, 284(5421): 1816–1819.

Fraley, C. and Raftery, A. E. 1998. How many clusters? which clustering method? answers via model-based cluster analysis. *The Computer Journal*, 41(8): 578–588.

Frary, A., Nesbitt, T. C., Frary, A., Grandillo, S., Van Der Knaap, E., Cong, B., Liu, J., Meller, J., Elber, R., Alpert, K. B., *et al.* 2000. fw2. 2: a quantitative trait locus key to the evolution of tomato fruit size. *Science*, 289(5476): 85–88.

Freckleton, R. P. 2012. Fast likelihood calculations for comparative analyses. *Methods in Ecology and Evolution*, 3(5): 940–947.

Gavryushkina, A., Heath, T. A., Ksepka, D. T., Stadler, T., Welch, D., and Drummond, A. J. 2017. Bayesian total-evidence dating reveals the recent crown radiation of penguins. *Systematic Biology*, 66(1): 57–73.

Germain, D. and Laurin, M. 2009. Evolution of ossification sequences in salamanders and urodele origins assessed through event-pairing and new methods. *Evolution & development*, 11(2): 170–190.

Gingerich, P. D. 1974. Stratigraphic record of early eocene hyopsodus and the geometry of mammalian phylogeny. *Nature*, 248(5444): 107.

Gingerich, P. D. 1979a. Paleontology, phylogeny, and classification: an example from the mammalian fossil record. *Systematic Zoology*, 28(4): 451–464.

Gingerich, P. D. 1979b. The stratophenetic approach to phylogeny reconstruction in vertebrate paleontology. In J. Cracraft and N. Eldredge, editors, *Phylogenetic Analysis and Paleontology*, volume 41, page 77. Columbia University Press New York.

Gingerich, P. D. 1983. Rates of evolution: effects of time and temporal scaling. *Science*, 222(4620): 159–161.

Gingerich, P. D. 1985. Species in the fossil record: concepts, trends, and transitions. *Paleobiology*, 11(1): 27–41.

Gingerich, P. D. 1993. Quantification and comparison of evolutionary rates. *Am. J. Sci.*, 293(A): 453–478.

Goloboff, P. A. and Catalano, S. A. 2016. Tnt version 1.5, including a full implementation of phylogenetic morphometrics. *Cladistics*, 32(3): 221–238.

Goloboff, P. A., Mattoni, C. I., and Quinteros, A. S. 2006. Continuous characters analyzed as such. *Cladistics*, 22(6): 589–601.

Goloboff, P. A., Torres, A., and Arias, J. S. 2017. Weighted parsimony outperforms other methods of phylogenetic inference under models appropriate for morphology. *Cladistics*.

Goloboff, P. A., Pittman, M., Pol, D., and Xu, X. 2018a. Morphological datasets fit a common mechanism much more poorly than dna sequences and call into question the mkv model. *Systematic Biology*.

Goloboff, P. A., Torres, A., and Arias, J. S. 2018b. Weighted parsimony outperforms other methods of phylogenetic inference under models appropriate for morphology. *Cladistics*, 34(4): 407–437.

González-José, R., Escapa, I., Neves, W. A., Cúneo, R., and Pucciarelli, H. M. 2008a. Cladistic analysis of continuous modularized traits provides phylogenetic signals in homo evolution. *Nature*, 453(7196): 775.

González-José, R., Escapa, I., Neves, W. A., Cúneo, R., and Pucciarelli, H. M. 2008b. Cladistic analysis of continuous modularized traits provides phylogenetic signals in homo evolution. *Nature*, 453(7196): 775.

Goswami, A. 2006. Cranial modularity shifts during mammalian evolution. *The American Naturalist*, 168(2): 270–280.

Goswami, A. and Finarelli, J. A. 2016. Emmli: A maximum likelihood approach to the analysis of modularity. *Evolution*, 70(7): 1622–1637.

Goswami, A., Weisbecker, V., and Sánchez-Villagra, M. 2009. Developmental modularity and the marsupial–placental dichotomy. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 312(3): 186–195.

Gould, S. 1980a. Is a new evolutionary theory emerging. *Paleobiology*, 6: 119–130.

Gould, S. J. 1970. Dollo on dollo's law: irreversibility and the status of evolutionary laws. *Journal of the History of Biology*, 3(2): 189–212.

Gould, S. J. 1977a. Bushes and ladders in human evolution. *Ever Since Darwin. Reflections in Natural History*, pages 56–62.

Gould, S. J. 1977b. *Ontogeny and Phylogeny*. Harvard University Press.

Gould, S. J. 1980b. The promise of paleobiology as a nomothetic, evolutionary discipline. *Paleobiology*, 6(1): 96–118.

Gould, S. J. and Eldredge, N. 1977. Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology*, 3(2): 115–151.

Guillerme, T. and Cooper, N. 2016. Effects of missing data on topological inference using a total evidence approach. *Molecular Phylogenetics and Evolution*, 94: 146–158.

Guindon, S. 2018. Accounting for calibration uncertainty: Bayesian molecular dating as a "doubly intractable" problem. *Systematic Biology*.

Guy, F., Lieberman, D. E., Pilbeam, D., de León, M. P., Likius, A., Mackaye, H. T., Vignaud, P., Zollikofer, C., and Brunet, M. 2005. Morphological affinities of the sahelanthropus tchadensis (late miocene hominid from chad) cranium. *Proceedings of the National Academy of Sciences*, 102(52): 18836–18841.

Hansen, T. F. 1997a. Stabilizing selection and the comparative analysis of adaptation. *Evolution*, pages 1341–1351.

Hansen, T. F. 1997b. Stabilizing selection and the comparative analysis of adaptation. *Evolution*, 51(5): 1341–1351.

Hansen, T. F. and Martins, E. P. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution*, pages 1404–1417.

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

Harmon, L. J., Losos, J. B., Jonathan Davies, T., Gillespie, R. G., Gittleman, J. L., Bryan Jennings, W., Kozak, K. H., McPeek, M. A., Moreno-Roark, F., Near, T. J., *et al.* 2010. Early bursts of body size and shape evolution are rare in comparative data. *Evolution: International Journal of Organic Evolution*, 64(8): 2385–2396.

Harvey, P. H. and Pagel, M. D. 1991. *The comparative method in evolutionary biology*, volume 239. Oxford University Press.

Hastings, W. K. 1970. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1): 97–109.

Hauser, D. L. and Presch, W. 1991. The effect of ordered characters on phylogenetic reconstruction. *Cladistics*, 7(3): 243–265.

Hawkins, J. A., Hughes, C. E., and Scotland, R. W. 1997. Primary homology assessment, characters and character states. *Cladistics*, 13(3): 275–283.

Heath, T. A., Huelsenbeck, J. P., and Stadler, T. 2014. The fossilized birth–death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences*, 111(29): E2957–E2966.

Hennig, C. 2015. What are the true clusters? *Pattern Recognition Letters*, 64: 53–62.

Hennig, W. 1965. Phylogenetic systematics. *Annual Review of Entomology*, 10(1): 97–116.

Hill, W. and Robertson, A. 1968. Linkage disequilibrium in finite populations. *Theor. Appl. Genet.*, 38(6): 226–231.

Hillis, D. and Huelsenbeck, J. 1992. Signal, noise, and reliability in molecular phylogenetic analyses. *J. Hered.*, 83(3): 189–195.

Ho, L. S. T. and Ané, C. 2014. Intrinsic inference difficulties for trait evolution with ornstein-uhlenbeck models. *Methods in Ecology and Evolution*, 5(11): 1133–1146.

Höhna, S., Landis, M. J., Heath, T. A., Boussau, B., Lartillot, N., Moore, B. R., Huelsenbeck, J. P., and Ronquist, F. 2016. Revbayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic Biology*, 65(4): 726–736.

Holloway, R. L. and Post, D. G. 1982. *The relativity of relative brain measures and hominid mosaic evolution*. Springer.

Hopkins, M. J. and Lidgard, S. 2012. Evolutionary mode routinely varies among morphological traits within fossil species lineages. *Proceedings of the National Academy of Sciences*, 109(50): 20520–20525.

Hubby, J. L. and Lewontin, R. C. 1966. A molecular approach to the study of genic heterozygosity in natural populations. i. the number of alleles at different loci in drosophila pseudoobscura. *Genetics*, 54(2): 577.

Hubert, L. and Arabie, P. 1985. Comparing partitions. *Journal of Classification*, 2(1): 193–218.

Huelsenbeck, J. P. 1991. When are fossils better than extant taxa in phylogenetic analysis? *Systematic Biology*, 40(4): 458–469.

Huelsenbeck, J. P. and Rannala, B. 1997. Maximum likelihood estimation of phylogeny using stratigraphic data. *Paleobiology*, 23(2): 174–180.

Huelsenbeck, J. P. and Suchard, M. A. 2007. A nonparametric method for accommodating and testing across-site rate variation. *Systematic Biology*, 56(6): 975–987.

Hunt, G. J., Guzmán-Novoa, E., Fondrk, M. K., and Page, R. E. 1998. Quantitative trait loci for honey bee stinging behavior and body size. *Genetics*, 148(3): 1203–1213.

Irish, J. D., Guatelli-Steinberg, D., Legge, S. S., de Ruiter, D. J., and Berger, L. R. 2013. Dental morphology and the phylogenetic "place" of australopithecus sediba. *Science*, 340(6129): 1233062.

Jones, E., Oliphant, T., Peterson, P., *et al.* 2016. Scipy: Open source scientific tools for python, 2001.

Jones, K. E., Smaers, J. B., and Goswami, A. 2015. Impact of the terrestrial-aquatic transition on disparity and rates of evolution in the carnivoran skull. *BMC evolutionary biology*, 15(1): 8.

Jukes, T. H. and Cantor, C. R. 1969. Evolution of protein molecules. *Mammalian Protein Metab.*, 3(21): 132.

Kimbel, W. H., Lockwood, C. A., Ward, C. V., Leakey, M. G., Rak, Y., and Johanson, D. C. 2006. Was australopithecus anamensis ancestral to a. afarensis? a case of anagenesis in the hominin fossil record. *Journal of Human Evolution*, 51(2): 134–152.

Kimura, M. *et al.* 1968. Evolutionary rate at the molecular level. *Nature*, 217(5129): 624–626.

Kirk, E. C., Cartmill, M., Kay, R. F., and Lemelin, P. 2003. Comment on" grasping primate origins". *Science*, 300(5620): 741–741.

Klingenberg, C. P. 2011. Morphoj: an integrated software package for geometric morphometrics. *Molecular ecology resources*, 11(2): 353–357.

Klopfstein, S., Vilhelmsen, L., and Ronquist, F. 2015. A nonstationary markov model detects directional evolution in hymenopteran morphology. *Systematic Biology*, 64(6): 1089–1103.

Kuhner, M. K. and Felsenstein, J. 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Mol. Biol. Evol.*, 11(3): 459–468.

Landis, M. J. and Schraiber, J. G. 2017. Pulsed evolution shaped modern vertebrate body sizes. *Proceedings of the National Academy of Sciences*, 114(50): 13224–13229.

Landis, M. J., Schraiber, J. G., and Liang, M. 2012. Phylogenetic analysis using lévy processes: finding jumps in the evolution of continuous traits. *Systematic biology*, 62(2): 193–204.

Landis, M. J., Schraiber, J. G., and Liang, M. 2013. Phylogenetic analysis using lévy processes: finding jumps in the evolution of continuous traits. *Systematic Biology*, 62(2): 193–204.

Laurin, M. 2014. Assessment of modularity in the urodele skull: an exploratory analysis using ossification sequence data. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 322(8): 567–585.

Leakey, M. G., Feibel, C. S., McDougall, I., and Walker, A. 1995. New four-million-year-old hominid species from kanapoi and allia bay, kenya. *Nature*, 376(6541): 565.

Lee, M. S. and Palci, A. 2015. Morphological phylogenetics in the genomic age. *Curr. Biol.*, 25(19): R922–R929.

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

Lee, M. S., Soubrier, J., and Edgecombe, G. D. 2013. Rates of phenotypic and genomic evolution during the cambrian explosion. *Curr. Biol.*, 23(19): 1889–1895.

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

Levinton, J. S. 2001. *Genetics, paleontology, and macroevolution*. Cambridge University Press.

Levinton, J. S. and Chris, M. S. 1980. A critique of the punctuated equilibria model and implications for the detection of speciation in the fossil record. *Systematic Biology*, 29(2): 130–142.

Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology*, 50(6): 913–925.

Lewis, P. O., Chen, M.-H., Kuo, L., Lewis, L. A., Fučíková, K., Neupane, S., Wang, Y.-B., and Shi, D. 2016. Estimating bayesian phylogenetic information content. *Systematic biology*, 65(6): 1009–1023.

Lewontin, R. C. 1967. An estimate of average heterozygosity in man. *American Journal of Human Genetics*, 19(5): 681.

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

Luo, A., Duchene, D., Zhang, C., Zhu, C.-D., and Ho, S. 2018. A simulation-based

evaluation of total-evidence dating under the fossilized birth-death process. *bioRxiv*, page 436303.

Lynch, M. 1990. The rate of morphological evolution in mammals from the standpoint of the neutral expectation. *The American Naturalist*, 136(6): 727–741.

Lynch, M., Walsh, B., *et al.* 1998. *Genetics and analysis of quantitative traits*, volume 1. Sinauer Sunderland, MA.

MacLatchy, L. M., Desilva, J., Sanders, W. J., and Wood, B. 2010. Hominini. In W. J. Sanders and L. Werdelin, editors, *Cenozoic Mammals of Africa*, pages 471–542. University of California Press, Oakland, CA.

MacLeod, N. 1991. Punctuated anagenesis and the importance of stratigraphy to paleobiology. *Paleobiology*, 17(2): 167–188.

MacLeod, N. 2001. The role of phylogeny in quantitative paleobiological data analysis. *Paleobiology*, 27(2): 226–240.

MacLeod, N. 2002. Phylogenetic signals in morphometric data. *Morphology, shape and phylogeny*, 100: 138.

Maddison, W. P. and Knowles, L. L. 2006. Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology*, 55(1): 21–30.

Maynard Smith, J. 1984. Palaeontology at the high table. *Nature*, 309(5967): 401–402.

Mayr, E. 1970. *Populations, species, and evolution: an abridgment of animal species and evolution*, volume 19. Harvard University Press.

McHenry, H. M. 1975. Fossils and the mosaic nature of human evolution. *Science*, 190(4213): 425–431.

McVean, G. 2007. The structure of linkage disequilibrium around a selective sweep. *Genetics*, 175(3): 1395–1406.

Mendes, F. K., Fuentes-Gonzalez, J. A., Schraiber, J. G., and Hahn, M. W. 2018. Evolutionary inferences about quantitative traits are affected by underlying genealogical discordance. *bioRxiv*, page 276642.

Mounier, A., Marchal, F., and Condemi, S. 2009. Is homo heidelbergensis a distinct species? new insight on the mauer mandible. *Journal of Human Evolution*, 56(3): 219–246.

Newell, N. D. 1949. Phyletic size increase, an important trend illustrated by fossil invertebrates. *Evolution*, pages 103–124.

Newell, N. D. 1952. Periodicity in invertebrate evolution. *Journal of Paleontology*, pages 371–385.

Nylander, J. A., Ronquist, F., Huelsenbeck, J. P., and Nieves-Aldrey, J. 2004. Bayesian phylogenetic analysis of combined data. *Systematic Biology*, 53(1): 47–67.

O'leary, M. A., Bloch, J. I., Flynn, J. J., Gaudin, T. J., Giallombardo, A., Giannini, N. P., Goldberg, S. L., Kraatz, B. P., Luo, Z.-X., Meng, J., *et al.* 2013. The placental mammal ancestor and the post–k-pg radiation of placentals. *Science*, 339(6120): 662–667.

Olson, E. C. and Miller, R. L. 1958. *Morphological Integration*. University of Chicago Press.

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

O'Reilly, J. E., Puttick, M. N., Parry, L., Tanner, A. R., Tarver, J. E., Fleming, J., Pisani, D., and Donoghue, P. C. 2016. Bayesian methods outperform parsimony but at the expense of precision in the estimation of phylogeny from discrete morphological data. *Biology Letters*, 12(4): 20160081.

Pagel, M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society B*, 255(1342): 37–45.

Palaisa, K., Morgante, M., Tingey, S., and Rafalski, A. 2004. Long-range patterns of diversity and linkage disequilibrium surrounding the maize y1 gene are indicative of an asymmetric selective sweep. *Proc. Natl. Acad. Sci. U. S. A.*, 101(26): 9885–9890.

Parins-Fukuchi, C. 2018a. Bayesian placement of fossils on phylogenies using quantitative morphometric data. *Evolution*, 72(9): 1801–1814.

Parins-Fukuchi, C. 2018b. Use of continuous traits can improve morphological phylogenetics. *Systematic Biology*, 67(2): 328–339.

Pattinson, D. J., Thompson, R. S., Piotrowski, A. K., and Asher, R. J. 2014. Phylogeny, paleontology, and primates: do incomplete fossils bias the tree of life? *Systematic Biology*, 64: 169–186.

Pease, J. B., Haak, D. C., Hahn, M. W., and Moyle, L. C. 2016. Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLoS Biology*, 14(2): e1002379.

Pennell, M. W. and Harmon, L. J. 2013. An integrative view of phylogenetic comparative methods: connections to population genetics, community ecology, and paleobiology. *Annals of the New York Academy of Sciences*, 1289(1): 90–105.

Pimentcl, R. A. and Riggins, R. 1987. The nature of cladistic data. *Cladistics*, 3(3): 201–209.

Pleijel, F. 1995. On character coding for phylogeny reconstruction. *Cladistics*, 11(3): 309–315.

Pomidor, B. J., Makedonska, J., and Slice, D. E. 2016. A landmark-free method for three-dimensional shape analysis. *PloS one*, 11(3): e0150368.

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

R Core Team 2016. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rabosky, D. L. and Adams, D. C. 2012. Rates of morphological evolution are correlated with species richness in salamanders. *Evolution: International Journal of Organic Evolution*, 66(6): 1807–1818.

Rabosky, D. L., Santini, F., Eastman, J., Smith, S. A., Sidlauskas, B., Chang, J., and Alfaro, M. E. 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nature communications*, 4: 1958.

Rae, T. C. 1998. The logical basis for the use of continuous characters in phylogenetic systematics. *Cladistics*, 14(3): 221–228.

Rambaut, A. and Drummond, A. 2013. Treeannotator v1. 7.0. *Available as part of the BEAST package at http://beast. bio. ed. ac. uk*.

Rannala, B., Zhu, T., and Yang, Z. 2011. Tail paradox, partial identifiability, and influential priors in bayesian branch length inference. *Molecular biology and evolution*, 29(1): 325–335.

Reich, D. E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P. C., Richter, D. J., Lavery, T., Kouyoumjian, R., Farhadian, S. F., Ward, R., *et al.* 2001. Linkage disequilibrium in the human genome. *Nature*, 411(6834): 199–204.

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

Revell, L. J. 2012b. phytools: an r package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.*, 3(2): 217–223.

Revell, L. J. and Harmon, L. J. 2008. Testing quantitative genetic hypotheses about the evolutionary rate matrix for continuous characters. *Evolutionary Ecology Research*, 10(3): 311–331.

Revell, L. J., Harmon, L. J., and Collar, D. C. 2008. Phylogenetic signal, evolutionary process, and rate. *Systematic Biology*, 57(4): 591–601.

Revell, L. J., Mahler, D. L., Reynolds, R. G., and Slater, G. J. 2015. Placing cryptic, recently extinct, or hypothesized taxa into an ultrametric phylogeny using continuous character data: A case study with the lizard anolis roosevelti. *Evolution*, 69(4): 1027–1035.

Rightmire, G. P. 1998. Human evolution in the middle pleistocene: the role of homo heidelbergensis. *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews*, 6(6): 218–227.

Robinson, D. F. and Foulds, L. R. 1981. Comparison of phylogenetic trees. *Math. Biosci.*, 53(1-2): 131–147.

Rohlf, F. J. 1998. On applications of geometric morphometrics to studies of ontogeny and phylogeny. *Systematic Biology*, 47(1): 147–158.

Rohlf, F. J. 2002. Geometric morphometrics and phylogeny. *Systematics Association Special Volume*, 64: 175–193.

Rohlf, F. J. and Marcus, L. F. 1993. A revolution in morphometrics. *Trends in Ecology and Evolution*, 8(4): 129–132.

Rohlfs, R. V., Harrigan, P., and Nielsen, R. 2013. Modeling gene expression evolution with an extended ornstein–uhlenbeck process accounting for within-species variation. *Molecular biology and evolution*, 31(1): 201–211.

Ronquist, F. and Huelsenbeck, J. P. 2003. Mrbayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19(12): 1572–1574.

Ronquist, F., Klopfstein, S., Vilhelmsen, L., Schulmeister, S., Murray, D. L., and Rasnitsyn, A. P. 2012. A total-evidence approach to dating with fossils, applied to the early radiation of the hymenoptera. *Systematic Biology*, 61(6): 973–999.

Ronquist, F., Lartillot, N., and Phillips, M. J. 2016. Closing the gap between rocks and clocks using total-evidence dating. *Phil. Trans. R. Soc. B*, 371(1699): 20150136.

Rosas, A. and De Castro, J. M. B. 1998. The mauer mandible and the evolutionary significance of homo heidelbergensis. *Geobios*, 31(5): 687–697.

Rose, C. S. 2003. The developmental morphology of salamander skulls. *Amphibian biology*, 5: 1684–1781.

Scales, J. A., King, A. A., and Butler, M. A. 2009. Running for your life or running for your dinner: what drives fiber-type evolution in lizard locomotor muscles? *The American Naturalist*, 173(5): 543–553.

Schlenke, T. A. and Begun, D. J. 2004. Strong selective sweep associated with a transposon insertion in drosophila simulans. *Proc. Natl. Acad. Sci. U. S. A.*, 101(6): 1626–1631.

Schraiber, J. G., Mostovoy, Y., Hsu, T. Y., and Brem, R. B. 2013. Inferring evolutionary histories of pathway regulation from transcriptional profiling data. *PLoS Computational Biology*, 9(10): e1003255.

Scotland, R. and Pennington, R. T. 2000. *Homology and systematics: coding characters for phylogenetic analysis*. CRC Press.

Scotland, R. W., Olmstead, R. G., and Bennett, J. R. 2003. Phylogeny reconstruction: the role of morphology. *Systematic Biology*, 52(4): 539–548.

Simões, T. R., Caldwell, M. W., Palci, A., and Nydam, R. L. 2017. Giant taxon-character matrices: quality of character constructions remains critical regardless of size. *Cladistics*, 33(2): 198–219.

Simpson, G. G. 1944. *Tempo and mode in evolution*. Columbia University Press.

Slater, G. J. 2013. Phylogenetic evidence for a shift in the mode of mammalian body size evolution at the cretaceous-palaeogene boundary. *Methods in Ecology and Evolution*, 4(8): 734–744.

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

Smith, A. B. 2000. Stratigraphy in phylogeny reconstruction. *Journal of Paleontology*, 74(5): 763–766.

Smith, S. A., Moore, M. J., Brown, J. W., and Yang, Y. 2015. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evolutionary Biology*, 15(1): 150.

Smith, S. A., Brown, J. W., Yang, Y., Bruenn, R., Drummond, C. P., Brockington, S. F.,

Walker, J. F., Last, N., Douglas, N. A., and Moore, M. J. 2018a. Disparity, diversity, and duplications in the caryophyllales. *New Phytologist*, 217(2): 836–854.

Smith, S. A., Brown, J. W., and Walker, J. F. 2018b. So many genes, so little time: A practical approach to divergence-time estimation in the genomic era. *PloS one*, 13(5): e0197433.

Smith, U. E. and Hendricks, J. R. 2013. Geometric morphometric character suites as phylogenetic data: extracting phylogenetic signal from gastropod shells. *Systematic Biology*, 62: 366–385.

Sneath, P. H. and Sokal, R. R. 1962. Numerical taxonomy. *Nature*, 193(4818): 855–860.

Solow, A. R. and Smith, W. 1997. On fossil preservation and the stratigraphic ranges of taxa. *Paleobiology*, 23(3): 271–277.

Soltis, D. E., Smith, S. A., Cellinese, N., Wurdack, K. J., Tank, D. C., Brockington, S. F., Refulio-Rodriguez, N. F., Walker, J. B., Moore, M. J., Carlsward, B. S., *et al.* 2011. Angiosperm phylogeny: 17 genes, 640 taxa. *American Journal of Botany*, 98(4): 704–730.

Soul, L. C. and Friedman, M. 2017. Bias in phylogenetic measurements of extinction and a case study of end-permian tetrapods. *Palaeontology*, 60(2): 169–185.

Stadler, T. 2010. Sampling-through-time in birth–death trees. *Journal of Theoretical Biology*, 267(3): 396–404.

Stamatakis, A. 2006. Raxml-vi-hpc: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, 22(21): 2688–2690.

Stamatakis, A. 2014. Raxml version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9): 1312–1313.

Stanley, S. M. 1979. *Macroevolution, pattern and process*. Johns Hopkins University Press.

Stebbins, G. 1983. Mosaic evolution: An integrating principle for the modern synthesis. *Experientia*, 39(8): 823–834.

Stebbins, G. L. 1984. Mosaic evolution, mosaic selection and angiosperm phylogeny. *Botanical journal of the Linnean Society*, 88(1-2): 149–164.

Strait, D. S. and Grine, F. E. 1999. Cladistics and early hominid phylogeny. *Science*, 285(5431): 1209–1209.

Strait, D. S., Grine, F. E., and Moniz, M. A. 1997. A reappraisal of early hominid phylogeny. *Journal of Human Evolution*, 32(1): 17–82.

Stringer, C. 2012. The status of homo heidelbergensis (schoetensack 1908). *Evolutionary Anthropology: Issues, News, and Reviews*, 21(3): 101–107.

Strotz, L. C. and Allen, A. P. 2013. Assessing the role of cladogenesis in macroevolution by integrating fossil and molecular evidence. *Proceedings of the National Academy of Sciences*, 110(8): 2904–2909.

Sukumaran, J. and Holder, M. T. 2010. Dendropy: a python library for phylogenetic computing. *Bioinformatics*, 26(12): 1569–1571.

Tavaré, S. 1986. Some probabilistic and statistical problems in the analysis of dna sequences. *Lectures on mathematics in the life sciences*, 17: 57–86.

Thiele, K. 1993. The holy grail of the perfect character: the cladistic treatment of morphometric data. *Cladistics*, 9(3): 275–304.

Thomas, G. H. and Freckleton, R. P. 2012. Motmot: models of trait macroevolution on trees. *Methods in Ecology and Evolution*, 3(1): 145–151.

Thompson, E. A. 1975. *Human Evolutionary Trees*. Cambridge University Press Archive.

Upchurch, P. 1995. The evolutionary history of sauropod dinosaurs. *Philosophical Transactions of the Royal Society B*, 349(1330): 365–390.

Uyeda, J. C., Caetano, D. S., and Pennell, M. W. 2015. Comparative analysis of principal components can be misleading. *Systematic Biology*, 64(4): 677–689.

Valdar, W., Solberg, L. C., Gauguier, D., Burnett, S., Klenerman, P., Cookson, W. O., Taylor, M. S., Rawlins, J. N. P., Mott, R., and Flint, J. 2006. Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nat. Genet.*, 38(8): 879–887.

Van Valen, L. 1973. A new evolutionary law. *Evolutionary Theory*, 1: 1–30.

Wagenmakers, E.-J. and Farrell, S. 2004. Aic model selection using akaike weights. *Psychonomic Bulletin & Review*, 11(1): 192–196.

Wagner, P. J. 1996. Contrasting the underlying patterns of active trends in morphologic evolution. *Evolution*, 50(3): 990–1007.

Wagner, P. J. 1998. A likelihood approach for evaluating estimates of phylogenetic relationships among fossil taxa. *Paleobiology*, 24(4): 430–449.

Wagner, P. J. 2012. Modelling rate distributions using character compatibility: implications for morphological evolution among fossil invertebrates. *Biology Letters*, 8(1): 143–146.

Ward, C. V., Leakey, M. G., and Walker, A. 2001. Morphology of australopithecus anamensis from kanapoi and allia bay, kenya. *Journal of human evolution*, 41(4): 255–368.

White, T. 2003. Early hominids–diversity or distortion? *Science*, 299(5615): 1994–1997.

Wickham, H. 2016. *ggplot2: elegant graphics for data analysis*. Springer.

Wiens, J. J. 2001. Character analysis in morphological phylogenetics: problems and solutions. *Systematic Biology*, 50(5): 689–699.

Wiens, J. J. 2005. Can incomplete taxa rescue phylogenetic analyses from long-branch attraction? *Systematic Biology*, 54(5): 731–742.

Wiens, J. J. 2009. Paleontology, genomics, and combined-data phylogenetics: can molecular data improve phylogeny estimation for fossil taxa? *Systematic Biology*, 58(1): 87–99.

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

Wilkinson, D. G., Bhatt, S., Cook, M., Boncinelli, E., and Krumlauf, R. 1989. Segmental expression of hox-2 homoeobox-containing genes in the developing mouse hindbrain. *Nature*, 341: 405–409.

Wilkinson, M. 1995. A comparison of two methods of character construction. *Cladistics*, 11(3): 297–308.

Wilson, J. A. and Sereno, P. C. 1998. Early evolution and higher-level phylogeny of sauropod dinosaurs. *Journal of Vertebrate Paleontology*, 18(S2): 1–79.

Wood, H. M., Matzke, N. J., Gillespie, R. G., and Griswold, C. E. 2012. Treating fossils as terminal taxa in divergence time estimation reveals ancient vicariance patterns in the palpimanoid spiders. *Systematic Biology*, 62(2): 264–282.

Wright, A. M. and Hillis, D. M. 2014. Bayesian analysis using a simple likelihood model

outperforms parsimony for estimation of phylogeny from discrete morphological data. *PLoS One*, 9(10): e109210.

Wright, A. M., Lloyd, G. T., and Hillis, D. M. 2015. Modeling character change heterogeneity in phylogenetic analyses of morphology through the use of priors. *Systematic Biology*, page syv122.

Wright, A. M., Lloyd, G. T., and Hillis, D. M. 2016. Modeling character change heterogeneity in phylogenetic analyses of morphology through the use of priors. *Systematic Biology*, 65(4): 602–611.

Yang, J.-R., Maclean, C. J., Park, C., Zhao, H., and Zhang, J. 2017. Intra and interspecific variations of gene expression levels in yeast are largely neutral:(nei lecture, smbe 2016, gold coast). *Molecular biology and evolution*, 34(9): 2125–2139.

Yang, Z. 1996. Among-site rate variation and its impact on phylogenetic analyses. *Trends in Ecology & Evolution*, 11(9): 367–372.

Yang, Z. 1998. On the best evolutionary rate for phylogenetic analysis. *Systematic Biology*, 47(1): 125–133.

Yang, Z. 2014. *Molecular evolution: a statistical approach*. Oxford University Press.

Yang, Z. and Zhu, T. 2018. Bayesian selection of misspecified models is overconfident and may cause spurious posterior probabilities for phylogenetic trees. *Proceedings of the National Academy of Sciences*, 115(8): 1854–1859.

Yang, Z., Kumar, S., and Nei, M. 1995. A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics*, 141(4): 1641–1650.

Zanne, A. E., Tank, D. C., Cornwell, W. K., Eastman, J. M., Smith, S. A., FitzJohn, R. G.,

McGlinn, D. J., O'Meara, B. C., Moles, A. T., Reich, P. B., *et al.* 2014. Three keys to the radiation of angiosperms into freezing environments. *Nature*, 506(7486): 89.

Zelditch, M. L., Fink, W. L., and Swiderski, D. L. 1995. Morphometrics, homology, and phylogenetics: quantified characters as synapomorphies. *Systematic Biology*, 44(2): 179–189.

Zhang, C., Stadler, T., Klopfstein, S., Heath, T. A., and Ronquist, F. 2016. Total-evidence dating under the fossilized birth–death process. *Systematic biology*, 65(2): 228–249.

Zollikofer, C. P., de León, M. S. P., Lieberman, D. E., Guy, F., Pilbeam, D., Likius, A., Mackaye, H. T., Vignaud, P., and Brunet, M. 2005. Virtual cranial reconstruction of sahelanthropus tchadensis. *Nature*, 434(7034): 755.

Zuckerkandl, E. and Pauling, L. 1965. Molecules as documents of evolutionary history. *Journal of Theoretical Biology*, 8(2): 357–366.