The Formation of Diversity - The Role of Environment and Biogeography in Dung Beetle Species Richness, and the Adequacy of Current Diversification Models:

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Orlando Manuel Schwery
December 2019

DEDICATION

I dedicate my work to my parents,

René and Carmen Schwery

for their continuous support throughout my life, and the sacrifice it involved from them, even when my path became more and more unrelatable to what they knew.

To my childhood mentor, and to my primary school teacher,

Paul "Böbi" Weber, and Ernst Giger

who sparked my love for nature and books, supported my interest to gain knowledge about the world around me, and supplied me with endless opportunities to do so.

And finally, to my high school teacher,

Bruno Koch

who taught me to have the hold myself to the highest standards, be critical of myself and others and abhor complacence, and to always strive to know and achieve more.

ACKNOWLEDGEMENTS

I owe many thanks to my advisor Brian O'Meara, for his steady support from the first inquiry email on, and throughout the following graduate school years. He always provided a good combination of criticism and encouragement, and went to great lengths to be available for questions and concerns of mine. Apart from sharing his knowledge and expertise to advise on my science, he also readily spent time discussing many other concerns, be it from adjusting to a new culture in the beginning, own and general graduate student concerns, to discussing private, academic, or global issues on my mind. This, together with his determination to advocate on behalf of me while holding me to high standards make him one of the most supportive advisors to have.

Furthermore, I want to thank my dissertation committee for their support and their critical thoughts throughout the process. More specifically: Jim Fordyce for his advice on sequencing and statistics; Kimberly Sheldon for inspiring me to pick dung beetles as my empirical study system and for taking me to the field to get acquainted with them; and Colin Sumrall for providing the paleontological perspective on diversification, and support in writing.

I thank the various members of the O'Meara lab, first and foremost Jeremy Beaulieu, for various inputs and advice along the way, and particularly for his help in letting me recognize the overarching story that unifies my dissertation. I thank my fellow grad students Sam Borstein, Katie Massana, and Jen Bosco, and Rosana Zenil-Ferguson for their comradery and shared insights along the path to graduation. Further, I thank Luna Sanchez-Reyes for various helpful comments on diversification models, phylogenetics resources, and for being a great conference travel companion; Nick Matzke for his advice on biogeography in general and BioGeoBEARS in particular; Dave Bapst and Michelle Lawing for their advice on paleontological problems; Sergei Tarasov for his help in the lab, and for sharing his knowledge and experience on dung beetle evolution; and finally Sandy Kawano, Nate Jackson, Dominic Evangelista, Tony Jhwueng, Megan Rua, and Jodie Wiggins, for stimulating discussions and support of all kinds.

I thank the members of the department of Ecology and Evolutionary Biology at UT for providing an environment that supported my progress, first and foremost all my grad school friends, especially Angela Chuang for her support and help in staying sane during the numerous challenges along the way to graduation. I further particularly thank Veronica Brown, Marisol Sanchez, Todd Pierson, Hailee Korotkin, Claire Winfrey, and Jess Dreyer for all their help and advice with sequencing and other molecular lab issues. I thank Elizabeth Johnson, Joel Corush, and Dan Simberloff their time discussing various ideas; Athmanathan Senthilnathan and Lucas Santana-Souza for helping me understand modified graph Laplacians; Jordan Bush for sharing her polygon code; Maggie Mamantov for being a field-companion and for sharing specimens; and Nina Fefferman for modeling insights and for taking the time to help me better understand academic careers and negotiations in the process of pursuing them.

For Chapter I, I additionally thank Frederik Matsen IV and two anonymous reviewers for well-considered criticism to improve the associated manuscript, Brian Looney and Sam Borstein for beta testing, and the members of the Tank lab, Arne Mooers, Karen Cranston, Bruce Cochrane and Daniel Gates for great ideas on increasing the usefulness of the MonoPhy R-package.

For Chapters II and III, I additionally want to thank Matt Pennell, Dean Anderson, and Emma Goldberg for helpful discussions and comments.

Finally, I thank the Society of Systematic Biologists, the Systematics Association, the department of Ecology and Evolutionary Biology at UT, their grad student association GREBE, and the Graduate Student Senate, for their financial support towards my projects and travel.

ABSTRACT

Model based approaches to study the driving factors behind diversification have become increasingly popular, but in the recent years, various weaknesses of these models have received increased attention. One way to ensure those issues do not affect one's inferences, is to test a model's adequacy as a way to judge its suitability to describe the data in an absolute sense. Here, I implement a simple adequacy test for diversification models in the R package BoskR, using metrics for tree shape. I demonstrate the method's ability to distinguish trees simulated under different models, and then use it to test the adequacy of a range of birth-death diversification models for a large set of empirical phylogenies. I find that while most models are adequate to describe a majority of the empirical trees, a few trees cannot be described by any of those models. Furthermore, the best fitting of a set of models may not always be adequate, highlighting the practical use of incorporating model adequacy tests in the standard procedures for diversification studies.

For the empirical parts of my dissertation, I investigate the diversification and biogeography of dung beetles. It has been hypothesized that their origin and distribution are either the result of Gondwanan vicariance, or out-of-Africa dispersal. Furthermore, dung beetle diversification is thought to have been affected by mammals – particularly large herds of herbivores inhabiting the vast grasslands after the Miocene – and potentially also by non-avian dinosaurs, if dinosaur dung-adapted beetles were affected by the K-Pg extinction of their dung producers. Crucial to answering these questions is to know whether dung beetles are of Mesozoic or Cenozoic origin. Thus, I construct a large dated phylogeny, and use model-based inference to estimate their ancestral area, and the influence of range evolution and diversity of dung producers on their diversification rates. My results suggest that dung beetles originated in Gondwana during the Mesozoic, but it remains unclear to which extent range evolution affected diversification. While adaptation to dinosaur dung and subsequent co-extinction are plausible, the available data cannot support a radiation with the rise of grasslands and herds of herbivores.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

An important part of the study of macroevolution is to understand the origin, accumulation and maintenance of biodiversity, and a compelling way to quantify those processes is though model-based diversification studies. Simple approaches estimate speciation and extinction rates using either the estimate according to Magallon and Sanderson (2001) which calculates net rates based on time and species richness; or using the Kendall-Moran estimate (Nee 2001) which relies on the sum of all branch lengths. However, the majority of recent work employs model-based approaches, simple examples of which just estimate constant speciation rates (λ) and extinction rates (μ) for a given phylogeny under a birth-death (λ & μ) or Yule (λ only) model. Extensions of this include time dependence or diversity dependence, or other covariates, with continuous environmental data over time being a more recent example (Condamine, Rolland & Morlon 2013). One line of work primarily concerned with identifying intrinsic and extrinsic factors affecting diversification and has produced methods that specifically model diversification rates in dependence of traits – the so-called 'SSE' models. They were implemented for binary traits (Maddison, Midford & Otto 2007; FitzJohn, Maddison & Otto 2009), multi-state traits (FitzJohn 2012), continuous traits (FitzJohn 2010), or geographical areas (Goldberg, Lancaster & Ree 2011), each allowing testing for differences in diversification rates between lineages associated with a particular trait state. A different line of work focused on detecting changes in diversification rates, regardless of the cause. Changes are either detected across the whole tree, happening at particular times (Stadler 2011), or along particular branches (Alfaro *et al.* 2009; Rabosky 2014).

While many of those approaches have enjoyed high popularity and application over the years, various caveats regarding the proper use of those models, and concerns about the shortcomings of the approaches themselves have received increased attention. These include conceptual concerns like the lack of replication (Maddison & FitzJohn 2015), the prevalence of false positive results (Rabosky & Goldberg 2015), or the reason for mismatches between the properties of empirical phylogenies and those derived from classic null models (Stadler, Degnan & Rosenberg 2016). Most recently, Louca and Pennell (2019) have proven that any given extant phylogeny can be explained by an infinite set of indistinguishable diversification scenarios. As a response to some of these issues, new approaches have been developed that would either alleviate the problems or allow them to be addressed (Beaulieu & O'Meara 2016; Rabosky & Goldberg 2017; Caetano, O'Meara & Beaulieu 2018). Explicit testing for model adequacy has been introduced in other phylogenetic methods, such as trait evolution (Pennell et al. 2015), molecular clock models (Duchêne *et al.* 2015), or substitution models (Bollback 2002; Brown & ElDabaje 2008; Brown 2014).

As part of this dissertation, I add to those efforts by designing and implementing an adequacy test for simple birth-death based diversification models. The test is based on a simulation approach using metrics for tree shape as test statistics. While a limited number of simple models is implemented, any model under which trees can be simulated can technically be tested this way. I further test the method on a large set of published empirical phylogenies in order to gain insights into how well those currently available models perform on real-world

examples, and to explore how well the test can be used to identify the reason behind model inadequacy.

For the empirical portion of my dissertation, I attempted to contribute to ongoing inquiries into the diversification and biogeography of dung beetles. This remarkable insect group, comprised of around 5,300 species with near-global distribution, exhibits the unusual life history trait of using other organisms' dung for food and reproduction (Hanski & Cambefort 1991). While not the most diverse group of beetles, their use of dung has a profound impact on their environment, making them one of the most important insect groups, both ecologically (Nichols *et al.* 2008) and economically (Losey & Vaughan 2006). There is ongoing debate on two major macroevolutionary questions surrounding dung beetles: their geographical origin and subsequent colonization of much of the world, and the factors that contributed to their diversification, particularly the question of whether it was driven by their dung producers. The answer to both questions depends on an accurate and reliable phylogeny (Tarasov & Génier 2015), and particularly on an age estimate for the whole group, as the main hypotheses all involve a major difference in time. Regarding their origin, it has been hypothesized that they either are of Gondwanan origin and that the major lineages primarily diversified through vicariance after the continents' breakup (Davis, Scholtz & Philips 2002), or that they originated in Africa and mainly diversified after they settled the rest of the world by long-distance dispersal (Sole & Scholtz 2010). As for the influence of dung producers on beetle diversification, a main dispute in the field revolves around whether dung beetles fed on non-avian dinosaur dung (and if so, potentially suffered diversity declines during the end-Cretaceous extinction) (Scholtz, Davis & Kryger 2009); or whether their most important resource was mammalian dung, and their increase in lineage diversity tracked that of mammals (Davis, Scholtz & Philips 2002). The answer to both questions depends partially on whether the origination of dung beetles lies in the Mesozoic (Davis, Scholtz & Philips 2002) or Cenozoic (Monaghan et al. 2007), as the latter would preclude both the option of a (post-Pangea breakup) Gondwanan origin and feeding on dinosaur dung.

I am thus providing a new phylogeny with expanded taxonomic sampling, and divergence time estimates using calibrations assuming different group ages. Furthermore, I am using recent model-based approaches to test hypotheses on their geographical origin and the influences on their diversification. Dung beetles are potentially a great model system for diversification, based on their taxonomic, morphological and ecological diversity, and the considerable amount of research that has already been performed on various aspects of them (Hanski & Cambefort 1991; Spector 2006; Scholtz, Davis & Kryger 2009; Moczek 2011).

References

Alfaro, M.E., Santini, F., Brock, C., Alamillo, H., Dornburg, A., Rabosky, D.L., Carnevale, G. & Harmon, L.J. (2009) Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences of the United States of America,* **106,** 13410-13414.

Beaulieu, J.M. & O'Meara, B.C. (2016) Detecting Hidden Diversification Shifts in Models of Trait-Dependent Speciation and Extinction. *Systematic Biology,* **65,** 583-601.

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

Brown, J.M. (2014) Detection of implausible phylogenetic inferences using posterior predictive assessment of model fit. *Systematic Biology,* **63,** 334-348.

Brown, J.M. & ElDabaje, R. (2008) PuMA: Bayesian analysis of p artitioned (and u npartitioned) m odel a dequacy. *Bioinformatics,* **25,** 537-538.

Caetano, D.S., O'Meara, B.C. & Beaulieu, J.M. (2018) Hidden state models improve state-dependent diversification approaches, including biogeographical models. *Evolution,* **72,** 2308-2324.

Condamine, F.L., Rolland, J. & Morlon, H. (2013) Macroevolutionary perspectives to environmental change. *Ecology Letters,* **16,** 72-85.

Davis, A.L.V., Scholtz, C.H. & Philips, T.K. (2002) Historical biogeography of scarabaeine dung beetles. *Journal of Biogeography,* **29,** 1217-1256.

Duchêne, D.A., Duchêne, S., Holmes, E.C. & Ho, S.Y. (2015) Evaluating the adequacy of molecular clock models using posterior predictive simulations. *Molecular Biology and Evolution,* **32,** 2986-2995.

FitzJohn, R.G. (2010) Quantitative Traits and Diversification. *Systematic Biology,* **59,** 619-633.

FitzJohn, R.G. (2012) Diversitree: comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution,* **3,** 1084-1092.

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

Goldberg, E.E., Lancaster, L.T. & Ree, R.H. (2011) Phylogenetic Inference of Reciprocal Effects between Geographic Range Evolution and Diversification. *Systematic Biology,* **60,** 451-465.

Hanski, I. & Cambefort, Y. (1991) Dung beetle ecology. *Dung beetle ecology***.,** i-xii, 1-481.

Losey, J.E. & Vaughan, M. (2006) The economic value of ecological services provided by insects. *Bioscience,* **56,** 311-323.

Louca, S. & Pennell, M.W. (2019) Phylogenies of extant species are consistent with an infinite array of diversification histories. *BioRxiv***,** 719435.

Maddison, W.P. & FitzJohn, R.G. (2015) The Unsolved Challenge to Phylogenetic Correlation Tests for Categorical Characters. *Systematic Biology,* **64,** 127-136.

Maddison, W.P., Midford, P.E. & Otto, S.P. (2007) Estimating a binary character's effect on speciation and extinction. *Systematic Biology,* **56,** 701-710.

Magallon, S. & Sanderson, M.J. (2001) Absolute diversification rates in angiosperm clades. *Evolution,* **55,** 1762-1780.

Moczek, A. (2011) Evolution and development: onthophagus beetles and the evolutionary developmental genetics of innovation, allometry and plasticity. *Ecology and evolution of dung beetles***,** 126-151.

Monaghan, M.T., Inward, D.J.G., Hunt, T. & Vogler, A.P. (2007) A molecular phylogenetic analysis of the Scarabaeinae (dung beetles). *Molecular Phylogenetics and Evolution,* **45,** 674-692.

Nee, S. (2001) Inferring speciation rates from phylogenies. *Evolution,* **55,** 661-668.

Nichols, E., Spector, S., Louzada, J., Larsen, T., Amequita, S., Favila, M.E. & Scarabaeinae Res, N. (2008) Ecological functions and ecosystem services provided by Scarabaeinae dung beetles. *Biological Conservation,* **141,** 1461-1474.

Pennell, M.W., FitzJohn, R.G., Cornwell, W.K. & Harmon, L.J. (2015) Model Adequacy and the Macroevolution of Angiosperm Functional Traits. *American Naturalist,* **186,** E33-E50.

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

Rabosky, D.L. & Goldberg, E.E. (2015) Model Inadequacy and Mistaken Inferences of Trait-Dependent Speciation. *Systematic Biology,* **64,** 340-355.

Rabosky, D.L. & Goldberg, E.E. (2017) FiSSE: A simple nonparametric test for the effects of a binary character on lineage diversification rates. *Evolution,* **71,** 1432-1442.

Scholtz, C.H., Davis, A.L.V. & Kryger, U. (2009) *Evolutionary biology and conservation of dung beetles*. Pensoft Sofia.

Sole, C.L. & Scholtz, C.H. (2010) Did dung beetles arise in Africa? A phylogenetic hypothesis based on five gene regions. *Molecular Phylogenetics and Evolution,* **56,** 631-641.

Spector, S. (2006) Scarabaeine dung beetles (Coleoptera: Scarabaeidae: Scarabaeinae): an invertebrate focal taxon for biodiversity research and conservation. *The Coleopterists Bulletin,* **60,** 71-83.

Stadler, T. (2011) Mammalian phylogeny reveals recent diversification rate shifts. *Proceedings of the National Academy of Sciences,* **108,** 6187-6192.

Stadler, T., Degnan, J.H. & Rosenberg, N.A. (2016) Does Gene Tree Discordance Explain the Mismatch between Macroevolutionary Models and Empirical Patterns of Tree Shape and Branching Times? *Systematic Biology,* **65,** 628-639.

Tarasov, S. & Génier, F. (2015) Innovative Bayesian and parsimony phylogeny of dung beetles (Coleoptera, Scarabaeidae, Scarabaeinae) enhanced by ontology-based partitioning of morphological characters. *Plos One,* **10,** e0116671.

CHAPTER I
MONOPHY: A SIMPLE PACKAGE TO FIND AND VISUALIZE MONOPHYLY ISSUES

A version of this chapter was originally published by Orlando Schwery and Brian C. O'Meara. Orlando Schwery's contributions to this paper include: (i) conceiving the idea, (ii) writing and testing the code for the package, (iii) preparing the example cases, figures and tables, (iv) writing the paper. Brian C. O'Meara's contributions include: (i) advice and discussions throughout, (ii) writing the paper.

Schwery, O., O'Meara, B.C. "MonoPhy: A simple R package to find and visualize monophyly issues." *PeerJ Computer Science* 2:e56. (2016) doi:10.7717/peerj-cs.56.

Abstract

The monophyly of taxa is an important attribute of a phylogenetic tree. A lack of it may hint at shortcomings of either the tree or the current taxonomy or can indicate cases of incomplete lineage sorting or horizontal gene transfer. Whichever is the reason, a lack of monophyly can misguide subsequent analyses. While monophyly is conceptually simple, it is manually tedious and time consuming to assess on modern phylogenies of hundreds to thousands of species. The R package MonoPhy allows assessment and exploration of monophyly of taxa in a phylogeny. It can assess the monophyly of genera using the phylogeny only, and with an additional input file any other desired higher order taxa or unranked groups can be checked as well. Summary tables, easily subsettable results and several visualization options allow quick and convenient exploration of monophyly issues, thus making MonoPhy a valuable tool for any researcher working with phylogenies.

Introduction

Phylogenetic trees are undoubtedly crucial for most research in ecology or evolutionary biology. Whether one is studying trait evolution (e.g. Coddington 1988; Donoghue 1989), diversification (e.g. Gilinsky & Good 1991; Hey 1992), phylogeography (Avise *et al.* 1987), or simply relatedness within a group (e.g. Shochat & Dessauer 1981; Sibley & Ahlquist 1981; Czelusniak *et al.* 1982), bifurcating trees representing hierarchically nested relationships are central to the analysis. Exactly because phylogenies are so fundamental to the inferences we make, we need tools that enable us to examine how reconstructed relationships compare with existing assumptions, particularly taxonomy. We have computational approaches to estimate confidence for parts of a phylogeny (Felsenstein 1985; Larget & Simon 1999) or measuring distance between two phylogenies (Robinson 1971), but assessing agreement of a new phylogeny with existing taxonomy is often done manually. This does not scale to modern phylogenies of hundreds to thousands of taxa. Modern taxonomy seeks to name clades: an ancestor and all of its descendants (the descendants thus form a monophyletic group). Discrepancies between the new phylogenetic hypothesis and the current taxonomic classification may indicate that the phylogeny is wrong or poorly resolved. Alternatively, a well-supported phylogeny that conflicts with currently recognized groups might suggest that the taxonomy should be reformed. To identify such discrepancies, one can simply assess whether the established taxa are monophyletic. A lack of group monophyly however, can also be an indicator for conflict

between gene trees and the species tree, which may be a result of incomplete lineage sorting or horizontal gene transfer. In any case, monophyly issues in a phylogeny suggest a potential error that can affect downstream analysis and inference. For example, it will mislead ancestral trait or area reconstruction or introduce false signals when assigning unsampled diversity for diversification analyses (e.g. in *diversitree* (FitzJohn 2012) or BAMM (Rabosky 2014)). In general, a lack of monophyly can blur patterns we might see in the data otherwise.
As this problem is by no means new, approaches to solve it have been developed earlier, particularly for large scale sequencing projects in bacteria and archaea, for which taxonomic issues are notoriously challenging. The program GRUNT (Dalevi *et al.* 2007) uses a tip to root walk approach to group, regroup, and name clades according to certain user defined criteria. The subsequently developed 'taxonomy to tree' approach (McDonald *et al.* 2012) matches existing taxonomic levels onto newly generated trees, allowing classification of unidentified sequences and proposal of changes to the taxonomic nomenclature based on tree topology. Finally, Matsen and Gallagher (2012) have developed algorithms that find mismatches between taxonomy and phylogeny using a convex subcoloring approach.
The new tool presented here, the R package *MonoPhy*, is a quick and user-friendly method for assessing monophyly of taxa in a given phylogeny. While the R package *ape* (Paradis, Claude & Strimmer 2004) already contains the helpful function is.monophyletic, which also enables testing for monophyly, the functionality of *MonoPhy* is much broader. Apart from assessing monophyly for all groups and focal taxonomic levels in a tree at once, *MonoPhy* is also not limited to providing a simple 'yes-or-no' output, but rather enables the user to explore underlying causes of non-monophyly. In the following, we outline the structure and usage of the package and provide examples to demonstrate its functionality. For a more usage-focused and application-oriented treatment, one should refer to the tutorial vignette (vignette("MonoPhyVignette")), which contains stepwise instructions for the different functions and their options. For any other package details consult the documentation (help("MonoPhy")).

Description

The package *MonoPhy* is written in R (R Development Core Team 2014, http://www.R-project.org/), an increasingly important language for evolutionary biology. It builds on the existing packages *ape* (Paradis, Claude & Strimmer 2004), *phytools* (Revell 2012), *phangorn* (Schliep 2011), *RColorBrewer* (Neuwirth 2014) and *taxize* (Chamberlain & Szocs 2013). A list of the currently implemented commands is given in Table I-1. Note that in the code and this paper, we distinguish between tips, the organisms at the tip of the tree, and higher order taxa. Functions with 'taxa' only return information about higher order taxa, not tips. The main function – AssessMonophyly – evaluates the monophyly of each higher order taxon by identifying the most recent common ancestor (MRCA) of a collection of tips (e.g. all species in a genus), and then returning all descendants of this node. The taxon is monophyletic if the number of its members (tips) equals the number of descendants of its MRCA. If there are more descendants than taxon members, the function will identify and list the tips that do not belong to the focal taxon and we then call these tips 'intruders'. Accordingly, we will further refer to the taxa whose monophyly was disrupted by these 'intruders' as 'intruded'. Note that if two taxa are

reciprocally disrupting each other's monophyly, certain tips of intruded taxa will often be intruders themselves: if the phylogeny is ((A1,B1),(A2,B2)), where A and B are genera, it's not clear if the A tips are intruding in B or the B tips are intruding in A.
Biologically, identifying a few intruders may suggest that the definition of a group should be expanded; observing some group members in very different parts of the tree than the rest of their taxon may instead suggest that these individuals were misidentified, that their placement is the result of contaminated sequences or due to horizontal gene transfer between members of two remote clades. Moreover, the approach as described above would suggest that the clades that are intruded by the outlier tips would in turn be intruders to the taxon the outliers belong to, which intuitively would not make sense. We thus implemented an option to specify a cutoff value, which defines the minimal proportion of tips among the descendants of a taxon's MRCA that are labeled as being actual members of that taxon. If a given group falls below this value, the function will find the 'core clade' (a subclade for which the proportion matches or exceeds the cutoff value) by moving tipward, always following the descendant node with the greater number of tips in the focal taxon (absolute, relative if tied), and at each step evaluating the subtree rooted at that node to see if it exceeds the cutoff value. Once such a subtree is found, it is then called the 'core clade', and taxon members outside this clade are then called 'outliers'. As there is no objective criterion to decide at what point individuals should be considered outliers, a reasonable cutoff value must be chosen by the user.
If the tree's tip labels are in the format '*Genus_speciesepithet*', the genus names will be extracted and used as taxon assignments for the tips. If the tip labels are in another format, or other taxonomic levels should be tested, taxon names can be assigned to the tips using an input file. To avoid having to manually compose a taxonomy file for a taxon-rich phylogeny, *MonoPhy* can automatically download desired taxonomic levels from ITIS or NCBI using *taxize* (Chamberlain & Szocs 2013).
All inference results are stored in a solution object, from which the other functions can extract information (e.g. summary tables, intruder and outlier lists) for one or more higher-level taxa of interest. PlotMonophyly reconstructs and plots the monophyly state of the tips using *phytools* (Revell 2012). Apart from the basic monophyly plot (Figure I-1), branches can be coloured according to taxonomic groups or to highlight intruders and outliers. Monophyletic groups can be collapsed, and plots can be saved directly to PDF to facilitate the visualization of large trees.
It is important to remember that the results produced by the package are merely the product of the used phylogeny and the available taxonomic information. It thus only makes the mismatches between those accessible but does not reveal any more than that. The decision of whether the result suggests problems in the phylogeny or the taxonomy, whether a tip should be considered a rogue taxon and be removed or whether gene tree – species tree conflicts should be investigated, is entirely up to the user's judgment.
MonoPhy is available through CRAN (https://cran.r-project.org/package=MonoPhy/) and is developed on GitHub (https://github.com/oschwery/MonoPhy). Intended extensions and fixes can be seen in the issues list of the package's GitHub page. Among the planned extensions of the package are: multiple trees, displaying the result for specific subtrees, proposing monophyletic subgroups, enabling formal tests for monophyly (incorporating clade support) and providing increased plot customizability.

Examples

Our first example makes use of the example files contained in the package. They come from a phylogeny of the plant family Ericaceae (Schwery *et al.* (2015) pruned to 77 species; original data see Schwery *et al.* (2014)) and two taxon files assigning tribes and subfamilies to the tips (in both files, errors have been introduced for demonstration purposes; see code and output for both examples in Supplementary Data). Running the main analysis command AssessMonophyly on genus level (i.e. tree only) and tribe level (i.e. tree plus taxonomy file) using standard settings took 0.045 and 0.093 seconds respectively on a MacBook Pro with 2.4 GHz Intel Core i5 and 8GB Ram. We could now use the remaining commands to extract the information of interest from the saved output object (e.g. summary tables, lists of problem taxa, etc.). The basic monophyly plot for the genus level analysis is displayed for a subclade of the tree in Figure I-1. For the second example, we demonstrate the package's performance on a tree of 31,749 species of Embriophyta (data see Zanne *et al.* 2013; Zanne *et al.* 2014), using an outlier-cutoff of 0.9 this time. Just checking monophyly for genera took 1.78 hours but revealed that 22% of genera on the tree are not monophyletic, while around half of all genera are only represented by one species each. Furthermore, we can see that the largest monophyletic genus is *Iris* (139 tips), that *Justicia* had the most intruders (13 tips) and that *Acacia* produced the most outliers (99 tips). Finally, with 2337 other tips as descendants of their MRCA, the 3 species of *Aldina* are most spread throughout the tree.

Citation

Researchers using *MonoPhy* in a published paper should cite this article and indicate the used version of the package. The citation information for the current package version can be obtained using citation("MonoPhy").

References

Avise, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reeb, C.A. & Saunders, N.C. (1987) Intraspecific Phylogeography - The Mitochondrial-DNA Bridge Between Population-Genetics and Systematics. *Annual Review of Ecology and Systematics,* **18,** 489-522.

Chamberlain, S.A. & Szocs, E. (2013) taxize: taxonomic search and retrieval in R. *F1000Research,* **2,** 191-191.

Coddington, J.A. (1988) Cladistic Tests Of Adaptational Hypotheses. *Cladistics-the International Journal of the Willi Hennig Society,* **4,** 3-22.

Czelusniak, J., Goodman, M., Hewettemmett, D., Weiss, M.L., Venta, P.J. & Tashian, R.E. (1982) Phylogenetic Origins and Adaptive Evolution of Avian and Mammalian Hemoglobin Genes. *Nature,* **298,** 297-300.

Dalevi, D., DeSantis, T.Z., Fredslund, J., Andersen, G.L., Markowitz, V.M. & Hugenholtz, P. (2007) Automated group assignment in large phylogenetic trees using GRUNT: GRouping, ungrouping, naming tool. *Bmc Bioinformatics,* **8**.

Donoghue, M.J. (1989) Phylogenies and the Analysis of Evolutionary Sequences, with Examples from Seed Plants. *Evolution,* **43,** 1137-1156.

Felsenstein, J. (1985) Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution,* **39,** 783-791.

FitzJohn, R.G. (2012) Diversitree: comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution,* **3,** 1084-1092.

Gilinsky, N.L. & Good, I.J. (1991) Probabilities of Origination, Persistence, and Extinction of Families of Marine Invertebrate Life. *Paleobiology,* **17,** 145-166.

Hey, J. (1992) Using Phylogenetic Trees to Study Speciation and Extinction. *Evolution,* **46,** 627-640.

Larget, B. & Simon, D.L. (1999) Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Molecular Biology and Evolution,* **16,** 750-759.

Matsen, F.A. & Gallagher, A. (2012) Reconciling taxonomy and phylogenetic inference: formalism and algorithms for describing discord and inferring taxonomic roots. *Algorithms for Molecular Biology,* **7**.

McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A., Andersen, G.L., Knight, R. & Hugenholtz, P. (2012) An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *Isme Journal,* **6,** 610-618.

Neuwirth, E. (2014) RColorBrewer: ColorBrewer Palettes.

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

R Development Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

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

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

Robinson, D.F. (1971) Comparison of labeled trees with valency three. *Journal of Combinatorial Theory, Series B*, **11,** 105-119.
Schliep, K.P. (2011) phangorn: phylogenetic analysis in R. *Bioinformatics*, **27,** 592-593.
Schwery, O., Onstein, R.E., Bouchenak-Khelladi, Y., Xing, Y., Carter, R.J. & Linder, H.P. (2014) Data from: As old as the mountains: the radiations of the Ericaceae. Dryad Data Repository.
Schwery, O., Onstein, R.E., Bouchenak-Khelladi, Y., Xing, Y., Carter, R.J. & Linder, H.P. (2015) As old as the mountains: the radiations of the Ericaceae. *New Phytologist*, **207,** 355-367.
Shochat, D. & Dessauer, H.C. (1981) Comparative Immunological Study of Albumins of Anolis Lizards of the Caribbean Islands. *Comparative Biochemistry and Physiology a-Physiology*, **68,** 67-73.
Sibley, C.G. & Ahlquist, J.E. (1981) The phylogeny and relationships of the ratite birds as indicated by DNA-DNA hybridization. *Evolution today. Proceedings of the second International Congress of Systematic and Evolutionary Biology*. (eds G.G.E. Scudder & J.L. Reveal), pp. 301-335.
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., Royer, D.L., Soltis, D.E., Stevens, P.F., Westoby, M., Wright, I.J., Aarssen, L., Bertin, R.I., Calaminus, A., Govaerts, R., Hemmings, F., Leishman, M.R., Oleksyn, J., Soltis, P.S., Swenson, N.G., Warman, L. & Beaulieu, J.M. (2014) Three keys to the radiation of angiosperms into freezing environments. *Nature*, **506,** 89-+.
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., Royer, D.L., Soltis, D.E., Stevens, P.F., Westoby, M., Wright, I.J., Aarssen, L., Bertin, R.I., Calaminus, A., Govaerts, R., Hemmings, F., Leishman, M.R., Oleksyn, J., Soltis, P.S., Swenson, N.G., Warman, L., Beaulieu, J.M. & Ordonez, A. (2013) Data from: Three keys to the radiation of angiosperms into freezing environments. Dryad Data Repository.

Appendix A

Table I-1: Functions of the Package MonoPhy.

Function name	Description
AssessMonophyly	Runs the main analysis to assess monophyly of groups on a tree
GetAncNodes	Returns MRCA nodes for taxa.
GetIntruderTaxa	Returns lists of taxa that cause monophyly issues for another taxon.
GetIntruderTips	Returns lists of tips that cause monophyly issues for a taxon.
GetOutlierTaxa	Returns lists of taxa that have monophyly issues due to outliers.
GetOutlierTips	Returns lists of tips that cause monophyly issues for their taxon by being outliers.
GetResultMonophyly	Returns an extended table of the results
GetSummaryMonophyly	Returns a summary table of the results
PlotMonophyly	Allows several visualizations of the result.

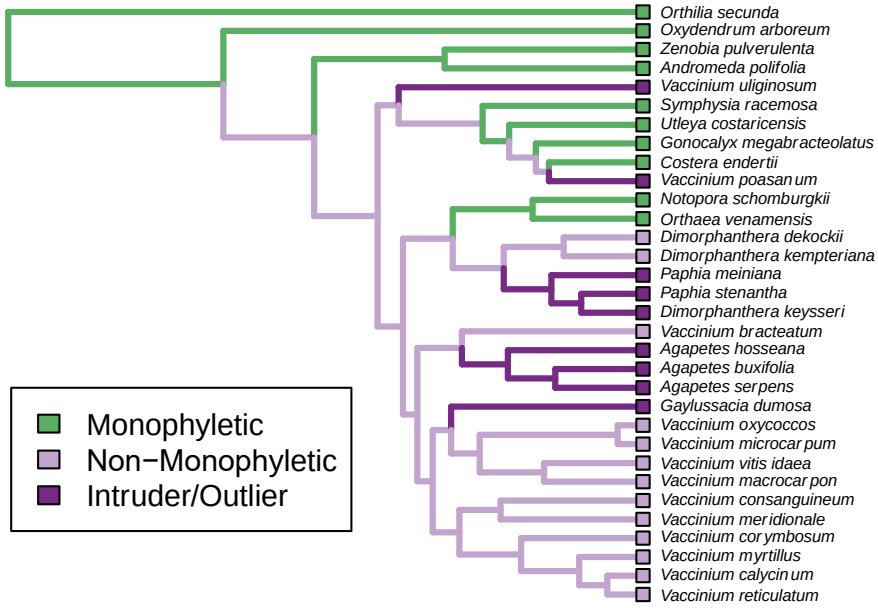

Figure I-1: Monophyly of the genera of Ericaceae.

Close-up on subfamily Vaccinioideae only. Branches of the tree colored according to monophyly status. We can see that *Vaccinium* has two outliers and that its intruders are *Paphia*, *Dimorphanthera*, *Agapetes* and *Gaylussacia*.

CHAPTER II
BOSKR – TESTING ADEQUACY OF DIVERSIFICATION MODELS USING TREE SHAPE

Abstract

The study of diversification largely relies on model-based approaches, estimating rates of speciation and extinction from phylogenetic trees. While a plethora of different models exist – all with different features, strengths and weaknesses – there is increasing concern about the reliability of the inference we gain from them. Apart from simply finding the model with the best fit for the data, we should find ways to assess a model's suitability to describe the data in an absolute sense. The R package BoskR implements a simple way of judging a model's adequacy for a given phylogeny using metrics for tree shape, assuming that a model is inadequate for a phylogeny if it produces trees that are consistently dissimilar in shape from the tree that should be analyzed. Tree shape is assessed via metrics derived from the tree's modified graph Laplacian spectrum, as provided by RPANDA. I exemplify the use of the method using simulated and empirical example phylogenies. BoskR was mostly able to correctly distinguish trees simulated under clearly different models and revealed that not all models are adequate for the empirical example trees. I believe the metrics of tree shape to be an intuitive and relevant means of assessing diversification model adequacy. Furthermore, by implementing the approach in an openly available R package, I enable and encourage researchers to adopt adequacy testing into their workflow.

Introduction

In the study of diversification, model-based estimation of diversification rates is probably the most common approach. Depending on the question, researchers will analyze diversification of their study group under different models and use some indicator for model fit (*e.g.* likelihood or Akaike Information Criterion (Akaike 1973)) to decide which model describes their data best and thus supports one or the other hypothesis. However, as those are only measures of relative fit, the best fitting model might still not actually be a good one to describe the data, hence the need for a way to assess its absolute fit. While we cannot assess a model's actual absolute fit, since we are usually lacking knowledge of the absolute truth behind our empirical data, we can assess whether a model is adequate for certain data, by exploring whether it can describe key properties of that data. This can be done by checking whether data generated under that model is similar to the data to be tested with regards to said key properties.

Testing model adequacy this way has been attempted before for different aspects of building phylogenies: Sullivan and Swofford (1997) demonstrated that using an inadequate substitution model can lead to false conclusions about phylogenetic relationships; Huelsenbeck *et al.* (2001) and Bollback (2002) used Bayesian posterior predictive simulations to test the adequacy of substitution models, with the frequency of site patterns as the one test statistic compared, and Foster (2004) used it to specifically test the adequacy of substitution models with regards to compositional heterogeneity of the nucleotide frequencies across the tree; Brown and ElDabaje (2008) extended their adequacy test to include partitioning (vs. not), and Rodrigue *et al.* (2009) included site-dependence between codons; Brown (2014) took the approach a step further, generating trees from the posterior predictive sequences, relying on inference-based comparison instead of data-based comparison, and also demonstrating that his inferential test statistics outperform the multinomial one used by Bollback (2002); and finally Lewis *et al.* (2013)

implement two new approaches that allow to dissect the overall verdict on substitution model adequacy.

Moving away from the substitution models, Duchêne *et al.* (2015) used posterior predictive simulations to test the adequacy of molecular clock models, also demonstrating the lower performance of a multinomial test statistic, and on a post-tree-inference state of phylogenetic analysis, Pennell *et al.* (2015) have introduced adequacy testing for models of continuous trait evolution, and explored how it evaluates common trait evolution models across many empirical data sets.

With focus on diversification, Revell, Harmon and Glor (2005) have shown that using under-parametrized substitution models will affect the assessment of diversification dynamics (as inferred by the γ statistic of Pybus and Harvey (2000)), erroneously suggesting they are slowing down. Rabosky and Goldberg (2015) addressed the adequacy of the actual diversification models, in the specific case of models for trait dependent speciation, exploring aspects that render those models inadequate and cautioning against using them without additional testing, *e.g.* using neutral trait simulations. Another way of addressing the issue of model inadequacy was introduced with the inclusion of hidden states in state dependent diversification models (Beaulieu & O'Meara 2016; Caetano, O'Meara & Beaulieu 2018), which allow to the signal of rate heterogeneity to be identified in the tree through these hidden states, instead of erroneously assigning it to the observed traits through accidental imperfect association. Finally, following the example of earlier attempts to test the adequacy of substitution models, Höhna, May and Moore (2015) explicitly highlighted capabilities to use posterior-predictive simulations for adequacy testing in their R package TESS (suggesting to use γ). However, the habit of actually testing model adequacy has not yet found its place in the standard procedures of diversification studies (Brown & ElDabaje 2008).

Here, I provide a way of assessing a model's adequacy using tree shape as a comparative metric. The dynamics of speciation and extinction determine the shape of a phylogeny, by affecting its topology and branch length distribution. In fact, these are the core data we use when estimating diversification rates. Different modes of diversification – described by different diversification models – should thus manifest in a range of different tree shapes, some of which are unique to a specific mode of diversification, whereas others are shared. Thus, if a model cannot generate phylogenies of similar shape to a given empirical phylogeny, it is probably not an adequate model to describe its underlying diversification process.

Material and Methods

Implementation of Adequacy Test

To assess adequacy, I rely on tree shape metrics derived from the tree's graph Laplacian (Lewitus & Morlon 2016) as implemented in the package RPANDA (Morlon et al. 2016) in R (R Development Core Team 2014). The method constructs a modified graph Laplacian from a phylogeny and creates a spectral density profile from its eigenvalues. The main aspects of tree shape are characterized from three properties of the spectral density profile including: principal

Eigenvalue (λ*; an indicator for species richness and phylogenetic diversity), asymmetry/skewness (ψ; indicating stemminess vs. tippiness), and peak height/kurtosis (η; indicating tree balance). A more detailed description of the method can be found in Lewitus and Morlon (2016).

The procedure to assess the adequacy of any diversification model for a given phylogeny then works as follows: First I calculate the shape metrics (λ*, ψ, η) of the empirical tree, as well as the rate estimates under the model being assessed. Next, I simulate a set of phylogenies under the same model using the same rates. I then assess the shape metrics for each of the resulting simulated trees, to gain a model- and parameter-specific distribution for each metric. The adequacy of a model for the empirical tree can then be assessed by comparing the shape metrics of the empirical tree to those of the simulated ones. This can be done in a number of ways, the following of which were tried: corrected or uncorrected p-values, 2D or 3D convex hulls, and Euclidean distances between and among simulated and empirical trees. A comparison of the results yielded by each method, suggested Bonferroni-corrected p-values and 2D convex hulls as a combination that provided a good compromise between performance and computational burden. In the first case, I calculate a p-value for each shape metric, based on the empirical distribution function of the metrics describing the simulated trees, and then apply a Bonferroni-correction to the three p-values. For the second method, I construct convex hulls around the points made up of a combination of two metrics describing the simulated trees. The resulting polygons from a minimal set of lines that connect the outermost points of the set and including all points of that set. I then test whether the point corresponding to the empirical tree lies within the convex hull or not, doing so for all three pairwise combinations of metrics. It is worth noting that if the cloud of simulated points has a non-convex shape, a potential could exist where the convex hull includes a significant amount of metric space that is not actually occupied by any of the simulated trees, but could be by the empirical tree.

The assumption is that a model which does not create trees of comparable shape to the empirical phylogeny is inadequate to describe that phylogeny: there is some aspect of the real data that is not predicted by the model, suggesting that a key biological process is not included. For an adequate model, one would expect the empirical tree to fall within the point cloud of the simulated trees, and thus to not significantly fall outside of the simulated distribution for each metric. If there is a significant difference between any of the shape properties of empirical and simulated trees, one has to assume that the model is not adequate, as in that there are processes underlying the generation of the empirical tree that are not accounted for by the employed model. While the focus here is on ultrametric trees of extant taxa, without the extinct lineages, it is worth noting that both the calculation of Laplacian spectra and the simulation of trees under a given model are generally possible, allowing the inclusion of extinct taxa (fossils) in the tree, if information on their placement is available.

The two methods of assessing whether the significant differences exist are expected to complement each other in two ways: the convex hull method tends to be more inclusive, only considering a model inadequate if the empirical tree metrics are entirely outside of the cloud of simulated tree metrics, whereas the Bonferroni-corrected p-values are more strict, potentially deeming a model inadequate if the empirical metrics lie within the simulated cloud of points, but

not falling within 95% of those values, thus only sharing similar shape features with a few simulation-outliers. On the other hand, the p-values are univariate, thus while an empirical tree's shape metrics could lie within 95% of each separate metric of the simulated trees, the convex hulls could reveal whether that particular combination of two metrics occurs within the simulations. Visually inspecting the results in a three-dimensional scatterplot can give us a more intuitive understanding of the data and can *e.g.* help clarifying cases where the two methods (Bonferroni and convex hulls) do not agree. Inspecting how exactly the tree shapes differ should allow one to explore where the inadequacy arises. An overview of all functions in the package can be found in Table II-1.

Test of Specificity

To assess whether the method works reliably, I simulated 300 trees under a constant rate birth death model, and subsequently tested whether said model is adequate to analyze those trees. Using sim.bd.taxa from the R package TreeSim (Stadler 2011), I simulated 300 trees with 200 extant taxa each, pruned extinct lineages, with diversification rates set to λ=1 and μ=0.2. As described above, I then for each tree estimated their rate parameters, simulated 1000 trees each with those parameters under a constant rate birth-death model, inferred the tree shape metrics for all trees and compared those of the sets of simulated 1000 trees with those of the initial trees they are based on.

Examples of Use

- **Simulated Examples**

First, I simulate ten trees each under a constant rate birth death model and a birth death model with mass extinctions and rate shifts. Both tree sets were simulated using the R package TreeSim (Stadler 2011), using sim.bd.taxa and sim.rateshift.taxa respectively. The goal was to obtain two sets of trees with shapes that are distinct enough to assure they do not result from the same processes, thus providing a positive control example. All trees were simulated to have 200 extant taxa, be completely sampled, and not include extinct lineages. The diversification rates for the pure birth death simulations were set to λ=1 and μ=0.2. The other set of trees started with λ=1.5 and μ=0.1 up to a mass extinction event at time 4.5, which only 0.1 of all lineages survived, followed by λ=0.2 and μ=0.1. I then subjected the trees to the same procedure as above to test whether the constant rate birth death model is adequate for them.

- **Empirical Examples**

Further, I use a set of three published empirical phylogenies including a tree of 87 whale species (Steeman *et al.* 2009), a tree of 77 species of Ericaceae (derived from Schwery *et al.* 2015), and a tree of 11 species of *Calomys* (Pigot, Owens & Orme 2012), all of which are part of the package's implemented example tree set. Using three of the currently implemented models – constant rate birth-death, time-dependent birth-death with exponential λ and constant μ, and diversity dependent birth-death with linear λ and constant μ – I inferred their rate estimates, simulated 1000 trees with those estimates under each model, inferred tree shape metrics for all trees and compared the empirical trees with their corresponding sets of simulated trees.

Results

Test of Specificity

Of the 300 trees simulated under a constant rate birth death model, 287 ran successfully, that is, without any issues that prevented the rate estimation or tree simulation. Of the 13 trees that failed to run, 10 failed completely (suggesting issues at the stage of rate estimation), while three only failed to calculate the metrics of the simulated trees. Among the successful runs, the constant rate birth death model was deemed adequate for 283 (98.6%) of which using Bonferroni corrected p-values (α=0.0139), and for 285 (99.3%) of which using 2D convex hulls (α=0.007). I expected the constant rate birth death model, as the generating model, to be adequate for all trees, though of course there is a chance that a rare simulated tree will incorrectly reject the true generating model. Of the sets of trees for which birth-death was inadequate, only one of the two identified through 2D convex hulls was also among those identified through p-values, potentially demonstrating the value of using them complementary. It is also worth noting that when only run with 100 simulations per empirical tree, p-values inferred the model correctly as adequate at a similar rate (98.3%), while 2D convex hulls performed much worse (85.5%), perhaps due to few points leading to bad estimation of the shape of the point cloud (with just 100 points, the number at the outline is very small with large spaces between them).

Examples of Use

- **Simulated Examples**

As would be expected, most trees simulated under birth-death fell well within the range of shape metrics of their corresponding simulations, whereas the shape metrics of all ten trees simulated with mass extinction events fell significantly outside the range of their simulations for at least one metric (Table II-2). However, the birth-death model also turned out to be inadequate for one of the initial birth-death trees (BD 2, see Table II-2). Both methods of assessing adequacy (Bonferroni-corrected p-values and 2D convex hulls) agreed in their verdict of which trees the birth-death model was adequate for.

- **Empirical Examples**

Both the birth-death and the density dependent model are adequate for all three phylogenies tested, while the time-dependent model seems to only be adequate for Tree 3 (*Calomys*), as the other two trees fall outside the simulated distributions for skewness and peakedness, and outside of all three polygons (Table II-3).

Discussion

I implemented a way to test model adequacy of diversification models, by comparing tree shape metrics (derived from Laplacian spectra) of a tree with those of trees simulated under the same model and parameters. Models are not adequate to describe the diversification process underlying a tree, if those models lead to trees of significantly different shape. I tested the approach on sets of simulated and empirical trees to demonstrate its use.

An advantage of using BoskR for empirical data is it allows us seeing which aspects of shape cause the mismatch of a tree with a model. When inspecting the source of the inadequacy of the

birth-death model for the mass extinction trees, it becomes apparent that the principal Eigenvalue is the source of mismatch. (Table II-2). The actual values of metrics (Table II-4) show that the high principal Eigenvalue of the mass extinction tree tends to be greater than that of the birth-death trees, and that the trees simulated from their parameters under the birth-death model have consistently and markedly lower principal Eigenvalues. The principal Eigenvalue of phylogenetic Laplacian spectra is an indicator for species richness and phylogenetic diversity (Lewitus & Morlon 2016). And indeed, while all initial trees had 200 tips, the simulated trees derived from the birth-death set had an average of 357.95 tips (min: 4, max: 3798), whereas those derived from the mass extinction set had an average of 20.46 tips (min: 2, max: 157). While counting the number of tips might have been sufficient in this particular case, it is worth noting that a difference in principal Eigenvalue between two trees could also be found if they had the same number of taxa. For the birth-death set, the estimated diversification rates were λ=1.014 and μ=0.254, but only λ=0.233 and μ=0 for the mass extinction set, leading to the difference in number of tips. The reason behind these differences derives from a constant-rate birth-death model generating (or assuming) an even distribution of speciation and extinction events per lineage over time. The continuous lineage accumulation in the initial birth-death tree set could be described adequately that way, but the mass extinction trees had a high lineage accumulation initially that plateaued after the mass extinction and the subsequent lower net diversification. This inflated the birth-death rate estimates, which cannot adequately represent this kind of dynamic.

The single birth-death tree for which the birth-death model turned out inadequate (BD2), differed significantly from the simulated set of 1000 trees in skewness (Table II-2). It shows a higher (positive) skewness (1.699) than all other initial trees (Table II-4), while the associated simulated trees have similar asymmetries as the simulated sets of the other initial birth-death sets (-2.504 to 2.338, mean of -1.051). Skewness describes the distribution of branching events over time, with negative values implying stemmy trees, and positive values implying tippy trees (Lewitus & Morlon 2016). Thus, BD2 is supposedly tippier than the other trees and the trees simulated from its parameters. When inspecting the tree and its associated Laplacian spectrum (Figure II-1), this tree stands out by having a small peak of higher Eigenvalues next to the main peak, which shifts the skewness of the whole spectrum to be higher, and which should correspond to the long branch clade at the root of the tree (sister to all the other taxa). A clade like this – with not just the diversity difference, but also the long-branch connection to the other taxa – would be rather uncommon under a constant rate birth-death tree but can occur by chance. Thus, it would be equally rare to get a similar-looking tree within the simulated set of 1000 birth-death trees, leading to the result that the model is inadequate, despite the initial tree being simulated under a birth-death model. Indeed, when dropping this clade from the tree and rerunning the adequacy test, birth-death comes out as adequate for this tree.

For the empirical examples, a comparison of Trees 1 and 2 and the tree sets derived from them under the time-dependent model, reveals that they mainly disagree in skewness and peakedness. As mentioned above, skewness describes whether a tree is tippy or stemmy, while peakedness describes how evenly the Eigenvalues are distributed, with low values indicating a homogeneous distribution, thus balanced trees, and high values accordingly imbalanced trees (Lewitus & Morlon 2016). For the two trees where the time dependent model is inadequate, both skewness

and peakedness are higher for the empirical trees than for the 1000 simulated ones derived from them (Table II-5). Thus, they should be more tippy and imbalanced than they would be under a time dependent process. Indeed, when looking at the actual trees, it becomes apparent that the simulated trees are extraordinarily stemmy and balanced when compared to their empirical counterpart (Figure II-2). The diversification rates estimated from the empirical trees (and used as input for the simulated sets) show that extinction is close to zero for all three trees, and that the speciation rate of whales and Ericaceae decrease over time (Figure II-3), which would be consistent with stemmy trees.

I believe that BoskR will be a useful tool for researchers attempting to identify which models are adequate or inadequate to describe the diversification dynamics of a study group. These insights should allow for a more informed analysis and increased confidence in the obtained results. Since tree shape (being topology and branch lengths) is essentially the primary data we use in model-based inference of diversification rates, it should an appropriate and intuitively relatable basis of assessing diversification model adequacy. Various other – arguably equally sensible – metrics of tree shape are of course available, and some of which might be assessed more easily. Indeed, for my simulated example, one could have simply used the number of tips to conclude that the birth-death model is inadequate for trees that evolved under a mass extinction model. However, number of tips would not have flagged BD2, which differed in skewness instead. Additionally, I have conditioned all simulation on crown age, which I deemed the most reasonable choice for most diversification questions. However, for cases where e.g. conditioning on number of surviving species, or both crown age and number of surviving species is more sensible, number of tips will not be a useful metric anymore. Using the Laplacian spectrum based metrics in BoskR thus has the advantage to include several aspects of tree shape, and thereby also allows to explore whether and how one or several of these aspects render a model inadequate for a tree. Finally, by implementing the approach in an openly available R package, I enable and encourage researchers to adopt adequacy testing into their workflow.

It is, however, worth noting an implicit assumption of my approach: that the empirical tree under scrutiny is the true tree (or a reasonably close approximation of it). Our approach does not account for any errors and biases introduced in the tree building process. It would thus be possible that *i.e.* a group in fact diversified in a time-dependent fashion, while BoskR would deem this an inadequate model, because the signature of time dependence was lost during tree building and divergence time estimation. Furthermore, this approach does currently not account for issues in rate estimation. If the diversification rates are not estimated accurately to begin with, BoskR would likely label the corresponding model inadequate, without that necessarily being the case.

Future versions of the package will aim to include a wider range of models, to better reflect the diversity of models currently used in the field. However, as long as any given model is able to both estimate parameters and simulate trees under those parameters, the resulting outputs can be used as input for BoskR and can thus be analyzed without problem.

References

Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle,[w:] proceedings of the 2nd international symposium on information, bn petrow, f. *Czaki, Akademiai Kiado, Budapest*.

Beaulieu, J.M. & O'Meara, B.C. (2016) Detecting Hidden Diversification Shifts in Models of Trait-Dependent Speciation and Extinction. *Systematic Biology,* **65,** 583-601.

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

Brown, J.M. (2014) Detection of implausible phylogenetic inferences using posterior predictive assessment of model fit. *Systematic Biology,* **63,** 334-348.

Brown, J.M. & ElDabaje, R. (2008) PuMA: Bayesian analysis of p artitioned (and u npartitioned) m odel a dequacy. *Bioinformatics,* **25,** 537-538.

Caetano, D.S., O'Meara, B.C. & Beaulieu, J.M. (2018) Hidden state models improve state-dependent diversification approaches, including biogeographical models. *Evolution,* **72,** 2308-2324.

Duchêne, D.A., Duchêne, S., Holmes, E.C. & Ho, S.Y. (2015) Evaluating the adequacy of molecular clock models using posterior predictive simulations. *Molecular Biology and Evolution,* **32,** 2986-2995.

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

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

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

Lewis, P.O., Xie, W., Chen, M.-H., Fan, Y. & Kuo, L. (2013) Posterior predictive Bayesian phylogenetic model selection. *Systematic Biology,* **63,** 309-321.

Lewitus, E. & Morlon, H. (2016) Characterizing and Comparing Phylogenies from their Laplacian Spectrum. *Systematic Biology,* **65,** 495-507.

Morlon, H., Lewitus, E., Condamine, F.L., Manceau, M., Clavel, J. & Drury, J. (2016) RPANDA: an R package for macroevolutionary analyses on phylogenetic trees. *Methods in Ecology and Evolution,* **7,** 589-597.

Pennell, M.W., FitzJohn, R.G., Cornwell, W.K. & Harmon, L.J. (2015) Model Adequacy and the Macroevolution of Angiosperm Functional Traits. *American Naturalist,* **186,** E33-E50.

Pigot, A.L., Owens, I.P.F. & Orme, C.D.L. (2012) Speciation and Extinction Drive the Appearance of Directional Range Size Evolution in Phylogenies and the Fossil Record. *Plos Biology,* **10**.

Pybus, O.G. & Harvey, P.H. (2000) Testing macro–evolutionary models using incomplete molecular phylogenies. *Proceedings of the Royal Society of London. Series B: Biological Sciences,* **267,** 2267-2272.

R Development Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rabosky, D.L. & Goldberg, E.E. (2015) Model Inadequacy and Mistaken Inferences of Trait-Dependent Speciation. *Systematic Biology,* **64,** 340-355.

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

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

Schwery, O., Onstein, R.E., Bouchenak-Khelladi, Y., Xing, Y., Carter, R.J. & Linder, H.P. (2015) As old as the mountains: the radiations of the Ericaceae. *New Phytologist,* **207,** 355-367.

Stadler, T. (2011) Simulating Trees with a Fixed Number of Extant Species. *Systematic Biology,* **60,** 676-684.

Steeman, M.E., Hebsgaard, M.B., Fordyce, R.E., Ho, S.Y.W., Rabosky, D.L., Nielsen, R., Rahbek, C., Glenner, H., Sorensen, M.V. & Willerslev, E. (2009) Radiation of Extant Cetaceans Driven by Restructuring of the Oceans. *Systematic Biology,* **58,** 573-585.

Sullivan, J. & Swofford, D.L. (1997) Are guinea pigs rodents? The importance of adequate models in molecular phylogenetics. *Journal of Mammalian Evolution,* **4,** 77-86.

Appendix B

Table II-1: Functions available in BoskR.

Function	Description
CombineTrees	combines several phylo objects into required format
GetMetricTreeSets	simulates sets of trees under a certain model using parameters estimated from an empirical tree
GetTreeMetrics	obtains shape metrics for sets of trees
GetTreeParams	estimates diversification model parameters of a set of trees
PvalMetrics	calculates a p-value for the empirical tree metrics based on the empirical cumulative density function of the simulated metrics
ScatterMetrics	generates a 3D-scatterplot of empirical and simulated tree metrics
TreeCorr	tests whether trees are ultrametric, bifurcating, and ordered cladewise, and corrects automatically if possible
plotPvalMetricsCDF	plots the empirical metrics and obtained p-values on their respective cumulative distribution functions
plotPvalMetricsPDF	plots empirical metrics on distribution of simulated metrics

Table II-2: Results simulated examples.

For each initially simulated tree, the three shape metrics (principal Eigenvalue, skewness, peakedness) are shown, followed by their corresponding Bonferroni-corrected p-value. The last three columns indicate for each pairwise combination of two metrics whether the initial tree is inside or outside of the polygon around the values of the simulated trees. P-values below 0.05 and points outside of 2D polygons are in bold and highlighted gray. BD= birth-death, ME= mass extinction, PrinE= principal Eigenvalue, Skew= skewness, Peak= peakedness.

Tree	Principal Eigenvalue	p	Skewness	p	Peakedness	p	PrinE-Skew	PrinE-Peak	Peak-Skew
BD 1	4122.028	1.000	-1.085	1.000	5.331	1.000	in	in	in
BD 2	5445.351	0.534	1.699	**0.018**	8.399	1.000	in	in	**out**
BD 3	5815.464	1.000	0.218	1.000	4.677	1.000	in	in	in
BD 4	4082.034	1.000	-0.948	1.000	4.436	1.000	in	in	in
BD 5	3985.050	1.000	-0.522	1.000	3.205	0.546	in	in	in
BD 6	5602.252	1.000	-1.031	1.000	4.279	1.000	in	in	in
BD 7	4886.781	0.930	-0.727	1.000	4.603	1.000	in	in	in
BD 8	4039.776	1.000	1.408	0.066	5.673	1.000	in	in	in
BD 9	5857.949	1.000	-0.775	1.000	4.863	1.000	in	in	in
BD 10	3383.424	1.000	-1.246	1.000	4.472	1.000	in	in	in
ME 1	6332.263	**0.000**	-0.867	0.156	3.435	0.180	**out**	**out**	in
ME 2	6881.577	**0.000**	-0.822	0.246	3.410	0.258	**out**	**out**	in
ME 3	6162.888	**0.000**	-0.961	0.084	3.517	0.126	**out**	**out**	in
ME 4	6196.274	**0.000**	-0.909	0.060	3.705	0.132	**out**	**out**	in
ME 5	6181.521	**0.000**	-0.939	0.096	3.476	0.174	**out**	**out**	in
ME 6	6445.562	**0.000**	-1.005	0.072	4.167	**0.048**	**out**	**out**	in
ME 7	6287.568	**0.000**	-0.821	0.216	2.934	0.522	**out**	**out**	in
ME 8	6123.175	**0.000**	-0.969	0.084	3.944	0.096	**out**	**out**	in
ME 9	6574.787	**0.000**	-1.007	0.102	3.909	0.138	**out**	**out**	in
ME 10	6908.047	**0.000**	-0.962	0.144	3.686	0.198	**out**	**out**	in

Table II-3: Results empirical examples.

For each of the three models, the three shape metrics (principal Eigenvalue, skewness, peakedness) of the empirical trees are shown, followed by their corresponding Bonferroni-corrected p-values. The last three columns indicate for each pairwise combination of two metrics whether the initial tree is inside or outside of the polygon around the values of the simulated trees. P-values below 0.05 and points outside of 2D polygons are in bold and highlighted gray. Tree 1= whales, Tree 2= Ericaceae, Tree 3= *Calomys,* BD= birth-death, TD= time dependent birth-death, DD= diversity dependent birth-death ME= mass extinction, PrinE= principal Eigenvalue, Skew= skewness, Peak= peakedness.

Model	Tree	Principal Eigenvalue	p	Skewness	p	Peakedness	p	PrinE-Skew	PrinE-Peak	Peak-Skew
BD	1	11340.708	1.000	0.592	1.000	2.065	0.936	in	in	in
	2	33280.806	1.000	0.129	1.000	1.989	0.738	in	in	in
	3	111.085	1.000	0.710	1.000	1.389	1.000	in	in	in
TD	1	11340.708	0.108	0.592	**0.000**	2.065	**0.000**	**out**	**out**	**out**
	2	33280.806	0.150	0.129	**0.000**	1.989	**0.000**	**out**	**out**	**out**
	3	111.085	1.000	0.710	1.000	1.389	1.000	in	in	in
DD	1	11340.708	1.000	0.592	0.786	2.065	0.462	in	in	in
	2	33280.806	1.000	0.129	1.000	1.989	0.336	in	in	in
	3	111.085	1.000	0.710	1.000	1.389	1.000	in	in	in

Table II-4: Metrics of Simulated Trees

For all three shape metrics (Principal Eigenvalue, Skewness, Peakedness) the values for each initially simulated tree are shown followed by the mean, minimal and maximal values of their corresponding set of 1000 trees simulated with the same parameters. BD= birth-death, ME= mass extinction.

Tree	Principal Eigenvalue				Skewness				Peakedness			
	initial	mean	min	max	initial	mean	min	max	initial	mean	min	max
BD 1	4122.028	3544.249	66.293	16677.736	-1.085	-0.471	-2.134	3.858	5.331	4.743	1.124	11.297
BD 2	5445.351	21648.058	522.668	103476.154	1.699	-1.051	-2.504	2.338	8.399	7.706	1.589	17.410
BD 3	5815.464	6948.965	239.777	31975.499	0.218	-0.553	-2.648	3.337	4.677	5.515	1.354	15.076
BD 4	4082.034	4400.988	108.527	20366.263	-0.948	-0.611	-2.173	2.782	4.436	5.119	1.501	13.664
BD 5	3985.050	7054.063	169.758	31808.846	-0.522	-0.820	-2.293	2.928	3.205	5.927	1.477	15.543
BD 6	5602.252	9924.124	228.569	49798.910	-1.031	-0.795	-2.470	3.539	4.279	6.166	1.250	16.263
BD 7	4886.781	13639.045	326.075	77492.732	-0.727	-0.984	-2.426	2.193	4.603	7.057	1.690	17.792
BD 8	4039.776	8845.887	106.573	51451.470	1.408	-0.883	-2.214	3.219	5.673	6.296	1.160	15.784
BD 9	5857.949	9063.837	182.765	43736.546	-0.775	-0.665	-2.471	3.940	4.863	5.924	1.167	16.875
BD 10	3383.424	2497.903	74.447	13926.700	-1.246	-0.471	-2.034	2.973	4.472	4.329	1.042	12.513
ME 1	6332.263	590.048	45.513	2921.191	-0.867	0.457	-1.176	2.889	3.435	1.872	0.756	4.579
ME 2	6881.577	796.181	49.063	5629.196	-0.822	0.339	-1.431	3.124	3.410	2.033	0.735	7.614
ME 3	6162.888	550.735	44.277	3147.839	-0.961	0.472	-1.096	3.131	3.517	1.873	0.745	6.045
ME 4	6196.274	538.247	44.712	2565.612	-0.909	0.479	-1.277	2.930	3.705	1.875	0.728	25.277
ME 5	6181.521	599.297	43.749	2577.140	-0.939	0.432	-1.309	2.773	3.476	1.918	0.789	7.058
ME 6	6445.562	698.300	46.356	3618.554	-1.005	0.368	-1.345	2.635	4.167	1.995	0.719	4.997
ME 7	6287.568	620.810	44.693	2482.020	-0.821	0.387	-1.268	3.198	2.934	1.935	0.720	5.692
ME 8	6123.175	658.443	43.344	2809.674	-0.969	0.377	-1.203	3.419	3.944	2.020	0.731	6.413
ME 9	6574.787	709.712	45.960	4468.763	-1.007	0.360	-1.300	2.687	3.909	2.039	0.721	11.556
ME 10	6908.047	752.806	49.307	3701.842	-0.962	0.432	-1.177	3.719	3.686	2.012	0.827	8.317

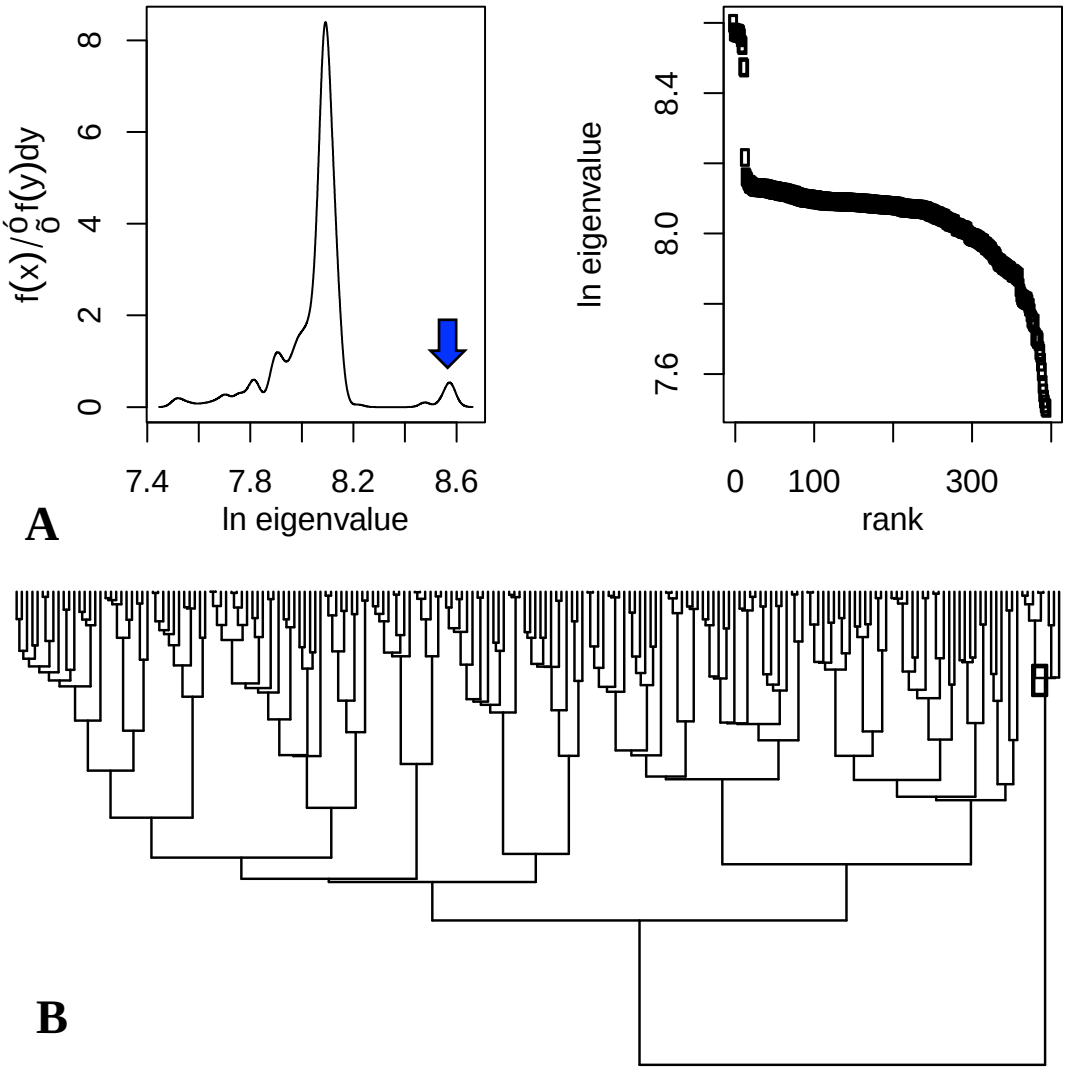

Figure II-1: Phylogeny BD2 and associated Laplacian spectrum.

A: Laplacian spectrum of the tree BD2, left, and eigenvalues sorted by size, right. **B:** The phylogeny BD2. The blue dot and arrow indicate the clade (and its associated peak in the spectrum) that caused the shift in skewness.

Table II-5: Metrics of Empirical Trees

For all three shape metrics (principal Eigenvalue, skewness, peakedness), the values for each empirical tree are shown followed by the mean, minimal and maximal values of their corresponding set of 1000 trees simulated with the same parameters under all three models. BD= birth-death, TD= time dependent birth-death, DD= diversity dependent birth-death.

Model	Tree	Principal Eigenvalue				Skewness				Peakedness			
		initial	mean	min	max	initial	mean	min	max	initial	mean	min	max
BD	1	11340.708	10491.411	320.914	47380.279	0.592	-0.218	-1.733	3.467	2.065	3.447	0.781	7.907
	2	33280.806	33454.247	981.070	181480.025	0.129	-0.228	-1.733	3.779	1.989	3.418	0.952	8.982
	3	111.085	115.525	14.961	565.026	0.710	0.809	-1.207	3.432	1.389	1.716	0.723	13.604
TD	1	11340.708	3851.769	179.289	17380.147	0.592	-0.090	-0.235	0.498	2.065	1.190	0.756	1.728
	2	33280.806	12583.687	586.423	78178.138	0.129	-0.052	-0.311	0.000	1.989	1.013	0.735	1.510
	3	111.085	106.331	14.961	435.420	0.710	0.377	-0.756	2.497	1.389	1.413	0.723	5.380
DD	1	11340.708	11447.491	554.804	24436.742	0.592	-0.370	-1.610	3.625	2.065	3.511	1.094	7.034
	2	33280.806	33913.463	1483.834	65119.609	0.129	-0.366	-1.502	3.299	1.989	3.373	1.059	8.771
	3	111.085	100.894	26.156	141.440	0.710	0.784	-0.601	3.164	1.389	1.706	0.810	13.650

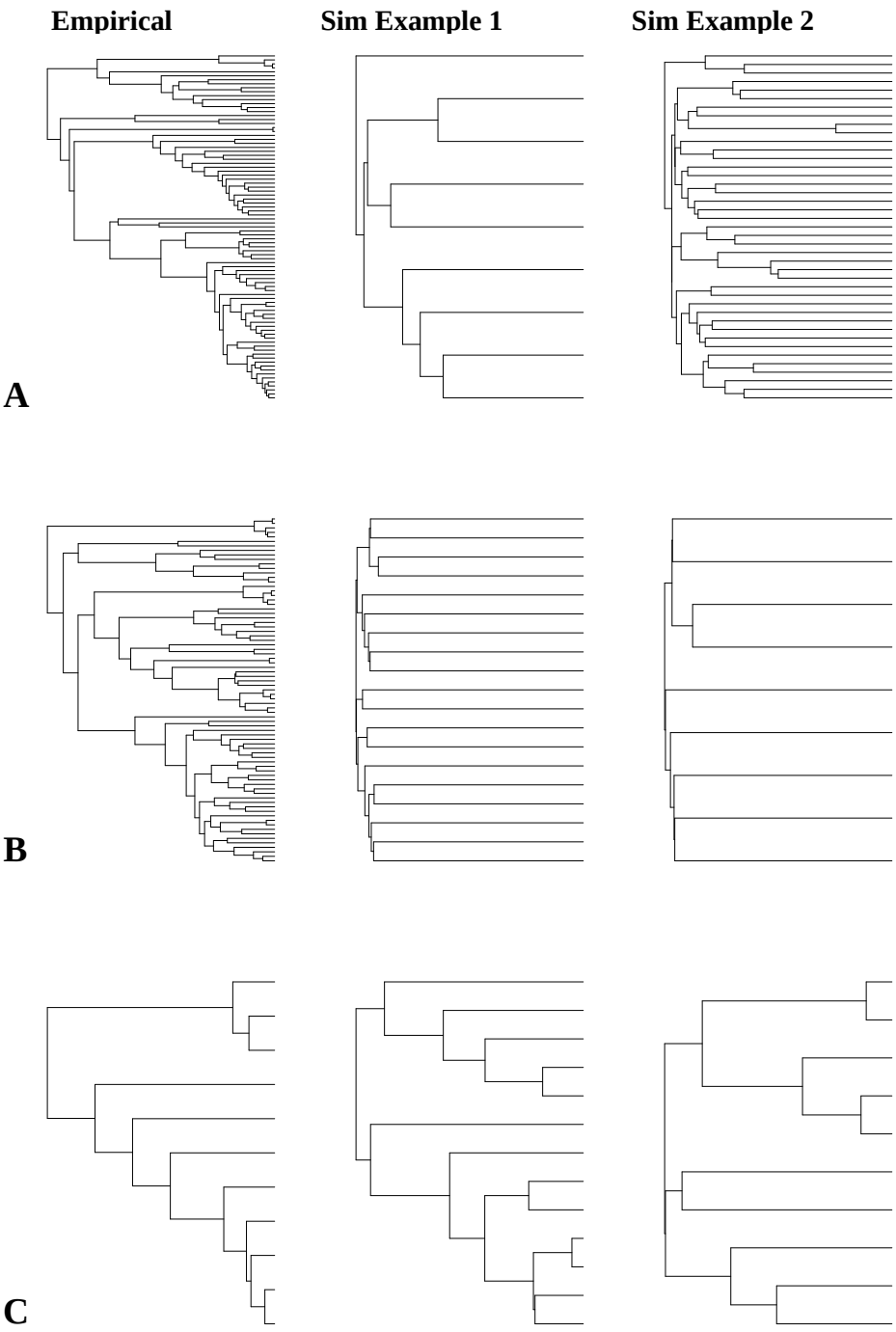

Figure II-2: Empirical Phylogenies and Examples Associated Simulations.

For each empirical phylogeny, two example trees are shown, simulated with the same parameters as the empirical tree, and under the time dependent model. **A**: whales, **B**: Ericaceae, **C**: *Calomys*.

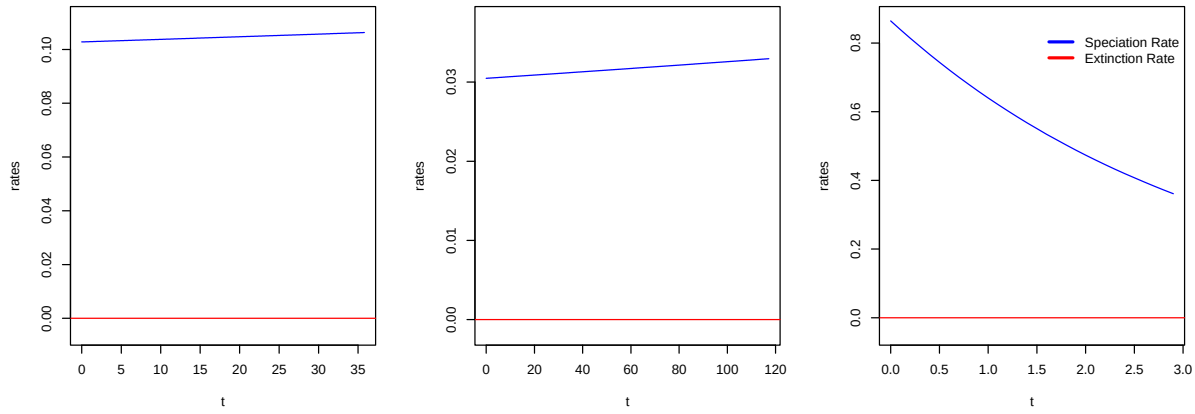

Figure II-3: Diversification Rates for Empirical Trees under Time-Dependent Model.

Speciation and extinction rates estimated from the empirical phylogenies under the time dependent model. **A**: whales, **B**: Ericaceae, **C**: *Calomys*, Blue: Speciation rate, red: Extinction rate.

CHAPTER III
THE SHAPE OF TREES – LIMITS OF CURRENT DIVERSIFICATION MODELS

Abstract

To investigate how biodiversity arose, the field of macroevolution largely relies on model-based approaches to estimate rates of diversification and what factors influence them. The number of available models is rising steadily, facilitation the modeling of an increasing number of possible diversification dynamics, and multiple hypotheses relating to what fueled or stifled lineage accumulation within groups of organisms. However, growing concerns about unchecked biases and limitations in the employed models suggest the need for rigorous validation of methods used to infer. Here, I address two points: the practical use of model adequacy testing, and what model adequacy can tell us about the overall state of diversification models. Using a large set of empirical phylogenies, and a new approach to test models using aspects of tree shape, I test how a set of staple models performs with regards to adequacy. Patterns of adequacy are described across trees and models and causes for inadequacy – particularly if all models are inadequate – are explored. The findings make clear that overall, only few empirical phylogenies cannot be described by at least one model. However, finding that the best fitting of a set of models might not necessarily be adequate makes clear that adequacy testing should become a step in the standard procedures for diversification studies.

Introduction

Among the main goals of the study of macroevolution is to explain large scale patterns of biodiversity, and to unveil the mechanisms behind it. In the case of the diversity of species, and particularly the heterogeneity of how that diversity is distributed both in space and across the tree of life, model-based diversification studies have become the method of choice. Since the use of models explicitly allows the inclusion of particular mechanisms and influencing factors into the analyses, making them particularly useful for testing specific hypotheses, rather than just describing patterns. Therefore, such models are especially promising to not only explain diversity patterns, but also better understand the forces that generate this diversity. This has led to considerable development of new approaches, ranging from simple estimation of constant speciation rates (λ) and extinction rates (μ), to via rates influenced by time or diversity, to more complex sets of models in which diversification rates depend on trait states (Maddison, Midford & Otto 2007; FitzJohn, Maddison & Otto 2009; FitzJohn 2010; FitzJohn 2012) or geographic areas (Goldberg, Lancaster & Ree 2011), or approaches that focused on the localization of shifts in diversification rates (Alfaro *et al.* 2009; Rabosky 2014), all of which are widely used.

However, all of these approaches have underlying assumptions, which may not always be satisfied by empirical data; and that they all have limitations, which can severely bias our results and mislead our interpretations thereof. Examples for this include biases introduced by under-parametrized substitution models (Revell, Harmon & Glor 2005), a lack of replication for categorical traits which only occur in one clade (Maddison & FitzJohn 2015), models erroneously inferring trait effects on diversification rates when the trait is actually neutral (Rabosky & Goldberg 2015), and the fact that most studies only rely on the explanatory power of a model to assess its adequacy, which is often not sufficient (Pennell et al. 2015). While ways to address some of these issues through more elaborate models (Beaulieu & O'Meara 2016;

Caetano, O'Meara & Beaulieu 2018) or explicit adequacy testing (Höhna, May & Moore 2015) have been proposed, a widespread standardized use of model adequacy testing has not yet established in the field (Brown & ElDabaje 2008), suggesting that the magnitude of model adequacy problems in the field of diversification studies is not yet known.

Here, I employ a simulation-based approach using tree shape metrics to assess the adequacy of standard diversification models across a large set of empirical trees. The method is implemented in the R package BoskR, which is described in detail in CHAPTER II. The results of the assessment should show us how well the models perform overall in comparison and reveal patterns on which aspects of the tree shapes they might fail to describe. However, trees for which all models fail might be indicative of common weaknesses of all current models and exploring the causes of inadequacy might bring to light properties of trees for which none of our current models can account. Exploring these could allow the identification of the kind of models we might still be missing and where the field has to develop further.

Material and Methods

Empirical Phylogenies

I assembled a collection of empirical trees of various sizes and taxonomic groups. Using the R package datelife (Stoltzfus *et al.* 2013; Nguyen *et al.* 2018), I assembled all chronograms from the Open Tree of Life (Hinchliff et al. 2015) that had between 30 and 200 taxa. I added the widely used tree of 87 whale species (Steeman et al. 2009), including its subtrees of the clades Balaenopteridae, Phocoenidae, and Phyllostomidae, and a tree of 11 species of *Calomys* (Pigot, Owens & Orme 2012). Finally, I added trees of 77 and 470 species respectively of Ericaceae (derived from Schwery et al. 2015), 575 species of Poales (Bouchenak-Khelladi, Muasya & Linder 2014), 584 species of Fagales (Xing *et al.* 2014), 309 species of Liverworts (Laenen *et al.* 2014), and 549 species of Angiosperms (O'Meara *et al.* 2016) coming to a total of 131 trees (the list of which is given in Table III-1).

Additionally, I obtained a tree set of 214 ultrametric trees of vertebrate families. This set was initially used by Lewitus and Morlon (2016b) to analyze patterns of diversification with their new method using Laplacian spectra (Lewitus & Morlon 2016a), which I am making use of here. The same set was subsequently used by Burin *et al.* (2019) to investigate how well we can estimate diversification rates under scenarios of diversity decline.

Combined, my set consisted of 345 trees spanning a broad range of taxonomic groups. Before subsequent analyses, I checked whether the trees were ultrametric and strictly bifurcating, and corrected both where necessary using BoskR's TreeCorr function.

Adequacy Test

From the diversification models available in BoskR, I chose a set of six basic models to represent some of the different rate dynamics that are implemented. Those were 1) constant speciation and no extinction (Yule), 2) constant speciation and extinction, time-dependent birth death with, 3) exponential speciation and constant extinction, 4) constant speciation and exponential extinction,

as well as diversity-dependent birth death with, 5) linear speciation and constant extinction, and 6) linear speciation and extinction.

To determine the adequacy of my candidate models to describe the diversification dynamics behind my empirical tree sets, I made use of the R package BoskR (as described in CHAPTER II of this dissertation). For each tree, I estimated diversification rates (and additional model parameters in case of time- and diversity-dependent models) under each model, and then used those estimates to simulate 1000 trees per model and tree. I then calculated the three shape metrics (principal Eigenvalue, skewness, and peakedness) from their spectral density profiles, and compared the metrics of each empirical tree with its corresponding simulated trees. Model inadequacy was determined both via Bonferroni-corrected p-values and 2D convex-hulls, with inadequate models showing significant p-values (<0.05) or having the empirical tree's metrics lie outside of at least one of the three polygons.

Trees without adequate Models

When all employed models are inadequate for a tree, this could mean that we have yet to develop suitable models to describe its underlying mode of diversification. While testing all available models and model-variations currently in existence is beyond the scope of this work, I wanted to expand the basic model set used by eleven, to get a sense of how many trees could actually not be accounted for by any model. These additional models were time dependent birth death with: 7) constant speciation and linear extinction, 8) exponential speciation and extinction, 9) exponential speciation and no extinction, 10) exponential speciation and linear extinction, 11) linear speciation and constant extinction, 12) liner speciation and extinction, 13) linear speciation and exponential extinction, 14) linear speciation and no extinction, as well as diversity-dependent birth death with: 15) exponential speciation and constant extinction, 16) constant speciation and linear extinction, 17) constant speciation and exponential extinction. I tested the adequacy of these models on the trees for which all previous models were inadequate, using the same procedure as described above.

For any trees for which still no model was adequate, I investigated the causes of inadequacy of the models by comparing the tree's metrics to those of the others, and by more closely inspecting its spectral density, and taxonomical background.

Adequacy and Fit

An important consideration is whether model adequacy is related to model fit, that is, whether adequate models tend to fit the data well, whereas inadequate models do not, or, in other words, whether mismatches are possible. To that end, I compared the fits of the different models by comparing their respective AIC. I first obtained the log-likelihood associated to parameter estimates under each model. Because different R packages calculate the likelihood differently, those values are not directly comparable (Stadler 2013). While I could transform the likelihoods of the models that were implemented in ape (Yule, constant-rate birth-death), to match those from DDD (the two diversity-dependent models) by dividing them by -2, the models implemented in RPANDA (the two time-dependent models) used a different likelihood equation altogether, which made it impossible to compare them by simple transformations (see Table 1 in Stadler 2013). I thus excluded those two models from the comparison. For the remaining four

models, I calculated the AIC from the adjusted log-likelihoods, determined the best model using ΔAIC, and tallied how often the best fitting model and the remaining models were either adequate or inadequate.

Results

Adequacy Test

For a total of 234 trees, all six initial models were successfully run. Among these, according to the corrected p-values (Table III-2), all six models had quite comparable numbers of trees for which they were adequate (82.5-89.3%), with the exception of the diversity dependent model with both λ and μ depending linearly on K being adequate for fewer trees (73%). The 2D convex hulls showed a similar picture (adequate for 78.2-87.6%), with Yule, constant-rate birth-death, and λ-linear μ-constant diversity dependent doing slightly better than the other three models (Table III-2).

Looking at the results from the perspective of trees, *all* models were adequate for a majority of them (61.1% or 58.1% respectively), with the percentage of trees for which fewer models were adequate declining quickly (Table III-3). In cases where only one model was adequate, it was either the λ-exponential μ-constant time dependent (66.6%/42.9%), or the λ-linear μ-constant diversity dependent model (33.3%, more clearly so for 2D convex-hulls with 42.9%). When only one model was inadequate, it was usually the λ-linear μ-linear diversity dependent model according to the corrected p-values (62.9%), and the λ-exponential μ-constant time dependent model according to 2D convex-hulls (46.3%). For a small number of trees (12 according to corrected p-values, 5 according to 2D-convex-hulls), none of the six initial models were adequate (see below).

When comparing how often combinations of models are inadequate (Table III-4), both assessments agree that Yule and constant rate birth-death highly associated (in case of p-values, Yule is always inadequate when birth-death is), and both are often inadequate together with the λ-constant μ-exponential time dependent model, but less so vice-versa. Finally, when the λ-linear μ-constant diversity dependent model is inadequate, often the λ-linear μ-linear diversity dependent model is as well, but not vice-versa.

Inspection of the metrics shows that very commonly the skewness was responsible for the inadequacy of all models (Table III-5). For Yule and constant-rate birth-death, it is followed by the principal Eigenvalue, while for exponential-λ constant-μ time-dependence and the two diversity dependent models, skewness was closely followed by peakedness. The constant-λ exponential-μ time-dependent model was mostly inadequate due to the principal Eigenvalue, followed by skewness. For the constant-λ exponential-μ time-dependent and the linear λ and μ diversity dependent model, the three metrics were more equally responsible than for the other four models.

In terms of actual metric values (Table III-5), skewness tends to be lower in the simulated trees than in their empirical counterparts, meaning the simulated trees are tippier. This is true for all

models, and whether or not they are adequate for a tree or not, with the difference being larger in trees modeled under inadequate models, and larger for trees for which more models are inadequate. Another global trend is that trees simulated under inadequate models tend to have a higher difference in principal Eigenvalue than those simulated under adequate models – in case of Yule and constant-rate birth-death, simulations under adequate models even have a lower principal Eigenvalue than the initial tree. This pattern is mirrored by the number of taxa in the trees. Also, peakedness and Eigengap (a measure of number of peaks in the spectrum, and indicator for how many distinct modes of diversification there are) while in general higher in all simulations, tend to be higher in trees simulated under an inadequate model.

We compared the metrics of the empirical trees between trees for which different numbers of models are adequate using an ANOVA and Tukey's post-hoc test (Figure III-1). Apart from slight patterns in age, principal Eigenvalue and Eigengap, the most striking signal comes from skewness, which seems to increase the chance that more models are inadequate for a tree if its skewness is more positive, meaning if the trees are very tippy.

A comparison of tree metrics for which a particular model is adequate or inadequate (Table III-6) shows that trees for which any model is inadequate tend to have a larger skewness than those for which the model is adequate. For Yule, birth-death, and linear-λ constant-μ density dependent, trees for which they are inadequate tend to additionally be larger, older, and have higher principal Eigenvalues – and a lower Eigengap in case of the diversity dependent model. Trees for which constant-λ exponential-μ time dependent models tend to be older than those for which it is adequate, as are trees for which linear-λ and μ is inadequate, with the addition that the latter also have a lower Eigengap. Finally, only for the exponential-λ constant-μ model there is a significant effect in peakedness (apart from the usual higher asymmetry and lower Eigengap), in that trees for which the model is inadequate have a lower peakedness.

Trees without adequate Models

Among trees for which all models ran, no model was adequate for twelve trees according to the Bonferroni-corrected p-values, and for five trees according to the 2D convex-hulls, with an overlap of four trees for which no model was adequate according to either way of assessing it. When running the additional set of models on those four trees, adequate models were found for two of them, while the other two still could not be adequately described by any model used (however, about half of the models used on those trees did not run successfully).

In terms of tree metrics, the two trees for which all models are inadequate, do not stand out in any particular way (Figure III-2), apart from their high skewness, as expected based on the results above. However, this would not explain what differentiates these two trees from the ones for which only the six initial models were inadequate, but for which an adequate one was found among the additional eleven models. The inferred phylogenies and associated Laplacian spectra (Figure III-3) suggest that the source of the high skewness might be that these trees have a small secondary peak at higher Eigenvalues than the main peak. These seem to result from the species/clade that is sister to the rest of each respective tree, and that they are not only

subtending a for this tree comparatively long branch, but that the rest of the tree is subtending a relatively long branch as well.

Adequacy and Fit

Overall, of 234 trees for which models were fitted, in 198 cases the best fitting model was adequate. In 22 cases the best model was inadequate, but at least one of the other models was adequate, and for 14 trees none of those four models were adequate (Figure III-4).

Discussion

We have seen a steady increase of models used to study diversification, and the equally increasing awareness of the issues and shortcomings those models have. In the light of this, steps need to be taken towards gaining insight into whether and how these issues affect the results inferred under these models, thereby both improving our inference and increasing confidence in our findings. This research addresses this issue by applying a new method to test the adequacy of diversification models to a large set of empirical phylogenies, in order to uncover patterns in model adequacy and the causes behind them.

Overall Adequacy Patterns

Overall, the results draw a promising picture of the adequacy of our models (assuming the adequacy test is accurate, and the metrics used as test statistics capture meaningful aspects of the trees). All models under investigation were adequate for a relatively high number of trees, and also from the tree's perspective, not only were all or many models adequate for most trees, even with the modest set of six basic models, for only a small number of trees could an adequate model not be found. It might come at a surprise that the models performed so evenly across the trees, as one could have imagined that *e.g.* overly simplistic models like Yule would only be adequate for a particular kind of tree. However, this seems not to be the case. This presumably demonstrates the flexibility of those models, partially resulting from the stochasticity underlying the processes – even when we use a model with one constant rate of lineage accumulation, a relatively wide range of branching patterns can still emerge from it by chance (Slowinski & Guyer 1989). However, it calls into question whether such a model has too much stochasticity to allow reliable inference, or in my case of adequacy testing. The wide range of tree shape space a model covers may make it seem adequate without necessarily meaning that the model really tells us anything reliable about the diversification process underlying a certain tree.

Models that are adequate when the others fail (λ-exponential μ-constant time dependent or λ-linear μ-constant diversity dependent), might suggest that these models are able to adequately describe particular branching patterns, for which the stochasticity of the other models could not account. For example, exponential variation in λ should be able to account for more extreme cases of increases or decreases in diversification on either extreme of time, whereas the diversity dependence model would allow for the kind of stagnation other models could not account for. Indeed, the challenge of being able to detect decreases in diversification rates have been noted previously (Liow, Quental & Marshall 2010; Burin *et al.* 2019).

Trees without adequate Models

It is both concerning and exciting to find phylogenies for which no model is adequate. The concern arises because the field lacks the tools to address come case studies. But those cases might reveal new aspects of diversification and drive the exploration of new models. Particularly with the limited number of models used here – and demonstrated by applying the extended set of models to the initially unmatched trees – it is conceivable that the adequate model for those cases might indeed already exist and is just waiting to be employed.

Alternatively, it is possible that the tree shape metrics used here are not appropriate – or not all of the appropriate necessary metrics – to use. Inadequate models maybe deemed adequate if these metrics fail to capture a crucial aspect of the trees lost on the model. Adequate models may seem inadequate if the metrics capture variation in an aspect of the tree that is not actually related to its underlying diversification process. However, while additional and alternative metrics should still be explored in the future, there is at least some confidence in the ones used here, as they have generally been shown to be able to distinguish different tree shapes (Lewitus & Morlon 2016a).

Another possibility, given that only for less than 1% of the tested trees tested is that BoskR erroneously failed to detect that. This would relegate these un-matchable trees simply be the product of type-I error. As mentioned before, there is a certain amount of stochasticity involved in the process, and as discussed in Chapter II, the results can be significantly influenced by things like *e.g.* insufficient simulations to generate a distribution of shape metrics that properly represents the properties of a certain pairing of model and parameters.

Returning to the two actual trees for which no model was adequate, their high value for skewedness appears to be the only aspect that, at least to some extent, differentiates them from trees for which models were found adequate. As it appears that the comparatively large amount of branch length that separates the bulk of the species from those sister to them (Figure III-3), the question arises whether these trees might simply include their outgroup. Tree 244 is of the bird family Dricuridae, tree 323 of the salamander family Hynobiidae. While the suspected outgroup of the Hynobiidae (*Onychodactylus*) is indeed part of the family, *Chaetorhynchus papuensis* only used to be placed in the Dricuridae. Alternatively, it has recently been grouped in the family Rhipiduridae (Barker *et al.* 2004; Irestedt *et al.* 2008), potentially placing it as an unintended outgroup, and explaining its distance to the rest of the group.

Finally, there is an implicit assumption in this approach, the violation of which could have a large impact on the validity of its results. We are assuming that the empirical trees are correct representations of the true trees (or a close enough representation thereof). Any model could, in theory, perfectly describe the diversification dynamics of a group of organisms but would still be marked inadequate if the empirical tree we judge it by does not represent the true diversification patterns of that group. On the more specific and technical side of this argument (and of the term ‘correct’), we are assuming that there is no relevant bias on the tree shape stemming from the process that we employed to infer the tree. Even if no mistakes in the strict sense are made, biases in tree inference or divergence time estimation could have a large effect on the shape of the tree, and it seems possible that *e.g.* the use of a Yule or birth-death prior in BEAST could

bias a tree towards looking like those models would be adequate for it. And lastly, while these models account for past diversity in now extinct lineages (that are thus missing from the tree) by estimating extinction rates, it is known that getting accurate estimates is possible, but challenging (Rabosky 2010; Beaulieu & O'Meara 2015; Rabosky 2016). Thus, without knowledge and explicit inclusion of past diversity in the trees, a degree of uncertainty of the models' actual adequacy will remain.

Relation of Adequacy and Model Fit

The finding that model fit and model adequacy do not necessarily coincide – or in other words, that an adequate model does not necessarily fit the data better than an inadequate one – is important in two ways. It demonstrates that the practice of model selection based on best fit is insufficient and might lead to false results. Adding adequacy testing to the procedure facilitates the consideration of only those models that provide meaningful results. Additionally, my example suggests the utility of adequacy testing in another way: there are cases when all models implemented in a certain framework (*e.g.* a specific R package) are inadequate. Knowing this would encourage a researcher to venture out and explore the models implemented elsewhere.

References

Alfaro, M.E., Santini, F., Brock, C., Alamillo, H., Dornburg, A., Rabosky, D.L., Carnevale, G. & Harmon, L.J. (2009) Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences of the United States of America,* **106,** 13410-13414.

Barker, F.K., Cibois, A., Schikler, P., Feinstein, J. & Cracraft, J. (2004) Phylogeny and diversification of the largest avian radiation. *Proceedings of the National Academy of Sciences,* **101,** 11040-11045.

Beaulieu, J.M. & O'Meara, B.C. (2015) Extinction can be estimated from moderately sized molecular phylogenies. *Evolution,* **69,** 1036-1043.

Beaulieu, J.M. & O'Meara, B.C. (2016) Detecting Hidden Diversification Shifts in Models of Trait-Dependent Speciation and Extinction. *Systematic Biology,* **65,** 583-601.

Bouchenak-Khelladi, Y., Muasya, A.M. & Linder, H.P. (2014) A revised evolutionary history of Poales: origins and diversification. *Botanical Journal of the Linnean Society,* **175,** 4-16.

Brown, J.M. & ElDabaje, R. (2008) PuMA: Bayesian analysis of p artitioned (and u npartitioned) m odel a dequacy. *Bioinformatics,* **25,** 537-538.

Burin, G., Alencar, L.R.V., Chang, J., Alfaro, M.E. & Quental, T.B. (2019) How Well Can We Estimate Diversity Dynamics for Clades in Diversity Decline? *Systematic Biology,* **68,** 47-62.

Caetano, D.S., O'Meara, B.C. & Beaulieu, J.M. (2018) Hidden state models improve state-dependent diversification approaches, including biogeographical models. *Evolution,* **72,** 2308-2324.

FitzJohn, R.G. (2010) Quantitative Traits and Diversification. *Systematic Biology,* **59,** 619-633.

FitzJohn, R.G. (2012) Diversitree: comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution,* **3,** 1084-1092.

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

Goldberg, E.E., Lancaster, L.T. & Ree, R.H. (2011) Phylogenetic Inference of Reciprocal Effects between Geographic Range Evolution and Diversification. *Systematic Biology,* **60,** 451-465.

Hinchliff, C.E., Smith, S.A., Allman, J.F., Burleigh, J.G., Chaudhary, R., Coghill, L.M., Crandall, K.A., Deng, J., Drew, B.T., Gazis, R., Gude, K., Hibbett, D.S., Katz, L.A., Laughinghouse, H.D., McTavish, E.J., Midford, P.E., Owen, C.L., Ree, R.H., Rees, J.A., Soltis, D.E., Williams, T. & Cranston, K.A. (2015) Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proceedings of the National Academy of Sciences of the United States of America,* **112,** 12764-12769.

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

Irestedt, M., Fuchs, J., Jonsson, K.A., Ohlson, J.I., Pasquet, E. & Ericson, P.G. (2008) The systematic affinity of the enigmatic Lamprolia victoriae (Aves: Passeriformes)--an example of avian dispersal between New Guinea and Fiji over Miocene intermittent land bridges? *Molecular Phylogenetics and Evolution,* **48,** 1218-1222.

Laenen, B., Shaw, B., Schneider, H., Goffinet, B., Paradis, E., Desamore, A., Heinrichs, J., Villarreal, J.C., Gradstein, S.R., McDaniel, S.F., Long, D.G., Forrest, L.L., Hollingsworth, M.L., Crandall-Stotler, B., Davis, E.C., Engel, J., Von Konrat, M., Cooper, E.D., Patino, J., Cox, C.J., Vanderpoorten, A. & Shaw, A.J. (2014) Extant diversity of bryophytes emerged from successive post-Mesozoic diversification bursts. *Nature Communications,* **5**.

Lewitus, E. & Morlon, H. (2016a) Characterizing and Comparing Phylogenies from their Laplacian Spectrum. *Systematic Biology,* **65,** 495-507.

Lewitus, E. & Morlon, H. (2016b) Natural Constraints to Species Diversification. *Plos Biology,* **14**.

Liow, L.H., Quental, T.B. & Marshall, C.R. (2010) When can decreasing diversification rates be detected with molecular phylogenies and the fossil record? *Systematic Biology,* **59,** 646-659.

Maddison, W.P. & FitzJohn, R.G. (2015) The Unsolved Challenge to Phylogenetic Correlation Tests for Categorical Characters. *Systematic Biology,* **64,** 127-136.

Maddison, W.P., Midford, P.E. & Otto, S.P. (2007) Estimating a binary character's effect on speciation and extinction. *Systematic Biology,* **56,** 701-710.

Nguyen, V.D., Nguyen, T.H., Tayeen, A.S.M., Laughinghouse, H.D., Sánchez-Reyes, L.L., Pontelli, E., Mozzherin, D., O'Meara, B. & Stoltzfus, A. (2018) Phylotastic: improving access to tree-of-life knowledge with flexible, on-the-fly delivery of trees. *BioRxiv,* 419143.

O'Meara, B.C., Smith, S.D., Armbruster, W.S., Harder, L.D., Hardy, C.R., Hileman, L.C., Hufford, L., Litt, A., Magallón, S. & Smith, S.A. (2016) Non-equilibrium dynamics and floral trait interactions shape extant angiosperm diversity. *Proceedings of the Royal Society B: Biological Sciences,* **283,** 20152304.

Pennell, M.W., FitzJohn, R.G., Cornwell, W.K. & Harmon, L.J. (2015) Model Adequacy and the Macroevolution of Angiosperm Functional Traits. *American Naturalist,* **186,** E33-E50.

Pigot, A.L., Owens, I.P.F. & Orme, C.D.L. (2012) Speciation and Extinction Drive the Appearance of Directional Range Size Evolution in Phylogenies and the Fossil Record. *Plos Biology,* **10**.

Rabosky, D.L. (2010) Extinction rates should not be estimated from molecular phylogenies. *Evolution: International Journal of Organic Evolution,* **64,** 1816-1824.

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

Rabosky, D.L. (2016) Challenges in the estimation of extinction from molecular phylogenies: A response to Beaulieu and O'Meara. *Evolution,* **70,** 218-228.

Rabosky, D.L. & Goldberg, E.E. (2015) Model Inadequacy and Mistaken Inferences of Trait-Dependent Speciation. *Systematic Biology,* **64,** 340-355.

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

Schwery, O., Onstein, R.E., Bouchenak-Khelladi, Y., Xing, Y., Carter, R.J. & Linder, H.P. (2015) As old as the mountains: the radiations of the Ericaceae. *New Phytologist,* **207,** 355-367.

Slowinski, J.B. & Guyer, C. (1989) Testing the stochasticity of patterns of organismal diversity: an improved null model. *The American Naturalist*, **134,** 907-921.
Stadler, T. (2013) How Can We Improve Accuracy of Macroevolutionary Rate Estimates? *Systematic Biology*, **62,** 321-329.
Steeman, M.E., Hebsgaard, M.B., Fordyce, R.E., Ho, S.Y.W., Rabosky, D.L., Nielsen, R., Rahbek, C., Glenner, H., Sorensen, M.V. & Willerslev, E. (2009) Radiation of Extant Cetaceans Driven by Restructuring of the Oceans. *Systematic Biology*, **58,** 573-585.
Stoltzfus, A., Lapp, H., Matasci, N., Deus, H., Sidlauskas, B., Zmasek, C.M., Vaidya, G., Pontelli, E., Cranston, K., Vos, R., Webb, C.O., Harmon, L.J., Pirrung, M., O'Meara, B., Pennell, M.W., Mirarab, S., Rosenberg, M.S., Balhoff, J.P., Bik, H.M., Heath, T.A., Midford, P.E., Brown, J.W., McTavish, E.J., Sukumaran, J., Westneat, M., Alfaro, M.E., Steele, A. & Jordan, G. (2013) Phylotastic! Making tree-of-life knowledge accessible, reusable and convenient. *Bmc Bioinformatics*, **14**.
Xing, Y.W., Onstein, R.E., Carter, R.J., Stadler, T. & Linder, H.P. (2014) Fossils and a large molecular phylogeny show that the evolution of species richness, generic diversity, and turnover rates are disconnected. *Evolution*, **68,** 2821-2832.

Appendix C

Table III-1: Collected Empirical Phylogenies.

Tree ID given in this study, number of taxa represented, crown age, and citation of the original study that published each tree.

TreeID	Taxa	Age	Source
Emp_1	87	35.86	Steeman, M.E., Hebsgaard, M.B., Fordyce, R.E., Ho, S.Y.W., Rabosky, D.L., Nielsen, R., Rahbek, C., Glenner, H., Sorensen, M.V. & Willerslev, E. (2009) Radiation of Extant Cetaceans Driven by Restructuring of the Oceans. Systematic Biology, 58, 573-585.
Emp_2	77	117.28	Schwery, O., Onstein, R.E., Bouchenak-Khelladi, Y., Xing, Y., Carter, R.J. & Linder, H.P. (2015) As old as the mountains: the radiations of the Ericaceae. New Phytologist, 207, 355-367
Emp_3	478	122.19	Schwery, O., Onstein, R.E., Bouchenak-Khelladi, Y., Xing, Y., Carter, R.J. & Linder, H.P. (2015) As old as the mountains: the radiations of the Ericaceae. New Phytologist, 207, 355-367
Emp_4	557	124.13	Bouchenak-Khelladi, Y., Muasya, A.M. & Linder, H.P. (2014) A revised evolutionary history of Poales: origins and diversification. Botanical Journal of the Linnean Society, 175, 4-16.
Emp_5	584	118.20	Xing, Y.W., Onstein, R.E., Carter, R.J., Stadler, T. & Linder, H.P. (2014) Fossils and a large molecular phylogeny show that the evolution of species richness, generic diversity, and turnover rates are disconnected. Evolution, 68, 2821-2832
Emp_6	549	136.00	O'Meara BC, Smith SD, Armbruster WS, Harder LD, Hardy CR, Hileman LC, Hufford L, Litt A, Magallón S, Smith SA, Stevens PF. Non-equilibrium dynamics and floral trait interactions shape extant angiosperm diversity. Proceedings of the Royal Society B: Biological Sciences. 2016 May 11;283(1830):20152304.
Emp_7	309	536.16	Laenen, B., Shaw, B., Schneider, H., Goffinet, B., Paradis, E., Desamore, A., Heinrichs, J., Villarreal, J.C., Gradstein, S.R., McDaniel, S.F., Long, D.G., Forrest, L.L., Hollingsworth, M.L., Crandall-Stotler, B., Davis, E.C., Engel, J., Von Konrat, M., Cooper, E.D., Patino, J., Cox, C.J., Vanderpoorten, A. & Shaw, A.J. (2014) Extant diversity of bryophytes emerged from successive post-Mesozoic diversification bursts. Nature Communications, 5
Emp_8	9	16.17	Steeman, M.E., Hebsgaard, M.B., Fordyce, R.E., Ho, S.Y.W., Rabosky, D.L., Nielsen, R., Rahbek, C., Glenner, H., Sorensen, M.V. & Willerslev, E. (2009) Radiation of Extant Cetaceans Driven by Restructuring of the Oceans. Systematic Biology, 58, 573-585.
Emp_9	11	2.99	Pigot, A.L., Owens, I.P.F. & Orme, C.D.L. (2012) Speciation and Extinction Drive the Appearance of Directional Range Size Evolution in Phylogenies and the Fossil Record. Plos Biology, 10
Emp_10	6	5.31	Steeman, M.E., Hebsgaard, M.B., Fordyce, R.E., Ho, S.Y.W., Rabosky, D.L., Nielsen, R., Rahbek, C., Glenner, H., Sorensen, M.V. & Willerslev, E. (2009) Radiation of Extant Cetaceans Driven by Restructuring of the Oceans. Systematic Biology, 58, 573-585.
Emp_11	150	30.69	Steeman, M.E., Hebsgaard, M.B., Fordyce, R.E., Ho, S.Y.W., Rabosky, D.L., Nielsen, R., Rahbek, C., Glenner, H., Sorensen, M.V. & Willerslev, E. (2009) Radiation of Extant Cetaceans Driven by Restructuring of the Oceans. Systematic Biology, 58, 573-585.

Table III-1 Continued

TreeID	Taxa	Age	Source
Emp_12	52	50.50	Selvatti, Alexandre Pedro, Ana Galvao, Anieli Guirro Pereira, Luiz Pedreira Gonzaga, Claudia Augusta de Moraes Russo. 2016. An African origin of the Eurylaimides (Passeriformes) and the successful diversification of the ground-foraging pittas (Pittidae). Molecular Biology and Evolution, p. msw250
Emp_13	118	78.49	Fine, Paul V. A., Felipe Zapata, Douglas C. Daly. 2014. Investigating processes of neotropical rain forest tree diversification by examining the evolution and historical biogeography of the Protieae (Burseraceae). Evolution 68 (7): 1988-2004
Emp_14	76	28.48	Arbabi, Tayebeh, Javier Gonzalez, Michael Wink. 2014. A re-evaluation of phylogenetic relationships within reed warblers (Aves: Acrocephalidae) based on eight molecular loci and ISSR profiles. Molecular Phylogenetics and Evolution 78: 304-313.
Emp_15	174	3.87	Ralph S. Peters, Lars Krogmann, Christoph Mayer, Alexander Donath, Simon Gunkel, Karen Meusemann, Alexey Kozlov, Lars Podsiadlowski, Malte Petersen, Robert Lanfear, Patricia A. Diez, John Heraty, Karl M. Kjer, Seraina Klopfstein, Rudolf Meier, Carlo Polidori, Thomas Schmitt, Shanlin Liu, Xin Zhou, Torsten Wappler, Jes Rust, Bernhard Misof, Oliver Niehuis, 2017, 'Evolutionary History of the Hymenoptera', Current Biology
Emp_16	148	425.00	A. E. Syme, T. H. Oakley, 2011, 'Dispersal between Shallow and Abyssal Seas and Evolutionary Loss and Regain of Compound Eyes in Cylindroleberidid Ostracods: Conflicting Conclusions from Different Comparative Methods', Systematic Biology, vol. 61, no. 2, pp. 314-336
Emp_17	103	72.81	Daniela Campanella, Lily C. Hughes, Peter J. Unmack, Devin D. Bloom, Kyle R. Piller, Guillermo Orti, 2015, 'Multi-locus fossil-calibrated phylogeny of Atheriniformes (Teleostei, Ovalentaria)', Molecular Phylogenetics and Evolution, vol. 86, pp. 8-23
Emp_18	68	49.71	Upham, N.S. & B.D. Patterson. 2015. Evolution of the caviomorph rodents: a complete phylogeny and timetree of living genera. Pp. 63-120 In: Biology of caviomorph rodents: diversity and evolution (A. I. Vassallo & D. Antenucci, eds.). SAREM Series A, Buenos Aires.
Emp_19	78	271.08	Vea, Isabelle M., David A. Grimaldi. 2016. Putting scales into evolutionary time: the divergence of major scale insect lineages (Hemiptera) predates the radiation of modern angiosperm hosts. Scientific Reports, 6: 23487 .
Emp_20	88	15.20	Hugall, Andrew F., Devi Stuart-Fox. 2012. Accelerated speciation in colour-polymorphic birds. Nature 485 (7400): 631-634.
Emp_21	111	30.22	Hugall, Andrew F., Devi Stuart-Fox. 2012. Accelerated speciation in colour-polymorphic birds. Nature 485 (7400): 631-634.
Emp_22	198	39.87	Hugall, Andrew F., Devi Stuart-Fox. 2012. Accelerated speciation in colour-polymorphic birds. Nature 485 (7400): 631-634.
Emp_23	53	27.86	Hugall, Andrew F., Devi Stuart-Fox. 2012. Accelerated speciation in colour-polymorphic birds. Nature 485 (7400): 631-634.
Emp_24	181	24.22	Hugall, Andrew F., Devi Stuart-Fox. 2012. Accelerated speciation in colour-polymorphic birds. Nature 485 (7400): 631-634.
Emp_25	89	125.75	Brown, Joseph W., Robert B. Payne, David P. Mindell. 2007. Comment. Nuclear DNA does not reconcile ‘rocks’ and ‘clocks’ in Neoaves: a comment on Ericson et al. Biology Letters 3 (3): 257-259.
Emp_26	75	30.61	Chris J. Law, Graham J. Slater, Rita S. Mehta, 2017, 'Lineage Diversity and Size Disparity in Musteloidea: Testing Patterns of Adaptive Radiation Using Molecular and Fossil-Based Methods', Systematic Biology
Emp_27	133	53.26	Andersen, Michael J., Jenna M. McCullough, William M. Mauck, Brian Tilston Smith, Robert G. Moyle. 2017. A phylogeny of kingfishers reveals an Indomalayan origin and elevated rates of diversification on oceanic islands. Journal of Biogeography

Table III-1 Continued

TreeID	Taxa	Age	Source
Emp_28	68	0.57	Ronquist, Fredrik, Seraina Klopfstein, Lars Vilhelmsen, Susanne Schulmeister, Debra L. Murray, Alexandr P. Rasnitsyn. 2012. A Total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. Systematic Biology, 61(6): 973-999.
Emp_29	68	0.27	Ronquist, Fredrik, Seraina Klopfstein, Lars Vilhelmsen, Susanne Schulmeister, Debra L. Murray, Alexandr P. Rasnitsyn. 2012. A Total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. Systematic Biology, 61(6): 973-999.
Emp_30	68	350.56	Ronquist, Fredrik, Seraina Klopfstein, Lars Vilhelmsen, Susanne Schulmeister, Debra L. Murray, Alexandr P. Rasnitsyn. 2012. A Total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. Systematic Biology, 61(6): 973-999.
Emp_31	68	388.88	Ronquist, Fredrik, Seraina Klopfstein, Lars Vilhelmsen, Susanne Schulmeister, Debra L. Murray, Alexandr P. Rasnitsyn. 2012. A Total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. Systematic Biology, 61(6): 973-999.
Emp_32	68	0.33	Ronquist, Fredrik, Seraina Klopfstein, Lars Vilhelmsen, Susanne Schulmeister, Debra L. Murray, Alexandr P. Rasnitsyn. 2012. A Total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. Systematic Biology, 61(6): 973-999.
Emp_33	68	0.37	Ronquist, Fredrik, Seraina Klopfstein, Lars Vilhelmsen, Susanne Schulmeister, Debra L. Murray, Alexandr P. Rasnitsyn. 2012. A Total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. Systematic Biology, 61(6): 973-999.
Emp_34	48	101.64	Jarvis, E. D., S. Mirarab, A. J. Aberer, B. Li, P. Houde, C. Li, S. Y. W. Ho, B. C. Faircloth, B. Nabholz, J. T. Howard, A. Suh, C. C. Weber, R. R. da Fonseca, J. Li, F. Zhang, H. Li, L. Zhou, N. Narula, L. Liu, G. Ganapathy, B. Boussau, M. S. Bayzid, V. Zavidovych, S. Subramanian, T. Gabaldon, S. Capella-Gutierrez, J. Huerta-Cepas, B. Rekepalli, K. Munch, M. Schierup, B. Lindow, W. C. Warren, D. Ray, R. E. Green, M. W. Bruford, X. Zhan, A. Dixon, S. Li, N. Li, Y. Huang, E. P. Derryberry, M. F. Bertelsen, F. H. Sheldon, R. T. Brumfield, C. V. Mello, P. V. Lovell, M. Wirthlin, M. P. C. Schneider, F. Prosdocimi, J. A. Samaniego, A. M. V. Velazquez, A. Alfaro-Nunez, P. F. Campos, B. Petersen, T. Sicheritz-Ponten, A. Pas, T. Bailey, P. Scofield, M. Bunce, D. M. Lambert, Q. Zhou, P. Perelman, A. C. Driskell, B. Shapiro, Z. Xiong, Y. Zeng, S. Liu, Z. Li, B. Liu, K. Wu, J. Xiao, X. Yinqi, Q. Zheng, Y. Zhang, H. Yang, J. Wang, L. Smeds, F. E. Rheindt, M. Braun, J. Fjeldsa, L. Orlando, F. K. Barker, K. A. Jonsson, W. Johnson, K.-P. Koepfli, S. O'Brien, D. Haussler, O. A. Ryder, C. Rahbek, E. Willerslev, G. R. Graves, T. C. Glenn, J. McCormack, D. Burt, H. Ellegren, P. Alstrom, S. V. Edwards, A. Stamatakis, D. P. Mindell, J. Cracraft, E. L. Braun, T. Warnow, W. Jun, M. T. P. Gilbert, G. Zhang. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science 346 (6215): 1320-1331.
Emp_35	95	97.11	Dahiana Arcila, R. Alexander Pyron, James C. Tyler, Guillermo Orti, Ricardo Betancur-R., 2015, 'An evaluation of fossil tip-dating versus node-age calibrations in tetraodontiform fishes (Teleostei: Percomorphaceae)', Molecular Phylogenetics and Evolution, vol. 82, pp. 131-145
Emp_36	43	390.58	Dornburg, Alex, Jeffrey P Townsend, Matt Friedman, Thomas J Near. 2014. Phylogenetic informativeness reconciles ray-finned fish molecular divergence times. BMC Evolutionary Biology 14(1): 169.

Table III-1 Continued

TreeID	Taxa	Age	Source
Emp_37	43	394.30	Dornburg, Alex, Jeffrey P Townsend, Matt Friedman, Thomas J Near. 2014. Phylogenetic informativeness reconciles ray-finned fish molecular divergence times. BMC Evolutionary Biology 14(1): 169.
Emp_38	45	26.84	Schweizer, Manuel, Timothy F. Wright, Joshua V. Penalba, Erin E. Schirtzinger, Leo Joseph. 2015. Molecular phylogenetics suggests a New Guinean origin and frequent episodes of founder-event speciation in the nectarivorous lories and lorikeets (Aves: Psittaciformes). Molecular Phylogenetics and Evolution 90: 34-48.
Emp_39	83	138.30	Sean G Brady, Brian L Fisher, Ted R Schultz, Philip S Ward, 2014, 'The rise of army ants and their relatives: diversification of specialized predatory doryline ants', BMC Evolutionary Biology, vol. 14, no. 1, p. 93
Emp_40	32	86.54	Garcia-R, Juan C., Gillian C. Gibb, Steve A. Trewick. 2014. Eocene diversification of crown group rails (Aves: Gruiformes: Rallidae). PLoS ONE 9 (10): e109635
Emp_41	124	514.26	Beaulieu, Jeremy M., Brian C. O'Meara, Peter Crane, Michael J. Donoghue. 2015. Heterogeneous rates of molecular evolution and diversification could explain the Triassic age estimate for Angiosperms. Systematic Biology 64 (5): 869-878
Emp_42	82	48.25	K. M. Kozak, N. Wahlberg, A. F. E. Neild, K. K. Dasmahapatra, J. Mallet, C. D. Jiggins, 2015, 'Multilocus Species Trees Show the Recent Adaptive Radiation of the Mimetic Heliconius Butterflies', Systematic Biology, vol. 64, no. 3, pp. 505-524
Emp_43	198	78.35	Prum, Richard O., Jacob S. Berv, Alex Dornburg, Daniel J. Field, Jeffrey P. Townsend, Emily Moriarty Lemmon, Alan R. Lemmon. 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. Nature 526, (7574): 569-573
Emp_44	198	72.91	Prum, Richard O., Jacob S. Berv, Alex Dornburg, Daniel J. Field, Jeffrey P. Townsend, Emily Moriarty Lemmon, Alan R. Lemmon. 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. Nature 526, (7574): 569-573
Emp_45	180	56.07	Dufort, Matthew J. 2015. An augmented supermatrix phylogeny of the avian family Picidae reveals uncertainty deep in the family tree. Molecular Phylogenetics and Evolution
Emp_46	54	750.70	dos Reis, Mario, Yuttapong Thawornwattana, Konstantinos Angelis, Maximilian J. Telford, Philip C.J. Donoghue, Ziheng Yang. 2015. Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. Current Biology
Emp_47	54	750.76	dos Reis, Mario, Yuttapong Thawornwattana, Konstantinos Angelis, Maximilian J. Telford, Philip C.J. Donoghue, Ziheng Yang. 2015. Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. Current Biology
Emp_48	54	819.73	dos Reis, Mario, Yuttapong Thawornwattana, Konstantinos Angelis, Maximilian J. Telford, Philip C.J. Donoghue, Ziheng Yang. 2015. Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. Current Biology
Emp_49	54	819.58	dos Reis, Mario, Yuttapong Thawornwattana, Konstantinos Angelis, Maximilian J. Telford, Philip C.J. Donoghue, Ziheng Yang. 2015. Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. Current Biology

Table III-1 Continued

TreeID	Taxa	Age	Source
Emp_50	54	776.48	dos Reis, Mario, Yuttapong Thawornwattana, Konstantinos Angelis, Maximilian J. Telford, Philip C.J. Donoghue, Ziheng Yang. 2015. Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. Current Biology
Emp_51	54	776.34	dos Reis, Mario, Yuttapong Thawornwattana, Konstantinos Angelis, Maximilian J. Telford, Philip C.J. Donoghue, Ziheng Yang. 2015. Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. Current Biology
Emp_52	54	813.44	dos Reis, Mario, Yuttapong Thawornwattana, Konstantinos Angelis, Maximilian J. Telford, Philip C.J. Donoghue, Ziheng Yang. 2015. Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. Current Biology
Emp_53	54	813.69	dos Reis, Mario, Yuttapong Thawornwattana, Konstantinos Angelis, Maximilian J. Telford, Philip C.J. Donoghue, Ziheng Yang. 2015. Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. Current Biology
Emp_54	88	68.80	Cushing, Paula E., Matthew R. Graham, Lorenzo Prendini, Jack O. Brookhart. 2015. A multilocus molecular phylogeny of the endemic North American camel spider family Eremobatidae (Arachnida: Solifugae). Molecular Phylogenetics and Evolution 92: 280-293
Emp_55	187	97.76	Toussaint, Emmanuel F. A., Lars Hendrich, Helena Shaverdo, Michael Balke. 2015. Mosaic patterns of diversification dynamics following the colonization of Melanesian islands. Scientific Reports 5: 16016
Emp_56	86	90.74	McCord, Charlene L., Mark W. Westneat. 2016. Phylogenetic relationships and the evolution of BMP4 in triggerfishes and filefishes (Balistoidea). Molecular Phylogenetics and Evolution 94: 397-409
Emp_57	48	96.57	Claramunt, Santiago, Joel Cracraft. 2015. A new time tree reveals Earth historys imprint on the evolution of modern birds. Science Advances 1 (11): e1501005-e1501005
Emp_58	102	62.58	Gibb, Gillian C., Ryan England, Gerrit Hartig, P.A. (Trish) McLenachan, Briar L. Taylor Smith, Bennet J. McComish, Alan Cooper, David Penny. 2015. New Zealand passerines help clarify the diversification of major songbird lineages during the Oligocene. Genome Biology and Evolution 7 (11): 2983-2995.
Emp_59	72	81.00	Ksepka, Daniel T., Matthew J. Phillips. 2015. Avian diversification patterns across the K-Pg boundary: influence of calibrations, datasets, and model misspecification. Annals of the Missouri Botanical Garden 100 (4): 300-328
Emp_60	72	81.00	Ksepka, Daniel T., Matthew J. Phillips. 2015. Avian diversification patterns across the K-Pg boundary: influence of calibrations, datasets, and model misspecification. Annals of the Missouri Botanical Garden 100 (4): 300-328
Emp_61	72	104.07	Ksepka, Daniel T., Matthew J. Phillips. 2015. Avian diversification patterns across the K-Pg boundary: influence of calibrations, datasets, and model misspecification. Annals of the Missouri Botanical Garden 100 (4): 300-328
Emp_62	72	97.63	Ksepka, Daniel T., Matthew J. Phillips. 2015. Avian diversification patterns across the K-Pg boundary: influence of calibrations, datasets, and model misspecification. Annals of the Missouri Botanical Garden 100 (4): 300-328
Emp_63	72	111.90	Ksepka, Daniel T., Matthew J. Phillips. 2015. Avian diversification patterns across the K-Pg boundary: influence of calibrations, datasets, and model misspecification. Annals of the Missouri Botanical Garden 100 (4): 300-328

Table III-1 Continued

TreeID	Taxa	Age	Source
Emp_64	72	111.07	Ksepka, Daniel T., Matthew J. Phillips. 2015. Avian diversification patterns across the K-Pg boundary: influence of calibrations, datasets, and model misspecification. Annals of the Missouri Botanical Garden 100 (4): 300-328
Emp_65	72	112.80	Ksepka, Daniel T., Matthew J. Phillips. 2015. Avian diversification patterns across the K-Pg boundary: influence of calibrations, datasets, and model misspecification. Annals of the Missouri Botanical Garden 100 (4): 300-328
Emp_66	72	85.96	Ksepka, Daniel T., Matthew J. Phillips. 2015. Avian diversification patterns across the K-Pg boundary: influence of calibrations, datasets, and model misspecification. Annals of the Missouri Botanical Garden 100 (4): 300-328
Emp_67	72	103.89	Ksepka, Daniel T., Matthew J. Phillips. 2015. Avian diversification patterns across the K-Pg boundary: influence of calibrations, datasets, and model misspecification. Annals of the Missouri Botanical Garden 100 (4): 300-328
Emp_68	72	81.35	Ksepka, Daniel T., Matthew J. Phillips. 2015. Avian diversification patterns across the K-Pg boundary: influence of calibrations, datasets, and model misspecification. Annals of the Missouri Botanical Garden 100 (4): 300-328
Emp_69	72	80.99	Ksepka, Daniel T., Matthew J. Phillips. 2015. Avian diversification patterns across the K-Pg boundary: influence of calibrations, datasets, and model misspecification. Annals of the Missouri Botanical Garden 100 (4): 300-328
Emp_70	72	106.29	Ksepka, Daniel T., Matthew J. Phillips. 2015. Avian diversification patterns across the K-Pg boundary: influence of calibrations, datasets, and model misspecification. Annals of the Missouri Botanical Garden 100 (4): 300-328
Emp_71	72	94.14	Ksepka, Daniel T., Matthew J. Phillips. 2015. Avian diversification patterns across the K-Pg boundary: influence of calibrations, datasets, and model misspecification. Annals of the Missouri Botanical Garden 100 (4): 300-328
Emp_72	72	81.08	Ksepka, Daniel T., Matthew J. Phillips. 2015. Avian diversification patterns across the K-Pg boundary: influence of calibrations, datasets, and model misspecification. Annals of the Missouri Botanical Garden 100 (4): 300-328
Emp_73	72	81.31	Ksepka, Daniel T., Matthew J. Phillips. 2015. Avian diversification patterns across the K-Pg boundary: influence of calibrations, datasets, and model misspecification. Annals of the Missouri Botanical Garden 100 (4): 300-328
Emp_74	72	81.01	Ksepka, Daniel T., Matthew J. Phillips. 2015. Avian diversification patterns across the K-Pg boundary: influence of calibrations, datasets, and model misspecification. Annals of the Missouri Botanical Garden 100 (4): 300-328
Emp_75	72	80.83	Ksepka, Daniel T., Matthew J. Phillips. 2015. Avian diversification patterns across the K-Pg boundary: influence of calibrations, datasets, and model misspecification. Annals of the Missouri Botanical Garden 100 (4): 300-328
Emp_76	35	104.99	Delsuc, Frederic, Gillian C. Gibb, Melanie Kuch, Guillaume Billet, Lionel Hautier, John Southon, Jean-Marie Rouillard, Juan Carlos Fernicola, Sergio F. Vizcaino, Ross D.E. MacPhee, Hendrik N. Poinar. 2016. The phylogenetic affinities of the extinct glyptodonts. Current Biology 26 (4): R155-R156
Emp_77	40	102.78	Gibb, Gillian C., Fabien L. Condamine, Melanie Kuch, Jacob Enk, Nadia Moraes-Barros, Mariella Superina, Hendrik N. Poinar, Frederic Delsuc. 2015. Shotgun mitogenomics provides a reference phylogenetic framework and timescale for living Xenarthrans. Molecular Biology and Evolution 33 (3): 621-642
Emp_78	100	78.41	A. C. Schneider, W. A. Freyman, C. M. Guilliams, Y. P. Springer, B. G. Baldwin, 2016, 'Pleistocene radiation of the serpentine-adapted genus Hesperolinon and other divergence times in Linaceae (Malpighiales)', American Journal of Botany, vol. 103, no. 2, pp. 221-232

Table III-1 Continued

TreeID	Taxa	Age	Source
Emp_79	32	4.37	Joel P. Olfelt, William A. Freyman, 2014, ' Relationships of North American members of Rhodiola (Crassulaceae) ', Botany, vol. 92, no. 12, pp. 901-910
Emp_80	94	54.58	Zhi-Yong Yuan, Wei-Wei Zhou, Xin Chen, Nikolay A. Poyarkov, Hong-Man Chen, Nian-Hong Jang-Liaw, Wen-Hao Chou, Nicholas J. Matzke, Koji Iizuka, Mi-Sook Min, Sergius L. Kuzmin, Ya-Ping Zhang, David C. Cannatella, David M. Hillis, Jing Che, 2016, 'Spatiotemporal Diversification of the True Frogs (Genus Rana): A Historical Framework for a Widely Studied Group of Model Organisms', Systematic Biology, p. syw055
Emp_81	93	83.32	Wang, Ning, Rebecca T. Kimball, Edward L. Braun, Bin Liang, Zhengwang Zhang. 2016. Ancestral range reconstruction of Galliformes: the effects of topology and taxon sampling. Journal of Biogeography
Emp_82	55	73.64	Ericson, Per GP, Seraina Klopfstein, Martin Irestedt, Jacqueline MT Nguyen and Johan AA Nylander. 2014. Dating the diversification of the major lineages of Passeriformes (Aves). BMC Evolutionary Biology, 14(8).
Emp_83	34	0.53	Slager, David L., C.J. Battey, Robert W. Bryson, Gary Voelker, John Klicka. 2014. A multilocus phylogeny of a major New World avian radiation: The Vireonidae. Molecular Phylogenetics and Evolution 80: 95-104
Emp_84	34	0.60	Slager, David L., C.J. Battey, Robert W. Bryson, Gary Voelker, John Klicka. 2014. A multilocus phylogeny of a major New World avian radiation: The Vireonidae. Molecular Phylogenetics and Evolution 80: 95-104
Emp_85	36	21.90	Sweet, Andrew D., Kevin P. Johnson. 2015. Patterns of diversification in small New World ground doves are consistent with major geologic events. The Auk 132 (1): 300-312.
Emp_86	76	223.77	Martin Malmstrom, Michael Matschiner, Ole K Torresen, Bastiaan Star, Lars G Snipen, Thomas F Hansen, Helle T Baalsrud, Alexander J Nederbragt, Reinhold Hanel, Walter Salzburger, Nils C Stenseth, Kjetill S Jakobsen, Sissel Jentoft, 2016, 'Evolution of the immune system influences speciation rates in teleost fishes', Nature Genetics
Emp_87	64	0.17	Wood, Jamie R., Kieren J. Mitchell, R. Paul Scofield, Vanesa L. De Pietri, Nicolas J. Rawlence, Alan Cooper. 2016. Phylogenetic relationships and terrestrial adaptations of the extinct laughing owl, Sceloglaux albifacies (Aves: Strigidae). Zoological Journal of the Linnean Society
Emp_88	106	4.43	Moyle, Robert G., Carl H. Oliveros, Michael J. Andersen, Peter A. Hosner, Brett W. Benz, Joseph D. Manthey, Scott L. Travers, Rafe M. Brown, Brant C. Faircloth. 2016. Tectonic collision and uplift of Wallacea triggered the global songbird radiation. Nature Communications 7: 12709
Emp_89	51	2.43	Fuchs, Jerome, Jan I. Ohlson, Per G. P. Ericson, Eric Pasquet. 2007. Synchronous intercontinental splits between assemblages of woodpeckers suggested by molecular data. Zoologica Scripta 36 (1): 11-25
Emp_90	33	126.66	Johnson, Jeff A., Joseph W. Brown, Jerome Fuchs, David P. Mindell, 2016, 'Multi-locus phylogenetic inference among New World Vultures (Aves: Cathartidae)', Molecular Phylogenetics and Evolution, vol. 105, pp. 193-199
Emp_91	33	126.76	Johnson, Jeff A., Joseph W. Brown, Jerome Fuchs, David P. Mindell, 2016, 'Multi-locus phylogenetic inference among New World Vultures (Aves: Cathartidae)', Molecular Phylogenetics and Evolution, vol. 105, pp. 193-199
Emp_92	195	339.35	Foster, Charles S. P., Herve Sauquet, Marlien van der Merwe, Hannah McPherson, Maurizio Rossetto, Simon Y. W. Ho. 2016. Evaluating the impact of genomic data and priors on Bayesian estimates of the angiosperm evolutionary timescale. Systematic Biology, p. 66(3)

Table III-1 Continued

TreeID	Taxa	Age	Source
Emp_93	107	17.11	Scofield, R. Paul, Kieren J. Mitchell, Jamie R. Wood, Vanesa L. De Pietri, Scott Jarvie, Bastien Llamas, Alan Cooper. 2016. The Origin and Phylogenetic Relationships of the New Zealand Ravens. Molecular Phylogenetics and Evolution 106: 136-143
Emp_94	84	65.00	Jennifer L. Fessler, Mark W. Westneat, 2007, 'Molecular phylogenetics of the butterflyfishes (Chaetodontidae): Taxonomy and biogeography of a global coral reef fish family', Molecular Phylogenetics and Evolution, vol. 45, no. 1, pp. 50-68
Emp_95	55	151.42	Harrington, Richard C., Brant C. Faircloth, Ron I. Eytan, W. Leo Smith, Thomas J. Near, Michael E. Alfaro, Matt Friedman. 2016. Phylogenomic analysis of carangimorph fishes reveals flatfish asymmetry arose in a blink of the evolutionary eye. BMC Evolutionary Biology 16: 224
Emp_96	108	56.36	Lohmann, L. G., Bell, C. D., Calio, M. F., & Winkworth, R. C. (2013). Pattern and timing of biogeographical history in the Neotropical tribe Bignonieae (Bignoniaceae). Botanical Journal of the Linnean Society, 171(1), 154-170.
Emp_97	36	43.40	Higdon J. W., Bininda-Emonds O. R. P., Beck R. M. D., Ferguson S. H. 2007. Phylogeny and divergence of the pinnipeds (Carnivora: Mammalia) assessed using a multigene dataset. BMC Evolutionary Biology 8: 216.
Emp_98	36	43.40	Higdon J. W., Bininda-Emonds O. R. P., Beck R. M. D., Ferguson S. H. 2007. Phylogeny and divergence of the pinnipeds (Carnivora: Mammalia) assessed using a multigene dataset. BMC Evolutionary Biology 8: 216.
Emp_99	36	43.40	Higdon J. W., Bininda-Emonds O. R. P., Beck R. M. D., Ferguson S. H. 2007. Phylogeny and divergence of the pinnipeds (Carnivora: Mammalia) assessed using a multigene dataset. BMC Evolutionary Biology 8: 216.
Emp_100	73	144.34	Uit de weerd D.R., & Gittenberger E. 2013. Phylogeny of the land snail family Clausiliidae (Gastropoda: Pulmonata). Molecular Phylogenetics and Evolution, 67(1): 201-216.
Emp_101	157	48.96	Hubert N., Paradis E., Bruggemann H., & Planes S. 2011. Community assembly and diversification in Indo-Pacific coral reef fishes. Ecology and Evolution, 1(3): 229-277.
Emp_102	37	0.17	de Sa R.O., Streicher J.W., Selenkoyela R., Forlani M.C., Loader S.P., Greenbaum E., Richards S., & Haddad C. 2012. Molecular phylogeny of microhylid frogs (Anura: Microhylidae) with emphasis on relationships among New World genera. BMC Evolutionary Biology 12: 241 .
Emp_103	68	39.92	Valtuena, Francisco J., Chris D. Preston, and Joachim W. Kadereit. "Phylogeography of a Tertiary relict plant, Meconopsis cambrica (Papaveraceae), implies the existence of northern refugia for a temperate herb." Molecular Ecology 21.6 (2012): 1423-1437.
Emp_104	87	35.86	Steeman, M., Hebsgaard M., Fordyce R., Ho S., Rabosky D., Nielsen R., Rahbek C., Glenner H., Sorensen M., & Willerslev E. 2009. Radiation of Extant Cetaceans Driven by Restructuring of the Oceans. Systematic Biology 58 (6): 573-585.
Emp_105	171	49.38	Lack J.B., & Van den bussche R.A. 2010. Identifying the Confounding Factors in Resolving Phylogenetic Relationships in Vespertilionidae. Journal of Mammalogy, .
Emp_106	154	41.92	Dumont E.R., Davalos L.M., Goldberg A., Santana S.E., Rex K., & Voigt C.C. 2012. Morphological innovation, diversification and invasion of a new adaptive zone. Proceedings of the Royal Society B: Biological Sciences, 279: 1797-1805.
Emp_107	32	28.07	Chan, L.M., Choi D., Raselimanana A.P., Rakotondravony H.E., & Yoder A.D. 2012. Defining spatial and temporal patterns of phylogeographic structure in Madagascar's iguanid lizards (genus Oplurus). Molecular Ecology 21 (15): 3839-3851.

Table III-1 Continued

TreeID	Taxa	Age	Source
Emp_108	34	76.40	Andujar, C., Serrano J., & Gomez-zurita J. 2012. Winding up the molecular clock in the genus Carabus (Coleoptera: Carabidae): assessment of methodological decisions on rate and node age estimation. BMC Evolutionary Biology 12: 40.
Emp_109	144	395.07	K. Mao, R. I. Milne, L. Zhang, Y. Peng, J. Liu, P. Thomas, R. R. Mill, S. S. Renner. 2012. Distribution of living Cupressaceae reflects the breakup of Pangea. Proceedings of the National Academy of Sciences 109(20):7793-7798.
Emp_110	191	21.00	Barker, F. K., K. J. Burns, J. Klicka, S. M. Lanyon, I. J. Lovette. 2013. Going to extremes: contrasting rates of diversification in a recent radiation of New World passerine birds. Systematic Biology 62 (2): 298-320.
Emp_111	115	100.00	Lovette, Irby J., Jorge L. Perez-Eman, John P. Sullivan, Richard C. Banks, Isabella Fiorentino, Sergio Cordoba-Cordoba, Maria Echeverry-Galvis, F. Keith Barker, Kevin J. Burns, John Klicka, Scott M. Lanyon, Eldredge Bermingham. 2010. A comprehensive multilocus phylogeny for the wood-warblers and a revised classification of the Parulidae (Aves). Molecular Phylogenetics and Evolution 57 (2): 753-770.
Emp_112	63	187.94	Sikes, D.S., & Venables C. 2013. Molecular phylogeny of the burying beetles (Coleoptera: Silphidae: Nicrophorinae). Molecular Phylogenetics and Evolution 69 (3): 552-565.
Emp_113	48	90.84	Gibb, Gillian C., Martyn Kennedy, David Penny. 2013. Beyond phylogeny: pelecaniform and ciconiiform birds, and long-term niche stability. Molecular Phylogentics and Evolution 68 (2): 229-238.
Emp_114	176	90.38	Bremer, B., & Eriksson, O. 2009. Time tree of Rubiaceae: Phylogeny and dating the family, subfamilies and tribes. Internat. J. Plant Sci. 170: 766-793.
Emp_115	44	81.74	Chakrabarty, Prosanta, Matthew P. Davis, W. Leo Smith, Zachary H. Baldwin, John S. Sparks. 2011. Is sexual selection driving diversification of the bioluminescent ponyfishes (Teleostei: Leiognathidae)? Molecular Ecology 20 (13): 2818-2834.
Emp_116	154	122.75	Friedman, M., B. P. Keck, A. Dornburg, R. I. Eytan, C. H. Martin, C. D. Hulsey, P. C. Wainwright, T. J. Near. 2013. Molecular and fossil evidence place the origin of cichlid fishes long after Gondwanan rifting. Proceedings of the Royal Society B: Biological Sciences 280 (1770): 20131733
Emp_117	95	349.79	Chen, Wei-Jen, Sebastien Lavoue, and Richard L. Mayden. 2013. Evolutionary origin and early biogeography of otophysan fishes (Ostariophysi: Teleostei). Evolution 67 (8): 2218-2239.
Emp_118	44	21.08	Koepfli, Klaus-Peter, Kerry A Deere, Graham J Slater, Colleen Begg, Keith Begg, Lon Grassman, Mauro Lucherini, Geraldine Veron, Robert K Wayne. 2008. Multigene phylogeny of the Mustelidae: Resolving relationships, tempo and biogeographic history of a mammalian adaptive radiation. BMC Biology 6 (1): 10.
Emp_119	53	556.63	Lee, M. S., Soubrier, J., & Edgecombe, G. D. (2013). Rates of phenotypic and genomic evolution during the Cambrian explosion. Current Biology 23 (19): 1889-1895.
Emp_120	137	16.64	Drummond, C. S., Eastwood, R. J., Miotto, S. T., & Hughes, C. E. (2012). Multiple continental radiations and correlates of diversification in Lupinus (Leguminosae): testing for key innovation with incomplete taxon sampling. Systematic Biology 61(3), 443-460.
Emp_121	45	74.27	Aggerbeck, Marie, Jon Fjeldsa, Les Christidis, Pierre-Henri Fabre, Knud Andreas Jonsson. 2014. Resolving deep lineage divergences in core corvoid passerine birds supports a proto-Papuan island origin. Molecular Phylogenetics and Evolution 70: 272-285.

Table III-1 Continued

TreeID	Taxa	Age	Source
Emp_122	109	1781.09	Parfrey, L. W., D. J. G. Lahr, A. H. Knoll, L. A. Katz. 2011. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. Proceedings of the National Academy of Sciences 108 (33): 13624-13629.
Emp_123	93	54.58	Smith, SA. 2009. Taking into account phylogenetic and divergence-time uncertainty in a parametric biogeographical analysis of the Northern Hemisphere plant clade Caprifolieae. Journal of Biogeography.
Emp_124	55	47.27	Cibois, Alice, Jean-Claude Thibault, Caroline Bonillo, Christopher E. Filardi, Dick Watling, Eric Pasquet. 2014. Phylogeny and biogeography of the fruit doves (Aves: Columbidae). Molecular Phylogenetics and Evolution 70: 442-453.
Emp_125	44	45.48	Drew, Bryan T., Kenneth J. Sytsma. 2013. The South American radiation of Lepechinia (Lamiaceae): phylogenetics, divergence times and evolution of dioecy. Botanical Journal of the Linnean Society 171 (1): 171-190.
Emp_126	100	50.00	Mahler, D. L., T. Ingram, L. J. Revell, J. B. Losos. 2013. Exceptional Convergence on the Macroevolutionary Landscape in Island Lizard Radiations. Science 341 (6143): 292-295.
Emp_127	176	96.13	Price, Trevor D., Daniel M. Hooper, Caitlyn D. Buchanan, Ulf S. Johansson, D. Thomas Tietze, Per Alstrom, Urban Olsson, Mousumi Ghosh-Harihar, Farah Ishtiaq, Sandeep K. Gupta, Jochen Martens, Bettina Harr, Pratap Singh, Dhananjai Mohan. 2014. Niche filling slows the diversification of Himalayan songbirds. Nature 509: 222-225.
Emp_128	34	244.52	Mitchell, K. J., B. Llamas, J. Soubrier, N. J. Rawlence, T. H. Worthy, J. Wood, M. S. Y. Lee, A. Cooper. 2014. Ancient DNA reveals elephant birds and kiwi are sister taxa and clarifies ratite bird evolution. Science 344 (6186): 898-900.
Emp_129	108	92.98	Antonelli, A. 2009. Have giant lobelias evolved several times independently? Life form shifts and historical biogeography of the cosmopolitan and highly diverse subfamily Loeblioideae (Campanulaceae). BMC Biology 7: 82 - see http://www.biomedcentral.com/1741-7007/7/82
Emp_130	73	43.00	Bergh N.G., & Linder H.P. 2009. Cape diversification and repeated out-of-southern-Africa dispersal in paper daisies. Molecular Phylogenetics and Evolution, 51(Special issue: origins and evolution of a biodiversity hotspot, the biodiversity of the African Cape Floristic Region): 5 - 18.
Emp_131	116	100.00	Johnson, M. T., FitzJohn, R. G., Smith, S. D., Rausher, M. D., & Otto, S. P. (2011). Loss of sexual recombination and segregation is associated with increased diversification in evening primroses. Evolution, 65(11), 3230-3240.

Table III-2: Model Inadequacy by Model.

Numbers and percentages of trees for which a particular model was adequate or inadequate for, assessed using either Bonferroni-corrected p-values (above) or 2D convex-hulls (below). BD=constant rate birth-death, T_ex_co=time-dependent birth-death with exponential λ and constant μ, T_co_ex=time-dependent birth-death with constant λ and exponential μ. DD_lin_co=diversity-dependent birth-death with linear λ and constant μ. DD_lin_lin=diversity-dependent birth-death with linear λ and linear μ.

Model	Yule	BD	T_ex_co	T_co_ex	DD_lin_co	DD_lin_lin
Bonferroni-Pvalues						
Adequate	201	209	199	193	201	171
Inadequate	33	25	35	41	33	63
% adequate	85.897	89.316	85.043	82.479	85.897	73.077
% inadequate	14.103	10.684	14.957	17.521	14.103	26.923
2D convex-hulls						
Adequate	201	205	183	185	203	187
Inadequate	33	29	51	49	31	47
% adequate	85.897	87.607	78.205	79.060	86.752	79.915
% inadequate	14.103	12.393	21.795	20.940	13.248	20.085

Table III-3: Model Inadequacy by Tree.

Number and percentage of trees for which a certain number of models were either adequate or inadequate.

# Models	0	1	2	3	4	5	6
Bonferroni-Pvalues							
adequate	12	3	7	12	22	35	143
inadequate	143	35	22	12	7	3	12
% adequate	5.128	1.282	2.991	5.128	9.402	14.957	61.111
% inadequate	61.111	14.957	9.402	5.128	2.991	1.282	5.128
2D convex-hulls							
adequate	5	7	15	14	16	41	136
inadequate	136	41	16	14	15	7	5
% adequate	2.137	2.991	6.410	5.983	6.838	17.521	58.120
% inadequate	58.120	17.521	6.838	5.983	6.410	2.991	2.137

Table III-4: Pairwise Model Inadequacy.

Pairwise counts of trees for which two models were inadequate for, and percentages of how often a model was inadequate together with another (by row), assessed using either Bonferroni-corrected p-values (above) or 2D convex-hulls (below). BD=constant rate birth-death, T_ex_co=time-dependent birth-death with exponential λ and constant μ, T_co_ex= time-dependent birth-death with constant λ and exponential μ. DD_lin_co= diversity-dependent birth-death with linear λ and constant μ. DD_lin_lin= diversity-dependent birth-death with linear λ and linear μ.

Model	Yule	BD	T_ex_co	T_co_ex	DD_lin_co	DD_lin_lin
Bonferroni-Pvalues [counts]						
Yule	33	25	19	25	16	22
BD	25	25	14	22	14	16
T_ex_co	19	14	35	24	18	24
T_co_ex	25	22	24	41	17	25
DD_lin_co	16	14	18	17	33	29
DD_lin_lin	22	16	24	25	29	63
Bonferroni-Pvalues [%]						
Yule	100.00	75.76	57.58	75.76	48.48	66.67
BD	100.00	100.00	56.00	88.00	56.00	64.00
T_ex_co	54.29	40.00	100.00	68.57	51.43	68.57
T_co_ex	60.98	53.66	58.54	100.00	41.46	60.98
DD_lin_co	48.48	42.42	54.55	51.52	100.00	87.88
DD_lin_lin	34.92	25.40	38.10	39.68	46.03	100.00
2D convex-hulls [counts]						
Yule	33	28	19	25	15	20
BD	28	29	15	22	15	17
T_ex_co	19	15	51	25	10	19
T_co_ex	25	22	25	49	16	22
DD_lin_co	15	15	10	16	31	25
DD_lin_lin	20	17	19	22	25	47
2D convex-hulls [%]						
Yule	100.00	84.85	57.58	75.76	45.45	60.61
BD	96.55	100.00	51.72	75.86	51.72	58.62
T_ex_co	37.25	29.41	100.00	49.02	19.61	37.25
T_co_ex	51.02	44.90	51.02	100.00	32.65	44.90
DD_lin_co	48.39	48.39	32.26	51.61	100.00	80.65
DD_lin_lin	42.55	36.17	40.43	46.81	53.19	100.00

Table III-5: Pairwise Model Inadequacy.

Difference between number of taxa, tree shape metrics, and Eigengap between an empirical tree and the tree set simulated based on them under the respective models. The rows with the model names indicate for the three shape metrics, how often they were significantly different (and thus the reason the model was deemed inadequate) according to Bonferroni-corrected p-values.

Model (times inadequate)	Ntax	Princ. Eigenv.	Skewness	Peakedness	Eigengap
Yule (33)		**11**	**22**	**3**	
Mean Diff	7.194	6504.625	-0.522	0.510	32.087
Mean Diff Ad.	-6.024	-993.316	-0.299	0.343	16.363
Mean Diff Inad.	87.704	52173.901	-1.884	1.529	127.857
BD (25)		**11**	**14**	**0**	
Mean Diff	-6.834	-1333.229	-0.429	0.401	18.180
Mean Diff Ad.	-8.781	-1739.287	-0.318	0.344	14.869
Mean Diff Inad.	9.444	2061.411	-1.349	0.880	45.862
T_ex_co (35)		**2**	**25**	**12**	
Mean Diff	41.717	9023.771	-0.363	1.135	36.268
Mean Diff Ad.	35.693	7746.381	-0.073	1.082	30.443
Mean Diff Inad.	75.967	16286.646	-2.011	1.439	69.388
T_co_ex (41)		**30**	**24**	**18**	
Mean Diff	10.576	6701.896	-0.490	0.463	27.750
Mean Diff Ad.	-0.768	3321.873	-0.238	0.397	19.213
Mean Diff Inad.	63.976	22612.734	-1.674	0.776	67.937
DD_lin_co (33)		**2**	**24**	**15**	
Mean Diff	4.4276	1371.9054	-0.8158	0.4076	25.6784
Mean Diff Ad.	2.2625	263.6969	-0.6035	0.2952	20.4348
Mean Diff Inad.	17.6151	8121.9021	-2.1085	1.0928	57.6166
DD_lin_lin (63)		**18**	**46**	**31**	
Mean Diff	2.9490	313.8179	-0.8562	0.2622	24.3850
Mean Diff Ad.	1.4881	74.2374	-0.5744	0.2121	18.2381
Mean Diff Inad.	6.9143	964.1081	-1.6212	0.3980	41.0696

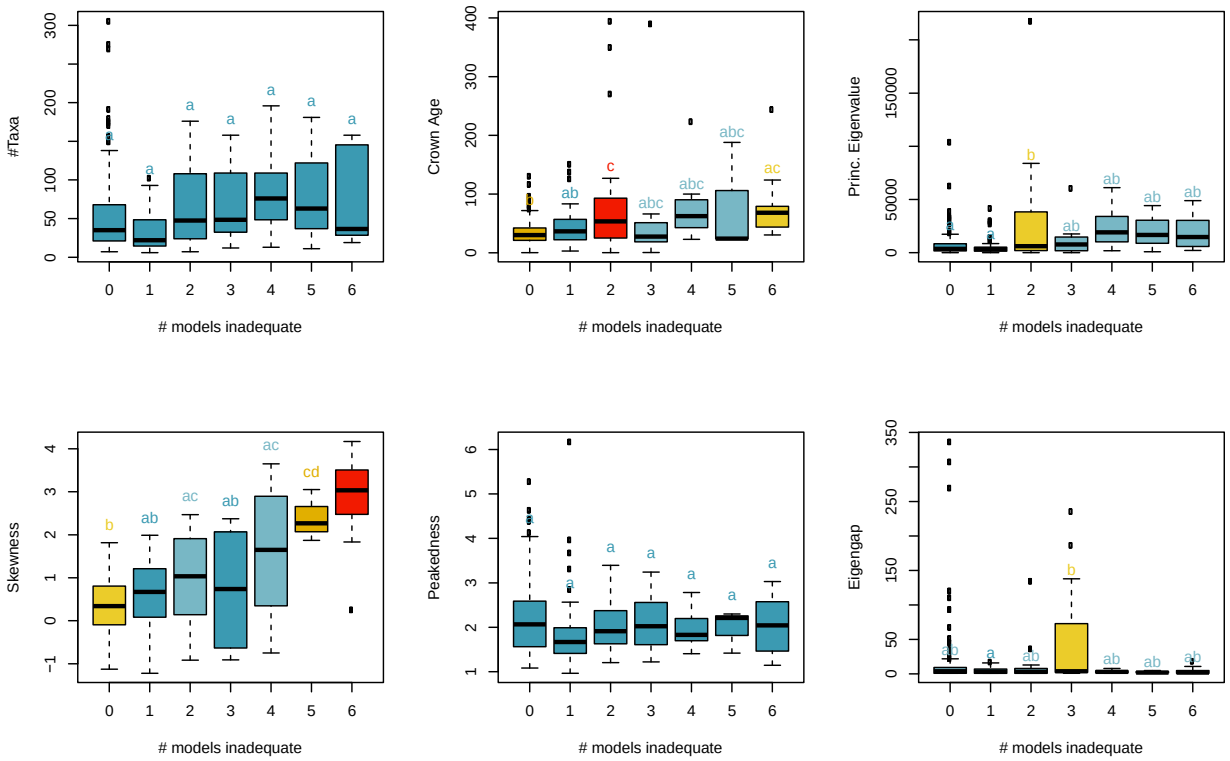

Figure III-1: Relation of Empirical Tree Metrics with Model Inadequacy.

Differences in number of taxa, crown age, and tree metrics for trees for which different numbers of the initial six models were inadequate. Colors and letters indicate the groupings based on a Tukey post-hoc test.

Table III-6: T-tests of Metrics between Adequate and Inadequate models.

Test for the difference between number of taxa, crown age, the tree shape metrics and Eigengap, between trees that were adequate and inadequate for a given model.

Model	Attribute	Statistic	Parameter	Pvalue	ConfInt1	ConfInt2	Est1	Est2
Yule	Ntax	-3.403	40.718	0.0015	-54.972	-14.021	50.413	84.909
	Tree Age	-3.174	33.776	0.0032	-85.731	-18.795	39.658	91.921
	Princ. Eigenv.	-2.948	33.180	0.0058	-34523.310	-6331.403	8181.396	28608.753
	Skewness	-3.936	34.146	0.0004	-1.773	-0.566	0.506	1.675
	Peakedness	-1.419	55.434	0.1616	-0.411	0.070	2.104	2.275
	Eigengap	-1.097	37.066	0.2795	-32.911	9.785	13.164	24.727
BD	Ntax	-2.951	28.324	0.0063	-59.777	-10.809	51.507	86.800
	Tree Age	-2.477	25.594	0.0202	-79.652	-7.377	42.380	85.895
	Princ. Eigenv.	-2.977	29.068	0.0058	-22644.131	-4203.324	9628.018	23051.746
	Skewness	-2.277	25.243	0.0315	-1.598	-0.081	0.581	1.420
	Peakedness	-0.988	35.191	0.3298	-0.416	0.143	2.114	2.250
	Eigengap	-1.375	26.066	0.1807	-46.148	9.145	12.818	31.320
T_ex_co	Ntax	0.671	50.395	0.5053	-11.499	23.038	56.141	50.371
	Tree Age	-1.896	38.623	0.0654	-52.043	1.685	43.263	68.442
	Princ. Eigenv.	-1.038	52.750	0.3042	-9998.894	3181.706	10552.345	13960.939
	Skewness	-7.373	38.529	0.0000	-2.042	-1.162	0.431	2.033
	Peakedness	2.898	58.147	0.0053	0.108	0.590	2.181	1.831
	Eigengap	3.852	211.673	0.0002	6.295	19.495	16.724	3.829
T_co_ex	Ntax	-0.525	54.380	0.6014	-23.867	13.954	54.409	59.366
	Tree Age	-2.481	47.035	0.0167	-53.118	-5.551	41.889	71.224
	Princ. Eigenv.	-1.476	68.935	0.1444	-10539.535	1574.252	10276.757	14759.399
	Skewness	-5.569	44.967	0.0000	-1.720	-0.806	0.449	1.713
	Peakedness	1.740	75.407	0.0859	-0.029	0.428	2.163	1.964
	Eigengap	-0.321	52.513	0.7493	-19.369	14.022	14.326	17.000
DD_lin_co	Ntax	-2.168	42.687	0.0358	-40.302	-1.456	52.333	73.212
	Tree Age	-2.486	35.172	0.0178	-65.819	-6.645	41.919	78.151
	Princ. Eigenv.	-2.960	43.189	0.0050	-18689.774	-3542.383	9494.525	20610.604
	Skewness	-5.719	37.200	0.0000	-1.734	-0.827	0.490	1.770
	Peakedness	1.804	62.352	0.0760	-0.022	0.422	2.157	1.957
	Eigengap	2.978	231.793	0.0032	3.512	17.248	16.259	5.879
DD_lin_lin	Ntax	-0.431	112.771	0.6675	-17.943	11.536	54.415	57.619
	Tree Age	-2.103	85.958	0.0384	-36.712	-1.033	41.948	60.820
	Princ. Eigenv.	-1.611	73.612	0.1114	-14285.236	1510.953	9342.562	15729.704
	Skewness	-5.581	77.529	0.0000	-1.314	-0.623	0.410	1.378
	Peakedness	1.293	118.487	0.1984	-0.080	0.380	2.169	2.019
	Eigengap	3.692	181.837	0.0003	6.644	21.901	18.637	4.365

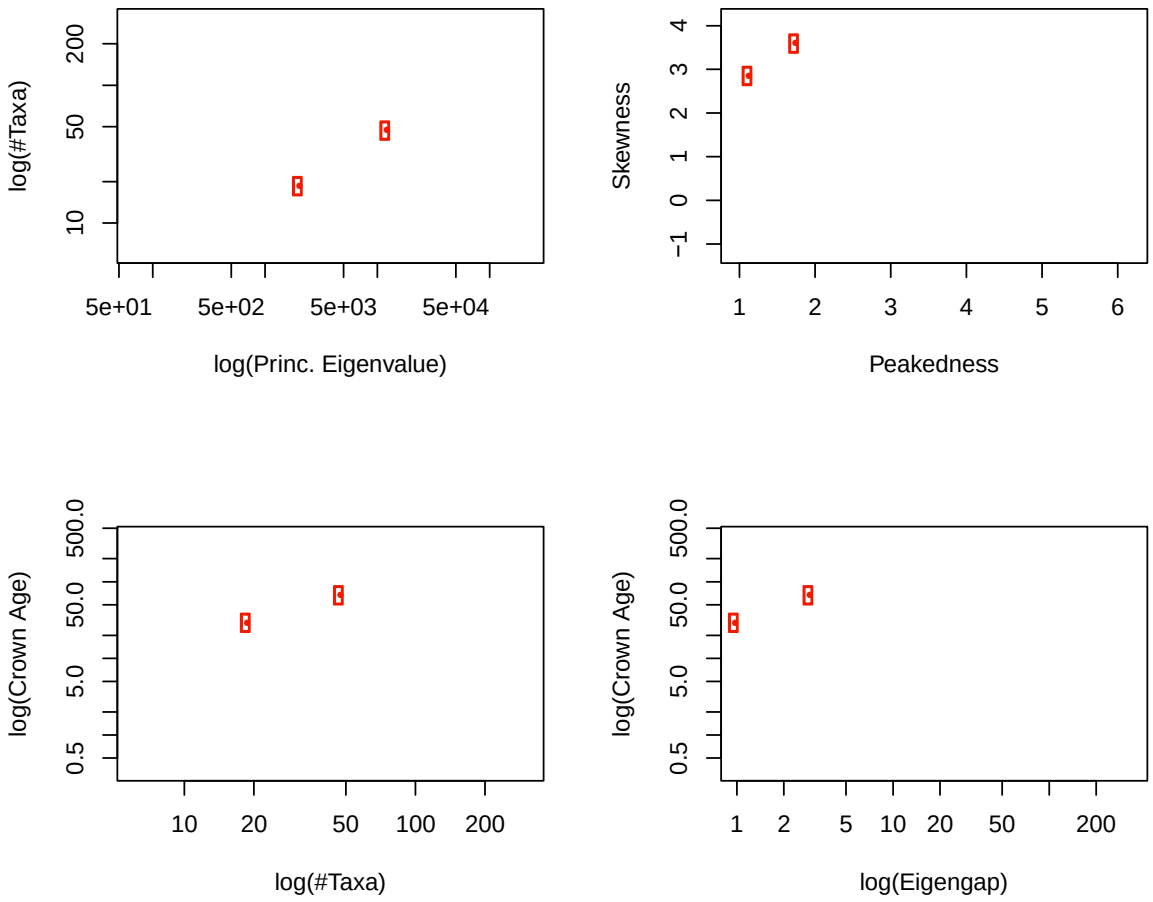

Figure III-2: Empirical Tree Metrics and Inadequate Models.

Scatterplots of number of taxa, crown ages and tree metrics, color coded by the number of models inadequate for each tree. Blue indicates at least one of the initial six models was adequate, red indicates all six were inadequate, and the two trees for which every model was inadequate are marked with red circles. The axes for number of taxa, crown age, principal Eigenvalue, and Eigengap are logarithmized.

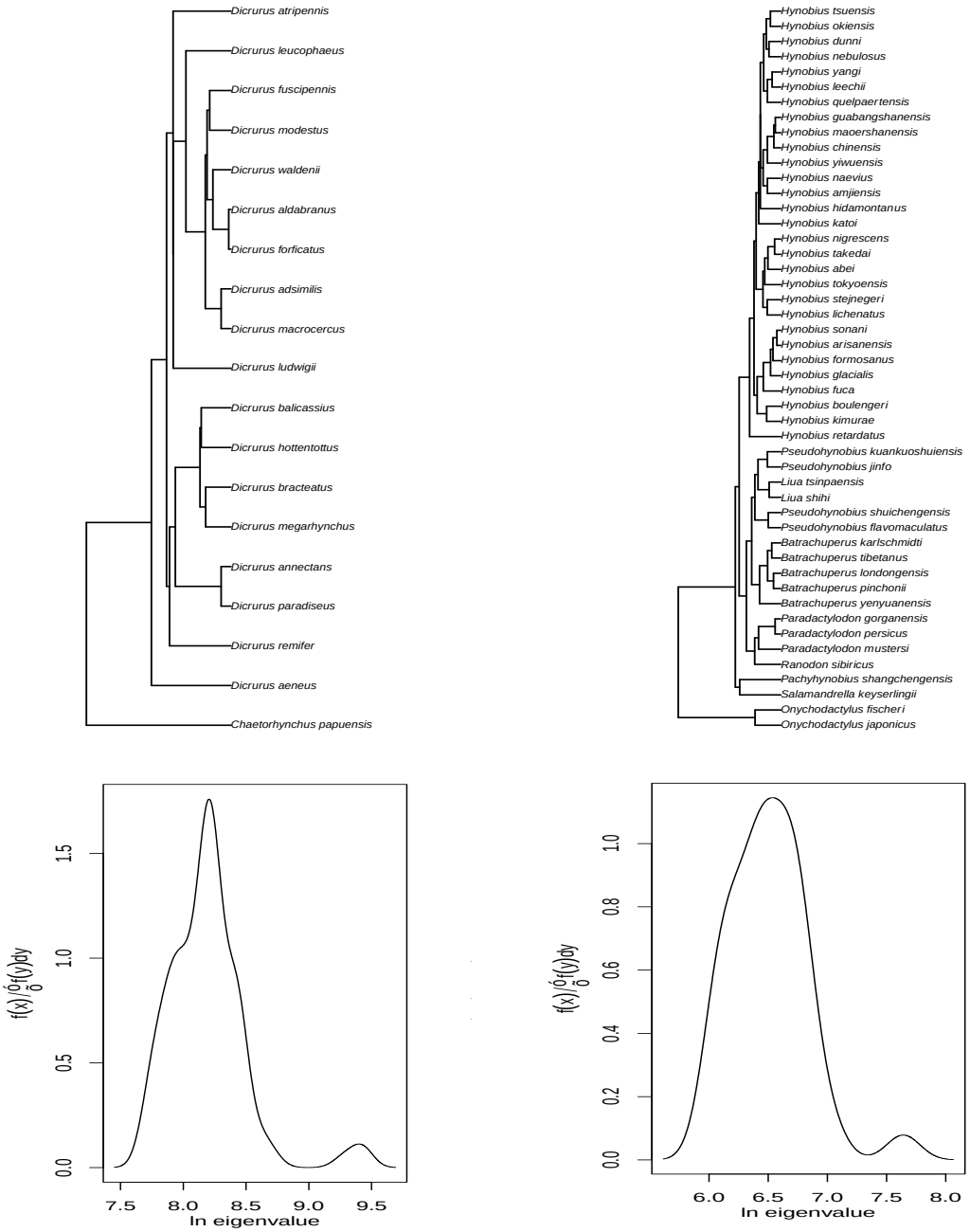

Figure III-3: Phylogenies and Spectra of Trees for which all Models fail.

The phylogenies (top) and corresponding Laplacian spectra (bottom) of the two trees for which all models are inadequate. On the left tree 244 (family Dricuridae), on the right tree 323 (family Hynobiidae).

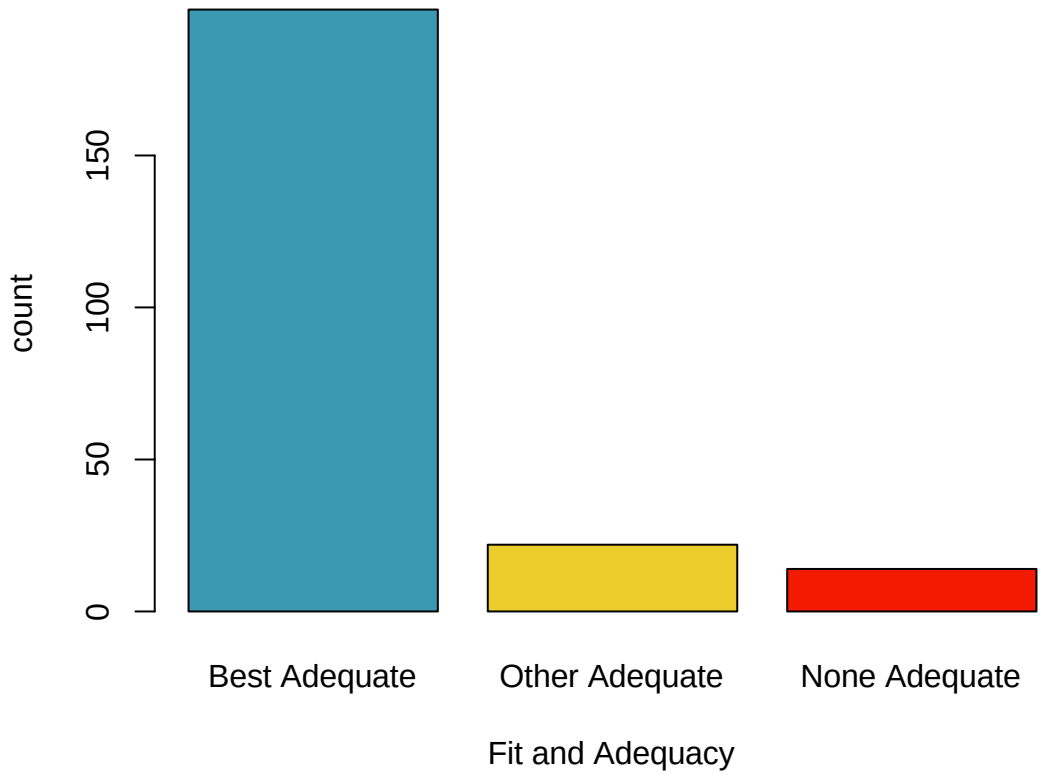

Figure III-4: Relation of Model Fit and Model Adequacy.

Numbers of trees for which their best fitting and other model was adequate, a different model than the best was adequate, or none were adequate.

CHAPTER IV
UNVEILING THE DIVERSITY OF DUNG BEETLES - BIOGEOGRAPHY

Abstract

The remarkable diversity and global distribution of dung beetles has long attracted the interest of researchers. However, there is still an ongoing debate on their origin, the reasons behind their diversity, and their path to global distribution. The two most prominent hypotheses regarding their origin and biogeographic history involve either vicariance events after the breakup of Gondwana, or an African origin and subsequent dispersal. One of the key reasons why the question is still disputed is a dependence on knowing the age of the dung beetles – a Mesozoic origin would favor the scenario of Gondwanan vicariance, a Cenozoic origin would suggest the out-of-Africa scenario. To help settle this longstanding question, I provide a taxonomically expanded phylogeny, with divergence times estimated under two calibration schemes suggesting an older or younger origin respectively. Using model-based inference, I estimate the ancestral area of the group and test for the influence of ranges on diversification rates. My results support the hypothesis of an old age for Scarabaeinae and Gondwanan origin, but remain ambiguous about the exact relation of range on lineage diversification.

Introduction

Dung beetles are a remarkable group of insects. Their unusual lifestyle requiring the dung of other animals to feed and reproduce gave rise to a host of morphological and behavioral specializations, as adaptations to the various ecological peculiarities they face in their worldwide distribution (Hanski & Cambefort 1991). While their diversity of around 5,300 species is comparatively modest among beetles, their dung-processing activity makes them one of the most important groups of insects, both ecologically (Nichols *et al.* 2008) and economically (Losey & Vaughan 2006). Their unusual life history has attracted considerable interest of researchers, in ecology and evolution (Hanski & Cambefort 1991; Scholtz, Davis & Kryger 2009), conservation (Spector 2006), and even developmental biology (Moczek 2011). Despite that, key aspects concerning their origin and the factors behind their diversity are still unknown and subject of debate in the field. The ecology of extant species has been well studied (Hanski & Cambefort 1991), but the main hindrance to understanding historical, evolutionary aspects of their biology is the lack of a comprehensive and reliable phylogeny (Tarasov & Génier 2015), and with that, validation for their taxonomy (Tarasov & Dimitrov 2016).

Probably the two most debated questions of dung beetle evolution revolve around their geographic origin and what led to their current diversity and distribution. A relation to their associated dung producers is suspected (Scholtz, Davis & Kryger 2009; Gunter *et al.* 2016), but disagreement exists over whether the success of dung beetles has always been an association with mammals, or whether dinosaurs were involved early in their history. Consequently, the main competing hypotheses regarding their origin and subsequent spread are whether it was Gondwanan-vicariance (Davis, Scholtz & Philips 2002) or dispersal out of Africa (Sole & Scholtz 2010). The answers to both partially hinge on the question on how old the group is. It has long been debated whether Scarabaeinae are of Mesozoic or Cenozoic origin – the former would make Gondwanan vicariance and feeding on dinosaur dung plausible, the latter would rule it out. Earlier attempts to determine the age of dung beetles using various approaches led to widely

different estimates, ranging from mid Mesozoic to early Cenozoic (Hanski & Cambefort 1991; Scholtz & Chown 1995; Krell 2000; Davis, Scholtz & Philips 2002; Krell 2006). More recent attempts using phylogenies also ranged from Cretaceous to Eocene-Oligocene (Wirta, Orsini & Hanski 2008; Ahrens, Schwarzer & Vogler 2014; Mlambo, Sole & Scholtz 2015; Gunter *et al.* 2016), shifting the support for early and late origins of Scarabaeinae over time (Scholtz, Davis & Kryger 2009). At this time, the field is still divided: an old Mesozoic origin of dung beetles, and with that a biogeographical scenario of Gondwanan vicariance and subsequent dispersal has been supported by some studies (Davis, Scholtz & Philips 2002; Gunter *et al.* 2016; Gunter *et al.* 2018), whereas a young Cenozoic origin and therefore an out-of-Africa dispersal scenario has been supported by others (Monaghan *et al.* 2007; Sole & Scholtz 2010; Davis, Scholtz & Sole 2016).

To address these questions, I provide an extended phylogeny of Scarabaeinae, with divergence times estimates based on calibrations reflecting two different assumptions of their maximal age. I use this phylogeny and the inferred age of the group to address the question of their geographical origin using model-based methods for ancestral range estimation; and the question of how it relates to their lineage diversification using range dependent diversification rate estimation. In particular, focusing on their Gondwanan origin and subsequent dispersals to areas of (what was formerly) Laurasia and Madagascar, I address the question of whether the dispersal into new areas was promoting dispersal, or whether their place of origin is the main source of diversification. Finally, I test a hypothesis that links dung beetle diversity to the dung producers present in different regions (Davis & Scholtz 2001; Scholtz, Davis & Kryger 2009). Specifically, linking beetle diversity to the size of mammal dung producers, and the diversity of different dung types they produced, suggesting that areas with larger mammals and more diverse droppings would allow for a more diverse dung beetle fauna.

Materials and Methods

Phylogenetic Analysis

I wrote a script in R (R Development Core Team 2014) using the packages reutils (Schöfl 2016) and seqinr (Charif *et al.* 2005) to query GenBank (Sayers *et al.* 2019) for any nucleotide sequences matching the organism label “Scarabaeinae” and that carried either “COI”, “16S”, “18S”, “28S”, or “CAD” in the title. I downloaded the resulting accessions using the packages ape (Paradis, Claude & Strimmer 2004) and insect (Wilkinson *et al.* 2018), removed any duplicates and saved them as FASTA files. The 28S accessions included both 28SD2 and 28SD3 sequences, which were sorted and aligned separately. The retrieved COI sequences largely covered two adjacent regions of the gene, and since only few accessions covered both regions, I decided to split them into two separate alignments (further called COI-1 and COI-2 respectively). If a marker had taxa with several accessions, the longest sequence was chosen unless it proved to be clearly different from the other accessions for that taxon. Taxa which were only determined to genus level were only included if they were the only representative of that genus.

The accessions of the seven markers were aligned separately using AliView v.1.26 (Larsson 2014): They were aligned using MAFFT globalpair, and then visually inspected and adjusted

manually where necessary. During this process, any sequences that showed considerable mismatch with the rest of the alignment were submitted to BLAST (Altschul *et al.* 1997) to detect any mislabeled sequences from other organismal groups. Similarly, quick RAxML (Stamatakis 2014) runs were performed for each aligned marker separately, and were tested for generic monophyly and long branches. Generic monophyly was tested using the package MonoPhy (Schwery & O'Meara 2016). Long branches were determined using the package ips; terminal branches were considered exceptionally long if their length was more than 0.25 times the maximum tip height in the tree, or if the product of their length and height was more than two times the interquartile range away from the third quartile of all tip length and tip height products in the tree. Any taxa that stood out by branch length or placement were checked using BLAST as well. The accession numbers of the sequences used in the final alignment can be found in Table IV-1.

Poorly aligned or divergent regions in each alignment were removed using Gblocks (Castresana 2000) (settings: minimum bases for conserved regions >0.5x alignment length, minimum for flanking regions >0.7x alignment length, maximum contiguous nonconserved sites: 8, minimum block length: 4 (noncoding) or 5 (coding), gap positions allowed with > half of sequences having gap). Finally, all alignments were concatenated using the package evobiR (Blackmon & Adams 2015). Partitioning the alignment by each marker (the two marker parts in case of COI), I used PartitionFinder v.2.1.1 (Lanfear *et al.* 2017) to determine the best substitution models for each partition, using the greedy algorithm (Lanfear *et al.* 2012), and PhyML (Guindon *et al.* 2010). Additionally, the package ClockstaR (Duchene, Molak & Ho 2014) was used to determine whether the partitions should have different clock models.

Using the concatenated alignment and the determined substitution models, a maximum likelihood phylogeny was constructed with RAxML v.8.2.4 (Stamatakis 2014), using the rapid hill climbing algorithm with 100 rapid bootstrap samples. The ingroup and four clades that would later be used for fossil calibrations (see below) were constrained to be monophyletic. Because of branch support issues, I used the online implementation of RogueNaRok (Aberer, Krompass & Stamatakis 2013) to find potential rogue taxa. Taxa whose exclusion would lead to a raw improvement of more than 1 were inspected and 136 sequences were eventually dropped and a final RAxML tree was built from this reduced alignment. To account for weakly supported clade relationships in subsequent analyses, ten sets of bootstrapped alignments were created, for each of which a separate tree was built in the same manner as from the full alignment. Monophyly on genus and tribe level was assessed for all trees using the R package MonoPhy (Schwery & O'Meara 2016).

Divergence Time Estimation

Reliable Scarabaeinae fossils that have recently been re-examined by Tarasov *et al.* (2016) were chosen to calibrate tree nodes during divergence time estimation. Of the 35 fossils previously assigned to Scarabaeinae, they considered only 21 to be assigned reliably on the basis of their morphological characters. From among these, the earliest fossils with confident generic placement were chosen for each genus that I could assign a node to. The minimal age of the fossils was used as minimum constraint for the stem age of the corresponding clade. The oldest

reliable fossil dung beetle, *Lobateuchus parisii*, was used to constrain the subfamily, and the minimal age estimate for *Juraclopus rohendorfi*, the oldest fossil of the family Scarabaeidae, was used as the minimal age constraint for the crown of the whole tree.

In order to get ages to constrain the maximal age of these clades, I consulted previous studies that had estimated the age of Scarabaeinae. On the younger end of the spectrum, some studies refer to Wirta, Orsini and Hanski (2008), or Mlambo, Sole and Scholtz (2015) for young estimates of the age of dung beetles (33.9ma or 56ma respectively). However, neither of those estimates are particularly useful to calibrate the age of the whole group, as Wirta, Orsini and Hanski (2008) focused on Helictopleurini of Madagascar, and Mlambo, Sole and Scholtz (2015) exclusively on African dung beetles. While some more dated phylogenies of clades within the subfamily exist, the only examples that include an age for the whole group are of higher taxa in which the dung beetles are nested. Ahrens, Schwarzer and Vogler (2014) constructed a tree of 146 taxa of Scarabaeoidea, in which the crown Scarabaeinae was estimated at 89.6ma (ranging from 83.5-98.1ma), while Gunter *et al.* (2016) in a phylogeny of 445 taxa of Scarabaeoidea estimated it to range from 118.8-131.6ma. While various studies estimated the ages within Scarabaeoidea (Ahrens, Schwarzer & Vogler 2014; McKenna *et al.* 2015; Toussaint *et al.* 2017), the ages of Scarabaeinae and Scarabaeidae could not always be obtained from them – consequently, I relied solely on Gunter *et al.* (2016) for an age estimate of Scarabaeidae (116.85-199.64ma).

Because no clade can be older than the clade in which it is nested, I constrained the maximal stem ages for all constrained genera (as well as the maximal crown age of all Scarabaeinae) to be the maximal estimated age of the Scarabaeinae, and the maximal crown age of the whole tree (being the stem age of Scarabaeinae) to be the maximal estimated age of Scarabaeidae. Given the disagreement of estimated ages in the literature, I set up two different sets of constraints: an 'old' one, with the Scarabaeinae and Scarabaeidae ages as estimated by Gunter *et al.* (2016), and a 'young' one, with the Scarabaeinae and Scarabaeidae ages as estimated by Ahrens, Schwarzer and Vogler (2014) – with the latter actually being the age of Scarabaeoidea, due to the lack of available ages for the actual Scarabaeidae (apart from that by Gunter *et al.* (2016)). While McKenna *et al.* (2015) actually estimated a younger age for Scarabaeoidea, their estimate is younger than the age of the oldest fossil of that group, which is why I chose to use the next older estimate by Ahrens, Schwarzer and Vogler (2014) ranging from 167.2-181.8ma. The different fossil constraints used can be seen in Table IV-5.

Those two sets of age constraints were then used to estimate divergence times for the ML tree obtained through RAxML and all ten bootstrap trees, using penalized likelihood in treePL (Smith & O'Meara 2012). I performed a thorough search with random cross validation, which was preceded by a preliminary run to estimate the best tuning parameters and smoothing factor.

Occurrence Data and Ancestral Range Estimation

I used the package rgbif (Chamberlain *et al.* 2016; Chamberlain & Boettiger 2017) to download occurrence data from GBIF (http://data.gbif.org) for all taxa determined to species level. Using an R script, the retrieved records were cleaned from empty or invalid entries, with regards to the

basis of record, identification date, country and coordinates, and sets with unique countries or coordinates per taxon respectively were made. Name validity of unavailable taxa was checked using the package taxize (Chamberlain & Szocs 2013), and eventually searched on genus level. For any taxa that were still missing, occurrence information was searched for in the literature.

I defined the areas taxa occur in as continents (Africa, Asia, Europe, North America, South America, Oceania), and defined Madagascar as a separate area, given the high number of taxa endemic to it (Miraldo, Wirta & Hanski 2011). However, I did not define India as a separate area, and considered it part of Asia. India has a peculiar tectonic history of initially staying connected to Madagascar until breaking off around 87ma and drifting northwards to collide with Asia around 55-43ma (Seton *et al.* 2012). Not explicitly including it as a separate area means a major simplification for the model, but also ignoring potentially relevant possible scenarios. For example, lineages that have entered India from Madagascar before their breakup, could have been evolving in relative isolation there until coming into contact with Eurasia. However, I suspect that testing that kind of scenario would require more detailed geographical resolution, and thus may warrant a dedicated separate study.

Each taxon's range was defined as the one or several areas it occurred in, based on the collected information from GBIF and the literature. Where information on continent was missing, it was inferred from coordinates and countries of occurrence, and inconsistencies between these were checked and corrected. Taxa occurring in many continents were inspected for the extent of overlap and reliability of the records. In doing so, I also paid attention to species that were introduced to areas by humans, and removed such occurrences from the species' range, thereby only assigning its presumed natural range.

Given the age of the group, it is conceivable that tectonic plate movement played a role in their dispersal and distribution. I thus constructed a stratified dispersal matrix representing the changing strength of dispersal barriers between the continents over time. I divided the last 200ma before the present into five time slices, following the tectonic events described in Seton *et al.* (2012), stretching from 200-150ma (Pangaea intact), 150-110ma (breakup of Pangaea into Gondwana and Laurasia, Madagascar breaking off of Africa, though still connected via Antarctica), 110-50ma (breakup of Gondwana into Africa, Australia-Antarctica, and South America, the latter still connected to Antarctica via a land bridge), 50-20ma (Australia separates from Antarctica, Laurasia breakup into Laurentia and Eurasia (Hosner, Braun & Kimball 2015), South America properly disconnected from Antarctica (Reguero *et al.* 2014)), and 20-0ma (land bridges (some temporary) establish at Beringia (Hosner, Braun & Kimball 2015), the Isthmus of Panama, and between Australia and Asia, and Eurasia and Africa).

Inspired by Toussaint, Bloom and Short (2017), I recognized dispersal barriers of different strengths: 1) directly adjacent areas (barrier of strength 0), 2) areas connected by land bridge (0.15), 3) areas separated by a small distance of water (0.25), and 4) by a large distance of water (0.75). Having to pass through another area was considered a barrier of strength 0.5, and the case of facing more than 3 barriers was assigned a strength of 0.95; the corresponding dispersal multiplier was 1 - strength of barriers. The route of smallest resistance was picked for each possible dispersal case, adding up the different barriers encountered (matrices see Table IV-6).

While it would make sense that different dispersal barriers at different times would affect the dispersal dynamics of the group, it is also possible that only time differences or only dispersal barriers did. Furthermore, the partially arbitrary choice of time intervals could affect the inference as well. Thus, I also calculated averaged dispersal matrices to be used without time stratification. For this purpose, the dispersal multipliers of each time slice were weighted by the length of that time slice relative to the total time from the beginning of that slice until the present. These weighted multipliers were added up for each transition and then divided by the sum of weights, so they would add to one. This weighting is intended to represent the fact that the more recent positioning of the continental plates should have had a higher impact on the current distribution of extant taxa. Two matrices (for 3 and 4 time slices respectively, see Table IV-7) were constructed that way, to be used for either the phylogeny dated with the younger or older calibration times.

I used the package BioGeoBEARS (Matzke 2013b; Matzke 2013a) to estimate the ancestral ranges of the dung beetles. The package implements some of the most popular models for ancestral range estimation, DEC (Ree & Smith 2008), DIVA (Ronquist 1997), and BayArea (Landis *et al.* 2013), in the same framework, allowing them to easily be compared against each other to test different biogeographical hypotheses. While the DEC model is a maximum likelihood implementation just as originally implemented in Lagrange, the DIVA model was originally implemented using parsimony, and in BioGeoBEARS the processes DIVA assumes are modeled under maximum likelihood. Similarly, BayArea was originally implemented as Bayesian, and is represented in BioGeoBEARS as a maximum likelihood interpretation of the same. Thus, these two models should be referred to as DIVALIKE and BAYAREALIKE respectively.

A popular feature of BioGeoBEARS is the addition of "jump-dispersal" to any of these models. By adding the additional jump parameter ("+j"), one allows for founder events in the model, meaning that at a speciation event, one descending lineage stays in the ancestral range, while the other descendant jumps to a new area (Matzke 2014). However, there has recently been some debate regarding the validity of the +j models (Ree & Sanmartín 2018; Klaus & Matzke 2019), and this type of cladogenetic event is believed to be more important in island systems than in non-island clades (Matzke 2013b). For those reasons, I decided not to employ the j parameter.

However, since the dispersal multiplier matrix defined above is somewhat arbitrary, I do pair it with a parameter (w) that scales the matrix and that can be optimized as well, constituting the +w models (Dupin *et al.* 2017). The parameter w is used to exponentiate the dispersal multiplier before it is multiplied by the dispersal rate and is 1 by default. Thus, if the multipliers play an important role in the biogeographical history of the group, it is likely to be estimated to be larger than 1, whereas it should be estimated to be lower than 1 if the dispersal matrix does not add much to explain the patterns. While this does not allow to modulate the relative strengths of the different multipliers, it seems reasonable to expect that w would be estimated to downweigh the importance of a grossly unrealistic set of multipliers.

I employed each of the three base models (DEC, DIVALIKE, BAYAREALIKE) in four different ways: 1) as a basic model (estimating d and e), 2) with the non-stratified manual

dispersal multiplier matrix and estimated w parameter, 3) as a time stratified model, and 4) as a time stratified model with a manual dispersal multiplier matrix for each time slice, and estimated w parameter. In all those analyses, I constrained the maximal range any lineage can occupy to 3, as none of the extant species occupy more than 3 areas. To account for the branch support issues, I ran all these models both on the trees with old and young calibrations, as well as the ten bootstrap trees each.

Finally, preliminary analyses produced curious results, particularly in the time stratified model, where lineages within clades that where entirely present in one area (e.g. Madagascar) would commonly disperse to a neighboring area (*e.g.* Africa) right after a speciation event, only to return to the ancestral area again. It was presumed that clades in which every species inhabited the same two areas (*e.g*. the Americas) would be responsible for this, as models as DEC do not include the required scenario (cladogenesis in widespread taxa where both daughters inherit widespread range). Thus, such cases required the above pattern of one daughter losing and re-gaining the widespread range, thus inflating the inferred dispersal rate and forcing other lineages to do the same. I thus created test data sets where I identified the area in which species most commonly co-occurred (the Americas), or the two pairs most co-inhabited areas (the Americas and Eurasia), and coded those as one, thus getting a 6-area and 5-area dataset to be analyzed separately.

Diversification Analysis

To test the two biogeography-related hypotheses: whether areas with larger mammals and more diverse droppings are associated with higher diversification rates of dung beetles, and whether diversification rates were raised when dung beetles gained access to new habitats, dispersing from Gondwana to Madagascar or Laurasia. The former hypothesis was derived from Davis and Scholtz (2001); (also elaborated in Scholtz, Davis & Kryger 2009). They classify mammalian dung into four types based on size and physico-chemical characteristics: 1) small dry pellets from small to medium herbivores, 2) small odiferous droppings from omni- or carnivores, 3) large, dry, course-fibered droppings from large non-ruminant herbivores, and 4) large, moist, fine-fibered pads from large ruminants. The number of those types of dung available, as well as the (fairly correlated) body size of mammals was tallied for different biogeographical regions, and shown to relate to aspects of dung beetle diversity (Davis & Scholtz 2001; Scholtz, Davis & Kryger 2009).

Explicitly biogeographic diversification models such as GeoSSE (Goldberg, Lancaster & Ree 2011) or GeoHiSSE (Beaulieu & O'Meara 2016; Caetano, O'Meara & Beaulieu 2018) only test diversification between two areas (and three ranges: endemic to one or the other area, or widespread), making explicitly testing hypotheses involving more areas (not to speak of all seven) impossible. However, in the case of only three areas, testing different combinations of two against each other can still yield relevant insights. To this end, I re-coded the distribution data from above to merge occurrences 1) of areas with large mammals and diverse droppings (Afro-Eurasia: Africa, Europe, and Asia), of medium sized mammals and less diverse droppings (North and South America), and those of those with small mammals and the least diverse droppings (East Gondwanan Fragments: Oceania and Madagascar).; and 2) those of the areas

formerly making up Gondwana (South America, Africa, and Oceania), those of former Laurasia (North America, Europe, and Asia), while keeping Madagascar as an area of its own. I then formatted these two data sets to suit the different GeoSSE tests: a set where I join the areas with medium sized mammals and intermediate droppings-diversity with the areas of large mammals and diverse droppings, and one where I join them with the small mammal and low droppings-diversity instead. Another set of where I join Madagascar with Gondwana, where I join it with Laurasia, and where I leave it as a separate area from the rest.

For each of the resulting five data sets (and for both the young and old tree respectively), I inferred maximum likelihood estimates with GeoSSE, under a set of different model constraints. The constraints used were: combinations of equal speciation between areas, equal extinction between areas, equal dispersal rates between areas, and variation where speciation was set to zero in one area and forced to be equal in the other and widespread lineages, and vice versa. The full and constrained models were then compared using likelihood ratio tests and their ΔAIC score. The best models for each combination of data set and tree were subsequently used to get posterior distributions of the parameter estimates using MCMC, with an exponential prior related to the Kendall-Moran estimate for net diversification rate (Kendall 1949; Moran 1951). A short preliminary chain of 100 generations was run and the distances between the 5% and 95% quantiles for each parameter were used to set the tuning parameter w for the slice sampler. Then, each dataset was run for 20,000 generations. Convergence was assessed by the convergence parameter of the function, by visual inspection of the log likelihood trace, and calculating effective sample size using effectiveSize from the R package coda (Plummer *et al.* 2004). I then compared the 95% quantiles of the posterior distribution for all estimated parameters to see whether they overlap.

Results

Phylogeny and Divergence Times

Out of 122 genera represented in this phylogeny, 33 were monophyletic, 28 were non-monophyletic, and 61 were monotypic (Table IV-2). When considering the bootstrap trees, 14 genera were consistently monophyletic, 22 consistently non-monophyletic, whereas the remaining 25 varied between trees. On the tribal level, I recovered Eucraniini, Ateuchini, Eurysterniini, Gymnopleurini, and Sisyphini as monophyletic (Table IV-3). Of these, only the last three consistently so, with Onitini being monophyletic in some bootstrap replicates. Upon closer inspection of the reasons for each tribe's non-monophyly (Table IV-4), we see that at least part of it results from taxonomic issues. For Scarabaeini, Phanaeini, and Onitini, non-monophyly is due to a few *incertae sedis* taxa, most of which monotypic. Oniticellini and Onthophagini have similar issues where one intruder of the former is a member of the latter, but also the whole of Oniticellini is nested within Onthophagini. Dichotomiini, Deltochilini and Canthonini are scattered in clumps across the tree, with the latter two intermingling a lot, while Coprini come out in three separate clades.

The estimated ages of the calibrated nodes for both the young and old calibration are given in Table IV-5. It is apparent that under both calibration schemes, the crown age of the Scarabaeinae

has hit its constrained maximum age, while the stem ages of the constrained nodes are not reaching their maximum constraints but are often meeting or exceeding the minimum age constraint of the subfamily's crown. The bootstrap branch support values can be seen in Figure IV-1.

Ancestral Range Estimation

For all analyses of the 7-area data set on the young and old maximum likelihood trees and their respective sets of 10 bootstrap trees, the analyses recovered the unstratified DEC model with manual dispersal multipliers (DEC+w) as the best fitting model. In all trees, w – the exponent for the dispersal multipliers – was estimated to be larger than 1 (1.71-2.83), indicating its weight in the dispersal process being higher than the initial dispersal matrix suggested. The estimated dispersal and extirpation rates vary between the trees but are comparable and at the very least in the same order of magnitude. For the 6-area and 5-area datasets, which model was inferred to fit best varied widely among the different trees.

The ancestral range of the whole subfamily was estimated to be Africa, Oceania, and South America for both the young and old tree. However, in the alternative trees based on bootstrap replicates, the estimated ancestral range could include Madagascar instead of South America or instead of Oceania, or could just include Africa and Oceania (Table IV-8). Interestingly, reconstructions with root states other than Africa, Oceania, and South America tended to be more ambiguous. Overall, the estimated ancestral ranges seemed to suggest that numerous clades mostly stay in the same areas (particularly Africa, Oceania, and Madagascar), with some more dispersal within the Americas and Eurasia, as well as within the clade comprising the Onthophagini and Oniticellini (Figure IV-2).

Diversification Analyses

The best fitting models for each data set and tree, according to likelihood ratio tests and ΔAIC scores, are given in Table IV-9. The MCMC analyses converged and yielded reasonably high effective sampling sizes with no value below 295. The 95% quantiles of the posterior distributions for each parameter estimate is given in Table IV-10.

Regardless of whether the areas with medium sized mammals and intermediate diversity in droppings were joined with the areas of large or small mammals, there was no significant difference between them in terms of diversification rates (where they were not constrained), and the latter attributes any difference in diversity between those areas to higher dispersal out of Afro-Eurasia into the other areas (Figure IV-3, Figure IV-4).

Despite some differences in which model fitted the data best, the results between the old tree (Figure IV-5) and the young tree (Figure IV-6) are largely consistent. They show that lineages in Gondwana have a significantly lower speciation rate than lineages outside of it, while lineages in Laurasia disperse into the other areas at a higher rate than vice-versa, while there are no significant differences in either diversification nor dispersal between Madagascar and the other areas.

Discussion

Phylogeny and Taxonomy

With 541 represented in-group species, this is to date the most inclusive dated species-level molecular phylogeny of Scarabaeinae. Most previously inferred phylogenies of the group were either constrained to specific sub-clades or regions (e.g. Davis & Scholtz 2001; Wirta, Orsini & Hanski 2008; Sole & Scholtz 2010; Wirta *et al.* 2010; Mlambo, Sole & Scholtz 2015; Breeschoten *et al.* 2016; Gunter *et al.* 2018), or part of a larger phylogeny where the main focus accordingly was not or not only on dung beetles (Ahrens, Schwarzer & Vogler 2014; Kim & Farrell 2015; Gunter *et al.* 2016; Toussaint *et al.* 2017). A recent large un-dated phylogeny by Tarasov and Dimitrov (2016) based on 8 gene regions had a similar amount of terminals (547 with outgroup), though it constitutes a smaller sample of actual species diversity, as many species were represented by multiple accessions and many were not determined to species level. The levels of generic and tribal monophyly in this new phylogeny (Table IV-2, Table IV-3) and the causes of it (Table IV-4) would suggest a reasonable level of agreement between this tree and current taxonomy. The lack of support for many groupings within the tree (Figure IV-1), particularly in the backbone, is cause for concern, as it not only suggests shortcomings in the inferred tree, but also casts doubt upon the reliability of analyses results derived from that tree. Other attempts at phylogenies of the overall Scarabaeinae were plagued with similar patterns of low branch support (Tarasov & Dimitrov 2016), which suggests a general issue in the study of this group. Molecular phylogenies have challenged the traditional classification, particularly the tribal monophyly of Canthonini and Dichotomiini (Monaghan *et al.* 2007), as well as Coprini, Onthophagini and Oniticellini, while the monophyly of the remaining tribes still seems supported (Scholtz, Davis & Kryger 2009). Tarasov and Dimitrov (2016) note how their own results are consistent with those of previous phylogenetic studies of Scarabaeinae (Ocampo & Hawks 2006; Monaghan *et al.* 2007; Vaz-de-Mello 2007; Wirta, Orsini & Hanski 2008; Sole & Scholtz 2010; Wirta *et al.* 2010; Mlambo, Sole & Scholtz 2014; Gunter *et al.* 2016), as well as with a large phylogeny based on morphology (Tarasov & Génier 2015). They furthermore observed that the studies to date tend to resolve old nodes and more recent nodes, but not intermediate ones, and that the same set of problematic tribes mentioned above are consistently not monophyletic. All of this seems to be reflected in my results as well, particularly when considering the extent of monophyly problems (Table IV-4). Using the suggested new classification by Tarasov and Dimitrov (2016) does not yield much improvement. While Eucraniini and Eurysternini are strictly monophyletic in both, monophyly gained through reclassification in Dichotomiini, Canthonini, and Scarabaeini is offset by the loss of it in Ateuchini, Gymnopleurini, and Sisyphini. Both classifications show similar consistency across the set of bootstrap trees.

Ancestral Range Estimation and the Origin of Scarabaeinae

The two main hypotheses regarding where dung beetles originated are an origin in Gondwana followed by vicariance events after the breakup of the supercontinent (Hanski & Cambefort 1991; Davis & Scholtz 2001; Davis, Scholtz & Philips 2002), and an origin in Africa and subsequent dispersal out of it (Sole & Scholtz 2010). One major point of conflict between those two ideas was the age of Scarabaeinae: Gondwanan vicariance would necessitate the group to be of Mesozoic, rather than Cenozoic origin, as they would have to exist and be widespread enough

before the continental breakup (110ma according to Sanmartin and Ronquist (2004), 93ma according to Scotese (1993)) in order for vicariance events to be plausible.

To answer the question of biogeographic origin in absence of an appropriate phylogeny, some workers relied on classification (Hanski & Cambefort 1991; Davis & Scholtz 2001), considering widespread tribes to be ancient and predating the Gondwana-breakup, in turn giving rise to younger, less widespread tribes. Later attempts combined the relative age of phylogenies (Monaghan *et al.* 2007; Wirta, Orsini & Hanski 2008) with fast and slow rates of molecular sequence divergence in insects to get maximal and minimal divergence time estimates (Scholtz, Davis & Kryger 2009), concluding that even the slowest known divergence rates would not support the idea of a pre-Gondwanan-breakup origin of dung beetles. Sole and Scholtz (2010) subsequently used a time calibrated phylogeny of the African representatives of Canthonini and Dicotomiini to address the question, finding the divergence times between dung beetles and their outgroup, and of the crown of dung beetles to be considerably younger than the breakup of Gondwana (56ma and 40ma respectively), thus further supporting the later out-of-Africa scenario.

As for the divergence times inferred in this study (Table IV-5), the older age of 131.6ma would place the origin of dung beetles well before the complete breakup of Gondwana, whereas the younger estimate of 98.1ma appears ambiguous, depending on when the actual separation of Gondwana was completed, and depending on the accuracy of this age estimate. In either case however, the inference of older origins, similar to other recent estimates that considered more than just a subset of Scarabaeinae (Ahrens, Schwarzer & Vogler 2014; Gunter *et al.* 2016), makes a Gondwanan origin and thus the potential for vicariance after its breakup seem like a plausible option again.

The reconstruction of what is essentially Gondwana as the ancestral range at the crown of my phylogenies seems to further support the idea of Gondwanan-vicariance. However, the DEC model is known to have a bias towards widespread ancestors (Clark *et al.* 2008; Ree & Smith 2008; Buerki *et al.* 2011; Matzke 2014), even if not as strongly as DIVA (Kodandaramaiah 2010). Therefore, since I constrained the number of areas a lineage can maximally inhabit to three, the inference of those three areas as the origin could possibly be an artefact. With regards to the branch support issues, it would appear that the consistency with which the same model was preferred across all trees using the 7-area dataset, and the relative stability of the inferred origin to be Gondwana, or parts thereof (Table IV-8), could be seen as a sign that the result is robust enough. However, the wildly varying best supported model under the reduced area datasets (6 and 5 areas) is cause for concern. It could be argued, that leaving those pairs areas separate (North and South America, Europe and Asia respectively), might lead to an area distribution across clades that is not broken up when they are rearranged in the bootstrap trees, thus implying the same dispersal mechanisms. On the contrary, lumping them could lead to single-distribution clades being broken up, thus changing the number of implied dispersal and vicariance events. However, this is rather speculative and requires further investigation. Finally, while the high estimates for w suggests that the specified dispersal multipliers are a relevant improvement of the model overall, this does not guarantee that they are an accurate representation of dung beetle dispersal probabilities at the given times. While they might capture

some large scale dispersal constraints, the relative magnitude of the dispersal multipliers between specific continents could still be inaccurate, e.g. because of the way the barriers were specified and the values they were assigned. Sensitivity tests, or even adding an approach using actual distance, could help to inform us about this.

In any case, given the branch support issues, I would refrain from a too detailed interpretation of the ranges reconstructed at internal nodes. However, it is notable that the reconstruction suggests that many clades can be found which seem to be predominantly confined to one (or few) areas, particularly to Oceania, Madagascar, and Africa, with some more dispersal in American or Eurasian clades. An exception seems to be the clade comprising Onthophagini and Oniticellini, which seems to have much more area changes. This could indicate that the Onthophagini and Oniticellini are different from the remaining dung beetles in their biogeographic history and maybe their dispersal abilities. It could however also just reflect a lack of sampling in that clade.

Diversification Analyses

The GeoSSE results for the hypothesis on whether mammal size and available dung-diversity in different biogeographical areas affected the diversification rates of Scarabaeinae seem to suggest that this is not the case. The increased dispersal from Afro-Eurasia into the other regions is certainly interesting and could be a reflection of different plausible dispersal events, but given that the test was not set up to address these alternatives, I would suggest exercising care not to overinterpret this pattern. When Davis and Scholtz (2001) reported on the patterns among mammals, the diversity of their droppings, and dung beetle diversity in different areas of the world, they noted that it was particularly related to tribal diversity (and generic diversity within tribes) in those areas, but less correlated to genus or species diversity. They interpret this as a sign that the influence of these patterns in mammal body size and droppings on dung beetle diversification happened earlier in their evolutionary history, around the time when the tribes split. While the idea that mammal dung availability could have been a crucial influence on beetle diversification in the past, it would seem that the arguments presented in this particular case are problematic. Firstly, it relates an extant pattern in mammals to past effects in beetles assuming that the distribution of mammals and their droppings across the world was comparable between then and now. However, those extant patterns were almost certainly not constant over the timespan of dung beetle diversification, with the most recent event that changed those patterns already being the Late Quaternary megafaunal extinctions (Stuart 2015). Furthermore, assuming from the ages in this current tree Table IV-5, as well as past estimates (e.g. Gunter *et al.* 2016), the ages of many tribes may either be older than the rise of mammal diversity, or would at least coincide with a time when mammal diversity would not be expected to compare well to the extant one. Finally, considering the fact that the pattern does not correlate well with species level diversity, the method I employed might also be suboptimal to test the implied scenario of past influence, and an approach that allows the influence of mammal size and droppings to be constrained to particular time-intervals might be more suitable.

The result of the Gondwanan origin GeoSSE analysis might initially seem surprising, as a large portion of Scarabaeinae species diversity is found in areas that were formerly Gondwana (Scholtz, Davis & Kryger 2009) and the former-Gondwanan lineages were not underrepresented

in comparison to the others (Gondwana: 320 taxa, Laurasia: 143 taxa, Madagascar: 114 taxa). But considering the possible Gondwanan origin of the whole group, and the perceived inertia of clades (diversifying in an area rather than dispersing more), one might suggest that the larger absolute number of species in former Gondwanaland results from diversifying at a lower rate but over longer timespans than the on average probably younger Laurasian or Malagasy lineages. This would also explain the lack of rate difference between the latter and Laurasia and Gondwana combined. The higher dispersal out of Laurasia was not expected, assuming the general direction of dung beetle dispersal was outward from Gondwana. But in the light of higher dispersal outside of Gondwana, this could reflect the dispersal of few lineages out of Gondwana, where they diversified, and subsequently a few of those Laurasian taxa returned to former Gondwana (*e.g.* in the case of taxa which returned from North America to South America after closing of the isthmus). The lack of difference in dispersal rate between Madagascar and the rest is plausible as well, knowing that most Malagasy dung beetles are part of one of few clades that entered Madagascar and diversified there, with very few dispersals back (Miraldo *et al.* 2011; Sole *et al.* 2011). This is also reflected by the fact that the inferred rate between Madagascar and the rest is comparatively low (Table IV-10). All in all, those results would support the hypothesis that access to new areas was associated with a rise of diversification rates in dung beetles. However, more fine-scale analyses are needed to confirm the implied scenarios behind this. Also, while overall sampling frequency in this phylogeny was considered in this analysis, it cannot be ruled out that some of these results are artefacts of uneven sampling between groups within the subfamily. Until an even more complete tree of Scarabaeinae is available, finding ways to correct for such sampling biases would be advisable.

References

Aberer, A.J., Krompass, D. & Stamatakis, A. (2013) Pruning Rogue Taxa Improves Phylogenetic Accuracy: An Efficient Algorithm and Webservice. *Systematic Biology,* **62,** 162-166.

Ahrens, D., Schwarzer, J. & Vogler, A.P. (2014) The evolution of scarab beetles tracks the sequential rise of angiosperms and mammals. *Proceedings of the Royal Society B-Biological Sciences,* **281**.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W. & Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research,* **25,** 3389-3402.

Beaulieu, J.M. & O'Meara, B.C. (2016) Detecting Hidden Diversification Shifts in Models of Trait-Dependent Speciation and Extinction. *Systematic Biology,* **65,** 583-601.

Blackmon, H. & Adams, R. (2015) EvobiR: tools for comparative analyses and teaching evolutionary biology.

Breeschoten, T., Doorenweerd, C., Tarasov, S. & Vogler, A.P. (2016) Phylogenetics and biogeography of the dung beetle genus Onthophagus inferred from mitochondrial genomes. *Molecular Phylogenetics and Evolution,* **105,** 86-95.

Buerki, S., Forest, F., Alvarez, N., Nylander, J.A.A., Arrigo, N. & Sanmartin, I. (2011) An evaluation of new parsimony-based versus parametric inference methods in biogeography: a case study using the globally distributed plant family Sapindaceae. *Journal of Biogeography,* **38,** 531-550.

Caetano, D.S., O'Meara, B.C. & Beaulieu, J.M. (2018) Hidden state models improve state-dependent diversification approaches, including biogeographical models. *Evolution,* **72,** 2308-2324.

Castresana, J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution,* **17,** 540-552.

Chamberlain, S., Ram, K., Barve, V. & Mcglinn, D. (2016) rgbif: Interface to the Global "Biodiversity" Information Facility "API". R package version 0.9. 8.

Chamberlain, S.A. & Boettiger, C. (2017) R Python, and Ruby clients for GBIF species occurrence data. PeerJ Preprints.

Chamberlain, S.A. & Szocs, E. (2013) taxize: taxonomic search and retrieval in R. *F1000Research,* **2,** 191-191.

Charif, D., Thioulouse, J., Lobry, J.R. & Perriere, G. (2005) Online synonymous codon usage analyses with the ade4 and seqinR packages. *Bioinformatics,* **21,** 545-547.

Clark, J.R., Ree, R.H., Alfaro, M.E., King, M.G., Wagner, W.L. & Roalson, E.H. (2008) A comparative study in ancestral range reconstruction methods: retracing the uncertain histories of insular lineages. *Systematic Biology,* **57,** 693-707.

Davis, A.L., Scholtz, C.H. & Sole, C.L. (2016) Biogeographical and co-evolutionary origins of scarabaeine dung beetles: Mesozoic vicariance versus Cenozoic dispersal and dinosaur versus mammal dung. *Biological Journal of the Linnean Society,* **120,** 258-273.

Davis, A.L.V. & Scholtz, C.H. (2001) Historical vs. ecological factors influencing global patterns of scarabaeine dung beetle diversity. *Diversity and Distributions,* **7,** 161-174.

Davis, A.L.V., Scholtz, C.H. & Philips, T.K. (2002) Historical biogeography of scarabaeine dung beetles. *Journal of Biogeography,* **29,** 1217-1256.

Duchene, S., Molak, M. & Ho, S.Y.W. (2014) ClockstaR: choosing the number of relaxed-clock models in molecular phylogenetic analysis. *Bioinformatics,* **30,** 1017-1019.
Dupin, J., Matzke, N.J., Särkinen, T., Knapp, S., Olmstead, R.G., Bohs, L. & Smith, S.D. (2017) Bayesian estimation of the global biogeographical history of the Solanaceae. *Journal of Biogeography,* **44,** 887-899.
Goldberg, E.E., Lancaster, L.T. & Ree, R.H. (2011) Phylogenetic Inference of Reciprocal Effects between Geographic Range Evolution and Diversification. *Systematic Biology,* **60,** 451-465.
Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010) New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology,* **59,** 307-321.
Gunter, N.L., Monteith, G.B., Cameron, S.L. & Weir, T.A. (2018) Evidence from Australian mesic zone dung beetles supports their Gondwanan origin and Mesozoic diversification of the Scarabaeinae. *Insect Systematics & Evolution,* **1,** 1-27.
Gunter, N.L., Weir, T.A., Slipinksi, A., Bocak, L. & Cameron, S.L. (2016) If Dung Beetles (Scarabaeidae: Scarabaeinae) Arose in Association with Dinosaurs, Did They Also Suffer a Mass Co-Extinction at the K-Pg Boundary? *Plos One,* **11**.
Hanski, I. & Cambefort, Y. (1991) Dung beetle ecology. *Dung beetle ecology***.,** i-xii, 1-481.
Hosner, P.A., Braun, E.L. & Kimball, R.T. (2015) Land connectivity changes and global cooling shaped the colonization history and diversification of New World quail (Aves: Galliformes: Odontophoridae). *Journal of Biogeography,* **42,** 1883-1895.
Kendall, D.G. (1949) Stochastic processes and population growth. *Journal of the Royal Statistical Society. Series B (Methodological),* **11,** 230-282.
Kim, S.I. & Farrell, B.D. (2015) Phylogeny of world stag beetles (Coleoptera: Lucanidae) reveals a Gondwanan origin of Darwin's stag beetle. *Molecular Phylogenetics and Evolution,* **86,** 35-48.
Klaus, K.V. & Matzke, N.J. (2019) Statistical Comparison of Trait-dependent Biogeographical Models indicates that Podocarpaceae Dispersal is influenced by both Seed Cone Traits and Geographical Distance. *Systematic Biology*.
Kodandaramaiah, U. (2010) Use of dispersal–vicariance analysis in biogeography–a critique. *Journal of Biogeography,* **37,** 3-11.
Krell, F.-T. (2000) The fossil record of Mesozoic and Tertiary Scarabaeoidea (Coleoptera: Polyphaga). *Invertebrate Systematics,* **14,** 871-905.
Krell, F.-T. (2006) Fossil record and evolution of Scarabaeoidea (Coleoptera: Polyphaga). *The Coleopterists Bulletin,* **60,** 120-143.
Landis, M.J., Matzke, N.J., Moore, B.R. & Huelsenbeck, J.P. (2013) Bayesian Analysis of Biogeography when the Number of Areas is Large. *Systematic Biology,* **62,** 789-804.
Lanfear, R., Calcott, B., Ho, S.Y.W. & Guindon, S. (2012) PartitionFinder: Combined Selection of Partitioning Schemes and Substitution Models for Phylogenetic Analyses. *Molecular Biology and Evolution,* **29,** 1695-1701.
Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T. & Calcott, B. (2017) PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. *Molecular Biology and Evolution,* **34,** 772-773.
Larsson, A. (2014) AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics,* **30,** 3276-3278.

Losey, J.E. & Vaughan, M. (2006) The economic value of ecological services provided by insects. *Bioscience,* **56,** 311-323.
Matzke, N.J. (2013a) BioGeoBEARS: BioGeography with Bayesian (and likelihood) evolutionary analysis in R Scripts. *R package, version 0.2,* **1,** 2013.
Matzke, N.J. (2013b) Probabilistic Historical Biogeography: New Models for Founder-Event Speciation, Imperfect Detection, and Fossils Allow Improved Accuracy and Model-Testing. *Frontiers of Biogeography,* **5,** 242-248.
Matzke, N.J. (2014) Model Selection in Historical Biogeography Reveals that Founder-Event Speciation Is a Crucial Process in Island Clades. *Systematic Biology*, **63,** 951-970.
McKenna, D.D., Farrell, B.D., Caterino, M.S., Farnum, C.W., Hawks, D.C., Maddison, D.R., Seago, A.E., Short, A.E.Z., Newton, A.F. & Thayer, M.K. (2015) Phylogeny and evolution of Staphyliniformia and Scarabaeiformia: forest litter as a stepping stone for diversification of nonphytophagous beetles. *Systematic Entomology,* **40,** 35-60.
Miraldo, A., Hewitt, G.M., Paulo, O.S. & Emerson, B.C. (2011) Phylogeography and demographic history of Lacerta lepida in the Iberian Peninsula: multiple refugia, range expansions and secondary contact zones. *Bmc Evolutionary Biology,* **11**.
Miraldo, A., Wirta, H. & Hanski, I. (2011) Origin and diversification of dung beetles in Madagascar. *Insects,* **2,** 112-127.
Mlambo, S., Sole, C.L. & Scholtz, C.H. (2014) Affinities of the Canthonini dung beetles of the Eastern Arc Mountains. *Organisms Diversity & Evolution,* **14,** 115-120.
Mlambo, S., Sole, C.L. & Scholtz, C.H. (2015) A molecular phylogeny of the African Scarabaeinae (Coleoptera: Scarabaeidae). *Arthropod Systematics & Phylogeny,* **73,** 303-321.
Moczek, A. (2011) Evolution and development: onthophagus beetles and the evolutionary developmental genetics of innovation, allometry and plasticity. *Ecology and evolution of dung beetles***,** 126-151.
Monaghan, M.T., Inward, D.J.G., Hunt, T. & Vogler, A.P. (2007) A molecular phylogenetic analysis of the Scarabaeinae (dung beetles). *Molecular Phylogenetics and Evolution,* **45,** 674-692.
Moran, P. (1951) Estimation methods for evolutive processes. *Journal of the Royal Statistical Society: Series B (Methodological),* **13,** 141-146.
Nichols, E., Spector, S., Louzada, J., Larsen, T., Amequita, S., Favila, M.E. & Scarabaeinae Res, N. (2008) Ecological functions and ecosystem services provided by Scarabaeinae dung beetles. *Biological Conservation,* **141,** 1461-1474.
Ocampo, F.C. & Hawks, D.C. (2006) Molecular phylogenetics and evolution of the food relocation behaviour of the dung beetle tribe Eucraniini (Coleoptera : Scarabaeidae : Scarabaeinae). *Invertebrate Systematics,* **20,** 557-570.
Paradis, E., Claude, J. & Strimmer, K. (2004) APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics,* **20,** 289-290.
Plummer, M., Best, N., Cowles, K. & Vines, K. (2004) CODA: Output Analysis and Diagnostics for MCMC, R package version 0.13-3, 2008. *CRAN: The Comprehensive R Archive Network*.
R Development Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ree, R.H. & Sanmartín, I. (2018) Conceptual and statistical problems with the DEC+ J model of founder-event speciation and its comparison with DEC via model selection. *Journal of Biogeography,* **45,** 741-749.
Ree, R.H. & Smith, S.A. (2008) Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology,* **57,** 4-14.
Reguero, M.A., Gelfo, J.N., López, G.M., Bond, M., Abello, A., Santillana, S.N. & Marenssi, S.A. (2014) Final Gondwana breakup: the Paleogene South American native ungulates and the demise of the South America–Antarctica land connection. *Global and Planetary Change,* **123,** 400-413.
Ronquist, F. (1997) Dispersal-vicariance analysis: A new approach to the quantification of historical biogeography. *Systematic Biology,* **46,** 195-203.
Sanmartin, I. & Ronquist, F. (2004) Southern hemisphere biogeography inferred by event-based models: plant versus animal patterns. *Systematic Biology,* **53**.
Sayers, E.W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K.D. & Karsch-Mizrachi, I. (2019) GenBank. *Nucleic Acids Research,* **47,** D94-D99.
Schöfl, G. (2016) reutils: Talk to the NCBI EUtils. *R package, version 0.2.3*.
Scholtz, C. & Chown, S. (1995) The evolution of habitat use and diet in the Scarabaeoidea: a phylogenetic approach [pp. 355–374]. *Biology, Phylogeny, and Classification of Coleoptera: Papers Celebrating the 80th Birthday of Roy A. Crowson. J. Pakaluk and SA Slipinski (eds.). Muzeum i Instytut Zoologii PAN, Warszawa, Poland*.
Scholtz, C.H., Davis, A.L.V. & Kryger, U. (2009) *Evolutionary biology and conservation of dung beetles*. Pensoft Sofia.
Schwery, O. & O'Meara, B.C. (2016) MonoPhy: a simple R package to find and visualize monophyly issues. *PeerJ Computer Science,* **2,** e56.
Scotese, C. (1993) Maps of generalized continental positions and orography produced by the palaeomap project, University of Texas, Arlington. *Terrestrial ecosystems through time*.
Seton, M., Müller, R., Zahirovic, S., Gaina, C., Torsvik, T., Shephard, G., Talsma, A., Gurnis, M., Turner, M. & Maus, S. (2012) Global continental and ocean basin reconstructions since 200 Ma. *Earth-Science Reviews,* **113,** 212-270.
Smith, S.A. & O'Meara, B.C. (2012) treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics,* **28,** 2689-2690.
Sole, C.L. & Scholtz, C.H. (2010) Did dung beetles arise in Africa? A phylogenetic hypothesis based on five gene regions. *Molecular Phylogenetics and Evolution,* **56,** 631-641.
Sole, C.L., Wirta, H., Forgie, S.A. & Scholtz, C.H. (2011) Origin of Madagascan Scarabaeini dung beetles (Coleoptera: Scarabaeidae): dispersal from Africa. *Insect Systematics & Evolution,* **42,** 29-40.
Spector, S. (2006) Scarabaeine dung beetles (Coleoptera: Scarabaeidae: Scarabaeinae): an invertebrate focal taxon for biodiversity research and conservation. *The Coleopterists Bulletin,* **60,** 71-83.
Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics,* **30,** 1312-1313.
Stuart, A.J. (2015) Late Quaternary megafaunal extinctions on the continents: a short review. *Geological Journal,* **50,** 338-363.

Tarasov, S. & Dimitrov, D. (2016) Multigene phylogenetic analysis redefines dung beetles relationships and classification (Coleoptera: Scarabaeidae: Scarabaeinae). *Bmc Evolutionary Biology,* **16**.
Tarasov, S. & Génier, F. (2015) Innovative Bayesian and parsimony phylogeny of dung beetles (Coleoptera, Scarabaeidae, Scarabaeinae) enhanced by ontology-based partitioning of morphological characters. *Plos One,* **10,** e0116671.
Tarasov, S., Vaz-de-Mello, F.Z., Krell, F.T. & Dimitrov, D. (2016) A review and phylogeny of Scarabaeine dung beetle fossils (Coleoptera: Scarabaeidae: Scarabaeinae), with the description of two Canthochilum species from Dominican amber. *Peerj,* **4**.
Toussaint, E.F., Bloom, D. & Short, A.E. (2017) Cretaceous West Gondwana vicariance shaped giant water scavenger beetle biogeography. *Journal of Biogeography,* **44,** 1952-1965.
Toussaint, E.F.A., Seidel, M., Arriaga-Varela, E., Hajek, J., Kral, D., Sekerka, L., Short, A.E.Z. & Fikacek, M. (2017) The peril of dating beetles. *Systematic Entomology,* **42,** 1-10.
Vaz-de-Mello, F. (2007) Revision taxonomica e analysis phylogenetico de la tribu Ateuchini. *Xalapa, Veracruz, Mexico: Instituto de Ecologia AC*.
Wilkinson, S.P., Davy, S.K., Bunce, M. & Stat, M. (2018) Taxonomic identification of environmental DNA with informatic sequence classification trees.
Wirta, H., Orsini, L. & Hanski, I. (2008) An old adaptive radiation of forest dung beetles in Madagascar. *Molecular Phylogenetics and Evolution,* **47,** 1076-1089.
Wirta, H., Viljanen, H., Orsini, L., Montreuil, O. & Hanski, I. (2010) Three parallel radiations of Canthonini dung beetles in Madagascar. *Molecular Phylogenetics and Evolution,* **57,** 710-727.

Appendix D

Table IV-1 GenBank Accession Numbers

This will be a table of all GenBank Accession numbers used in the final alignments.

Name	16S	18S	28S-D2	28S-D3	CAD	COI-1	COI-2
Paraphytus sp.		KX512471.1	KX512561.1	KX512638.1		KJ867466.1	
Frankenbergerius armatus	GQ289676.1	KX512504.1	KX512593.1	KX512664.1	KX530439.1		GQ290031.1
Sarophorus costatus	AY131523.1	KX512496.1	GQ289814.1	AY131712.1	GQ289965.1	KF956280.1	AY131883.1
Sarophorus tuberculatus	AY131524.1	JN619266.1		AY131713.1			AY131884.1
Delopleurus pullus	GQ289700.1		GQ289833.1	GQ289918.1	GQ289976.1		GQ290052.1
Coptorhina nitidipennis	GQ289683.1		GQ289818.1	GQ289901.1			GQ290037.1
Odontoloma pusillum	AY131468.1	JN619208.1	GQ289790.1	AY131661.1	GQ289930.1		AY131839.1
Aliuscanthoniola similaris	JN804741.1		JN804810.1	KC928076.1			JN804667.1
Outenikwanus tomentosus	GQ289748.1		GQ289798.1	GQ289886.1	GQ289948.1		GQ290024.1
Peckolus alpinus	GQ289731.1		GQ289780.1	GQ289869.1	GQ289953.1		GQ290004.1
Endroedyolus paradoxus	GQ289702.1		GQ289752.1	GQ289838.1	GQ289921.1		GQ289981.1
Silvaphilus oubosiensis	JN804748.1		JN804822.1	KC928077.1	JN804606.1		JN804677.1
Namakwanus davisi	GQ289745.1		GQ289802.1	GQ289883.1	GQ289956.1		GQ290021.1
Byrrhidium convexum	GQ289732.1		GQ289782.1	GQ289871.1	GQ289938.1		GQ290006.1
Byrrhidium namaquensis	GQ289724.1		GQ289771.1	GQ289860.1	GQ289931.1		GQ289998.1
Dicranocara deschodti	EF656672.1	JN969180.1	GQ289793.1	EF656714.1	GQ289944.1		DQ667017.1
Dicranocara tatasensis							DQ667010.1
Dicranocara inexpectata							DQ667014.1
Boletoscapter cornutus	AY131441.1	JN619225.1		AY131632.1			AY131813.1
Uroxys pygmaeus	AY131529.1			EF656712.1			EF656761.1
Bdelyropsis bowditchi	EF656654.1	JN619144.1		EF656696.1			EF656745.1

Table IV-1 Continued

Name	16S	18S	28S-D2	28S-D3	CAD	COI-1	COI-2
Uroxys micros	AY131528.1			AY131717.1		AY144774.1	AY131886.1
Diorygopyx tibialis	AY847522.1						
Diorygopyx simpliciclunis	AY131456.1	JN619181.1		KF802118.1			KF801954.1
Pseudignambia sp							
Diorygopyx incomptus	AY847524.1			KF802119.1			KF801955.1
Diorygopyx niger	AY847523.1	KX512476.1	KX512566.1	KF802106.1			KF801942.1
Onthobium cookii							
Ignambia fasciculata	AY131463.1	JN619175.1		AY131654.1			AY131834.1
Pseudonthobium fracticolloides	AY131477.1	JN619182.1		AY131669.1			AY131847.1
Paronthobium simplex	AY131473.1	JN619173.1		AY131665.1			AY131843.1
Anonthobium tibiale	AY131439.1	JN619174.1		AY131630.1			AY131811.1
Demarziella interrupta	AY131511.1	JN619178.1		AY131700.1			
Demarziella mirifica	AY131512.1			AY131701.1			AY131872.1
Amphistomus complanatus	AY131436.1	JN619228.1					AY131808.1
Coptodactyla meridionalis							MG588139.1
Coptodactyla storeyi	AY131497.1						
Coptodactyla glabricollis	AY131496.1	JN619207.1		AY131687.1			AY131863.1
Ochicanthon punctatus	AY131474.1	JN619188.1		AY131666.1			AY131844.1
Epactoides mangabeensis	EU030497.1			EU030542.1			EU030586.1
Epactoides hanskii	EU030504.1	GQ341988.1		EU030549.1			EU030593.1
Epactoides helenae	EU030507.1	GQ341989.1		EU030552.1			EU030595.1
Epactoides mahaboi	EU030515.1	GQ341991.1		EU030560.1			EU030599.1
Epactoides semiaeneus	EU030529.1			EU030573.1			EU030610.1
Epactoides tiinae	EU030533.1	GQ341995.1		EU030576.1			EU030614.1
Epactoides rahagai	EU030527.1	GQ341994.1		EU030570.1			EU030608.1

Table IV-1 Continued

Name	16S	18S	28S-D2	28S-D3	CAD	COI-1	COI-2
Epactoides major	EU030519.1	GQ341992.1		DQ369542.1			EU030601.1
Epactoides femoralis	EU030501.1			EU030546.1			EU030589.1
Epactoides viridicollis	DQ369609.1	DQ369660.1		DQ369543.1			EU030617.1
Epactoides lissus	EU030514.1			EU030557.1			EU030598.1
Epactoides incertus	EU030509.1	GQ341990.1		EU030554.1			EU030597.1
Epactoides perrieri	EU030524.1			EU030569.1			EU030606.1
Epactoides spinicollis	EU030532.1			EU030574.1		KF309738.1	EU030612.1
Epactoides frontalis	DQ369607.1	DQ369661.1		DQ369541.1	KF309818.1		EU030590.1
Coproecus hemisphaericus	AY131451.1	JN619177.1		AY131641.1			AY131821.1
Lepanus occidentalis	KF801733.1			KF802062.1			KY784156.1
Lepanus villosus							KY784174.1
Lepanus dichrous							KY784155.1
Lepanus nitidus	AY131464.1			AY131655.1			AY131835.1
Lepanus australis	KF801821.1			KF802149.1			KF801986.1
Lepanus palumensis							KY784177.1
Lepanus politus	AY847533.1						KY784151.1
Lepanus pygmaeus						AY144789.1	KY784172.1
Lepanus monteithi							KY784157.1
Lepanus globulus	KF801822.1			KF802150.1		AY144776.1	KF801987.1
Lepanus ustulatus		JN619265.1	KX512567.1	AY131656.1	KX530432.1		KY784152.1
Sauvagesinella palustris	KF801735.1			KF802064.1			KF801906.1
Sauvagesinella becki	KF801737.1			KF802066.1			KF801899.1
Sauvagesinella monstrosa	KF801736.1			KF802065.1			KF801898.1
Monoplistes curvipes	AY131467.1	JN619186.1		AY131659.1			
Canthonosoma macleayi		KX512474.1	KX512564.1	KX512641.1			

Table IV-1 Continued

Name	16S	18S	28S-D2	28S-D3	CAD	COI-1	COI-2
Canthonosoma castelnaui	AY131447.1			AY131638.1			AY131818.1
Saphobius squamulosus	AY131480.1			AY131672.1			
Saphobius setosus	AY131479.1	JN619310.1		AY131671.1			
Cephalodesmius armiger	AY131448.1	AY745577.1					
Cephalodesmius quadridens	AY131449.1			AY131639.1			AY131819.1
Monteithocanthon glaber	EF656646.1			EF656688.1		AY144792.1	KY784166.1
Temnoplectron boucomonti						AY144791.1	AY144792.1
Temnoplectron major						AY144788.1	AY144791.1
Temnoplectron disruptum						AY144785.1	AY144788.1
Temnoplectron cooki						AY144787.1	AY144785.1
Temnoplectron reyi						AY144786.1	AY144787.1
Temnoplectron politulum	AY131484.1			AY131676.1		AY144782.1	AY144786.1
Temnoplectron aeneopiceum						AY144783.1	AY144782.1
Temnoplectron subvolitans						AY144779.1	AY144783.1
Temnoplectron monteithi						AY144790.1	AY144794.1
Temnoplectron finnigani	AY131483.1	JN619220.1		AY131675.1		AY144771.1	AY131851.1
Temnoplectron lewisense							AY144795.1
Janssensantus sp							
Tanzanolus sp							
Gyronotus pumilus	GQ289714.1	GQ342009.1	GQ289765.1	GQ289851.1	GQ289950.1		GQ342148.1
Canthodimorpha lawrencei	GQ289744.1		GQ289797.1	GQ289889.1			GQ290018.1
Anachalcos convexus	AY131437.1	GQ341942.1	GQ289768.1	AY131628.1	GQ289929.1		AY131809.1
Anachalcos suturalis	AY131438.1			AY131629.1			AY131810.1
Paragymnopleurus maurus	AY131545.1	JN619216.1		AY131733.1			AY131902.1
Paragymnopleurus striatus	AY131546.1			AY131734.1			AY131903.1

Table IV-1 Continued

Name	16S	18S	28S-D2	28S-D3	CAD	COI-1	COI-2
Allogymnopleurus thalassinus	JN804680.1	AY821521.1	JN804751.1	AY131729.1	JN804566.1		AY131898.1
Garreta nitens	AY131542.1		JN804773.1	AY131730.1	JN804579.1		AY131899.1
Gymnopleurus virens	AY131543.1	JN619219.1		AY131731.1			AY131900.1
Gymnopleurus humanus	JN804700.1		JN804772.1				JN804627.1
Gymnopleurus flagellatus						KT454096.1	AY039364.1
Gymnopleurus mopsus	KJ721833.1			KJ721895.1		DQ369595.1	AY039363.1
Arachnodes kelifelyi	DQ369620.1	DQ369673.1		DQ369554.1			GQ342120.1
Arachnodes grossepunctatus	GQ341890.1			GQ342052.1			GQ342119.1
Arachnodes pusillus	GQ341901.1	GQ341970.1		GQ342064.1			GQ342126.1
Arachnodes philippi	GQ341899.1	GQ341968.1		GQ342062.1			
Arachnodes micheli	GQ341925.1	GQ342004.1		GQ342091.1		DQ369590.1	GQ342143.1
Arachnodes andriai	DQ369614.1	DQ369666.1		DQ369548.1		DQ369598.1	GQ342115.1
Arachnodes manaitrai	DQ369625.1	GQ341964.1		DQ369559.1			GQ342124.1
Pseudoarachnodes mantillerii	GQ341937.1			GQ342107.1		DQ369600.1	
Pseudoarachnodes hanskii	GQ341935.1	DQ369678.1		GQ342105.1			GQ342155.1
Pseudoarachnodes semichalceus	GQ341938.1			GQ342108.1			GQ342156.1
Pseudoarachnodes insularis	GQ341936.1	GQ342029.1		GQ342106.1			
Arachnodes genieri	GQ341923.1	GQ342002.1		GQ342089.1		DQ369591.1	GQ342141.1
Arachnodes apotolamproides	DQ369616.1	DQ369669.1	KX512574.1	DQ369550.1			GQ342136.1
Arachnodes antoetrae	DQ369615.1	DQ369667.1		DQ369549.1			EU248061.1
Arachnodes morio	GQ341897.1	GQ341967.1		GQ342060.1			GQ342125.1
Arachnodes colasi		GQ341957.1				DQ369597.1	GQ342116.1
Arachnodes sicardi	DQ369623.1	DQ369677.1		DQ369557.1			GQ342130.1
Arachnodes emmae	GQ341920.1	GQ342001.1		GQ342086.1			GQ342138.1
Arachnodes globuloides	GQ341889.1			GQ342051.1			GQ342118.1

Table IV-1 Continued

Name	16S	18S	28S-D2	28S-D3	CAD	COI-1	COI-2
Arachnodes seminitidus	GQ341904.1	GQ341974.1		GQ342068.1			GQ342129.1
Arachnodes delphinensis	GQ341919.1	GQ342000.1		GQ342085.1			EU248060.1
Arachnodes splendidus	EF656678.1	GQ342007.1		EF656720.1			GQ342146.1
Arachnodes prasinus	GQ341926.1	GQ342005.1		GQ342092.1			GQ342144.1
Arachnodes ruteri	GQ341927.1			GQ342093.1		DQ369593.1	GQ342145.1
Arachnodes cuprarius	DQ369618.1	DQ369671.1		DQ369552.1			DQ369478.1
Arachnodes mantasoae	GQ341924.1	GQ342003.1		GQ342090.1			GQ342142.1
Arachnodes mahafalyensis	GQ341893.1	GQ341963.1		GQ342056.1			GQ342123.1
Arachnodes saprinoides	DQ369621.1	DQ369675.1		DQ369555.1			GQ342128.1
Arachnodes purpuricollis	GQ341900.1	GQ341969.1		GQ342063.1			
Arachnodes luctuosus				GQ342054.1			GQ342121.1
Arachnodes robinsoni		GQ341971.1		GQ342065.1			GQ342127.1
Arachnodes dichrous	DQ369624.1	DQ369668.1		DQ369558.1			
Litocopris muticus	JN804711.1		JN804782.1		JN804585.1		JN804639.1
Copris jacchoides							JN804618.1
Copris hispanicus							AY039366.1
Copris sonensis						KM441825.1	MG642091.1
Copris lunaris							AY039365.1
Copris agnus	AY131490.1			AY131682.1			AY131857.1
Copris laeviceps	AY131492.1						
Paracopris punctulatus		EF188042.1					EF188219.1
Copris lugubris	AY131493.1	AY821529.1		AY131684.1			AY131860.1
Copris aeneus	AY131489.1			AY131681.1			AY131856.1
Copris cornifrons	JN804686.1		JN804759.1				JN804613.1
Copris amyntor	AY131491.1		DQ430882.1	AY131683.1			AY131858.1

Table IV-1 Continued

Name	16S	18S	28S-D2	28S-D3	CAD	COI-1	COI-2
Copris confucius		EF187978.1		EF188048.1			EF188135.1
Copris sinicus	AY131495.1			AY131686.1			AY131862.1
Deltochilum carinatum	AY131453.1			AY131644.1			AY131824.1
Deltochilum barbipes				AY131643.1			AY131823.1
Deltochilum pseudoparile	AY131455.1			AY131646.1			AY131826.1
Deltochilum mexicanum		KX512461.1	DQ430879.1	KX512630.1			
Deltochilum amazonicum	AY131452.1			AY131642.1			AY131822.1
Deltochilum gibbosum	AY131454.1	DQ012274.1	DQ430878.1	DQ430930.1			AY131825.1
Eudinopus dytiscoides	AY131461.1	DQ430832.1	DQ430880.1	AY131652.1			AY131832.1
Hansreia affinis	AY131462.1	JN619191.1		AY131653.1			AY131833.1
Megathoposoma candezei	AY131465.1	JN619142.1		AY131657.1			AY131836.1
Megathopa villosa			DQ430875.1	DQ430927.1			
Malagoniella puncticollis		DQ430830.1	DQ430874.1	DQ430926.1			
Scybalophagus plicatipennis		KX512440.1	DQ430876.1	DQ430928.1	KX530423.1		
Scybalophagus rugosus			DQ430877.1	DQ430929.1			
Canthon cyanellus	KX807615.1						KX807648.1
Canthon lamprimus	EF656648.1			EF656690.1			EF656739.1
Canthon perseverans	GQ341916.1	GQ341985.1		GQ342080.1			
Scybalocanthon pygidialis	AY131481.1	JN619189.1		AY131673.1			AY131849.1
Canthon deyrollei	GQ341915.1	GQ341984.1		GQ342078.1			
Canthon indigaceus	AY131443.1			AY131634.1			AY131814.1
Canthon aequinoctialis	GQ341914.1	GQ341983.1	KX512553.1	KX512631.1			
Canthon luteicollis	AY131444.1			AY131635.1			AY131815.1
Canthon viridis	AY131446.1	AY821531.1		AY131637.1		MG321659.1	AY131817.1
Phanaeus sororibispinus						EU477307.1	MG321647.1

Table IV-1 Continued

Name	16S	18S	28S-D2	28S-D3	CAD	COI-1	COI-2
Phanaeus alvarengai			EU432232.1			EU477294.1	
Phanaeus haroldi			EU432221.1			EU477296.1	
Phanaeus melibaeus			EU432224.1			EU477309.1	
Phanaeus kirbyi			EU432234.1			EU477313.1	
Phanaeus paleano			EU432235.1			MG321655.1	
Oxysternon conspicillatum	AY131608.1	JN619205.1	MG321631.1	AY131792.1		EU477358.1	AY131948.1
Oxysternon durantoni			EU432273.1			MG321670.1	
Oxysternon macleayi						EU477357.1	
Oxysternon festivum			EU432272.1			EU477291.1	
Phanaeus splendidulus			EU432222.1			EU477292.1	
Phanaeus dejeani			EU432223.1			EU477362.1	
Oxysternon silenus			EU432271.1			MG321656.1	
Diabroctis mirabilis			MG321632.1				MG321645.1
Sulcophanaeus batesi			DQ430856.1	DQ430910.1		MG321673.1	
Coprophanaeus bellicosus						MG321651.1	MG321638.1
Coprophanaeus bonariensis			MG321627.1			MG321667.1	MG321640.1
Coprophanaeus lancifer	AY131604.1		MG321633.1	AY131788.1		MG321650.1	AY131945.1
Coprophanaeus ensifer			MG321624.1				MG321641.1
Dendropaemon bahianus							
Coprophanaeus chiriquensis		KX512455.1	KX512545.1	KX512628.1	KX530427.1	EU477349.1	
Coprophanaeus pluto			EU432265.1			EU477353.1	
Coprophanaeus ignecinctus			EU432267.1			MG321671.1	
Coprophanaeus telamon	AY131605.1	JN619140.1		AY131789.1		MG321663.1	AY131946.1
Coprophanaeus dardanus						EU477300.1	
Phanaeus achilles			EU432228.1			EU477321.1	

Table IV-1 Continued

Name	16S	18S	28S-D2	28S-D3	CAD	COI-1	COI-2
Phanaeus prasinus			EU432240.1			EU477299.1	
Phanaeus chalcomelas			EU432226.1			EU477306.1	
Phanaeus lecourti			EU432231.1			EU477304.1	
Phanaeus meleagris			EU432230.1				
Phanaeus demon	AY131610.1						AY131950.1
Phanaeus amethystinus amethystinus						EU477329.1	
Phanaeus lunaris			EU432249.1			EU477337.1	
Phanaeus howdeni			EU432255.1			EU477338.1	
Phanaeus sallei	AY131611.1	AY821524.1	EU432256.1	AY131793.1		EU477333.1	AY131951.1
Phanaeus amithaon			EU432253.1				
Phanaeus wagneri wagneri						EU477319.1	
Phanaeus pyrois			EU432239.1			EU477317.1	
Phanaeus endymion			EU432237.1			EU477345.1	
Phanaeus igneus		DQ430828.1	DQ430855.1	DQ430909.1		EU477347.1	
Phanaeus vindex			EU432264.1				
Phanaeus triangularis texensis						EU477341.1	
Phanaeus quadridens			EU432259.1			EU477336.1	
Phanaeus yecoraensis			EU432254.1			EU477323.1	
Phanaeus nimrod			EU432241.1			EU477324.1	
Phanaeus furiosus			EU432244.1				
Ennearabdus lobocephalus	AY131532.1	DQ430829.1	DQ430861.1	AY131721.1			AY131889.1
Eucranium arachnoides	AY131533.1	AY821527.1	DQ430859.1	AY131722.1			AY131890.1
Eucranium planicolle			DQ430860.1	DQ430913.1			
Anomiopsoides cavifrons			DQ430865.1	DQ430918.1			
Glyphoderus sterquilinus	AY131534.1		DQ430863.1	AY131723.1			AY131891.1

Table IV-1 Continued

Name	16S	18S	28S-D2	28S-D3	CAD	COI-1	COI-2
Anomiopsoides biloba	AY131530.1		DQ430864.1	AY131719.1			AY131887.1
Glyphoderus centralis			DQ430862.1	DQ430915.1			
Anomiopsoides heteroclyta	AF499693.1	JN619203.1	DQ430866.1	AY131720.1			AY131888.1
Canthidium haroldi	AY131506.1			AY131695.1			AY131868.1
Kheper subaeneus	AF499696.1						
Kheper nigroaeneus	AF499695.1	AY821523.1	DQ430870.1	AY131795.1			AY131953.1
Kheper bonellii	JN804740.1		JN982317.1		JN819271.1	AY258257.1	JN804665.1
Pachysoma valeflorae						AY258247.1	AY258257.1
Pachysoma schinzi						AY258213.1	AY258247.1
Pachysoma aesculapius						AY258223.1	AY258213.1
Pachysoma endroeydi						AY258226.1	AY258223.1
Pachysoma glentoni						AY258215.1	AY258226.1
Pachysoma hippocrates	AF499699.1		DQ430868.1	DQ430920.1		AY258253.1	AY258215.1
Pachysoma denticollis						AY258241.1	AY258253.1
Pachysoma rotundigenus						AY258244.1	AY258241.1
Pachysoma rodriguesi						AY258250.1	AY258244.1
Pachysoma striatus						AY258231.1	AY258250.1
Pachysoma gariepinus						AY258238.1	AY258231.1
Pachysoma bennigseni	AF499698.1						AY258238.1
Scarabaeus cicatricosus							AY039362.1
Drepanopodus costatus	AY131612.1	JN619206.1		AY131794.1			AY131952.1
Scarabaeus galenus	AF499704.1			AY131798.1		AY965239.1	AY131956.1
Drepanopodus proximus	AF499694.1						AY965239.1
Scarabaeus bohemani	AF499701.1						
Scarabaeus flavicornis	AF499702.1						

Table IV-1 Continued

Name	16S	18S	28S-D2	28S-D3	CAD	COI-1	COI-2
Scarabaeus brittoni	AY131618.1	JN619308.1		AY131800.1			AY131958.1
Scarabaeus adamastor	AF499711.1						
Scarabaeus hippias	AY131619.1			AY131801.1			AY131959.1
Pachylomera femoralis							
Scarabaeus rugosus	AF499706.1						FJ763730.1
Scarabaeus caffer	JN804737.1		JN804806.1		JN819270.1		
Scarabaeus westwoodi	AF499709.1					GU305943.1	
Scarabaeus viettei	GU305941.1						GU305943.1
Scarabaeus goryi	AF499705.1						
Scarabaeus satyrus	AF499708.1					GU305942.1	
Scarabaeus radama	GU305940.1					KP752076.1	GU305942.1
Scarabaeus sacer						KP752063.1	KP752076.1
Scarabaeus pius						KP752070.1	KP752055.1
Scarabaeus thphon							KP752068.1
Neateuchus proboscideus	AF499697.1						
Scarabaeus deludens		KP419281.1	DQ430869.1	GU226585.1			
Scarabaeus zambezianus	AF499710.1		JN804808.1		JN819268.1		
Leotrichillum sp							
Trichillidium pilosum		KX512459.1	KX512549.1				
Canthidium guanacaste	AY131505.1	AY821530.1		AY131694.1			AY131867.1
Canthidium rufinum	AY131507.1			AY131696.1			AY131869.1
Canthidium thalassinum	AY131508.1			AY131697.1			AY131870.1
Dichotomius boreus	AY131514.1			AY131703.1		HQ824542.1	AY131874.1
Dichotomius laevicollis	HQ824536.1						
Dichotomius yucatanus	AY131516.1		DQ430850.1	AY131705.1			AY131876.1

Table IV-1 Continued

Name	16S	18S	28S-D2	28S-D3	CAD	COI-1	COI-2
Dichotomius parcepunctatus	AY131515.1			AY131704.1		HQ824539.1	AY131875.1
Dichotomius geminatus	HQ824533.1					HQ824541.1	
Dichotomius semisquamosus	HQ824535.1					HQ824540.1	
Dichotomius sericeus	HQ824534.1						
Dichotomius nisus	HQ824538.1					HQ824543.1	
Dichotomius bos	HQ824537.1						
Ateuchus viduus							
Ateuchus ecuadorense	EF656650.1			EF656692.1			EF656741.1
Ateuchus chrysopyge	AY131502.1	JN619143.1		AY131692.1			AY131866.1
Ateuchus floridensis							
Eurysternus inflexus	AY131538.1			AY131726.1			AY131895.1
Eurysternus plebejus	AY131539.1			AY131727.1			AY131896.1
Eurysternus angustulus	AY131535.1	AY821533.1		AY131724.1			AY131892.1
Eurysternus velutinus	AY131540.1			AY131728.1			AY131897.1
Eurysternus caribaeus	AY131536.1			AY131725.1			AY131893.1
Eurysternus hamaticollis	AY131537.1			EF656708.1		KF309775.1	AY131894.1
Nanos minutus	EU247965.1	GQ342019.1		EU248016.1	KF309836.1	KF309746.1	EU248068.1
Nanos bicoloratus	EU247994.1			EU248047.1	KF309819.1	KF309755.1	EU248089.1
Nanos humeralis				KF309915.1	KF309825.1	KF309777.1	
Nanos rubrosignatus		GQ342022.1		EU248013.1	KF309838.1	KF309753.1	EU248065.1
Nanos hanskii	DQ369632.1	DQ369679.1		DQ369566.1	KF309824.1	KF309792.1	EU248074.1
Nanos nitens	EU247989.1			EU248041.1	KF309829.1	DQ369604.1	EU248182.1
Nanos viettei	DQ369631.1	DQ369684.1		DQ369565.1	KF309842.1	KF309751.1	EU248153.1
Nanos clypeatus	EU247970.1	GQ342013.1		EF656718.1	KF309822.1	KF309780.1	EU248106.1
Nanos dubitatus	DQ369629.1	DQ369682.1		DQ369563.1	KF309834.1	KF309767.1	EU248099.1

Table IV-1 Continued

Name	16S	18S	28S-D2	28S-D3	CAD	COI-1	COI-2
Nanos occidentalis	DQ369633.1	DQ369680.1		DQ369567.1	KF309830.1	KF309769.1	EU248087.1
Nanos peyrierasi	DQ369630.1	DQ369683.1		DQ369564.1	KF743767.1	KF309782.1	EU248092.1
Nanos vadoni	EU248003.1	GQ342024.1		EU248054.1	KF309841.1	KF309759.1	EU248097.1
Nanos manomboensis	EU247981.1	GQ342015.1		EU248033.1	KF743770.1	KF309747.1	EU248078.1
Nanos bimaculatus	DQ369628.1	DQ369681.1		DQ369562.1	KF309820.1	KF309750.1	EU248071.1
Nanos binotatus					KF309821.1	DQ369596.1	HM029151.1
Arachnodes semipunctatus	DQ369622.1	DQ369676.1		DQ369556.1		KF309739.1	DQ369481.1
Nanos semiscribosus	GQ341903.1	GQ341973.1		KF309931.1	KF309839.1	KF309703.1	GQ342150.1
Apotolamprus cyanescens	GQ341886.1	GQ341958.1		KF309887.1	KF309803.1	KF309707.1	GQ342111.1
Apotolamprus latipennis	GQ341873.1	GQ341946.1		GQ342034.1	KF309807.1	KF309733.1	GQ342113.1
Apotolamprus zombitsyensis	GQ341879.1	GQ341950.1		KF309904.1	KF309816.1	KF309710.1	
Apotolamprus metallicus				KF309894.1	KF309810.1	DQ369592.1	
Arachnodes balianus	DQ369617.1	DQ369670.1		DQ369551.1		KF309726.1	DQ369477.1
Apotolamprus milloti	GQ341876.1			KF309895.1	KF309811.1	KF309716.1	
Apotolamprus vadoni				KF309902.1		KF309702.1	
Apotolamprus ambohitsitondronensis				KF309885.1	KF309802.1	KF309709.1	
Apotolamprus marojejyensis	GQ341874.1			GQ342035.1	KF309809.1	KF309728.1	
Apotolamprus quadrimaculatus	GQ341877.1			KF309899.1	KF309813.1	KF309715.1	
Apotolamprus sericeus				KF309901.1	KF309815.1	KF309706.1	
Apotolamprus helenae	DQ369611.1	DQ369663.1		DQ369545.1	KF309806.1	KF309705.1	EU248063.1
Apotolamprus hanskii	DQ369610.1	DQ369662.1		DQ369544.1	KF309805.1	KF309714.1	EU248062.1
Apotolamprus quadrinotatus	DQ369612.1	DQ369664.1		DQ369546.1	KF309814.1	KF309704.1	EU030619.1
Apotolamprus darainaensis	GQ341872.1	GQ341943.1		KF309888.1	KF309804.1	KF309712.1	GQ342112.1
Apotolamprus orangeaensis				KF309896.1		KF309713.1	
Apotolamprus pseudomanomboensis				KF309898.1	KF309812.1	KF309708.1	

Table IV-1 Continued

Name	16S	18S	28S-D2	28S-D3	CAD	COI-1	COI-2
Apotolamprus manomboensis	GQ341895.1	GQ341965.1		GQ342058.1	KF309808.1	KF309779.1	GQ342114.1
Apotolamprus sahatezaensis				KF309932.1	KF309840.1	KF309737.1	
Cambefortantus ranomafanensis	GQ341913.1	GQ341982.1		KF309905.1	KF309817.1		
Bohepilissus subtilis			GQ289784.1	GQ289873.1	GQ289940.1		GQ290008.1
Heliocopris andersoni	AY131518.1	AY821526.1		AY131707.1			AY131878.1
Heliocopris hamadryas	AY131519.1		GQ289827.1	AY131708.1	GQ289971.1		AY131879.1
Metacatharsius opacus	AY131498.1	JN619202.1		AY131688.1			AY131864.1
Metacatharsius exiguiformis	JN804720.1		JN804788.1		JN804589.1		JN804647.1
Metacatharsius marani	JN804721.1		JN804791.1		JN804591.1		JN804650.1
Catharsius philus	AY131487.1			AY131679.1			AY131854.1
Catharsius calaharicus	AY131485.1	JN619192.1		AY131677.1			AY131852.1
Catharsius sesostris	AY131488.1			AY131680.1		JQ855856.1	AY131855.1
Catharsius molossus	AY131486.1			KJ721870.1			AY131853.1
Epirinus aeneus	AY131458.1	JN619201.1	GQ289777.1	AY131649.1	GQ289935.1		AY131829.1
Epirinus hilaris	AY131459.1			AY131650.1			AY131830.1
Epirinus convexus	GQ289705.1		GQ289755.1	GQ289841.1	GQ289924.1		GQ289984.1
Epirinus silvestris	GQ289750.1		GQ289800.1	GQ289882.1	GQ289946.1		GQ290019.1
Epirinus aquilus			HQ289943.1		HQ289923.1		HQ289975.1
Epirinus sebastiani	HQ289917.1		HQ289965.1		HQ289939.1		HQ289976.1
Epirinus relictus	HQ289915.1		HQ289964.1		HQ289937.1		HQ290003.1
Epirinus ngomae	HQ289909.1		HQ289957.1		HQ289933.1		HQ289987.1
Epirinus hluhluwensis	HQ289906.1		HQ289955.1		HQ289931.1		HQ289995.1
Epirinus comosus	HQ289899.1		HQ289946.1		HQ289926.1		HQ289991.1
Epirinus pygidialis			HQ289963.1				HQ289994.1
Epirinus obtusus	HQ289912.1		HQ289962.1		HQ289935.1		HQ289982.1

Table IV-1 Continued

Name	16S	18S	28S-D2	28S-D3	CAD	COI-1	COI-2
Epirinus scrobiculatus	HQ289902.1		HQ289948.1		HQ289927.1		HQ289979.1
Epirinus sulcipennis	HQ289918.1		HQ289969.1		HQ289941.1		HQ289988.1
Epirinus flagellatus	HQ289903.1		HQ289951.1		HQ289928.1		HQ289983.1
Epirinus validus	HQ289920.1		HQ289971.1				HQ289984.1
Onitis caffer	AY131600.1		JN804798.1	EF656713.1			EF656762.1
Onitis ion							AY039340.1
Bubas bison	AY131595.1			AY131779.1			AY131938.1
Bubas bubalus	AY131596.1			AY131780.1			AY131939.1
Cheironitis scabrosus	JN804690.1		JN804763.1		JN804573.1		JN804616.1
Cheironitis hoplosternus	AY131597.1	AY821528.1		AY131781.1			AY131940.1
Onitis alexis	AY131599.1	DQ430835.1	DQ430888.1	AY131783.1			AY131942.1
Onitis falcatus	AY131601.1			AY131785.1			AY131943.1
Onitis subopacus		EF188030.1					EF188205.1
Hammondantus psammophilus	GQ289743.1	KX512501.1	GQ289796.1	GQ289881.1			GQ290017.1
Pycnopanelus krikkeni	GQ289708.1		GQ289761.1	GQ289845.1			GQ289987.1
Xinidium dentilabris	GQ289670.1	KX512493.1	GQ289805.1	GQ289890.1	GQ289958.1		GQ290025.1
Xinidium dewitzi	GQ289690.1		GQ289824.1	GQ289908.1	GQ289969.1		GQ290042.1
Macroderes amplior	GQ289698.1		GQ289834.1	GQ289916.1	GQ289978.1		GQ290050.1
Macroderes minutus	GQ289699.1		GQ289835.1	GQ289915.1	GQ289977.1		GQ290051.1
Macroderes mutilans	GQ289673.1		GQ289808.1	GQ289893.1	GQ289959.1		GQ290028.1
Macroderes mutilatus							
Dwesasilvasedis medinae	GQ289711.1	KX512508.1	GQ289762.1	GQ289848.1	GQ289926.1		GQ289990.1
Sisyphus fasciculatus							
Sisyphus crispatus	AY131624.1			AY131805.1		KM452277.1	AY131963.1
Sisyphus schaefferi	KJ721831.1			KJ721893.1			AY039367.1

Table IV-1 Continued

Name	16S	18S	28S-D2	28S-D3	CAD	COI-1	COI-2
Sisyphus seminulum	AY131627.1						AY131966.1
Neosisyphus fortuitus	AY131621.1	AY821534.1		AY131803.1			AY131961.1
Sisyphus gazanus	AY131626.1			AY131807.1			AY131965.1
Phalops rufosignatus	JN804732.1		JN804801.1		JN804601.1		JN804660.1
Phalops ardea	AY131592.1	JN619194.1		AY131776.1			AY131935.1
Onthophagus signatus	JN804727.1	EF188038.1		EF188124.1		EU162450.1	EF188215.1
Digitonthophagus gazella	AY131563.1	EF188036.1	DQ430884.1	FJ817954.1			AY131918.1
Proagoderus bicallosus							
Proagoderus schwaneri	AY131594.1			AY131778.1			AY131937.1
Parascatonomus penicillatus	DQ369540.1	DQ369585.1		DQ369522.1			EF188221.1
Onthophagus semiareus	AY131589.1			AY131773.1			AY131932.1
Onthophagus seniculus							
Serrophorus seniculus		EF188045.1		EF188131.1			EF188225.1
Onthophagus diabolicus							
Onthophagus avocetta		EF188031.1		EF188115.1			
Onthophagus elegans		EF188033.1		EF188117.1			EF188208.1
Proagoderus aciculatus	JN804735.1		JN804804.1		JN804602.1	EU162465.1	JN804663.1
Onthophagus rangifer	EU162562.1						
Onthophagus lanista	EU162551.1						
Onthophagus tersidorsis	EU162574.1						
Oniticellus planatus		EF188028.1		EF188113.1			EF188203.1
Liatongus vertagus		EF188025.1		EF188110.1			EF188201.1
Tiniocellus spinipes	AY131556.1	EF188046.1	JN804821.1	AY131743.1			AY131912.1
Proagoderus sapphirinus							
Tragiscus dimidiatus	AY131557.1	EF188047.1		AY131744.1			AY131913.1

Table IV-1 Continued

Name	16S	18S	28S-D2	28S-D3	CAD	COI-1	COI-2
Oniticellus egregius	AY131553.1		JN804795.1	AY131740.1	JN804595.1		AY131909.1
Drepanocerus orientalis							
Drepanocerus kirbyi	AY131549.1			AY131737.1			AY131906.1
Drepanocerus laticollis		EF187980.1		EF188049.1			EF188137.1
Drepanocerus bechynei	AY131548.1	JN619217.1		AY131736.1			AY131905.1
Eodrepanus bechynei						KM439695.1	
Euoniticellus fulvus	AY131554.1	AY821522.1		AY131741.1			AY131910.1
Euoniticellus triangulatus		EF187983.1		EF188050.1			EF188138.1
Euoniticellus intermedius	AY131550.1	JN619209.1	KX512609.1	AY131738.1			
Euoniticellus africanus	JN804699.1		JN804771.1		JN804578.1		JN804626.1
Tiniocellus sarawacus	AY131555.1	JN619222.1		AY131742.1			AY131911.1
Liatongus militaris	AY131552.1	EF188024.1	JN804783.1	AY131739.1	JN804587.1		EF188199.1
Liatongus phanaeoides	KJ721841.1			KJ721903.1			
Cyptochirus ambiguus	AY131547.1		JN804766.1	AY131735.1	JN804574.1		AY131904.1
Helictopleurus fungicola	EF187937.1	DQ369570.1		DQ369507.1			EF188157.1
Heterosyphus sicardi	EF187964.1	EF188013.1		EF188097.1			EF188188.1
Helictopleurus cribricollis	EF187922.1			EF188054.1			EF188143.1
Helictopleurus rudicollis	EF656675.1	DQ369580.1		FJ817927.1			EF188183.1
Helictopleurus quadripunctatus	AY131551.1	JN619263.1		EF656698.1			AY131907.1
Helictopleurus unifasciatus	DQ369539.1	DQ369584.1		DQ369521.1			EF188196.1
Helictopleurus fulgens	DQ369534.1	DQ369579.1		DQ369516.1			EF188156.1
Helictopleurus perrieri	DQ369532.1	DQ369577.1		DQ369514.1			EF188175.1
Helictopleurus splendidicollis	DQ369538.1	DQ369583.1		DQ369520.1			EF188191.1
Helictopleurus fissicollis	DQ369524.1	DQ369569.1		DQ369506.1			EF188151.1
Helictopleurus neuter	EF187951.1			EF188082.1			EF188170.1

Table IV-1 Continued

Name	16S	18S	28S-D2	28S-D3	CAD	COI-1	COI-2
Helictopleurus obscurus	DQ369531.1	DQ369576.1		DQ369513.1			DQ369458.1
Helictopleurus multimaculatus	DQ369528.1	DQ369573.1		DQ369510.1			DQ369455.1
Helictopleurus fasciolatus	EF187927.1	EF187989.1		EF188060.1			EF188149.1
Helictopleurus carbonarius	EF187918.1	EF187984.1		EF188052.1			EF188140.1
Helictopleurus marsyas	DQ369527.1	DQ369572.1		DQ369509.1			EF188163.1
Helictopleurus nicollei	DQ369530.1	DQ369575.1		DQ369512.1			EF188172.1
Helictopleurus corruscus	DQ369523.1	DQ369568.1		DQ369505.1			EF188141.1
Helictopleurus sinuaticornis	DQ369537.1	DQ369582.1		DQ369519.1			EF188189.1
Helictopleurus giganteus	DQ369526.1	DQ369571.1		DQ369508.1			EF188161.1
Helictopleurus steineri	EF187969.1	EF188017.1		EF656716.1			EF188193.1
Helictopleurus semivirens	EF187962.1	DQ369581.1		DQ369518.1			EF188185.1
Helictopleurus dorbignyi	EF187925.1	EF187986.1		EF188057.1			EF188146.1
Helictopleurus minutus	EF187946.1			EF188078.1			
Helictopleurus viridiflavus	EF187973.1	EF188022.1		EF188106.1			
Helictopleurus littoralis							EU429042.1
Helictopleurus neoamplicollis	DQ369529.1	DQ369574.1		DQ369511.1			EF188168.1
Onthophagus maki			KC294253.1				KC294242.1
Onthophagus hirtus							AY039346.1
Onthophagus alcyon							
Hyalonthophagus alcyon	AY131565.1	JN619184.1		AY131752.1		EU162439.1	AY131920.1
Onthophagus alcyonides	EU162533.1						
Hyalonthophagus alcyonides	JN804706.1		JN804777.1				JN804633.1
Onthophagus fuliginosus	EU162543.1					KU665398.1	
Onthophagus auritus	AY847529.1						
Onthophagus furcaticeps	AY131576.1			AY131761.1			

Table IV-1 Continued

Name	16S	18S	28S-D2	28S-D3	CAD	COI-1	COI-2
Cleptocaccobius convexifrons	AY131561.1	JN619210.1		AY131748.1			AY131917.1
Caccobius nigritulus	AY131559.1			AY131746.1		MH020304.1	AY131915.1
Caccobius schreberi	AY131560.1			AY131747.1			AY131916.1
Onthophagus depressus	EF187975.1	EF188032.1		EF188116.1			EF188207.1
Onthophagus hinnulus		EF188037.1		EF188122.1			EF188214.1
Onthophagus mije	AY131581.1			AY131765.1			
Onthophagus ochropygus	JN804726.1						
Onthophagus fimetarius	AY131575.1			AY131760.1			AY131925.1
Onthophagus vulpes	AY131591.1			AY131775.1			MF944216.1
Onthophagus pedator	MF944129.1						MF962927.1
Onthophagus rorarius	AY131586.1			AY131770.1			AY131929.1
Onthophagus orientalis	MF944127.1						MF944206.1
Onthophagus maniti	MF944125.1						MF944204.1
Euonthophagus carbonarius	AY131564.1	JN619193.1		AY131751.1			AY131919.1
Onthophagus vinctus		KX512514.1	KX512603.1	KX512673.1			
Onthophagus interstitialis	JN804698.1		JN804770.1		JN804575.1	EU162442.1	JN804623.1
Onthophagus binodis	EU162536.1						
Onthophagus punctatus							AY039348.1
Caccobius binodulus	AY131558.1	JN619183.1		AY131745.1		EU162459.1	AY131914.1
Onthophagus nigriventris	EU162555.1					EU162475.1	
Onthophagus sugillatus	EU162572.1						
Euonthophagus amyntas							AY039342.1
Onthophagus borneensis	MF944122.1						MF944202.1
Onthophagus furcatus							AY039357.1
Onthophagus variegatus		EF188040.1	JN804797.1	EF188125.1		KM441496.1	EF188217.1

Table IV-1 Continued

Name	16S	18S	28S-D2	28S-D3	CAD	COI-1	COI-2
Onthophagus coenobita							AY039350.1
Onthophagus opacicollis			KC294251.1			KM441390.1	KC294243.1
Onthophagus fracticornis			KC294254.1			LN832281.1	KC294245.1
Onthophagus similis	AY131590.1		KC294256.1	AY131774.1		LN554677.1	AY131933.1
Onthophagus lemur							AY039353.1
Onthophagus melitaeus							AY039349.1
Onthophagus latigena							AY039356.1
Onthophagus ruficapillus			KC294255.1			KM445886.1	KC294246.1
Onthophagus ovatus							AY039351.1
Onthophagus grossepunctatus						KM439793.1	AY039347.1
Onthophagus verticicornis			KC294258.1				KC294249.1
Onthophagus nutans							AY039344.1
Onthophagus stylocerus						LN554680.1	AY039352.1
Onthophagus andalusicus						LN832282.1	
Onthophagus vacca			KC294250.1			KM447997.1	KC294236.1
Onthophagus medius							
Onthophagus marginalis	KJ721838.1			KJ721900.1		LN554683.1	
Onthophagus nebulosus							
Onthophagus merdarius						KU916082.1	AY039355.1
Onthophagus nuchicornis	EU162556.1		KC294252.1				KC294244.1
Onthophagus glabratus	AY131577.1			AY131762.1			AY131926.1
Onthophagus pronus	AY847525.1					EU162451.1	
Onthophagus granulatus	EU162545.1						
Onthophagus phanaeides	MF944130.1						MF944208.1
Milichus apicalis	AY131566.1	JN619196.1	JN804787.1	AY131753.1			AY131921.1

Table IV-1 Continued

Name	16S	18S	28S-D2	28S-D3	CAD	COI-1	COI-2
Onthophagus rubicundulus	AY131587.1			AY131771.1		EU162452.1	AY131930.1
Onthophagus haagi	EU162546.1					EU162477.1	
Onthophagus vermiculatus	EU162575.1						
Onthophagus dunningi							MG588143.1
Onthophagus mulgravei	AY131582.1			AY131766.1			AY131927.1
Onthophagus turrbal	AY847531.1						
Onthophagus quadripustulatus	AY131585.1			AY131769.1			
Onthophagus consentaneus	AY131573.1			AY131758.1		EU162456.1	
Onthophagus laminatus	EU162550.1			AY131764.1		EU162458.1	
Onthophagus mjobergi	EU162553.1					EU162468.1	
Onthophagus sloanei	EU162565.1					EU162449.1	
Onthophagus ferox	EU162542.1					EU162463.1	
Onthophagus pentacanthus	EU162559.1					EU162448.1	
Onthophagus evanidus						KP978430.1	
Onthophagus ochromerus							
Onthophagus neostenocerus	AY847530.1					EU162441.1	
Onthophagus australis	EU162535.1						
Onthophagus mammillatus	AY847526.1						
Onthophagus pugnax	AY847527.1					EU162443.1	
Onthophagus capella	EU162537.1			AY131756.1		KM450900.1	
Onthophagus illyricus							
Onthophagus bivertex	KJ721867.1					EU162476.1	
Onthophagus taurus	EU162573.1		DQ430885.1	DQ430937.1	XM23060308.1		KC294248.1
Onthophagus taurinus	MF944135.1						MF944213.1
Onthophagus solivagus	KJ721822.1			KJ721871.1			JX064146.1

Table IV-1 Continued

Name	16S	18S	28S-D2	28S-D3	CAD	COI-1	COI-2
Onthophagus sinicus	KJ721837.1			KJ721899.1			
Onthophagus viduus	KJ721823.1			KJ721872.1			JX064148.1
Onthophagus obscurior	AY131584.1			AY131768.1			AY131928.1
Onthophagus babirussa	MF944116.1						MF944197.1
Onthophagus babirussoides	AY131568.1			AY131754.1		EU162440.1	AY131922.1
Onthophagus asperulus	EU162534.1						
Onthophagus fodiens	KJ721824.1			KJ721873.1			JX064151.1
Onthophagus lenzi	JX269021.1			KJ721874.1		EU162444.1	JX064153.1
Onthophagus clypeatus	EU162538.1			EF656709.1		EU162478.1	EF656758.1
Onthophagus xanthomerus	EU162576.1					EU162464.1	
Onthophagus praecellens	EU162560.1					EU162437.1	
Onthophagus acuminatus	EU162531.1					EU162474.1	
Onthophagus stockwelli							
Onthophagus batesi	EF656647.1			EF656689.1		EU162455.1	EF656738.1
Onthophagus incensus	EU162549.1					EU162467.1	
Onthophagus sharpi	EU162564.1						
Onthophagus championi	EF656651.1			EF656693.1			EF656742.1
Onthophagus bidentatus	AY131569.1			AY131755.1		EU162457.1	AY131923.1
Onthophagus marginicollis	EU162552.1					EU162453.1	
Onthophagus haematopus	EU162547.1			EF656711.1		MG059101.1	EF656760.1
Onthophagus orpheus	EU162557.1					EU162454.1	
Onthophagus hecate	EU162548.1		DQ430887.1	DQ430939.1		EU162445.1	
Onthophagus coscineus	EU162539.1					EU162447.1	
Onthophagus crinitus	EU162541.1	AY821535.1		AY131759.1		EU162462.1	AY131924.1
Onthophagus pennsylvanicus	EU162558.1						

Table IV-2 Generic Monophyly Dung Beetles.

Monophyly for all genera represented in the phylogeny, assessed for the original data set and each of the trees made from the bootstrapped data respectively. The last three columns sum how often a genus was monophyletic ("Yes") or non-monophyletic ("No") across all trees. Monot.=Monotypic.

Genus	Full	BS0	BS1	BS2	BS3	BS4	BS5	BS6	BS7	BS8	BS9	Yes	No
Byrrhidium	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	11	0
Catharsius	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	11	0
Dicranocara	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	11	0
Eucranium	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	11	0
Eurysternus	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	11	0
Gymnopleurus	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	11	0
Heliocopris	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	11	0
Kheper	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	11	0
Macroderes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	11	0
Metacatharsius	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	11	0
Pachysoma	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	11	0
Paragymnopleurus	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	11	0
Sarophorus	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	11	0
Xinidium	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	11	0
Ateuchus	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	10	1
Cheironitis	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	10	1
Coptodactyla	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	10	1
Epactoides	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	10	1
Epirinus	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	10	1
Sauvagesinella	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	10	1
Anachalcos	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	9	2
Demarziella	Yes	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	9	2
Lepanus	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	9	2
Saphobius	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes	9	2
Cephalodesmius	Yes	No	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	8	3
Pseudoarachnodes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No	Yes	8	3
Scybalophagus	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	8	3
Temnoplectron	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	8	3
Uroxys	No	Yes	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	8	3
Deltochilum	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	7	4

Table IV-2 Continued

Genus	Full	BS0	BS1	BS2	BS3	BS4	BS5	BS6	BS7	BS8	BS9	Yes	No
Canthonosoma	Yes	Yes	Yes	Yes	No	No	No	No	No	No	Yes	5	6
Euoniticellus	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	5	6
Dichotomius	Yes	Yes	No	No	No	No	Yes	No	No	Yes	No	4	7
Bubas	Yes	No	No	Yes	No	No	Yes	No	No	No	No	3	8
Diorygopyx	No	Yes	Yes	No	No	No	No	No	No	Yes	No	3	8
Phalops	No	No	Yes	No	Yes	Yes	No	No	No	No	No	3	8
Canthidium	No	Yes	No	Yes	No	No	No	No	No	No	No	2	9
Drepanopodus	No	No	No	No	No	No	No	Yes	No	No	No	1	10
Glyphoderus	No	No	No	No	No	No	Yes	No	No	No	No	1	10
Aliuscanthoniola	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Allogymnopleurus	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Amphistomus	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Anonthobium	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Bdelyropsis	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Bohepilissus	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Boletoscapter	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Cambefortantus	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Canthodimorpha	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Cleptocaccobius	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Coproecus	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Coptorhina	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Cyptochirus	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Delopleurus	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Dendropaemon	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Diabroctis	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Digitonthophagus	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Dwesasilvasedis	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Endroedyolus	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Ennearabdus	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Eodrepanus	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Eudinopus	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Frankenbergerius	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Garreta	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Gyronotus	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Hammondantus	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0

Table IV-2 Continued

Genus	Full	BS0	BS1	BS2	BS3	BS4	BS5	BS6	BS7	BS8	BS9	Yes	No
Hansreia	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Heterosyphus	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Ignambia	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Janssensantus	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Leotrichillum	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Litocopris	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Malagoniella	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Megathopa	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Megathoposoma	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Milichus	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Monoplistes	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Monteithocanthon	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Namakwanus	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Neateuchus	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Neosisyphus	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Ochicanthon	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Odontoloma	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Onthobium	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Outenikwanus	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Pachylomera	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Paracopris	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Paraphytus	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Parascatonomus	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Paronthobium	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Peckolus	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Pseudignambia	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Pseudonthobium	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Pycnopanelus	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Scybalocanthon	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Serrophorus	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Silvaphilus	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Sulcophanaeus	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Tanzanolus	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Tragiscus	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0
Trichillidium	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	Monot.	0	0

Table IV-2 Continued

Genus	Full	BS0	BS1	BS2	BS3	BS4	BS5	BS6	BS7	BS8	BS9	Yes	No
Anomiopsoides	No	No	No	No	No	No	No	No	No	No	No	0	11
Apotolamprus	No	No	No	No	No	No	No	No	No	No	No	0	11
Arachnodes	No	No	No	No	No	No	No	No	No	No	No	0	11
Caccobius	No	No	No	No	No	No	No	No	No	No	No	0	11
Canthon	No	No	No	No	No	No	No	No	No	No	No	0	11
Copris	No	No	No	No	No	No	No	No	No	No	No	0	11
Coprophanaeus	No	No	No	No	No	No	No	No	No	No	No	0	11
Drepanocerus	No	No	No	No	No	No	No	No	No	No	No	0	11
Euonthophagus	No	No	No	No	No	No	No	No	No	No	No	0	11
Helictopleurus	No	No	No	No	No	No	No	No	No	No	No	0	11
Hyalonthophagus	No	No	No	No	No	No	No	No	No	No	No	0	11
Liatongus	No	No	No	No	No	No	No	No	No	No	No	0	11
Nanos	No	No	No	No	No	No	No	No	No	No	No	0	11
Oniticellus	No	No	No	No	No	No	No	No	No	No	No	0	11
Onitis	No	No	No	No	No	No	No	No	No	No	No	0	11
Onthophagus	No	No	No	No	No	No	No	No	No	No	No	0	11
Oxysternon	No	No	No	No	No	No	No	No	No	No	No	0	11
Phanaeus	No	No	No	No	No	No	No	No	No	No	No	0	11
Proagoderus	No	No	No	No	No	No	No	No	No	No	No	0	11
Scarabaeus	No	No	No	No	No	No	No	No	No	No	No	0	11
Sisyphus	No	No	No	No	No	No	No	No	No	No	No	0	11
Tiniocellus	No	No	No	No	No	No	No	No	No	No	No	0	11

Table IV-3 Tribal Monophyly Dung Beetles.

Monophyly for all tribes represented in the phylogeny, assessed for the original data set and each of the trees made from the bootstrapped data respectively.

Tribe	Full	BS0	BS1	BS2	BS3	BS4	BS5	BS6	BS7	BS8	BS9
Phanaeini	No	No	No	No	No	No	No	No	No	No	No
Dichotomiini	No	No	No	No	No	No	No	No	No	No	No
Eucraniini	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Deltochilini	No	No	No	No	No	No	No	No	No	No	No
Canthonini	No	No	No	No	No	No	No	No	No	No	No
Coprini	No	No	No	No	No	No	No	No	No	No	No
Ateuchini	Yes	Yes	No	Yes	Yes	Yes	No	No	Yes	No	No
Scarabaeini	No	No	No	No	No	No	No	No	No	No	No
Eurysternini	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gymnopleurini	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Onthophagini	No	No	No	No	No	No	No	No	No	No	No
Oniticellini	No	No	No	No	No	No	No	No	No	No	No
Onitini	No	Yes	No	No	No	No	Yes	No	No	No	Yes
Sisyphini	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table IV-4 Tribal Monophyly-Issues Dung Beetles

Monophyly status and reasons for non-monophyly for all tribes in the full matrix trees. Mon.=monophyly-status, #Tips=number of taxa assigned to this tribe, ΔTips=number of additional taxa in this clade (descendants of same MRCA), #Intr=number of tribes among intruder taxa, Intruders=Genera or Tribes intruding (numbers in parentheses are numbers of tips, all taxa *incertae sedis* unless indicated otherwise), #Outl.=number of taxon-members outside the main clade.

Tribe	**Mon.**	**#Tips**	**ΔTips**	**#Intr.**	**Intruders**	**#Outl.**	**Outliers**
Eucraniini	Yes	8	0	0		NA	
Ateuchini	Yes	2	0	0		NA	
Eurysternini	Yes	6	0	0		NA	
Gymnopleurini	Yes	8	0	0		NA	
Sisyphini	Yes	6	0	0		NA	
Scarabaeini	No	37	1	1	*Neateuchus* (1)	0	
Phanaeini	No	44	2	1	*Diabroctis* (1), *Dendropaemon* (1)	0	
Onitini	No	7	2	1	*Cheironitis* (2)	0	
Oniticellini	No	36	10	2	*Tiniocellus* (2), *Tragiscus* (1), *Drepanocerus* (4), *Cyptochirus* (1), *Heterosyphus* (1), *Proagoderus* (1, Onthophagini)	0	
Onthophagini	No	119	53	1	*Hyalonthophagus* (2), *Cleptocaccobius* (1), *Euonthophagus* (2), *Milichus* (1)	16	*Onthophagus diabolicus* and 15 more.
Deltochilini	No	132	414	1	Canthonini (2)	112	*Arachnodes emmae* and 111 more.
Canthonini	No	52	494	1	Deltochilini (7)	25	*Megathopa villosa* and 24 more.
Coprini	No	23	501	1	*Paracopris* (1)	10	*Coptodactyla storeyi* and 9 more.
Dichotomiini	No	36	510	0		24	*Canthidium haroldi* and 23 more.

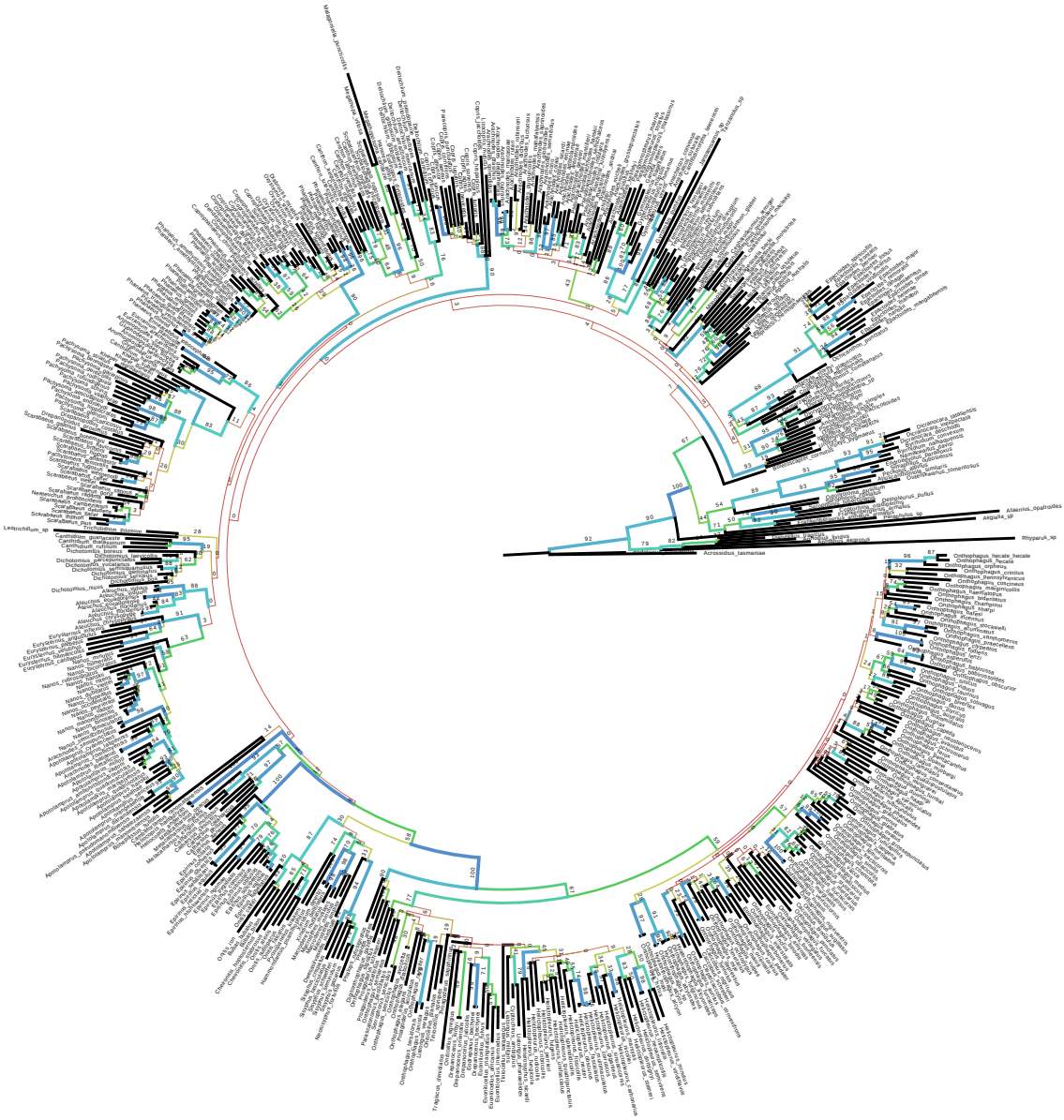

Figure IV-1: Branch Support Dung Beetle Phylogeny.

Bootstrap branch length support on the un-dated RAxML phylogeny. Colder colors indicating higher branch support, decreasing branch width with decreasing support (terminal branches are black).

Table IV-5 Node Calibrations and Estimated Ages.

For each node age calibration, the corresponding clade, fossil, location on the branch (crown or stem) are given. For each calibration scheme (‘young’ vs. ‘old’), the age constraints used and the estimated node ages on crown and stem, are indicated for each calibrated node, followed by the difference between the estimates.

Clade	Fossil	Stem/ crown	Young				Old				Age Diff.	
			Min cons.	Max cons.	Crown age	Stem age	Min cons.	Max cons.	Crown age	Stem age	Min	Max
Copris	*Copris kartlinus*	stem	3.6	98.1	**85.80**	**85.89**	3.6	131.6	59.80	**59.87**	26.00	26.02
Gymnopleurus	*Gymnopleurus sisyphus*	stem	13.5	98.1	**35.53**	**52.97**	13.5	131.6	**50.45**	**74.30**	14.91	21.33
Onthophagus	*Onthophagus bisontinus*	stem	13.5	98.1	**57.74**	**70.81**	13.5	131.6	**85.03**	**100.58**	27.29	29.77
Heliocopris	*Heliocopris antiquus*	stem	18.7	98.1	**1.46**	**78.79**	18.7	131.6	**2.09**	**110.90**	0.63	32.11
Scarabaeinae	*Lobateuchus parisii*	crown	53	98.1	**98.10**		53	131.6	**131.60**		33.50	
Scarabaeidae	*Juraclopus rohdendorfi*	crown	152	181.8			152	199.64				

Table IV-6 Manual Dispersal Rate Multiplier Matrices.

Manual dispersal rates for dispersal between all areas, for each defined epoch (duration see header row). The colors represent large sea barriers (dark blue), short sea barriers (light blue), land barriers (yellow), direct adjacency (green), and maximal barrier strength (light gray). Af=Africa, As=Asia, Eu=Europe, Na=North America, Oc= Oceania, Sa=South America, Ma=Madagascar.

20-0 ma	Af	As	Eu	Na	Oc	Sa	Ma
Af	1	0.85	0.85	0.25	0.25	0.25	0.75
As	0.85	1	1	0.85	0.85	0.25	0.25
Eu	0.85	1	1	0.75	0.05	0.25	0.25
Na	0.25	0.85	0.75	1	0.05	0.85	0.25
Oc	0.25	0.85	0.05	0.05	1	0.25	0.25
Sa	0.25	0.25	0.25	0.85	0.25	1	0.25
Ma	0.75	0.25	0.25	0.25	0.25	0.25	1

50-20 ma	Af	As	Eu	Na	Oc	Sa	Ma
Af	1	0.75	0.85	0.25	0.25	0.25	0.75
As	0.75	1	1	0.85	0.25	0.25	0.25
Eu	0.85	1	1	0.75	0.05	0.25	0.25
Na	0.25	0.85	0.75	1	0.05	0.75	0.25
Oc	0.25	0.25	0.05	0.05	1	0.25	0.25
Sa	0.25	0.25	0.25	0.75	0.25	1	0.25
Ma	0.75	0.25	0.25	0.25	0.25	0.25	1

110-50 ma	Af	As	Eu	Na	Oc	Sa	Ma
Af	1	0.35	0.85	0.35	0.25	0.75	0.75
As	0.35	1	1	1	0.25	0.25	0.25
Eu	0.85	1	1	1	0.25	0.25	0.25
Na	0.35	1	1	1	0.05	0.75	0.25
Oc	0.25	0.25	0.25	0.05	1	0.35	0.25
Sa	0.75	0.25	0.25	0.75	0.35	1	0.25
Ma	0.75	0.25	0.25	0.25	0.25	0.25	1

150-110 ma	Af	As	Eu	Na	Oc	Sa	Ma
Af	1	0.35	0.85	0.75	0.5	1	0.75
As	0.35	1	1	1	0.25	0.35	0.25
Eu	0.85	1	1	1	0.25	0.35	0.25
Na	0.75	1	1	1	0.05	0.85	0.05
Oc	0.5	0.25	0.25	0.05	1	0.5	1
Sa	1	0.35	0.35	0.85	0.5	1	0.5
Ma	0.75	0.25	0.25	0.05	1	0.5	1

200-150 ma	Af	As	Eu	Na	Oc	Sa	Ma
Af	1	0.5	1	1	0.5	1	1
As	0.5	1	1	1	0.25	0.5	0.25
Eu	1	1	1	1	0.25	0.5	0.5
Na	1	1	1	1	0.25	1	0.5
Oc	0.5	0.25	0.25	0.25	1	0.5	1
Sa	1	0.5	0.5	1	0.5	1	0.5
Ma	1	0.25	0.5	0.5	1	0.5	1

Table IV-7 Averaged Manual Dispersal Rate Multiplier Matrices.

Manual dispersal rates for dispersal between all areas, averaged over all epochs (for 4 or 3 epochs, see header row). Af=Africa, As=Asia, Eu=Europe, Na=North America, Oc= Oceania, Sa=South America, Ma=Madagascar.

4-Epoch averaged							
	Af	As	Eu	Na	Oc	Sa	Ma
Af	1.00	0.66	0.85	0.33	0.28	0.45	0.75
As	0.66	1.00	1.00	0.90	0.50	0.26	0.25
Eu	0.85	1.00	1.00	0.83	0.12	0.26	0.25
Na	0.33	0.90	0.83	1.00	0.05	0.80	0.23
Oc	0.28	0.50	0.12	0.05	1.00	0.30	0.33
Sa	0.45	0.26	0.26	0.80	0.30	1.00	0.28
Ma	0.75	0.25	0.25	0.23	0.33	0.28	1.00

3-Epoch averaged							
	Af	As	Eu	Na	Oc	Sa	Ma
Af	1.00	0.69	0.85	0.28	0.25	0.38	0.75
As	0.69	1.00	1.00	0.89	0.53	0.25	0.25
Eu	0.85	1.00	1.00	0.81	0.10	0.25	0.25
Na	0.28	0.89	0.81	1.00	0.05	0.80	0.25
Oc	0.25	0.53	0.10	0.05	1.00	0.28	0.25
Sa	0.38	0.25	0.25	0.80	0.28	1.00	0.25
Ma	0.75	0.25	0.25	0.25	0.25	0.25	1.00

Table IV-8 BioGeoBEARS Root Range Probabilities.

Probabilities of the ten ancestral ranges with the on average highest relative probability at the root node, for each tree and bootstrap tree. The most probable area for each tree is highlighted in green. Af=Africa, Oc=Oceania, Sa=South America, Ma=Madagascar, As=Asia, Na=North America.

Tree	AfOcSa	AfSaMa	AfOc	AfSa	AfOcMa	Af	AfAsOc	AfNaSa	AfNaOc	AfMa
Young	0.9184	0.0001	0.0129	0.0039	0.0006	0.0001	0.0002	0.0000	0.0634	0.0000
Old	0.9170	0.0001	0.0122	0.0030	0.0008	0.0001	0.0002	0.0000	0.0663	0.0000
Young BS0	0.2215	0.0000	0.6561	0.0007	0.1016	0.0066	0.0024	0.0000	0.0099	0.0003
Young BS1	0.0577	0.6080	0.0008	0.2930	0.0013	0.0138	0.0000	0.0077	0.0027	0.0073
Young BS2	0.8633	0.0002	0.0913	0.0146	0.0081	0.0127	0.0003	0.0018	0.0019	0.0002
Young BS3	0.8369	0.0010	0.0347	0.0888	0.0002	0.0301	0.0003	0.0021	0.0002	0.0003
Young BS4	0.9573	0.0010	0.0195	0.0121	0.0003	0.0024	0.0001	0.0002	0.0065	0.0000
Young BS5	0.0300	0.5061	0.0128	0.0189	0.2487	0.0312	0.0210	0.0001	0.0003	0.0629
Young BS6	0.9183	0.0001	0.0132	0.0015	0.0021	0.0001	0.0041	0.0000	0.0602	0.0000
Young BS7	0.3352	0.3389	0.0082	0.1850	0.0192	0.0256	0.0001	0.0269	0.0267	0.0065
Young BS8	0.0020	0.0000	0.2655	0.0003	0.3904	0.0534	0.2495	0.0000	0.0067	0.0054
Young BS9	0.2381	0.4228	0.0030	0.1638	0.0000	0.0111	0.0000	0.1298	0.0002	0.0043
Old BS0	0.1342	0.0000	0.7518	0.0017	0.0610	0.0329	0.0037	0.0000	0.0112	0.0010
Old BS1	0.0979	0.5331	0.0040	0.2713	0.0033	0.0438	0.0000	0.0086	0.0044	0.0198
Old BS2	0.6501	0.0009	0.1738	0.0430	0.0132	0.0873	0.0012	0.0055	0.0049	0.0013
Old BS3	0.7257	0.0020	0.0467	0.1317	0.0003	0.0772	0.0004	0.0040	0.0003	0.0009
Old BS4	0.8966	0.0013	0.0509	0.0226	0.0008	0.0127	0.0002	0.0003	0.0132	0.0002
Old BS5	0.0151	0.4344	0.0156	0.0172	0.2953	0.0394	0.0218	0.0000	0.0002	0.0829
Old BS6	0.9002	0.0002	0.0279	0.0027	0.0049	0.0003	0.0085	0.0000	0.0540	0.0001
Old BS7	0.3453	0.2859	0.0148	0.1946	0.0114	0.0532	0.0003	0.0218	0.0241	0.0108
Old BS8	0.0020	0.0001	0.2837	0.0010	0.3020	0.1604	0.1803	0.0000	0.0048	0.0160
Old BS9	0.2318	0.3596	0.0041	0.1849	0.0000	0.0194	0.0000	0.1634	0.0001	0.0054

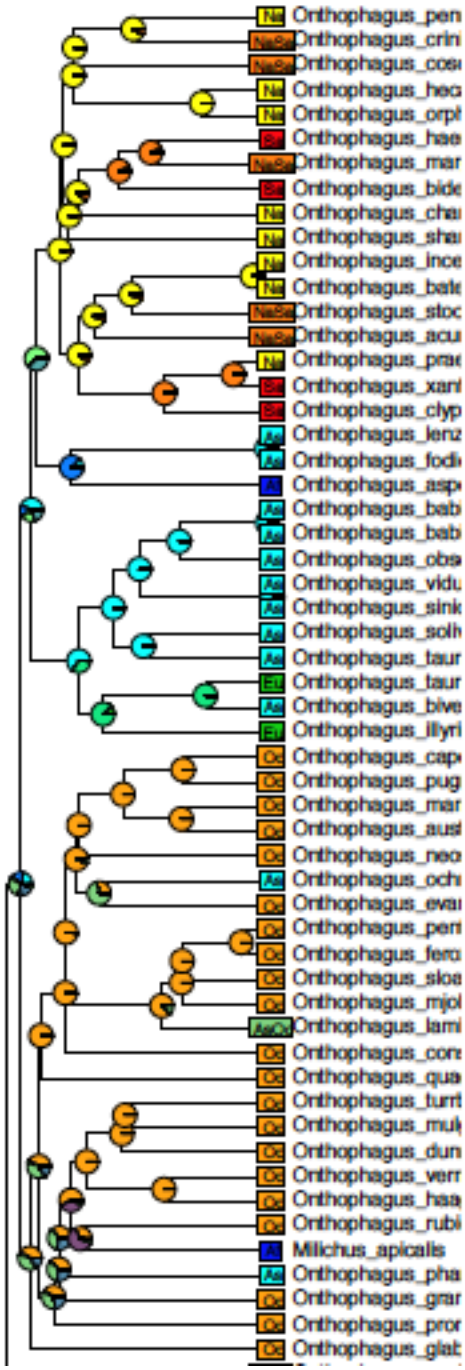

Figure IV-2: BioGeoBEARS Ancestral Range Estimation Young Tree.

Relative probabilities of estimated ancestral ranges at each node indicated by pie charts. Model: DEC+w. Af=Africa, Oc=Oceania, Sa=South America, Ma=Madagascar, As=Asia, Na=North America.

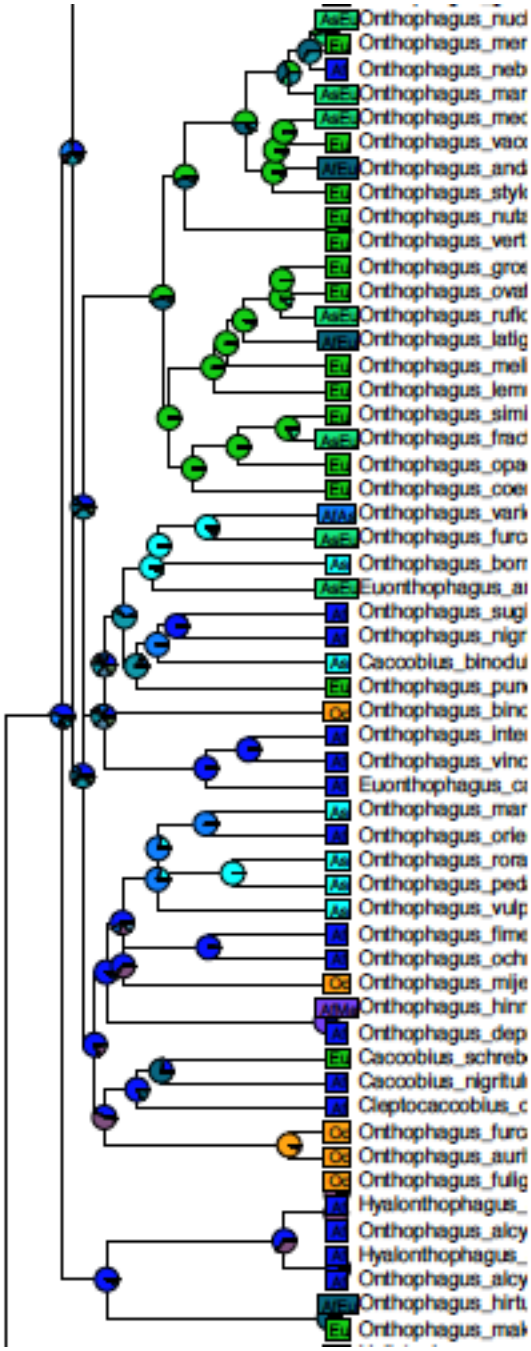

Figure IV-2 Continued

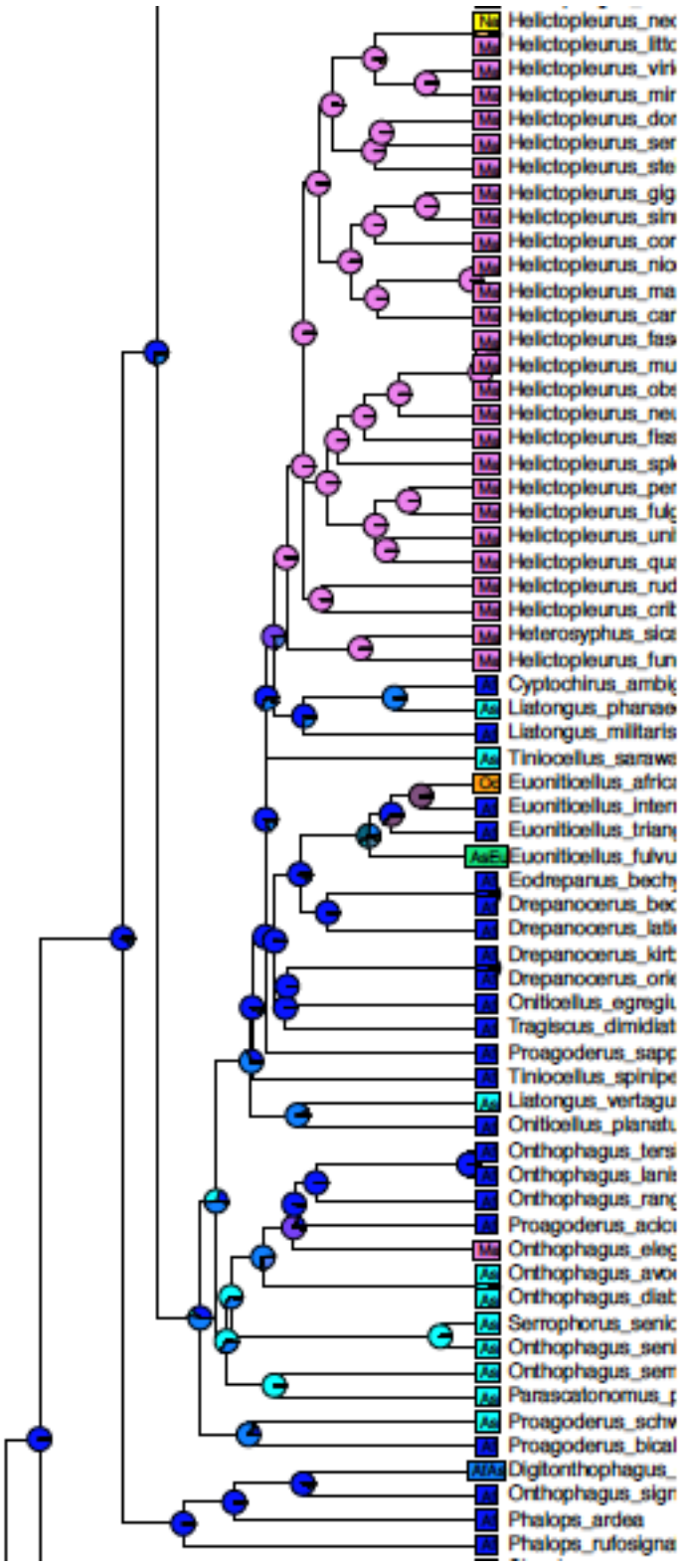

Figure IV-2 Continued

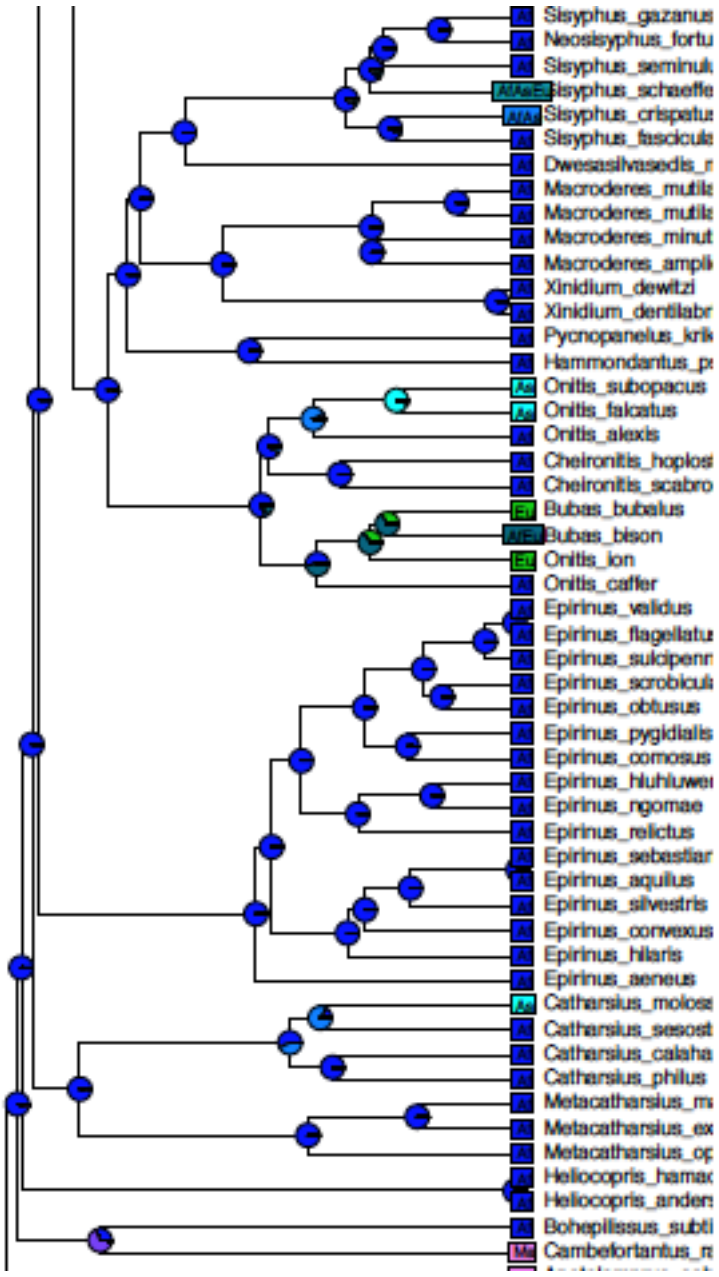

Figure IV-2 Continued

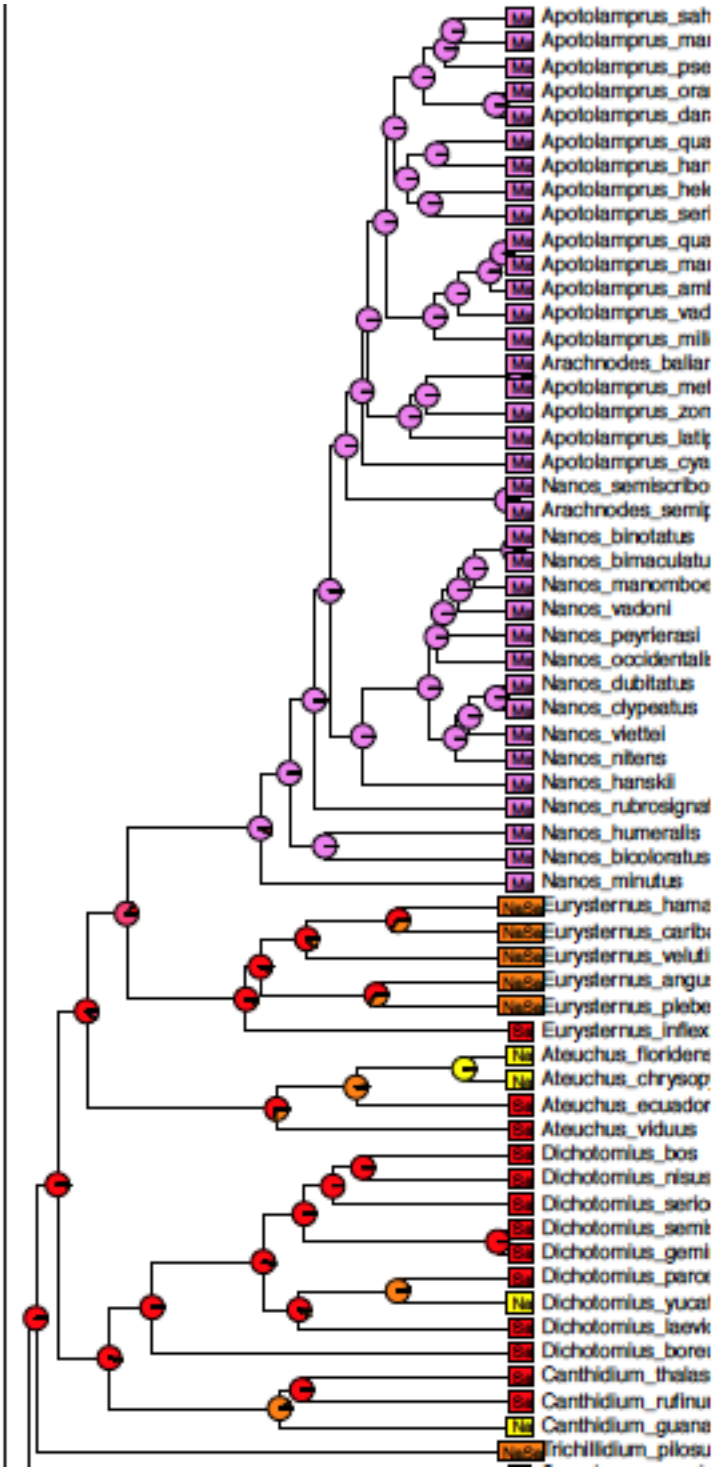

Figure IV-2 Continued

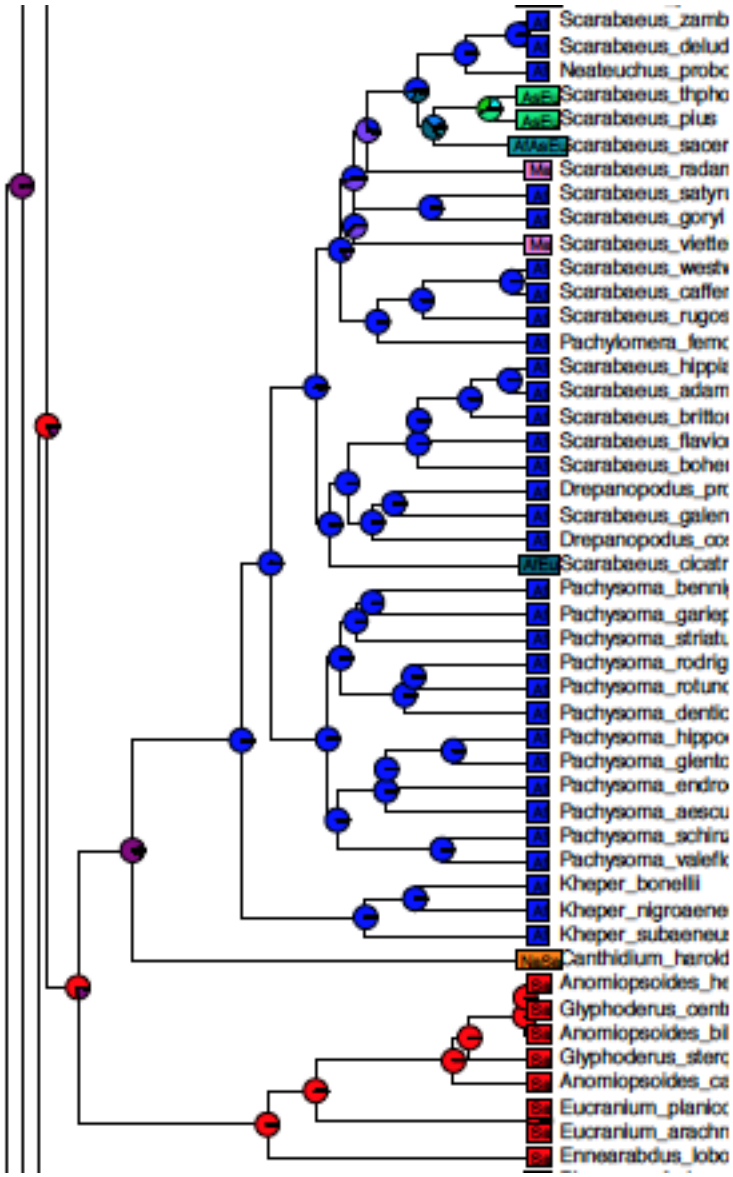

Figure IV-2 Continued

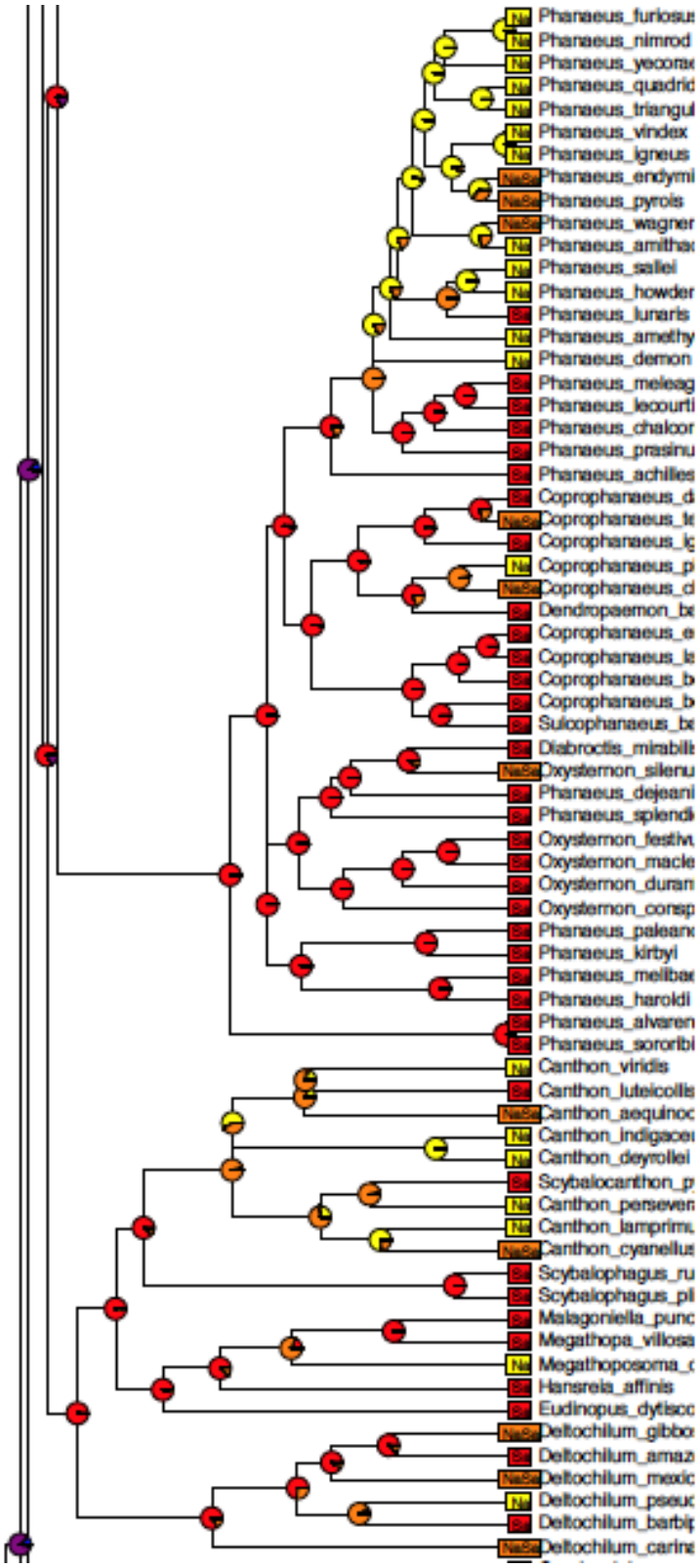

Figure IV-2 Continued

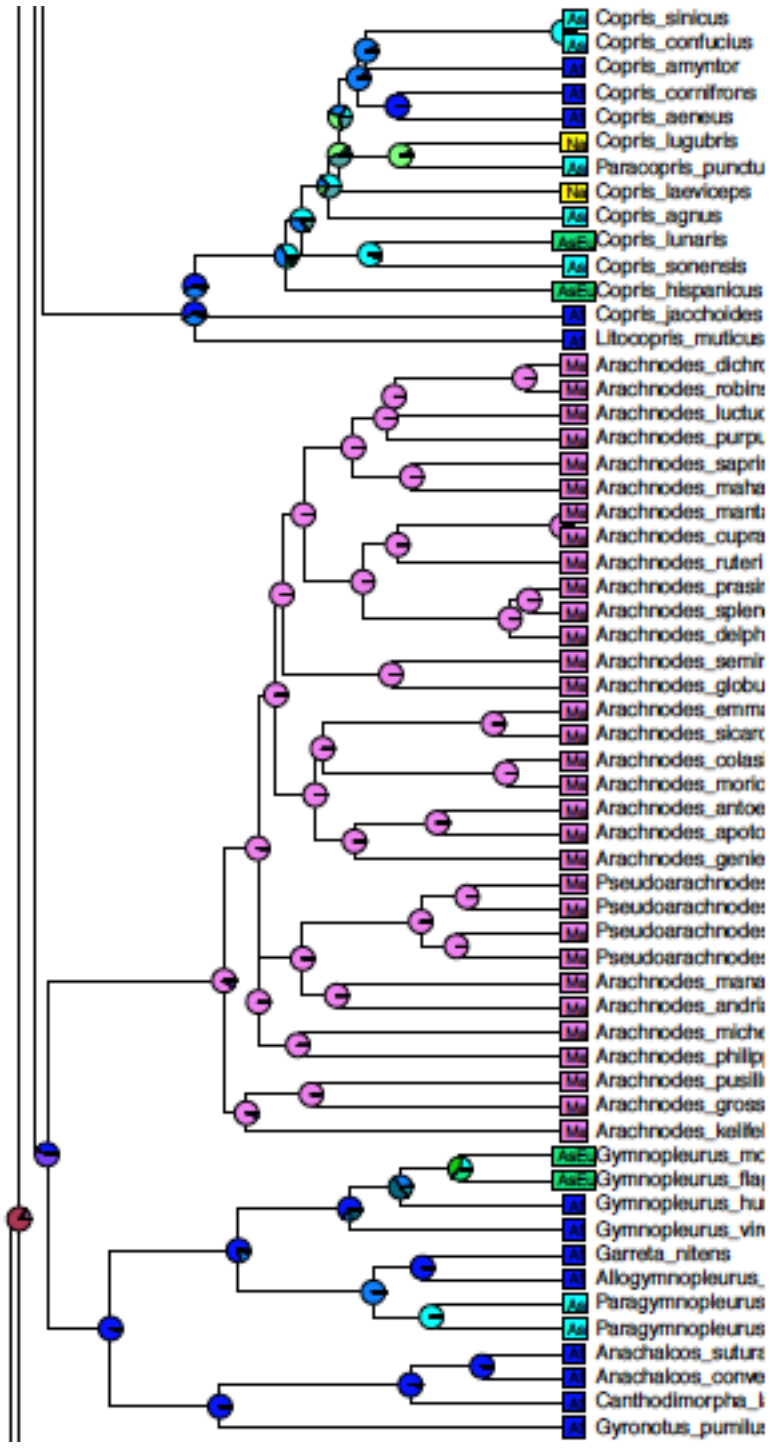

Figure IV-2 Continued

Figure IV-2 Continued

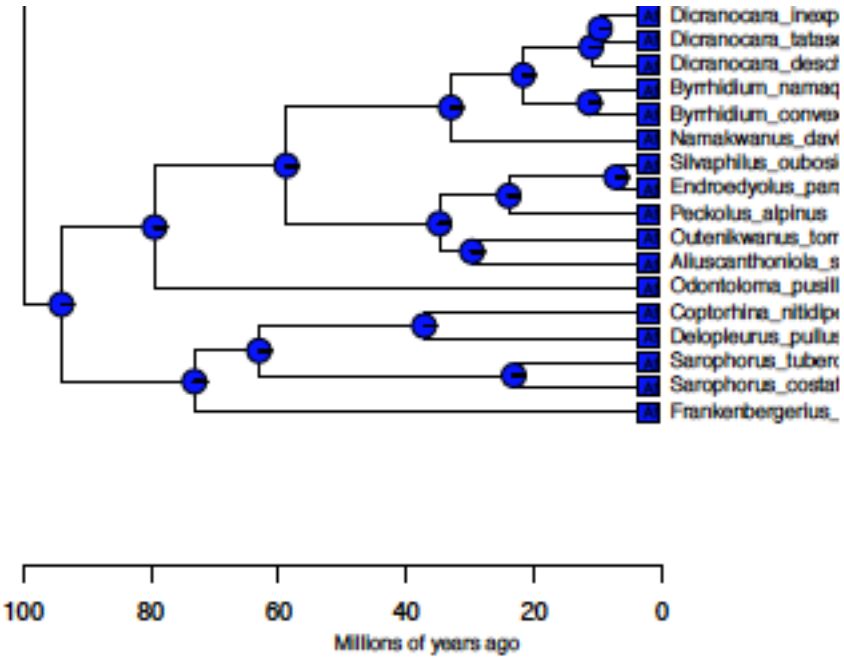

Figure IV-2 Continued

Table IV-9 Best Fitting Models for Biogeographcal Diversification Hypotheses.

Best fitting models of ML GeoSSE analyses for each data set and each tree, as determined by likelihood ratio tests and ΔAIC scores. Models where lambda was constrained to be equal in both areas, also constrained lambda to be zero for widespread taxa.

Tested Pairing	Young Tree	Old Tree
Large Mammals vs. rest	equal mu, equal dispersal	equal mu, equal dispersal
Small Mammals vs. rest	equal lambda, equal mu	equal lambda, equal mu
Gondwana vs. rest	equal mu	equal mu, equal dispersal
Laurasia vs . rest	full model	equal lambda, equal mu
Madagascar vs. rest	equal dispersal	equal mu, equal dispersal

Table IV-10 Posterior Distribution of GeoSSE Rate Estimates.

95% quantiles of posterior distributions for diversification and dispersal rates estimated under the best scoring model for each data set and each tree. s=speciation rate, x=extinction rate, r=net diversification rate (s-x), d= dispersal rate.

Pairings	Quant.	sA	sB	sAB	xA	xB	rA	rB	rAB	dA	dB	Likelihood
	Young Tree											
Only Large Mam. Vs. rest	2.50%	0.0968	0.0942	0.0382		0.0192	0.0636	0.0600	0.0382		0.0008	-2436.8114
	97.50%	0.1449	0.1437	3.9306		0.0785	0.0803	0.0799	3.9306		0.0021	-2428.7662
Only Small Mam. vs. rest	2.50%		0.1076			0.0342		0.0613		0.0002	0.0015	-2436.8855
	97.50%		0.1552			0.0922		0.0756		0.0014	0.0042	-2431.4368
Madagascar vs. rest	2.50%	0.1041	0.0930	0.1085	0.0006	0.0191	0.0770	0.0602	0.1085		0.0003	-2381.0849
	97.50%	0.2102	0.1423	8.0402	0.1276	0.0798	0.1158	0.0756	8.0402		0.0010	-2373.0650
Laurasia vs . rest	2.50%	0.0819	0.0863	0.0016	0.0006	0.0160	0.0581	0.0528	0.0016	0.0223	0.0003	-2619.1510
	97.50%	0.1308	0.1409	0.0651	0.0674	0.0857	0.0928	0.0724	0.0651	0.0577	0.0036	-2610.9780
Gondwana vs. rest	2.50%	0.0964	0.0683	0.0002		0.0008	0.0863	0.0593	0.0002	0.0051	0.0028	-2629.9069
	97.50%	0.1228	0.0924	0.0309		0.0305	0.1054	0.0721	0.0309	0.0156	0.0056	-2622.8756
Pairings	**Old Tree**											
Only Large Mam. vs. rest	2.50%	0.0636	0.0619	0.0286		0.0093	0.0448	0.0426	0.0286		0.0005	-2632.4797
	97.50%	0.0952	0.0942	1.5142		0.0486	0.0560	0.0557	1.5142		0.0013	-2624.6484
Only Small Mam. vs. rest	2.50%		0.0698			0.0183		0.0435		0.0001	0.0010	-2633.0250
	97.50%		0.0994			0.0546		0.0530		0.0009	0.0029	-2627.6812
Madagascar vs. rest	2.50%	0.0726	0.0575	0.0952		0.0043	0.0558	0.0437	0.0952		0.0002	-2578.7214
	97.50%	0.1066	0.0887	6.9666		0.0438	0.0753	0.0544	6.9666		0.0007	-2571.5387
Laurasia vs . rest	2.50%		0.0602			0.0103		0.0394		0.0137	0.0002	-2810.4804
	97.50%		0.0851			0.0443		0.0513		0.0327	0.0027	-2805.3483
Gondwana vs. rest	2.50%	0.0627	0.0477	0.0001		0.0003	0.0580	0.0437	0.0001		0.0028	-2829.6036
	97.50%	0.0780	0.0611	0.0174		0.0152	0.0702	0.0512	0.0174		0.0045	-2822.8190

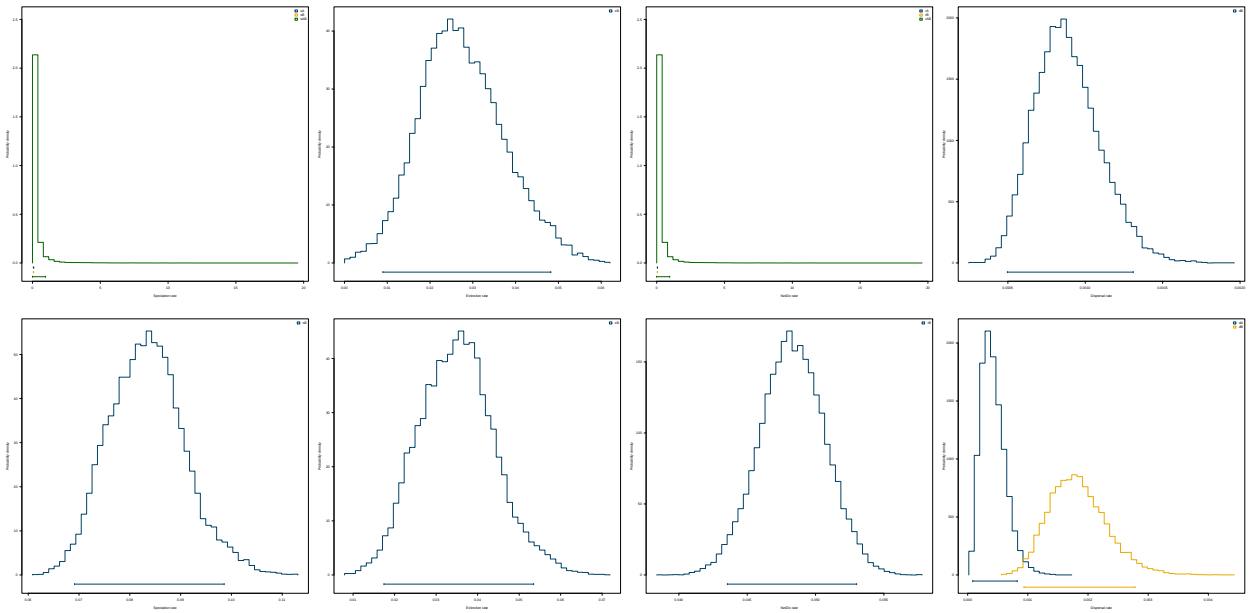

Figure IV-3: GeoSSE Results Dung Producers Old Tree.

Posterior distributions of area-dependent rates estimated by GeoSSE. Left to right: speciation rates, extinction rates, net diversification rates (speciation – extinction), dispersal rates. **Top**: rates in areas with large and medium sized mammals and high and intermediate droppings-diversity (Afro-Eurasia and Americas, blue), and rates in areas with small sized mammals and low droppings-diversity (East Gondwanan Fragments, yellow). **Bottom**: rates in areas with large sized mammals and high droppings-diversity (Afro-Eurasia, yellow), and rates of areas with low and medium sized mammals and low and intermediate droppings-diversity (Americas and East Gondwanan Fragments, blue). Rates of widespread taxa are colored green; where only one rate was estimated, it is colored blue. Dispersal rates of an area reflect dispersal *out* of said area.

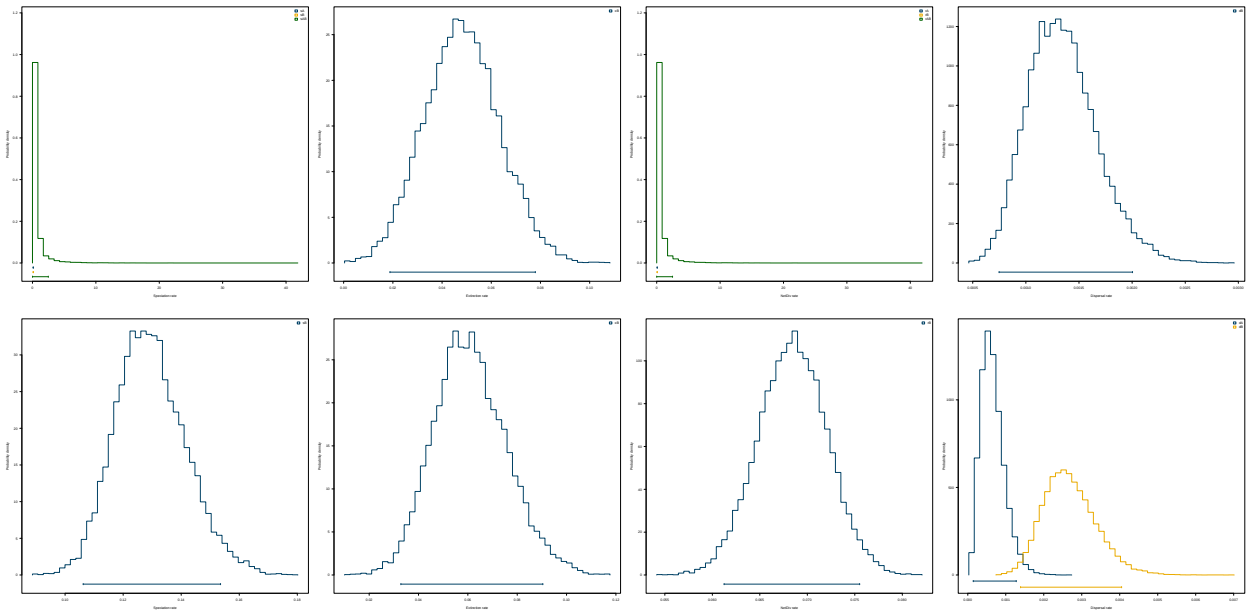

Figure IV-4: GeoSSE Result Dung Producers Young Tree.

Posterior distributions of area-dependent rates estimated by GeoSSE. Left to right: speciation rates, extinction rates, net diversification rates (speciation – extinction), dispersal rates. **Top**: rates in areas with large and medium sized mammals and high and intermediate droppings-diversity (Afro-Eurasia and Americas, blue), and rates in areas with small sized mammals and low droppings-diversity (East Gondwanan Fragments, yellow). **Bottom**: rates in areas with large sized mammals and high droppings-diversity (Afro-Eurasia, yellow), and rates of areas with low and medium sized mammals and low and intermediate droppings-diversity (Americas and East Gondwanan Fragments, blue). Rates of widespread taxa are colored green; where only one rate was estimated, it is colored blue. Dispersal rates of an area reflect dispersal *out* of said area.

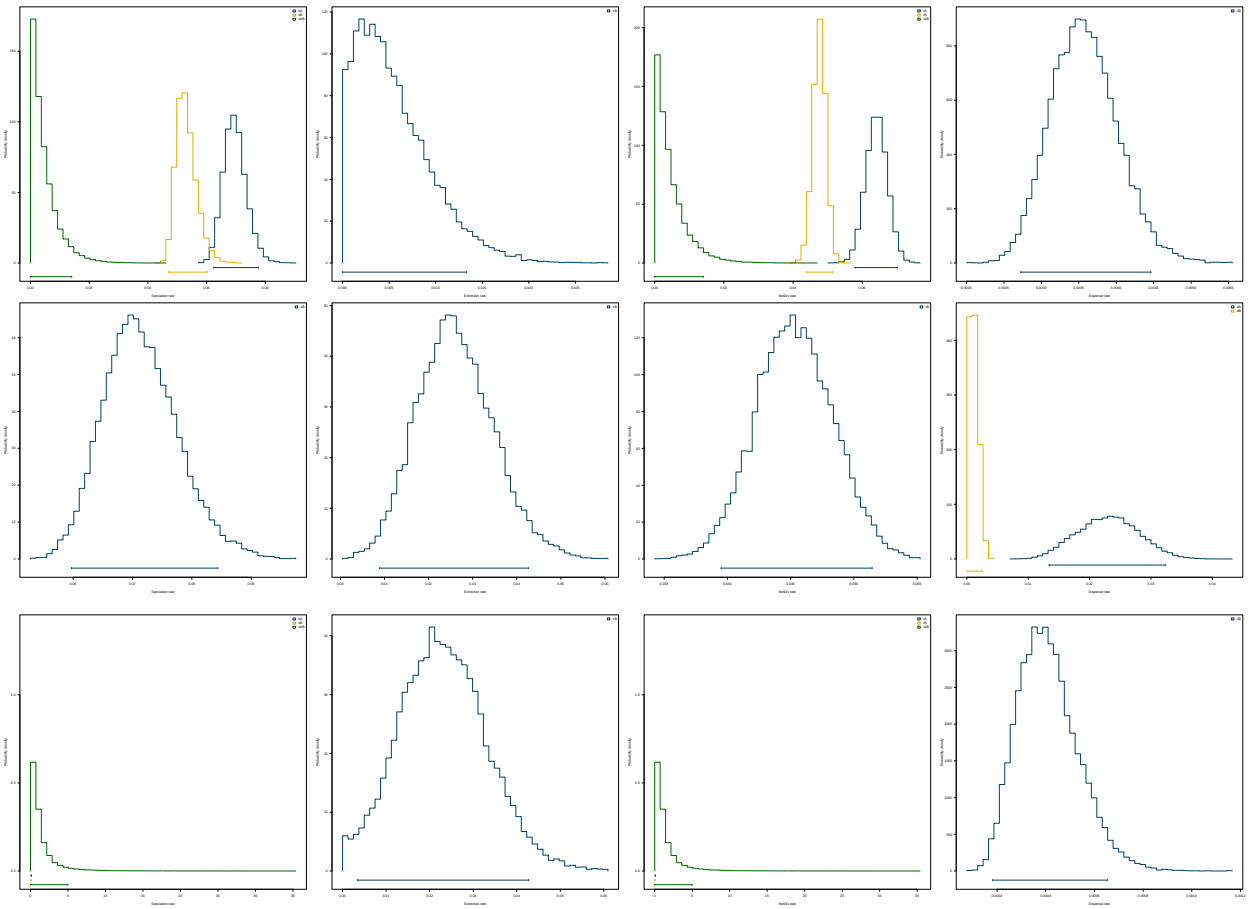

Figure IV-5: GeoSSE Results Out-Of-Gondwana Old Tree.

Posterior distributions of area-dependent rates estimated by GeoSSE. Left to right: speciation rates, extinction rates, net diversification rates (speciation – extinction), dispersal rates. **Top**: rates in Laurasia and Madagascar (blue), and rates in Gondwana (yellow). **Center**: rates in Laurasia (blue), and rates in Gondwana and Madagascar (yellow). **Bottom**: rates in Madagascar (blue), and rates in Gondwana and Laurasia. Rates of widespread taxa are colored green; where only one rate was estimated, it is colored blue. Dispersal rates of an area reflect dispersal *out* of said area.

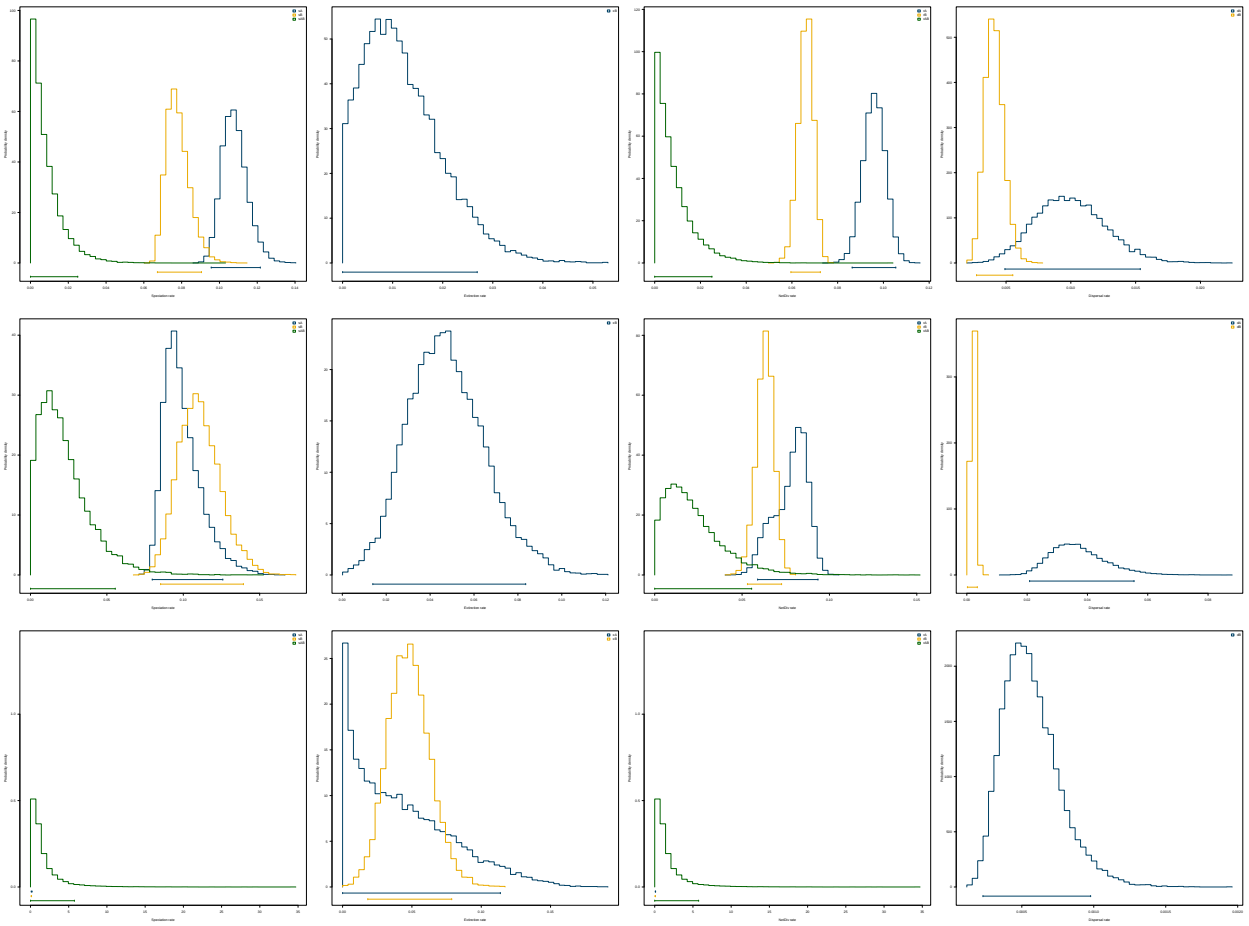

Figure IV-6: GeoSSE Results Out-Of-Gondwana Young Tree.

Posterior distributions of area-dependent rates estimated by GeoSSE. Left to right: speciation rates, extinction rates, net diversification rates (speciation – extinction), dispersal rates. **Top**: rates in Laurasia and Madagascar (blue), and rates in Gondwana (yellow). **Center**: rates in Laurasia (blue), and rates in Gondwana and Madagascar (yellow). **Bottom**: rates in Madagascar (blue), and rates in Gondwana and Laurasia. Rates of widespread taxa are colored green; where only one rate was estimated, it is colored blue. Dispersal rates of an area reflect dispersal *out* of said area.

CHAPTER V
UNVEILING THE DIVERSITY OF DUNG BEETLES – THE RISE OF THE GRASSLANDS

Abstract

Dung beetles show a high diversity and near global distribution and exhibit the remarkable life history trait of depending on other organisms' dung for feeding and reproduction. Given the rich diversity of dung beetle species living in different habitats and showing various degrees of specialization concerning the kind of dung they prefer, it has been hypothesized that the accumulation of that diversity might have been directly influenced by the diversity in their main resource – dung – and its producers. In particular, there has been debate over whether dung beetles – now mostly specialized on mammal droppings – evolved to feed on dinosaur dung, and whether this caused them to experience losses in diversity as a result of the non-avian dinosaurs extinction, and whether the vast diversity of extant dung beetles has, in part, evolved in response to the increased dung availability from large herds of ungulates after the rise of grasslands during the Miocene aridification. Using an updated phylogeny and recently developed comparative methods, I show that while an association with dinosaurs and subsequent co-extinction is plausible, a diversity increase in response to the rise of grassland-inhabiting herbivores during the Miocene cannot be supported given available data.

Introduction

The subfamily Scarabaeinae – the 'true dung beetles' show high diversity of approximately 5,300 species and have a near-global distribution (Hanski & Cambefort 1991). Their remarkable life history trait using other organisms' dung to feed and reproduce makes them one of the most ecologically (Nichols *et al.* 2008) and economically (Losey & Vaughan 2006) important insect groups. It has been hypothesized that the impressive diversity of dung beetles must, at least in part, have been influenced by the diversity of what provided them with the resource for their specialized lifestyle – the dung producers (Scholtz, Davis & Kryger 2009; Gunter *et al.* 2016).

It has been argued that the evolution of mammals, and particularly their increase in body size, could have led to radiations in dung beetles (Davis, Scholtz & Philips 2002). A particular scenario suggests that the rise of grasslands during the Miocene (Strömberg 2011) led to the evolution of larger-bodied herd-living mammalian herbivores (Wing & Sues 1992; Toljagić *et al.* 2017), and that the ensuing increased availability of highly suitable dung would have allowed for a rise in dung beetle diversity. The idea is appealing and intuitive – many dung beetles have specific preferences for dung of particular producers or particular properties (Scholtz, Davis & Kryger 2009), more available ecological niches provide opportunity for lineage diversification (Wellborn & Langerhans 2015), and the implications of grassland evolution for other organisms is rather well studied (Meadows & Linder 1993).

In addition, the possibility was discussed whether dung beetles were feeding on dinosaur dung at the time when those were the dominant dung producers. While the thought seems reasonable in principle, there are arguments for and against it, surrounding the actual feeding strategies of early dung beetles, the nature and suitability of dinosaur dung, and the question of whether dinosaurs were endo- or ectothermic (Davis, Scholtz & Philips 2002; Scholtz, Davis & Kryger 2009). But a

more fundamental question surrounds the age of dung beetles themselves, and whether they were contemporaneous with dinosaurs. The age of dung beetles has been the subject of disagreement, with estimates of their origin ranging from as old as Cretaceous (118.8-131.6ma: Gunter *et al.* (2016), 86.6-100.2ma: Ahrens, Schwarzer and Vogler (2014)), through Paleocene-Eocene (56ma: Mlambo, Sole and Scholtz (2015)) to as recent as Eocene-Oligocene (33.9ma: Wirta, Orsini and Hanski (2008)), complicated by reliability issues with some of the already scarce fossils available for calibration (Tarasov et al. 2016). Finally, if dung beetles were indeed depending on dinosaur dung, the question arises whether they also experienced the effects of the end-Cretaceous mass extinction, that would have robbed them of their major food providers (Scholtz, Davis & Kryger 2009; Gunter *et al.* 2016).

Previous studies attempting to address these questions have struggled with the dispute over the Cenozoic or Mesozoic origin of dung beetles, the lack of a representative molecular phylogeny of the whole subfamily, and (partially associated to that) the lack of appropriate methods to address the question. Thus, I will use a recently developed model-based tool for estimating diversification rates, and an updated phylogeny, to address whether dung beetle diversity was affected by the end-Cretaceous mass extinction, and by the rise of grasslands and large herds of mammalian dung producers during the Miocene aridification.

Materials and Methods

Phylogenetic Data and Diversity of Co-Diversifying Groups

For the dung beetles I used the phylogenies inferred in CHAPTER IV, dated with the young and old calibrations respectively.

We used the grass phylogeny by Spriggs, Christin and Edwards (2014), who built a dated phylogeny of 3,595 species of Poaceae (dated using macrofossil evidence), to evaluate whether the evolution of C4 grasses was associated with a rise in their diversification rates. Since the radiations of C4 grasses may have been an important element in the Miocene grassland expansion, I also used the C3-C4 assignment for the taxa in the Spriggs, Christin and Edwards (2014) phylogeny, and dropped all C3 taxa to obtain a C4-only tree comprising of 1341 species.

We used the mammal supertree by Bininda-Emonds *et al.* (2007), which contains 4,510 of the 4,554 extant species of mammals, and which was used by them to evaluate whether the end-Cretaceous mass extinction had a delayed, rather than an immediate effect on the rise of extant mammal groups. Because the dung of herbivores is presumed to be particularly desirable for dung beetles (Scholtz, Davis & Kryger 2009), I created a sub-tree of herbivores only. We used the dietary classification from Price *et al.* (2012) into herbivores, carnivores, and omnivores. Price *et al.* (2012) is based on a 1530 species mammal phylogeny, meaning that species-level dietary information was only available for part of the larger tree by Bininda-Emonds *et al.* (2007). We thus tallied the number of carnivores, omnivores and herbivores in the list for each available genus, family, and order, and calculated the ratio of herbivores at each level. We then considered a taxon herbivorous if more than 75% of its members were herbivorous, while

excluding the order Sirenia. We then assigned herbivory status to the taxa of the 4510 species tree based on the status of the lowest taxonomic level available (preferring genus over family over order). Finally, I dropped all non-herbivorous taxa to obtain a herbivore tree comprising of 1,264 species. To further narrow it down to the large-bodied herd-living mammal dung producers, I pruned the tree further to only include members of Perissodactyla and Artiodactyla (that is, Cetartiodactyla sans Cetacea), obtaining a tree of 229 species.

Both in case of C4-grasses and the herbivores, my approach of representing diversity through time of groups with those traits (C4 photosynthesis and herbivory), assumes that they actually evolved in a way that meant all the direct ancestors its extant carriers also had it (or at least had whatever property of it was crucial for the assumed effect on dung beetle diversification). This does not necessarily have to be the case.

Because all the groups considered above have a crown age that is much younger than that of dung beetles, and to address hypotheses regarding whether or not herds of herbivorous dinosaurs provided dung that dung beetles could rely on for nutrition, I also considered the diversity of non-avian dinosaurs. To this end, I used a phylogeny of 614 non-avian dinosaurs from Sakamoto, Benton and Venditti (2016), and dropped the largely carnivorous Theropods, leaving us with a 366 species tree of the mostly herbivorous clades Ornithischia and Sauropodomorpha.

Codiversification Analyses

To test whether diversification of dung beetles was influenced by the diversity of their dung producers, and particularly by the rise of grasslands and herds of large-bodied ruminants, I used a model of environmental dependent diversification, which is implemented as the function fit_env (Condamine, Rolland & Morlon 2013) in the R package RPANDA (Morlon *et al.* 2016). The function requires a continuous environmental variable over time, on which the estimated diversification rate depends. We thus prepared the lineage-through-time (LTT) information derived from the grass, mammal, and dinosaur phylogenies above to fit the required format.

Because the dung beetle phylogeny is only about 10% complete, and because this should affect the diversification estimates, I randomly dropped the tips of the other phylogenies down to the same sampling level (0.10085). Based on the total number of species (11367 *spp.* of Poaceae (Spriggs, Christin & Edwards 2014), 4554 *spp.* of mammals (Bininda-Emonds *et al.* 2007), and an estimated 1021 *spp.* of Ornithischia and Sauropodomorpha (Starrfelt & Liow 2016); and assuming that the sampling of C4-grasses and herbivorous mammals in their respective trees are proportional to the overall sampling), those trees were reduced to 1146 *spp.* of Poaceae, 428 *spp.* of C4-grasses, 459 *spp.* of mammals, 126 *spp.* of herbivorous mammals, 23 *spp.* of Artiodactyla and Perissodactyla, and 103 *spp.* of herbivorous dinosaurs.

We constructed the LTT data for each tree using the ltt function of phytools (Revell 2012), and transformed them to fit the format of the environmental curves, in the process extending them from their respective crown age to 200ma, by adding 1s, as the environmental data needs to cover at least the same time as the associated phylogeny (in the dinosaur case, I extended their LTT data with 0s from 66ma to present). We then combined the LTT-sets of each extant group with that of the dinosaurs by matching up their time intervals and then adding their LTT values.

The maximal number of lineages vastly differs among the different taxa, but my assumption would be the relative magnitude of the diversity of those groups (mediated by their actual abundance or ecological importance), rather than the absolute number of lineages. However, a case can also be made that the dung of past and extant producers (dinosaurs and mammals, respectively) was not of equal suitability for the dung beetles, and that the better suited dung of mammals fueled dung beetle diversification during the rise of grasslands and mammals. Thus, I created two versions of each LTT data set: one where the maximal number of lineages was standardized to 1 for both dinosaurs and mammals/grasses (subsequently called "equal-weighted"), and one where the dinosaurs were standardized to 1, whereas the mammals and grasses were standardized to 2, giving them higher relative influence on the beetles' diversification rates (subsequently called "extant-heavy").

Finally, I fitted the environmental diversification model using both the equal-weighted and extant-heavy version of the LTT data of dinosaurs plus Poaceae, C4-grasses, mammals, herbivores, and Artyo- and Perissodactyla respectively to both dung beetle phylogenies (young calibrations and old calibrations). Issues with the standard spline fitting function (sm.spline from the package pspline) used by both the function fitting the environmental diversification model and the plotting function, necessitated that I modified these functions to use smooth.spline from the package stats instead. The latter is more stable at the cost of potentially worse fitting, but visual inspection of the fitted spline convinced me that it is adequate. To test whether there is any environmental dependence to begin with, I also fitted a selection of constant rate and time dependent birth death models (constant birth death, linearly time-dependent speciation, linearly time-dependent speciation and extinction, exponentially time dependent speciation and constant extinction, and exponentially time dependent speciation and linear extinction) to both trees. The models were then compared using delta AICc.

Results

The diversification analyses drew two slightly different pictures for the different tree age calibrations. For the younger tree (Table III-1) diversification depends on the extant-heavy diversity of dinosaurs and herbivorous mammals (with speciation depending exponentially and extinction linearly). This model is closely followed by extant heavy C4-grass diversity (linear λ, linear μ), and extant heavy mammal diversity (linear λ, linear μ), with a larger ΔAICc separating it from the same model with exponential λ, and an even larger ΔAICc separating those from the first non-environmental model (with the rates depending on time, λ exponentially, μ linearly). After that, the differences in AICc increase gradually for the remaining models. The AICc-weights that the top three models (herbivores, C4-grasses, and mammals with linear dependence) bear by far most of the weight relatively evenly among them, with the next lower model an order of magnitude removed from them.

For the older tree, I get a similar picture (Table V-2), with three environment dependent models being separated by a small jump in ΔAICc from an exponentially and linearly time dependent model. However, those leading three models are all equal-weighted, rather than extant-heavy, and all have λ depending linearly on the dung-producers' diversity, and μ constant. And finally,

while they also feature herbivorous mammals (scoring second best) and mammals overall (third best), the Artiodactyla and Perissodactyla are fitting best for this tree. The AICc-weights assign the most importance to the top two models, with the third markedly lower, and the highest not LTT-dependent model again an order of magnitude behind the rest.

When looking at the parameters and visualizing the inferred rates through time for the young tree (Figure V-1), a common feature of all models is that they infer a decrease in net diversification around the time of the K-Pg extinction at around 66ma (while for the old tree (Figure V-2), the decrease happens before and diversification has already risen beyond its previous maximum by that time). But strikingly, they also all show a strong decrease in net diversification rate around 20ma (slightly earlier and slightly more gradual in case of the old tree). In the young tree, this dynamic is mostly achieved by a positive dependence of the diversity of the other groups with the dung beetles' speciation rate, and a negative relation of it with the extinction rate, which changes signs once the relative diversity exceeds 1 around 20ma (with the exception of the exponential-linear dependence on mammals, where also speciation has a negative relation to the mammal diversity). In the old tree, with the constant extinction rates, the dynamic is exclusively achieved by a negative dependence of dung beetle speciation rates with the diversity of the dung producers.

Discussion

The results show that tying dung beetle diversification rates to the diversity curves of dung producers or grasses fits the trees better than simple time dependent rate variation or constant rates. While this is generally encouraging, it is worth deconstructing those results a bit further. The visualizations of the fitted diversification curves (Figure V-1, Figure V-2) are indeed showing diversification slow-downs around the K-Pg boundary, but also after 20ma. The latter would particularly contradict the idea that the rise of grasslands and herds of herbivores was fueling dung beetle diversification in the Miocene, as it would suggest the exact opposite scenario.

It is worth recognizing that the decrease around 66ma is implied by all the LTT-based predictor variables, meaning that unless diversification rates are inferred to only have a weak dependence on the predictors (or e.g. a negative one), the predictor curves essentially enforce such a decrease to be present in the fitted rates. While better support for models with the KT-slowdown of rates over those without could be seen as support for the slowdown being real, it would be prudent to consider that other rate dynamics which neither model properly accounts for could be the cause for that support. On the other hand, the rate decrease at 20ma is not explicitly implied by the LTT curves of grasses or mammals (although C4-grasses and Artio- and Perissodactyla could imply strong rate changes around that time) yet seems to be inferred in all of the well-fitting models. Indeed, even the best fitting model that does not employ LTT-data (exponential-λ linear-μ time-dependent) recovers a rate decrease after 20ma. This all suggests that the signal in the data for that diversification slow-down after 20ma is substantial enough to overrule most other dynamics that are present, as is exemplified by the high support the model employing the extant-

heavy C4-grass curve in the young tree (Figure V-1), estimating a fairly consistently high diversification rate followed by a massive drop after ~10ma; or the fact that negative dependence on the diversity of dung producers is favored in the old tree (Figure V-2), despite the rapid rate increase it suggests to have taken place during the K-Pg mass extinction.

When considering the accumulation of (surviving) lineages through time (Figure V-3), the decreasing slope towards the present is apparent, and in the actual trees (Figure V-4) clearly have a lot of long terminal branches extending to 20ma and beyond. Decreasing diversification rates (Rabosky & Lovette 2008), and particularly a downturn in the LTT curve towards the present (Etienne & Rosindell 2012) are a well-known phenomenon, meaning this pattern could be real. However, the decrease in these two trees seems pronounced enough to suggest the possibility of additional biases behind it, potentially through missing sampling, issues in divergence time estimation, or a combination of the two. However, LTT plots of Scarabaeinae and other groups in Gunter *et al.* (2016) show a similar or even more pronounced stagnation between 20ma and the present, suggesting that this pattern might either be real or that whatever biases caused it are not more pronounced here than in comparable studies. While I suggested that the inferred quick increase of the diversification rate during the K-Pg mass-extinction might be spurious, and simply an artefact of the rate decrease after 20ma necessitating the negative relation of diversification rates on dung producer diversity, the old tree and its LTT plot (Figure V-3, Figure V-4) do indeed suggest a marked increase in diversity right post K-Pg, with a number of extant clades seemingly originating at that time. However, also the latter could potentially be an artefact of the imposed maximum age constraints during the divergence time estimation.

With these caveats in mind, a few careful conclusions can still be made. Firstly, dung beetle extinction at the K-Pg boundary seems plausible, based not only on the inferred diversification rate dynamics (mainly in the young tree), but also based on the LTT curves, suggesting a decreased lineage accumulation around that time under both tree dating schemes. Gunter *et al.* (2016), have attempted to answer the same question before, and came to the same conclusion. As they had to rely on a much more sparsely sampled Scarabaeinae phylogeny (embedded in a larger tree of Scarabaeidae), explicit tests for rate shifts have not been able to support such an extinction, thus they relied on its qualitative signature in the LTT curves as their supporting evidence. Arguably, such a signature is even clearer in the LTT curves inferred in this study.

Gunter *et al.* (2016) further go on to hypothesize that this extinction was due to the loss of dinosaurs as dung producers, and that the surviving ancestors of extant lineages were either generalists or have been pre-adapted to the use of mammal dung by feeding on that of early Cretaceous mammals. However, whether or not this extinction was in fact caused by the loss of non-avian dinosaurs as the major dung producers, rather than potential direct effects of the events that caused the mass extinction cannot be determined here. Either dating scheme places the origin of dung beetles well further back in time the K-Pg boundary, and while mammals were already present at that time, maybe even already experiencing their first radiation (Bininda-Emonds *et al.* 2007), and dinosaurs were already in decline (Sakamoto, Benton & Venditti 2016), this does lend some credibility to the idea that early dung beetles might have subsisted on dinosaur excrement. But while it is not implausible that early dung beetles would have utilized dinosaur dung, there is still debate about how likely it actually is, considering various lines of

evidence (Scholtz, Davis & Kryger 2009 and therein). Also, the trace fossils interpreted as dung beetle tunnels associated with a dinosaur coprolite (Chin & Gill 1996), and often brought up in support of dung beetles specialized on dinosaur dung, turn out to now be considered inconclusive and of questionable taxonomic identity (Scholtz, Davis & Kryger 2009; Tarasov *et al.* 2016).

Finally, as for the hypothesis that dung beetle diversification was fueled by the rise of grasslands in the Miocene and the subsequent prevalence of large herds of ungulates, it appears that this study cannot find any support for this. In case of the old tree, the results suggest the opposite. When a positive relation of diversification rates to dung producer diversity is suggested, the results are still reversed just around 20ma, the time the Miocene aridification was supposed to take place. Whether this is an accurate inference or obscured by dating or sampling artefacts, it does not provide support to the idea. The scenario of grassland and ungulate fueled diversification for dung beetles also does not match the time when dung beetle lineages started to increase post-KT (~60ma or ~40ma according to LTT plots for old and young tree respectively, or ~60ma and then more ~30ma according to the model fits for the young tree). The start of grassland as a dominant ecosystem, and subsequently the establishment of grassland-specific faunas, did not occur at the same time across the world (Strömberg 2011), and specifically the step from established grasslands to grassland adapted ruminants was slow (Toljagić *et al.* 2017), and previously assumed relations were not as straight forward as assumed (Strömberg 2002; Strömberg 2005; Jardine *et al.* 2012; Strömberg *et al.* 2013). Even if there was a causal sequence from grasslands over ungulates to dung beetles (and even if there were no potential issues in data availability and methods that would obscure it), one could expect the signal of this process in dung beetle diversification rates to be diffuse to some extent.

However, considering the results of the dung-producer-dependent model fitting and the changes in species accumulation as inferred by the LTT plots, an influence of the increase in diversity of mammals in general, or herbivores in particular on the diversification of dung beetles still seems plausible. However, if one was to try and establish such a connection with greater certainty, it might be necessary to include more of the relevant information. Connections between extant diversity of mammals and dung beetles in different parts of the world might support the idea of codiversification (Davis & Scholtz 2001; Gunter *et al.* 2016), but do not by themselves provide conclusive evidence that this is a historical pattern. However, they may hint at that including biogeographical patterns could help to get a clearer picture. Incorporating fossil data on dung producers would allow to gain a more adequate measure of their past diversity, which would remove further bias from any estimates of their effect on dung beetle diversification. In particular, it would allow to distinguish any effect of codiversification from that of the pull of the recent (the overestimation of diversity towards the present, due to only sampling surviving crown clades). Furthermore, in order to get a better sense of whether any of the relations found are coincidental or actually causal, more appropriate alternative models would be required. And finally, a more extensively and representatively sampled phylogeny might let us deal better with any sampling or dating related biases and might allow for less ambiguity in divergence time estimates.

References

Ahrens, D., Schwarzer, J. & Vogler, A.P. (2014) The evolution of scarab beetles tracks the sequential rise of angiosperms and mammals. *Proceedings of the Royal Society B-Biological Sciences,* **281**.

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

Chin, K. & Gill, B.D. (1996) Dinosaurs, dung beetles, and conifers: participants in a Cretaceous food web. *Palaios***,** 280-285.

Condamine, F.L., Rolland, J. & Morlon, H. (2013) Macroevolutionary perspectives to environmental change. *Ecology Letters,* **16,** 72-85.

Davis, A.L.V. & Scholtz, C.H. (2001) Historical vs. ecological factors influencing global patterns of scarabaeine dung beetle diversity. *Diversity and Distributions,* **7,** 161-174.

Davis, A.L.V., Scholtz, C.H. & Philips, T.K. (2002) Historical biogeography of scarabaeine dung beetles. *Journal of Biogeography,* **29,** 1217-1256.

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

Gunter, N.L., Weir, T.A., Slipinksi, A., Bocak, L. & Cameron, S.L. (2016) If Dung Beetles (Scarabaeidae: Scarabaeinae) Arose in Association with Dinosaurs, Did They Also Suffer a Mass Co-Extinction at the K-Pg Boundary? *Plos One,* **11**.

Hanski, I. & Cambefort, Y. (1991) Dung beetle ecology. *Dung beetle ecology***.,** i-xii, 1-481.

Jardine, P.E., Janis, C.M., Sahney, S. & Benton, M.J. (2012) Grit not grass: concordant patterns of early origin of hypsodonty in Great Plains ungulates and Glires. *Palaeogeography, Palaeoclimatology, Palaeoecology,* **365,** 1-10.

Losey, J.E. & Vaughan, M. (2006) The economic value of ecological services provided by insects. *Bioscience,* **56,** 311-323.

Meadows, M.E. & Linder, H.P. (1993) A paleoecological perspective on the origin of afromontane grasslands. *Journal of Biogeography,* **20,** 345-355.

Mlambo, S., Sole, C.L. & Scholtz, C.H. (2015) A molecular phylogeny of the African Scarabaeinae (Coleoptera: Scarabaeidae). *Arthropod Systematics & Phylogeny,* **73,** 303-321.

Morlon, H., Lewitus, E., Condamine, F.L., Manceau, M., Clavel, J. & Drury, J. (2016) RPANDA: an R package for macroevolutionary analyses on phylogenetic trees. *Methods in Ecology and Evolution,* **7,** 589-597.

Nichols, E., Spector, S., Louzada, J., Larsen, T., Amequita, S., Favila, M.E. & Scarabaeinae Res, N. (2008) Ecological functions and ecosystem services provided by Scarabaeinae dung beetles. *Biological Conservation,* **141,** 1461-1474.

Price, S.A., Hopkins, S.S., Smith, K.K. & Roth, V.L. (2012) Tempo of trophic evolution and its impact on mammalian diversification. *Proceedings of the National Academy of Sciences,* **109,** 7008-7012.

Rabosky, D.L. & Lovette, I.J. (2008) Density-dependent diversification in North American wood warblers. *Proceedings of the Royal Society B: Biological Sciences,* **275,** 2363-2371.

Revell, L.J. (2012) phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, **3,** 217-223.
Sakamoto, M., Benton, M.J. & Venditti, C. (2016) Dinosaurs in decline tens of millions of years before their final extinction. *Proceedings of the National Academy of Sciences*, **113,** 5036-5040.
Scholtz, C.H., Davis, A.L.V. & Kryger, U. (2009) *Evolutionary biology and conservation of dung beetles*. Pensoft Sofia.
Spriggs, E.L., Christin, P.-A. & Edwards, E.J. (2014) C4 photosynthesis promoted species diversification during the Miocene grassland expansion. *Plos One*, **9,** e97722.
Starrfelt, J. & Liow, L.H. (2016) How many dinosaur species were there? Fossil bias and true richness estimated using a Poisson sampling model. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **371,** 20150219.
Strömberg, C.A. (2002) The origin and spread of grass-dominated ecosystems in the late Tertiary of North America: preliminary results concerning the evolution of hypsodonty. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **177,** 59-75.
Strömberg, C.A. (2005) Decoupled taxonomic radiation and ecological expansion of open-habitat grasses in the Cenozoic of North America. *Proceedings of the National Academy of Sciences*, **102,** 11980-11984.
Strömberg, C.A. (2011) Evolution of grasses and grassland ecosystems. *Annual Review of Earth and Planetary Sciences*, **39,** 517-544.
Strömberg, C.A., Dunn, R.E., Madden, R.H., Kohn, M.J. & Carlini, A.A. (2013) Decoupling the spread of grasslands from the evolution of grazer-type herbivores in South America. *Nature Communications*, **4,** 1478.
Tarasov, S., Vaz-de-Mello, F.Z., Krell, F.T. & Dimitrov, D. (2016) A review and phylogeny of Scarabaeine dung beetle fossils (Coleoptera: Scarabaeidae: Scarabaeinae), with the description of two Canthochilum species from Dominican amber. *Peerj*, **4**.
Toljagić, O., Voje, K.L., Matschiner, M., Liow, L.H. & Hansen, T.F. (2017) Millions of years behind: slow adaptation of ruminants to grasslands. *Systematic Biology*, **67,** 145-157.
Wellborn, G.A. & Langerhans, R.B. (2015) Ecological opportunity and the adaptive diversification of lineages. *Ecology and Evolution*, **5,** 176-195.
Wing, S.L. & Sues, H.-D. (1992) Mesozoic and early Cenozoic terrestrial ecosystems. *Terrestrial ecosystems through time: evolutionary paleoecology of terrestrial plants and animals*.
Wirta, H., Orsini, L. & Hanski, I. (2008) An old adaptive radiation of forest dung beetles in Madagascar. *Molecular Phylogenetics and Evolution*, **47,** 1076-1089.

Appendix E

Table V-1: Diversification Analyses Young Age Calibration.

Estimated diversification rates, likelihood and ΔAICc/AICc-weights for each model applied to the dung beetle phylogeny with younger age calibrations, ordered by increasing ΔAICc/decreasing AICc-weights. Following terminology from RPANDA, parameters λ 1 and μ 1 are the base speciation or extinction rates respectively, whereas λ 1 and μ 2 are either the exponent or coefficient, depending on whether the model is exponential or linear (those values are NA for models with constant dependences, where no such parameter is estimated).

Model	λ1	λ2	μ1	μ2	LogLik	ΔAICc	AICcwt
Herb extant-heavy exp_lin	0.149	0.012	0.174	-0.173	-2322.390	0.000	0.346
C4 extant-heavy lin_lin	0.188	0.030	-0.142	0.306	-2322.442	0.104	0.329
Mammal extant-heavy lin_lin	0.136	0.005	-0.139	0.140	-2322.550	0.319	0.295
Mammal extant-heavy exp_lin	0.144	-0.057	0.134	-0.128	-2324.915	5.049	0.028
bd exp_lin	0.083	0.025	0.077	-0.007	-2328.273	11.766	0.001
APDactyl exp_lin	0.094	-0.003	-0.045	0.058	-2328.788	12.794	0.001
APDactyl lin_lin	0.082	0.027	-0.030	0.060	-2328.906	13.031	0.001
Herb lin_lin	0.076	0.027	-0.026	0.043	-2330.221	15.660	0.000
Mammal lin_lin	0.077	0.023	-0.022	0.035	-2330.375	15.969	0.000
APDactyl extant-heavy exp_lin	0.113	-0.083	0.062	-0.028	-2331.114	17.447	0.000
Mammal exp_lin	0.115	-0.019	-0.053	0.021	-2333.131	21.482	0.000
Grass extant-heavy exp_const	0.100	0.027	-0.018	NA	-2334.163	21.516	0.000
APDactyl lin_const	0.108	0.011	-0.042	NA	-2334.177	21.543	0.000
Mammal lin_const	0.108	0.011	-0.042	NA	-2334.215	21.620	0.000
Herb exp_lin	0.114	-0.014	-0.053	0.022	-2333.243	21.706	0.000
Herb lin_const	0.108	0.011	-0.042	NA	-2334.332	21.853	0.000
bd const_const	0.124	NA	0.053	NA	-2335.381	21.929	0.000
Grass lin_const	0.117	0.014	-0.048	NA	-2334.377	21.943	0.000
bd exp_const	0.109	-0.004	0.023	NA	-2334.491	22.171	0.000
C4 lin_const	0.114	0.012	-0.044	NA	-2334.592	22.374	0.000
bd lin_const	0.106	0.000	0.019	NA	-2334.644	22.478	0.000
Grass extant-heavy lin_const	0.111	0.012	-0.043	NA	-2334.685	22.559	0.000
Grass exp_const	0.119	0.013	-0.046	NA	-2334.694	22.578	0.000
C4 extant-heavy exp_const	0.118	0.012	-0.046	NA	-2334.695	22.580	0.000
C4 exp_const	0.118	0.012	-0.046	NA	-2334.971	23.131	0.000
C4 exp_lin	0.107	-0.063	-0.042	0.054	-2334.050	23.319	0.000
C4 extant-heavy lin_const	0.110	0.012	-0.043	NA	-2335.110	23.409	0.000
Grass exp_lin	0.110	-0.071	-0.044	0.053	-2334.173	23.565	0.000
Herb extant-heavy exp_const	0.119	0.017	-0.052	NA	-2335.539	24.268	0.000

Table V-1 Continued

Model	λ1	λ2	μ1	μ2	LogLik	ΔAICc	AICcwt
Mammal extant-heavy exp_const	0.119	0.016	-0.052	NA	-2335.554	24.298	0.000
bd lin_lin	0.110	0.001	0.035	0.001	-2334.562	24.344	0.000
Herb exp_const	0.116	0.012	-0.045	NA	-2335.585	24.361	0.000
Mammal exp_const	0.116	0.012	-0.045	NA	-2335.589	24.367	0.000
APDactyl extant-heavy exp_const	0.119	0.014	-0.051	NA	-2335.592	24.375	0.000
C4 extant-heavy exp_lin	0.104	-0.040	-0.036	0.029	-2334.585	24.390	0.000
APDactyl exp_const	0.116	0.012	-0.045	NA	-2335.612	24.413	0.000
APDactyl extant-heavy lin_const	0.067	0.015	0.002	NA	-2335.765	24.719	0.000
Herb extant-heavy lin_const	0.049	0.032	0.000	NA	-2336.012	25.215	0.000
Grass extant-heavy exp_lin	0.113	-0.025	-0.046	0.018	-2335.187	25.593	0.000
C4 lin_lin	0.130	-0.011	-0.065	0.026	-2335.465	26.150	0.000
Grass lin_lin	0.133	-0.013	-0.069	0.029	-2335.833	26.884	0.000
Herb extant-heavy lin_lin	0.047	0.038	-0.008	0.012	-2336.254	27.728	0.000
Grass extant-heavy lin_lin	0.101	0.047	-0.025	-0.052	-2338.645	32.508	0.000
APDactyl extant-heavy lin_lin	0.097	-0.010	0.038	-0.021	-2339.788	34.796	0.000
Mammal extant-heavy lin_const	0.020	0.058	0.000	NA	-2375.908	105.006	0.000

Table V-2: Diversification Analyses Old Age Calibration.

Estimated diversification rates, likelihood and ΔAICc/AICc-weights for each model applied to the dung beetle phylogeny with older age calibrations, ordered by increasing ΔAICc/decreasing AICc-weights. Following terminology from RPANDA, parameters λ 1 and μ 1 are the base speciation or extinction rates respectively, whereas λ 1 and μ 2 are either the exponent or coefficient, depending on whether the model is exponential or linear (those values are NA for models with constant dependences, where no such parameter is estimated).

Model	λ1	λ2	μ1	μ2	LogLik	ΔAICc	AICcwt
APDactyl lin_const	0.115	-0.030	-0.056	NA	-2527.934	0.000	0.352
Herb lin_const	0.118	-0.037	-0.055	NA	-2528.046	0.224	0.315
Mammal lin_const	0.118	-0.034	-0.057	NA	-2528.569	1.270	0.186
bd exp_lin	0.054	0.018	0.043	-0.003	-2529.247	4.656	0.034
C4 lin_const	0.084	-0.016	-0.032	NA	-2530.848	5.827	0.019
bd const_const	0.078	NA	0.029	NA	-2532.371	6.851	0.011
C4 extant-heavy lin_const	0.098	-0.012	-0.048	NA	-2531.582	7.295	0.009
Grass extant-heavy lin_lin	0.093	0.000	-0.038	-0.026	-2530.743	7.647	0.008
Grass extant-heavy lin_const	0.087	-0.008	-0.037	NA	-2531.804	7.740	0.007
bd exp_const	0.078	0.000	0.028	NA	-2532.371	8.873	0.004
Grass extant-heavy exp_const	0.079	0.000	-0.030	NA	-2532.374	8.879	0.004
bd lin_const	0.078	0.000	0.028	NA	-2532.378	8.887	0.004
Grass exp_const	0.077	0.001	-0.027	NA	-2532.381	8.894	0.004
Herb extant-heavy exp_const	0.077	0.007	-0.028	NA	-2532.485	9.101	0.004
Mammal extant-heavy exp_const	0.077	0.007	-0.028	NA	-2532.488	9.108	0.004
Grass lin_const	0.071	-0.003	-0.017	NA	-2532.498	9.128	0.004
Herb exp_const	0.080	0.008	-0.031	NA	-2532.504	9.139	0.004
Mammal exp_const	0.080	0.008	-0.031	NA	-2532.506	9.143	0.004
APDactyl extant-heavy exp_const	0.077	0.008	-0.028	NA	-2532.515	9.161	0.004
APDactyl exp_const	0.080	0.008	-0.031	NA	-2532.530	9.192	0.004
C4 lin_lin	0.070	-0.016	-0.017	0.036	-2531.585	9.331	0.003
C4 exp_const	0.081	0.008	-0.031	NA	-2532.631	9.394	0.003
Herb extant-heavy lin_lin	0.137	-0.042	0.082	-0.029	-2531.671	9.502	0.003
C4 extant-heavy exp_const	0.081	0.008	-0.031	NA	-2532.902	9.935	0.002
bd lin_lin	0.078	0.000	0.033	0.000	-2532.369	10.900	0.002
Mammal extant-heavy exp_lin	0.067	0.004	-0.028	0.024	-2533.523	13.208	0.000
Grass extant-heavy exp_lin	0.073	-0.026	-0.024	0.011	-2533.746	13.653	0.000
Herb extant-heavy exp_lin	0.065	-0.013	-0.021	0.013	-2533.769	13.699	0.000
Grass lin_lin	0.070	-0.002	-0.019	0.013	-2534.156	14.474	0.000
C4 exp_lin	0.065	-0.050	-0.017	0.029	-2534.857	15.875	0.000

Table V-2 Continued

Model	λ1	λ2	μ1	μ2	LogLik	ΔAICc	AICcwt
Grass exp_lin	0.075	-0.017	-0.026	0.020	-2535.044	16.249	0.000
Mammal exp_lin	0.069	0.019	-0.028	0.021	-2535.104	16.369	0.000
Mammal extant-heavy lin_lin	0.058	0.005	-0.004	-0.004	-2535.123	16.407	0.000
APDactyl extant-heavy lin_const	0.056	0.002	0.002	NA	-2536.144	16.419	0.000
Herb exp_lin	0.070	0.005	-0.030	0.022	-2535.142	16.445	0.000
APDactyl extant-heavy exp_lin	0.065	-0.002	0.023	-0.016	-2535.367	16.894	0.000
Mammal extant-heavy lin_const	0.050	0.010	0.003	NA	-2536.913	17.958	0.000
C4 extant-heavy exp_lin	0.073	-0.040	-0.027	0.028	-2535.950	18.062	0.000
C4 extant-heavy lin_lin	0.065	0.023	-0.007	-0.038	-2536.609	19.379	0.000
Herb extant-heavy lin_const	0.053	0.014	0.012	NA	-2538.896	21.923	0.000
Herb lin_lin	0.060	0.022	-0.003	-0.030	-2537.967	22.095	0.000
APDactyl exp_lin	0.064	0.001	-0.029	0.035	-2538.869	23.899	0.000
APDactyl lin_lin	0.059	0.021	0.000	-0.030	-2538.953	24.068	0.000
Mammal lin_lin	0.057	0.019	-0.013	0.004	-2539.919	25.999	0.000
APDactyl extant-heavy lin_lin	0.066	0.070	0.012	-0.104	-2576.790	99.741	0.000

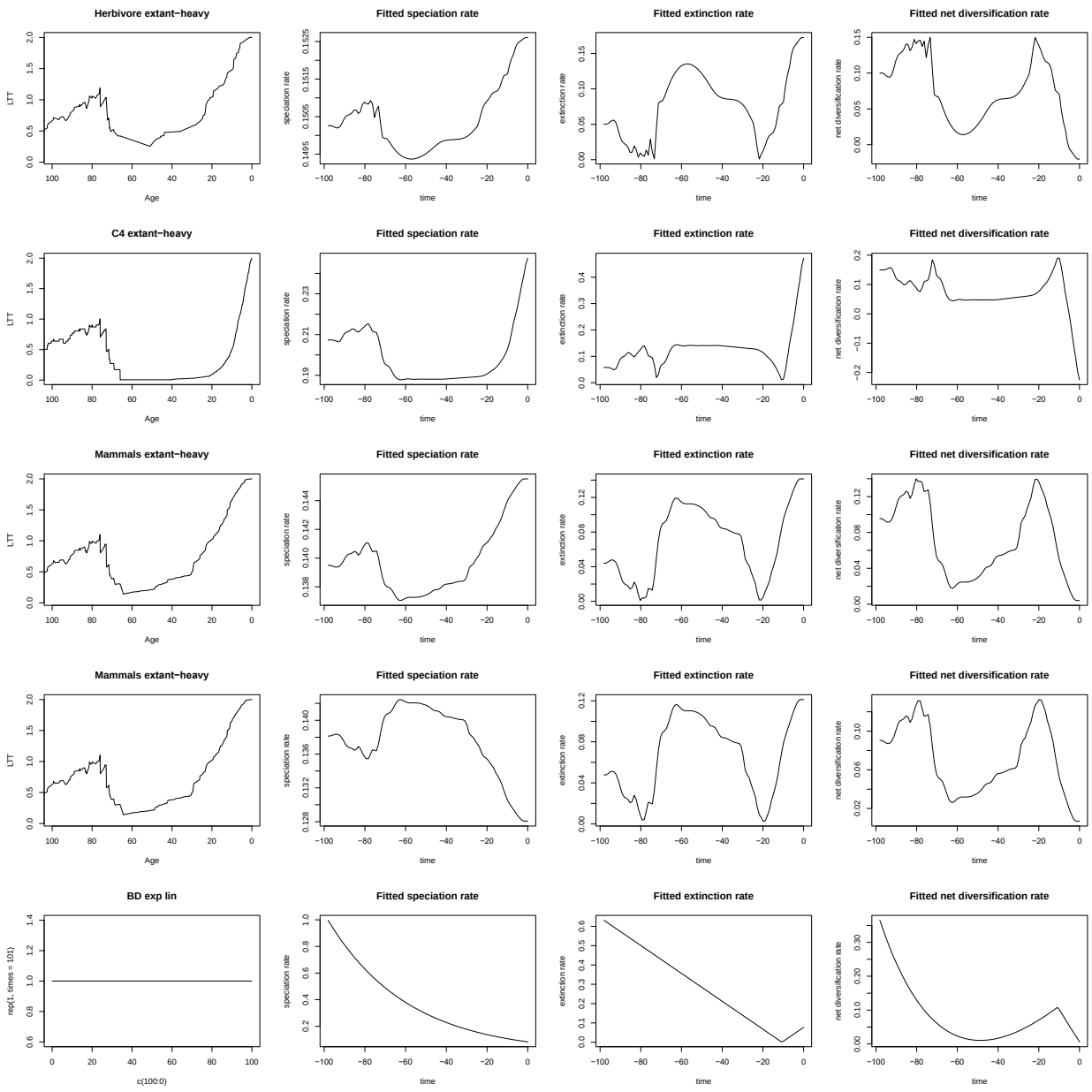

Figure V-1: Inferred Diversification Rates Young Tree.

Estimated diversification rates and the environmental data they are dependent on, for the five best scoring models from top to bottom. Left to right: underlying standardized LTT data, speciation rate, extinction rate, and net diversification rate.

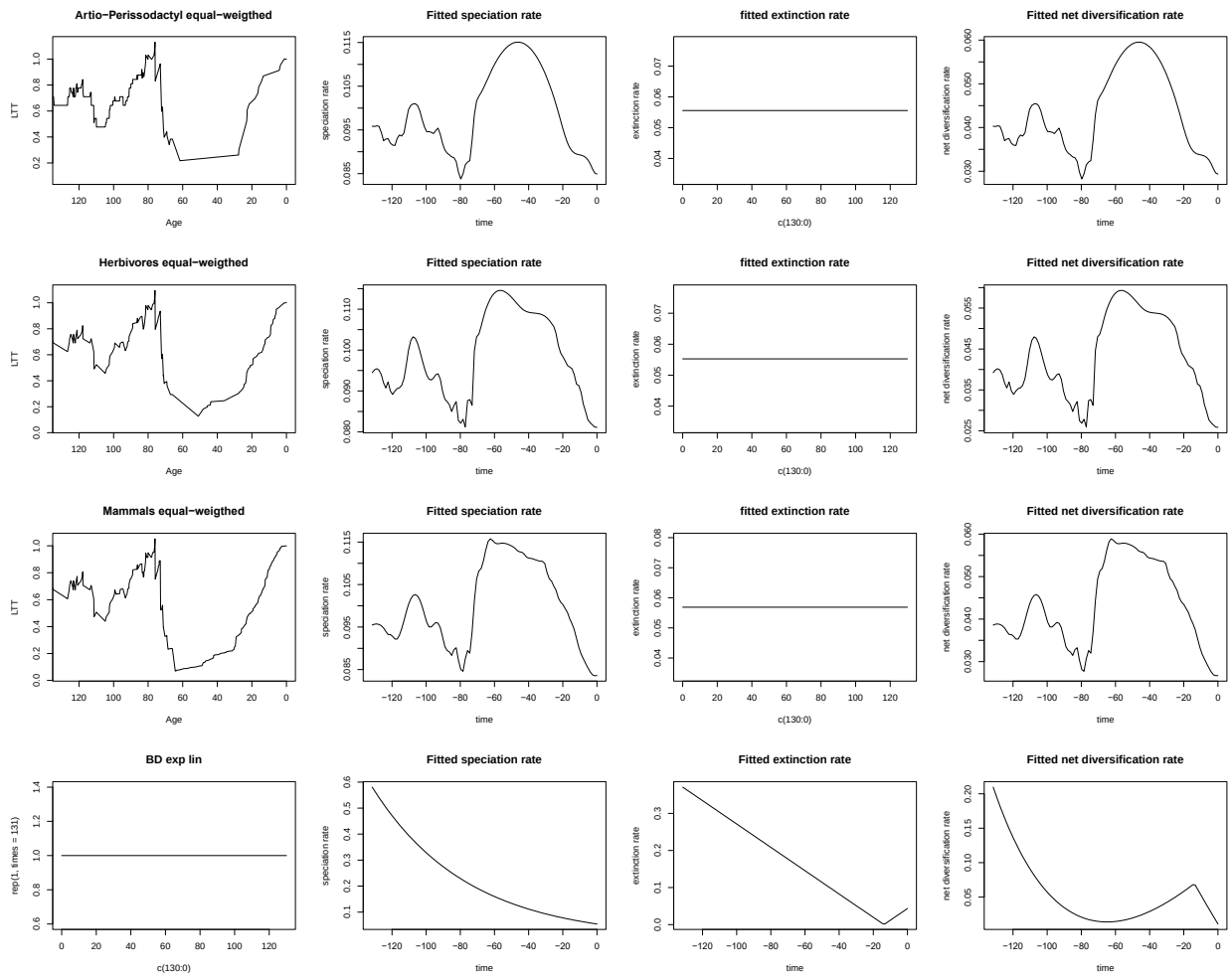

Figure V-2: Inferred Diversification Rates Old Tree.

Estimated diversification rates and the environmental data they are dependent on, for the four best scoring models from top to bottom. Left to right: underlying standardized LTT data, speciation rate, extinction rate, and net diversification rate.

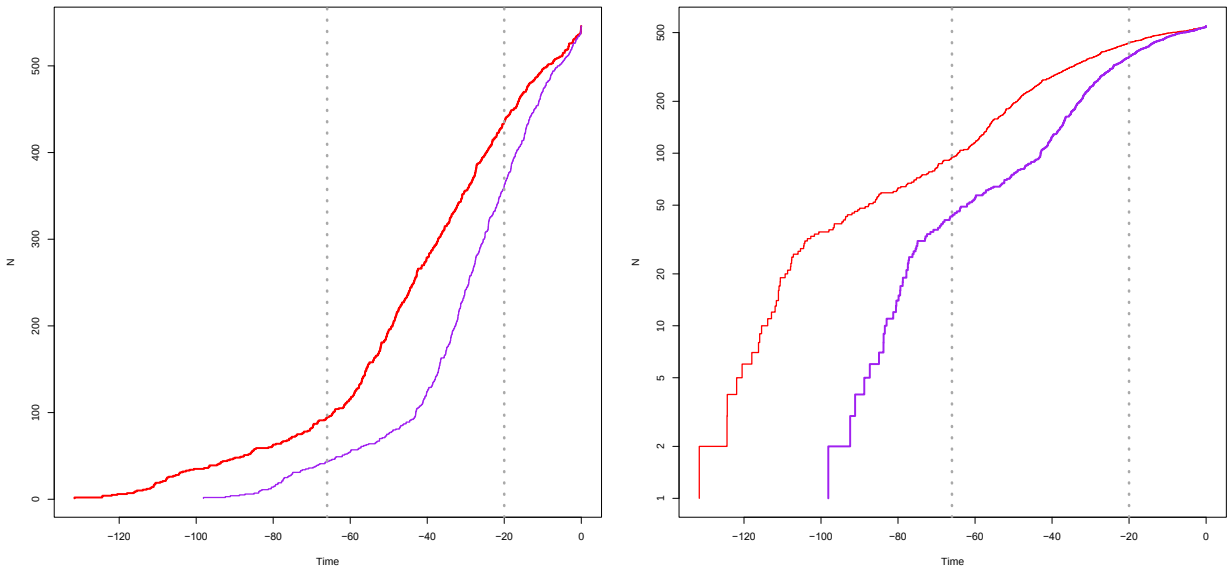

Figure V-3: Dung Beetle Lineages Through Time.

Lineage Through Time plots for the young (purple) and old (red) dung beetle phylogeny. The y-axis of the plot on the right is on a log-scale. Dotted gray lines indicate the end-Cretaceous extinction at 66mya and the time of decreasing diversification rates at 20mya.

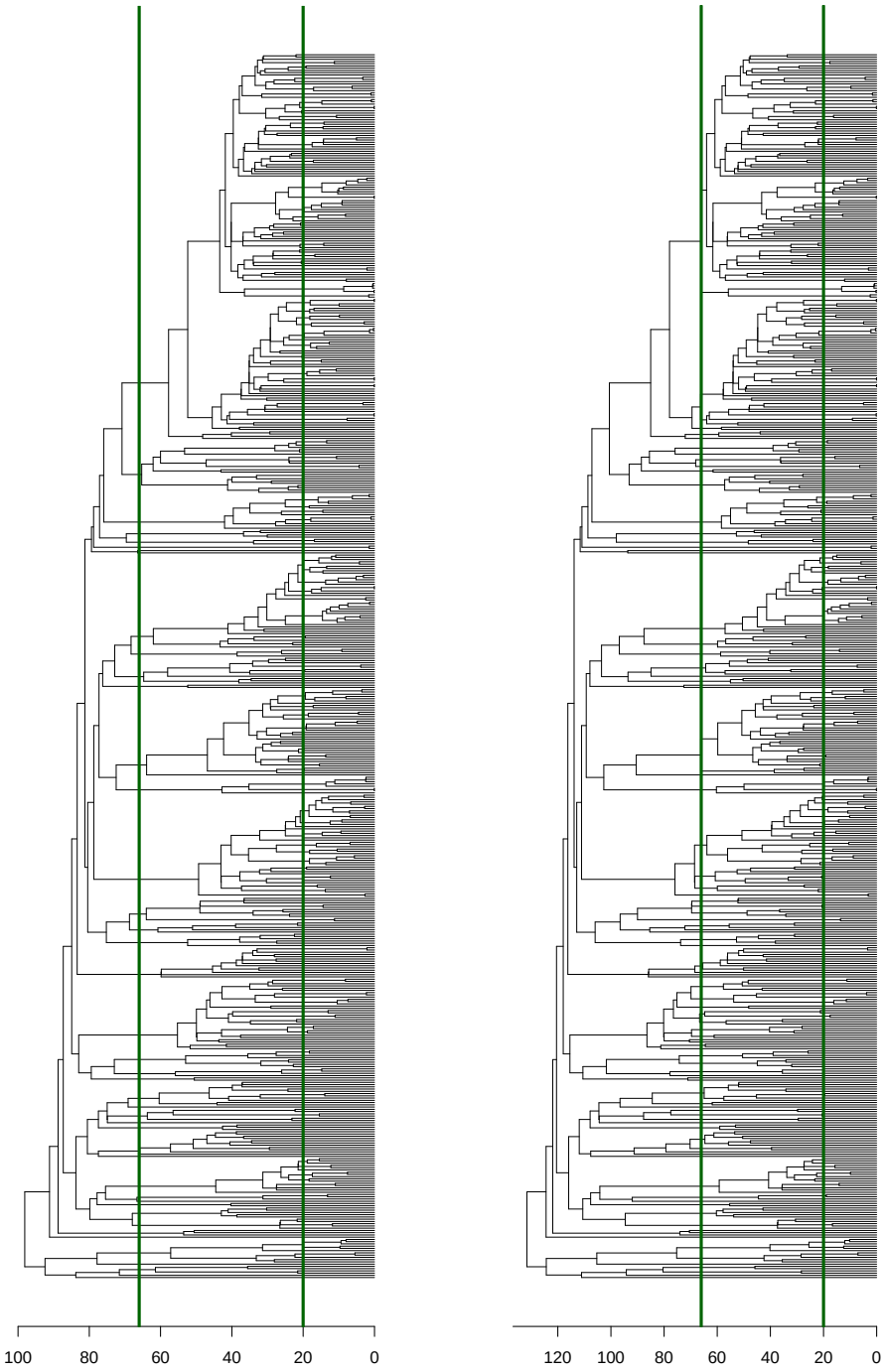

Figure V-4: Young and Old Dung Beetle Phylogeny.

Dated dung beetle phylogenies, derived from young time-calibrations on the left and old time-calibrations on the right. Solid green lines indicate the end-Cretaceous extinction at 66mya and the time of decreasing diversification rates at 20mya.

CONCLUSION

The factors that influence the diversification dynamics of a clade are manifold, and so are the potential biases that can affect both our data and our inference methods when we attempt to investigate those factors. With increasing awareness of the issues that plague the field, it is crucial to work towards tools that either allow to resolve them or at least detect them. Here, I presented some tools to add to the arsenal of future phylogeneticists, in the hope they may prove useful in the quest to derive more accurate and reliable conclusions from phylogenetic analyses.

The R package MonoPhy was designed as a helper tool to judge to which extent a new phylogeny agrees or disagrees with taxonomic hypotheses of how species are related. Allowing to use automated taxonomical assignment or input of one or several levels of customized taxonomic information should give a researcher the needed flexibility to asses monophyly of any taxonomic level or other taxon-associated aspects desired. Detailed, explorable output and different plotting options allow for both quick overviews and detailed explorations.

Adding model adequacy to the standard procedures of diversification analyses is a crucial step towards better results and more confidence in one's conclusions. The R package BoskR tests model adequacy of a set of birth-death related diversification models but can in theory be applied to any model under which phylogenies can be simulated. Using measures of tree shape (as estimated via the tree's modified graph Laplacian spectrum) as the test statistic to judge the adequacy of a model for a certain tree, it provides a good starting point to explore in what aspects of tree shape the model differs from the empirical tree it is supposed to describe. When applied to a large set of empirical phylogenies, BoskR reveals that even a modest set of basic available models can adequately describe diversification in a large majority of trees. However, it also shows that model fit does not have to coincide with adequacy, and if used by itself could give false confidence in flawed results.

In my empirical chapters, I tried to contribute to two contended questions pertaining dung beetle evolution: from which origin the Scarabaeinae spread through the world, and which role mammals and other dung producers played in the diversification of dung beetle lineages. Using a new dated phylogeny and up to date model-based methods, I gain support for some of the disputed hypotheses in the field. The inferred age of Scarabaeinae, as well as their estimated ancestral range supports the hypothesis of Gondwanan vicariance with some indication that access to new areas may have fueled diversification. Furthermore, while the age of dung beetles as well as inferred relations of diversification rates and diversity of dung producers would seem to support the hypothesis that dung beetles fed on dinosaur droppings and were affected by the end-Cretaceous extinction, they could not support a connection with the rise of grasslands.

VITA

Orlando Schwery was born in Basel, BS, Switzerland. He attended high school at the Kantonsschule Frauenfeld, TG and graduated in 2006. After graduation he attended the University of Zurich, ZH where he received a Bachelor of Science in Biology in 2011. Subsequently, he joined the Institute of Systematic Botany and received a Master of Science in Systematics & Evolution in 2013. He accepted a graduate teaching assistantship at the University of Tennessee, Knoxville, in the Ecology and Evolutionary Biology Department. He graduated with a Doctor of Philosophy in Fall 2019.