Species delimitation and diversification history in Rhinoceros beetles: how many and why so many species?

by

Jen-Pan Huang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Ecology and Evolutionary Biology)
in The University of Michigan
2016

Doctoral Committee:

Professor L. Lacey Knowles, Chair
Associate Professor Catherine E. Badgley
Associate Professor Gregory J. Dick
Assistant Professor Stephen A. Smith

© Jen-Pan Huang 2016

ACKNOWLEDGEMENTS

I would like to thank all the different people that helped us acquire samples, including: Diego Alvarado-Serrano, Marco Bolaños, Fortuné Chalumeau, Wei-Yun Chen, Francis Deknuydt, Patrick Demez, Douglas Emlen, Wei Hu, Michael Ivie, Cliff Keil, Hsin-Ping Ko, Jonathan Lai, Elisa Levy, Edwin Levy, Javier Lamber, Chung-Ping Lin, Hervé Magnin, Rajindra Mahabir, Franklin Neira, Diego Peña, J. Mark Rowland, Felix Stumpe, Ping Feng Tsai, William Wallin, Chia-Hsuan Wei, and Guy Van-Laere, as well as the Ministerio del Ambiente of Ecuador, Pontificia Universidad Católica del Ecuador, Parc national de la Guadeloupe, Yanayacu Biological Station, and Parque Nacional Yasuní. I also thank Johnnie Chong, Hector Gasca, and Brett Ratcliffe for providing occurrence data for ecological niche models. I would also like to thank all the funding sources for supporting my study – Graduate Student Research Awards from Society of Systematic Biologists, Rosemary Grant from the Society for the Study of Evolution, Block Grants and Hinsdale-Walker scholarship from the department of Ecology and Evolutionary Biology, Pre- and Post Candidate fellowships from the Rackham Graduate School, NSF Dissertation improvement grant DEB-15-01462, and Ammerman Fund from the Insect Division of Ruthven Museums.

I became an ardent beetle hunter since my childhood. The diversity in size, shape, and color of beetles always amazes me; it seems like there are always new forms to be found in beetles. The question about how to designate species to different life forms and why are they so different has always puzzled me. Questions such as how many different beetles are there in my collection and what are the relationships between these beetles (in short, how to place different beetles in the same box that makes sense to me) have been important yet difficult to me. It has been a great experience for me to use the biological system, that I fancy for a long time to tackle the two long lasting challenges in biology in my thesis. I will explain more detail why the two Rhinoceros beetle groups are chosen to test different sets of questions in each of the following chapter more explicitly and discuss what my results mean and their broader impacts on the field of biological study.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ii
LIST OF FIGURES vii
LIST OF TABLES xii
ABSTRACT xiii
CHAPTER 1: Introduction 1
CHAPTER 2: The species versus subspecies conundrum: quantitative delimitation from integrating multiple data types within a single Bayesian approach in Hercules beetles 4
2.1 Abstract 4
2.2 Introduction 5
2.3 Materials and Methods 12
Sample collection and DNA sequencing 12
Morphological divergence 15
Species delimitation 17
Tests of niche divergence and assessments of geographic overlap 21
2.4 Results 24
DNA sequence data 24
Geographic overlap and niche similarity 26
Morphological differences 32
Species tree 35
Species delimitation 36
2.5 Discussion 41
Species and subspecies are statistically equivalent 42
Different positions along the speciation continuum 44
Insights from integrating genetic and phenotypic data 46

Quantitative delimitation methods and taxonomic designations 47
2.6 Conclusion 51
CHAPTER 3: The Great American Biotic Interchange and diversification history in *Dynastes* beetles (Scarabaeidae; Dynastinae) 53
3.1 Abstract 53
3.2 Introduction 54
3.3 Materials and Methods 60
Species trees 60
Reconstructing biogeographic history 61
Estimating speciation/diversification rate 61
Testing the effects of different geographic distributions and altitudinal preferences on species diversification 64
3.4 Results 65
Biogeographic history reconstruction 65
Changes in diversification rates 65
The effects of different geographic distributions and altitudinal preferences 69
3.5 Discussion 70
Biogeography 71
Diversification process 73
3.6 Conclusion 76
CHAPTER 4: Tests of divergence across multiple levels of biodiversity in *Xylotrupes* beetles: impact of oceanic and forest barriers on diversification dynamics across the Indo-Australian Archipelago 78
4.1 Abstract 78
4.2 Introduction 79
4.3 Material and Methods 83
DNA extraction, sequencing and alignment 83
Estimates of phylogenetic relationships and divergence times 86
Regional differentiation and dispersal 88
Diversification models 89
Speciation duration 91
Estimates of population divergence 92

4.4 Results 93
Species-tree estimate 93
Regional differentiation and dispersal 95
Changes in diversification rates through space and time 95
Speciation duration 100
Population subdivisions 101
4.5 Discussion 102
Differences in the processes structuring population versus species-level divergence 103
The duration of speciation as a filter for the processes that contribute to species diversity patterns 106
Implications from an integrative phylogeographic and biogeographic study 108
4.6 Conclusion 109
CHAPTER 5: Conclusion 111
BIBLIOGRAPHY 114

LIST OF FIGURES

Fig. 2.1. Representative major males of each of the white Hercules (upper five images) and Giant Hercules (lower ten images) beetle taxa with the corresponding abbreviations used throughout the text and other tables and figures. Images of *D. h. paschoali* and *D. h. trinidadensis* courtesy of Jonathan Lai (2008). 9

Fig. 2.2. Landmarks used for thoracic (blue) and cephalic (red) horns, where the numbers refer to homologous points that within the three distinct types of cephalic horns of the White and the Giant Hercules beetles (shape differences are not compared between species groups given the obvious distinctiveness of the two groups; the numbered landmarks differ between the White and the Giant Hercules taxa). 16

Fig. 2.3. Pairwise sequence divergence (K2P distances) between (grey) and within (white) putative taxa of White and Giant Hercules beetles. 25

Fig. 2.4. Projected distributions that contrasts the allopatry of all White Hercules taxa (i.e., gr, ty, hy, mo, and ma) to the parapatrically distributed taxa within the Giant Hercules group of South and Central America (i.e., names in bold and distributions shown in color: lic, ecu, sep, and occ) and the island endemics or allopatric Giant Hercules taxa (blu, tre, her, rei, mor, and pas). Circles indicate actual collection records used to estimate distributions of each taxon using the program MAXENT (see methods). Abbreviations and their corresponding taxonomic names can be found in Fig. 2.1. 26

Fig. 2.5. Comparison of predicted current and past distributions for the White Hercules taxa. Light grey: gr, dark grey: hy, green: ty, orange: mo, and dark red: ma. 28

Fig. 2.6. Comparison of predicted current and past distributions for the Giant

Hercules taxa. Green: sep, light green: occ, dark red: lic, yellow: ecu, and dark grey: pas. mor and blu, and island endemic her, rei, and tri are excluded from this map.. 29
Fig. 2.7. Results from niche background similarity tests. In most taxa the observed niche similarity, calculated by the index D (and shown as the thin vertical line), did not differ among putative sister taxa (see Fig.) any more than expected from the difference in niche similarity that can be attributed to differences in the geographic distribution of taxa, or tended towards supporting niche conservatism, based on a similarity index that was higher than expected (where the histograms show the expected frequency distribution of niche similarity scores based on random samples of background points across the projected distribution of the taxon involved in the pairwise comparison). Only two cases of significant niche divergence were identified (marked as *Div). In both cases, significant divergence was only evident when the observed niche similarity was compared against one of the two backgrounds (e.g., niche divergence is apparent between occ and lic when the environmental data at collection localities of lic are compared to those from random samples across the projected distribution of occ, but not when environmental data at collection localities of occ is compared to those from random samples across the projected distribution of lic). Also note that the lic and occ comparison represents divergence between highland Andean cloud forest and the Choco ecoregion, and the other case of niche divergence involves comparison of taxa between subtropical Southeast United States and neotropical Central America (i.e., ty and ma). Abbreviations and the corresponding taxonomic names are listed in Fig. 2.1. 31
Fig. 2.8. Position of taxa in morphospace based on geometric morphometric analyses of the cephalic and thoracic male horn shape based on the first two PCs, which explained most of the differences in shape among taxa (i.e., between 60 and 82% of the variance). Landmarks in tangent space are shown for the most extreme shapes on upper left and lower right corners of each PCA. Permutation tests based on Euclidean distances between samples from PC space indicate significant differences in horn shape across taxa in all four analyses ($P < 0.01$). Different taxa of White Hercules beetles are shown with black = gr, red = hy, blue = mo, green = ma, and cyan = ty, and different taxa of Giant Hercules beetles (excluding rei because of its extremely divergent morphology) are shown with dark green = blu,

red = ecu, blue = lic, green = her, cyan = mor, yellow = pas, magenta = occ, black = sep, and orange = tri. Taxonomic abbreviations and the species names can be found in Fig. 2.1. 34
Fig. 2.9. Species tree of *Dynastes* estimated from *BEAST; posterior probabilities above 50% are shown on nodes. Divergence times with confidence intervals estimated using a rate of 0.0115/site/million years are shown at each node. Note that outlines of major male horn shape are not drawn to scale (see Fig. 2.1 for relative sizes). 36
Fig. 2.10. Results from iBPP analyses using combined genetic and morphological datasets; note that outlines of major male horn shape shown adjacent to the abbreviated taxonomic names are not drawn to scale (see Fig. 2.1 for relative body sizes). Support values reported for each node are based on the algorithm setting 0 for the rjMCMC, the algorithm setting 1 for the Brownian motion model, and for four different priors corresponding to large (versus small) ancestral population sizes with relatively deep (versus shallow) divergence times. Specifically, the support values in each box correspond to analyses with the following different priors: upper left, θ = G(1, 10) and τ = G(1, 10); lower left, θ = G(1, 10) and τ = G(2, 2000); upper right, θ = G(2, 2000) and τ = G(1, 10); lower right, θ = G(2, 2000) and τ = G(2, 2000). Taxonomic abbreviations and the corresponding names can be found in Fig. 2.1. .. 37
Fig. 2.11. Comparison of BPP results for molecular data using both 0 and 1 rjMCMC algorithms and a reconstructed species tree from *BEAST as guide tree. Splits without absolute supports (<1.0) are shown in red boxes. 39
Fig. 2.12. Comparison of iBPP results for analyses based on morphological data only using rjMCMC algorithm 0 and a fine tune setting of 0 and 1 for BM model and a reconstructed species tree from *BEAST as guide tree. Splits without absolute supports (<1.0) are shown in red boxes. 39
Fig. 3.1. Results from BAMM and LAGRANGE analyses. Branch color represents estimated speciation rate, where a warmer color indicates a faster rate. A red dot on the branch leading to subgenus *Dynastes* indicates a speciation rate shift event. Black and white squares denote geographic states of South and North America, respectively. Black and white squares located on branches denote reconstructed ancestral geographic area, while those that next to the taxon abbreviations current

geographic states (Note that, the reconstructed ancestral state for the common ancestor of subgenus *Dynastes* can be either South America or widespread). White triangles indicate inferred dispersal events into North America. A grey shaded area indicates the time frame when the Isthmus of Panama was completely formed (3.4 – 3.6 MYA). Representative samples of *D. neptunus* (subgenus *Theogenes*) and *D. hercules ecuatorianus* (subgenus *Dynastes*) are shown with a scale bar of 1 cm. 58
Figure 3.2. BAMM results from analysis incorporating information about possible missing taxa in the data. (A) The *Dynastes* species tree with branches colored according to estimated speciation rate. (B) Distinct rate shift patterns and their frequencies determined by applying a threshold of Bayes factor = 3. A grey circle on the branch indicates the position where rate shift occurred. 63
Fig. 3.3. Results from rate through time plots. The x-axis is in a scale of million years. Solid lines indicate the mean rates, while the grey areas represent the 5% to 95% Bayesian credible regions for the distributions of the rates. 67
Fig. 3.4. Results from macroevolutionary cohort analysis. A correlation matrix based on speciation rates between tip lineages of the phylogeny is plotted, where each correlation is a posterior frequency that the two compared species are found in the same macroevolutionary rate regime. A warmer color represents a higher correlation than a colder color. The correlation between any two species can be found by locating their intersection in the matrix. 68
Fig. 3.5. Comparisons of estimated speciation rates between trait states. Left panel: the estimated rates between South (black) and North (white) American *Dynastes* beetles. Right panel: the estimated tip rates between lowland living (black) and highland living (white) species. 70
Fig. 4.1. The map of Indo-Australian Archipelago and representatives of major males from six *Xylotrupes* species groups. Solid lines indicate geographic breaks between zoological regions and colored dashed lines show the geographic distributions of each species groups. Geographic areas that would have connected isolated landmasses during glacial periods (Hall 2012) are shown in light greyish blue color. 82
Fig. 4.2. Distribution of ancestral areas (shown above nodes and color coded by region) across the estimated species tree (with posterior probabilities given for nodes

with greater than 50%); the names of *Xylotrupes* taxa color coded according to their distribution across regions (Wallacea in black, Sundaland in red, Australasia in blue, Indo-China in orange, the Philippines in green, India in grey, and the Himalayan region in brown; see Fig. 4.1 for names of species groups). Branches with multiple color squares represent composite ancestral areas and curved arrows mark inferred dispersal events. Estimated divergence times for the nodes, as well as 95% confidence intervals, for each node are shown, where the corresponding time of divergence in shown by the legend (in mya). A red arrow identifies the branch where a diversification rate shift is inferred using MEDUSA. 94

Fig. 4.3. Posterior probability densities of estimated diversification parameters (speciation rate, extinction rate, and transition rate, which refers to the frequency of transitions between geographic states) for the two geographic states – islands (shown in white) versus continents (shown in black) – using GeoSSE speciation... 97

Fig. 4.4. Lineage through time plots (lower panel) and estimations of switch points between diversification rates under a model that favors two diversification rates (upper panel) and three diversification rates (middle panel). For the LTT plot (lower panel) the median (solid line) and 95% intervals (grey area) of LTT plot from 8,000 post-burnin species trees are shown..... 99

Fig. 4.5. Degree of population subdivision (as measured by F_{ST}) associated with (A) oceanic barriers (for each type of landmass; see methods for details) and (B) and geographic distance. F_{ST} values estimated between continental populations are shown in black, temporary island is in light grey, and island state is in dark grey. 102

LIST OF TABLES

Table 2.1. Summary of different usages of species versus subspecies designations among animal clades. 8
Table 2.2. Samples used for molecular analyses. 12
Table 2.3. Mean morphological Euclidean distance between White Hercules taxa. 33
Table 2.4. Mean morphological Euclidean distance between White Hercules taxa. 33
Table 2.5. Results from BPP v.3 analyses of molecular data that integrate over uncertainty in the species tree. 40
Table 3.1. Studied taxa. 57
Table 4.1. Summaries of taxa studied in this chapter. 83
Table 4.2. Summary of sequenced loci, sample sizes, and DNA variation. 86
Table 4.3. Comparison of the fit of different models from ML GeoSSE analyses. Specifically, models that assume equal or different speciation (λ), extinction (μ), and transition (δ) rates between geographic states are tested. 96
Table 4.4. Results from ANOVA for isolation by oceanic barrier. Specifically, the pairwise F_{ST}-values between populations isolated without oceanic barrier and those with oceanic barriers (both temporal and oceanic, see materials and methods) are compared. 101

ABSTRACT

My thesis study focuses on systematic biology, biogeography, and more generally evolutionary biology. Specifically, I address two grand challenges in biology – species delimitation and speciation process. Furthermore, I also attempt to bridge the gap between macro- and micro-evolutionary studies. The characteristics of two groups of rhinoceros beetles (genus *Dynastes* and *Xylotrupes*) I study make them ideal for addressing these questions – geographically widespread taxa with local morphological forms – also make them challenging to study. These groups exhibit inconsistency in taxonomic designations and the their geographic distributions imply complex historical processes in their diversification process. My research highlights both the power, but also the necessity, of an integrative framework that considers different data types, as well as quantitative approaches to test different hypotheses about species boundaries and the diversification process. For example, my first chapter revealed the arbitrariness in taxonomic decisions, even between closely related taxa from the same lineage, by demonstrating that species boundaries were statistically equivalent among taxa even though some were assigned as subspecies by previous taxonomic treatments. By establishing this taxonomic foundation, my studies on the effects of ecological and

geographic isolation on species diversification in the following chapters avoid the biases introduced by taxonomic ambiguity and inconsistency.

In my second chapter, I show that the effects of habitat stability/instability (e.g., changes in the geographic distribution and size of the habitats) outweighs the contributions of geological events that connect previously isolated biotas (e.g., the rise of the Isthmus of Panama) in promoting rapid diversification in Hercules beetles. Following the general theme of the contribution of different barriers to divergence, in Chapter 3 I extend the investigation to consider whether their effects are similar across different levels of biological organization – that is, in the structuring of patterns of genetic diversity among population, species, and faunal communities. This work shows that oceanic barriers between landmasses in the Indo-Australian Archipelago delineates zoological regions by structuring distinct faunal communities and promotes population subdivision in *Xylotrupes* beetles. However, the rate of species diversification appears to be associated with shifts in forest fragmentation across geological times (i.e., between Miocene and Pliocene). As such, this work highlights the decoupling of processes contributing to micro- and macro- evolutionary patterns, which only became evident because of my integrative approach to research that involves consideration of alternative mechanisms and study of divergence at multiple levels of biodiversity.

CHAPTER 1: Introduction

The diversity of life is a spectacular, and arguably the most fascinating, features of our planet. Generations of scientists have been pursuing not only the logic of the hierarchical structure among, but also the origin and maintenance of the endless forms of biodiversity (Howard & Berlocher 1998). However, after generations of studies, disagreement about how species should be designated persists (de Queiroz 2007). Moreover, because different speciation processes may proceed from divergence along different possible axes (e.g., along the ecological preference and/or the morphological axes) (Coyne & Orr 2004; Nosil et al. 2009), recognized species may differ depending upon the type of data used to delimit taxa (e.g., genetic versus phenotypic data). Differences in delimited taxa in turn may significantly affect our interpretation of the diversification history of lineages (Smith et al. 2013). That is, the studies of speciation and diversification processes are dependent on the results of species delimitation.

My dissertation research addresses two of the grand challenges in biology: species delimitation and speciation. Although individual chapters focus predominantly on one of the two conceptual areas, my research is motivated by the links between these areas. For example, I test whether currently designated species and subspecies can be statistically

supported under coalescent by integrating information from multiple datasets (molecular, phenotypic and ecological data). If not, can such inconsistencies be identified based on statistical results **(Chapter 2)**. I then use the statistical species delimitation results to test predictions from hypothesized historical processes that might have shaped the current species diversity pattern in order to understand what could be the driving and maintaining forces of biological diversification **(Chapter 3)**. Finally, I investigate the effects of different types of barriers on genetic, species, and faunal diversities. Specifically, I test whether the predominate mechanism that structures biodiversity can be not only evolutionary lineage specific, but also biodiversity level specific **(Chapter 4)**. My dissertation study thus demonstrates how to achieve a comprehensive understanding on the origin and maintenance of biodiversity pattern by integrating studies of species delimitation and diversification process in the same system.

My system of interest, the Rhinoceros beetle (Dynastinae), is a speciose lineage that comprises more than 1,000 species widely distributed throughout the world (Rowland & Miller 2012). The Rhinoceros beetle system provides a great opportunity to study species delimitation by integrating multiple types of data and diversification process at different biodiversity levels. For example, in addition to using neutral molecular dataset that represents demographic effects on the observed divergences between taxa, the male horn structure is highly diverse in size and shape among species and populations of the same species, which represents an divergence axis potentially reacts to the strength of sexual and/or natural selection. Additionally, the diverse geographic forms, e.g., island endemic

taxa, found in Rhinoceros beetles are great candidates to investigate effects of various historical/macroevolutionary processes on structuring the current day biodiversity patterns – population genetic structure, species diversity, and regional specific fauna.

CHAPTER 2: The species versus subspecies conundrum: quantitative delimitation from integrating multiple data types within a single Bayesian approach in Hercules beetles

2.1 Abstract

With the recent attention and focus on quantitative methods for species delimitation, an overlooked but equally important issue regards what has actually been delimited. This study investigates the apparent arbitrariness of some taxonomic distinctions, and in particular how species and subspecies are assigned. Specifically, I use a recently developed Bayesian model-based approach to show that in the Hercules beetles (genus *Dynastes*) there is no statistical difference in the probability that putative taxa represent different species, irrespective of whether they were given species or subspecies designations. By considering multiple data types, as opposed to relying exclusively on genetic data alone, I also show that both previously recognized species and subspecies represent a variety of points along the speciation spectrum (i.e., previously recognized species are not systematically further along the continuum than subspecies). For example, based on evolutionary models of divergence, some taxa are statistically distinguishable on more than one axis of differentiation (e.g., along both phenotypic and genetic

dimensions), whereas other taxa can only be delimited statistically from a single data type. Because both phenotypic and genetic data are analyzed in a common Bayesian framework, my study provides a framework for investigating whether disagreements in species boundaries among data types reflect (i) actual discordance with the actual history of lineage splitting, or instead (ii) differences among data types in the amount of time required for differentiation to become apparent among the delimited taxa. I discuss what the answers to these questions imply about what characters are used to delimit species, as well as the diverse processes involved in the origin and maintenance of species boundaries. With this in mind, I then reflect more generally on how quantitative methods for species delimitation are used to assign taxonomic status.

2.2 Introduction

The importance of species delimitation extends beyond immediate taxonomic goals in systematics (Mayr 1942; de Queiroz 2007; Wiens 2007) – that is, it is not simply a semantic issue. Taxonomic treatments have profound implications for other fields, ranging from studies in evolution and ecology to conservation biology (Wiens 2007). For example, conservation priorities that focused on a widespread and abundant lineage left a separate endangered lineage of the Preble's meadow jumping mouse unprotected (Malaney & Cook 2013). Likewise, whether ecological isolation is viewed as important in speciation can be dependent upon how species are delimited (e.g., Smith et al. 2013),

and differences in taxonomic practices can bias patterns of species diversity, making comparisons among taxa difficult (Isaac et al. 2004). However, species delimitation is itself a challenging endeavor. The actual biological properties used in identifying species have evolved with shifts in the availability of different data types and with recognition of the limitations of different species concepts. As a consequence, putative taxa may differ depending on not only the data used in analyses, but also the criteria used to delimit taxa (e.g., Edwards & Knowles 2014). For example, multiple species might be recognized based on the monophyly of mitochondrial gene trees, whereas only one species may be delimited based on morphological differences (Hebert et al. 2004a; 20004b). The expansion of genetic data to multiple loci has also highlighted the impact of decisions about data choice on inferences about species boundaries. For example, unrecognized taxa might be detected from consideration of multiple loci, even when there has not been enough time for the random sorting of gene lineages to produce monophyletic gene trees at any individual locus (Carstens & Knowles 2007). In other words, even for a given species concept (e.g., that species are independently evolving lineages; de Queiroz 2007), a critical question remains—how do we practically identify species, and what information do we need to consider to confidently delimit different species?

This quagmire is in many ways exemplified by the use of subspecies designations in systematic study. Subspecies conventionally have been used to denote geographic forms of polytypic species under the biological species concept (Mayr 1942; Mayr 1963), perhaps with additional evidence provided by a region's geological history (Mayr 1942).

Yet many subspecies delimitations more often than not appear to be arbitrary (Wilson & Brown 1953; Mayr & Ashlock 1991). This is especially evident when we consider that the use of species versus subspecies designations not only differs across taxonomic groups (Table 2.1), but may also show pronounced differences within taxa depending upon the region where the described taxa were collected. Such differences are highlighted in the charismatic Hercules beetles (genus *Dynastes* MacLeay, 1819), which are the focus of this study, where species versus subspecies designations appear to reflect differences in systematic/taxonomic practices, rather than differences in actual biological entities being delimited (Fig. 2.1). Specifically, the Hercules beetles are composed of two major groups: (i) the Giant Hercules group (which refers to its large body size of up to 15cm), which includes multiple subspecies within a single species *D. hercules* distributed across the Neotropics, and (ii) the White Hercules group (which refers to a general whitish body coloration), which has five recognized North American species (Ratcliffe 2003; Moron 2009). In both the Giant and White Hercules groups, species and subspecies delimitations are based on differences in elytral coloration and horn shape in adult males (Chalumeau & Reid 2002; Ratcliffe 2003; Moron 2009), given a general lack of variation in genitalic structures.

Table 2.1. Summary of different usages of species versus subspecies designations among animal clades.

	#Species	#Subspecies	#Subspecies/Species
Birds	11312	18969	2.677
Mammals	4811	6704	2.393
Reptile	10486	4325	1.412
Insects	622424	70515	1.113
Fishes	30015	513	1.017

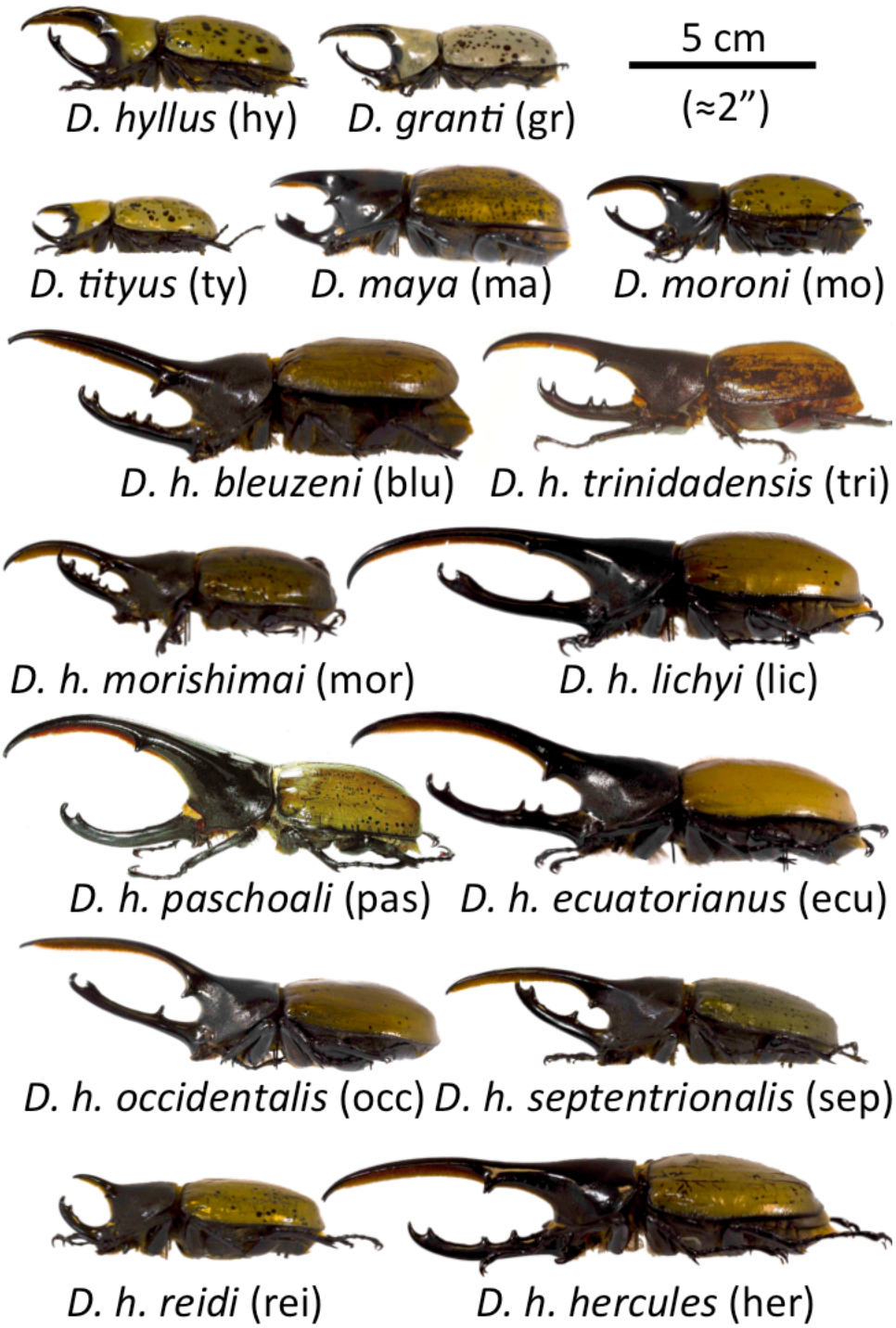

Fig. 2.1. Representative major males of each of the white Hercules (upper five images) and Giant Hercules (lower ten images) beetle taxa with the corresponding abbreviations used throughout the text and other tables and figures. Images of *D. h. paschoali* and *D. h. trinidadensis* courtesy of Jonathan Lai (2008).

In this chapter I evaluate whether the statistical support for taxonomic distinctiveness differs between species and subspecies of Hercules beetles as an example of how such analyses can be used to reveal whether different taxonomic units are justified biologically. Moreover, unlike all the other model-based statistical approaches for species delimitation that rely exclusively on genetic data to identify independently evolving evolutionary lineages (e.g., Carstens & Knowles 2007; Yang & Rannala 2010; Fujita et al. 2012), I apply an approach that integrates expectations for genetic distinctiveness derived from the coalescent with those for patterns of morphological divergence under a Brownian motion model of continuous quantitative traits (Solis-Lemus et al. 2015). Although quantitative methods for species delimitation have become dominated by an exclusive reliance upon genetic data in recent systematic studies (Fujita et al. 2012; Carstens et al. 2013), consideration of multiple data types can increase the accuracy of inferences (Dayrat 2005), especially under divergence scenarios where expectations for species distinctiveness under the coalescent can be compromised, such as when divergence occurs with gene flow or when selectively driven differentiation outpaces the resolution of neutral markers (Solis-Lemus et al. 2015). By accommodating multiple data types, the newly developed iBPP method (Solis-Lemus et al. 2015) also provides a context for delimiting species when divergence proceeds along different axes (Nosil et al. 2009) without sacrificing the benefits of model-based approaches – genetic and phenotypic data can be analyzed in a common Bayesian framework with iBPP. In additional to evaluating whether divergence between species versus subspecies of

Hercules beetles is statistically equivalent, I also consider the extent to which taxa within the Giant Hercules and White Hercules groups represent different levels of divergence along the speciation continuum (de Quiroz 2007). Specifically, I assess divergence patterns along different possible axes of differentiation (e.g., ecological, morphological, and geographical isolations; Nosil et al. 2009).

Although there has been some advocacy for seeking consensus across multiple genetic-based delimitation methods (Carstens et al. 2013), this approach is not adopted here because such a criterion imposes a very conservative, and in my view, inappropriate rationale for species delimitation. Both the power and sensitivities to violations of underlying model assumptions are known to differ among methods (Solis-Lemus et al. 2015). Seeking consensus therefore comes at the expense of decreasing our ability to recognize taxa. Not only can divergence proceed by different processes (e.g, with or without gene flow, and perhaps via selection, which are modes of divergence that are not accommodated by any of the methods reviewed in Carstens et al. 2013), but because recent speciation is associated with low levels of genetic distinctiveness, taxa may also go undetected by methods with limited power to distinguish taxa until they have accumulated appreciable levels of genetic differences (Hudson & Coyne 2002; Knowles & Carstens 2007).

2.3 Materials and Methods

Sample collection and DNA sequencing

Specimens were field collected, with a subset of tissues obtained from local collectors or insect dealers (Table 2.2). A total of 173 specimens were studied across five White Hercules and ten Giant Hercules taxa (Fig. 2.1 & Table 2.2), corresponding to an average of about 10 specimens per putative taxon. Three samples from two species of the subgenus *Theogenes* (genus *Dynastes*; Hwang 2011) were included as outgroups. Samples are vouchered in the Insect division of the Museum of Zoology at the University of Michigan (UMMZ).

Table 2.2. Samples used for molecular analyses.

Group	Scientific name (abr.)	Collection sites (sample size)
White	*D. granti* (gr)	Arizona (2)[2], Star Valley, Payson, Arizona (15), Reserve, New Mexico (9)
	D. hyllus (hy)	Mexico (3)[2], El Palmito, Sinaloa, Mexico (1)[1], Chiapas, Mexico (1)[1], Poterillo, Sinaloa, Mexico (2)[1], Puebla, Mexico (2)[2]
	D. maya (ma)	Honduras (4)[1], El Cosuco, Cortez, Honduras (1)[2]
	D. moroni (mo)	Volcan San Martin, Sierra de Los Tuxtlas, Veracruz, Mexico (4)[1], Sierra de Los Tuxtlas, Veracruz, Mexico (2)[2]
	D. tityus (ty)	North Carolina (2)[2], Fort White, Florida (2), Lexington, Fayette, Kentucky (1), Chase, Maryland (1), Franklin, North Carolina (2), Montgomery,

		Tennessee (1), Rock Island, Tennessee (2), Suburb Atlanta, Georgia (1)
Giant (*D. hercules*)	*D. h. occidentalis* (occ)	Rio Chuchuvi, Esmeraldas, Ecuador (4), Los Bancos, Pichincha, Ecuador (5)[1]
	D. h. septentrionalis (sep)	Near Chiripo, Costa Rica (1)[1], Finca La Firmeza, Sierra de Coral, Guatemala (4)[1], Santa Barbara, Honduras (3)[1], El Cosuco, Cortez, Honduras (3)[1], Cerro Azul, Panama (1)[1]
	D. h. paschoali (pas)	Bahia, Brazil (4)[2]
	D. h. reidi (rei)	Martinique (1), Soufriere, Saint Lucia (2)
	D. h. hercules (her)	Dominique (3), Saint-Claude, Guadeloupe (3), Guadeloupe (1)[2]
	D. h. lichyi (lic)	Cosanga, Napo, Ecuador (8), Cosanga, Napo, Ecuador (1)[1], Cayagama, Sucumbios, Ecuador (3), La Bellela, Santander, Colombia (2)[1], Selva Central, Peru (1)[1], Satipo, Junin, Peru (4)[1], El Reventador, Ecuador (3), La Bonita, Sucumbios, Ecuador (2), Surrounding mountains/hills near Lumbaqui, Ecuador (2), Santa Clara, Ecuador (4), Near Sumaco Park Entrance, Ecuador (6), San Pablo (45 minutes drive N. from Tena), Ecuador (6), Cabañas San Isidro, Cosanga, Ecuador (1), Yanayacu Station, Ecuador (1)
	D. h. trinidadensis (tri)	Morne Bleu, Trinidad (6), Trinidad (2)[1]
	D. h. bleuzeni (blu)	Venezuela (1)[2], Cerro Sarisariñama, Bolivar, Venezuela (1)[1]
	D. h. ecuatorianus (ecu)	Colombia (3)[1], Misahualli, Napo, Ecuador (1)[1], Misahualli, Napo, Ecuador (5), Iquitos, Loreto, Peru (5)[1], Oso G Petrolane station, Loreto, Ecuador (1), Pompeya, Orellana, Ecuador (1), Yasuni Station, Ecuador (2)
	D. h. morishimai (mor)	Bolivia (2)[2]
Outgroup	*D. neptunus* (Dn)	Cosanga, Napo, Ecuador (1), Los Bancos, Pichincha,

		Ecuador (2)
Outgroup	*D. satanas* (Ds)	Bolivia (1)[2]

[1]Samples obtained from online insect specimen dealers

[2]Samples from pet stores

Tissues were preserved in 100% EtOH and stored at -80°C at the Museum of Zoology, University of Michigan. Genomic DNAs were extracted from either thoracic or leg muscles using the DNeasy Blood & Tissue kit (Qiagen, Germany). One mitochondrial and four nuclear regions were amplified: COI, *argK*, *cad*, *h3*, and *its1* (Folmer et al. 1994; Richards et al. 1997; Colgan et al. 1998). I designed four *Dynastes* specific primers for *argK* and *cad*: DyAKF: 5'- GACATCCACCAAAGAACTGGGGC -3', DyAKR: 5'- CCTTGTTGGAGGCTAATTTGGGC -3', DyCDF: 5'- GCCGTTGGTCCCGGAATATGTAG -3', DyCDR: 5'- GCTGGGTTCGAAGCAAGCTGTTG -3'. Each reaction contained 0.5µl of extracted DNA, 1µl of 10x buffer, 0.75µl of MgCl2, 0.5µl of 10mM dNTPS, 0.2µl of 1% BSA, 0.4µl of each primer, 0.04µl of Tag DNA polymerase (Invitrogen, USA), and ddH2O to make a total of 10µl reaction. A standard PCR profile with one-minute duration for each step, a total of 35 cycles, and a finally extension of 10 minutes at 72°C was followed. The annealing temperatures for COI, *argK*, *cad*, *h3*, and *its1* were 52°C, 58°C, 60°C, 60°C, and 52°C, respectively. PCR products were sequenced on a ABI^{3700} sequencer by the Sequencing Core, University of Michigan. Sequences were checked using SeqMan (DNAStar Inc., USA) and edited sequences were imported into MegAlign (DNAStar Inc., USA) for multiple sequence alignment using clustalW. Aligned sequences were

converted into both NEXUS and PHYLIP formats using Mesquite (Maddison & Maddison 2011).

Morphological divergence

The shape of the cephalic and thoracic horns was characterized using landmark-based geometric morphometrics (Fig. 2.2; digital images for the specimens are available from Dryad: doi: 10.5061/dryad.8p6m0). Note that differences in horn shape and elytra coloration are typically used to distinguish the taxa. However, elytral coloration was not analyzed here because it is not only sensitive to the background humidity when specimens are dried, but also how the specimens were preserved (dried versus in alcohol). Additionally, specimens from historical collections often have lighter body coloration, which may reflect how the specimens were initially preserved and/or degradation of pigments.

Landmarks were identified from digital images and analyzed using the R package geomorph (Adams et al. 2013). Specifically, generalized Procrustes analysis was performed to remove the effects of location, size, and rotation of the relative positions of landmarks among specimens. This superimposition method minimizes the sum-of-squared distances between landmarks across samples (Rohlf & Slice 1990). The residuals from the mean shape of the cephalic and thoracic horns were used for analyses of shape differences among the taxa. Specifically, an analysis of variance (ANOVA) was

conducted based on the Euclidean distances separating taxa. Pairwise comparisons were used as post hoc tests to investigate if significantly different horn shapes are found between pairs of taxa. For both ANOVA and pairwise analyses, I used 999 permutations to test for statistical significance.

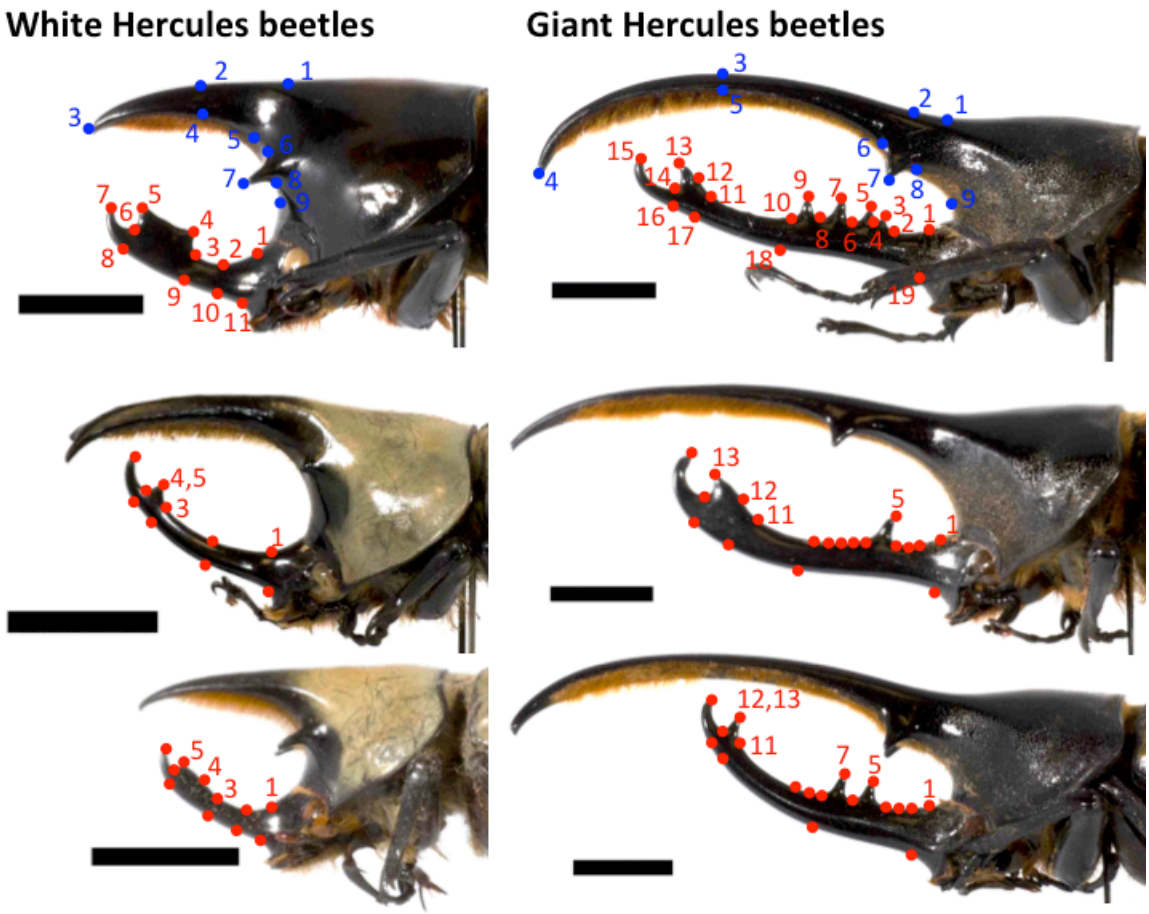

Fig. 2.2. Landmarks used for thoracic (blue) and cephalic (red) horns, where the numbers refer to homologous points that within the three distinct types of cephalic horns of the White and the Giant Hercules beetles (shape differences are not compared between species groups given the obvious distinctiveness of the two groups; the numbered landmarks differ between the White and the Giant Hercules taxa).

The data from the geometric morphometric analyses were also used in joint

Bayesian inference of species boundaries with the genetic data in the program iBPP (Solis-Lemus et al. 2015). Specifically, I extracted the values from PC1 score from the Geometric Morphometric analyses for both cephalic and thoracic horns, and used these values as input for the iBPP analyses.

Species delimitation

Species trees obtained from *BEAST version 2.0.2 (Drummond et al. 2012) were used as the guide tree for BPP (Yang & Rannala 2010) and iBPP (Solis-Lemus et al. 2015) analyses discussed below, with traditional taxonomic criteria used to assign individuals to putative taxa (e.g., species designations were based on distinguishing morphological characters and geographic information about taxonomic distributions based on Chalumeau & Reid 2002, Hwang 2011, and Ratcliffe et al. 2013; see Yeates et al. 2010). The species tree was estimated from $1x10^9$ generations with species trees sampled every 1×10^5 generations and parameters sampled every 1×10^4 generations. The first 20% of trees were discarded as burnin and the remaining trees were imported into TreeAnnotator to reconstruct a Majority Clade Credibility (MCC) tree. A Yule speciation model with a linear and constant root population model was specified. A mutation rate of 0.0115 per site per million years was set for COI (following Brower 1994), with other locus-specific rates estimated relative to COI, to estimate divergence times, and models of nucleotide evolution were identified for each locus using the Akaike Information Criterion in jModelTest 2 (Darriba et al. 2012). Heterozygous sites were kept

in the analyses; the IUPAC nucleotide ambiguity code was used for heterozygous sites.

The program BPP (version 2.2, Yang & Rannala 2010) was used to delimit taxa using genetic data only. BPP analyses were conducted for Giant and White Hercules beetle clades separately to avoid sensitivities to potentially different demographic histories between the clades that can affect the program's performance (see Zhang et al. 2011). For example, the opportunities for gene flow differ given the allopatric distribution among White Hercules beetle taxa compared to the parapatry found among certain taxa of Giant Hercules beetles. Because different combinations of parameter values can produce similar levels of discord among loci, four different demographic scenarios corresponding to relatively large (or small) ancestral population sizes with relatively deep (or shallow) divergence times were used to estimate the support of putative taxa. Specifically, the parameter setting used were: $\theta = G(1, 10)$ and $\tau = G(1, 10)$, $\theta = G(1, 10)$ and $\tau = G(2, 2000)$, $\theta = G(2, 2000)$ and $\tau = G(1, 10)$, and $\theta = G(2, 2000)$ and $\tau = G(2, 2000)$, where θ and τ refers to the ancestral population sizes and divergence times, respectively, and G specifies a gamma distribution for the prior. Locus-specific rates of evolution were estimated and default settings for parameter tuning and scaling were used in the analyses. The MCMC chains for these (and the analyses described below with iBPP) were run for 5×10^5 generations with parameters sampled every five generations and a burnin period of 5×10^4 generations. Likewise, both algorithms 0 and 1 (tuning parameter of 1 and 5) of reversible-jump MCMC searches (rjMCMC) were used to determine if the results were robust to different searching algorithms. Because the

results were not significantly different, only the results from the runs with the algorithm 0 setting are presented.

Because the relationships among Giant Hercules beetles are not well resolved from the *BEAST analysis (see result section), which could affect the taxa delimited with BPP (McKay et al. 2013; but see Zhang et al. 2014), uncertainty in the species-tree estimate was considered by integrating across possible species trees when evaluating the probability of different putative taxa (see BPP version 3, Yang & Rannala 2014). I employed both algorithm 0 and 1 for rjMCMC searches with the same set of four different demographic scenarios used in the previous section. A burnin period of 5×10^4 generations was used and each of the MCMC searches were run for 5×10^5 generations with trees sampled every 5 generations. The probabilities of different numbers of delimited species and the frequency of identified putative taxa were calculated.

A joint Bayesian inference based on genetic and phenotypic data was also used to delimit species using the newly developed program iBPP (Solis-Lemus et al. 2015). This integrative program is an expansion of the original program BPP (hence, the name iBPP), and differs in that it incorporates models of evolution for continuous quantitative traits under a Brownian motion process (Solis-Lemus et al. 2015). As with the analyses based exclusively on the genetic data, I confirmed the robustness of the results to different algorithms (e.g., we analyzed the data with both the fine tune setting of 0 and 1, with the algorithm 0 for rjMCMC searches). The λ-value for each continuous trait and locus-specific rates of evolution were estimated for each trait and locus, respectively;

note that the λ-value can differ among phenotypic traits, thereby accommodating different rates of evolution among continuous quantitative traits (Solis-Lemus et al. 2015). The same four different demographic settings corresponding to relatively large (or small) ancestral population sizes with relatively deep (or shallow) divergence times described above for the analyses of the genetic data alone were also used in independent runs of iBPP based on the combined genetic and phenotypic data. I also conducted an analysis based just on the phenotypic data with similar run times and settings as those used for the joint inference based on genetic and phenotypic data. Note that given the role of the beetle horns in male fighting, the horns may not have evolved according to the assumptions of a Brownian motion model; however, see Solis-Lemus et al. (2015) for a discussion of the robustness of results from iBPP when characters were simulated under non-Brownian motion models.

As with the analyses of the genetic data alone, I also accounted for uncertainty in the species relationships among the Giant Hercules beetles in iBPP analyses. Specifically, all species tree topologies among the 8,000 post-burnin *BEAST species trees, and their probabilities of being observed, were summarized using the trprobs function from the SumTrees program of the DendroPy package (Sukumaran & Holder 2010). A total of 1739 unique topologies were imported as guide trees for 1739 independent iBPP analyses each with a pre-burnin of 1×10^5 generations. The probability of a specific topology and the probability of different delimited taxa were calculated across the 1×10^6 retained post-burnin samples.

Tests of niche divergence and assessments of geographic overlap

Species occurrence data were obtained from my field collections and local researcher Hector Gasca for taxa from Colombia, published records (Grossi & Arnaud 1993; Ratcliffe & Cove 2006; Ratcliffe et al. 2013), and the GBIF database for *D. hercules septentrionalis* (a complete list can be found from Dryad doi:10.5061/dryad.8p6m0). Ecological niche models (ENMs) were projected for each taxon from selected bioclimatic layers (see below) at a spatial resolution of 2.5 arc minutes for the present (WorldClim database: http://www.worldclim.org/) using MAXENT (version 3.3.3k, Phillips 2006) with the following parameters: regularization multiplier = 1, max number of background points = 10,000, replicates = 20, replicated run type = cross-validate. I used a spatial scale of 2.5 arc minutes instead of a finer scale (e.g., 30 arc seconds) to accommodate spatial uncertainty in some historical records (e.g., collection records reported at a spatial scale of a town or village). To generate a predicted distribution during the last glacial maximum (LGM), paleoclimatic data for the same selected bioclimatic layers were used (Paleoclimate Modelling Intercomparison Project Phase II: Community Climate System Model [CCSM]). Jackknife analyses for testing the robustness of models and area under the ROC curve (AUC) was used to investigate

model performance for each ENM. Note that ENMs were not estimated for all subspecies of the Giant Hercules because of a scarcity of collection records (i.e., no ENMs were estimated for blu, her, mor, rei, and tri; see Fig. 2.1 for full subspecies names).

To guard against the inherent difficulties involved in extrapolating distributions into novel climates (reviewed in Alvarado-Serrano & Knowles 2014), an iterative approach was used to generate ENMs for the LGM (see also Knowles et al. 2015). Specifically, multivariate environmental similarity surfaces (MESS maps) were used to identify which of the 19 bioclimatic variables resulted in areas of low-reliability predictions due to the variables being outside of the range present in the present-day environmental data (Elith et al. 2010). MAXENT was rerun excluding these out of range variables, and this process of analysis with MESS maps was repeated until no LGM variables were out-of-range compared to present-day bioclimatic variables. Additionally, a present-day ENM was generated using the subset of variables with greater than 10% importance (determined by jackknifing) across at least five different taxa (greater than 5% for all taxa included for ENMs). Based on these analyses, a reduced number of variables were used to generate present-day and LGM ENMs (specifically, the 5 bioclimatic variables: annual mean temperature, Bio1, isothermality,Bio3, temperature seasonality, Bio4, mean temperature of the coldest quarter, Bio11, and precipitation of the warmest quarter, Bio18). Note that it is possible that the procedures used here based on statistical rational, while guarding against inaccuracies with the modeling (see Alvarado-Serrano & Knowles 2014), might nonetheless mask aspects of the species'

ecologies relevant to identifying suitable geographic areas.

Geographic overlap between parapatric taxa was assessed using the threshold of a 10% training presence (average across 20 replicates) to create binary maps calculated in ENMTools (Warren et al 2010). A 10% training presence threshold was used, which always yielded the highest cutoff value. Note that geographic overlap in both the present and past (i.e., based on the ENM for the LGM) was used because I wanted to assess whether taxa that are distributed allopatrically today might have overlapped in the past. Genetic distinctiveness among taxa that either overlap today (or in the past) represents one axis of the speciation continuum that is relevant for evaluating whether species and subspecies designations are biologically justified (e.g., species overlap in distribution, whereas subspecies remain allopatric).

To measure the niche similarity between ENMs across taxa (Warren et al. 2008), background similarity tests (based on Index D; Warren et al. 2008) were used to statistically assess if the niches were significantly more or less or similar between taxa than expected based on geographic aspects of their ranges (i.e., whether despite allopatric distributions, the similarity/differences of the niches exceed expectations based on expectations from geography alone)(Warren et al. 2008; McCormack et al. 2010). Specifically, climatic variables associated with the predicted niche of one species were compared to 50 randomly selected points from the other species' geographic distribution estimated from the ENM; significance was assessed from 100 permutations using ENMTools. Tests were done between sister taxa as well as geographically proximate taxa;

geographically distant taxa are not presented because the similarity index for such comparisons were 0 or nearly 0. Note that predicted geographic areas of suitable habitat that were not actually part of a target taxon's known range because of barriers posed by environmentally inhospitable intervening geographic regions (Monon 2009; Ratcliffe et al. 2013) were not included in niche similarity tests. That is, the distribution map used for the background similarity test corresponds to the actual geographic range species occupy given barriers to dispersal (see Glor & Warren 2010).

2.4 Results

DNA sequence data

Of the 173 specimens, sequences were successfully generated for COI, *argK*, *cad*, *h3*, and *its1* regions in 173, 73, 114, 134, and 98 specimens respectively. Aligned sequences contain 659, 575, 789, 353, and 653 sites, respectively for each locus (GenBank accession # KT183708-184299). The numbers of parsimony-informative sites are 170, 14, 34, 8, and 32, respectively (Summarized using PAUP* ver. 4.0a146; Swofford 2002). Average uncorrected pairwise sequence divergence for White Hercules beetles ranged from 0.37% (± 0.10 SE) within putative taxa to 1.12% (± 0.19 SE) between putative taxa; average pairwise sequence divergence for Giant Hercules beetles ranged from 0.43% (± 0.13 SE) within putative taxa to 1.06% (± 0.08 SE) between

putative taxa (Fig. 2.3). Note that with multiple individuals sequenced per putative taxa (Knowles 2010), and even with low levels of variation (Lanier et al. 2014), accurate species tree can be estimated without large numbers of loci (McCormack et al. 2009), especially when uncertainty in estimates of the gene trees of individual loci is taken into account (Huang et al. 2010).

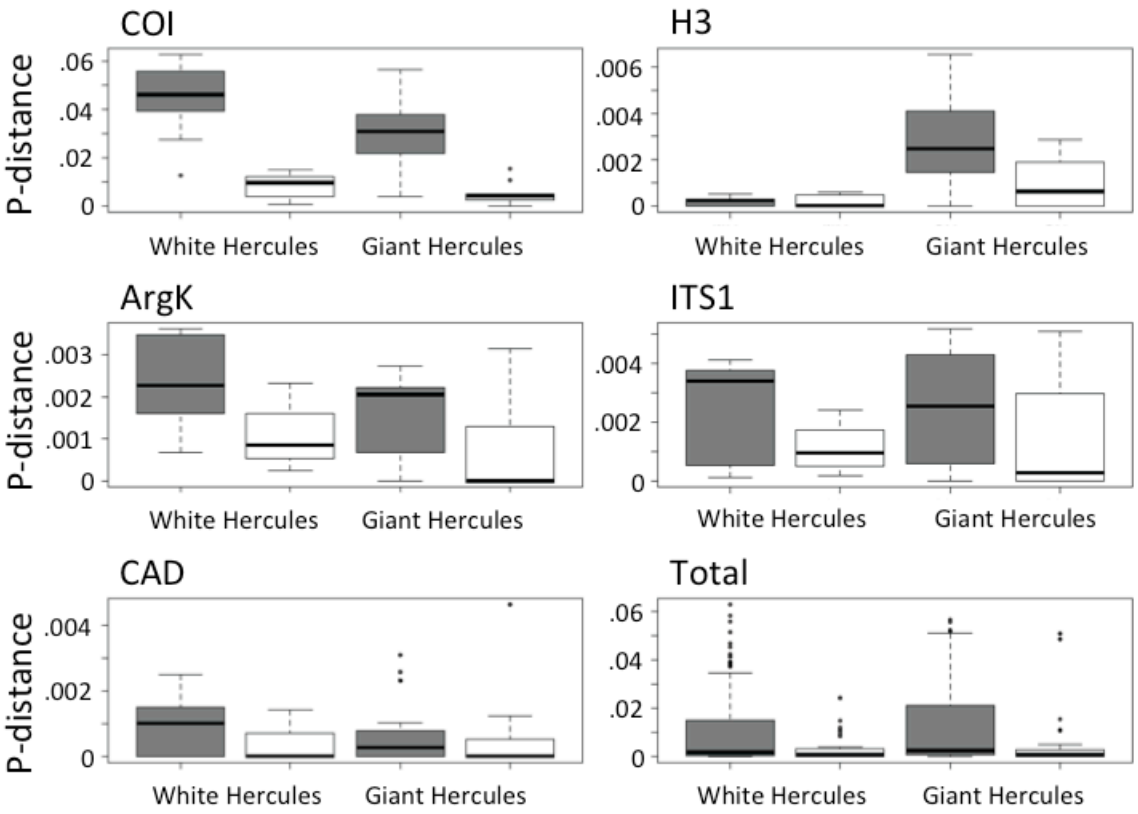

Fig. 2.3. Pairwise sequence divergence (K2P distances) between (grey) and within (white) putative taxa of White and Giant Hercules beetles.

Geographic overlap and niche similarity

Geographic overlap differs among taxa and between the White and Giant Hercules groups (Fig. 2.4). None of the taxa within the White Hercules group overlap. Several taxa within the Giant Hercules group, in contrast, have overlapping distributions (i.e., those distributed throughout the Andes and the adjoining region of the Amazon; Fig. 2.4).

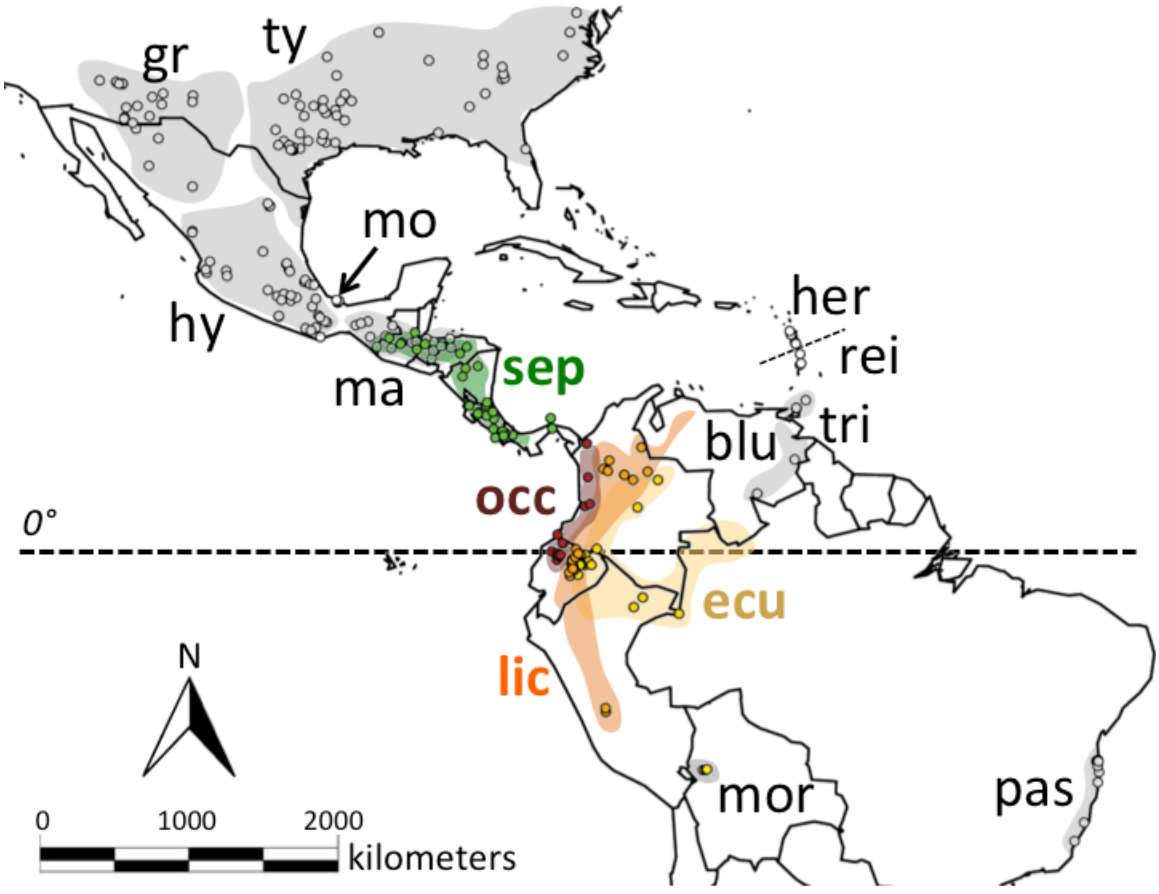

Fig. 2.4. Projected distributions that contrasts the allopatry of all White Hercules taxa (i.e., gr, ty, hy, mo, and ma) to the parapatrically distributed taxa within the Giant Hercules group of South and Central America (i.e., names in bold and distributions shown in color: lic, ecu, sep, and occ) and the island endemics or allopatric Giant Hercules taxa (blu, tre, her, rei, mor, and pas). Circles indicate actual collection records used to estimate

distributions of each taxon using the program MAXENT (see methods). Abbreviations and their corresponding taxonomic names can be found in Fig. 2.1.

In contrast to the current allopatric distributions of taxa within the White Hercules group, ENMs based on climatic conditions during the LGM suggests that some taxa may have come into contact. However, the degree of overlap varied among species (e.g., projected geographic overlap of 0.4 % between gr and hy, 34.9% between hy and ma, 8.3% between hy and mo, and 37.4%b between ma and mo). Within the Giant Hercules group, in addition to the present geographic overlap among taxa (i.e., among occ, lic, and ecu; see Fig. 2.4), overlap during the LGM is predicted as well (i.e., 54.4% between ecu and lic, 61.3% between lic and occ, and 5.7% between occ and sep).

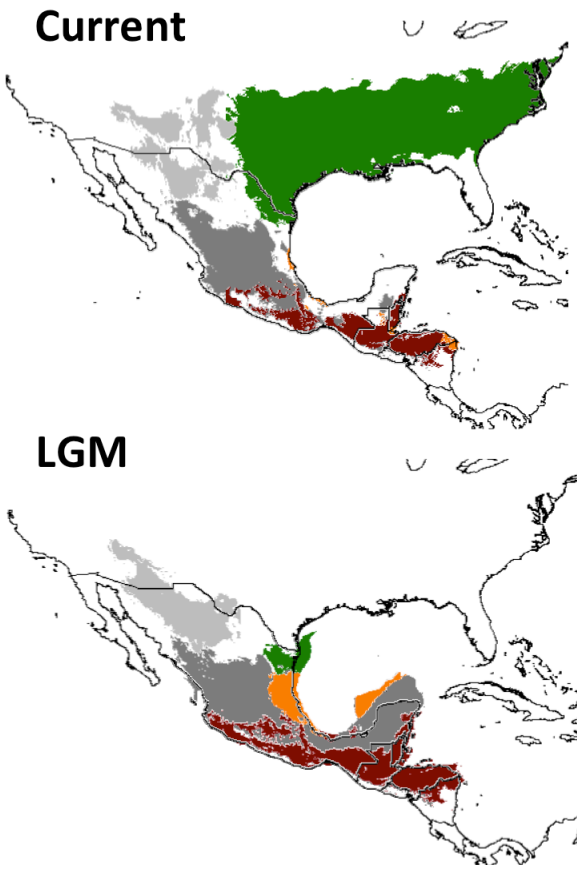

Fig. 2.5. Comparison of predicted current and past distributions for the White Hercules taxa. Light grey: gr, dark grey: hy, green: ty, orange: mo, and dark red: ma.

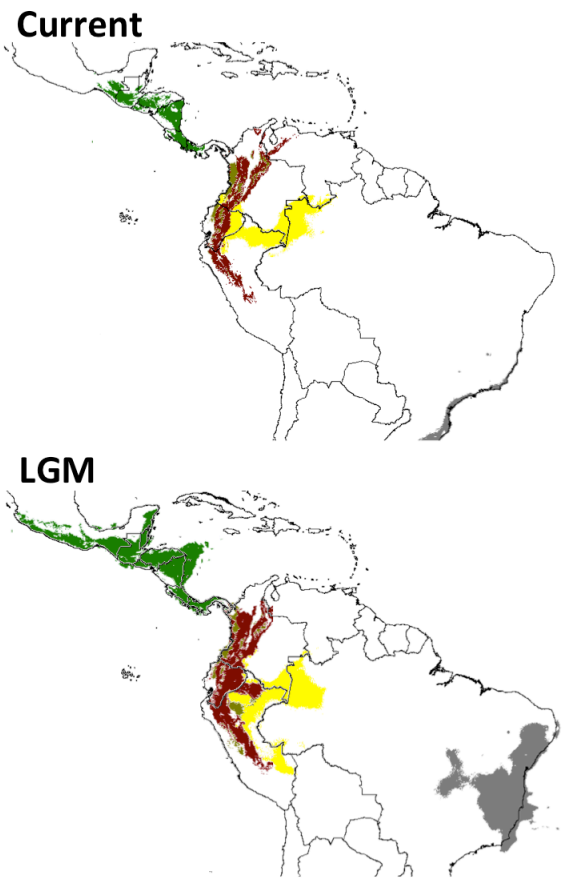

Fig. 2.6. Comparison of predicted current and past distributions for the Giant Hercules taxa. Green: sep, light green: occ, dark red: lic, yellow: ecu, and dark grey: pas. mor and blu, and island endemic her, rei, and tri are excluded from this map.

The observed niche similarity among most taxa (Fig. 2.7) did not differ any more than expected based on their geographic distribution, or it supported niche conservatism, irrespective of whether putative sister taxa of White or Giant Hercules beetles were compared. In two cases there was marginal evidence for significant niche divergence (i.e., occ and lic, and ty and ma). In both cases, significant divergence was only detectable when observed similarity indices were compared with random samples of environmental variables from one of the two backgrounds (e.g., niche divergence is apparent between

occ and lic when the environmental data at collection localities of lic are compared to those from random samples across the projected distribution of occ, but not when environmental data at collection localities of occ is compared to those from random samples across the projected distribution of lic; Fig. 2.7).

As with any study based on ENMs, the projected distributions, and hence, tests of niche overlap, depend on the extent to which the models accurately capture the variables relevant to determining a species' distribution. For example, different models might capture aspects of the species' ecologies that might impact a species' past distribution. For the models used here, statistical criteria were considered to minimize errors (e.g., removing non-analog sets of climate conditions for LGM modeling; see details in methods) and for consistency across taxa (e.g., non-analog climatic conditions were not used in the ENMs for any taxa), which is critical given that comparisons of taxa are integral for assessing whether taxa in the White and Giant groups are consistently more of less advanced along the speciation continuum, where geographic overlap of distributions (either in the present or past), as opposed to strict allopatry, would provide evidence that taxa have achieved sufficient reproductive isolation to maintain genetic distinctiveness.

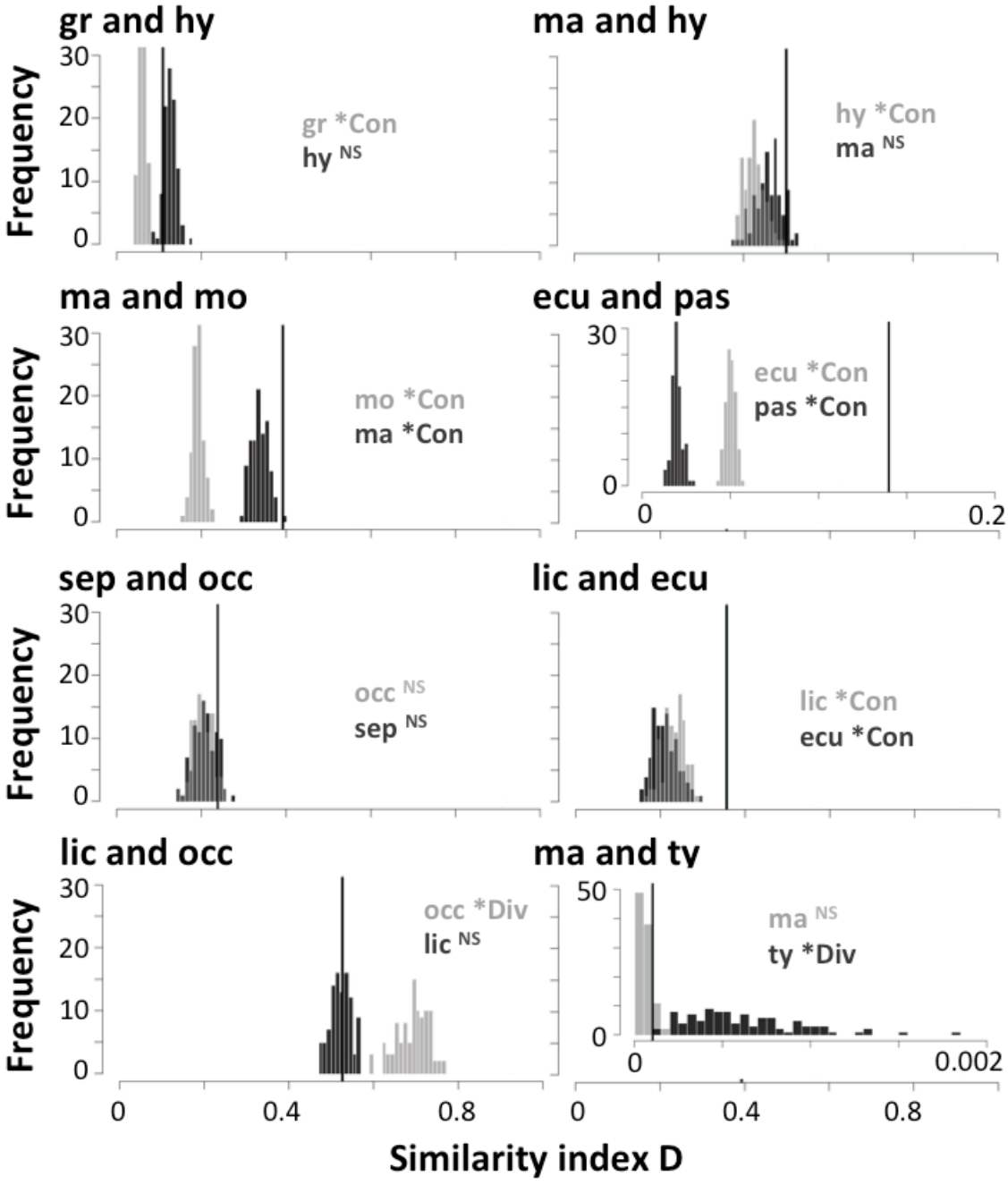

Fig. 2.7. Results from niche background similarity tests. In most taxa the observed niche similarity, calculated by the index D (and shown as the thin vertical line), did not differ among putative sister taxa (see Fig.) any more than expected from the difference in niche similarity that can be attributed to differences in the geographic distribution of taxa, or tended towards supporting niche conservatism, based on a similarity index that was higher than expected (where the histograms show the expected frequency distribution of

niche similarity scores based on random samples of background points across the projected distribution of the taxon involved in the pairwise comparison). Only two cases of significant niche divergence were identified (marked as *Div). In both cases, significant divergence was only evident when the observed niche similarity was compared against one of the two backgrounds (e.g., niche divergence is apparent between occ and lic when the environmental data at collection localities of lic are compared to those from random samples across the projected distribution of occ, but not when environmental data at collection localities of occ is compared to those from random samples across the projected distribution of lic). Also note that the lic and occ comparison represents divergence between highland Andean cloud forest and the Choco ecoregion, and the other case of niche divergence involves comparison of taxa between subtropical Southeast United States and neotropical Central America (i.e., ty and ma). Abbreviations and the corresponding taxonomic names are listed in Fig. 2.1.

Morphological differences

Taxa within both the White and Giant Hercules groups exhibit significant morphological differences in both cephalic and thoracic horn shapes (Fig. 4; $P < 0.01$). Even though individuals from the same taxon cluster together in morphospace, there is some overlap between most taxa in both groups (Fig. 2.8), as is also evident in pairwise comparison of mean morphological Euclidean distances among taxa (Tables 2.3 and 2.4).

Table 2.3. Mean morphological Euclidean distance between White Hercules taxa.

Taxa	gr	hy	ma	mo	ty
gr		0.046	0.375**	0.075	0.265**
hy	0.094		0.374**	0.059	0.254**
ma	0.189**	0.136*		0.343**	0.208**
mo	0.078	0.066	0.165*		0.219**
ty	0.226**	0.177**	0.097	0.218**	

Results for the cephalic horn and thoracic horn shapes are shown above and below the diagonal, respectively.

* means $P < 0.05$ and ** means $P < 0.01$ based on 999 permutations.

Table 2.4. Mean morphological Euclidean distance between White Hercules taxa.

Taxa	blu	ecu	her	lic	mor	occ	pas	sep	tri
blu		0.069	0.114	0.119*	0.121	0.102	0.088	0.117	0.099
ecu	0.092		0.087*	0.118**	0.090*	0.099**	0.112*	0.113**	0.081
her	0.140	0.206**		0.160**	0.104	0.141**	0.108	0.122**	0.098
lic	0.121	0.150**	0.102*		0.168**	0.056	0.142**	0.129**	0.152**
mor	0.072	0.088	0.148*	0.128*		0.140**	0.146*	0.095*	0.051
occ	0.166*	0.085*	0.288**	0.228**	0.161**		0.125*	0.100**	0.121**
pas	0.110	0.152*	0.093	0.051	0.109	0.230**		0.133*	0.118
sep	0.169*	0.100*	0.290**	0.241**	0.154*	0.054	0.233**		0.085
tri	0.071	0.116*	0.113	0.101	0.065	0.190**	0.074	0.185**	

Results for the cephalic horn and thoracic horn shapes are shown above and below the diagonal, respectively.

* means $P < 0.05$ and ** means $P < 0.01$ based on 999 permutations.

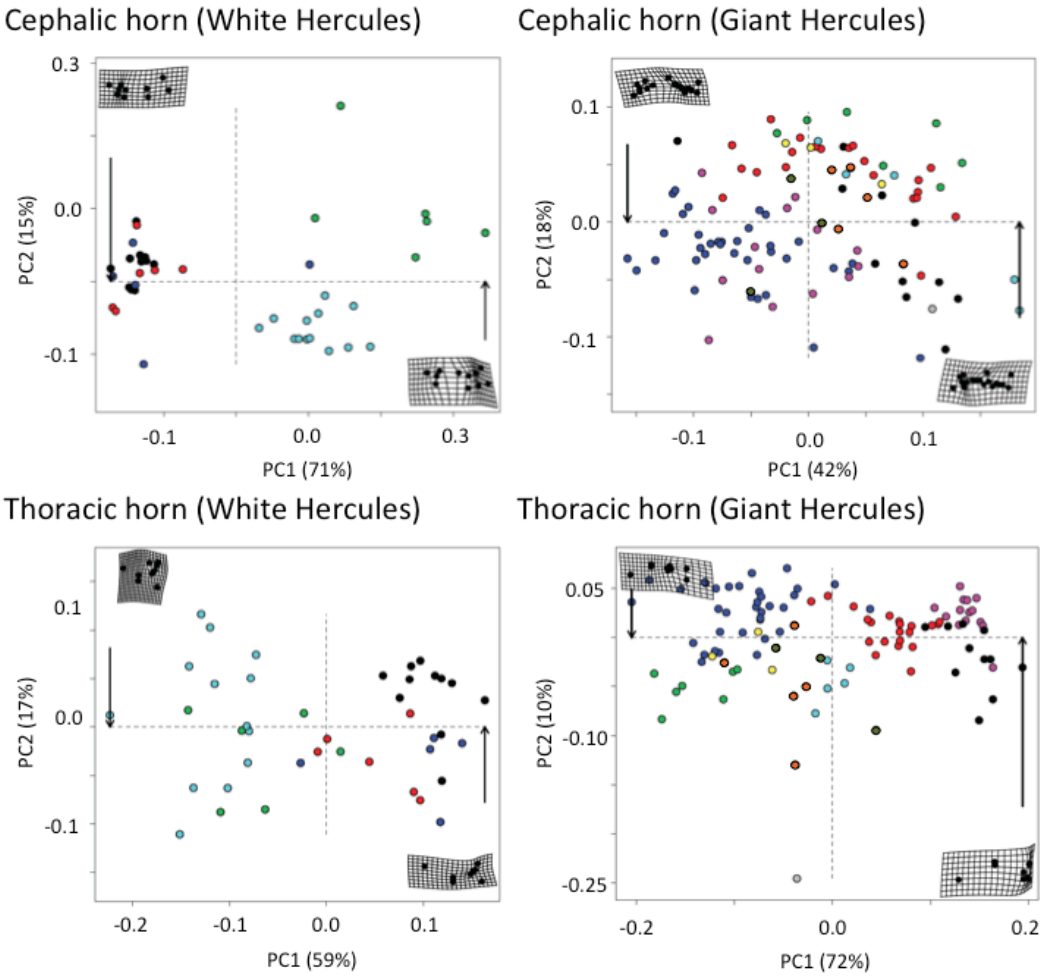

Fig. 2.8. Position of taxa in morphospace based on geometric morphometric analyses of the cephalic and thoracic male horn shape based on the first two PCs, which explained most of the differences in shape among taxa (i.e., between 60 and 82% of the variance). Landmarks in tangent space are shown for the most extreme shapes on upper left and lower right corners of each PCA. Permutation tests based on Euclidean distances between samples from PC space indicate significant differences in horn shape across taxa in all four analyses ($P < 0.01$). Different taxa of White Hercules beetles are shown with black = gr, red = hy, blue = mo, green = ma, and cyan = ty, and different taxa of Giant Hercules beetles (excluding rei because of its extremely divergent morphology) are shown with dark green = blu, red = ecu, blue = lic, green = her, cyan = mor, yellow = pas, magenta = occ, black = sep, and orange = tri. Taxonomic abbreviations and the species names can be found in Fig. 2.1.

Species tree

The *BEAST MCMC run reached a likelihood plateau soon after generation 10,000,000 and all estimated parameters had effective sample sizes (ESS) larger than 2,000. The estimated rates of evolution for *argK*, *cad*, *h3*, and *its1* are 0.00047 (0.00025 – 0.00072), 0.00149 (0.00089 – 0.00216), 0.00034 (0.00017 – 0.00053), and 0.00060 (0.00035 – 0.00086) per site per million years, respectively. Nodal support (Fig. 2.9) is strong for all taxa within the White Hercules group and among six of the taxa from the Giant Hercules group, whereas support is low among some of the non-allopatric Andean and Amazonian taxa and the allopatric coastal Brazilian taxa (i.e., lic, ecu, mor, pas; see Fig. 2.4). Note that low support values are not necessarily associated with the recency of diversification, nor are they strictly associated with parapatric taxa. For example, the sister relationship of the recently derived putative taxa blu and tri within the Giant Hercules group has strong support, as does the sister relationship of the putative taxa occ and sep, both of which overlap with other taxa (Fig. 2.9).

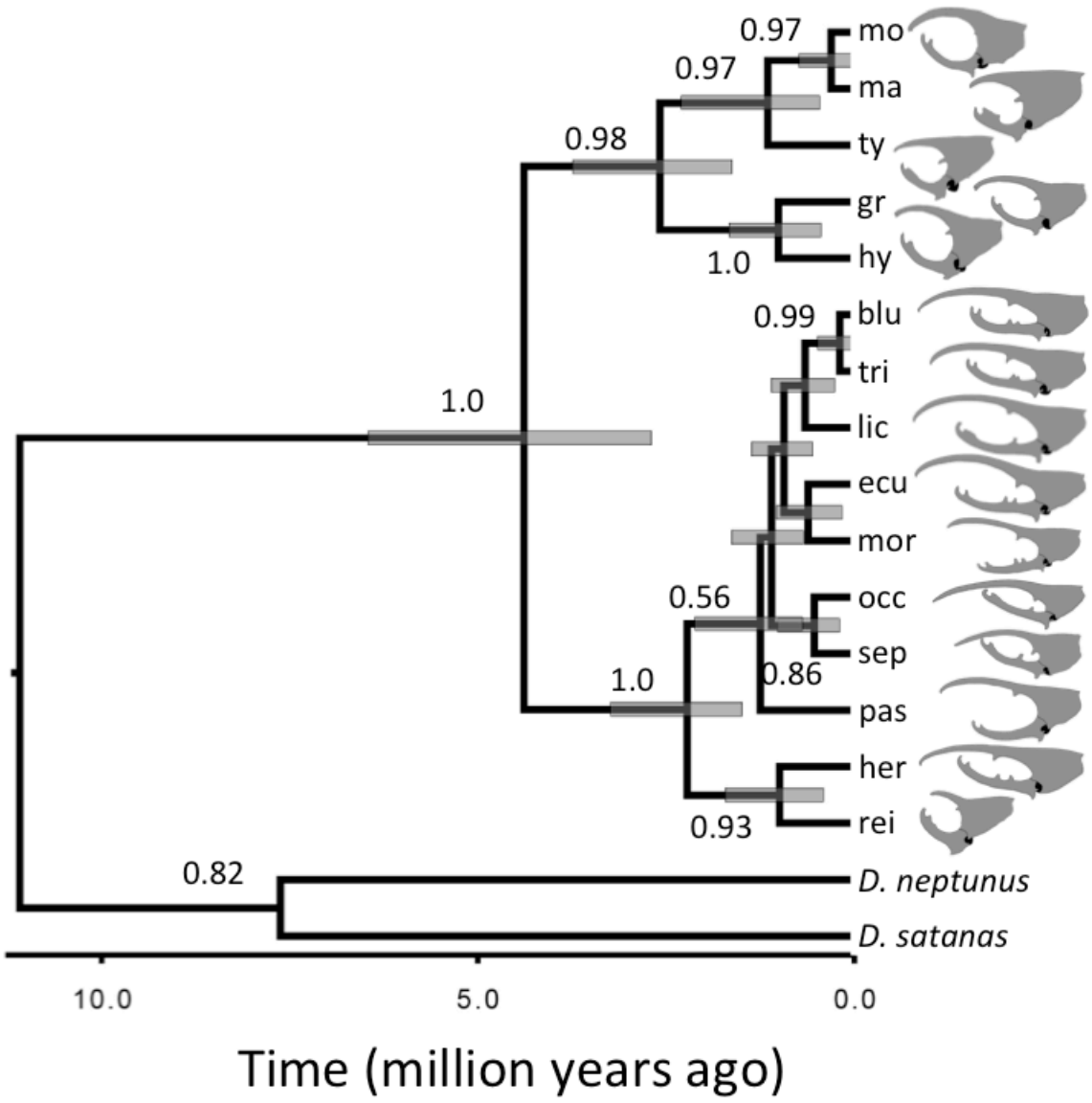

Fig. 2.9. Species tree of *Dynastes* estimated from *BEAST; posterior probabilities above 50% are shown on nodes. Divergence times with confidence intervals estimated using a rate of 0.0115/site/million years are shown at each node. Note that outlines of major male horn shape are not drawn to scale (see Fig. 2.1 for relative sizes).

Species delimitation

Irrespective of whether genetic data alone, phenotypic data alone, or genetic and

phenotypic data combined are analyzed, all the analyses agree on the number of taxa: 5 species of White Hercules and 10 species of Giant Hercules beetles (Figs. 2.10, 2.11, and 2.12; with the exception of the gr-hy divergence based on morphological data alone). The results are also robust to uncertainty in species-tree estimates for the Giant Hercules beetles (Table 2.5).

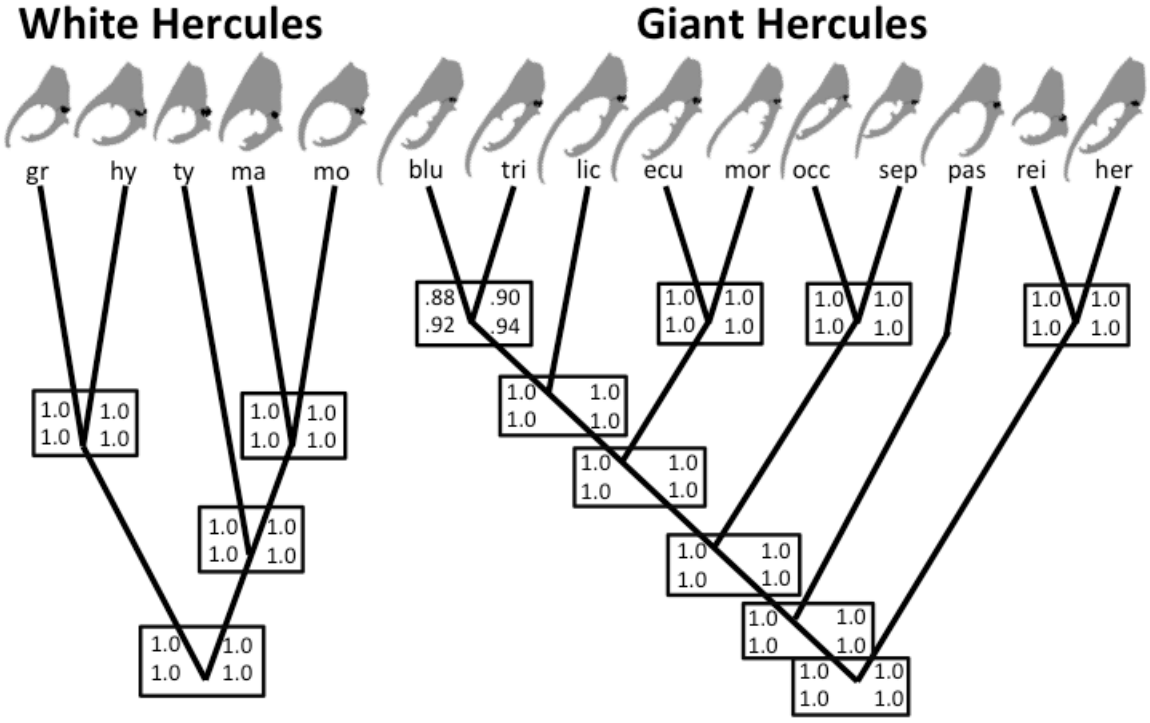

Fig. 2.10. Results from iBPP analyses using combined genetic and morphological datasets; note that outlines of major male horn shape shown adjacent to the abbreviated taxonomic names are not drawn to scale (see Fig. 2.1 for relative body sizes). Support values reported for each node are based on the algorithm setting 0 for the rjMCMC, the algorithm setting 1 for the Brownian motion model, and for four different priors corresponding to large (versus small) ancestral population sizes with relatively deep (versus shallow) divergence times. Specifically, the support values in each box correspond to analyses with the following different priors: upper left, $\theta = G(1, 10)$ and $\tau = G(1, 10)$; lower left, $\theta = G(1, 10)$ and $\tau = G(2, 2000)$; upper right, $\theta = G(2, 2000)$ and τ

= G(1, 10); lower right, θ = G(2, 2000) and τ = G(2, 2000). Taxonomic abbreviations and the corresponding names can be found in Fig. 2.1.

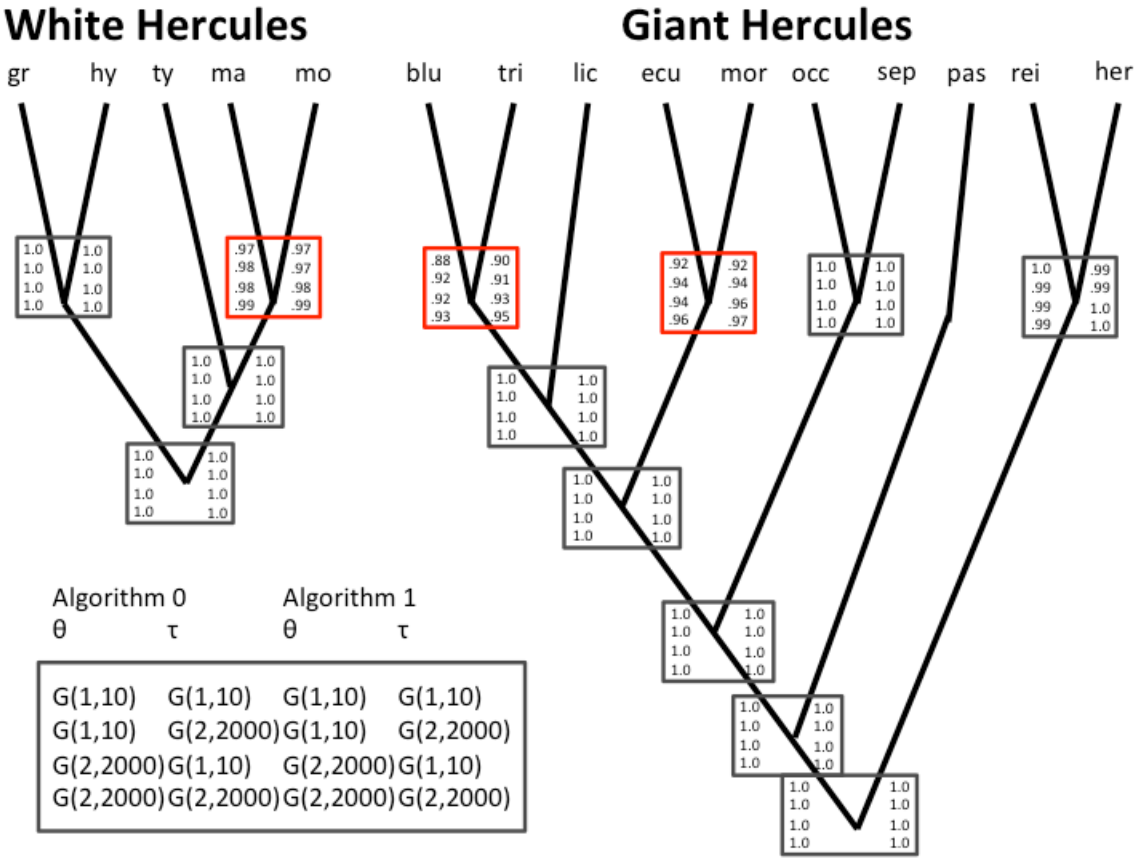

Fig. 2.11. Comparison of BPP results for molecular data using both 0 and 1 rjMCMC algorithms and a reconstructed species tree from *BEAST as guide tree. Splits without absolute supports (<1.0) are shown in red boxes.

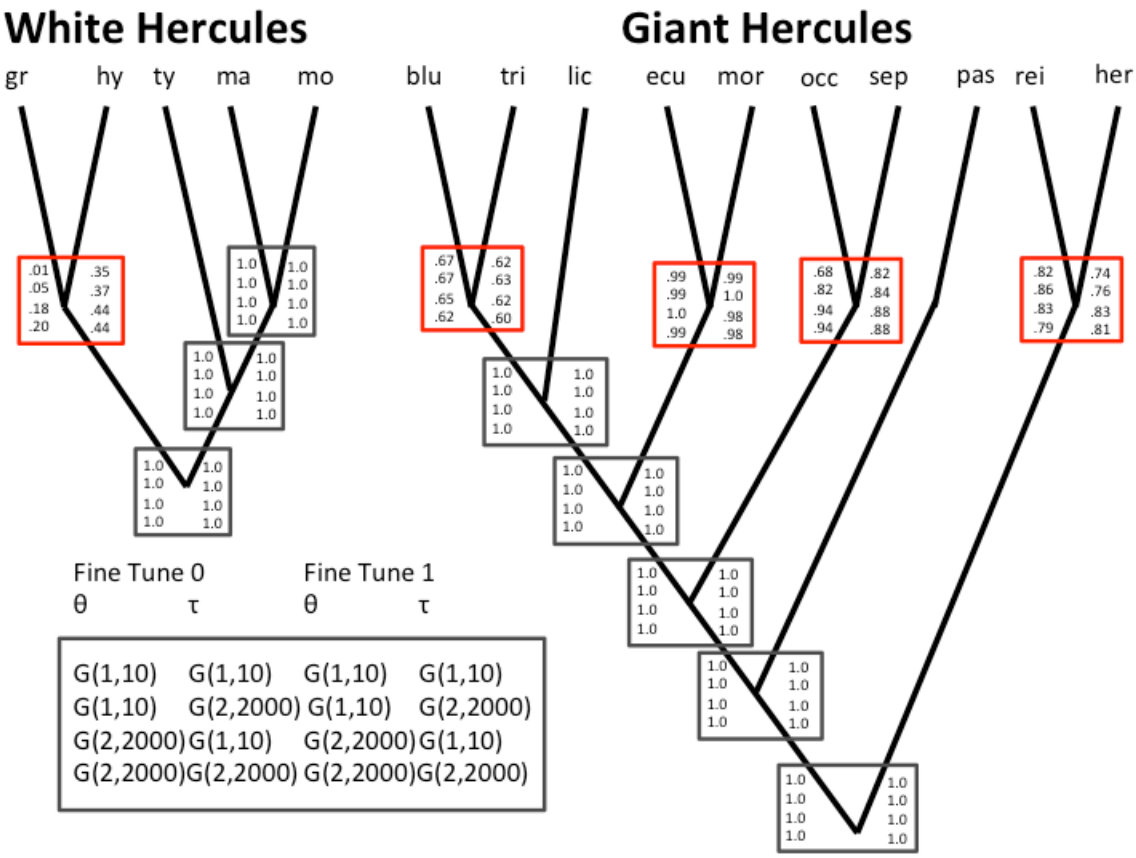

Fig. 2.12. Comparison of iBPP results for analyses based on morphological data only using rjMCMC algorithm 0 and a fine tune setting of 0 and 1 for BM model and a reconstructed species tree from *BEAST as guide tree. Splits without absolute supports (<1.0) are shown in red boxes.

Table 2.5. Results from BPP v.3 analyses of molecular data that integrate over uncertainty in the species tree.

Algorithm (rjMCMC)	Demographic settings	#Topologies (95% Cum)	P(10sp)	P(9sp)	P(blu)	P(tri)	P(blu+tri)
0	θ = G(1,10) τ = G(1,10)	8091	0.945	0.054	0.962	0.962	0.037
0	θ = G(1,10) τ = G(2,2000)	8779	0.935	0.063	0.966	0.966	0.034
0	θ = G(2,2000) τ = G(1,10)	15303	0.941	0.058	0.944	0.947	0.053
0	θ = G(2,2000) τ = G(2,2000)	15042	0.950	0.049	0.954	0.957	0.043
1	θ = G(1,10) τ = G(1,10)	9629	0.954	0.045	0.974	0.975	0.025
1	θ = G(1,10) τ = G(2,2000)	13760	0.923	0.077	0.961	0.966	0.033
1	θ = G(2,2000) τ = G(1,10)	14806	0.901	0.089	0.933	0.935	0.057
1	θ = G(2,2000) τ = G(2,2000)	14503	0.955	0.045	0.960	0.963	0.037

Results from iBPP that jointly analyze genetic and phenotypic data, however, show increased support for taxa compared to analyses based only on genetic or phenotypic data alone. Moreover, the results from the combined genetic and phenotypic data are robust to different priors on the demography of species divergence (i.e., large versus small ancestral population sizes with relatively deep versus shallow divergence times; Fig. 2.10)

and under different algorithms for the rjMCMC (results not shown), with the exception of the taxonomic split of blu and tri within the Giant Hercules group, whose posterior support varied from 0.88 to 0.94 (Fig. 2.10). Based on results from the ibpp analyses that incorporated the effect of uncertainty in the species tree of the Giant Hercules group, 78.87 % of analyses from the sampled species trees support 10 delimited species, while 20.59 % support 9 species. Among the analyses that favor the 9 delimited species, 99.45 % of the sampled species trees collapsed blu and tri into one species. The results were generally robust (e.g., based on analyses of either the genetic and phenotypic data alone, and across different priors and algorithms for the rjMCMC and/or algorithms for modeling trait evolution) with any differences being limited to slight changes in the posterior probabilities of a few splits (Figs. 2.11 and 2.12).

2.5 Discussion

My study not only highlights the arbitrariness associated with some species and subspecies designations, but also demonstrates how a quantitative framework can be used to identify such taxonomic practices and evaluate the statistical equivalency of different taxonomic designations. The results confirm that an integrative species delimitation approach can more effectively and objectively identify evolutionarily independent lineages, compared to relying upon a single data type. Moreover, by analyzing different data types under a common statistical framework, I am able to accommodate the

multidimensional and continuous aspects of the speciation processes when delimiting species. Divergences between putative taxa of Hercules beetles are found to represent different positions along the speciation continuum. When coupled with estimates of relative divergence times, the findings highlight differences in the rates of divergence along genetic, morphological, ecological and geographic axes. Referring specifically to the results for the Hercules beetles, below I discuss the ramification of inconsistent taxonomy on biodiversity studies. I also discuss why inconsistencies in delimited species may arise because of the speciation process, and why relying upon single data type for species delimitation can be particularly problematic. Lastly, I reflect more generally on how quantitative methods for species delimitation are used to assign taxonomic designations when taxa are statistically distinguishable, but exhibit differing degrees of divergence.

Species and subspecies are statistically equivalent

Although both subspecies and biological species are defined as evolutionarily independent lineages (Mayr 1963), the subspecies designation (as opposed to species) is presumed to reflect evolutionary ephemerality (e.g., incipient divergences will be lost if secondary contact occurs between subspecies). Nevertheless, my results show that putative taxa in both Giant Hercules (subspecies) and White Hercules (species) beetles are equivalent in terms of being statistically distinguishable from patterns of character

divergence. Moreover, instead of reflecting some underlying biological difference, assignments of subspecies status to the South American Giant Hercules taxa versus species status to the North American White Hercules taxa most likely reflect differences in taxonomic practices (Wilson & Brown 1953). For example, despite the dichotomy of species versus subspecies designations, taxa from the respective groups do not appear to consistently represent points early or later on the speciation continuum (Nosil et al. 2009) when I consider divergence along morphological, ecological and genetic axes (Fig. 2.7, 2.10). Moreover, only the Giant Hercules taxa from South America overlap geographically (Fig. 2.4), despite being designated as subspecies. As such, the taxonomic designations given to the White and Giant Hercules taxa are inconsistent with the view that geographic overlap is indicative of reproductive isolation (Mayr 1963). Because all the taxa, irrespective of being given species versus subspecies, are statistically identifiable as independent lineages (Fig. 2.10), and based on divergence across multiple dimensions considered here, all the putative Hercules beetle taxa arguably merit distinct species status (a taxonomic revision that formally raises all the subspecies to species status is in review [Huang 2015]; also see below for discussion).

Failure to recognize independently evolving biological lineages as species can have profound effects on evolutionary and conservation studies. For example, currently all putative taxa of Giant Hercules beetles are synonymized according to the Catalogue of Life database (http://www.catalogueoflife.org/). As a consequence, it might be inferred that speciation rate in Hercules beetles, which have a South American origin (Dutrillaux

& Dutrillaux 2013), has increased drastically after the colonization of North America because of the five recognized species in North America. However, my results show that there are potentially at least twice as many species (but see discussion below) in South America as in North America (Fig. 2.10), indicating that evidence for an increased speciation rate associated with the colonization of North America is an artifact of biases in taxonomic practice, given equivalent times for species to accumulate within the two lineages (Fig. 2.9). Likewise, without recognizing the statistical equivalency of taxa previously assigned as species versus subspecies, misguided conservation priorities can result. For example, the decrease in population size of an endemic Brazilian Atlantic Forest species (e.g., *D. hercules paschoali*; see pas in Figs. 2.1, 2.4, & 2.10) due to habitat loss (Ribeiro et al. 2009) may go unrecognized, or viewed as the loss of a geographic population. Understanding potential biases in species delimitation within and between groups of organisms (Table 2.1) is therefore urgently needed for biodiversity studies.

Different positions along the speciation continuum

Speciation can proceed along different possible axes with different rates depending on the cause of speciation (Mayr 1963; Nosil et al. 2009). For example, allopatric speciation with niche conservatism can result in genetically distinct, but morphologically similar evolutionary lineages, whereas ecological speciation can lead to rapid divergence

in phenotypes, without appreciable differentiation at neutral loci (Coyne & Orr 2004). The degree of divergence among the Hercules taxa does differ. However, White Hercules taxa are not consistently more differentiated than the Giant Hercules taxa, despite the taxonomic designation of species versus subspecies, respectively, which again suggests the dichotomy simply reflects differences in taxonomic practices.

As noted above, divergence in morphology (Fig. 2.8) and ecology (as captured by an ENM; Fig. 2.7) varies across taxa. Geographic overlap among taxa isn't observed until taxa are diverged along both morphological and genetic axes, supporting the speciation continuum hypothesis. Moreover, the position of taxa along multiple dimensions of divergence (Warren et al. 2008; Nosil et al. 2009) suggests that allopatric divergence predominates in the Hercules beetles. Most taxa are allopatricly distributed (Fig. 2.4), and there is little support of niche divergence (Fig. 2.7). Divergence in morphology also does not simply reflect an accumulation over time. Instead, some taxa show high levels of morphological differentiation or very little (as evidenced by the support or lack thereof for independent lineages when delimiting taxa based on morphology alone; Fig. 2.12), despite similarly high degrees of genetic differentiation (based on delimited lineages from analyses of genetic data; Fig. 2.11). This suggests that processes other than drift, and most likely sexual or natural selection (Jarman & Hinton 1974; Emlen et al. 2005), are governing rates of divergence in horn shape (Fig. 2.8).

Insights from integrating genetic and phenotypic data

Incorporating multiple data types to delimit species boundaries has had a long tradition (Wiens & Graham 2005; Wiens 2007), although it has been questioned recently given the relative ease of collecting genetic data (e.g., Carstens et al. 2014; but see Edwards and Knowles 2014; Solis-Lemus et al. 2015). One complication of analyzing multiple data types arises when different data types give apparently conflicting results, which can lead to delimitation decisions that can be arbitrary (McKay et al. 2013). However, as my study demonstrates, by accommodating different rates of evolution within a model-based framework, tests of distinguishable species lineages can be made, even if genetic divergences exceed those in morphological data (or vice versa; see Solis-Lemus et al. 2015). It has been known that geographically well-differentiated lineages may show few morphological differences due to similar selective regime, or niche conservatism (allopatric sibling species; Mayr 1963).

My genetic data does clearly support the delimited species in both White and Giant Hercules beetles (Fig. 2.11), without the addition of morphological data (Fig. 2.10), which might be used to argue for genetic species delimitation, as has been recently promoted (Fujita et al. 2012; Carstens et al. 2013; but see Olave et al. 2014). However, I also demonstrate that delimitation based solely on genetic data does not always result in better support compared to phenotypic data (e.g., see the divergence between White Hercules species ma and mo; Fig. 2.11 & 2.12). Moreover, integrative species

delimitation starts with assigning individuals by ecological, behavioral, morphological or geographic cohesiveness (Edwards & Knowles 2014) to putative taxa, whereas mistakes in this initial assignment when based on genetic data will compound downstream analyses (Olave et al. 2014; but see Zhang et al. 2014). Likewise, by estimating the probability of divergence using data from all possible axes along which speciation can proceed, independent lineages arising by diverse speciation processes will be better accommodated (Solis-Lemus et al. 2015).

Quantitative delimitation methods and taxonomic designations

Several aspects of this study raise some critical issues regarding the practical implications of the results from quantitative delimitation methods more generally (i.e., to methods other than, and including, iBPP). The accuracy of delimiting lineages may differ for several reasons. For example, the accuracy of delimited lineages may differ depending upon whether a method has sufficient power to delimit lineages (see Carmargo et al. 2012) and is robust to violations of the assumptions of a model (see Solis-Lemus et al. 2015), or whether the assignment of individuals to putative candidate taxa are accurate prior to analyses of delimitation (see O'Meara 2010; Olave & Knowles 2014). However, these factors all pertain to just one (albeit important) dimension to consider when

applying quantitative species delimitation methods - statistical accuracy of a method for delimiting lineages. An equally important, and often under emphasized, aspect is the relationship between delimited lineages and taxonomic designation - that is, what has actually been delimited by a statistical delimitation method?

The findings of particular relevance, which are documented in the Hercules beetles but are no doubt common to many other taxa, regarding the relationship between delimited lineages and taxonomic designations are that (i) species exhibit divergences that suggest they are positioned along different points of the speciation continuum, as identified by differing degrees of divergence among taxa in genetic, morphological, ecological and geographic data (past and present), and (ii) taxa assigned as species are not consistently more divergent across the different axes of divergence than taxa assigned to subspecies; (iii) nevertheless, all taxa are equivalent (whether assigned species or subspecies status) with respect to being statistically distinguishable.

The interpretation of differing degrees of divergence among data types and across taxa is directly relevant to assigning a particular taxonomic designation, but several different interpretations might be justified. For example, such differences are not only acceptable, but are also expected, under the general lineage concept for a species (de Queiroz 2007), in which case, an argument can be made for assigning all the statistically distinguishable taxa as species. However, many populations within species are also distinguishable with genetic data, and with the increasing availability of genomic data ever more fine geographic structuring of genetic variation is possible (e.g., Spinks et al.

2014). Consequently, without the reliance on some arbitrary threshold (such as the 10× rule of DNA barcoding; Hebert 2004a), analyses of genetic data by themselves are not sufficient to evaluate whether a statistically distinguishable lineage is a population versus species. Even when applying coalescent methods (including BPP or iBPP applied here), whether the genetically distinguishable lineage corresponds to a population or species is not discernable because what these methods distinguish are groups of individuals that have remained without (or with very little) gene flow among them for a sufficient amount of time for divergence in neutral genetic markers to accumulate. At least by considering multiple data types, and demonstrating their significant contribution to species delimitation (e.g., many of the beetle taxa can be distinguished using morphological data alone) it can be shown that differentiation has progressed beyond the inherent property of any isolated population - divergence in neutral genetic markers. Moreover, when coupled with evidence that some taxa overlap geographically (or are projected to have overlapped in the past based on ENMs for the LGM), the genetic distinctiveness of such taxa implies they have diverged sufficiently to be reproductively isolated. That is, only with consideration of multiple data types can we assess whether divergence is above and beyond what is expected when populations are geographically isolated. One caveat with respect to the Hercules beetles is that geographic overlap between sister taxa (when it exists) tends to be low compared to the amount of geographic overlap observed between non-sister taxa. This observation too is subject to different interpretations (see also Pontarp et al. 2015). It may reflect either insufficient time for taxa to achieve

reproductive isolation (which would imply allopatric taxa should not be interpreted as species), or it may reflect that candidate species have not had sufficient time to become sympatric (which would imply that allopatric taxa may indeed warrant species designation). I note that with the delimited taxa here being separated by inhospitable environments (Fig. 2.4) and a general lack of evidence for frequent niche divergence (Fig. 2.7), the allopatric distribution of taxa itself is not evidence for the lack of appreciable reproductive isolation. Again, note that such a determination would not have been possible if I had relied upon genetic data alone to delimit taxa.

While not everyone may accept that I have sufficient evidence to argue that all the subspecies of Giant Hercules should be elevated to species status, or even that all the White Hercules taxa actually warrant the species status they have been granted (especially if someone choses to focus on one data type, such as Fig. 2.8, which I caution against), I would like to close by discussing another critical aspect of quantitative species delimitation that has not received sufficient attention (and applies here even if you think that only some of the taxa warrant species status). What comes next?

Irrespective of whether researchers accept all the results from a study that applies quantitative methods for species delimitation there are two points to consider: (i) the delimited taxa are often used as the input for addressing ecological and evolutionary questions (e.g., Bond et al. 2001; Pons et al. 2006; Esselstyn et al. 2012), yet (ii) the results from delimitation analyses are rarely used in formal treatments on taxonomic status (e.g., for monographic revision and/or species description)(as reviewed in Carstens

et al. 2013). This poses a number of problems. For example, the data used in analyses of biodiversity patterns will differ depending upon whether a researcher adheres to formally described species or relies upon the lineages distinguished from a quantitative delimitation method. Given that the practitioners of such delimitation approaches are typically not the same researchers conducting taxonomic revisionary work, it also means that calls for "future study" are unlikely to impact the actual taxonomy of a particular focal group. This is distressingly ironic when quantitative species delimitation methods are viewed unequivocally as superior or preferred to traditional taxonomic description. This is not to discount the cases where the practicalities of delimitation justify analyses of genetic data alone (e.g., large-scale biodiversity assessments of exceptionally diverse, yet unstudied groups). Nevertheless, for quantitative delimitation methods to realize their full potential, the field needs to come to terms with how the results from quantitative delimitation studies more generally should be applied to advance the systematics of focal taxa (e.g., revisionary study on Hercules beetles based on the results presented here; Huang 2015).

2.6 Conclusion

My finding from chapter 2 highlights how the potential for arbitrary taxonomic decision regarding species/subspecies designations can be identified using an objective statistical framework. Specifically, by analyzing multiple data types within the common

Bayesian framework of the program iBPP, I show that taxa of Hercules beetles are statistically equivalent as distinguishable independent lineages, even though the degree of differentiation between genetic and morphological data varies among taxa. I argue that all the Hercules beetles studied merit distinct species status. By reference to some specific examples, I highlight how inconsistent taxonomic practices across the genus would also result in erroneous interpretation of the factors affecting their rates of diversification. My study also highlights how the apparent conflict of different types of data when delimiting taxa may simply represent variation in the rates that such differences accumulate between species. Moreover, I show that multiple data types can not only be effective for quantitative tests of independently evolving lineages when delimiting species, but can also reveal where taxa lie along the speciation continuum, better capturing the species divergence as a continuous process that occurs across multiple axes. With these points in mind, it is clear that the methods used for quantitative species delimitation (especially those based on genetic data alone) need to consider carefully how results from such methods are used to inform taxonomic status, but that the results from such analyses also need to be incorporated into formal taxonomic treatments if such work is going to have a lasting impact on the systematics and ecological and evolutionary study of the focal taxa.

CHAPTER 3: The Great American Biotic Interchange and diversification history in *Dynastes* beetles (Scarabaeidae; Dynastinae)

3.1 Abstract

Biotic interchange between geographic regions can promote rapid diversification. However, what are the important factors that determine the rate of diversification vary between study systems. The evolutionary history of *Dynastes* beetles, which can be found in both North and South Americas and exhibit two different altitudinal preferences (highland and lowland) is tested for the effects of biotic interchange between continents and different ecological preferences on the rate of species diversification. Additionally, the hypotheses of geological time-dependent and lineage specific diversification rate are tested. My results indicate that in *Dynastes* beetles a pre-landbridge dispersal hypothesis from South to North America is preferred and that the speciation rates are similar between lineages of different geographic origins and different altitudinal preferences. On the other hand, my result from marcoevolutionary cohort analysis reveals that the rate of speciation in *Dynastes* beetles is, instead of trait (geographic and ecological) dependent, lineage specific. Furthermore, a steadily increasing speciation rate can be found in

Pliocene and Pleistocene, which implies that geological and climatic events, i.e., colonizing North America, habitat reformation in the Amazonia, and forest contraction in Pleistocene, have together shaped the current biodiversity pattern in *Dynastes* beetles.

3.2 Introduction

Biotic interchange between geographic realms creates opportunities for species diversification (Wallace 1876; Cody et al. 2010; Gillespie et al. 2012). However, how did intercontinental biotic interchange occur differ significantly between taxa. Closely related lineages found in both North and South Americas characterize the Great American Biotic Interchange (GABI), which provide excellent candidates for studying the effects of biotic interchange on generating biodiversity. It is hypothesized that the closure of the Isthmus of Panama around 3.5 million years ago initiated GABI (Marshall 1988). Many terrestrial lineages expanded their geographic ranges into previously unreachable region by traveling through the newly emerged landbridge (H_1). However, recent studies have revealed additional events that can also account for, and may have more significant effects on, GABI – they are traveling (e.g., rafting) across the marine barrier before the closure of the Isthmus of Panama (H_2 [Bacon et al. 2015]) and island hopping via the Antilles Archipelago (H_3 [e.g., Ali 2012]). Currently lineage diversification patterns across different groups of organisms living in both North and South Americas differ significantly (Cody et al. 2010), which may reflect different historical contingencies

resulted from colonization via different routes.

In addition to the colonization history, different species diversification patterns associated with GABI also intrigue generations of evolutionary biologists. For example, multiple mammalian lineages of North American origin experienced radiation after colonizing South America (Simpson 1950; Marshall 1988), while plant lineages of South American origin diversified after colonizing North America (Cody 2010; Bacon et al. 2015). The colonization into a new geographic area could have resulted in a sudden increase in species number simply because of the newly founded habitats containing multiple open niches. In addition, multiple forest ecoregions emerged contemporaneously with GABI in South America could also promote species diversification. Specifically, habitat reformation occurred in Amazonia, where the Andean mountain extent northward and distinct forest ecoregions were formed in Pliocene (Morrone 2006; Hoorn et al. 2010). Furthermore, forest contraction during Pleistocene could further accelerate allopatric divergence between forest dwelling taxa (Garzón-Orduña et al. 2014). However, the answer to if species diversification is trait dependent (geographical and ecological), geological time dependent (Miocene vs. Pliocene-Pleistocene), or evolutionary lineage specific (which are not mutually exclusive and can be analyzed under the same framework; Rabosky 2014; Rabosky et al. 2014), although has been tested intensively in macroevolutionary studies, can vary significantly depending on the studied systems.

The evolutionary history of a group of Giant beetles (genus *Dynastes* MacLeay,1819; Table 3.1) with a hypothesized South American origin (Dutrillaux &

Dutrillaux 2013), from the Americas is utilized in this study to investigate the biogeographic history of GABI and to test the effects of different macorevolutionary processes that may drive species diversification. There are two major lineages in this genus: (1) Subgenus *Theogenes*, Burmeister 1847, which includes Neptune, *D. neptunus*, and Satan, *D. satanas*, beetles. These two species are restricted to the highland Andes of South America and can be distinguished from species of the other subgenus by completely black elytral coloration and distinct tarsal morphology (Hwang 2011); and (2) Subgenus *Dynastes*, which is the Hercules beetle (Fig. 3.1). The Hercules beetles are distributed throughout North and South Americas and composed of two major groups that can be found in both highland and lowland forest habitats: (2.1) the Giant Hercules group, which includes at least ten evolutionarily independent lineages and can be found in the Neotropics and the Lesser Antilles islands (Chalumeau & Reid 2002; Huang & Knowles 2015); (2.2) The White Hercules group, which includes five evolutionarily independent lineages and can be found in forested habitats of North and Central America (Moron 2009, Huang & Knowles 2015).

Table 3.1. Studied taxa.

Subgenus	Taxa	Abbr.	States*	Geographic distribution
Dynastes	*D. granti*	Dg	0,1	Highland of the southern edge of the Rockies
	D. hyllus	Dhy	1,1	Sierra Madre of Mexico
	D. moroni	Dmo	0,1	Sierra de Los Tuxtlas
	D. maya	Dma	1,1	Central American rainforest
	D. tityus	Dty	1,1	Southeast of North America
	D. h. hercules	Dhh	1,0	Guadeloupe and Dominique
	D. h. reidi	Dhr	1,0	Saint Lucia and Martinique
	D. h. paschoali	Dhp	1,0	Northern Atlantic Forest of Brazil
	D. h. occidentalis	Dho	0,0	The Chocó-Darién
	D. h. septentrionalis	Dhs	0,1	Cloud forests of Central America
	D. h. lichyi	Dhl	0,0	Highland cloud forest of the Andes
	D. h. bleuzeni	Dhb	1,0	Orinoco delta
	D. h. trinidadensis	Dht	1,0	Trinidad and Tobago
	D. h. morishimai	Dhm	0,0	The Yungas of Bolivia
	D. h. ecuatorianus	Dhe	1,0	Lowland Amazonian rain forest
Theogenes	*D. neptunus*	Dn	0,0	Highland cloud forest of the Andes
	D. satanas	Ds	0,0	The Yungas of Bolivia

*Altitudinal (before comma) and geographic (after comma) states of each taxon are represented by 0 (highland or South America) or 1 (lowland or North America).

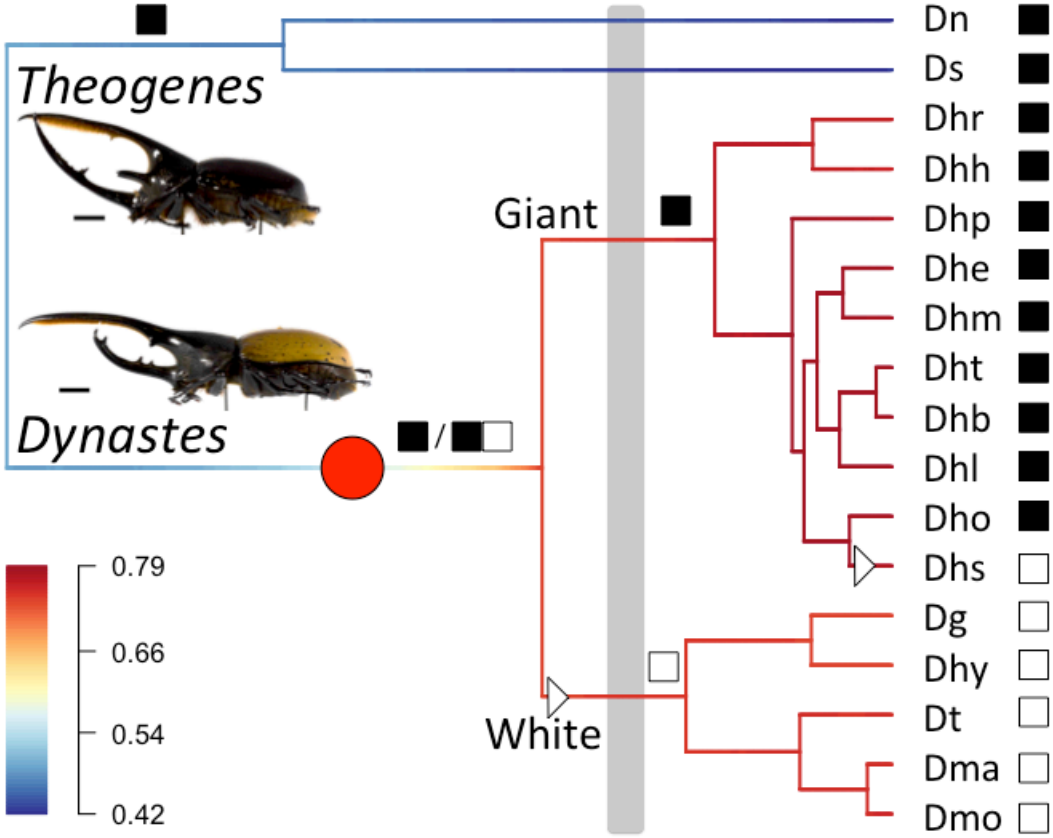

Fig. 3.1. Results from BAMM and LAGRANGE analyses. Branch color represents estimated speciation rate, where a warmer color indicates a faster rate. A red dot on the branch leading to subgenus *Dynastes* indicates a speciation rate shift event. Black and white squares denote geographic states of South and North America, respectively. Black and white squares located on branches denote reconstructed ancestral geographic area, while those that next to the taxon abbreviations current geographic states (Note that, the reconstructed ancestral state for the common ancestor of subgenus *Dynastes* can be either South America or widespread). White triangles indicate inferred dispersal events into North America. A grey shaded area indicates the time frame when the Isthmus of Panama was completely formed (3.4 – 3.6 MYA). Representative samples of *D. neptunus* (subgenus *Theogenes*) and *D. hercules ecuatorianus* (subgenus *Dynastes*) are shown with a scale bar of 1 cm.

The difference in preferring lowland or highland habitats is important because intercontinental dispersal is most likely via the newly emerged lowland landbridge or rafting between lowland coastal regions, while highland distribution can be associated with restricted geographic distribution and local endemism (Hoorn et al. 2010). Furthermore, changes in lowland and highland forest habitats are found drastically in the recent geological history (Hoorn et al. 2010, Garzón-Orduña et al. 2014), which may have impacted the associated diversification rates differently. Therefore, the geographic distribution and the difference in altitudinal preference make *Dynastes* beetles an excellent system to study how species diversification proceeds when intercontinental biotic interchange occurred. Additionally, rafting (H_2), island hopping (H_3), and walking across the Isthmus of Panama (H_1) are all possible explanations for the current distribution of *Dynastes* beetles. For example, the ability to raft across oceans on drifting wood has been demonstrated in arthropods (Coulson et al. 2002). In addition to having a mobile adult stage, Hercules beetles have larval periods where they are constrained to rotten wood, which may serve as overwater dispersal vessels for *Dynastes* beetles. The Hercules beetles nevertheless have the highest species diversity in Central America (Moron 2009), which implies that Central America might be the first colonized region and favors H_1 and H_2 over H_3. This study hence focuses on testing whether the complete formation of the Isthmus of Panama had promoted the colonization of North America (H_1), or did Hercules beetles travel across narrow oceanic strait before 3.5 MYA (H_2).

The reconstructed species trees of *Dynastes* beetles (Fig. 2.9; Huang & Knowles

[2015]), where the evolutionary independence of each tip taxon is quantitatively tested using multiple data types, is used in this chapter to study the biogeographic and diversification history. Ancestral area reconstruction and macroevolutionary comparative methods are utilized to answer the following questions: (1) Does the closure of the Isthmus of Panama promote the colonization of North America? (2) Do North American taxa have a higher diversification rate? (3) Does speciation rate differ according to different altitudinal preferences? And (4) what are the major factors affecting the diversification history in *Dynastes* beetles?

3.3 Materials and Methods

Species trees

A reconstructed *Dynastes* species tree, the majority clade credibility tree (Fig. 2.9), generated by *BEAST analysis was used in this chapter. Taxa included in this study and their associated geographic areas and ecoregions are summarized in table 3.1. Note that species and subspecies in *Dynastes* beetles can be statistically equivalent based on molecular, morphological, and ecological data (Chapter 2; Huang & Knowles 2015). In this chapter, species and subspecies are all treated as different species (they all merit different species status following general lineage concept).

Reconstructing biogeographic history

A geographic state, North or South America, was assigned to each taxon (Table 3.1). Central American taxa were assigned into the North America category, while taxa from the Lesser Antilles Islands were assigned with the South America category. The dispersal, extinction, and cladogenesis model (DEC, Ree & Smith 2008) was utilized to reconstruct the biogeographic history for *Dynastes* beetles. Unconstrained dispersal rate was first used to reconstruct the biogeographic history. I further tested two additional models to investigate if dispersal happened after the closure of the Isthmus of Panama. The first alternative model assumed that the dispersal rate between North and South America was 50% less before the closure (3.5 MYA) than after; the second alternative model assumed that there was no dispersal prior to the completion of the landbridge. To account for uncertainty in molecular dating, two additional sets of analyses that applied different times of rate switch (specifically, 4 and 5 MYA) were also performed. Differences in fitting these models to explain the diversification pattern of the species tree were directly compared using the estimated likelihood values between models.

Estimating speciation/diversification rate

The net-diversification, speciation and extinction rates based on the *Dynastes* species tree were estimated using the program BAMM (Rabosky 2014; Rabosky et al.

2014). Specifically, a total of 2×10^8 generations of rjMCMC searches with samples stored every 1×10^5 generations was performed using the speciation-extinction analyses via BAMM. A total of 1000 post burnin samples (50%) were retained. Note that a separate BAMM analysis that incorporated the information about possible missing taxa, specifically one in the *Theogenes* group and three in the Giant Hercules beetles (i.e., *D. neptunus rouchei*, *D. hercules takakuwai*, and *D. hercules tuxtlaensis* [Hwang 2011] and an genetically distinct *D. h. reidi* lineage from the island of Martinique [Huang 2015]), resulted in a similar pattern of diversification rate through time (Fig. 3.2); therefore, only the results that assume complete taxon sampling from the current species tree were shown. The estimated speciation rate of the *Dynastes* species tree was plotted using the plot.bammdata function from the R package BAMMtools (Rabosky et al. 2014). Additionally, the number of post burnin MCMC samples that support a significant rate shift on the species tree was calculated using a Bayes factor threshold of 3. The estimated net-diversification and speciation rates through time were then plotted using the plotRateThroughTime function. Furthermore, a macroevolutionary cohort analysis (Rabosky et al. 2014) was utilized to test if the estimated speciation rate is highly correlated between closely related lineages using the function getCohortMatrix and cohorts from BAMMtools. Although the extinction rate was estimated as a model component in BAMM, extinction rate was not reported in this study and the interpretation of changes in extinction rate was avoided due to controversies regarding estimating extinction rates from molecular phylogenies (Rabosky 2010).

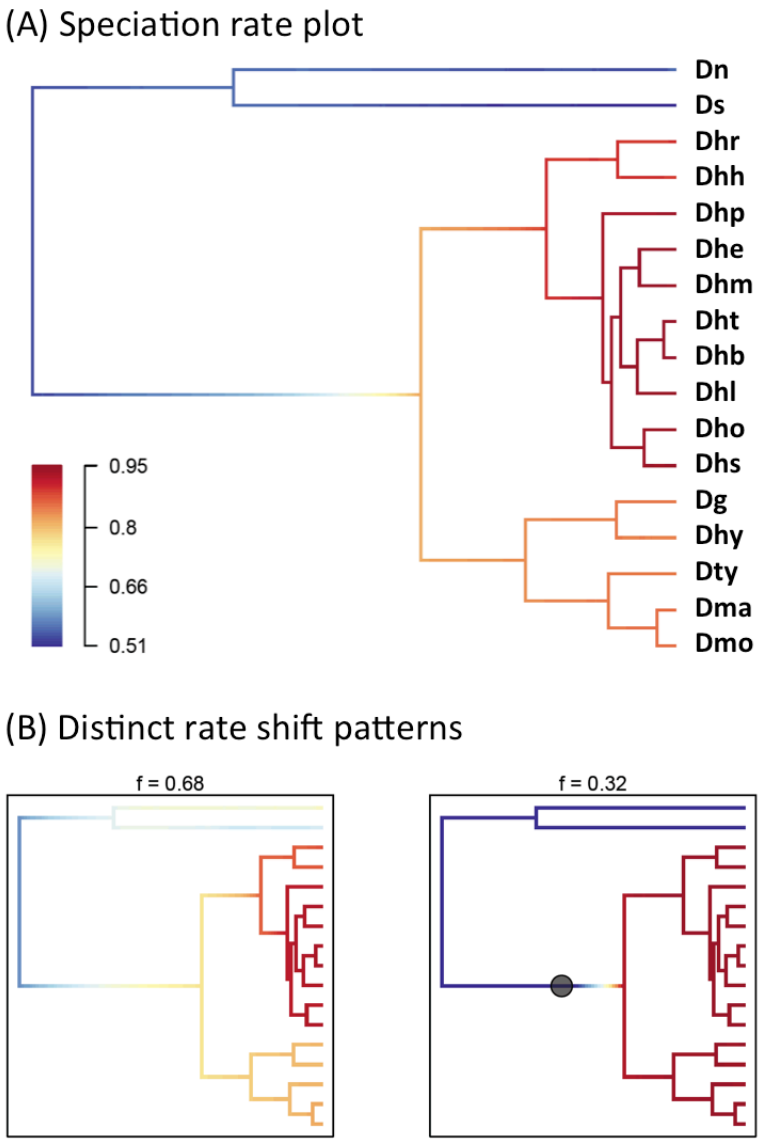

Figure 3.2. BAMM results from analysis incorporating information about possible missing taxa in the data. (A) The *Dynastes* species tree with branches colored according to estimated speciation rate. (B) Distinct rate shift patterns and their frequencies determined by applying a threshold of Bayes factor = 3. A grey circle on the branch indicates the position where rate shift occurred.

Testing the effects of different geographic distributions and altitudinal preferences on species diversification

Because speciation rate can not only change through time, but also correlated with phylogenetic relatedness, trait dependent evolutionary analyses that do not take these factors into consideration can result in erroneous inferences (Rabosky & Goldberg 2015). The structured rate permutations on phylogenies (STRAPP) analysis was developed to cope with such problems by comparing the observed difference in speciation rate between species that exhibit different trait states to a background speciation rate through randomizing the estimated tip speciation rates from the BAMM outputs (Rabosky & Huang 2015). STRAPP analyses for testing trait dependent speciation rate in this study were performed using the traitDependentBAMM function from BAMMtools. Specifically, if speciation rate in *Dynastes* beetles is correlated with different altitudinal preferences or with different geographic origins was accessed by 1×10^4 permutations.

3.4 Results

Biogeographic history reconstruction

The model without setting any dispersal constraint performs best among all three models (-lnL = 9.453; alternative models: –lnLs = 9.545 & 9.868 for 50% & 100% less dispersal prior to 3.5 MYA, respectively), implying that the closure of the Isthmus of Panama may not have significant effect on the colonization of North America in *Dynastes* beetles. Analyses that assumed different times of rate switch lead to the same results. The ML reconstructed ancestral area for each branch is shown based on results from the best (unconstrained) model (Fig. 3.1). South America is inferred as the ancestral state. The ancestral state for the lineage leading to Hercules beetles (subgenus *Dynastes*) can either be South America (lnL = -10.18, $P = 0.4837$) or widespread (lnL = -10.35, $P = 0.4076$) (Fig. 3.1). Two inferred dispersal events can be found on branches leading to the White Hercules taxa and to Dhs (Fig. 3.1). The estimated dispersal and extinction rates are 0.03285 and 4.285×10^{-9}, respectively.

Changes in diversification rates

The rjMCMC searches in the BAMM analysis reached plateau soon after the first 1000 generations. By using a Bayes factor of 3 as threshold, 33 % of the post burnin

samples indicate a significant rate shift, and this rate shift, which is an increase in diversification rate, is located on the branch leading to subgenus *Dynastes* (Fig. 3.1). The RTT plots unveil steadily increasing speciation and net-diversification rates through time, where a sudden increase in rates can be found around 4 MYA (Fig. 3.3). The results from macroevolutionary cohort analysis reveal that the phylogenetic distance between taxa is highly correlated with the estimated speciation rate. For example, species from the White Hercules beetles share a highly similar speciation rate, whereas the estimated speciation rates between species from White and Giant Hercules beetles are less similar (Fig. 3.4).

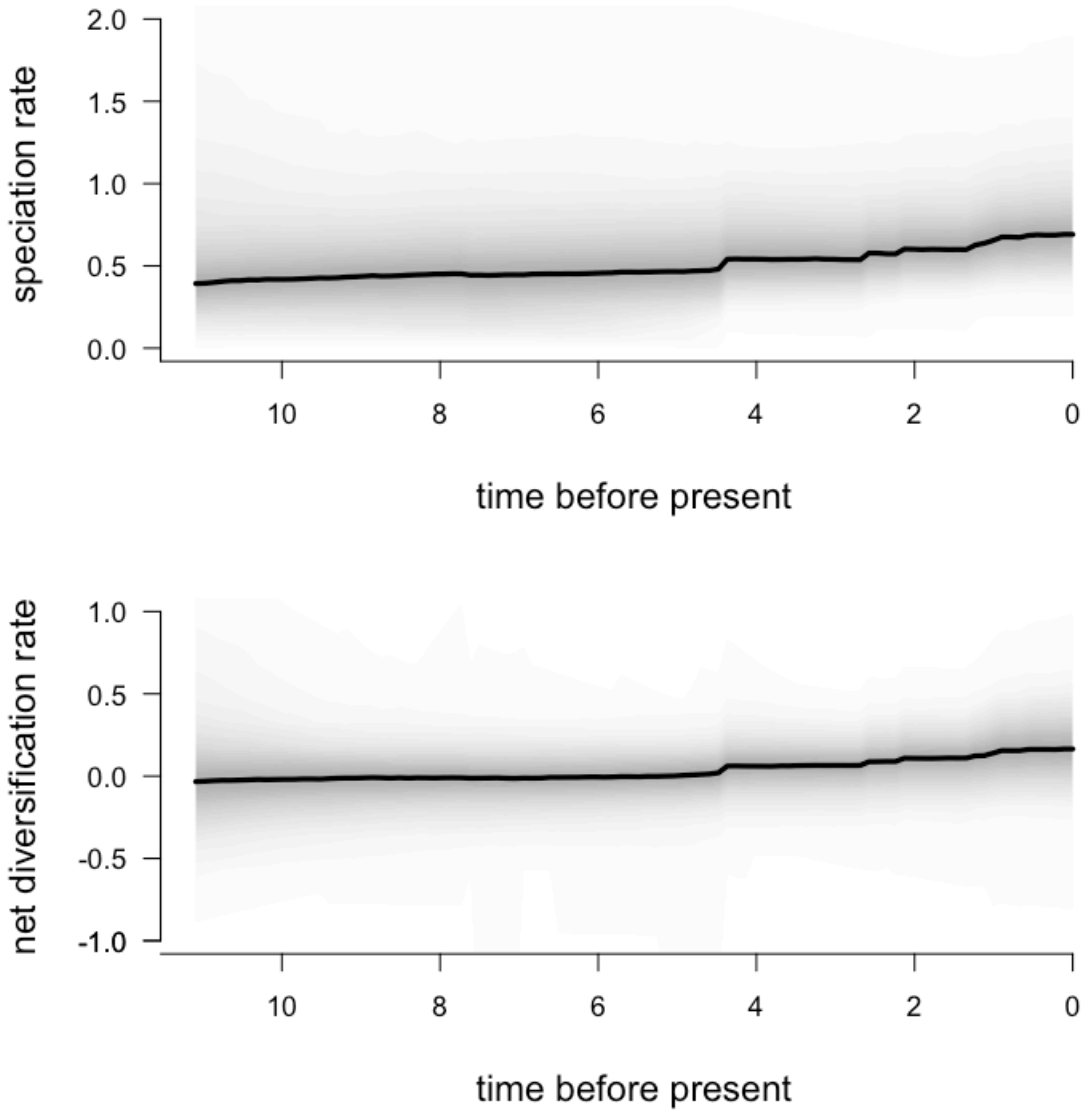

Fig. 3.3. Results from rate through time plots. The x-axis is in a scale of million years. Solid lines indicate the mean rates, while the grey areas represent the 5% to 95% Bayesian credible regions for the distributions of the rates.

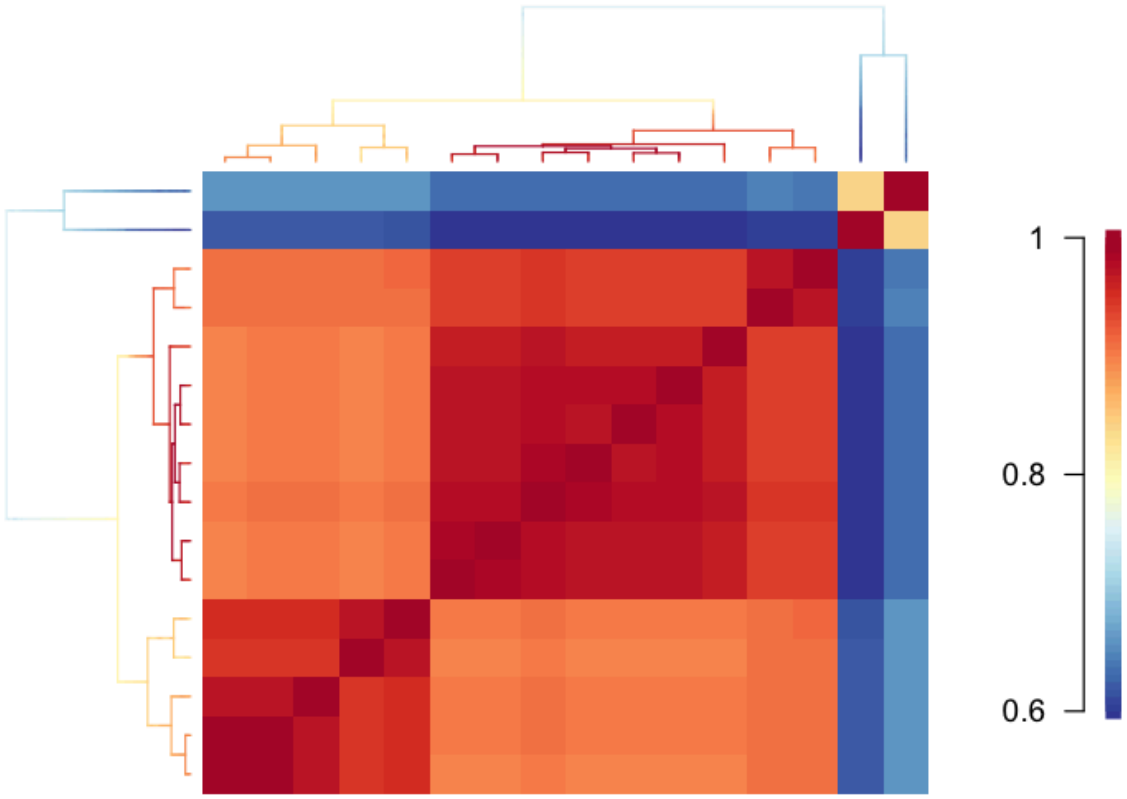

Fig. 3.4. Results from macroevolutionary cohort analysis. A correlation matrix based on speciation rates between tip lineages of the phylogeny is plotted, where each correlation is a posterior frequency that the two compared species are found in the same macroevolutionary rate regime. A warmer color represents a higher correlation than a colder color. The correlation between any two species can be found by locating their intersection in the matrix.

The effects of different geographic distributions and altitudinal preferences

The estimated speciation rates for North and South American taxa are 0.75 ± 0.006 (SE) and 0.71 ± 0.043, respectively (Fig. 3.5). A *t*-test assuming unequal variance shows insignificant difference in estimated speciation rates (t = -0.9226, df = 10.403, P = 0.3771). In addition, the STRAPP result also indicates that the estimated speciation rate is not significantly dependent on the geographic states (P = 0.919). Similarly, the estimated speciation rates between taxa of different altitudinal preferences (0.76 ± 0.006 and 0.68 ± 0.057 for lowland and highland species, respectively [Fig. 3.5]) are not statistically different (t = 1.4672, df = 7.146, P = 0.1849). The STRAPP result also reveals insignificant support for altitudinal preference dependent speciation rate (P = 0.9105).

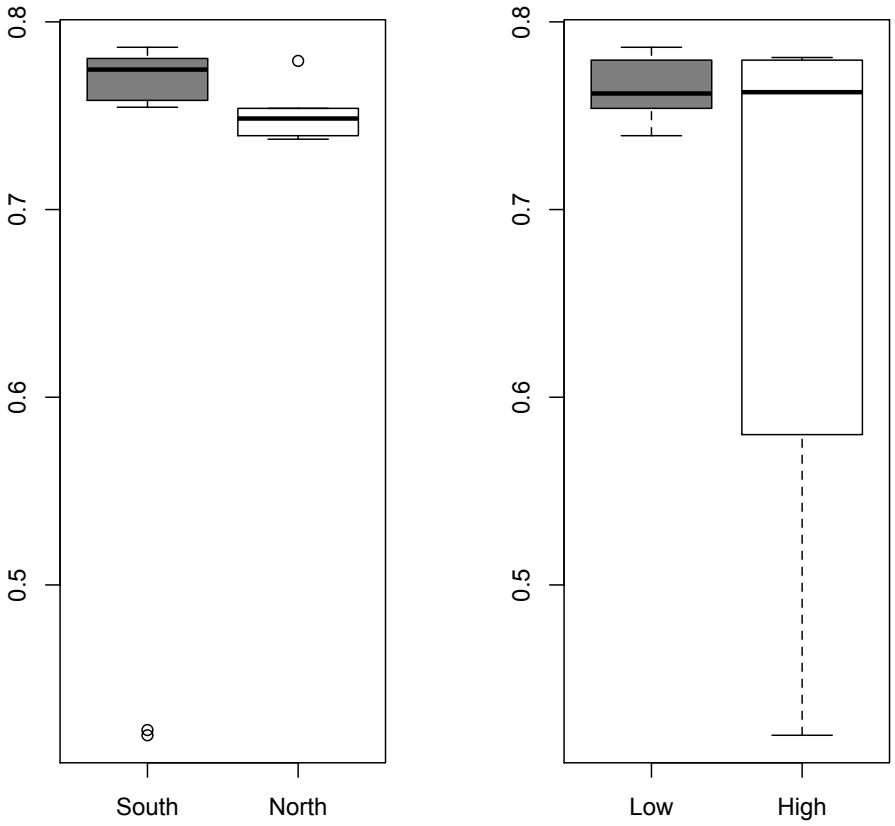

Fig. 3.5. Comparisons of estimated speciation rates between trait states. Left panel: the estimated rates between South (black) and North (white) American *Dynastes* beetles. Right panel: the estimated tip rates between lowland living (black) and highland living (white) species.

3.5 Discussion

A recently reconstructed *Dynastes* phylogeny, where all taxa are statistically supported as evolutionary independent lineages, is used in this study to test biogeographic hypotheses about processes that may promote species diversification. It

has been shown that fail to recognize true biological entities in a macroevolution study can severely affect the interpretation of mechanisms that lead to the current biodiversity (Smith et al. 2013). Given a fairly complete sampling of taxa that are statistically delimited as different species (see chapter 2; Huang & Knowles 2015), results and inferences made from this study should be robust. It is unraveled here that *Dynastes* beetles have a South American origin and that the GABI in Hercules beetles predates the closure of the Isthmus of Panama. The estimated speciation rate is highly lineage specific and a rate increase can be identified on the branch leading to Hercules beetles. My results from RTT plots further indicate an increase, although slightly, in diversification rate in Pliocene and Pleistocene, which corresponds to a slowing down in geological activity and an increased frequency in climatic fluctuation and reformation of ecoregions in the Amazonia (Hoorn et al. 2010, Garzón-Orduña et al. 2014). Furthermore, I show that the geographic states of North and South America have similar effect on species diversification and that different ecological states of preferring lowland and highland habitats also have similar effect on speciation rate. The biogeographic history and factors affecting the diversification in *Dynastes* beetles are discussed in the following sections.

Biogeography

A South American origin with subsequent dispersal events into North America is inferred for the *Dynastes* beetles using the DEC model. My finding is in congruence with

a recent study done by investigating karyotypes (Dutrillaux & Dutrillaux 2013). In their study, ancestral and derived chromosomal types are found in *Dynastes* taxa from South America, but the ancestral type is absent in North American taxa. It is also true that both subgenus *Dynastes* and *Theogenes* are distributed in South America, while there is no *Theogenes* taxa in North America. A South American origin of the genus *Dynastes* is thus a favored hypothesis and supported by multiple lines of evidence.

Although the reconstructed time calibrated species tree of *Dynastes* beetles reveals that lineages leading to North American taxa originated after the closure of the Isthmus of Panama (Fig. 3.1), my DEC results indicate that a model assuming constant dispersal rate between the Americas through time fits the species tree better. My results presented here may suffer from the effect of low statistic power to discriminate between models because of a small sampling size (only 17 tip taxa), but the clear trend of decreasing likelihood value for models assuming constrained dispersals before 3.5 MYA implies that the completion of the Isthmus of Panama may not be a major driving force for GABI in *Dynastes* beetles. Two Giant Hercules taxa (Dhh and Dhr) have successfully colonized islands of the Lesser Antilles, i.e., Saint Lucia, Martinique, Dominique, and Guadeloupe (Chalumeau & Reid 2002). A historical record indicates that they might have made it to Hispaniola as well (Wetherbee 1985). The mobile adult stage, which could fly for a decent geographic distance, and a potential dispersing larval stage via drifting wood could have enabled Hercules beetles to travel across the narrow oceanic strait before the closure of the Isthmus of Panama (see Leigh et al. [2014] for a map of the Isthmus

between 12 and 6 MYA). Such inference has also been reported in many terrestrial organisms, where specifically a South to North America biotic introduction became apparent around 6 MYA (Bacon et al. 2015). Given the fact that Central America has the highest species diversity from the Hercules beetle lineage and a pre-landbridge dispersal model is favored, the most likely historical scenario explaining the biogeography of *Dynastes* is $\mathbf{H_2}$: pre-landbridge dispersal through rafting across Central America.

Conventional molecular biogeographic studies focusing on similar question tend to infer a predominate role of landbridge in inter-continental biotic dispersal if the estimated age of the common ancestor between North and South American lineages are found generally less than or close to 3.5 MYA based on a dated phylogeny. While pre-landbridge dispersal are inferred if the estimated common ancestor between lineages of North and South America significantly predates 3.5 MYA. Few studies have tried fitting different models on the reconstructed phylogeny and compared the goodness of fits between models before making inferences (Bacon et al. 2013). My results from studying *Dynastes* beetles presented here indicate that interpreting deterministic process by observing molecular phylogenetic patterns in a biogeographic study without applying statistic tests between alternative explanations should be taken with cautions.

Diversification process

It is revealed in this chapter that speciation rate in *Dynastes* beetles is not trait

dependent. An interesting finding here is that the geographic state of living in North America is not correlated with a higher speciation rate. Although rapid diversification is commonly observed after successful biotic introduction, it has been shown that mammal species having a South American origin tend to have limit success to diversify in North America (Simpson 1950; Marshall 1988; Webb 1991). The rain forest habitat, for example, only covers a small proportion of North America, allowing only a small area for species of South American origin to successfully diversify assuming phylogenetic niche conservatism. Additionally, repeated glaciations during Pleistocene in North America may have exterminated decedents from lineages of tropical South American origin. It is also intriguing to point out that according to my results here the evolution of different ecological preferences does not correlate to differences in speciation rate. Although lowland living may have facilitated the colonization into previously isolated continents, speciation rate can be dependent on other factors, which may not be associated with the specific trait state that facilitate dispersal. Specifically, the number of available niches that may promote speciation in *Dynastes* beetles can be highly correlated to the number of allopatrically/parapatrically distributed forest ecoregion (speciation predominated by allopatric process; see Chapter 2 [Huang & Knowles 2015]), which can be independent from being a highland or lowland geographic state. That is, both highland and lowland regions have many distinct ecoregions.

Colonizing North America however did result in biological diversification in White Hercules beetles (i.e., five species are generated within 3 million years), and thus

the lack of opportunity, which is often invoked to explain the lack of successful diversification in mammals of South American origin after GABI, can not fully explain my result of similar speciation rates between *Dynastes* lineages from North and South Americas. Speciation rate in White Hercules beetles is not comparable to that of its South American counter part, Giant Hercules beetles (i.e., at least 10 species are generated within 3 million years; Fig. 2.9 & 3.1). Whereas speciation rate in White Hercules beetles is faster than that in the subgenus *Theogenes* (Fig. 3.1). Because South American lineages are composed of fast diverging Giant Hercules beetles and slow diverging *Theogenes* taxa (Fig. 3.1 & 3.4), the estimated speciation rate for all South American lineages as a whole can be misleading. Colonization into a new continent indeed resulted in species diversification in Hercules beetles; however, there were contemporaneous events occurring in South America (Hoorn et al. 2010; Garzón-Orduña et al. 2014), which could have resulted in an even faster speciation rate. Specifically, the recent formation of a variety of ecoregions in Amazonia (Hoorn et al. 2010) and subsequently the forest contraction because of a drier climate condition in Pleistocene (Garzón-Orduña et al. 2014) could together lead to an increase in speciation rate in the South American Giant Hercules lineage. This inference can be further supported by my results from RTT plots, where an increased species diversification rate can be found in Pliocene and Pleistocene. Additionally, my results from macorevolutionary cohort analysis clearly suggest that speciation rate is highly lineage specific and that the fastest diverging lineage in *Dynastes* is composed of taxa that live in ecoregions that are geographically very close to the

northern Andes (Dhl, Dhb, Dhm, Dhe, Dho, and Dhs; table 3.1; Fig. 2.4, 3.1, & 3.4), where the habitats changed most drastically in the recent history. Comparing to geographical and ecological explanations of different diversification patterns found between biological systems in the Americas, the importance of lineage specific properties and the formation of ecoregions in Pliocene and forest contraction during Pleistocene (Garzón-Orduña et al. 2014) have received less attention. However, it is clear based on my results that geological and climatic events play important roles in shaping different diversity patterns in different *Dynastes* beetle lineages.

3.6 Conclusion

In chapter 3, the biogeographic and diversification history of *Dynastes* beetles is studied and a potential problem of inferring historical process by observing divergence patterns and times from a reconstructed phylogeny is revealed. Although all North American lineages are formed after 3.5 MYA, dispersals between Americas do not necessarily have to occur after 3.5 MYA. In fact, a model assuming a constant dispersal rate before and after the closure of the Isthmus of Panama fits better the reconstructed *Dynastes* species tree according to my results. Thus, for historical biogeographic studies, cautions should be undertaken should alternative scenarios that can explain the observed phylogenetic patterns were not tested and compared statistically. I also show that speciation rate in *Dynastes* beetles may not be dependent on specific geographic or

ecological trait states. Instead, it changes in a lineage specific manner, which could be resulted from different lineage specific historical processes. Specifically, the determinant of diversification in *Dynastes* beetles is likely the availability of different forest ecoregions, which is the result of changes in climatic condition and geological activity and varies across geological times. It is clear that, although conventionally trait-dependent, lineage- specific, and time- dependent evolutionary patterns have been investigated as independent subjects of interest, an integrative model that can investigate the effects from all these potential diversification drivers can help us understand the biodiversity pattern more comprehensively. Studies that only investigate the effect of one specific factor can result in biased interpretations.

CHAPTER 4: Tests of divergence across multiple levels of biodiversity in *Xylotrupes* beetles: impact of oceanic and forest barriers on diversification dynamics across the Indo-Australian Archipelago

4.1 Abstract

The effect of different barriers on genetic divergence, and whether the impact is similar on multiple levels of biodiversity, is investigated using *Xylotrupes* beetles from Indo-Australasian Archipelago. Specifically, tests of the role of oceanic barriers and forest fragmentation in promoting population subdivision, species diversification, and community differences are conducted. Two mitochondrial (COI and 16S in 304 and 275 individuals, respectively) and two nuclear (ITS1 and ITS2 in 128 and 251 individuals, respectively) loci were sequenced from 81 populations sampled across all taxa from 8 species groups of *Xylotrupes* beetles. A phylogenetic history of divergence was estimated using the program *BEAST, as well as speciation duration under a protracted birth-death model, and the timing of divergence using a molecular clock. The role of vicariance and dispersal in structuring the beetle communities was investigated using a dispersal-extinction-cladogenesis model. A correspondence between shifts in

diversification rates associated with historical changes in oceanic barriers and forest fragmentation was tested using the Geographic State Speciation and Extinction model, lineage-through-time plots, and tests of competing models of evolutionary diversification using stepwise AIC. At the population level, pairwise F_{ST} were used to test for divergence within species associated with the different barrier types. My results show that different zoological regions isolated by oceanic trenches are characterized by endemic lineages and rare dispersal between regions. Isolation by oceanic barriers correlates with significant population subdivision, but not with speciation rate, whereas historical increases of forest fragmentation due to past climatic change are associated with an increase in diversification rate. The effects of barriers on divergence across the multiple levels of biodiversity differ and depend upon the barrier type. These differences may relate to the duration of speciation and the time required for species to evolve coexistence in sympatry, which may mediate the effects of processes contributing to divergence and species diversity.

4.2 Introduction

Although it is clear that different types of barriers can impact species divergence, the effects of a specific barrier type on differentiation at multiple levels of biodiversity has rarely been investigated (Wiens 2012). Because of this knowledge gap, it is unclear the extent to which biogeographic patterns of species diversity at the macroevolutionary level

can be informed by phylogeographic studies of microevolutionary phenomena and *vice versa*. For example, although physical barriers often promote population subdivision, this incipient divergence may not persist long enough to become new species (Sukumaran & Knowles 2016), or populations may merge upon secondary contact, such that the effects of barriers on speciation rate may vary depending upon the persistence of within-species lineages (Dynesius & Jansson 2014). Moreover, the links between speciation rate, and hence, macroevolutionary patterns, and processes operating at the microevolutionary scale may also depend on factors that influence the number of divergent populations within species (Dynesius & Jansson 2014). For example, with multiple types of barriers, a higher number of divergent populations might contribute to increased diversification. However, differences in the persistence of populations under different barrier types might also mean that a barrier only contributes to microevolutionary divergence (Papadopoulou & Knowles 2016).

Both physical barriers (e.g., oceanic barriers) and changes in climatic conditions that alter island connectivity are hypothesized as main forces that structure both phylogeographic and biogeographic patterns (particularly in island/archipelago systems; Lohman et al. 2011; Papadopoulou & Knowles 2015). However, the relative contribution of different barrier types at a given level of biodiversity (i.e., population versus species divergence) and whether they will have a concerted impact on divergence across levels remains poorly understood.

Here I present an integrative biogeographic and phylogeographic study to

investigate (i) the effects of different barrier types on divergence in rhinoceros beetles (genus *Xylotrupes*, Dynastinae, Scarabaeidae) from the Indo-Australian Archipelago (IAA), and (ii) whether their impact is similar on multiple levels of biodiversity. The IAA encompasses multiple zoological regions of high biodiversity (Lohman et al. 2011). The highly fragmented distribution of islands, as well as the complex history of repeated connections during its dynamic climatic and geological history, is hypothesized to drive diversification in the region (Lohman et al. 2011; Hall 2012). In addition to geographic explanations, forest fragmentation during drier climatic conditions has also been hypothesized to promote diversification (e.g., Bendiksby et al. 2010; Morley 2012). Six *Xylotrupes*, species groups, which are restricted to forests, are widely distributed throughout the IAA (Fig. 4.1)(Rowland 2011). Within the species groups, morphological systematic studies have revealed localized geographic subspecies (Rowland 2003; 2011), suggesting the potential importance of geographic isolation as a driver of diversification in *Xylotrupes*. However, given that the taxa are associated with forests, it is also possible that habitat fragmentation has contributed to divergence in the beetles. In this study, I test the effects of oceanic barriers and forest fragmentation associated with past climatic change on divergence at three different biodiversity levels: tests of the role of oceanic barriers and forest fragmentation in promoting population subdivision, species diversification, and community differentiation. I also estimate the speciation duration (Etienne & Rosindell 2012) in *Xylotrupes*, which provides a context for interpreting why the barriers might (or might not) act in a concerted manner in terms of the effects of

divergence at different levels of biodiversity (as discussed above).

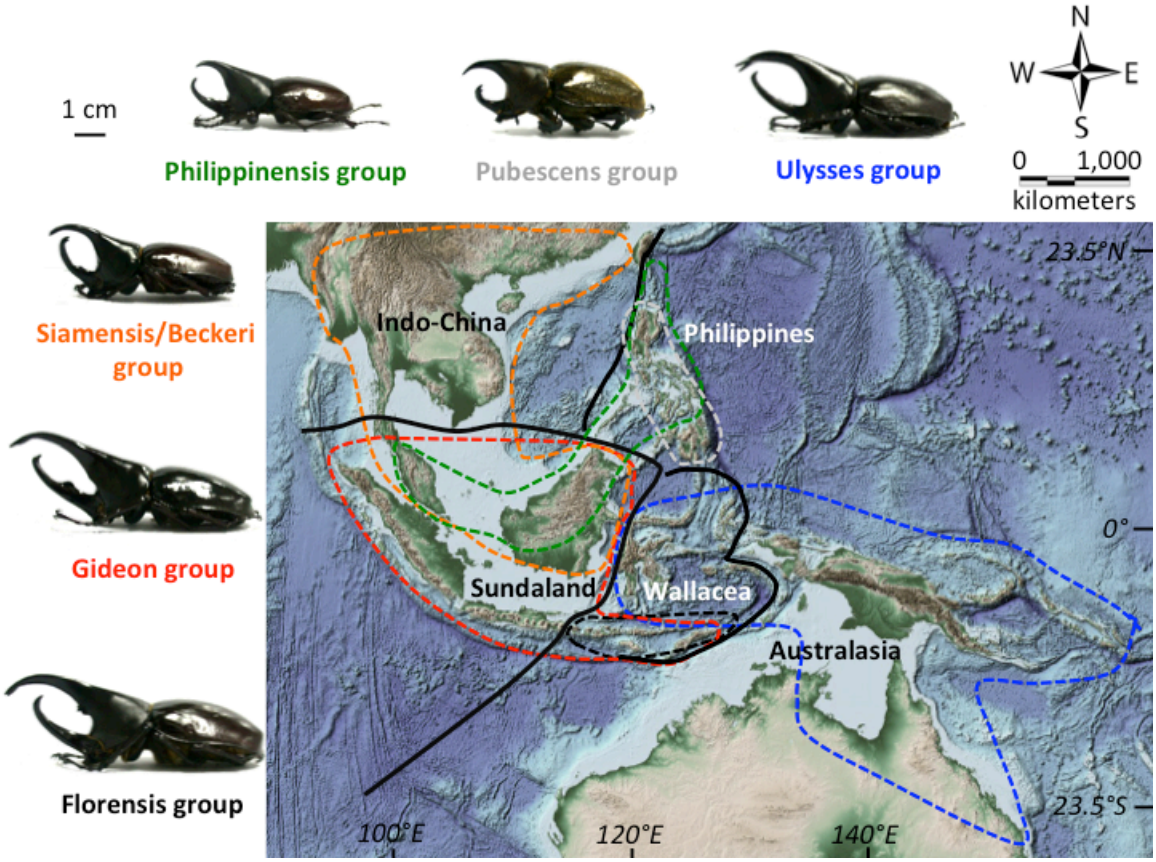

Fig. 4.1. The map of Indo-Australian Archipelago and representatives of major males from six *Xylotrupes* species groups. Solid lines indicate geographic breaks between zoological regions and colored dashed lines show the geographic distributions of each species groups. Geographic areas that would have connected isolated landmasses during glacial periods (Hall 2012) are shown in light greyish blue color.

4.3 Material and Methods

DNA extraction, sequencing and alignment

DNA was extracted from thoracic muscles using the DNeasy Blood & Tissue Kit (QIAGEN, USA) in 304 beetles sampled from multiple populations of 30 species (including seven un-described species) and six subspecies from all the eight *Xylotrupes* species groups (Table 4.1). Species and subspecies nomenclature used here follows J. M. Rowland (2003, 2011, and personal communication); all currently recognized species were included in this study. Voucher specimens were deposited in the Insect division of the Museum of Zoology at the University of Michigan (UMMZ), the University of New Mexico (UNM), and Tunghai University (THU).

Table 4.1. Summaries of taxa studied in this chapter.

Species Group	Species/Subspecies	Collection Localities (CODE[#])	Zoological Region
Florensis	*X. florensis*	Wetar Island (WET-F), Flores Island (XF), Pantar Island (XFP), Sangeang Island (XFS), Mt. Gumitir, E. Java? (GUM-F)	Sundaland/Wallacea
	X. florensis tanimbar	Larat Island (XFL), Tanimbar Island (XFT)	Wallacea
Meridionalis	*X. meridionalis*	Mudigere, India (XME), Kerala, India (KER-M)	India
Mniszechii	*X. mniszechii*	Pallandri/Mang Azad/Rawalakot, Kashmir (XMN), Mt. Gaoligong, Yunnan (XMNYN), Medog, Tibet (XMNT)	Himalaya
Siamensis/ Beckeri	*X. siamensis*	Hainan Island (HAI-T), Kanchanaburi (KAN-S), Sri Sawat, Kanchanaburi (KAN-T), Kho Chang Island (KHO-S), Xieng Khouang, N. Laos (LAO-S), Chiang	Indo-China

		Mai (SIA), Hong Kong (SIAHK), Xishuangbanna, Yunnan (SIAYN), Shaoguang, Guangdong (TON), Ngoc Linh Mt., Kon Tum (TONKT), Ba Nam, Ba To, Quang Ngai (TONQG), Tr'hy, Tay Giang, Quang Nam (TONQN), Captive breed from Vietnam (TONV)	
	X. beckeri	Cameron Highlands (XB)	Sundaland
	X. beckeri 2 (Lowland species)	Pagai Island (PAG-B), Siberut Island (SIB-B), Kluang (KLU-B)	Sundaland
	X. wiltrudi	Central Kalimantan (XWIK), Sambas, W. Kalimantan, Indonesia (XWIS)	Sundaland
	X. sp.1	E. Java (EJA-X)	Sundaland
	X. inarmatus	Mt. Semeru, E. Java (SEM-I/XINS), Mt. Argopuro (XIN)	Sundaland
Gideon	*X. gideon*	Mt. Argopuro, E. Java (G), Bali (GB), Lombok (GL), Lampung, S. Sumatra (LAM-G), W. Java (WJA-G), Gunung Halimun, W. Java (XGH*)	Sundaland/Wallacea
	X. damarensis	Yamdena (GY), Tanimbar Island (XGI)	Wallacea
	X. tadoana 1	Wetar Island (GW/WET-G), Timor Island (SOT/TIM-T)	Wallacea
	X. tadoana 2	Sumba Island (SOB), Flores Island (SOF), Sangeang Island (SOS), Sumbawa Island (SOW)	Wallacea
	X. pachycera	Kalimantan, Borneo (BO)	Sundaland
	X. sumatrensis	Maninjau Lake, Sumatra (SU)	Sundaland
	X. sumatrensis tanahmalayu	Cameron Highlands (BE)	Sundaland
Ulysses	*X. lorquini*	Buton Island (BUT-L), Captive breed sample from Sulawesi (BAUS), Muna Island (MUNA)	Wallacea
	X. rindaae	Selayar Island (SEL-R)	Wallacea
	X. falcatus	Sanghir Island (XFA)	Wallacea
	X. telemachos	Makian Island, Maluku (MAK-T), Halmahera (XST)	Wallacea
	X. clinias 1	Ceram Island (CER-C), Ambon Island (AMB-G)	Wallacea
	X. ulysses	Duke of York Island (YOR-U), Kimbe (KIM-U)	Australasia
	X. clinias 2	Kei Islands (BAU)	Wallacea
	X. australicus	Garradunga, Queensland (AUS/XAA), Brisbane,	Australasia

		Queensland (XAB)	
	X. australicus darwinia	Darwin, Northern Territory (XAD)	Australasia
	X. carinulus	Fak-Fak, W. Irian (U), Wau, Morobe, Papua New Guinea (LA)	Australasia
	X. macleayi	Misima Island (MIS-M), Trobriand Island (TRO-M)	Australasia
	X. macleayi szekessyi	Buin District, PNG (BOU-S), Buka Island (BUK-S), Guadalcanal, Solomon (MA)	Australasia
Pubescens	*X. pubescens*	Mt. Apo, Mindanao (XP), Mt. Musuan, Mindanao (XPM)	Philippines
	X. pubescens beaudeti	Mt. Balocane, S. Leyte (XPUL)	Philippines
	X. sp.2	Luzon (XSP)	Philippines
Philippinensis	*X. philippinensis*	Dinalungan, Aurora, Luzon (XPHA), St. Thomas, Luzon (PHEST), Marinduque Island (PHM), Maripipi Island, Leyte (PLE), Bulacan, Luzon (BUL-P)	Philippines
	X. philippinensis peregrinus	Green Island, Taiwan (PHG), Orchid Island, Taiwan (PHO)	Philippines
	X. pauliani	Cameron Highlands (CAM-P)	Sundaland
	X. sp.3	Malaysia (MAL-X)	Sundaland

Mitochondrial COI (C1-J-1718/C1-N-2454; Bell et al. 2004) and 16S (16Sar/16Sbr; Simon et al. 1994), as well as nuclear ITS1 (TW81/HITR; Richards et al. 1997) and ITS2 (FB5.8SFWD/FB28SREV; Jenkins et al. 2007) loci, were amplified on an Ependorf thermocycler (Mastercycler Gradient, Hamburg, Germany) with *Taq* DNA polymerase (Invitrogen, USA). The PCR profile for all amplifications involved denaturing at 94°C for three minutes, followed by 35 cycles of a 94°C one minute denature step, a 52°C one minute annealing, and a 72°C one minute extension, and a final 10 minute 72°C extension. PCR products were purified using a QIAquick PCR Purification Kit (QIAGEN,

USA) or ExoSap (Affymetrix, USA).

DNA was sequenced on an ABI PRISM^{TM} 377 automatic sequencer (Perkin Elmer, USA) at the University of Michigan sequencing core (see Table 4.2 for details about sequences; GenBank accession # xxxxx-xxxxx [not publically available yet]). Initial multi-sequence alignments from ClustalW (implemented in MegAlign; DNAStar package, Madison, USA) were imported into SeaView (version 4.4.0; Gouy et al. 2010) for alignment using the program Muscle (Edgar 2004). COI sequences were translated into amino acid sequences using a mitochondrial genetic code of *Drosophila* in Mesquite (version 2.75, Maddison & Maddison 2011) to confirm the absence of stop codons.

Table 4.2. Summary of sequenced loci, sample sizes, and DNA variation.

Loci	# individuals	Alignment length	# PI* sites
COI	304	604	192
16S	275	496	100
ITS1	128	957	203
ITS2	251	541	96

*Parsimony informatics sites

Estimates of phylogenetic relationships and divergence times

To account for genealogical discord arising from random gene-lineage coalescence, a species-tree approach was used to estimate a history of divergence for the beetles. With

a taxonomic revision of *Xylotrupes* in progress (personal communication with J.M. Rowland), I considered both *Xylotrupes* species and subspecies as terminal units in the species-tree analysis (see Carstens & Knowles, 2007, for application of species-tree approaches below the species level). A species tree was estimated using the Bayesian program *BEAST (Heled & Drummond 2010) implemented in BEAST2 (version 2.0.1; Bouckaert et al. 2014) under a relaxed lognormal clock. A GTR + I + Γ model of nucleotide substitution was parameterized for COI, 16S, and ITS1, and a HKY + I + Γ for ITS2; for estimates of gene trees, the mitochondrial loci COI and 16S were treated as a single locus independent of the nuclear locus comprised of the linked ITS1 and ITS2 regions. A stepwise linear model with a constant root under a Yule speciation process was used in initial species-tree analyses to optimize prior and operator settings. A species tree was estimated from a run with 1×10^9 generations, in which parameters and trees were stored every 1×10^4 generation, and the first 20% of the runs discarded as burn-in, using the BEAGLE library (Ayres et al. 2012). The Effective Sampling Sizes (ESSs) for estimated parameters and convergence between runs was determined using Tracer (version 1.50). Species trees were resampled every 1×10^5 generations using LogCombiner and a maximum clade credibility tree (hereafter referred to simply as the species tree) was generated from the 8,000 sampled trees, as implemented in TreeAnnotator.

To estimate the timing of divergence, I used a molecular clock calibrated specifically for beetles with a rate of 0.0177 (per site per million year; Papadopoulou et al.

2010). While factors impacting estimates of absolute divergence times have been studied and discussed extensively (e.g., Arbogast et al. 2002), justification for the rate of substitution applied here include: (1) the estimated clock rate of COI is consistent across multiple lineages of beetles, (2) the geological events used for clock calibration have been well studied, and (3) the generation times and population sizes of taxa used to estimate the substitution rate are similar to our study system (see details in Papadopoulou et al. 2010). While some of the analyses and inferences depend upon the molecular clock, others are not (e.g., the estimation of speciation duration), which I highlight in the discussion.

Regional differentiation and dispersal

The IAA and nearby regions was partitioned into seven regions for estimating ancestral areas, dispersal, and extinction rates – specifically, India, Himalaya, Indo-China, Sundaland, Wallacea, Australasia, and the Philippines (Figs. 1 & 2). A dispersal, local extinction, and cladogenesis model (DEC) was estimated in the program LAGRANGE (ver. 20130526; Ree & Smith 2008) with the species tree using two different dispersal parameters to account for pronounced differences in the degree of forest fragmentation in the region during different geologic periods (Morley 2012). Specifically, a scaling factor of 0.5 on dispersal between adjacent regions was used to reflect the relative isolation of fragmented forests for the recent past, whereas no constraint on dispersal was used for the

period 5-20 MYA when a widely distributed forest is hypothesized to have covered the entire region during the Miocene Thermal Maximum (Morley 2012). Likewise, a constraint of 0 dispersal between India and Sundaland during the last 5 million years was used, in contrast to prior periods when a connection between Sundaland and India was possible (although debated; Hall 2002; Replumaz & Tapponnier 2003). No disjunct distributions were considered as potential ancestral ranges and a maximum of five geographic regions for the 7 regions studied were allowed for ancestral area reconstruction.

Diversification models

To test if diversification rates differed when divergence occurred across oceanic islands or within continents, a maximum likelihood method Geographic State Speciation and Extinction (GeoSSE; Goldberg et al. 2011) was used. Each taxon in the species tree was assigned a geographic state – island or continent; taxa from islands of the Sunda and the Sahul Shelves were assigned an intermediate state, because connections among these islands formed major landmasses throughout most of the Pleistocene (i.e., divergence may be driven by both island and continental states). Although our sampling included all current *Xylotrupes* taxa, I caution against the direct interpretation of estimated speciation and extinction rates because a significant rate shift can be detected (see the results section), and rate shifts may effect inferences from trait-dependent speciation model

(Rabosky & Goldberg 2015). Instead, I rely on inferences about the relative differences in estimated diversification rates to assess if speciation rate is statistically higher in one state than the other (i.e., across islands or within continents). In particular, the likelihood of a constraint model, which assumes no difference in speciation, extinction, and transition rates between states (i.e., islands versus continents), was compared with the likelihoods of additional models, which include free parameters to estimate state specific speciation and extinction rates, using likelihood ratio tests (calculations were made using the R package diversitree; FitzJohn 2012). Estimates of the probability distribution of speciation, extinction, and transition rates for each of the different states were also obtained from an un-constrained full parameterized model run for 1×10^5 generations of MCMC searches.

To test for a significant increase in speciation rate around 5 MYA (at the Pliocene-Miocene boundary; Lohman et al. 2011; Morley 2012), which is predicted if forest fragmentation promoted diversification, lineage through time (LTT) plots were estimated from all the post-burnin *Xylotrupes* species trees (8,000 trees) to accommodate phylogenetic uncertainty, and was implemented in the laser package in R (Rabosky 2006). The trees were fit to models assuming one constant versus two or three different diversification rates and the fit of the models were compared using AIC. In addition, maximum likelihood estimates of the point of a diversification rates shift (i.e., switch points; see Rabosky 2006) was calculated for models with multiple diversification rates, as was the probability of a specific rate shift at 5 MYA for only those trees that support a

rate switch model (because trees that fit a constant diversification rate model cannot be used to evaluate the probability of a rate shift).

Lastly, I tested if patterns of diversification where similar across all clades using estimates of lineage specific diversification rate and rate shifts as estimated using MEDUSA (Alfaro et al. 2009) and implemented in the R package geiger (version 2; Harmon et al. 2008). Specifically, a model assuming a single diversification rate across the species tree was fit and then stepwise comparisons between models assuming more and more complex patterns in rate shift(s) at different nodes and stems on the species tree were calculated and compared using AIC. The best-fit model was chosen based on an AIC threshold (see Alfaro et al. 2009).

Speciation duration

I estimated the speciation duration (the time required for an isolated lineage to become an independent species; Dynesius & Jansson 2014) under a protracted speciation model (Etienne & Rosindell 2012) for all taxa in the estimated species tree, except Indian and Himalayan taxa, which were excluded because only the IAA taxa are relevant to evaluating the role the speciation duration may play in mediating the effects of oceanic and forest barriers on *Xylotrupes* diversification dynamics. Specifically, the speciation initiation rate (how often are new isolated lineages formed), the speciation completion rate (how often the isolated lineages become evolutionarily independent species), and the

extinction rates of the incipient lineages and independent species were estimated and these four parameters were used to calculate the speciation duration. The analysis was performed using the R package PBD (Etienne et al. 2014). In our case it refers to an independently evolving lineage including both species and subspecies).

Estimates of population divergence

Pairwise F_{ST}-values were calculated using the program DnaSP (version 5.10.1; Librado & Rozas 2009) from the concatenated mitochondrial dataset (insufficient sequencing of multiple individuals across all populations precluded analysis of the two nuclear ITS regions). Average F_{ST} -values were calculated for the different oceanic barriers for the different types of landmasses, namely, oceanic island, temporary island (those that were connected to other landmasses during glacial periods), and continental areas. Analysis of variance (ANOVA; implemented in R) was also used to test for significant difference in F_{ST}-values among populations associated with the different landmass types, as was a pattern of IBD for each of the landmass types, separately, using linear regression (implemented in R with geographic distance calculated using Google Earth, version 6.2).

4.4 Results

Species-tree estimate

The estimated species tree shows a good correspondence with the current taxonomy of *Xylotrupes* based on morphological characters (Rowland 2003; 2011), except not all the species groups are monophyletic (Fig. 4.2). Although the lack of monophyly is relevant to future systematic investigations (the genus is under revision; J. M. Rowland personal communication), with respect to the analyses on the biogeographic history and diversification rates, the lack of monophyly does not necessarily represent a problem (i.e., analyses focus on geographic regions, as opposed to clade specific measures)
Posterior support is generally high among taxa within species groups, compared with moderate support for relationships among certain species groups (Fig. 4.2); note that phylogenetic uncertainty was accommodated by including all the post-burnin *Xylotrupes* species trees (i.e., 8,000 trees) in the LTT analyses.

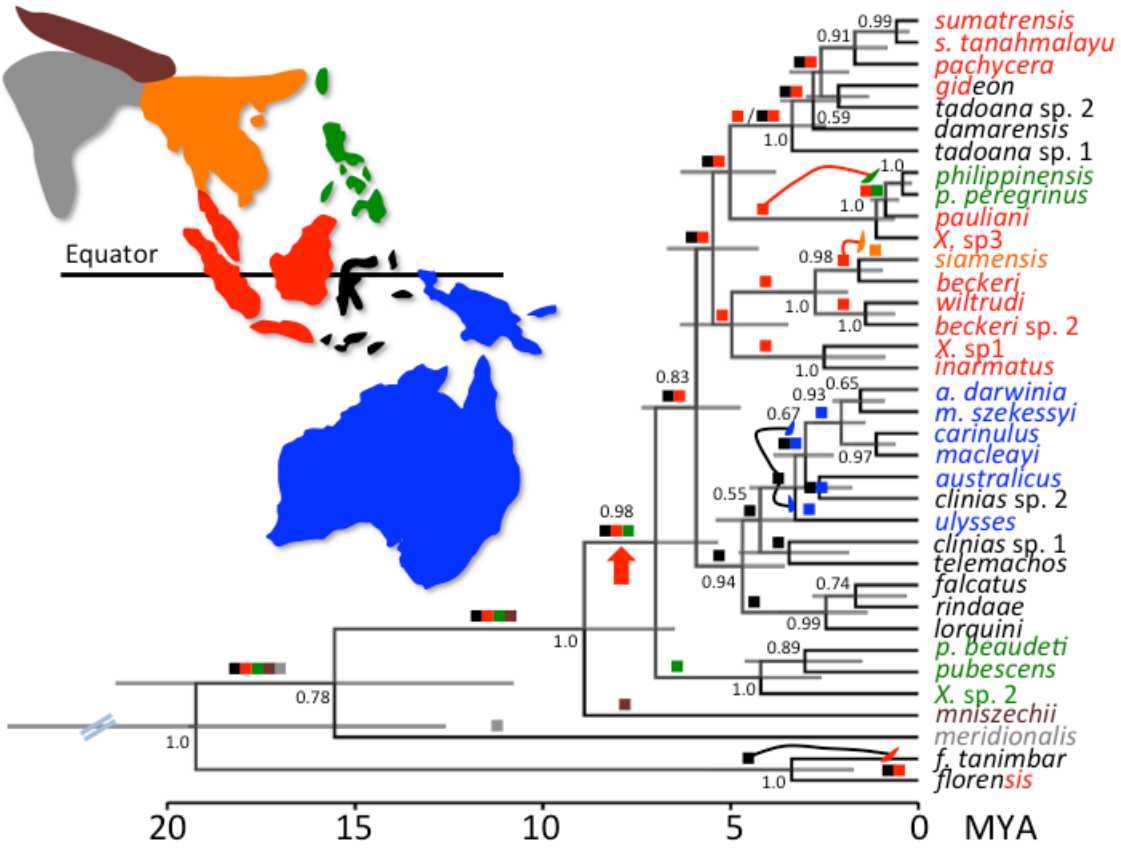

Fig. 4.2. Distribution of ancestral areas (shown above nodes and color coded by region) across the estimated species tree (with posterior probabilities given for nodes with greater than 50%); the names of *Xylotrupes* taxa color coded according to their distribution across regions (Wallacea in black, Sundaland in red, Australasia in blue, Indo-China in orange, the Philippines in green, India in grey, and the Himalayan region in brown; see Fig. 4.1 for names of species groups). Branches with multiple color squares represent composite ancestral areas and curved arrows mark inferred dispersal events. Estimated divergence times for the nodes, as well as 95% confidence intervals, for each node are shown, where the corresponding time of divergence in shown by the legend (in mya). A red arrow identifies the branch where a diversification rate shift is inferred using MEDUSA.

Regional differentiation and dispersal

The species tree revels clear geographic structuring of lineages, with taxa within a region generally being closely related and the DEC analyses show that vicariance events predominate the diversification history, with only two inferred dispersal events (Fig. 4.2). The estimated global dispersal and extinction rates are 0.01722 and 0.006541, respectively (-lnL = 54.98). In contrast to the recent diversification history, there is considerable uncertainty surrounding the ancestral state reconstruction during the initial diversification of the taxa (Fig. 4.2).

Changes in diversification rates through space and time

There was no detectable significant difference in diversification rate between lineages associated with and without oceanic barriers, given the similar likelihoods across models with constrained parameters (e.g., those equal speciation, extinction, and dispersal rates between geographic states) and those without constraints (Table 4.3). In fact, the estimated probability densities of speciation, extinction, and transition rates between geographic states overlap (Fig. 4.3).

Table 4.3. Comparison of the fit of different models from ML GeoSSE analyses. Specifically, models that assume equal or different speciation (λ), extinction (μ), and transition (δ) rates between geographic states are tested.

	d.f.	lnL	AIC	χ^2*	*P* value*
Full model	7	-123.28	260.57		
Constrained	3	-125.49	256.97	4.4012	0.35
Equal λ	6	-123.84	259.69	1.1162	0.29
Equal μ	6	-123.98	259.96	1.3888	0.24
Equal δ	6	-124.58	261.16	2.5851	0.11

*these values are derived by comparing to the full model.

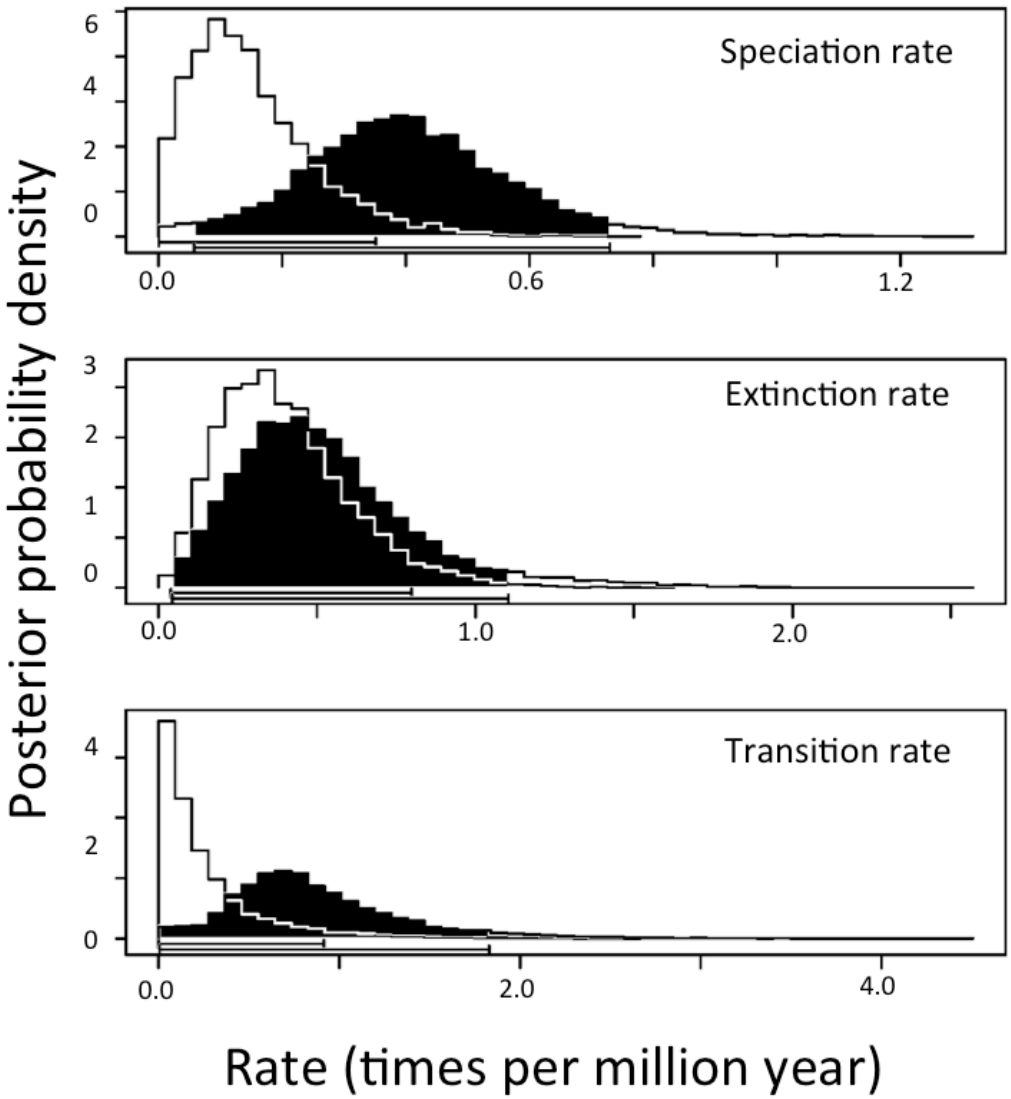

Fig. 4.3. Posterior probability densities of estimated diversification parameters (speciation rate, extinction rate, and transition rate, which refers to the frequency of transitions between geographic states) for the two geographic states – islands (shown in white) versus continents (shown in black) – using GeoSSE speciation

Estimates of the diversification rate from the LTT plots suggest that the diversification rate in *Xylotrupes* may have shifted, with a rate increase around the Miocene-Pliocene boundary (Fig. 4.4). Of the 8,000 post-burnin species trees analyzed, 172 trees (2.2%) favor a two-diversification-rates model and 7805 (97.5%) trees favor a

three-rates model. Only 23 trees (≈ 0.3%) fit a single-rate model. From the 172 trees favoring a two-rate diversification model, a decrease in diversification rate within the last 1 million years (mean = 0.79 and sd = 0.18) is suggested (Fig. 4.4), whereas for the trees supporting a three-rate diversification model, a decrease in diversification is estimated around 2 MYA (mean = 1.85 and sd = 1.18), and an increase in diversification rate about 6 MYA (mean = 5.82 and sd = 2.04) relative to the initial diversification rate of taxa in the region (Fig. 4.4). The probability of 5MYA as the time point of rate increase is 0.344 based on probability density estimated from the trees supporting a three rate diversification model.

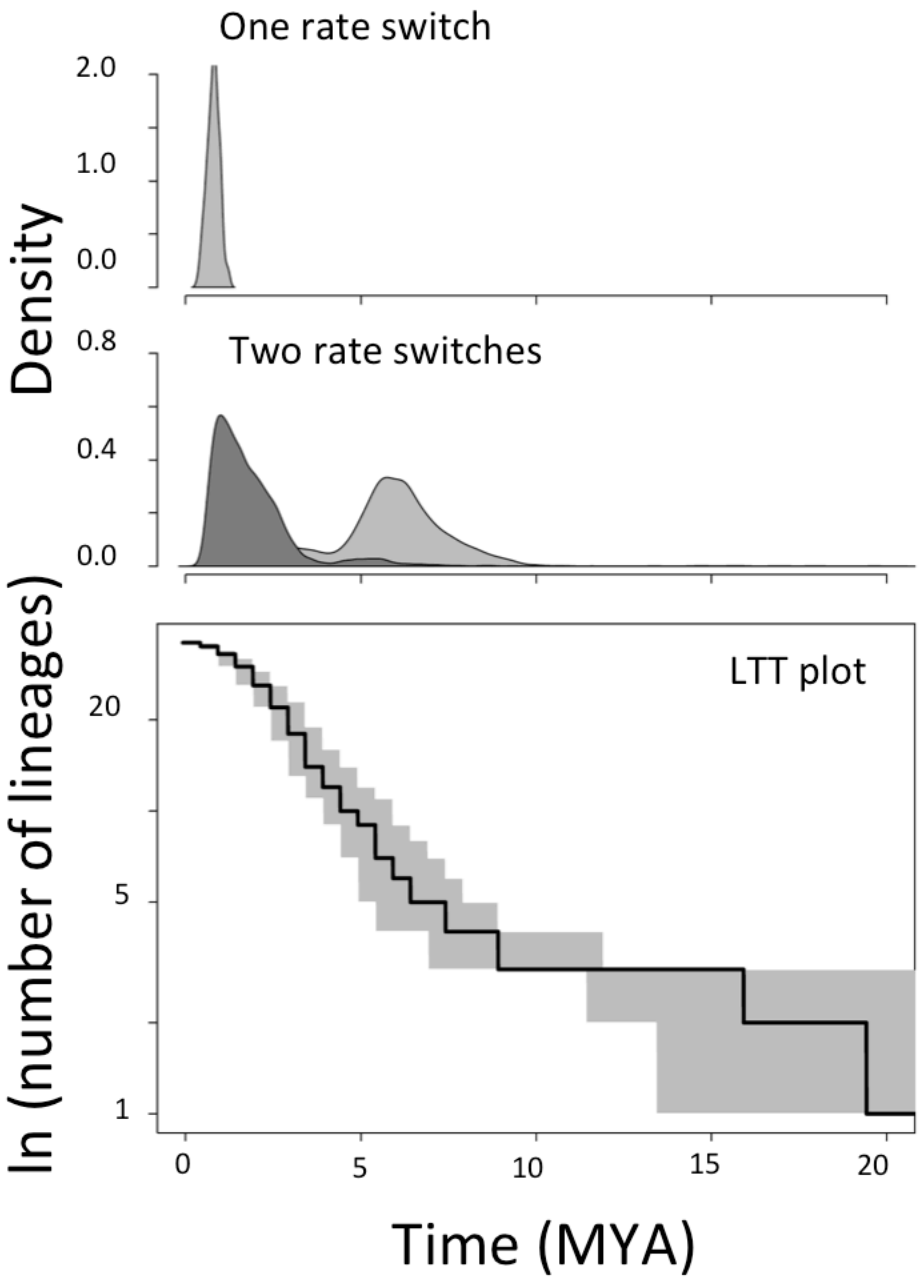

Fig. 4.4. Lineage through time plots (lower panel) and estimations of switch points between diversification rates under a model that favors two diversification rates (upper panel) and three diversification rates (middle panel). For the LTT plot (lower panel) the median (solid line) and 95% intervals (grey area) of LTT plot from 8,000 post-burnin species trees are shown.

The result from the MEDUSA analysis supports a significant shift in diversification rate. This rate shift model fits the species tree significantly better than a constant diversification rate model (lnL = -85.7751, AICc = 173.608; AICc threshold for a tree of 36 tips = 1.927). The diversification rate estimated from the best model before the rate shift is 0.2166046 and is 0.4389856 after the shift. The position of the rate shift on the branch leading to the common ancestor of the Pubescens, Philippinensis, Siamensis, Gideon, and Ulysses species groups (lnL = -78.6854, AICc = 163.729; Fig. 4.2), is consistent with the predicted pattern if climatic events were a major driver of diversification (i.e., the rate shift would be global and affect all contemporaneous lineages). Specifically, if climatic events were a major driver of diversification, then shifts in diversification rate should be global and observed across all contemporaneous lineages.

Speciation duration

The maximum likelihood estimates based on the protracted speciation model for the speciation initiation rate, the speciation completion rate, and the extinction rates of independent species and incipient lineages are 1.232, 0.379, 1.098, and 1.098, respectively (lnL = -72384). The expected duration of speciation calculated from the above parameters is 0.929.

Population subdivisions

The existence of an oceanic barrier is associated with a significantly higher pairwise F_{ST}-value (Table 4.4; Fig. 4.5A). This correlation is consistent regardless of different types of landmass. The effect of geographic distance on population subdivision is not significant for any of the three groups of geographic categories (F = 2.589, 0.262, and 3.677; P = 0.129, 0.629, and 0.071; and N = 17, 10, and 19 for continental, temporary island, and island, respectively) (Fig. 4.5B). The results presented here are robust to both logarithm and square root transformations of geographic distance (results not shown).

Table 4.4. Results from ANOVA for isolation by oceanic barrier. Specifically, the pairwise F_{ST}-values between populations isolated without oceanic barrier and those with oceanic barriers (both temporal and oceanic, see materials and methods) are compared.

	d.f.	S.S.	F	P value
groups	2	2.158	20.6	**4.09×10^{-7}**
Residuals	46	2.409		

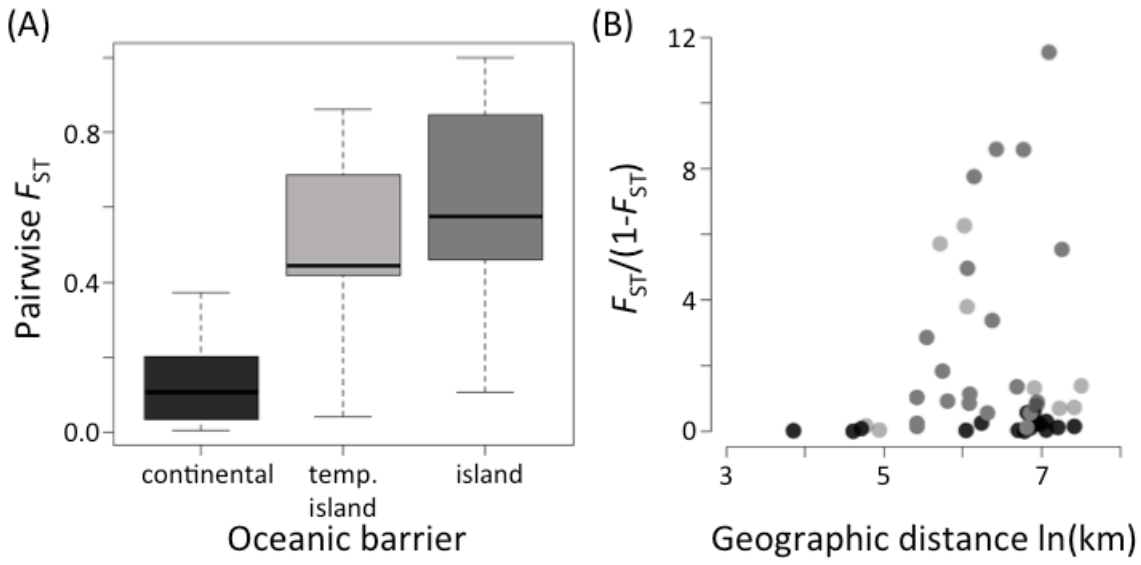

Fig. 4.5. Degree of population subdivision (as measured by F_{ST}) associated with (A) oceanic barriers (for each type of landmass; see methods for details) and (B) and geographic distance. F_{ST} values estimated between continental populations are shown in black, temporary island is in light grey, and island state is in dark grey.

4.5 Discussion

My study not only highlights that different types of geographic barrier may have their own biodiversity level specific effects, but also demonstrates how phylogeographic and biogeographic patterns can be studied simultaneously and how such integrative study can help us bridging the gap between microevolutionary and macroevolutionary studies. I show that evolutionary lineages in *Xylotrupes* beetles are geographically structured, where closely related taxa are often found in the same zoological region and dispersal events between regions are rare. Nevertheless, while an oceanic barrier can effectively structure populations into local genetic clusters (Table 4.4), the diversification rate in

Xylotrupes beetles is only correlated to different hypothesized levels of forest fragmentation across geological times (Fig. 4.4). I further reveal prolonged speciation duration estimated from the *Xylotrupes* species tree that uncovers an ephemeral, albeit strong, effect of oceanic barrier on genetic differentiation within zoological region. Below, I discuss why different types of geographic barrier may have different effects on structuring patterns of genetic diversity at different biodiversity levels and the implication from an integrative phylogeographic and biogeographic study.

Differences in the processes structuring population versus species-level divergence

My results clearly demonstrate that the effect of a barrier type is not consistent on patterns of genetic variation across multiple levels of biodiversity. The effect of an oceanic barrier on population subdivision is significant irrespective to different types of islands (i.e., temporary versus true oceanic islands; Fig. 4.5A). It has also been shown that distinct island forms, especially in male horn morphology, can rapidly evolve within the same *Xylotrupes* species (Rowland 2003; 2011). Together these findings indicate that isolation by an oceanic barrier can be a main mechanism promoting biological diversification as hypothesized in previous studies (Mayr 1942; Lohman et al. 2011; Wiens 2012). On the contrary, however, the GeoSEE results reveal that isolation by an

oceanic barrier is not correlated with a higher rate of biological diversification at species level (table 4.3; Fig. 4.3). Note that, although the BiSSE model family has been shown prone to result in false positive inference of trait-dependent speciation (Rabosky & Goldberg 2015), especially when there were significant rate shift in the past, my result of non-significant difference in speciation rate between geographic states is less likely to be affected.

The rate of species diversification in *Xylotrupes* beetles on the other hand is correlated with different hypothesized levels of forest fragmentation between geological times (Fig. 4.4). This pattern is also found in many other studies using forest associated species (e.g., Bendiksby et al. 2010), where an increase in diversification rate is associated with a geological time period that drastic decrease in forest distribution in IAA is hypothesized (Morley 2012). My results not only significantly support an increase in diversification rate from more than 95% of the post burnin species trees, but also confidently place the time of rate increase at the Miocene-Pliocene boundary (Fig. 4.4). The increase in diversification rate however is only shared among the five most speciose groups of *Xylotrupes* beetles from IAA based on our MEDUSA result (Fig. 4.2), which implies a major effect of forest fragmentation specific to lineages from IAA. This lineage specific rate shift can be due to the specific geographic and geological features of IAA (Lohman et al. 2011) – i.e., it might not be a global event. Additionally, life history features such as the evolution of a shorter generation time or a smaller population size due to smaller habitat size (e.g., islands) may also contribute to this pattern of lineage

specific rate shift. Unfortunately, based on my current data and limited information about the natural history of *Xylotrupes* beetles, the relative contribution between climate and geography and the evolution in life history traits on lineage specific diversification rates can not be properly evaluated.

The Indo-Australian Archipelago has conventionally been subdivided into multiple zoological regions because of distinct faunal communities (Lohman et al. 2011), which are delineated by oceanic trenches. My results support this long history in systematic studies by showing that species from the same zoological region are more closely related to one another than species from distant regions (Fig. 4.2; except for those from Wallacea, which appears to accommodate taxa from multiple distinct lineages). Vicariance and rare dispersal events out of Sundaland and Wallacea can result in the formation of independently evolving lineages that will be isolated from its sister lineage for millennia because landbridges have rarely been formed between these zoological regions. My results here indicate that even for the same type of geographic barrier, different effects on phylogeographic and biogeographic patterns can result because of differences in the evolutionary time scale that different barriers can endure – i.e., oceanic barriers that separate islands versus those delineating zoological regions.

The duration of speciation as a filter for the processes that contribute to species diversity patterns

I demonstrate in the previous section that oceanic barrier between landmasses can promote population subdivision between oceanic islands and lead to faunal community differentiation between zoological regions, but does not correlate with the proliferation of new species. It has been hypothesized that the persistence of an isolated population that may become a new species can be determined by the rate of extinction and the frequency of secondary contact (Dynesius & Jansson 2014). Since my GeoSSE result shows that extinction rates are similar between lineages of island and continental states (table 4.3; Fig. 4.3), the complete formation of a new species is likely controlled by the duration of the physical barrier. In my study system, lowered sea level during glacial periods could result in frequent secondary contacts between island populations (Lohman et al. 2011). For example, the relative time for islands being disconnected in the Sunda Shelf is short comparing to the time of being connected to other larger landmasses during Pleistocene, because a large proportion of the Sunda Shelf is only 50 meters below current sea level (Lohman et al. 2011; Hall 2012). Since the estimated speciation duration based on the protracted speciation model is around one myr, the isolation time provided by sea level rise during interglacial periods of Pleistocene (usually tens of or maybe hundreds of thousands of years) may be insufficient for recent divergence to be completed.

Additionally, the time required for evolving coexistence in sympatry can also has significant implication on current biodiversity patterns. As expected for speciation driven predominantly by geographic isolation, closely related *Xylotrupes* taxa do not coexist on the same island. For example, on Flores Island there are *X. florensis* and *X. tadoana* that have been diverged from each other for more than 19 myrs based on our reconstructed species tree; surrounding the Cameron Highland of Malay Peninsular there are *X. sumatrensis*, *X. beckeri*, and *X. pauliani*, which are species from three different species groups diverged from one another for at least 5 myrs. In other words, I find that the minimum time for evolving coexistence between *Xylotrupes* species can be as long as 5 myrs. A recent decrease in diversification rate revealed from 99% of the post burnin species trees from our LTT results (Fig. 4.4) further support the importance of the evolution of coexistence. Specifically, this decrease in diversification rate may reflect that the available niches for evolving new species have been depleted (Rabosky 2013). *Xylotrupes* beetles that feed on all kinds of fruit plant as well as native tree species and are commonly recognized as rampant plant pests (Firake et al. 2013) are expected to exhibit little niche partitions between species. Since almost every island, no matter how small the island size, from IAA has been occupied by species of *Xylotrupes* beetles (Rowland 2003) the recent decrease in diversification rate implies that the current species number of *Xylotrupes* beetles may have reached the carrying capacity (Rabosky 2013). The fact that incipient species (Mayr 1942) can only be found on different oceanic islands from our results supports the idea that strong physical barrier, e.g, oceanic barrier, can

maintain the genetic differentiation between recently diverged evolutionarily independent lineages. On the other hand, my results also reveal that divergence initiated by physical barrier without ecological/biological differentiation can hardly be completed once the barrier no longer exists.

Implications from an integrative phylogeographic and biogeographic study

The transition between microevolutionary and macroevolutionary patterns has been a study focus in evolutionary biology (Etienne & Rosindell 2012) – i.e., if these patterns are results of the same process, but differ in the evolutionary time that process has proceeded. For example, if there are fundamental differences between populations and biological species (Mallet 2008). The Darwinian view of species implies that population subdivision is the initial phase of speciation (Mallet 2008), but Mayr's view of species hypothesizes that a distinct process, i.e., the evolution of reproductive isolation, separates species level diversification from population level subdivision. My results imply that an additional filter may exist (speciation duration) that determines how many of the structured populations can become distinct biological species (as discussed in the previous section). Therefore, species and populations can be fundamentally different evolutionary hierarchies in the *Xylotrupes* beetle system.

My findings from an integrative phylogeographic and biogeographic study not only have significant implications on philosophical arguments such as the population-species relationship, but also empirical practices such as species delimitation and conservation. Island endemic taxa, populations or species, often exhibit distinct genetic property that can easily be distinguished from other such taxa (e.g., based on a monophyletic mitochondrial gene tree). Island populations in addition to species can be recognized as distinct species with very high statistical supports when molecular species delimitation methods are applied (e.g., GMYC; Papadopoulou et al. 2008). I caution against such taxonomic practice, which can obscures biodiversity patterns governed by different evolutionary processes. Additionally, detrimental effects on biological conservation can result because of taxonomic over splitting (see Frankham et al. 2012 and Zachos 2013).

4.6 Conclusion

I demonstrate the apparent biodiversity level dependent effects of different biogeographic processes on structuring genetic diversity. Specifically, while population subdivision is evidently enhanced by the existence of oceanic barrier, the rate for the proliferation of different species is only correlated to the hypothesized level of forest fragmentation. However, oceanic barriers that delineate zoological regions, which are effective and persistent, result in regional biological communities. I uncover a great merit to study intraspecific and interspecific genetic diversities in the same system when testing

phylogeographic and biogeographic hypotheses that although can be achieved by many studies has seldom been applied. Future studies focusing on bridging the gap between mechanisms for microevolutionary and macroevolutionary patterns are needed to understand how biodiversity is geographically generated and structured.

CHAPTER 5: Conclusion

My thesis provides an integrative way of combining information from multiple data types to investigate the consistency in defining different biological hierarchies, and the evolutionary and biological differences between the levels of biodiversity. My work also bridges micro- and macroevolutionary study by testing whether the same processes drive patterns at different levels of biodiversity. The major implications drawn from my study are that (i) species designations can be statistically evaluated and consistently applied across multiple evolutionary lineages, and (ii) such statistical evaluation is important because population versus species levels of divergence, at least in the rhinoceros beetle systems, can be generated and maintained by different processes. This finding reinforces the general (and debated) view that populations and biological species are fundamentally different evolutionary units.

My results have broader significance on conservation, systematics and biogeography, and evolutionary biology in general. For example, geographic overlap occurs between certain independently evolving lineages, where their divergence times can be used as a yard-stick for species level divergence and are strong evidence in supporting the merit of naming and protecting allospecies. On the other hand, my finding cautions against the

naming of allospecies based solely on molecular data. For example, island endemic *Xylotrupes* taxa may acquire genetic distinctiveness, but such divergence, can be erased easily when the physical barrier no longer exists.

For systematics and biogeography, my study results imply that different practice in defining operational taxonomic units (OTUs) and the resulted differences in evolutionary inferences (e.g., whether ecological differentiation is viewed as diversification driver) bare evolutionary significances. For example, if geographic populations are sampled as OTUs, the historical processes that lead to the divergences of recently formed taxa, which are microevolutionary processes, will be revealed as diversification drivers. On the other hand, if only distinct biological species are represented as OTUs, only those mechanisms that pertain to macroevolutionary processes will be recognized as diversification drivers. It is therefore critical to understand what level of diversity pattern is the focus (used as OTU) of a study and compare only results from using the same level of biodiversity as OTUs between studies.

Finally, my study results adhere to the classic view of speciation in evolutionary biology, where species and populations are fundamentally different evolutionary hierarchies (Mayr 1942). Speciation is decoupled from adaptation and neutral population subdivision – i.e., they can be driven and maintained by different evolutionary processes. In some cases, mechanisms that are responsible for speciation and population subdivision can be the same (e.g., distinct female mate choices lead to reproductive isolation between color morphs/species in lake Victoria cichlids), but it does not always have to be the case.

Some contemporary theories utilize mechanisms underlying population subdivision to explain species diversity and to test macroevolutionary patterns (e.g., speciation rate depends on the evolution of new traits or population sizes). However, although these mechanisms can indeed facilitate the emergence of new evolutionary entities (populations and local forms – whatever these taxa are called), how do these mechanisms help to complete the speciation process is still poorly known (e.g., Nosil et al. 2009). My study findings suggest that extrapolations between micro- and macroevolutionary patterns should be undertaken with caution. Furthermore, by recognizing structured populations as species at the early stage of speciation, the definition of species is arbitrary – species are simply populations that have proceeded further along the speciation continuum. Additionally, if structured populations are species, then genetic exchange between species or during the process of speciation (e.g., speciation with gene flow and introgression), and hybridization between species can become a widespread and common phenomenon. These patterns could have been recognized as gene flows between populations if species are defined differently.

BIBLIOGRAPHY

Adams D.C., Otárola-Castillo E. 2013. Geomorph: an R package for the collection and analysis of geometric morphometric shape data. Methods in Ecology and Evolution, 4: 393-399.

Alfaro M.E., Santini F., Brock C.D., 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 Science, USA, 106: 13410-13414.

Ali J.R. 2012. Colonizing the Caribbean: is the GAARlandia land-bridge hypothesis gaining a foothold? Journal of Biogeography 39: 431-433.

Alvarado-Serrano D.F., Knowles L.L. 2014. Ecological niche models in phylogeographic studies: applications, advances and precautions. Molecular Ecology Resources, 14: 233-248.

Arbogast B.Sl., Edwards S.V., Wakeley J., Beerli P., Slowinski J.B. 2002. Estimating divergence times from molecular data on phylogenetic and population genetic timescales. Annual Reviews in Ecology and Systematics, 33: 707-740.

Ayres D.L., Darling A., Zwickl D.J., Beerli P., Holder M.T., Lewis P.O., Huelsenbeck J.P., Ronquist F., Swofford D.L., Cummings M.P., Rambaut A., Suchard M.A. 2012. BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics. Systematic Biology, 61: 170-173.

Bacon C.D., Mora A., Wagner W.L., Jaramillo C.A. 2013. Testing geological models of evolution of the Isthmus of Panama in a phylogenetic framework. Botanical Journal of the Linean Society 171: 287-300.

Bacon C.D., Silvestro D., Jaramillo C., Smith B.T., Chakrabarty P., Antonelli A. 2015. Biological evidence supports an early and complex emergence of the Isthmus of Panama. Proceedings of the National Academy of Sciences USA 112: 6110-6115.

Bell K.L., Yeates D.K., Moritz C., Monteith G.B. 2004. Molecular phylogeny and biogeography of the dung beetle genus *Temnoplectron* Westwood (Scarabaeidae: Scarabaeinae) from Australia's wet tropics. Molecular Phylogenetics and Evolution, 31: 741-753.

Bendiksby M., Schumacher T., Gusarova G., Nais J., Mat-Salleh K., Sofiyanti N., Madulid D., Smith S.A., Barkman T. 2010. Elucidating the evolutionary history of the Southeast Asian, holoparasitic, giant-flowered Rafflesiaceae: Pliocene vicariance, morphological convergence and character displacement. Molecular Phylogenetics and Evolution, 57: 620-633.

Bond J.E., Hedin M.C., Ramirez M.G., Opell B.D. 2001. Deep molecular divergence in the absence of morphological and ecological change in the Californian coastal dune endemic trapdoor spider *Aptostichus simus*. Molecular Ecology, 10: 899-910.

Bouckaert R., Heled J., Kühnert D., Vaughan T.G., Wu C-H., Xie D., Suchard M.A., Rambaut A., Drummond A.J. 2014. BEAST2: a software platform for Bayesian evolutionary analysis. PLoS Computational Biology, 10: e1003537.

Brower A.V.Z. 1994. Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. Proceedings of the National Academy of Sciences, USA, 91: 6491-6495.

Camargo A., Morando M., Avila L.J., Sites Jr. J. W. 2012. Species delimitation with ABC and other coalescent-based methods: a test of accuracy with simulations and an empirical example with lizards of the *Liolaemus darwini* complex (Squamata: Liolaemidae). Systematic Biology, 66: 2834-2849.

Carstens B.C., Knowles L.L. 2007. Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from *Melanoplus* grasshoppers. Systematic Biology, 56: 400-411.

Carstens B.C., Pelletier T.A., Reid N.M., Satler J.D. 2013. How to fail at species delimitation. Molecular Ecology, 22: 4369-4383.

Chalumeau F., Reid W. 2002. Aperçus sur le complexe *hercules* et statut du *Dynastes alcides* (Coleoptera, Dynastidae). Nouvelle Revue d'Entomologie, 19: 83-91.

Cody S., Richardson J.E., Rull V., Ellis C., Pennington R.T. 2010. The Great American Biotic Interchange revisited. Ecography 33: 326-332.

Colgan D.J., McLauchian A., Wilson G.D.F., Livingston S., Macaranas J. Edgecombe G.D., Gassis G., Gray M.R. 1998. Molecular phylogenetics of the Arthropoda: relationships based on histone H3 and U2 snRNA DNA sequences. Australian Journal of Zoology, 46: 419-437.

Coulson S.J., Hodkinson I.D., Webb N.R., Harrison J.A. 2002. Survival of terrestrial soil-dwelling arthropods on and in seawater: implications for trans-oceanic dispersal. Functional Ecology 16: 363-356.

Coyne J.A., Orr H.A. 2004. Speciation. Sinaur Associates, MA.

Davis M.P., Midford P.E., Maddison W. 2013. Exploring power and parameter estimation of the BiSSE method for analyzing species diversification. BMC Evolutionary Biology, 13: 38.

Darriba D., Taboada G.L., Doallo R., Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nature Methods, 9: 772.

Dayrat B. 2005. Towards integrative taxonomy. Biological Journal of the Linnean Society, 85: 407-415.

Drummond A.J., Suchard M.A., Xie D., Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology and Evolution, 29: 1969-1973.

Dutrillaux B., Dutrillaux A.M. 2013. A South American origin of the genus *Dynastes* (Coleoptera: Scarabaeidae: Dynastinae) demonstrated by chromosomal analyses. Cytogenetic and Genome Research, 141:37-42.

Dynesius M., Jansson R. 2014. Persistence of within-species lineages: a neglected control of speciation rates. Evolution, 68: 923-934.

Edgar R.C. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics, 5: 113.

Edwards D.L., Knowles L.L. 2014. Species detection and individual assignment in species delimitation: can integrative data increase efficacy? Proceedings of the Royal Society B: Biological Sciences, 281: 20132765.

Elith J., Kearney M., Philips S. 2010. The art of modeling range-shifting species. Methods in Ecology and Evolution, 1: 330-342.

Emlen D.J., Marangelo J., Bell B., Cunningham C.W. 2005. Diversity in the weapons of sexual selection: horn evolution in the beetle genus *Onthophagus* (Coleoptera: Scarabaeidae). Evolution, 59: 1060-1084.

Esselstyn J.A., Evans B.J., Sedlock J.L., Khan F.A.A., Heaney L.R. 2012. Single-locus species delimitation: a test of the mixed Yule-coalescent model, with an empirical application to Philippines round-leaf bats. Proceedings of the Royal Society, B, 279: 3678-3686.

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: 204-213.

Etienne R.S., Morlon H., Lambert A. 2014. Estimating the duration of speciation from phylogenies. Evolution, 68: 2430-2440.

Firake D.M., Deshmukh N.A., Behere G.T., Azad Thakur N.S., Ngachan S.V., Rowland J.M. 2013. First report of elephant beetles in the genus *Xylotrupes* Hope (Coleoptera: Scarabaeidae) attacking guava. The Coleopterists Bulletin, 67: 608-610.

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

Folmer O., Black M., Hoeh W., Lutz R., Vrijenhoek R. 1994. DNA primers for amplication of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology, 3: 294-299.

Frankham R., Ballou J.D., Dudash M.R., Eldridge M.D.B., Fenster C.B., Lacy R.C., Mendelson III J.R., Porton I.J., Ralls K., Ryder O.A. 2012. Implications of different species concepts for conserving biodiversity. Biological Conservation, 153: 25-31.

Fujita M.K., Leaché A.D., Burbrink F.T., McGuire J.A., Moritz C. 2012. Coalescent-based species delimitation in an integrative taxonomy. Trends in Ecology and Evolution, 27: 480-488.

Garzón-Orduña I.J., Benetti-Longhini J.E., Brower A.V.Z. 2014. Timing the diversification of the Amazonian biota: butterfly divergences are consistent with Pleistocene refugia. Journal of Biogeography 41: 1631-1638.

Gillespie R.G., Baldwin B.G., Waters J.M., Fraser C.I., Nikula R., Roderick G.K. 2012. Long-distance dispersal: a framework for hypothesis testing. Trends in Ecology and Evolution 27: 47-56.

Glor R.E., Warren D. 2010. Testing ecological explanations for biogeographic boundaries. Evolution, 65: 673-683.

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.

Gouy M., Guindon S., Gascuel O. 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Molecular Biology and Evolution, 27: 221-224.

Grossi E., Arnaud P. 1993. Description d'une nouvelle sous-espece de *Dynastes hercules*. Bulletin de la Société Sciences Nat, 78: 13-14.

Hall R. 2002. Cenozoic geological and plate tectonic evolution of SE Asia and the SW Pacific: computer-based reconstructions, models and animations. Journal of Asian Earth Sciences, 20: 353-431.

Hall R. 2012. Sundaland and Wallacea: geology, plate tectonics and palaeogeography. *Biotic Evolution and Environmental Change in Southeast Asia* (ed. by D.J. Gower, J.E. Richard, B.R. Rosen, L. Rüber & S.T. Williams), pp. 32-78. Cambridge University Press, UK.

Harmon L., Weir J.T., Brock C.D., Glor R.E., Challenger W. 2008. GEIGER: investigating evolutionary radiations. Bioinformatics, 24: 129-131.

Hebert P.D.N., Stoeckle M.Y., Zemlak T.S., Francis C.M. 2004a. Identification of birds through DNA barcodes. PLoS Biology, 2: e312.

Hebert P.D.N., Penton E.H., Burns J.M., Janzen D.H., Hallwachs W. 2004b. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proceedings of the National Academy of Sciences, USA, 101: 14812-14817.

Heled J., Drummond A.J. 2010. Bayesian inference of species trees from multilocus data. Molecular Biology and Evolution, 27: 570-580.

Hoorn C., Wesselingh F.P., ter Steege H., Bermudez M.A., Mora A., Sevink J., Sanmartín I., Sanchez-Meseguer A., Anderson C.L., Figueiredo J.P., Jaramillo C., Riff D., Negri F.R., Hooghiemstra H., Lundberg J., Stadler T., Särkinen T., Antonelli A.. 2010. Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. Science 330: 927-931.

Howard D.J., Berlocher S.H. 1998. Endless Forms: Species and Speciation. Oxford University Press, Oxford.

Huang H., He Q., Kubatko L.S., Knowles L.L. 2010. Sources of error for species-tree estimation: Impact of mutational and coalescent effects on accuracy and implications for choosing among different methods. Systematic Biology, 59: 573-583.

Huang J.-P. 2014. Évaluation génétique du statut taxonomique des populations de dynaste hercule des Petites Antilles (Coleoptera, Dynastinae). Coléoptères des Petites Antilles Tome 2: 30-36.

Huang J-P. 2015. Hercules Beetles (Genus *Dynastes*, Dynastidae): a revisionary study based on molecular, morphological, ecological differences, and geographic distribution. University of Michigan Museum of Zoology Miscellaneous Publications, in review.

Huang J.-P., Knowles L.L. 2015. The species versus subspecies conundrum: quantitative delimitation from integrating multiple data types within a single Bayesian approach in Hercules beetles. Systematic Biology in press.

Hudson, R. R., Coyne J. A. 2002. Mathematical consequences of the genealogical species concept. Evolution, 56:1557-1565.

Hwang, S-M-R. 2011. The Dynastini of the World. Nature & Ecology, Seoul, Korea.

Isaac N.J.B., Mallet J., Mace G.M. 2004. Taxonomic inflation: its influence on macroecology and conservation. Trends in Ecology and Evolution, 19: 464-469.

Jarman G.M., Hinton H.E. 1974. Some defence mechanisms of the Hercules beetle, *Dynastes hercules*. Journal of Entomology Series A, General Entomology, 49: 71-80.

Jenkins T.M., Jones S.C., Lee C-Y., Forshler B.T., Chen Z., Lopez-Martinez G., Gallagher N.T., Green G., Neal M., Thistleton B., Kleinschmidt S., Harvey C. 2007. Phylogeography used to illuminate maternal origins of exotic invasions of *Coptotermes gestori* (Isoptera: Rhinotermitidae). Molecular Phylogenetics and Evolution, 42: 612-621.

Knowles L.L. 2010. Sampling strategies for species-tree estimation. In: Estimating Species Trees: Practical and Theoretical Aspects (L. L. Knowles and L. S. Kubatko, eds.), pp. 163-172. Wiley-Blackwell.

Knowles L.L., Carstens B.C. 2007. Delimiting species without monophyletic gene trees. Systematic Biology, 56:887-895.

Knowles, L.L., Massatti R., He Q., Olson L.E., Lanier H.C. 2015. Quantifying the similarity between genes and geography across Alaska's alpine small mammals. Journal of Biogeography, in revision.

Lai J., Ko H-P. 2008. For the Love of Rhinoceros and Stag Beetles. 2nd Ed., Morning Star Publisher, Taipei, TW.

Lanier H.C., Huang H., Knowles L.L. 2014. How low can you go? The effects of mutation rate on the accuracy of species-tree estimation. Molecular Phylogenetics and Evolution, 70: 112-119.

Leigh E.G., O'Dea A., Vermeij G.J. 2014. Historical biogeography of the Isthmus of Panama. Biological Reviews 89: 148-172.

Librado P., Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics, 25: 1451-1452.

Lohman D.J., de Bruyn M., Page T., von Rintelen K., Hall R., Ng P.K.L., Shih H-T., Carvalho G.R., von Rintelen T. 2011. Biogeography of the Indo-Australian Archipelago. Annual Review of Ecology, Evolution, and Systematics, 42: 205-226.

Maddison W.P., Maddison D.R. 2011. Mesquite: a modular system for evolutionary analysis. Version 2.75 http://mesquiteproject.org

Malaney J.L., Cook J.A. 2013. Using biogeographical history to inform conservation: the case of Preble's meadow jumping mouse. Molecular Ecology, 22: 6000-6017.

Mallet J. 2008. Mayr's view of Darwin: was Darwin wrong about speciation? Biological Journal of the Linnean Society, 95: 3-16.

Marshall L.G. 1988. Land mammals and the Great American Interchange. American Scientist 76: 380-388.

Mayr E. 1942. Systematics and the Origin of Species from the Viewpoint of a Zoologist. Columbia University Press, NY.

Mayr E. 1963. Animal Species and Evolution. Harvard University Press, MA.

Mayr E., Ashlock P. D. 1991. Principles of Systematic Zoology. 2nd Ed., McGraw-Hill, inc., NY.

McCormack J.E., Huang H., Knowles L.L. 2009. Maximum-likelihood estimates of species trees: how accuracy of phylogenetic inference depends upon the divergence history and sampling design. Systematic Biology, 58: 501-508.

McCormack J.E., Zellmer A.J., Knowles L.L. 2010. Does niche divergence accompany allopatric divergence in Aphelocoma jays as predicted under ecological speciation? Insights from tests with niche models. Evolution, 64: 1231-1244.

McKay B.D., Mays Jr. H.L., Wu Y., Li H., Yao C-T., Nishiumi I., Zou F. 2013. An empirical comparison of character-based and coalescent-based approaches to species delimitation in a young avian complex. Molecular Ecology, 22: 4943-4957.

Morley R.J. 2012. A review of the Cenozoic palaeoclimate history of Southeast Asia. *Biotic Evolution and Environmental Change in Southeast Asia* (ed. by D.J. Gower, J.E. Richard, B.R. Rosen, L. Rüber & S.T. Williams), pp. 79-114. Cambridge University Press, UK.

Morón M.Á. 2009. El género *Dynastes* MacLeay, 1819 en la zona de transición Mexicana (Coleoptera: Melolonthidae: Dynastinae). Boletín Sociedad Entomológica Aragonesa, 45: 23-38.

Morrone J.J. 2006. Biogeographic areas and transition zones of Latin America and the Caribbean islands based on panbiogeographic and cladistics analyses of the entomofauna. Annual Review of Entomology 51: 467-494.

Nosil P. 2008. Speciation with gene flow could be common. Molecular Ecology, 17: 2103-2106.

Nosil P., Harmon L.J., Seehausen O. 2009. Ecological explanations for (incomplete) speciation. Trends in Ecology and Evolution, 24: 145-156.

Olave M., Solà E., Knowles L.L. 2014. Upstream analyses create problems with DNA-based species delimitation. Systematic Biology, 63: 263-271.

Papadopoulou A., Bergsten J., Fujisawa T., Monaghan M.T., Barraclough T.G., Vogler A.P. 2008. Speciation and DNA barcodes: testing the effects of dispersal on the formation of discrete sequence clusters. Philosophical Transactions of the Royal Society, B: Biological Sciences, 363: 2987-2996.

Papadopoulou A., Knowles L.L. 2015. Genomic tests of the species-pump hypothesis: recent island connectivity cycles drive population divergence but not speciation in Caribbean crickets across the Virgin Islands. Evolution, 69: 1501-1517.

Papadopoulou A., Knowles L.L. 2016. Refined hypotheses based on taxon-specific traits in comparative phylogeography. Proceedings of the National Academy Sciences, USA, in review.

Papadopoulou A., Anastasiou I., Vogler A.P. 2010. Revisiting the insect mitochondrial molecular clock: the mid-Aegean Trench calibration. Molecular Biology and Evolution, 27: 1659-1672.

Pons J., Barraclough T.G., Gomez-Zurita J., Cardoso A., Duran D.P., Hazell S., Kamoun S., Sumlin W.D., Vogler A.P. 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Systematic Biology, 55: 595-609.

Pontarp M., Ripa J., Lundberg P. 2015. The biogeography of adaptive radiations and the geographic overlap of sister species. The American Naturalist, 186: 565-581.

Pyron R.A., Hsieh F.W., Lemmon A.R., Lemmon E.M., Hendry C.R. 2016. Integrating phylogenomic and morphological data to assess candidate species-delimitation models in brown and red-bellied snakes (*Storeria*). Zoological Journal of the Linnean Society, advance access: doi: 10.1111/zoj.12392.

de Queiroz K. 2007. Species concepts and species delimitation. Systematic Biology, 56: 879-886.

Rabosky D.L. 2006. LASER: a maximum likelihood toolkit for detecting temporal shifts in diversification rates from molecular phylogenies. Evolutionary Bioinformatics Online, 2: 247-250.

Rabosky D.L. 2010. Extinction rates should not be estimated from molecular phylogenies. Evolution 64: 1816-1824.

Rabosky D.L. 2013. Diversity-dependence, ecological speciation, and the role of competition in macroevolution. Annual Review of Ecology, Evolution, and Systematics, 44: 481-502.

Rabosky D.L. 2014. Automatic detection of key innovations, rate shifts, and diversity dependence on phylogenetic trees. PLoS One 9: e89543.

Rabosky D.L., Donnellan S.C., Grundler M., Lovette I.J. 2014. Analysis and visualization of complex macroevolutionary dynamics: an example from Australian Scincid Lizards. Systematic Biology 63: 610-627.

Rabosky D.L., Goldberg E.E. 2015. Model inadequacy and mistaken inferences of trait-dependent speciation. Systematic Biology 64: 340-355.

Rabosky D.L., Huang H. 2015. A robust semi-parametric test for detecting trait-dependent diversification. Systematic Biology in press.

Ratcliffe B.C. 2003. The Dynastine Scarab Beetles of Costa Rica and Panama. Bulletin UNSM, Vol. 16, Lincoln, NE.

Ratcliffe B.C., Cave R.D. 2006. The Dynastine Scarab Beetles of Honduras, Nicaragua and El Salvador. Bulletin UNSM, Vol. 21, Lincoln, NE.

Ratcliffe B.C., Cave R.D., Cano E.B. 2013. The Dynastine Scarab Beetles of Mexico, Guatemala, and Belize. Bulletin UNSM, Vol. 27, Lincoln, NE.

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.

Replumaz A., Tapponnier P. 2003. Reconstruction of the deformed collision between India and Asia by backward motion of lithospheric blocks. Journal of Geophysical Research, 108: 2285.

Ribeiro M.C., Metzger J.P., Martensen A.C., Ponzoni F.J., Hirota M.M. 2009. The Brazilian Atlantic Forest: how much is left, and how is the remaining forest distributed? Implications for conservation. Biological Conservation, 142: 1141-1153.

Richards N.K., Glare T.R., Hall D.C.A. 1997. Genetic variation in grass grub, *Costelytra zealandica*, from several regions. Proceedings of the 50th New Zealand Plant Protection Conference: 338-343.

Rohlf F.J., Slice D. 1990. Extensions of the Procrustes method for the optimal superimposition of landmarks. Systematic Zoology, 39: 40-59.

Rowland J.M. 2003. Male horn dimorphism, phylogeny and systematics of rhinoceros beetles of the genus *Xylotrupes* (Scarabaeidae, Coleoptera). Australian Journal of Zoology, 51: 213-258.

Rowland J.M. 2011. Notes on nomenclature in *Xylotrupes* Hope (Scarabaeidae: Dynastinae: Dynastini). Insecta Mundi, 0176: 1-10.

Rowland J.M., Miller K.B. 2012. Phylogeny and systematics of the giant rhinoceros beetles (Scarabaeidae: Dynastini). Insecta Mundi, 0236: 1-15.

Simon C., Frati F., Beckenback A., Crespi B., Liu H., Flook P. 1994. Evolution, weighting and phylogenetic utility of mitochondrial gene sequences and compilation

of conserved polymerase chain reaction primers. Entomological Society of America, 87: 651-701.

Simpson G.G. 1950. History of the fauna of Latin America. American Scientist 38: 361-389.

Smith R.T., Ribas C.C., Whitney B.M., Hernández-Baños B.E., Klicka J. 2013. Identifying biases at different spatial and temporal scales of diversification: a case study in the Neotropical parrotlet genus *Forpus*. Molecular Ecology, 22: 483-494.

Solis-Lemus C., Knowles L.L., Ané C. 2015. Bayesian species delimitation combining multiple genes and traits in a unified framework. Evolution, 69: 492-507.

Spinks P.Q., Thomson R.C., Shaffer B. 2014. The advantages of going large: genome-wide SNPs clarify the complex population history and systematics of the threatened western pond turtle. Molecular Ecology, 23: 2228-2241.

Sukumaran J., Holder M.K. 2010. DendroPy: a Python library for phylogenetic computing. Bioinformatics, 26: 1569-1571.

Sukumaran J., Knowles L.L. 2016. Species vs. structure delimitation under the multispecies coalescent. Proceedings of the National Academy of Sciences, USA, in review.

Swofford D.L. 2002. PAUP*. Phylogenetic Analysis Using Parsimony. (*and Other Methods). Version 4. Sinaur Associates, Sunderland, MA.

Wallace A.R. 1876. The Geographical Distribution of Animals. Harper & Brothers, NY, USA.

Warren D.L., Glor R.E., Turelli M. 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. Evolution, 62: 2868-2883.

Warren D.L., Glor R.E., Turelli M. 2010. ENMTools: a toolbox for comparative studies of environmental niche models. Ecography, 33: 607-611.

Webb S.D. 1991. Ecogeography and the Great American Interchange. Paleobiology 17: 266-280.

Wetherbee D.K. 1985. The former occurrence of the endangered giant beetle, *Dynastes hercules* in Hispaniola. Caribbean Journal of Science 211: 83-84.

Wiens J.J., Graham C.H. 2005. Niche conservatism: integrating evolution, ecology, and conservation biology. Annual Review of Ecology, Evolution, and Systematics, 36: 519-539.

Wiens J.J. 2007. Species delimitation: new approaches for discovering diversity. Systematic Biology, 56: 875-878.

Wiens J.J. 2012. Why biogeography matters: historical biogeography vs. phylogeography and community phylogenetics for inferring ecological and evolutionary processes. Frontiers of Biogeography, 4.3: 128-135.

Wilson E.O., Brown W.L. 1953. The subspecies concept and its taxonomic application. Systematic Zoology, 2: 97-111.

Yang Z., Rannala B. 2010. Bayesian species delimitation using multilocus sequence data. Proceedings of the National Academy of Sciences, 107: 9264-9269.

Yang Z., Rannala B. 2014. Unguided species delimitation using DNA sequence data from multiple loci. Molecular Biology and Evolution, 31: 3125-3135.

Yeates D.K., Seago A., Nelson L., Cameron S.L., Joseph L., Trueman J.W.H. 2010. Integrative taxonomy, or iterative taxonomy? Systematic Entomology, 36: 209-217.

Zachos F.E. 2013. Taxonomy: species splitting puts conservation at risk. Nature, 494: 35.

Zhang C., Rannala B., Yang Z. 2014. Bayesian species delimitation can be robust to guide-tree inference errors. Systematic Biology, 63: 993-1004.

Zhang C., Zhang D-X., Zhu T., Yang Z. 2011. Evaluation of a Bayesian coalescent method of species delimitation. Systematic Biology, 60: 747-761.