Systematics and Species Delimitation in New Guinea Skink Species Complexes (Squamata: Scincidae)

Eric Nikolaus Rittmeyer
Louisiana State University and Agricultural and Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_dissertations

Recommended Citation
https://digitalcommons.lsu.edu/gradschool_dissertations/287

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
SYSTEMATICS AND SPECIES DELIMITATION IN NEW GUINEA SKINK SPECIES COMPLEXES (SQUAMATA: SCINCIDAE)

A Dissertation

Submitted to the Graduate Faculty of the Louisiana State University and Agricultural and Mechanical College in partial fulfillment of the requirements for the degree of Doctor of Philosophy in

The Department of Biological Sciences

by

Eric Nikolaus Rittmeyer
B.S., Cornell University, 2008
August 2014
ACKNOWLEDGEMENTS

First and foremost, I would like to thank my major professor, Dr. Christopher C. Austin; without his support, advice, guidance, and mentorship, none of this work would have been possible. I would also thank my committee members, both past and current, including Dr. Jeremy Brown, Dr. Robb Brumfield, Dr. Bryan Carstens, Dr. Michael Hellberg, Dr. Slawo Lomnicki, and Dr. Robert Tague, all of whom have been of great benefit towards my development as a researcher, evolutionary biologist, and herpetologist, as well as in the formation of this research. This work would not have come to fruition without the guidance, support and advice of all the aforementioned.

Prior to my arrival at LSU, my undergraduate research mentors, Dr. Harry Greene and Dr. Kelly Zamudio, my undergraduate academic advisor, Dr. Richard Harrison, and the advisor to the Cornell Herpetological Society, Dr. Kraig Adler, all played critical roles in my development as a scientist, and in forming my knowledge base in herpetology and evolutionary biology. I have no doubt that I would not have gone to graduate school, and would not have undertaken this dissertation research were it not for them. I would also like to thank Michael Gründler and Derrick Thompson, who accompanied me on my first expedition to Papua New Guinea in June 2008, which solidified my fascination with the diverse and underexplored island of New Guinea, Dr. Ed Scholes, who motivated our lofty decision to undertake an expedition to Papua New Guinea as a group of three undergraduate students, and Dr. Allen Allison, whose support helped make this trip possible.

My gratitude also goes out to the members of the Austin lab, past and present, graduate and undergraduate, including Dr. Alison Hamilton, Dr. Nathan Jackson, Cathy Newman, Zach Rodriguez, Lauren Oliver, John Andermann, and Jeff Weinell. All have been excellent labmates who provided helpful feedback, discussions, and ideas that greatly benefited my research. I also would thank my other herpetologist friends and colleagues in Baton Rouge, including Dr. Jeff Boundy, Dr. John McVay, Chris Williams, and Melina Lopez, as well as those outside of Baton Rouge, including Amanda Talaba, Nicholas Sly, and Shawn Billerman.

The phenomenal assistance of various other faculty and staff members at Louisiana State University has also been critical to this research. In particular, Tammie Jackson, Laura Savage, and Priscilla Milligan have all been absolutely incredibly helpful during my graduate career. I have also had excellent discussion with, received guidance from, and been otherwise aided by a number of others, including Steve Cardiff, Dr. Prosanta Chakrabarty, Donna Dittman, Dr. Jake Esselstyn, Dr. Mark Hafner, Dr. Kyle Harms, Dan Lane, Dr. Fred Sheldon, and Dr. Phil Stouffer, among others.

My immense gratitude also goes out to the students and postdoctoral researchers in the Museum of Natural Science and the Department of Biological Sciences, who I have been fortunate enough to call friends and colleagues over the past several years, and who have been everything from sounding boards for ideas, sources of advice and guidance, reviewers for talks and manuscripts, and friendly companions. These include Matt Brady, Dr. Gustavo Bravo, Clare Brown, Kristin Bryzyski, Vivien Chua, Dr. Andrés Cuervo, Melissa DeBiasse, Valerie Derouen, Dr. Liz Derryberry, Caroline Duffie, Jesús Fernández, Meche Gavilanez, Richard Gibbons, Michael
Harvey, Dr. Sarah Hird, Bill Ludt, Dr. James Maley, Dr. Verity Mathis, Dr. John McCormack, Caleb McMahan, Dr. John McVay, John Mittermeier, Brad Nelson, Lori Patrick, Carlos Prada, Tara Pelletier, Dr. Noah Reid, César Sánchez, Jordan Satler, Dr. Brian Smith, Hector Urbina, and Jared Wolfe, among others.

This research would not have been possible without the assistance of a number of colleagues and collaborators who provided tissue samples, including Dr. Allen Allison, Dr. Fred Kraus, Dr. Steven Donnellan, Dr. Jim McGuire, Alex Stubbs, Ben Karins, Dr. Ben Evans, and Pumehana Imada. Dr. Jonathan Losos, José Rosado, Tsuyoshi Takahashi, Jonathan Woodward, David Kizirian, and Dr. Christopher Raxworthy were all helpful in facilitating museum collections visits, and Ross Sadlier also provided loans of voucher specimens.

I would also like to thank the large number of people who assisted with the fieldwork associated with this research. In particular, Bulisa Iova, Jimmy Anamiato, and Iliah Bigilale of the Papua New Guinea National Museum were immensely helpful in fieldwork, as were Barbra Roy, Veri Kula, and Barnabus Wilmot from the PNG Department of Environment and Conservation, and Jim Robins from the PNG National Research Institute. My great gratitude also goes out to the many villagers who also assisted with fieldwork, either as guides or in other capacities.

This work also represents many years of computational time, a task that would not have been possible without the service of multiple high performance computing clusters. These include the Louisiana State University High Performance Computing cluster (LSUHPC), the Cyber Infrastructure for Phylogenetic Research (CIPRES), and the Computational Biology Service Unit (CBSU) at Cornell University, which is partially funded by the Microsoft Corporation.

I would also like to thank my family, including, in particular, my parents, Deb and Tom Rittmeyer, my sister, Ellen Rittmeyer, her son, Gavin, stepson, Jeremy, husband, Matt Hendershott, my maternal grandfather, Richard Gleason, and my aunts and uncles, including Anne Foster, Mike Foster, Andres Rittmeyer, Ruby Rittmeyer, and others. All have been exceptionally loving and supportive, and I would not be where I am today without them.

Last, but most certainly not least, I would like to thank my fiancée, Zeynep Altinay, whose unconditional love and support has truly made this dissertation a reality. While our time together has, thus far, been limited, she has been instrumental the completion of this work, and I have no doubt she will be a source of love, support, and inspiration for decades to come.
TABLE OF CONTENTS

ACKNOWLEDGEMENTS ........................................................................................................... ii

ABSTRACT ....................................................................................................................................... v

CHAPTER 1. INTRODUCTION: SYSTEMATICS, SPECIES DELIMITATION, AND NEW GUINEA BIODIVERSITY ......................................................... 1

CHAPTER 2. THE EFFECTS OF SAMPLING ON DELIMITING SPECIES FROM MULTI-LOCUS SEQUENCE DATA .......................................................................................... 10

CHAPTER 3. DIVERSIFICATION IN THE SAVANNAS OF THE SAHUL: SYSTEMATICS, SPECIES DELIMITATION, AND NICHE EVOLUTION IN THE CARLIA BICARINATA GROUP (SQUAMATA: SCINCIDAE) .................................................................................................... 29

CHAPTER 4. SYSTEMATICS AND BIOGEOGRAPHY OF THE CARLIA FUSCA GROUP (SQUAMATA: SCINCIDAE) ................................................................................................................. 50

CHAPTER 5. COMBINED NEXT-GENERATION SEQUENCING AND MORPHOLOGY REVEAL FINE-SCALE SPECIATION IN CROCODILE SKINKS (SQUAMATA: SCINCIDAE: TRIBOLONOTUS) ........................................................................................................ 70

CHAPTER 6. CONCLUSIONS ........................................................................................................... 90

REFERENCES ............................................................................................................................... 96

APPENDIX A. PERMISSIONS FOR PORTIONS OF CHAPTER 1 .................................................. 118

APPENDIX B. PERMISSIONS FOR CHAPTER 2 ........................................................................ 119

APPENDIX C. SUPPLEMENTAL MATERIALS FOR CHAPTER 3 .............................................. 120

APPENDIX D. SUPPLEMENTAL MATERIALS FOR CHAPTER 4 .............................................. 123

APPENDIX E. PERMISSIONS FOR PORTIONS OF CHAPTER 4 .............................................. 133

APPENDIX F. SUPPLEMENTAL MATERIALS FOR CHAPTER 5 .............................................. 134

VITA .............................................................................................................................................. 143
ABSTRACT

Though among the most controversial topics in systematic and evolutionary biology, species are a fundamental unit in biology, and are utilized by and critical to a wide variety of studies in the life sciences. Despite this importance, little work has focused on developing and examining objective methods for species delimitation until recently. Further, New Guinea and the surrounding regions are among the most diverse and geologically complex regions globally, yet the region remains poorly explored biologically, and little work has examined the evolutionary history of the fauna in the region.

To investigate the influence of factors such as sampling intensity, species richness, and phylogenetic structure on discovery methods for species delimitation, I combine simulated and empirical data. In Chapter 1, I use simulated data to examine the accuracy of three discovery methods for species delimitation under a variety of different sampling strategies. I find that genetic clustering algorithms, such as Structurama, can be highly accurate in identifying even recent divergences with limited sampling of individuals and of loci, and that Gaussian clustering can be similarly accurate, though somewhat less sensitive to detecting recent divergences. However, my results show that nonparametric delimitation is highly sensitive to errors in gene genealogy estimation, and generally fails to delimit species accurately when true coalescent gene genealogies are unknown, as in empirical applications.

In Chapters 3 and 4, I apply these methods empirically to examine the species boundaries, as well as the phylogeny and other aspects of the evolutionary history of, scincid lizards of the C. bicarinate and C. fusca groups, respectively. My results in Chapter 3 indicate that species delimitation analyses may be prone to underestimating the number of species by identifying only higher levels of clustering in systems with deep phylogenetic structure. I additionally find evidence for several cryptic species in the group, including deep, species-level divergence among the populations of C. storri from Australia, the Aru Islands, and New Guinea, despite their recent connectivity via Sahul Shelf emergence during Pleistocene glaciations. Through also examining niche evolution in the group, I find evidence for niche conservatism among most species in the group, but two species, C. bicarinate and C. sp. Amaw from eastern Papua New Guinea, show evidence for environmental niche divergence.

Analyses of the C. fusca group in Chapter 4 provide further evidence for a tendency of discovery methods for species delimitation to under-detected species in groups with high diversity or deep phylogenetic structure. Genetic clustering algorithms based on the complete dataset only identify a small number of clusters that correspond largely to deep phylogenetic clades, but when restricted to within these clades, this method identifies clusters that correspond well to finer, putative species-level structure. I also find evidence for extensive cryptic diversity in this group, identifying 28 distinct species among my sampling of 16 currently recognized species, as well as other incongruence with current taxonomy, including synonymous species and mis-assigned populations, supporting previous evidence of the need for extensive taxonomic revision in the C. fusca group. My biogeographic analyses also provide evidence that the C. fusca group likely evolved in Australia or Australia and New Guinea before diversifying in New Guinea, dispersing at least twice across Lydekker’s line into Wallacea, and possibly also recolonizing Australia.
Finally, in Chapter 5, I take a more comprehensive approach, and combine genomic and morphological data to test the validity of and examine the demographic history of two putative species of *Tribolonotus* from the islands of Buka and Bougainville in the northwestern Solomon Archipelago. I use next-generation sequencing to collect a genomic dataset of several thousand loci, and apply species discovery (genetic clustering algorithms) and species validation (Bayes factor delimitations) to test for speciation between these populations. My results support this speciation event, despite the recent connectivity between these islands. I also collect a suite of morphological characters for this group and provide evidence for morphological divergence and diagnosibility. Demographic analyses applied using approximate Bayesian computation and diffusion analysis further provide evidence for a complex demographic scenario in which migration between these populations continued for some time following their initial divergence, but subsequently decreased in rate or ceased entirely.

Combined, these results yield extensive insight into the utility of several methods for species delimitation, the taxonomy and systematics of *CARLIA* and *Tribolonotus* in New Guinea and the surrounding regions, and the complex processes responsible for driving the generation and maintenance of the phenomenal diversity in the Sahul shelf region.
1.1. The Importance of Species and Species Delimitation

In a broad sense, the primary goals of evolutionary and systematic biology are to document biodiversity, to estimate the phylogenetic relationships among organisms, and to understand the processes responsible for the generation and maintenance of this diversity (Futuyma 2009). As a fundamental unit in biology, species play a critically important role in this, and in a wide variety of other subfields, including ecology, physiology, microbiology and undeniably every other subfield of the life sciences (De Queiroz & Gauthier 1992; de Queiroz & Gauthier 1994; Rieseberg & Burke 2001; Lee 2003; Coyne & Orr 2004; Agapow et al. 2004; Agapow 2005; de Queiroz 2007; Bortolus 2008; Wheeler 2008). Indeed, errors in species delimitation, which can occur via misidentification, over-lumping, or over-splitting of species, can have negative consequences that cascade through the various subfields of biology and can compromise a wide variety of studies. For example, poor species delimitations can result in over- or under-estimations of biodiversity, reducing the efficiency of conservation plans aimed at maximizing the protection of biodiversity, prevent the detection of invasive species, or even result in the inadvertent introduction of invasive species during restoration projects (Isaac et al. 2004; Bickford et al. 2007; Bortolus 2008). Such errors can also drive misinterpretations of the variability within or among species, including morphological, ecological, physiological, or genetic variability (Isaac et al. 2004; Bickford et al. 2007; Bortolus 2008). Further, accurate phylogenetic estimations are important to a wide variety of studies in biology, particularly evolutionary biology (Huelsenbeck et al. 2001; Felsenstein 2004; Edwards et al. 2007). However, heterogeneity among gene genealogies - that is, different genomic regions showing different phylogenetic histories due to processes such as incomplete lineage sorting and hybridization (Maddison 1997; Degnan & Rosenberg 2006, 2009; Edwards et al. 2007; Heled & Drummond 2010) - complicates the problem of accurately estimating the phylogenetic history. While recently developed species tree estimation methods allow the underlying species tree to be estimated while accounting for this gene tree heterogeneity, these methods require accurate species delimitations (Liu 2008; Kubatko et al. 2009; Knowles 2009; Heled & Drummond 2010). Thus, delimiting species in a way that accurately represents independently evolving lineages, as well as understanding the processes responsible for the generation and maintenance of this diversity, are vitally important to biology.

1.1.1. The Species Concept Controversy

While species are an exceptionally important aspect of biology, the problem of defining what a species is and how best to delimit species is among the most controversial areas of systematic biology (Sokal & Crovello 1970; Mallet 2001; Coyne & Orr 2004; de Queiroz 2005, 2007; Hausdorf 2011). Numerous species concepts have been proposed in the literature, using a wide variety of criteria to define species, including reproductive isolation (biological species concept; Mayr 1942, 1995), ecological differentiation (ecological species concept; Van Valen 1976), reciprocal monophyly (genealogical species concept; Baum & Donoghue 1995; Baum & Shaw 1995; Shaw 1998), or morphological diagnosibility (diagnostic or phylogenetic species concept; Cracraft 1989), among others. More recently, de Queiroz (2005, 2007) argued for a
general lineage species concept, defining species as independently evolving metapopulation lineages, and arguing that the criteria used to define species concepts are not useful in defining species, but instead represent characteristics that evolve gradually in diverging lineages and can be useful in identifying distinct species (Fig. 1.1). I do not here intend to contribute to the already extensive debate into the positives and negatives of the wide variety of species concepts. Instead, I follow de Queiroz in recognizing that many criteria used in species concepts evolve gradually as lineages diverge, and, for the purposes of this work, define species as independently evolving metapopulation lineages.

**Figure 1.1.** Schematic of the general lineage species concept, depicting how various species criteria utilized by other species concept evolve gradually as lineages diverge. Reproduced from De Queiroz (2007) with permission from Oxford University Press.

**1.1.2. The Importance of and Approaches to Species Delimitation**

Despite the critical importance of species, species delimitation is far from a trivial task, particularly in broadly distributed groups, where distinguishing between isolation by distance or local adaptation and species boundaries can be difficult, or in groups that exhibit a low level of morphological diversity or high levels of intraspecific variation, where inter- and intraspecific variation can be difficult to differentiate. Genetic data contain a wealth of information that can be tapped in attempts to delimit species in taxonomically difficult groups, as evidenced by the numerous studies that have used genetic resources to delimit morphologically cryptic species. The majority of these studies use relatively subjective criteria, such as genetic divergence thresholds (Hebert *et al.* 2004; Lefébure *et al.* 2006) or reciprocal monophyly, which takes a long time to evolve, particularly with large effective population sizes, and thus may be overly conservative and is likely to fail to identify recently diverged species (Hudson & Coyne 2002; Hudson & Turelli 2003; Knowles & Carstens 2007). Further, many of these studies rely on a single marker which may not accurately represent the species phylogeny and species boundaries,
even in absence of hybridization (Maddison 1997; Degnan & Rosenberg 2006, 2009; Edwards et al. 2007). Therefore, a more accurate method of delimiting species from molecular data would incorporate multiple loci and account for the stochasticity of the coalescent process, as discussed below in Section 1.2.

Methods for species delimitation can be broadly categorized in two ways: based on the aims of the analyses (species validation versus species discovery), and based on the type of analyses (phylogeny-based or non-phylogeny-based). Species validation approaches, such as Bayesian species delimitation (Yang & Rannala 2010), Bayes factor delimitations (Leaché et al. 2013; Grummer et al. 2014) or SpeDeSTEM (Ence & Carstens 2011), aim to test the validity of putative species designated a priori, frequently in the context of comparing species trees estimated using the multi-species coalescent under competing species delimitation models. However, the accuracy of species delimitation using these approaches is reliant on the accuracy of the a priori putative species designations. Further, Bayesian species delimitations use a reversible jump MCMC algorithm to collapse or split nodes representing putative speciation events in a guide tree; thus, the accuracy of this method also relies on the accuracy of the guide tree specified a priori (Leaché & Fujita 2010; Yang & Rannala 2010). Frequently in empirical systems, it is difficult to accurately and objectively designate putative species a priori. Species discovery methods, such as Gaussian clustering (Hausdorf & Hennig 2010; Edwards & Knowles 2014), genetic clustering algorithms (Pritchard et al. 2000; Falush et al. 2003; Huelsenbeck & Andolfatto 2007; Huelsenbeck et al. 2011), the generalized mixed Yule coalescent (Pons et al. 2006; Fontaneto et al. 2007; Reid & Carstens 2012; Fujisawa & Barraclough 2013) or fields for recombination (Doyle 1995), aim to identify species without the need for a priori assignments of samples to putative species. Yet many of these species discovery approaches do not generally delimit species in a directly phylogenetic framework that accounts for heterogeneity among gene genealogies, such as via the multi-species coalescent; those that do suffer from the exceptional computational challenge associated with both species tree estimation, and the delimitation of species without a priori putative species, and do not account for uncertainty in gene genealogy estimation (O’Meara 2010).

Alternatively, species delimitation methods can be categorized based on the type of analyses - that is, whether the analysis uses a phylogeny-based approach, such as nonparametric delimitation (O’Meara 2010), Bayesian species delimitation (Yang & Rannala 2010), or the generalized mixed Yule coalescent model (Pons et al. 2006; Fontaneto et al. 2007; Reid & Carstens 2012), or a non-phylogeny-based approach, such as fields for recombination (Doyle 1995), genetic clustering algorithms (Pritchard et al. 2000; Falush et al. 2003; Huelsenbeck & Andolfatto 2007; Huelsenbeck et al. 2011), or Gaussian clustering (Hausdorf & Hennig 2010; Edwards & Knowles 2014). Some authors have argued against the use of phylogeny-based methods of species delimitation, due to the inability of gene topologies to track the tokogenetic relationships below the species level and the expectation that individual gene genealogies may not track organismal relationships (Doyle 1995). More recent work has further shown that the accuracy of phylogeny-based methods for species delimitation can be decreased substantially if uncertainty in phylogeny is not accounted for (Reid & Carstens 2012; Rittmeyer & Austin 2012). While more recently developed methods account for some of these problems, these methods introduce other difficulties. The Bayesian implementation of the generalized mixed Yule coalescent model accounts for phylogenetic uncertainty, but requires either single locus gene
data or a concatenation approach (Reid & Carstens 2012). Bayesian species delimitation (Leaché & Fujita 2010; Yang & Rannala 2010) and Bayes factor delimitations (Leaché et al. 2013; Grummer et al. 2014) use multi-locus data and explicitly account for heterogeneity among gene genealogies via the multi-species coalescent; however, these methods require a priori hypotheses of putative species. Finally, nonparametric delimitation and KC delimitation are species discovery methods that account for heterogeneity among gene genealogies and do not require a priori assignments of samples to putative species; however, these methods use point estimates of gene genealogies and thus do not account for phylogenetic uncertainty (O’Meara 2010). While non-phylogenetic methods of species delimitation do not generally require a priori hypotheses of putative species, these methods typically use genetic distances (Gaussian clustering, Hausdorf & Hennig 2010; Edwards & Knowles 2014) or allele frequencies (Structurama, Huelsenbeck & Andolfatto 2007; Huelsenbeck et al. 2011; Structure, (Pritchard et al. 2000; Falush et al. 2003), and do not directly incorporate evolutionary history or coalescent modeling in species delimitation. Thus, a variety of methods are available for species delimitation, each of which has its own assumptions, requirements, approach, and limitations. Despite the shortcomings of each of these methods, studies have suggested that many of these analyses may be useful for delimiting species from genetic data (Hausdorf & Hennig 2010; Leaché & Fujita 2010; Reid & Carstens 2012; Grummer et al. 2014). However, studies specifically aimed at addressing the impact of sampling strategy and of the true number of species are currently lacking.

1.2. Phylogenetics and the Problem of Heterogeneity Among Gene Genealogies

Estimating the phylogenetic relationships among species is a fundamental goal of evolutionary biology (Wiens 2007; Futuyma 2009). Beyond this, failure to accurately account for shared history can yield strongly biased results (Felsenstein 1985; Harvey & Pagel 1991; Martins & Hansen 1997; Schluter et al. 1997; Pagel 1999), a problem not restricted to evolutionary biology, but pervasive throughout the life sciences. With the advent of modern molecular techniques and increased computational power, our ability to reconstruct phylogenies has vastly improved (Felsenstein 2004; Edwards et al. 2007; Heled & Drummond 2010; Wiley & Lieberman 2011; Ronquist et al. 2012). Until recently, the vast majority of phylogenetic studies relied on single locus data - frequently, at least in the case of animals, single copy mitochondrial DNA (Zink & Barrowclough 2008; Beheregaray 2008; Hickerson et al. 2010) - or on concatenating multiple independently evolving nuclear loci and analyzing them under the assumption of a single shared phylogenetic history (Felsenstein 2004; Edwards et al. 2007). However, as multi-locus datasets have become more available, the problem of heterogeneity among gene topologies has become more apparent (Doyle 1992; Maddison 1997; Degnan & Rosenberg 2006, 2009; Edwards et al. 2007). This problem has often been overlooked, and gene genealogies are often assumed to reflect the species phylogeny (Edwards et al. 2007). However, particularly in systems with short internal branches or large effective populations sizes, individual gene genealogies can differ substantially from the underlying species tree, and in some cases in the ‘anomaly zone,’ the most common gene genealogy can differ from the underlying organismal phylogeny (Degnan & Rosenberg 2006, 2009). Further, some work has shown that concatenated analysis of multiple loci is statistically inconsistent, and can yield strong support for incorrect relationships (Kubatko & Degnan 2007). Recently developed species tree estimation methods, such as *BEAST (Bouckaert et al. 2013), BEST (Liu 2008), STEM (Kubatko et al. 2009), STEAC (Liu et al. 2009), and others, account for this heterogeneity among gene genealogies while estimating the
species-level phylogenetic relationships. However, these methods assume accurate species delimitations (Edwards et al. 2007; Liu 2008; Kubatko et al. 2009; Heled & Drummond 2010), which can be particularly difficult in morphologically conservative groups with cryptic species. While studies on the impact of errors in species delimitation on species tree estimation are lacking, it is reasonable to expect that errors in species delimitation could strongly bias these inferences, particularly in cases where errors involve non-sister taxa, but even errors involving sister taxa are likely to cause underestimated divergence times and overestimated population size. Therefore, delimiting species in a biologically realistic way is increasingly important, and critical not only for alpha taxonomic purposes, but for phylogenetic estimations and a wide variety of other biological studies.

1.3. Biodiversity of New Guinea

New Guinea and the surrounding areas (Fig. 1.2) is a particularly rewarding region for studying the processes responsible for generating and maintaining biodiversity. Despite occupying 0.6% of the global land area, New Guinea harbors an impressive 5-7% of the world’s biodiversity (Allison 1996; Mittermeier et al. 2003, 2005), and has been identified as one of five High Biodiversity Wilderness areas (Mittermeier et al. 2003), while the adjacent islands of Northern Melanesia, including the Admiralty, Bismarck, and Solomon Archipelagoes (Fig. 1.2) have been identified as one of 35 global biodiversity hotspots (Mittermeier et al. 2005). Remarkably, the diversity of this region remains vastly underestimated: large numbers of species

Figure 1.2. Map of New Guinea and the surrounding regions. AI: Admiralty Islands; Ar: Aru Islands; CY: Cape York, Queensland, Australia; H: Halmahera; HP: Huon Peninsula; New Guinea; K: Kei Islands; NB: New Britain; Bismarck Archipelago; NI: New Ireland, Bismarck Archipelago; NT: Arnhem Land, Northern Territory, Australia; PM: Port Moresby, New Guinea; S: Seram Island; SI: Solomon Archipelago; TF: Trans-Fly region, New Guinea; V: Vogelkopf Peninsula, New Guinea.
remain to be discovered in the region (Allison 1996, 2007; Menzies 2006; Austin et al. 2008), and numerous new species (Kraus 2008, 2009, 2013; Günther et al. 2012, 2014; Oliver et al. 2012; Rittmeyer et al. 2012; Zug & Fisher 2012; Menzies 2014) and even new genera (Kraus 2010; Günther et al. 2010) are described annually. While species accumulation curves are expected to plateau as the cumulative number of described species in a region approaches the total number of species in the region, in New Guinea, the species accumulation curves for squamates (snakes and lizards, Fig. 1.3A) and anurans (frogs, Fig. 1.3B) continue to increase at a higher than linear rate, reflective of the vast number of undescribed species in the region. Further, while only a handful of the many broadly distributed taxa have been subject to rigorous phylogeographic analysis, these studies have generally revealed extensive cryptic diversity (Rawlings & Donnellan 2003; Austin et al. 2011; Macqueen et al. 2011; Oliver et al. 2013). For example, Oliver et al. (2013) showed the widespread terrestrial frog species Mantophryne lateralis instead represents a complex of at least nine distinct species, while Rawlings and Donnellan (2003) revealed two deeply divergent, potentially species-level lineages within the large, arboreal snake Morelia viridis. This incredible biodiversity is due in part to the geologic and tectonic complexity of the region (Hall 2002; Heads 2002; Polhemus 2007), as well as historic climate and sea level fluctuations driving shifts in the distributions of various habitat types and altering the connectivity among now isolated islands (Bowler et al. 1976; Allison 1996; Hope & Aplin 1997; Voris 2000; Hope 2007). The southern portion of New Guinea is the leading edge of the Australian plate (Hall 2002); as this plate moved northwards, first the Inner Melanesian Island Arc was accreted approximately 45 to 50 million years ago (Mya, Petterson et al. 1999; Hall 2002; Mann & Taira 2004). Approximately five to ten Mya, the subsequent continued northward movement of the Australian plate resulted in a collision with the western edge of the Outer Melanesian Island Arc (Fig. 1.4; Abbott et al. 1994; Tregoning et al. 1999;
 Portions of this historic island arc now form much of the north coast of New Guinea, while the eastern portions of it form the adjacent Melanesian Islands of the Admiralty, Bismarck, and Solomon Archipelagoes (Abbott et al. 1994; Allison 1996; Tregoning et al. 1999; Hall 2002; Heads 2002). This collision with the Outer Melanesian Island Arc further drove the uplift of the main east-to-west cordillera that exceeds 5,000m in height.

Historic fluctuations in climate and sea level, particularly during Pleistocene glacial cycles, also likely played an important role in driving the generation of the impressive biodiversity of New Guinea. During Pleistocene glaciations, sea levels repeatedly declined at least 100 m below present levels; a decrease of less than 10 m results in the formation of a land bridge connecting New Guinea with Cape York, Australia, while a further decrease of only 40 m yields a subaerial connection with the Aru Islands (Hope & Aplin 1997; Voris 2000). These sea level declines also resulted in the formation of subaerial land bridges between numerous other islands in the region, including joining many of the islands of the Solomon Archipelago. Further, modern New Guinea is dominated by tropical rainforests. Seasonally xeric, eucalypt savanna or woodland habitats occur in only two disjunct patches: one in the Trans-Fly region of south central New Guinea, and one in the vicinity of Port Moresby in southeastern New Guinea (Allison 1996, 2006; Menzies 2006; Marshall & Beehler 2007). However, during Pleistocene glaciations and other historically drier periods, these seasonally xeric savanna habitats were far

![Figure 1.4](image-url). Tectonic history of the New Guinea region, modified from Hall (2002). A. 55 million years before present (Mya), B. 30 Mya, C. 10 Mya, D. present. Shades of green indicate portions of the Australian plate, blue indicates portions of the Inner Melanesian Island Arc, and red indicates portions of the Outer Melanesian Island Arc, light grey indicates shallow seas that were subaerial during periods of low sea level, and dark grey indicates subaerial land of other geologic origins.
more widespread throughout much of southern New Guinea. Montane rainforest habitats and alpine grasslands also shifted to lower elevations during the historically cooler and drier periods of Pleistocene glaciations (Bowler et al. 1976; Allison 1996; Hope & Aplin 1997; Hope 2007). The complex geologic history, habitat shifts, and sea level fluctuations, along with other factors, have combined to drive the generation of the substantial biodiversity of the region, providing exceptional opportunity for elucidating the patterns and processes of diversification.

1.4. Overview of Chapters

This dissertation focuses primarily on two important aspects of my research interests: examining the utility of various methods for species delimitation from molecular data and their sensitivity to sampling strategy, and examining the systematics and evolution of scincid lizards in New Guinea, including species boundaries, phylogenetics, and various other aspects of their evolution, such as ecological niche evolution, biogeography, and historical demography.

In Chapter 2, I examine the accuracy and sensitivity to sampling strategy of three recently developed methods that aim to delimit species from multi-locus DNA sequence data without a priori assignments of samples to putative species. Specifically, I simulate data for 100 five-taxon species trees at each of two species tree depths (6 Ne generations and 12 Ne generations) and under a variety of sampling strategies ranging from five alleles per species and five loci to 20 alleles per species and 100 loci to test (1) Structurama (Huelsenbeck & Andolfatto 2007; Huelsenbeck et al. 2011), (2) Gaussian clustering (Hausdorf & Hennig 2010), and (3) nonparametric delimitation (O’Meara 2010). I also investigate the types of error in species delimitation from each of these three methods as errors due to misidentification of samples, due to over-lumping of species, or due to over-splitting of species. Finally, in the case of nonparametric delimitation, I examine the impact of phylogenetic uncertainty and errors in gene genealogy estimation on resultant species delimitations through comparing results obtained using estimated gene genealogies to those obtained using the true, coalescent genealogies.

In Chapter 3, I examine the phylogeny, species boundaries, and niche evolution in the Carlia bicarinata group, a small clade consisting of three species of primarily savanna-specialist scincid lizards distributed throughout southern New Guinea, the Aru Islands, and eastern Australia. I collect multi-locus sequence data via Sanger sequencing for 43 samples spanning all known species, morphotypes, and major distributional areas in the group, and apply multiple methods for species delimitation, including Structurama (Huelsenbeck & Andolfatto 2007; Huelsenbeck et al. 2011), Gaussian clustering (Hausdorf & Hennig 2010), Bayesian species delimitation (Yang & Rannala 2010), and Bayes factor delimitation (Grummer et al. 2014) to test species boundaries in the group, and used *BEAST (Heled & Drummond 2010) to estimate species level relationships while accounting for heterogeneity among gene genealogies. I additionally use collection localities from HerpNET (www.herpnet.org) and OZCAM (www.ozcam.org), and GIS layers of environmental variables along with background similarity tests (Warren et al. 2008) and multivariate analyses of niche similarity (McCormack et al. 2010) to examine environmental niche evolution in the group.

In Chapter 4, I use multi-locus sequence data to examine the phylogeny, species boundaries, and biogeographic history of the Carlia fusca group. As currently recognized, this
group is comprised of 18 species, with the majority of the diversity distributed largely parapatrically throughout New Guinea (Zug 2004; Zug & Allison 2006), but that also includes three species in parts of northern Australia (Donnellan et al. 2009), and four species that occur east of Lydekker’s line (Lydekker 1896) in Wallacea (Zug 2004). However, some molecular phylogeographic work has suggest extensive incongruence between major phylogenetic clades and the currently recognized species in the group (Austin et al. 2011). Further, most tests of and empirical applications of methods for species delimitation in diverse clades have utilized the generalized mixed Yule coalescent model (Pons et al. 2006; Monaghan et al. 2009), that does not account for heterogeneity among gene genealogies, and conservatively assumes species are monophyletic (Carstens et al. 2013). Little work has examined the utility of other methods for species delimitation in highly speciose clades. Therefore, I additionally use this dataset to test the utility of Gaussian clustering (Hausdorf & Hennig 2010), genetic clustering under a Dirichlet process prior (Huelsenbeck & Andolfatto 2007; Huelsenbeck et al. 2011), and Bayesian species delimitation (Yang & Rannala 2010) in complexes with relatively deep phylogenetic structure and high species richness. Finally, I use dispersal extinction cladogenesis (Ree et al. 2005; Ree & Smith 2008) and statistical dispersal vicariance analyses (Yu et al. 2010) to examine the biogeographic history of the C. fusca group, and specifically examine 1) the geographic origin for the group, 2) the number of dispersal events between New Guinea and Australia, and 3) the number of dispersal events across Lydekker’s line.

Analysis of relatively small numbers of loci collected via Sanger sequencing, as in the above chapters, can yield extensive useful information on species boundaries, phylogenetic relationships and evolutionary history. The recent development and rapidly expanding availability of next generation sequencing (NGS) technologies enables researchers to sequence hundreds to thousands of independently evolving loci, and thus obtain far more precise, robust, and detailed inference of species boundaries and evolutionarily important parameters, such as effective population sizes, divergence times, and migration rates. However, in studies endeavoring to examine species boundaries, it is critically important to test the validity of putative species identified via molecular data with other types of data, such as morphology. In Chapter 5, I use NGS to obtain several thousand loci for two divergent populations of Tribolonotus pseudoponceleti complex skinks from currently isolated, but historically connected islands in the northwestern Solomon Archipelago. I combine these data with a morphological dataset consisting of eight meristic and eight mensural characters to test the validity of these populations as distinct species using multiple data sources and multiple analytical approaches. I additionally use approximate Bayesian computation and diffusion approximation of the allele frequency spectrum to test a variety of demographic models, and infer the history of divergence, migration, and population size.

Finally, in Chapter 6, I conclude with a discussion synthesizing the results of this research, and its implications regarding species delimitation, systematics, and the taxonomy of the focal groups. I also discuss lingering questions and future avenues for research related to the foci of this dissertation.
CHAPTER 2
THE EFFECTS OF SAMPLING ON DELIMITING SPECIES FROM MULTI-LOCUS SEQUENCE DATA

2.1. Introduction

Species are a fundamental unit in biology important to every subfield in biology, and the inaccurate delimitation of species can compromise the integrity, relevance, and conclusions of research (Coyne & Orr 2004; de Queiroz 2007; Bickford et al. 2007; Bortolus 2008). Despite the importance of species and of delimiting species in a biologically meaningful manner, species concepts remain a controversial topic subject to extensive debate (Coyne & Orr 2004; de Queiroz 2007). A variety of criteria, including reproductive isolation (biological species concept; Mayr, 1942, 1995), reciprocal monophyly (genealogical species concept; Baum & Donoghue, 1995; Baum & Shaw, 1995), and diagnostic characters (phylogenetic species concept; Cracraft, 1989), among others, have been proposed for delimiting species; however, it is unlikely for many of these criteria to evolve instantaneously with speciation and the order in which they evolve is likely to vary among systems (De Queiroz 2005; de Queiroz 2007). Arguably the most inclusive species concept is the unified species concept, which defines species as independently evolving metapopulation lineages, and argues that many species concepts represent criteria that evolve as lineages diverge and that may be used to help delimit species, rather than definitions of species (De Queiroz 2005; de Queiroz 2007). Regardless of the specific species concept used, errors in species delimitation may come in three forms: over-splitting (i.e. a single species is treated as multiple species), over-lumping (i.e. multiple species are treated as a single species), or incorrect assignment of individuals or populations (i.e. samples of one species is treated as a member of a different, though valid, species). Over-splitting of species inflates measures of biodiversity, potentially biasing harvest or conservation strategies (Bickford et al. 2007). Over-splitting can also result in underestimates of intraspecific variation and viability, and overestimates of interspecific gene flow (Funk & Omland 2003). Over-lumping of species can cause the opposite problems, and potentially the failure to recognize and protect species of conservation concern (Bickford et al. 2007). Depending on the sampling strategy and questions investigated, all of these problems may also arise as a result of the incorrect assignment of populations to species. With increased use of molecular markers and increased sophistication of analyses for molecular data, it is becoming apparent that many traditionally recognized species, particularly those with broad geographic distributions, are actually complexes of multiple species with little or no gene flow among them, often recently diverged and morphologically conservative (Bickford et al. 2007). Thus, of the three delimitation errors, recent research suggests that over-lumping is of major concern to biodiversity research (Bickford et al. 2007).

Further, recently developed species tree inference methods (e.g. *BEAST, Heled & Drummond, 2010; BEST,Liu, 2008; minimize deep coalescences, Maddison, 1997; STEM Kubatko, Carstens, & Knowles, 2009) attempt to identify the underlying species-level phylogeny while accounting for heterogeneity among gene genealogies due to incomplete lineage sorting

(Edwards et al. 2007). However, these methods assume accurate delimitations of species a priori, and errors in these assignments are likely to result in unreliable species tree estimates, particularly if mis-assigned or over-lumped samples involve non-sister species.

Several methods that attempt to delimit species from molecular data rely on fixed divergence thresholds (Hebert et al. 2004; Lefébure et al. 2006) or reciprocal monophyly (Sites Jr. & Marshall 2004). Many such methods, such as generalized mixed Yule coalescent model (Pons et al. 2006; Monaghan et al. 2009) and statistical parsimony networks (Templeton et al. 1992; Clement et al. 2000), also use only single markers. Selection of a threshold of divergence for species delimitation is highly subjective, and a single threshold is unlikely to be appropriate for all systems (Moritz & Cicero 2004; Knowles & Carstens 2007). Additionally, while reciprocal monophyly may be useful for identifying species with older divergences, it may take a substantial amount of time for lineages to sort to reciprocal monophyly, particularly at multiple loci and in species with large effective population sizes (Degnan & Rosenberg 2006, 2009). Thus, reciprocal monophyly is highly conservative and likely to fail to identify recently diverged species (Hudson & Coyne 2002; Hudson & Turelli 2003; Knowles & Carstens 2007). Further, processes such as incomplete lineage sorting can result in a single marker not accurately representing the species phylogeny and species boundaries, particularly for recently diverged species, and in rapid radiations where the interval between speciation events is short (Degnan & Rosenberg 2006, 2009). Therefore, a more accurate method of delimiting species from molecular data would incorporate multiple loci and account for the stochasticity of the coalescent process.

While some powerful methods are available that delimit species from multi-locus data under a coalescent framework (e.g. SpedeSTEM, Ence & Carstens, 2011; Bayesian species delimitation, Yang & Rannala, 2010), these and other methods require a priori assignment of samples to putative species. These species validation methods are not be appropriate in all situations; even in well-studied systems, processes such as convergent evolution or morphological conservatism may make it impossible to accurately and objectively assign all populations to putative species (e.g. Pantherophis obsoletus complex, Burbrik, Lawson, & Slowinski, 2000; Sceloporus undulatus complex, Leaché & Reeder, 2002; Leaché, 2009; Carlia fusca group, Austin et al., 2011). In such situations, errors in assignment would likely result in errors in species delimitation, potentially resulting in misleading inferences.

Several methods are available for delimiting species from multi-locus molecular data without a priori putative species assignments. Of particular promise are Structurama (Huelsenbeck & Andolfatto 2007; Huelsenbeck et al. 2011), Gaussian clustering (Hausdorf & Hennig 2010), and nonparametric delimitation (O’Meara 2010). One additional method for delimiting species from multi-locus molecular data without a priori assignments is fields for recombination (FFR), which attempts to delimit species from non-overlapping sets of heterozygous individuals (Doyle 1995; Sites Jr. & Marshall 2003). However, this method performed extremely poorly in a previous test, correctly assigning less than 27% of individuals (Hausdorf & Hennig 2010). We thus exclude FFR from our analyses.

Structurama was developed for the purpose of detecting intraspecific population structure from genetic data (Huelsenbeck & Andolfatto 2007; Huelsenbeck et al. 2011) by combining the Bayesian clustering algorithm implemented in Structure (Pritchard et al. 2000; Falush et al.
with a Dirichlet process prior that allows the number of populations to be treated as a random variable (Huelsenbeck & Andolfatto 2007; Huelsenbeck et al. 2011). The algorithm thus aims to estimate both the number of populations and the composition of these populations by minimizing linkage disequilibrium and maximizing Hardy-Weinberg equilibrium (Pritchard et al. 2000; Falush et al. 2003; Huelsenbeck & Andolfatto 2007; Huelsenbeck et al. 2011). The processes driving neutral differentiation among species are similar to those driving neutral differentiation among intraspecific populations (i.e. genetic drift coupled with restricted gene flow); thus Structurama may also be useful for species delimitation. Indeed, the method has previously been shown to be informative for this application (Hausdorf & Fujita 2010; Pinzón & LaJeunesse 2011; Salicini et al. 2011). While both Structure and Structurama utilize the same algorithm, Structure assumes a fixed number of populations ($K$), whereas Structurama treats the number of populations as a random variable estimated via a Dirichlet process prior. Although metrics have been proposed to estimate the most appropriate $K$ using Structure (Evanno et al. 2005), estimating $K$ remains difficult and may be somewhat ambiguous (Hausdorf & Hennig 2010), and confidence in the values of $K$ generated by these methods cannot be statistically assessed (Pritchard et al. 2000; Evanno et al. 2005; Huelsenbeck et al. 2011). Thus, it seems unlikely that Structure would significantly outperform Structurama, and we here focus on testing Structurama and do not include Structure in this study.

Gaussian clustering was first applied to the problem of species delimitation by Hausdorf and Hennig (2010) for use with dominant and co-dominant allelic data (e.g., AFLPs, microsatellites) by using multidimensional scaling to convert a genetic distance matrix to a series of similarity vectors, from which clusters (i.e., species) are estimated. In the previous implementation of this method for species delimitation, it performed relatively well (Hausdorf & Hennig 2010), correctly assigning 73-93% of individuals. We include this method in this study to more thoroughly test its accuracy using multilocus DNA sequence data.

Nonparametric delimitation and KC delimitation are two additional approaches that attempt to jointly estimate species assignments and species trees without a priori data on putative species (O’Meara 2010). Unlike the methods described above, which use either a distance matrix or genetic data directly in a non-genealogical context, both of these methods are topology-based; that is, these methods attempt to delimit species from a set of gene genealogies. Nonparametric delimitation attempts to identify the species tree and the species delimitations that minimize both excess structure within species and the number of deep coalescent events among species. As this method uses gene genealogies as input, errors in species delimitation may result from one of two sources: errors due to gene tree uncertainty, and errors due to the algorithm itself. Therefore, to both test the empirical utility of the method and to tease apart the sources of error, we apply nonparametric delimitation both on estimated gene trees and on simulated coalescent gene trees.

While KC delimitation is a theoretically intriguing method that attempts to identify the species delimitations and species tree that maximize the probability of a set of gene genealogies, the method is extremely computationally intensive, and is thus unfeasible for use with datasets larger than a few samples or loci (O’Meara 2010). Further, in previous tests of the method, KC delimitation performed poorly (O’Meara 2010), possibly due to an inability to efficiently search the parameter space. Thus, we focus instead on nonparametric delimitation and do not further test KC delimitation in this study.
While all the methods discussed above have been applied to the problem of species delimitation, their accuracy and sensitivity to sampling intensity has not been thoroughly examined. We use simulated datasets at a variety of sampling intensities to assess the performance of a variety of species delimitation methods and to investigate their robustness to a range of sampling strategies. Specifically, we focus on testing Structurama (Huelsenbeck & Andolfatto 2007; Huelsenbeck et al. 2011), Gaussian clustering (Hausdorf & Hennig 2010), and nonparametric delimitation (O’Meara 2010) due to their fulfillment of two primary criteria: 1) they can be applied to multi-locus data, and 2) they do not require a priori assignments to putative species.

2.2. Methods

2.2.1. Data Simulations

To represent groups in which lineage sorting is expected to be complete for many loci between the most divergent species, as well as groups in which lineage sorting is expected to be incomplete at many loci among all species, we simulated data for two different levels of divergence or tree depths: 6N and 12N generations, where N is the effective population size. The mean time for lineage sorting to complete for a given locus is 4N ± 2N generations (Degnan & Rosenberg 2006, 2009); thus, the shallower trees (6N) represent the mean time to lineage sorting + one standard deviation, and some lineage sorting would be expected among all species, whereas for the deeper trees, lineage sorting should be complete for the majority of loci across the deeper divergences. For each of the two total tree depths, 100 species trees, each with five species, were simulated under a uniform Yule speciation model in Mesquite v.2.73 (Maddison 2009). For each species tree, 100 gene genealogies were simulated with 20 tips per species (i.e. 100 operational taxonomic units total) and θ equal to 0.01 in ms (Hudson & Coyne 2002). In our simulations, we assume no migration following speciation, thus the species simulated represent reproductively isolated species (biological species concept, Mayr, 1942, 1995). DNA sequence data, 500 bp in length for each gene, were then simulated on each gene genealogy in Seq-Gen (Rambaut & Grassly 1997) under an HKY+G model with a transition-transversion ratio of 3.0, base frequencies as 0.3 A, 0.2 C, 0.3 T, 0.2 G, and a discrete gamma distribution with a shape parameter α of 0.8, as in (McCormack et al. 2009). We refer to these simulated sequences as alleles, regardless of whether each is unique, such that alleles refers to the number of sampled sequences, and datasets with 20 alleles sample per species may (and typically do) include less than 20 unique sequences. To test the sensitivity of each method of species delimitation to sampling effort, we randomly reduced the number of alleles and loci sampled to obtain 18 total datasets per species tree: 5, 10 or 20 alleles per species sampled at 5, 10, 25, 50, 75 or 100 loci. While we do not explicitly assign alleles to diploid individuals in this study, the methods tested herein do not take intra-individual variation (i.e. heterozygosity) into account in delimiting species, thus our results are still applicable to diploid or higher ploidy level organisms. To prevent biases due to particularly informative loci or alleles, random reductions were performed such that all samples included in the smaller datasets were included in all larger datasets (i.e. datasets were nested).

To examine the extent of incomplete lineage sorting in the simulated data, genealogical sorting indices (gsi, Cummings, Neel, & Shaw, 2008) were calculated for each species based on
the true coalescent gene genealogies using the genealogicalSorting package (Bazinet, Neel, Shaw, & Cummings, unpublished; Cummings et al., 2008) in R v. 2.14.1. The gsi quantifies the amount of common ancestry of a group of operational taxonomic units (OTUs) on a phylogenetic tree, and varies from zero to one, where larger values represent more complete lineage sorting, up to a maximum value of 1.0 for a monophyletic group (i.e. complete lineage sorting). To examine the extent of variation within the simulated sequence data, we calculated the number of unique alleles for each locus both for each species and for each species tree (i.e. combining the five simulated species). We similarly calculated the number of segregating sites for each locus both for each species and for each species tree. The numbers of unique alleles were calculated using the pegas package (Paradis 2010) in R; the numbers of segregating sites were calculated using the ape package (Paradis et al. 2004) in R.

2.2.2. Species Delimitation using Structurama

Structurama (Huelsenbeck & Andolfatto 2007; Huelsenbeck et al. 2011) assumes loci are unlinked allelic markers and thus requires multi-locus sequence data to be converted to alleles (coded by integers). We used SNAP Map (Price & Carbone 2005; Aylor et al. 2006) to convert each locus to numbered alleles. The number of populations (K) was set as a random variable to implement the Dirichlet process prior; the prior distribution on the number of populations was set as a gamma distribution with a shape of 1.0 and a scale of 1.0. Markov chains were each run for one million iterations, sampling every 100 iterations; the first 1000 samples (10%) were discarded as burn-in. To ensure consistency of the results, a subset of 360 analyses (10 at each sampling intensity and tree depth) were repeated.

2.2.3. Species Delimitation using Gaussian Clustering

Genetic distance matrices for each locus were calculated using maximum likelihood as implemented in PAUP* ver. 4.0b10 (Swofford 2003) and the model of sequence evolution under which the data was simulated (HKY+G). Single locus distance matrices were then combined using standardized distances to create a multi-locus distance matrix in pofad ver. 1.03 (Joly & Bruneau 2006). This method scales distance matrices for each locus by the largest distance at that locus to prevent highly variable loci from having an excessive impact on the combined distance matrix. Gaussian clustering was then implemented in R v. 2.12.0 using the prabclus (Hausdorf & Hennig 2010) and mclust (Fraley & Raftery 2006) packages. Kruskal’s non-metric multidimensional scaling (Kruskal 1964) was used to convert the multi-locus distance matrix into similarity vectors, with a tuning constant of four (as suggested by Hausdorf and Hennig (2010) for identifying clusters containing a minimum of five individuals). Nearest neighbor-based noise detection was used with a tuning constant equal to the smallest integer greater than or equal to the number of samples divided by 40, as suggested by Hausdorf and Hennig (2010). Gaussian clustering was then implemented for all clustering models implemented in mclust; the best-fit model was selected using the Bayesian information criterion. To ensure consistency of the results, a subset of 360 analyses (10 at each sampling intensity and tree depth) were repeated.

2.2.4. Species Delimitation using Nonparametric Delimitation

Because nonparametric delimitation (NP) is a topology-based species delimitation method, we first estimated gene genealogies for each locus using maximum likelihood in RAxML ver. 7.0.3 (Stamatakis 2007). The model of sequence evolution was set to GTR+G, as the simpler, HKY+G model under which the data were simulated cannot be implemented in
RAxML. Three search replicates were conducted for each locus, and the tree with the highest log likelihood was retained for subsequent analyses.

While nonparametric delimitation based on these estimated gene genealogies would be more comparable to empirical implementations of the method, it would remain unclear if errors in species delimitation based on these estimated genealogies were due to errors in gene genealogy estimation or to poor performance of the nonparametric delimitation method. Thus, to control for errors in species delimitation due to gene tree uncertainty, we also implemented nonparametric delimitation using the true coalescent gene genealogies from which the sequence data were simulated.

Nonparametric delimitation was implemented with both estimated (NP.E) and coalescent (NP.C) gene genealogies in Brownie ver. 2.1.3 (O’Meara 2010) under default parameters of a structure weight of 0.5, and a P Threshold of 1.0. All nonparametre delimitation analyses consisted of five search replicates to ensure the best solution had been found. Although nonparametric delimitation jointly estimates the species tree and species delimitations, we here focus only on the accuracy of the species assignments, as the accuracy of a number of species tree estimation methods have previously been examined elsewhere (Linnen & Farrell 2008; McCormack et al. 2009; Heled & Drummond 2010; Leaché & Rannala 2011).

2.2.5. Statistical Tests
We calculated the accuracy of each delimitation method for every sampling strategy by calculating the percent of the samples correctly assigned to species, and averaging these values across all species trees at each sampling intensity for each of the two total species tree depths. Thus in a case with 100 alleles, if two species are lumped into a single species, and no samples are mis-assigned to a different species, the accuracy would be 0.8 (80/100), but the proportion of incorrectly assigned samples would be 0 (0/100). Similarly, in a case with 50 alleles, if two species are lumped into a single species, and two samples from a third species are lumped within this single lumped species, the accuracy would be 0.76 (38/50), but the proportion inaccurate would be 0.04 (2/50). To further examine the specific sources of error in species delimitations, we also calculated the number of over-split species, the number of over-lumped species, and the proportion of incorrectly assigned samples. We considered species as over-split if greater 20% of the alleles (i.e. at least two for tests with five alleles sampled, at least three for tests with 10 alleles sampled, or at least five for tests with 20 alleles sampled) were assigned to each of two distinct species. Similarly, we considered species as over-lumped if greater than 20% of the alleles from two different species were assigned to the same species. We calculated the proportion of incorrectly assigned alleles as the proportion of alleles assigned to a cluster along with only 20% or less of the conspecific alleles.

To test for significant differences among methods on the accuracy of species delimitations, as well as to determine specific impact of sampling intensity of the accuracy of species delimitation, we conducted pairwise t-tests in R ver. 2.14.1. P-values were adjusted for multiple comparisons via Bonferroni correction, that is, by multiplying the p-values by the number of comparisons: two for examining the impact of number of alleles sampled (i.e. increases from five to ten alleles and from ten to 20 alleles), five for examining the impact of the
number of loci sampled (i.e. increases from five to ten loci, ten to 25 loci, 25 to 50 loci, 50 to 75 loci, and 75 to 100 loci).

2.3. Results

Incomplete lineage sorting was extensive in the simulated datasets, as expected given the depth of the simulated species trees. The shallower, 6N species trees had an average gsi of 0.808 (± 0.192 standard deviation, SD). Further, an average 16.4% of simulated loci had gsi less than 1.0 for all five species (as expected given that most simulated divergences were more recent than 6N generations, and many were much more recent), and only 1.4% of the simulated loci were monophyletic (i.e. gsi = 1.0) for all five simulated species. Each locus showed complete lineage sorting for an average of 1.6 (± 1.1 SD) species for the shallower species trees. Lineage sorting was much more complete for the deeper, 12N species trees, though still prevalent. The average gsi for these trees was 0.877 (± 0.162 SD). Only an average of 4.1% of simulated loci had gsi of less than 1.0 for all five species, and lineage sorting was complete (i.e. gsi = 1.0) for all five species for 5.4% of the simulated loci in the deeper species trees. Lineage sorting had completed for an average of 2.4 (± 1.2 SD) species on the deeper species trees.

The simulated sequence data included an average of 39.9 (± 7.8 SD) segregating sites per locus for the complete datasets (i.e. 100 alleles) for the shallower species trees, but an average of only 8.7 (±4.2 SD) segregating sites within each simulated species. The total number of unique alleles for the complete datasets averaged 26.7 (± 4.5 SD), whereas within each species, the average number of unique alleles was 6.0 (±1.8 SD). The simulated sequence data for the deeper species trees averaged 50.6 (± 8.6 SD) segregating sites per locus for the complete datasets, but only 8.7 (± 4.2 SD) average segregating sites per locus within each simulated species. The complete datasets included an average of 27.9 (± 4.3 SD) unique alleles for the deeper species trees, and each species contained, on average, 6.0 (± 1.8 SD) unique alleles.

Structurama and NP.C performed significantly better (p<0.001) than other tested methods under nearly all sampling strategies, and under both tree depths (Fig. 2.1), though Structurama only moderately outperformed NP.E under the lowest sampling intensity (5 alleles, 5 loci) for the shallower trees (p=0.047). Three exceptions to this are the largest datasets (20 alleles, 100 loci) for the deeper tree (in which NP.C was not significantly better than Gaussian clustering, p=0.463), the smallest datasets for the deeper tree (in which Structurama was not significantly better than NP.E, p=0.126), and datasets including 10 alleles and 100 loci for the deeper tree (in which NP.C was not significantly better than Gaussian clustering, p=0.105). However, in all these cases, these equivalent methods were significantly outperformed by another method (i.e. Structurama or NP.C, p<0.001). With smaller numbers of alleles sampled per locus, NP.C typically outperformed Structurama, whereas with larger numbers of alleles, Structurama typically outperformed NP.C. Similarly, with smaller numbers of alleles, NP.E typically outperformed Gaussian clustering, but with larger numbers of alleles Gaussian clustering outperformed NP.E.

Despite the higher accuracy of NP.C at lower numbers of sampled alleles, Structurama had the lowest percent of incorrectly assigned samples, regardless of sampling strategy (Fig. 2.2). Further, all methods except Structurama failed in some cases to delimit even the most
1.7 deeply divergent species (i.e. those that diverged from all other species 6N or 12N generations ago), lumping them with other species at the exclusion of lineages more closely related to the latter species. However, Structurama only failed to delimit these deeply divergent species under the least intense sampling strategies (i.e. five loci for any number of alleles or 10 loci and five alleles). In all sampling strategies with 10 or more loci sequenced, except when five alleles were sequenced for 10 loci, Structurama successfully detected all divergences greater than approximately 2N generations. With at least 25 loci sequenced, Structurama successfully detected all divergences greater than approximately 1N.

2.3.1. Species Delimitation using Structurama

The effect of number of alleles on the accuracy of species delimitation by Structurama varied depending on the number of loci sampled (Fig. 2.3, Table 2.1, Table 2.2). With small numbers of loci (<25), accuracy generally increases with increasing numbers of alleles. However, when sampling a large number of loci (>50), the accuracy generally decreases with
increasing numbers of alleles. When sampling only five alleles, there was generally a greater
sensitivity to the number of loci sampled, and accuracy increased significantly for all increases in
numbers of loci except from 75 to 100 loci (p<0.024; Fig. 2.3, Table 2.3, Table 2.4). With
greater than five alleles sampled, a significant increase in accuracy was detected for the increases
from five to 10 loci (p<0.001), and no significant increase in accuracy was detected by
increasing the number of loci beyond 25 (p>0.514).

The majority of the errors in species delimitations with Structurama were a result of over-
lumping of species (Figs. 2.2, 2.4, 2.5, 2.6), typically involving recently diverged (<1.5N
generations) sister species. Over 90% of the species over-lumped by Structurama were sister
taxa, and most other cases of over-lumped species involved lumping of closely related, three-
species clades. We detected some instances of over-splitting of species and of incorrectly
assigned samples with small datasets (five alleles and <25 loci, or 10 alleles and <10 loci);
however, when sampling larger datasets, we found almost no instances of incorrectly assigned
samples or of over-split species (Figs. 2.2, 2.4, 2.5, 2.6).
2.3.2. Species Delimitation using Nonparametric Delimitation

In the case of nonparametric delimitation using the true coalescent gene genealogies, the effect of increasing the number of alleles varied depending on the scale of the increase, and, to a lesser extent, the tree depth (Fig. 2.3, Table 2.1, Table 2.2). In the case of the shallower trees, accuracy generally decreased with increasing numbers of alleles, though this effect was not significant for the increase from five to 10 alleles when sampling less than fifty loci (p>0.077), or when sampling only five loci (p>0.531). For the deeper tree depths, the increase from five to 10 alleles increased accuracy when sampling five or ten loci (though non-significantly, p>0.074), but decreased accuracy when sampling large numbers of loci (>25). The increase from 10 to 20 significantly decreased accuracy when sampling five loci (p=0.005), but had no significant impact on accuracy when sampling greater numbers of loci (p>0.498).

The effect of number of loci sampled on delimitation via NP.C also varied dependent on the tree depth and the number of alleles sampled (Fig. 2.3, Table 2.3, Table 2.4). For the
Table 2.1. Significance of change in accuracy of species delimitation with increasing numbers of alleles per species for shallower species trees, 6N total tree depth. T-scores greater than zero indicate an increase in accuracy with increased sampling intensity, t-scores less than zero indicate decreased sampling intensity. P-values are corrected for multiple comparisons via Bonferroni correction (p*-value = p-value X number of comparisons (2)). Values significant at the α=0.05 level after Bonferroni correction are in bold.

<table>
<thead>
<tr>
<th>Δ No.</th>
<th>All #s Loci</th>
<th>5 Loci</th>
<th>t</th>
<th>p*-value</th>
<th>10 Loci</th>
<th>t</th>
<th>p*-value</th>
<th>25 Loci</th>
<th>t</th>
<th>p*-value</th>
<th>50 Loci</th>
<th>t</th>
<th>p*-value</th>
<th>75 Loci</th>
<th>t</th>
<th>p*-value</th>
<th>100 Loci</th>
<th>t</th>
<th>p*-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gaussian Clustering</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5→10</td>
<td>4.013</td>
<td>0.001</td>
<td>2.041</td>
<td>0.219</td>
<td>0.100</td>
<td>1.000</td>
<td>0.089</td>
<td>1.000</td>
<td>2.422</td>
<td>0.086</td>
<td>0.072</td>
<td>1.000</td>
<td>3.705</td>
<td>0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10→20</td>
<td>2.314</td>
<td>0.114</td>
<td>0.776</td>
<td>1.000</td>
<td>-1.258</td>
<td>1.000</td>
<td>-2.625</td>
<td>0.050</td>
<td>-3.002</td>
<td>0.017</td>
<td>-0.620</td>
<td>1.000</td>
<td>-1.349</td>
<td>0.889</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Structurama</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5→10</td>
<td>1.920</td>
<td>0.289</td>
<td>2.468</td>
<td>0.077</td>
<td>2.411</td>
<td>0.089</td>
<td>-0.907</td>
<td>1.000</td>
<td>-2.805</td>
<td>0.030</td>
<td>-3.262</td>
<td>0.008</td>
<td>0.922</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10→20</td>
<td>4.813</td>
<td>&lt;0.001</td>
<td>2.352</td>
<td>0.103</td>
<td>-1.616</td>
<td>0.546</td>
<td>-1.136</td>
<td>1.000</td>
<td>-1.750</td>
<td>0.416</td>
<td>-1.915</td>
<td>0.292</td>
<td>1.626</td>
<td>0.523</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NP.C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5→10</td>
<td>-0.895</td>
<td>1.000</td>
<td>-0.379</td>
<td>1.000</td>
<td>-1.946</td>
<td>0.273</td>
<td>-3.398</td>
<td>0.005</td>
<td>-3.481</td>
<td>0.004</td>
<td>-5.163</td>
<td>&lt;0.001</td>
<td>-5.944</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10→20</td>
<td>-1.631</td>
<td>0.531</td>
<td>-3.553</td>
<td>0.003</td>
<td>-3.971</td>
<td>0.001</td>
<td>-4.004</td>
<td>0.001</td>
<td>-3.334</td>
<td>0.006</td>
<td>-3.549</td>
<td>0.003</td>
<td>-8.234</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NP.E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5→10</td>
<td>-2.818</td>
<td>0.029</td>
<td>-4.930</td>
<td>&lt;0.001</td>
<td>-8.370</td>
<td>&lt;0.001</td>
<td>-13.450</td>
<td>&lt;0.001</td>
<td>-13.604</td>
<td>&lt;0.001</td>
<td>-14.095</td>
<td>&lt;0.001</td>
<td>-21.370</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10→20</td>
<td>-7.593</td>
<td>&lt;0.001</td>
<td>-11.115</td>
<td>&lt;0.001</td>
<td>-13.826</td>
<td>&lt;0.001</td>
<td>-13.681</td>
<td>&lt;0.001</td>
<td>-17.020</td>
<td>&lt;0.001</td>
<td>-16.884</td>
<td>&lt;0.001</td>
<td>-30.959</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.2. Significance of change in accuracy of species delimitation with increasing numbers of alleles per species for deeper species trees, 12N total tree depth. T-scores greater than zero indicate an increase in accuracy with increased sampling intensity, t-scores less than zero indicate decreased sampling intensity. P-values are corrected for multiple comparisons via Bonferroni correction (p*-value = p-value X number of comparisons (2)). Values significant at the α=0.05 level after Bonferroni correction are in bold.

<table>
<thead>
<tr>
<th>Δ No.</th>
<th>All #s Loci</th>
<th>5 Loci</th>
<th>t</th>
<th>p*-value</th>
<th>10 Loci</th>
<th>t</th>
<th>p*-value</th>
<th>25 Loci</th>
<th>t</th>
<th>p*-value</th>
<th>50 Loci</th>
<th>t</th>
<th>p*-value</th>
<th>75 Loci</th>
<th>t</th>
<th>p*-value</th>
<th>100 Loci</th>
<th>t</th>
<th>p*-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gaussian Clustering</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5→10</td>
<td>-0.817</td>
<td>1.000</td>
<td>-0.471</td>
<td>1.000</td>
<td>1.656</td>
<td>0.505</td>
<td>1.423</td>
<td>0.790</td>
<td>1.322</td>
<td>0.946</td>
<td>1.369</td>
<td>0.870</td>
<td>1.834</td>
<td>0.335</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10→20</td>
<td>-1.641</td>
<td>0.519</td>
<td>-2.247</td>
<td>0.134</td>
<td>-2.442</td>
<td>0.082</td>
<td>-1.460</td>
<td>0.737</td>
<td>-1.564</td>
<td>0.605</td>
<td>-1.492</td>
<td>0.694</td>
<td>-4.474</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Structurama</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5→10</td>
<td>3.606</td>
<td>0.002</td>
<td>1.926</td>
<td>0.285</td>
<td>1.830</td>
<td>0.351</td>
<td>-1.817</td>
<td>0.361</td>
<td>-3.288</td>
<td>0.007</td>
<td>-3.498</td>
<td>0.004</td>
<td>0.160</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10→20</td>
<td>3.543</td>
<td>0.003</td>
<td>3.282</td>
<td>0.007</td>
<td>-0.445</td>
<td>1.000</td>
<td>-1.136</td>
<td>1.000</td>
<td>-1.915</td>
<td>0.292</td>
<td>-3.000</td>
<td>0.017</td>
<td>0.991</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NP.C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5→10</td>
<td>2.481</td>
<td>0.074</td>
<td>0.391</td>
<td>1.000</td>
<td>-4.927</td>
<td>&lt;0.001</td>
<td>-6.271</td>
<td>&lt;0.001</td>
<td>-4.465</td>
<td>&lt;0.001</td>
<td>-7.288</td>
<td>&lt;0.001</td>
<td>-7.359</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10→20</td>
<td>-3.418</td>
<td>0.005</td>
<td>-1.662</td>
<td>0.498</td>
<td>-1.397</td>
<td>0.828</td>
<td>0.402</td>
<td>1.000</td>
<td>-1.295</td>
<td>0.992</td>
<td>1.025</td>
<td>1.000</td>
<td>-2.517</td>
<td>0.060</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NP.E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5→10</td>
<td>-0.528</td>
<td>1.000</td>
<td>-4.987</td>
<td>&lt;0.001</td>
<td>-12.395</td>
<td>&lt;0.001</td>
<td>-16.801</td>
<td>&lt;0.001</td>
<td>-18.165</td>
<td>&lt;0.001</td>
<td>-20.181</td>
<td>&lt;0.001</td>
<td>-24.346</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10→20</td>
<td>-11.807</td>
<td>&lt;0.001</td>
<td>-14.434</td>
<td>&lt;0.001</td>
<td>-13.808</td>
<td>&lt;0.001</td>
<td>-18.857</td>
<td>&lt;0.001</td>
<td>-23.304</td>
<td>&lt;0.001</td>
<td>-20.781</td>
<td>&lt;0.001</td>
<td>-39.820</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
shallower trees, the number of loci generally had a stronger effect when sampling a small number of alleles: with five alleles sampled, accuracy increased up to fifty loci, whereas with 20 alleles sampled, the number of loci had no significant effect. Results were similar for the deeper trees, though with five alleles sampled, accuracy improved up to 25 loci.

Table 2.3. Significance of change in accuracy of species delimitation with increasing numbers of loci for shallower species trees, 6N total tree depth. T-scores greater than zero indicate an increase in accuracy with increased sampling intensity, t-scores less than zero indicate decreased sampling intensity. P-values are corrected for multiple comparisons via Bonferroni correction (p*-value = p-value X number of comparisons (5). Values significant at the α=0.05 level after Bonferroni correction are in bold.

<table>
<thead>
<tr>
<th>Δ No. Loci</th>
<th>5 Alleles</th>
<th>10 Alleles</th>
<th>20 Alleles</th>
<th>All #s Alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t</td>
<td>p*-value</td>
<td>t</td>
<td>p*-value</td>
</tr>
<tr>
<td>Gaussian Clustering</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5→10</td>
<td>2.490</td>
<td>0.072</td>
<td>1.211</td>
<td>1.000</td>
</tr>
<tr>
<td>10→25</td>
<td>2.019</td>
<td>0.231</td>
<td>0.103</td>
<td>1.000</td>
</tr>
<tr>
<td>25→50</td>
<td>3.317</td>
<td>0.006</td>
<td>3.978</td>
<td>0.001</td>
</tr>
<tr>
<td>50→75</td>
<td>-0.621</td>
<td>1.000</td>
<td>2.140</td>
<td>0.174</td>
</tr>
<tr>
<td>75→100</td>
<td>2.348</td>
<td>0.104</td>
<td>1.068</td>
<td>1.000</td>
</tr>
<tr>
<td>Structurama</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5→10</td>
<td>9.451</td>
<td>&lt;0.001</td>
<td>8.133</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>10→25</td>
<td>5.599</td>
<td>&lt;0.001</td>
<td>5.181</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>25→50</td>
<td>4.214</td>
<td>&lt;0.001</td>
<td>1.136</td>
<td>1.000</td>
</tr>
<tr>
<td>50→75</td>
<td>2.879</td>
<td>0.024</td>
<td>0.705</td>
<td>1.000</td>
</tr>
<tr>
<td>75→100</td>
<td>1.990</td>
<td>0.247</td>
<td>0.815</td>
<td>1.000</td>
</tr>
<tr>
<td>NP.C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5→10</td>
<td>2.869</td>
<td>0.025</td>
<td>2.352</td>
<td>0.103</td>
</tr>
<tr>
<td>10→25</td>
<td>6.114</td>
<td>&lt;0.001</td>
<td>4.255</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>25→50</td>
<td>2.627</td>
<td>0.050</td>
<td>0.219</td>
<td>1.000</td>
</tr>
<tr>
<td>50→75</td>
<td>0.623</td>
<td>1.000</td>
<td>-0.018</td>
<td>1.000</td>
</tr>
<tr>
<td>75→100</td>
<td>0.887</td>
<td>1.000</td>
<td>-1.256</td>
<td>1.000</td>
</tr>
<tr>
<td>NP.E</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5→10</td>
<td>1.819</td>
<td>0.359</td>
<td>0.024</td>
<td>1.000</td>
</tr>
<tr>
<td>10→25</td>
<td>2.505</td>
<td>0.069</td>
<td>-1.284</td>
<td>1.000</td>
</tr>
<tr>
<td>25→50</td>
<td>2.305</td>
<td>0.116</td>
<td>-2.094</td>
<td>0.194</td>
</tr>
<tr>
<td>50→75</td>
<td>2.624</td>
<td>0.050</td>
<td>1.065</td>
<td>1.000</td>
</tr>
<tr>
<td>75→100</td>
<td>0.262</td>
<td>1.000</td>
<td>-1.362</td>
<td>0.881</td>
</tr>
</tbody>
</table>

The most frequent source of error in species delimitations with NP.C was over-lumped species (Figs. 2.2, 2.5, 2.6). As with other delimitation methods, many of these over-lumped species involved recently diverged sister species. However, many cases of species over-lumping with NP.C also involved more deeply divergent, non-sister species (occasionally involving even the deepest divergences in the simulated species trees of 6N or 12N generations), often at the exclusion of other, more closely related species. While less prevalent than over-lumping species, over-splitting species and incorrectly assigning species were also common sources of error in species delimitations based on NP.C, regardless of the sampling strategy (Figs. 2.2, 2.6).

With one exception (the smallest datasets for the deeper trees), the accuracy of nonparametric delimitation based on estimated gene genealogies decreased significantly with increasing number of alleles sampled, regardless of the number of loci sampled or the total tree depths (p<0.029; Fig. 2.3, Table 2.3, Table 2.4). In general, when sampling a small number of
Further, unlike other methods of species delimitation, overcombined, with several samples from each of multiple species lumped into a single species. Errors in species delimitations from NP.E resulted from overcombination of species and incorrectly assigning samples were increasingly common. The prevalence of these errors of over-splitting species and incorrectly assigning samples increases with larger numbers of loci, to the point that over 20% of the samples were incorrectly assigned with the larger datasets (Fig. 2.2, Table 2.3, Table 2.4).

Table 2.4. Significance of change in accuracy of species delimitation with increasing numbers of loci for deeper species trees, 12N total tree depth. T-scores greater than zero indicate an increase in accuracy with increased sampling intensity, t-scores less than zero indicate decreased sampling intensity. P-values are corrected for multiple comparisons via Bonferroni correction (p*-value = p-value X number of comparisons (5). Values significant at the α=0.05 level after Bonferroni correction are in bold.

<table>
<thead>
<tr>
<th>Δ No. Loci</th>
<th>5 Alleles t</th>
<th>p* -value</th>
<th>10 Alleles t</th>
<th>p* -value</th>
<th>20 Alleles t</th>
<th>p* -value</th>
<th>All #s Alleles t</th>
<th>p* -value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian Clustering</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5→10</td>
<td>1.642</td>
<td>0.519</td>
<td>2.592</td>
<td>0.055</td>
<td>2.227</td>
<td>0.141</td>
<td>3.549</td>
<td>0.002</td>
</tr>
<tr>
<td>10→25</td>
<td>2.024</td>
<td>0.228</td>
<td>4.775</td>
<td>&lt;0.001</td>
<td>4.439</td>
<td>&lt;0.001</td>
<td>6.198</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>25→50</td>
<td>1.877</td>
<td>0.317</td>
<td>3.091</td>
<td>0.013</td>
<td>4.395</td>
<td>&lt;0.001</td>
<td>4.925</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>50→75</td>
<td>1.498</td>
<td>0.687</td>
<td>2.391</td>
<td>0.093</td>
<td>2.875</td>
<td>0.025</td>
<td>3.328</td>
<td>0.005</td>
</tr>
<tr>
<td>75→100</td>
<td>0.819</td>
<td>1.000</td>
<td>2.073</td>
<td>0.204</td>
<td>2.932</td>
<td>0.021</td>
<td>2.478</td>
<td>0.069</td>
</tr>
<tr>
<td>Structurama</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5→10</td>
<td>10.325</td>
<td>&lt;0.001</td>
<td>7.212</td>
<td>&lt;0.001</td>
<td>5.898</td>
<td>&lt;0.001</td>
<td>13.321</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>10→25</td>
<td>5.105</td>
<td>&lt;0.001</td>
<td>4.573</td>
<td>&lt;0.001</td>
<td>1.682</td>
<td>0.479</td>
<td>6.689</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>25→50</td>
<td>2.855</td>
<td>0.026</td>
<td>0.000</td>
<td>1.000</td>
<td>-0.815</td>
<td>1.000</td>
<td>1.129</td>
<td>1.000</td>
</tr>
<tr>
<td>50→75</td>
<td>2.915</td>
<td>0.022</td>
<td>0.705</td>
<td>1.000</td>
<td>0.000</td>
<td>1.000</td>
<td>2.148</td>
<td>0.162</td>
</tr>
<tr>
<td>75→100</td>
<td>2.435</td>
<td>0.083</td>
<td>1.647</td>
<td>0.514</td>
<td>-0.575</td>
<td>1.000</td>
<td>2.247</td>
<td>0.127</td>
</tr>
<tr>
<td>NP.C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5→10</td>
<td>3.692</td>
<td>0.002</td>
<td>1.722</td>
<td>0.441</td>
<td>2.903</td>
<td>0.023</td>
<td>4.821</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>10→25</td>
<td>5.267</td>
<td>&lt;0.001</td>
<td>0.212</td>
<td>1.000</td>
<td>0.155</td>
<td>1.000</td>
<td>2.898</td>
<td>0.020</td>
</tr>
<tr>
<td>25→50</td>
<td>0.937</td>
<td>1.000</td>
<td>-0.284</td>
<td>1.000</td>
<td>1.741</td>
<td>0.424</td>
<td>1.454</td>
<td>0.735</td>
</tr>
<tr>
<td>50→75</td>
<td>0.571</td>
<td>1.000</td>
<td>2.094</td>
<td>0.194</td>
<td>-0.680</td>
<td>1.000</td>
<td>0.680</td>
<td>1.000</td>
</tr>
<tr>
<td>75→100</td>
<td>1.276</td>
<td>1.000</td>
<td>-2.458</td>
<td>0.079</td>
<td>1.348</td>
<td>0.904</td>
<td>0.447</td>
<td>1.000</td>
</tr>
<tr>
<td>NP.E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5→10</td>
<td>3.185</td>
<td>0.010</td>
<td>-1.579</td>
<td>0.588</td>
<td>-6.430</td>
<td>&lt;0.001</td>
<td>-1.798</td>
<td>0.366</td>
</tr>
<tr>
<td>10→25</td>
<td>3.103</td>
<td>0.012</td>
<td>-4.849</td>
<td>&lt;0.001</td>
<td>-4.673</td>
<td>&lt;0.001</td>
<td>-3.306</td>
<td>0.005</td>
</tr>
<tr>
<td>25→50</td>
<td>1.440</td>
<td>0.766</td>
<td>-3.928</td>
<td>0.001</td>
<td>-7.148</td>
<td>&lt;0.001</td>
<td>-4.570</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>50→75</td>
<td>0.493</td>
<td>1.000</td>
<td>-0.934</td>
<td>1.000</td>
<td>-4.316</td>
<td>&lt;0.001</td>
<td>-2.097</td>
<td>0.184</td>
</tr>
<tr>
<td>75→100</td>
<td>1.136</td>
<td>1.000</td>
<td>-0.365</td>
<td>1.000</td>
<td>-0.474</td>
<td>1.000</td>
<td>0.346</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Errors in species delimitations from NP.E varied dependent upon the sampling strategy, but frequently involved over-lumped species, over-split species, and incorrectly assigned samples (Figs. 2.2, 2.5, 2.6). When sampling only five alleles per species, most of the errors in species delimitations from NP.E resulted from over-lumping of species or incorrectly assigning samples, though over-split species were also frequently detected. With larger datasets (10 or 20 alleles per species), over-lumping of species was still a common source of error; however, over-splitting species and incorrectly assigning samples were increasingly common. The prevalence of these errors of over-splitting species and incorrectly assigning samples increases with larger numbers of loci, to the point that over 20% of the samples were incorrectly assigned with the larger datasets (Fig. 2.2). Errors in species delimitations from NP.E were also frequently combined, with several samples from each of multiple species lumped into a single species. Further, unlike other methods of species delimitation, over-lumping of species in NP.E analyses...
frequently involved non-sister species, and often lumped species across the deepest divergences (6N or 12N generations) simulated in the species trees, regardless of the sampling strategy.

2.3.3. Species Delimitation using Gaussian Clustering

The effect of sampling strategy on species delimitation by Gaussian clustering is somewhat more complicated than other examined methods. For the deeper total tree depths, there was generally no significant effect of increasing the number of alleles, though the increase from 10 to 20 alleles tended to decrease accuracy (Fig. 2.3, Table 2.2). This decrease was significant when sampling a moderate number of loci (10 to 25). For the shallower trees, the increase from five to 10 alleles tended to increase accuracy (though not significant for all numbers of alleles examined; Fig. 2.3, Table 2.1). However, the increase from 10 to 20 alleles tended to increase accuracy of species delimitation when sampling a small number of loci (<25; Fig. 2.3, Table 2.1), but decrease accuracy when sampling a larger number of alleles (>25; Fig. 2.3, Table 2.1). Increasing the number of loci generally resulted in an increase in the accuracy of delimitations,
regardless of the total tree depth or the number of alleles sampled (Fig. 2.3, Table 2.3, Table 2.4). However, these increases were not significant in a number of cases.

As with species delimitation analyses using Structurama, the most prevalent source of error for species delimitation with Gaussian clustering was over-lumping of relatively recently diverged (<2.5N generations) species. Over 80% of the species over-lumped by Gaussian clustering were sister species, and many other over-lumped species were grouped with one or both other members of a relatively recently diverged, three species clade, though more deeply divergent species, even those over the deepest divergences in the species tree (i.e. 6N or 12N generations) were lumped in some, albeit rare, instances. With smaller datasets (<25 loci), over-splitting of species and incorrectly assigning samples was also an important source of error in species delimitations via Gaussian clustering (Figs. 2.2, 2.5, 2.6). While both over-splitting and incorrect assignments were both detected at all sampling intensities, both these sources of error were rare when sampling larger numbers (25 or more) of loci.
2.4. Discussion

When sampling 25 or more loci, Structurama always successfully delimited species greater than approximately 1N generations divergent and typically delimited species greater than 1.5N generations divergent, regardless of the sampling strategy. When sampling 25 or more loci, at least 90% of the species not properly delimited by Structurama were sister taxa, typically with shallow divergences. Similarly, while Gaussian clustering occasionally failed to delimit even the most divergent species, species greater than 2.5N generations divergent were typically delimited successfully when sampling 25 or more loci, and at least 80% of those species not detected were sister species with relatively shallow divergences. Thus, the imperfect performance of these methods is largely due to over-lumping of extremely shallow (less than 2N) divergences between sister species. Lineage sorting is expected to take an average of 4N ± 2N generations per locus (Degnan & Rosenberg 2006, 2009), thus the shallow divergences examined here would be expected to exhibit extensive incomplete lineage sorting for nearly all loci. Indeed, incomplete
lineage sorting was abundant in the simulated data: the average gsi for the shallower species trees was 0.808, and only an average of 1.6 species were monophyletic per locus, while the average gsi was only 0.877 for the deeper species trees, and each locus had, on average, 2.4 monophyletic species. Further, with the exception of NP.E, the most frequent source of error in species delimitations was over-lumping of closely related species. The failure of these methods to delimit species with shallow divergences is likely the result of insufficient time for lineage sorting to occur and therefore a lack of detectable differences between species. Thus, delimiting species with extremely shallow divergences should rely on other types of data, such as morphology, ecology, and reproductive isolation, or on identifying specific diagnostic loci responsible for maintaining and driving lineage divergence.

Nonparametric delimitation performs relatively well when the true, coalescent gene genealogies are known. Indeed, when sampling only five alleles, NP.C generally outperforms all other tested methods. However, when using estimated gene genealogies, nonparametric delimitation performs rather poorly, and, when sampling 10 or more alleles, performs significantly worse than any other examined method of species delimitation. The poor performance of NP.E therefore appears to be a result of errors in gene tree estimation and gene tree uncertainty, rather than poor performance of the nonparametric delimitation method itself. Regardless, nonparametric delimitation’s use in empirical study systems is limited, since all researchers will only have estimated gene trees. True coalescent gene trees can never be known with certainty and are particularly difficult to accurately estimate in recently diverged species, where species delimitation is likely to be most problematic. As such, NP.C is empirically impossible, and the problems caused by uncertainty or errors in gene tree estimation suggest that nonparametric delimitation is an ineffective method for species delimitation. As nonparametric delimitation assumes accurate point estimates of gene trees, relaxing this assumption to accommodate gene tree uncertainty, such as through repeated sampling from a distribution of gene trees rather than using a single fixed topology per locus, may improve the utility of nonparametric delimitation and improve its accuracy when using estimated gene genealogies.

In general, the accuracy of NP.E decreases with increased sampling (Fig. 2.1), particularly when sampling a large number of alleles, a somewhat unexpected and troubling observation, as with an accurate and powerful method, accuracy should increase with increasing amounts of data. Further, these decreases in accuracy occur in a complex, non-linear pattern. For example, for the shallower species trees, when sampling 20 alleles per species, the increases from five to 10 and from 25 to 50 loci result in significant decreases in accuracy, whereas the increase from 10 to 25 loci, while still resulting in decreased accuracy, is not significant. The cause of this complex pattern is not entirely clear, and may be a result of particularly misleading loci resulting in substantial decreases in accuracy, or, perhaps more likely, it may be the result of stochasticity and noise in the dataset overpowering any signal of species identity. Regardless, the decreasing accuracy of species delimitations from NP.E with increased sampling intensity is apparently due to the accumulation of errors in estimated gene genealogies, resulting in a combination of increased over-splitting of species (Fig. 2.6), increased over-lumping of species (Fig. 2.5), and increased numbers of incorrectly assigned samples (Fig. 2.2). The mean number of species identified by NP.E when sampling 100 loci and 20 alleles is 8.890 ± 0.32 for the shallower trees or 10.410 ± 0.287 for the deeper trees (Fig. 2.4), far higher than the true number of five species, or the number of species identified by any other method: the
largest number of species identified by any other method is $5.940 \pm 0.194$ for the shallower trees or $6.680 \pm 0.183$ for the deeper trees (Fig. 2.4; both from Structurama with very small datasets of five loci and five alleles).

Similarly, when sampling 100 loci and 20 alleles, the average proportion of samples incorrectly assigned by NP.E is 27.3% for the deeper trees or 23.6% with the shallower trees. The only other methods with an average proportion of incorrectly assigned samples higher than 5% error rates are Structurama when sampling five alleles and five loci for the deeper trees (8.9%) or Gaussian clustering when sampling five loci at five or 20 alleles for the shallower trees (12.0% and 12.3%, respectively). The only method to, on average, incorrectly assign more than 3.8% of samples when sampling greater than five loci was NP.E, and, on average, Structurama incorrectly assigns less than 1% of samples when sampling 10 or more loci.

Despite performing significantly worse than Structurama and NP.C under most sampling strategies, Gaussian clustering performs moderately well, as most species not delimited properly are sister taxa with relatively shallow divergences. Further, while the proportions of incorrectly assigned samples are generally somewhat higher than Structurama, they are still relatively low, particularly when sampling greater than five loci. Proportions of samples incorrectly assigned by Gaussian clustering are also generally lower than the proportions with NP.C when sampling more than five alleles, or comparable to those with NP.C when sampling only five alleles and 10 or more loci (incorrect assignments are, however, generally rather high with Gaussian clustering when sampling only five loci). Similarly, while the number of over-split species were higher for Gaussian clustering than for Structurama, this number was still low under most sampling strategies, and was far lower than for NP.E. The lower accuracy of Gaussian clustering is thus apparently largely a result of lower sensitivity of the method, as evidenced by the general failure to detect divergences between 1N and 2.5N generations divergent, that are generally detected by Structurama, as well as the occasional failure to delimit deeply divergent species at the exclusion of other, more closely related species. However, the relatively complex response of the method to sampling strategy suggests it may be highly sensitive to the amount of information in the loci included. Loci with higher levels of incomplete lineage sorting may cause a strong response in terms of decreased accuracy, whereas those with low levels of incomplete lineage sorting may cause a similarly strong response in increased accuracy.

In conclusion, our study suggests that Structurama may be the most promising method among those tested herein for species delimitation. While NP.C significantly outperforms Structurama when the number of alleles sampled is small, the true coalescent gene trees are never known in empirical studies, thus NP.C is not empirically applicable. Further, Structurama has the lowest rates of incorrectly assigned samples and of over-split species among tested methods, and deeply divergent species were always detected when sampling at least 10 loci, unlike any other method examined. We acknowledge an important caveat, however. The algorithm implemented in Structurama was designed to detect intraspecific population structure by defining clusters in a way that minimizes linkage disequilibrium and maximizes Hardy-Weinberg equilibrium (Pritchard et al. 2000; Falush et al. 2003; Huelsenbeck & Andolfatto 2007; Huelsenbeck et al. 2011). The simulation strategy implemented herein did not include intraspecific phylogeographic structure, yet it is probable that in some empirical applications, divergent but conspecific populations may be identified as distinct clusters. Thus, clusters
defined by Structurama (and other methods tested in this study) are perhaps most appropriately
treated as putative genetic lineages that should be further tested, such as using methods of
species verification (e.g. Bayesian species delimitation, SpeDeSTEM). Additionally,
Structurama and the other methods tested herein provide a means to identify distinct species –
i.e. independently evolving lineages – based on available genetic data. However, genetic data
alone should not be used for the identification and description of cryptic species. Wherever
possible and informative, we recommend combining the genetic species delimitation methods
examined herein with other types of data, such as morphology, ecology, sonograms, behavior,
and reproductive data, as perhaps the most promising approach to species delimitation in
taxonomically difficult complexes.
CHAPTER 3
DIVERSIFICATION IN THE SAVANNAS OF THE SAHUL:
SYSTEMATICS, SPECIES DELIMITATION, AND NICHE EVOLUTION
IN THE CARLIA BICARINATA GROUP (SQUAMATA: SCINCIDAE)

3.1. Introduction

It is widely recognized that due to processes such as incomplete lineage sorting, heterogeneity among gene genealogies is commonplace in biological systems (Doyle 1992; Maddison 1997; Edwards et al. 2007; Knowles 2009). Further, under some biologically realistic conditions, such as short internal branch lengths, the most frequent gene genealogy may not reflect the true species phylogeny (Degnan & Rosenberg 2006, 2009; Rosenberg 2013). A result of this discordance among loci has been a shift away from single locus or concatenated analyses, and towards multi-locus species tree analyses under the multi-species coalescent (e.g. *BEAST, Heled & Drummond 2010; BEST, Liu 2008; STEM, Kubatko et al. 2009), that account for this gene genealogy heterogeneity while estimating the underlying species tree. However, the models underlying these methods of species tree estimation assume that species designations assigned a priori are accurate representations of biological reality (Liu 2008; Heled & Drummond 2010; Leaché & Fujita 2010; Rittmeyer & Austin 2012). While work focusing on the impact of errors in species delimitation on species tree estimates is limited, it is reasonable to expect that these sorts of model violations could strongly bias results. Errors in species delimitation may have the greatest impact on the resultant phylogeny when inaccuracies involve non-sister taxa, but even misassignments involving sister taxa could cause overestimated genetic diversity within species, resulting in overestimates of population sizes, and underestimates of divergence times. In part due to its importance for phylogenetic studies utilizing species tree methods and an increased interest in more objective means of delimiting species, the issue of species delimitation has recently received increased attention (Sites Jr. & Marshall 2003, 2004; Wiens 2007; O’Meara 2010; Yang & Rannala 2010; Rittmeyer & Austin 2012; Grummer et al. 2014). However, species delimitation remains a frequently overlooked aspect of phylogenetic studies, and many studies utilizing species tree methods do not specifically test the validity of the assumed species delimitation model.

Species delimitation analyses are an important aspect of phylogenetic studies, particularly studies involving recently divergent or morphologically similar species where errors in species delimitations are more frequent. However, while species delimitation and phylogenetic analyses are critical for revealing patterns of biodiversity, inferences of the specific processes that played a role in speciation, as well as those responsible for maintaining species boundaries, are limited under these sorts of analyses. One factor frequently invoked as playing an important role in driving speciation is ecological divergence (Orr & Smith 1998; Schluter 2001, 2009; Funk et al. 2006; Zink 2014), and several studies have shown ecological divergence between closely related species (Raxworthy et al. 2007; Rissler & Apodaca 2007; Kalkvik et al. 2012; Wooten & Gibbs 2012; Ahmadzadeh et al. 2013). However, some authors have argued that niche conservatism, the tendency for related species to occupy similar ecological niches, may be a more frequent pattern (Peterson et al. 1999; Wiens 2004; Wiens & Graham 2005; Wiens et al. 2010), and several studies have shown a predominance of niche conservatism (Peterson et al. 1999; Kozak
The possible role of niche divergence in driving diversification is clear, but even niche conservatism may play a role in driving diversification by isolating populations in environmentally similar refugia, limiting gene flow among populations isolated by mutually unsuitable habitat (Kozak & Wiens 2006). While limited to examining niche divergence at broad, environmental scales, rather than at the microhabitat level, environmental niche modeling, background similarity tests (Warren et al. 2008), and multivariate niche similarity analyses (McCormack et al. 2010) provide a powerful means of assessing niche divergence while accounting for difficulties due to spatial autocorrelation and differences in habitat availability between allopatrically distributed species. Despite the availability of methods such as these to assess niche similarity, the relative importance of the various ways in which niche evolution can drive or limit divergence remains unclear.

Here, we use multilocus sequence data to delimit species and estimate the phylogeny of scincid lizards in the _Carlia bicarinata_ group of the Sahul Shelf region of New Guinea, Australia, and the Aru Islands. We also combine these sequence data with occurrence information from museum records and geographic information systems (GIS) layers of broad scale environmental variables to test niche conservatism among lineages. As currently recognized, the _C. bicarinata_ group consists of three species distributed in parts of southern New Guinea, the Aru Islands, and throughout eastern Australia (Fig. 3.1), and is largely restricted to seasonally xeric savanna and woodland habitats (Ingram & Covacevich 1989; Swan & Wilson 2013). Though currently separated by marine barriers, the terrestrial biotas of New Guinea, Australia, and the Aru Islands were connected by the Sahul Shelf emergence during periods of lower sea levels, such as during Pleistocene glaciations (Hope & Aplin 1997; Voris 2000), resulting in a pattern of strong biotic similarities among these regions. Indeed, the similarity of the Aru Islands fauna to that of Australia and New Guinea, combined with its dissimilarity to that of the geographically more proximate Kei Islands played an important role in the development of early theories on biogeography (Wallace 1857, 1860). Further, within New Guinea, suitable savanna and woodland habitat is currently restricted to two isolated patches: one in the Port Moresby region of southeastern New Guinea, and one in the Trans-Fly region of south central New Guinea. However, during historically dryer periods, such as Pleistocene glaciations, suitable xeric habitats were far more widespread throughout southern New Guinea, and broad corridors of suitable habitat connected these regions (Bowler et al. 1976; Allison 1996; Hope 2007).

Among the three species of the focal group, _C. schmeltzii_ occurs in eastern Australia, from Cape York to approximately the border of Queensland and New South Wales (Fig. 3.1), and is comprised of two morphotypes: southern populations (including the type locality of _C. schmeltzii_) are robust, tricarinate (i.e. each scale has three keels), and distinctly patterned (Ingram & Covacevich 1989; Swan & Wilson 2013), while the northern populations (described as _C. prava_ by Covacevich & Ingram (1975), but subsequently synonymized with _C. schmeltzii_ by Ingram & Covacevich (1989)) are gracile, bicarinate (i.e. each scale has two keels), and largely without pattern. However, where these morphotypes come into contact in the vicinity of Townsville, Australia, both forms can be found, as well as intermediate individuals with a mix of bicarinate and tricarinate scales (Ingram & Covacevich 1989). _Carlia storri_ occurs in three geographically isolated populations: Cape York, Australia; the Aru Islands, Indonesia; and the...
Trans-Fly region of south central New Guinea (Fig. 3.1). *Carlia bicarinata* is restricted to the relatively xeric eucalypt savannas and woodlands in the vicinity of Port Moresby in southeastern New Guinea, and is currently isolated from New Guinea *C. storri* by lowland rainforest habitat, unsuitable to both species (Fig. 3.1). Recently, a putative fourth species of *C. bicarinata* group skink was collected in Amau, southeastern New Guinea (*C. sp. Amau*, Fig. 3.1), ecologically distinct from all other species in the group in its occurrence in disturbed patches of lowland rainforest habitat, east of the distribution of *C. bicarinata*.

Here, we use multilocus sequence data to 1) estimate the phylogeny of the *Carlia bicarinata* group, 2) test the monophyly of the group, specifically whether *C. schmeltzii* forms a monophyletic clade along with the remainder of the group, and 3) examine species boundaries in the group, specifically testing for speciation between the northern and southern morphotypes of *C. schmeltzii*, and among the three geographically isolated populations of *C. storri*, and determining if the putatively undescribed species, “*C. sp. Amau*,” is distinct from *C. bicarinata*. We additionally combine occurrence data and environmental GIS layers to test for environmental niche divergence among species and populations of the *C. bicarinata* group.

![Figure 3.1. Distributions and collection locality data from HerpNet and OZCAM for *C. bicarinata* group skinks. Distributions obtained from environmental niche models in maxent, converted to binary predicted presences using a threshold of the minimum training value. Note that points represent occurrence data from museum records used in analyses of environmental niche, but do not all represent sampling localities included in genetic analyses. A. Distribution of northern and southern morphotypes of *C. schmeltzii*. Yellow indicates the northern morphotype, red indicates the southern morphotype, and orange indicates the predicted occurrence of both morphotypes. B. Distributions of *C. bicarinata* group skinks, exclusive of *C. schmeltzii*. Teal is the locality for *C. sp. Amau*, green indicates *C. bicarinata*, purple indicates New Guinea *C. storri*, pink refers to Aru Islands *C. storri*, and blue indicates Australian *C. storri.*](image-url)
3.2. Methods

3.2.1. Sampling and DNA Sequencing

We collected 41 ingroup samples spanning the geographic distribution of the *Carlia bicarinata* group (Fig. 3.2, Table C.1), including all three recognized species, both the northern and southern forms of *C. schmeltzii*, and the Australian and New Guinean populations of *C. storri*, as well as two undescribed species, one from southeastern Papua New Guinea, and one from the Aru Islands, Indonesia. We also collected 23 outgroup samples, including representatives of two closely related genera, *Lygisaurus* and *Liburnascincus*, formerly included within *Carlia* (Stuart-Fox et al. 2002; Dolman & Hugall 2008), a representative of the related genus *Emoia*, and seven of the nine recognized species groups of the genus *Carlia* (Zug 2010). Sequences of an eighth species group, the *C. rhomboidalis* group, were obtained from GenBank (Table C.1). Unfortunately, we were unable to obtain samples for the ninth species group, the *C. peronii* group.

Figure 3.2. Sampling localities for tissue samples included in the present study. Red triangles: southern morphotype *C. schmeltzii*; yellow triangle: northern morphotype *C. schmeltzii*; blue circle: Australian *C. storri*; pink circle: Aru Islands *C. storri*; purple circle: New Guinea *C. storri*; green square: *C. bicarinata*; teal square: *C. sp. Amau*. 
Whole genomic DNA was extracted from all samples using Qiagen DNeasy Blood & Tissue Kits (Valencia, CA, USA), as per manufacturer’s instructions, or via salt extractions (Fetzner 1999). One mitochondrial and eight nuclear loci were amplified for all samples as in Austin et al. (2010), using the primers and annealing temperatures in Table C.2. Amplicons were then Sanger sequenced in both directions using the amplification primers by Beckman Coulter (Danvers, MA, USA). Sequences were visually verified and complementary strands assembled in Geneious ver. 6.1.2; heterozygous sites in the nuclear loci were identified via visual inspection of the chromatograms, and using the heterozygotes plugin in Geneious.

Cleaned sequences were then aligned in MUSCLE ver. 3.8.31 (Edgar 2004), with a maximum of 1000 iterations. Alleles for nuclear loci were then identified in PHASE ver. 2.1.1 (Stephens & Donnelly 2003), using a custom python script, phaser, to facilitate converting aligned sequences to PHASE input, and PHASE output to aligned sequence files. PHASE analyses were run with 1000 iterations of burnin, followed by 5000 iterations, sampling every fifth generation. Standard IUPAC ambiguity codes were retained for any heterozygous sites that could not be phased with high posterior probability (>0.95). The best-fit model of sequence evolution was then estimated for each nuclear locus and for the mitochondrial locus, partitioned by codon position, using the corrected Akaike information criterion (cAIC) in jModelTest ver. 2.1.4 (Posada 2008).

### 3.2.2. Concatenated Phylogenetic Analyses

Phylogenetic relationships were estimated from the concatenated mitochondrial and unphased nuclear sequences via both Bayesian and maximum likelihood analyses, partitioned by locus and, in the case of the mitochondrial locus, codon position. Maximum likelihood analyses were conducted in GARLI ver. 2.01 (Zwickl 2006) using the best-fit models of sequence evolution for each partition, as estimated in jModelTest, with 48 search replicates to ensure the maximum likelihood solution had been found. Branch support was assessed via 1000 nonparametric bootstrap replicates. Bayesian phylogenetic analyses were conducted via two replicate runs, each consisting of four chains, for 20 million iterations, sampling every 1000 iterations, in MrBayes ver. 3.2.2 (Huelsenbeck & Ronquist. 2001; Ronquist & Huelsenbeck 2003; Ronquist et al. 2012). Models of sequence evolution for each partition were sampled from general-time-reversible (GTR) model space simultaneously with estimation of the phylogeny (Huelsenbeck et al. 2004). The model of among-site rate heterogeneity (i.e. none, proportion of invariant sites, or a gamma distribution) was set based on the best-fit model of sequence evolution from jModelTest. Bayesian phylogenetic analyses have been shown to be prone to branch-length inflation under the default prior settings of MrBayes (Marshall 2010; Brown et al. 2010). Therefore, we used a compound Dirichlet prior on branch- and tree-lengths, with an unconstrained gamma distribution on branch lengths ($\alpha = \beta = 1.0$). This prior has been shown to be more robust for estimating tree lengths and avoiding branch length inflation (Rannala et al. 2012; Zhang et al. 2012). We assessed convergence by inspecting the effective sample sizes (ESSs), all of which were in excess of 200, and traces of all parameters in Tracer ver. 1.6, and by comparing the posterior probabilities of all splits between runs in AWTY (Nylander et al. 2008).

### 3.2.3. Species Delimitation and Species Tree Inference

While concatenated phylogenetic analyses can often provide useful information on the phylogenetic relationships among taxa, in many cases, particularly those involving short internal
branches, these analyses may not accurately reflect true phylogenetic relationships due to processes such as incomplete lineage sorting (Maddison 1997; Edwards et al. 2007). Therefore, we also estimated the species-level phylogeny for the *Carlia bicarinata* group under the multi-species coalescent in *BEAST* ver. 2.1.0 (Heled & Drummond 2010; Bouckaert et al. 2013) in order to account for heterogeneity among gene genealogies. Prior to species tree analyses, we used several methods for species delimitation, both to compare among species delimitation methods and to ensure accurate species delimitations, since the multi-species coalescent model implemented in *BEAST* is dependent on accurate species delimitations (Heled & Drummond 2010). We focus on four methods for species delimitation, including Bayes factor delimitations (BFD, Grummer et al. 2014) and Bayesian species delimitations (Yang & Rannala 2010), both of which estimate species limits in a multi-species coalescent framework, but require *a priori* assignments of samples to putative species, as well as Bayesian inference of population structure under a Dirichlet process prior (Huelsenbeck & Andolfatto 2007; Huelsenbeck et al. 2011) and Gaussian clustering (Hausdorf & Hennig 2010), both of which use clustering algorithms to group samples into populations, though not under a multi-species coalescent framework.

Species delimitation via Gaussian clustering involves combining genetic distance matrices from each locus into a single distance matrix, using non-metric multidimensional scaling (NMDS) to convert the distance matrix into a similarity matrix, and then using Gaussian clustering to delimit species (Hausdorf & Hennig 2010; Rittmeyer & Austin 2012). We first calculated genetic distances for the each codon position of the mitochondrial locus using the best-fit models of sequence evolution from jModelTest in PAUP* ver. 4.0b10 (Swofford 2003), and combined these into a single distance matrix using standardized distances in pofad ver 1.03 (Joly & Bruneau 2006). We then calculated genetic distances for each of the phased nuclear loci using the best-fit models of sequence evolution from jModelTest in PAUP*, and combined these with the mitochondrial distance matrix using standardized distances in pofad to create a single multilocus genetic distance matrix. We then used the MASS package (Venables & Ripley 2002) in R ver. 3.0.2 to implement NMDS on this multilocus genetic distance matrix with a dimensionality of three, which resulted in a stress value of less than 10%. Gaussian clustering was then used to delimit species using the prabclus (Hausdorf 2012) and mclust (Fraley & Raftery 2006) packages in R. Analyses were conducted both excluding noise detection, and including nearest neighbor-based noise detection with a tuning constant of one (approximately equivalent to the value of the smallest integer greater than or equal to the number of samples divided by 40, as suggested by Hausdorf & Hennig, 2010).

Bayesian analyses of population structure under a Dirichlet process prior were implemented in Structurama (Huelsenbeck & Andolfatto 2007; Huelsenbeck et al. 2011), and were run for 1.5 million iterations, sampling every 100 iterations, with the first 5000 samples discarded as burnin. Structurama requires allelic data, and assumes that loci are unlinked; therefore, we collapsed each locus into allele calls using a custom Python script, Seq2Struct. Five analyses were run, with varying values for the $\alpha$ parameter of the Dirichlet process prior, set to provide prior mean numbers of populations ($E(k)$) of three, six, nine, 12, and 15, in order to test the sensitivity of the posterior to a variety of prior distributions on the number of populations.
Bayesian species delimitation (BSD) uses a reversible-jump MCMC algorithm to move among species delimitation models by collapsing or splitting nodes on a guide tree. The guide tree was estimated for *C. bicarinata* and *C. fusca* group samples in *BEAST* under a Yule speciation model and independent uncorrelated lognormal (UCLD) relaxed clocks for each locus and mitochondrial partition, with species delimitations set based on the results of Structurama. Priors on the UCLD mean rates were set as diffuse gamma distributions with a shape of 0.001 and a scale of 1000. Analyses to infer the guide tree were run for 500 million iterations, sampling every 10,000, with the first 80% discarded as burnin. We assessed convergence by examining the effective sample sizes (ESSs), all of which were in excess of 200, and traces of all parameters in Tracer ver. 1.6. The resultant maximum clade credibility tree was used as the guide tree in BSD analyses. We performed BSD analyses in bpp ver. 2.2 (Yang & Rannala 2010) using species delimitation algorithm 0 with a fine-tuning parameter $\varepsilon = 15.0$, following Leaché and Fujita (2010). Analyses were run for 1 million iterations, sampling every 100 iterations, following a burnin period of 20,000 iterations. BSD analyses can be sensitive to the prior distribution on effective population sizes ($\theta$) and divergence times ($\tau$). Therefore, we replicated analyses using three sets of prior distributions on these parameters, following Leaché and Fujita (2010): $\theta$~Gamma(1,10) and $\tau$~Gamma(1,10), which sets large prior means on the effective population sizes and divergence times; $\theta$~Gamma(2,2000) and $\tau$~Gamma(2,2000), which sets small prior means on these parameters; and $\theta$~Gamma(1,10) and $\tau$~Gamma(2,2000), which sets a large prior mean on the effective population sizes, and a small prior mean on the divergence times.

BSDs assume a fixed guide tree topology. Therefore, we also implemented Bayes factor delimitations (BFD), which can test among species delimitation models while accounting for phylogenetic uncertainty. Bayes factor delimitations were conducted using *BEAST* ver. 2.1.0 (Heled & Drummond 2010; Bouckaert *et al.* 2013) with the BEASTii version 1.1.0 add-on for marginal likelihood estimation. Seven species delimitation models were developed based on the results of the Structurama analyses (Table 3.1). Marginal likelihoods for each species delimitation model were estimated in *BEAST* using path sampling (Lartillot & Philippe 2006) with 48 steps, each consisting of 50 million iterations, which was sufficient to obtain suitable effective sample sizes (ESSs). Bayes factors (BF) were calculated as twice the difference in log marginal likelihoods between competing models (Kass & Raftery 1995). Specifically, we used BFDs to test for speciation (1) between *C. bicarinata* and the Amau populations, (2) among the Australian, New Guinean, and Aru Islands populations of *C. storrii*, and (3) between the northern and southern populations of *C. schmeltzii* (Table 3.1).

Species trees were then estimated from the full dataset, in *BEAST* ver. 2.1.0 under a Yule speciation model and independent uncorrelated lognormal (UCLD) relaxed clocks for each locus and mitochondrial partition. Priors on the UCLD mean rates were set as diffuse gamma distributions with a shape of 0.001 and a scale of 1000. Species tree analyses were run for one billion iterations, sampling every 10,000 iterations, with the first 80% of samples discarded as burn-in. Convergence was assessed by examination of the ESSs, all of which were in excess of 200, and traces of all parameters in Tracer ver. 1.6.

### 3.2.4. Environmental Niche Conservatism Analyses

We used two approaches to test for environmental niche conservation among the focal taxa. The first approach, background similarity tests (BGST), uses niche overlap metrics, either
Schoener’s $D$ or Hellinger’s $I$, to test if the environmental niche models (ENMs) for two species are more or less similar than between one species and the available ‘background’ habitat of the other species (Warren et al. 2008). The second approach, multivariate analyses of niche similarity (MVNS), uses principal components analyses to determine if the environmental conditions occupied by two species are more or less similar than expected given the difference in available ‘background’ habitat between the two species (McCormack et al. 2010). Specifically, we test for environmental niche conservatism between $C$. bicarinata and the peripatrically distributed $C$. sp. Amau (Figs. 3.1, 3.2), between $C$. bicarinata and the New Guinea populations of $C$. storri, among $C$. storri from New Guinea, the Aru Islands, and Australia, and between the northern and southern morphotypes of $C$. schmeltzii.

Occurrence data for all focal taxa were obtained from HerpNet (www.herpnet.org) and OZCAM (www.ozcam.org.au), resulting in a total of 111 localities for Australian $C$. storri, 34 localities for New Guinea $C$. storri, and 69 localities for $C$. bicarinata. Unfortunately, $C$. sp. Amau is known only from a single locality, and we were only able to obtain 3 localities for Aru Islands $C$. storri. The northern and southern morphotypes of $C$. schmeltzii come into contact in Townsville, Australia (Ingram & Covacevich 1989; Swan & Wilson 2013); therefore, localities north of this region were considered as the northern morphotype, while localities south of this region were considered the southern morphotype. Because both morphotypes, as well as

Table 3.1. Marginal likelihoods and Bayes factors for various alternative species delimitation models. The 'Full Model' is the fully delimited model, including nine species: $Carlia$ bicarinata, $C$. sp. Amau, Australian $C$. storri, Aru Islands $C$. ”storri,” New Guinea $C$. ”storri,” Northern $C$. schmeltzii, Southern $C$. schmeltzii, $C$. ailanpalai, and $C$. eothen. Bayes factors for all other tested species delimitation models were calculated relative to this fully delimited model using the formula $2*(\ln ML_0 - \ln ML_1)$, where $\ln ML_0$ is the natural logarithm of the marginal likelihood of the alternate, lumped model, and $\ln ML_1$ is the natural logarithm of the marginal likelihood of the fully delimited model.

<table>
<thead>
<tr>
<th>Species Delimitation Model</th>
<th>ln Marg. Like.</th>
<th>Bayes Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Model</td>
<td>-13246.53677</td>
<td>--</td>
</tr>
<tr>
<td>$C$. schmeltzii</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lump North + South</td>
<td>-13270.82805</td>
<td>-48.58255</td>
</tr>
<tr>
<td>$C$. bicarinata</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lump $C$. bicarinata + $C$. sp. Amau</td>
<td>-13330.63286</td>
<td>-168.19217</td>
</tr>
<tr>
<td>$C$. storri</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lump all 3 pop'n's</td>
<td>-13474.00049</td>
<td>-454.92744</td>
</tr>
<tr>
<td>Lump Australia + New Guinea</td>
<td>-13352.13037</td>
<td>-211.18719</td>
</tr>
<tr>
<td>Lump Aru Islands + New Guinea</td>
<td>-13340.41918</td>
<td>-187.76482</td>
</tr>
<tr>
<td>Lump Australia + Aru Islands</td>
<td>-13349.56852</td>
<td>-206.06351</td>
</tr>
</tbody>
</table>

Table 3.1. Marginal likelihoods and Bayes factors for various alternative species delimitation models. The 'Full Model' is the fully delimited model, including nine species: $Carlia$ bicarinata, $C$. sp. Amau, Australian $C$. storri, Aru Islands $C$. ”storri,” New Guinea $C$. ”storri,” Northern $C$. schmeltzii, Southern $C$. schmeltzii, $C$. ailanpalai, and $C$. eothen. Bayes factors for all other tested species delimitation models were calculated relative to this fully delimited model using the formula $2*(\ln ML_0 - \ln ML_1)$, where $\ln ML_0$ is the natural logarithm of the marginal likelihood of the alternate, lumped model, and $\ln ML_1$ is the natural logarithm of the marginal likelihood of the fully delimited model.
morphologically intermediate individuals, are known from the Townsville area (Ingram & Covacevich 1989), localities within 25 km of Townsville were considered as belonging to both morphotypes, resulting in a total of 47 localities for the northern morphotype, and 48 for the southern morphotype.

We used a set of 22 GIS layers of environmental variables at a spatial resolution of 30 arc-seconds (~ 1 km), including the mean and standard deviation of the normalized difference vegetation index (NDVI), elevation, and 19 frequently used BioClim layers that include variables describing biologically relevant climatic factors such as precipitation, temperature, and seasonality (Table 3.2). NDVI measures were obtained from MODIS (NASA-MODIS/Terra dataset, available at http://modis-atmos.gsfc.nasa.gov/NDVI/) at 16-day intervals for a ten-year period spanning 1 January 2001 to 31 December 2010. These layers were then combined to calculate the mean and standard deviation of NDVI over this time period.

Table 3.2. Environmental layers included in environmental niche modeling analyses and analyses of niche conservatism.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>alt</td>
<td>Elevation above sea level</td>
<td>Hijmans et al. 2005</td>
</tr>
<tr>
<td>bio1</td>
<td>Annual mean temperature</td>
<td>Hijmans et al. 2005</td>
</tr>
<tr>
<td>bio2</td>
<td>Mean diurnal temperature range</td>
<td>Hijmans et al. 2005</td>
</tr>
<tr>
<td>bio3</td>
<td>Isothermality (Diurnal/Annual Temperature Range)</td>
<td>Hijmans et al. 2005</td>
</tr>
<tr>
<td>bio4</td>
<td>Temperature seasonality</td>
<td>Hijmans et al. 2005</td>
</tr>
<tr>
<td>bio5</td>
<td>Max. temperature of warmest month</td>
<td>Hijmans et al. 2005</td>
</tr>
<tr>
<td>bio6</td>
<td>Min. temperature of coldest month</td>
<td>Hijmans et al. 2005</td>
</tr>
<tr>
<td>bio7</td>
<td>Temperature annual range</td>
<td>Hijmans et al. 2005</td>
</tr>
<tr>
<td>bio8</td>
<td>Mean temperature of wettest quarter</td>
<td>Hijmans et al. 2005</td>
</tr>
<tr>
<td>bio9</td>
<td>Mean temperature of driest quarter</td>
<td>Hijmans et al. 2005</td>
</tr>
<tr>
<td>bio10</td>
<td>Mean temperature of warmest quarter</td>
<td>Hijmans et al. 2005</td>
</tr>
<tr>
<td>bio11</td>
<td>Mean temperature of coldest quarter</td>
<td>Hijmans et al. 2005</td>
</tr>
<tr>
<td>bio12</td>
<td>Annual precipitation</td>
<td>Hijmans et al. 2005</td>
</tr>
<tr>
<td>bio13</td>
<td>Precipitation of wettest month</td>
<td>Hijmans et al. 2005</td>
</tr>
<tr>
<td>bio14</td>
<td>Precipitation of driest month</td>
<td>Hijmans et al. 2005</td>
</tr>
<tr>
<td>bio15</td>
<td>Precipitation seasonality</td>
<td>Hijmans et al. 2005</td>
</tr>
<tr>
<td>bio16</td>
<td>Precipitation of wettest quarter</td>
<td>Hijmans et al. 2005</td>
</tr>
<tr>
<td>bio17</td>
<td>Precipitation of driest quarter</td>
<td>Hijmans et al. 2005</td>
</tr>
<tr>
<td>bio18</td>
<td>Precipitation of warmest quarter</td>
<td>Hijmans et al. 2005</td>
</tr>
<tr>
<td>bio19</td>
<td>Precipitation of coldest quarter</td>
<td>Hijmans et al. 2005</td>
</tr>
<tr>
<td>mean NDVI</td>
<td>Average Normalized Differential Vegetation Index (2001-2010)</td>
<td>MODIS</td>
</tr>
<tr>
<td>sd NSVI</td>
<td>Standard Deviation in Normalized Differential Vegetation Index (2001-2010)</td>
<td>MODIS</td>
</tr>
</tbody>
</table>

With the exception of the Aru Islands *C. storri* and *C. sp. Amau*, available background habitats were determined by sampling random points from a convex hull drawn around all
localities. Due to the limited number of sampling localities for the Aru Islands species, and the reasonable extent of available habitat, the entire Aru Archipelago was used as the available region for this species. For *C. sp. Amau*, because the species is only known from a single locality, we sampled available habitat from a 25 km buffer drawn around the locality.

BGST were conducted using ENMtools ver. 1.4.3 (Warren et al. 2010) with maxent ver. 3.3.3k (Phillips et al. 2006; Phillips & Dudí 2008). The limited number of sampling localities for the Aru Islands *C. storri* and *C. sp. Amau* makes it difficult to construct ENMs for these species (Pearson et al. 2006; Wisz et al. 2008). Therefore, we focused on comparing the environmental niches of the two morphotypes of *C. schmeltzii*, between Australian and New Guinea *C. storri*, and between *C bicarinata* and New Guinea *C. storri* for BGST analyses. For each comparison, a null distribution of overlap values was created by 100 pseudoreplicates of calculating the Schoener’s *D* niche similarity metric between the ENM of species A with the ENM from *n* localities sampled at random from the available background habitat for species B, where *n* is the number of observed collection localities for species B. The empirical niche similarity metrics between species A and B were then compared to these null overlap distributions to test for niche conservatism or divergence. BGST analyses were conducted comparing species A to the background habitat of species B, as well as comparing species B to the background habitat of species A; therefore, for each pair of species examined, two null distributions and two statistical tests were conducted.

For MVNS analyses, background environmental availability data for each of the 22 climatic variables were extracted from 1000 points, sampled at random from within the distribution of each species, with the exception of *C. sp. Amau*, for which 200 random points were selected from within a 25 km buffer. These data were combined with the climatic data extracted from the localities of each species, and reduced using principal components analysis (PCA) of the correlation matrix in R ver. 3.0.2, resulting in five principal components (PCs), each explaining at least 3% of the variation, and that, combined, explained over 94% of the variation. For each pair of taxa being compared, the observed difference in mean niche values was compared to a null model of background divergence along each of these PC axes. We assessed significance using 1000 jackknife replicates resampling 75% of the background points.

### 3.3. Results

The final aligned dataset consisted of 707 bp of mitochondrial sequence data, and 4,440 bp of nuclear data, with loci ranging from 212 to 813 bp (Table C.2). Phylogenetic analyses of the concatenated dataset reveal a largely strongly supported tree (Fig. 3.3), though relationships among outgroup samples are poorly supported by maximum likelihood bootstraps (MLBS). These results are also concordant with previously published phylogenies of the genus *Carlia* (Stuart-Fox et al. 2002; Dolman & Hugall 2008), with the exception of several nodes that were poorly supported in the previous studies, by maximum likelihood bootstraps in the current analysis, or both. The monophyly of the *C. bicarinata + C. fusca* group is strongly supported in the concatenated phylogeny. However, while previous studies placed *C. schmeltzii* as the sister taxon to the rest of the *C. bicarinata* group, our analyses place this species as sister to a clade composed of the remainder of the *C. bicarinata* group and the *C. fusca* group with moderate (MLBS) to strong (Bayesian posterior probability, PP) support. The two morphotypes of *C.*
Schmeltzii were covered as deeply divergent sister taxa with strong support (MLBS=100%, PP=1.0). Within the ingroup, we find strong support for a deep divergence between *C. bicarinata* + *C. sp. Amau* and the three populations of the *C. storrī* complex. Within each of these clades, we recover a deep divergence between *C. bicarinata* and *C. sp. Amau* (mean p-distance = 0.1303 at the mitochondrial ND4 locus), as well as among the New Guinea, Australia, and the Aru Islands populations of *C. storrī* (mean p-distance = 0.066 to 0.112 among the three populations at the mitochondrial ND4 locus). Finally, concatenated phylogenetic analyses reveal a strongly supported (MLBS=87.2%, PP=1.0) sister relationship between the Aru Islands *C. storrī* and Australian *C. storrī*.
Species tree analyses were largely congruent with the concatenated phylogenetic analyses, with much of the discordance occurring at poorly supported nodes among outgroup taxa (Fig. 3.4). However, within the ingroup, we recover an important discordance with the concatenated phylogeny. Rather than recovering *C. schmelzii* as sister to a clade consisting of the *C. fusca* group and the rest of the *C. bicarinata* group, the *C. fusca* group is recovered as sister to the entire *C. bicarinata* group, including *C. schmelzii*. However, this relationship is poorly supported in species tree analyses (PP=0.38). Additionally, while not discordant between concatenated and species tree analyses, the relationships among the three populations of *C. storri* is poorly supported in the species tree analysis (PP=0.46).

Figure 3.4. Species tree of *C. bicarinata* group skinks from analysis in *BEAST*. Coloration on bar corresponds to that in Figure 3.3.

3.3.1. Species Delimitation

Species delimitation analyses in Structurama were largely insensitive to the selection of prior distribution on the number of populations (Fig. 3.5); under four of the five priors on number of populations, nine clusters were recovered, corresponding to: *C. ailanpalai*, *C. eothen*, *C. bicarinata*, *C. sp. Amau*, the northern morphotype of *C. schmelzii*, the southern morphotype of *C. schmelzii*, Australian *C. storri*, New Guinea *C. storri*, and Aru Islands *C. storri*. However, when using the prior distribution with the largest mean number of populations (i.e. E(k)=15), a further subdivision between the two samples of the southern morphotype of *C. schmelzii* was recovered.
Gaussian clustering species delimitation analyses were less sensitive to detecting more recent divergence, but, when including all *C. bicarinata* and *C. fusca* group samples, were not affected by the inclusion of noise detection (Fig. 3.5). With this full dataset, four clusters were detected, corresponding to *C. fusca* group samples, both morphotypes of *C. schmeltzii*, *C. bicarinata + C. sp. Amau*, and all samples of *C. storri* (including Australian, New Guinea, and Aru Islands samples). To determine if these results represent a lack of sensitivity of the method for detecting recent divergences (as suggested by Rittmeyer and Austin, 2012), or a tendency of the method to only detect higher levels of structuring (similar to the ΔK statistic for Structure,
Evanno et al., 2005), we reanalyzed each cluster independently. These second sets of analyses were substantially incongruent between including and excluding noise detection. Within clusters, Gaussian clustering performed poorly when including noise detection, with numerous samples unassigned to clusters, and extensive subdivision of morphologically and genetically similar, syntopically collected samples (Figs. 3.2, 3.5). Due to limited sampling, Gaussian clustering with noise detection could not be performed within the C. schmeltzii cluster. However, excluding noise detection, Gaussian clustering analyses within clusters yielded results comparable to the results of Structure analyses (Fig. 3.5); the two differences were the failure of Gaussian clustering analysis to identify the northern and southern morphotypes as distinct, and the subdivision of C. eother into three clusters.

Species delimitation analyses via BSD in bpp were consistent across the tested prior distributions and starting topologies. All eight nodes separating the nine putative species (C. ailanpalai, C. eother, C. bicarinata, C. sp. Amau, northern C. schmeltzii, southern C. schmeltzii, Australian C. storrí, New Guinea C. storrí, and Aru Islands C. storrí) were split, indicating a predicted speciation event, with a posterior probability of 1.0 across all BSD analyses.

BFD strongly supports treating each of the three populations of C. storrí (Australia; Trans-Fly, New Guinea; Aru Islands) as distinct species over the alternative of lumping all three as a single species, or lumping any combination of two of these populations as a single species, by a BF of at least 187.76 (Table 3.1). BFs in excess of 10 are considered ‘decisive’ support for a model over an alternative (Kass & Raftery 1995). Thus, this represents exceptionally strong support for the treatment of the three populations as distinct species. BFD analyses also show support for splitting the northern and southern morphotypes of C. schmeltzii into distinct species with a BF of 48.58 (Table 3.1). Finally, we detect strong support for the species delimitation model treating C. sp. Amau as distinct from C. bicarinata, with a BF of 168.19 (Table 3.1).

3.3.2. Environmental Niche Conservatism Analyses
ENMs provided a good fit to the occurrence data for each of the five species modeled, with an area under the curve (AUC) statistic in excess of 0.99 for each species. However, there was some minor apparent over-prediction: parts of northern New Guinea, south of the Huon Peninsula were predicted as suitable for C. bicarinata, and parts of the Kimberley and Arnhem Land areas of northwestern Australia, and parts of Timor Island were predicted for northern C. schmeltzii and Australian C. storrí (Fig. 3.1). Analyses of environmental niche conservatism via background similarity tests (BGST) detect significant niche conservatism between each of the three pairs of species compared by rejection of one of the two null distributions (Table 3.3). However, in each comparison, the empirical niche overlap was not significantly different than the second null distribution (Table 3.3). That is, in each comparison, the niche overlaps of species A and species B were significantly more similar than expected given the overlap of species A with the background of species B, but no more similar than expected given the overlap of species B with the background of species A, which may be indicative of niche conservatism coupled with differential availability of ideal environmental conditions (Nakazato et al. 2010). No comparisons tested revealed significant niche divergence relative to background environmental availability.
Table 3.3. Results of Background similarity tests of niche conservatism. Values in left column correspond to comparison of the first species to the background of the second species, values in the right column correspond to the reverse.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Niche Overlap</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. schmeltzii North vs. C. schmeltzii South</td>
<td>0.352</td>
</tr>
<tr>
<td>Null Distribution</td>
<td>(0.207 - 0.325)</td>
</tr>
<tr>
<td>P-value</td>
<td>0.004</td>
</tr>
<tr>
<td>C. bicarinata vs. NG C. storri</td>
<td>0.26</td>
</tr>
<tr>
<td>Null Distribution</td>
<td>(0.242 - 0.304)</td>
</tr>
<tr>
<td>P-value</td>
<td>0.413</td>
</tr>
<tr>
<td>Aus C. storri vs. NG C. storri</td>
<td>0.067</td>
</tr>
<tr>
<td>Null Distribution</td>
<td>(0.039 - 0.056)</td>
</tr>
<tr>
<td>P-value</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

The first five principal components (PC) axes of environmental variables, each of which explained at least 3% of the variation and over 94% of the variation in environmental variables combined, were used for multivariate niche similarity analyses (MVNS). Details of the percent variation explained by each axis, and the primary contributing variables is provided in Table 3.4. Between the two morphotypes of C. schmeltzii, significant niche conservation was detected on four of the five PC axes, but significant divergence was detected on the fifth axis, PC5, which corresponds to vegetation and elevation (Table 3.4). *Carlia bicarinata* and *C. sp. Amau* were significantly divergent on three of five PC axes. We detected no significant difference along a fourth PC axis. However, along PC3, which described precipitation regimes, *C. sp. Amau* and *C. bicarinata* were significantly more similar than expected (Table 3.4), despite the occurrence of *C. sp. Amau* in areas with substantially higher and less seasonal precipitation than areas occupied by *C. bicarinata* (Fig. 3.6). Among *C. storri* lineages, significant niche conservatism was detected on two to three PC axes per comparison, while significant niche divergence was detected on zero to two PC axes (Table 3.4). Specifically, Australian *C. storri* were significantly divergent from the other two lineages on PC3, which relates to precipitation, as well as from New Guinea *C. storri* on PC5, which describes vegetation. No significant divergence was found between the New Guinea and Aru Islands populations. Finally, *C. bicarinata* and New Guinea *C. storri* were significantly conserved on two PC axes, PC2 and PC3, which correspond to high temperatures and to precipitation, respectively. However, these species were significantly divergent on two other axes, PC1 and PC4, which describe a combination of temperature seasonality and minimum temperatures, and isothermality (the relationship between diurnal and annual temperature ranges) and minimum precipitation, respectively (Table 3.4). Overall, MVNS analyses showed mixed results: most species examined were divergent on at least one PC axis, but all species compared were conserved on at least one PC axis. However, only the comparison between *C. bicarinata* and *C. sp. Amau* was characterized by a predominance of environmental niche divergence; all other comparisons show either a predominance of niche conservatism, or no predominant pattern.
3.4. Discussion

3.4.1. Species Delimitation

Species delimitations were concordant among three of the four methods applied, identifying seven species within the *C. bicarinata* group: *C. bicarinata*, *C. sp. Amau*, northern *C. schmeltzii*, southern *C. schmeltzii*, Australian *C. storri*, New Guinea *C. storri*, and Aru Islands *C. storri*. However, Structurama analyses using the largest prior mean number of populations (i.e. a prior mean of 15 populations) split the two samples of southern *C. schmeltzii* into distinct clusters. Gaussian clustering, based on the dataset of all *C. bicarinata* and *C. fusca* group
samples, was much less sensitive to more recent divergences, and only identified four clusters, corresponding to deeper levels of structure within the dataset. Previous work showed a similar pattern in which Gaussian clustering was less sensitive to detecting recent divergences than other methods of species delimitation, such as Structurama (Rittmeyer & Austin 2012; but see Edwards & Knowles 2014). However, our subsequent Gaussian clustering analyses within each of these clusters were more concordant with other methods of species delimitation, suggesting that this method may detect higher levels of structure initially, similar to Evanno et al.’s ΔK statistic with Structure (Evanno et al. 2005), and that species delimitation via Gaussian clustering should involve hierarchical analyses in order to delimit species level structure at multiple levels. Further, within each of the initially identified clusters, including nearest neighbor noise detection appears to be more problematic than beneficial for species delimitation via Gaussian clustering. When including noise detection, Gaussian clustering within clusters yielded extensive apparent over-splitting of species identified via other methods, and numerous samples were not assigned to clusters. It is plausible that the limited sampling within each cluster plays a role in the difficulties in applying noise detection. However, these results clearly indicate a need for investigating the consistency of species delimitations via Gaussian clustering under multiple noise detection tuning constants, and without noise detection.

Figure 3.6. Availability of environmental niche space between *C. bicarinata* (red) and *C. sp. Amau* (blue) for each principal component axis, showing the overlapping environmental availability between the two species. Vertical lines represent mean values for the observed occurrence localities for each species.
3.4.2. Phylogenetics of the C. bicarinata group

Previous phylogenetic studies of the genus *Carlia*, while including all currently recognized species in the group, were limited in their sampling, including only one morphotype of *C. schmelzii* and only one of the three disjunct populations of *C. storri* (Stuart-Fox et al. 2002; Dolman & Hugall 2008). While our sampling outside the *C. bicarinata* group was limited, our results corroborate these previous studies in recovering short, poorly supported internal branches, particularly among outgroup *Carlia* samples, suggesting that the genus *Carlia* represents a rapid radiation. Within the *C. bicarinata* group, relationships are largely well supported, with two important exceptions: relationships among the three clades of *C. storri*, and relationships among the *C. fusca* group, the two morphotypes of *C. schmelzii*, and the remainder of the *C. bicarinata* group.

Concatenated phylogenetic analyses strongly support a sister relationship between Australian *C. storri* and *C. storri* from the Aru Islands (MLBS=87.2, PP=1.0). Species tree analyses also recover this sister relationship, but with poor support (PP=0.46). These three regions are separated by shallow seas, and were connected as recently as the last glacial maximum, approximately ten to fifteen thousand years before present (Hope & Aplin 1997; Voris 2000). Thus, the rapid diversification of these three lineages is not surprising.

Unfortunately, no useful fossil calibrations are available to robustly estimate the timing of this divergence in years. However, the mean pairwise divergence among these species at the mitochondrial locus is 6.6-11.2%. Assuming the frequently used mitochondrial divergence rate of two percent per million years, this divergence likely occurred on the order of three to six million years before present. Thus, the divergence among these three populations, as well as the similarly deep divergences between *C. bicarinata* and *C. sp. Amau* and between the two morphotypes of *C. schmelzii*, likely occurred during the Pliocene, and almost certainly predates the most recent glacial cycle, during which the emergence of the Sahul shelf formed land bridges among New Guinea, Australia, and the Aru Islands. Few studies have examined divergence across the Torres Strait separating New Guinea from Australia. Of these, all found little to no divergence, but focused on species with higher vagility (e.g., Green Tree Python, *Morelia viridis*, Rawlings & Donnellan 2003; elapid snakes, Wüster et al. 2005). Thus, our observation of speciation across this barrier highlights the need for more taxonomically comprehensive examination of this barrier to determine its broader impact on speciation.

Concatenated phylogenetic analyses also show moderate to strong support (MLBS=78.5, PP=1.0) for both morphotypes of *C. schmelzii* as sister to a clade comprised of the *C. fusca* group and the rest of the *C. bicarinata* group. However, species tree analyses recover the *C. fusca* group as sister to a monophyletic *C. bicarinata* group, although the relationships among these three clades was not well resolved in species tree analyses (PP=0.37). Previous work similarly found poor support for these relationships (Dolman & Hugall 2008), and the internal branch length is short (Figs. 3.3, 3.4), suggesting that these three lineages diverged rapidly. Further, all other members of the *C. bicarinata* group are gracile species with bicarinate scales (Ingram & Covacevich 1989). While the northern morphotype of *C. schmelzii* is similar to the rest of the *C. bicarinata* group, the southern morphotype is larger, more robust, and has tricarinate scales (Ingram & Covacevich 1989; Zug 2010), more similar to *C. fusca* group skinks (Ingram & Covacevich 1989; Zug 2004, 2010). Thus, although the northern and southern morphotypes of *C. schmelzii* are strongly supported as sister taxa, the northern morphotype is
morphologically similar to the *C. bicarinata* group, while the southern morphotype is morphologically similar to the *C. fusca* group.

### 3.4.3. Taxonomic Implications

The northern morphotype of *C. schmeltzii* was described as *C. prava* by Covacevich and Ingram (1975). With additional sampling, Ingram and Covacevich (1989) subsequently synonymized *C. prava* with *C. schmeltzii*, largely on the basis of the high variation in individuals from the vicinity of Townsville, Australia, including bicarinate (northern morphotype), tricarinate (southern morphotype) and mixed individuals. Our analyses reveal deep divergence between these morphotypes (mean p-distance = 0.0945 at the mitochondrial ND4 locus), and three of the four species delimitation methods strongly support the genetic distinctiveness of these morphotypes (Fig. 3.5). The fourth method, Gaussian clustering, was previously shown to be less sensitive to recent divergences than other methods (Rittmeyer & Austin, 2012; but see Edwards & Knowles, 2014). Thus, the failure of Gaussian clustering to delimit these morphotypes as distinct may be reflective of the recency of their divergence. We therefore resurrect the taxon *Carlia prava* for the northern morphotype. We note that the occurrence of both species as well as intermediate individuals in the Townsville area (Ingram & Covacevich 1989; Swan & Wilson 2013) suggests that these species may come into contact and form a hybrid zone in this area. Distinct species frequently form hybrid zones in nature (Harrison 1993; Barton 2001); however, our current sampling is insufficient to test this hypothesis, or to estimate parameters of the possible hybrid zone, such as cline width. Further work is necessary to examine this potential hybrid zone in detail, and determine the extent to which the two species hybridize.

All four species delimitation methods also provide strong support for the recognition of the putative new species from southeastern Papua New Guinea (*C. sp. Amau*), as well as for the recognition of each of the three isolated populations of *C. storri* as distinct species. The type locality for *C. storri* is in Cape York, Australia (Dulhunty River Crossing on Telegraph Road, 110 km S of Bamaga, Cape York, Queensland, Australia; -11.833°, 142.5°; Ingram & Covacevich, 1989), thus we suggest the name *C. storri* should be restricted to the Australian population. However, no names are currently available for the New Guinea or Aru Islands populations, or for *C. sp. Amau*. We are in the process of formally describing these new species.

### 3.4.4. Environmental Niche Evolution

For each of the three pairs of taxa examined via BGST analyses, we observed an asymmetrical pattern of no significant difference in one comparison, but significant niche conservatism in the reverse. While this result may initially appear as counterintuitive, it likely reflects niche conservatism between the species, coupled with differential availability of environmental conditions (Nakazato et al. 2010). That is, this may reflect a pattern in which the two species prefer similar environmental niches, but where this preferred set of conditions is more available to one of the two species, resulting in the observed pattern of the two species being more similar than expected given the habitat availability to the more limited species, but not significantly different given the habitat availability to the species for which the preferred climatic conditions are available.
For most comparisons, MVNS analyses corroborate a pattern of limited environmental divergence among species in the *C. bicarinata* group, showing either a predominance of niche conservatism across PC axes (e.g. between *C. schmeltzii* morphotypes, Aru Islands versus New Guinea *C. storri*), or no predominant pattern of niche evolution (e.g. New Guinea *C. storri* versus Australian *C. storri*). These results, combined with the results of the BGSTs, suggest that *C. bicarinata* group skinks generally occupy environmental niches more similar than expected given habitat availability, and that niche divergence has not accompanied allopatric divergence within this group. The only comparison that did show a predominance of niche divergence was between *C. bicarinata* and *C. sp. Amau*, that were significantly divergent on three of the five PC axes, and only significantly conserved on a single PC axis (Table 3.4). However, along this PC axis, which describes precipitation and precipitation seasonality, there was very little overlap between the conditions available to the species, with regions available to *C. bicarinata* showing substantially lower and more seasonal precipitation.

Due to the limited number of sampling localities for *C. sp. Amau*, as well as for Aru Islands *C. storri*, these results should be interpreted cautiously. However, on the majority of PC axes, there is broad overlap between the available background conditions for *C. bicarinata* and for *C. sp. Amau* (Fig. 3.6), the only exception being the third PC axis, on which the two species were significantly niche conserved, suggesting that along the majority of the environmental PC axes, similar conditions are available to both species, but not occupied. Further, ecological niche models for *C. bicarinata* do not predict the area around Amau as suitable. Finally, while *C. bicarinata* and other members of the species group are restricted to eucalypt savannas and woodlands, *C. sp. Amau* was collected from disturbed habitats in lowland rainforest. Thus, this apparent environmental niche divergence is not surprising, and suggests that the apparent environmental niche divergence between *C. bicarinata* and *C. sp. Amau* is reflective of a role of niche evolution in the divergence of these species, rather than an artifact of limited sampling. It is also plausible that this niche divergence has played an important role in maintaining the speciation of these taxa. *Carlia sp. Amau* is distributed peripatrically to *C. bicarinata*, with no obvious physical barrier to dispersal. Further, xeric savanna habitat, suitable for *C. bicarinata*, was historically much more widespread in southern New Guinea during historically cooler and drier periods, such as Pleistocene glaciations (Bowler et al. 1976; Allison 1996; Hope 2007). While the level of divergence between these taxa (mean p-distance 0.1303 at the mitochondrial ND4 locus) almost certainly predates the Pleistocene, the niche divergence between these taxa likely played a role in maintaining their isolation during these periods, when the distribution of *C. bicarinata* may have expanded and encroached upon that of *C. sp. Amau*.

### 3.4.5. Conclusions

Our species delimitation analyses largely corroborated each other, generally delimiting the *C. bicarinata* group into the same seven species. Our analyses suggest that Gaussian clustering tends to initially detect only higher levels of structure, and should be applied in a hierarchical manner to detect structure at multiple levels. Finally, our results suggest that including noise detection hinders species delimitation via Gaussian clustering, at least when sample sizes are limited. We therefore suggest that, while Gaussian clustering appears to be a useful method for species delimitation, multiple analyses, including noise detection with multiple tuning constants and excluding noise detection, as well as subsequent within-cluster analyses, are critical to the use of this method.
Our analyses also reveal that the diversity of the *C. bicarinata* group is substantially underestimated: *C. schmeltzii* is a complex of at least two species (the northern morphotype, *C. prava*, and the southern morphotype, *C. schmeltzii*), *C. storri* is a complex of at least three distinct species, and a further, recently discovered species occurs in southeastern Papua New Guinea. Our phylogenetic analyses further reveal an ambiguous relationship of *C. schmeltzii* (including the northern morphotype) to the rest of the genus. While species tree analyses suggest these taxa form the sister clade to the rest of the *C. bicarinata* group, this relationship is poorly supported, and concatenated analyses suggest this taxon is sister to a clade comprised of the *C. fusca* group and the remainder of the *C. bicarinata* group. Interestingly, the northern morphotype, *C. prava*, is morphologically more similar to the *C. bicarinata* group in having a gracile body and bicarinate scales, while the southern morphotype, *C. schmeltzii*, is more similar to the *C. fusca* group in having a robust body and tricarinate scales. All analyses strongly support this sister relationship between the two morphotypes. These results highlight the need for further work to examine the relationships among these three clades, and to study the morphological evolution within this group. Further, our analyses reveal speciation among the populations assigned to *C. storri* from New Guinea, Australia, and the Aru Islands, despite their relatively recent connections during periods of low sea levels, and the previous work with other taxa that has found limited divergence across the Torres Strait separating these populations. More work is necessary to determine the extent to which this barrier has played a role in driving diversification.

Finally, analyses of niche conservatism reveal substantial environmental niche divergence between *C. bicarinata* and *C. sp. Amau*, suggesting that niche evolution may have played an important role in driving and maintaining this speciation event. However, comparisons between other species of the *C. bicarinata* group reveal either niche conservatism or no predominant pattern of niche conservatism or divergence, suggesting that, in general, niche evolution has not played a particularly important role in driving diversification in this group.
CHAPTER 4
SYSTEMATICS AND BIOGEOGRAPHY OF THE CARLIA FUSCA
GROUP (SQUAMATA: SCINCIDAE)

4.1. Introduction

4.1.1. Species Delimitation in Diverse Clades

The concept of species is among the most controversial and debated topics in modern evolutionary biology (Sokal & Crovello 1970; Mallet 2001; Lee 2003; Coyne & Orr 2004; de Queiroz 2007; Baum 2009; Hausdorf 2011). Yet as a fundamental unit in biology, delimiting species in a biologically meaningful and accurate way is critically important to a wide variety of studies (Rieseberg & Burke 2001; Agapow et al. 2004; Agapow 2005; Bickford et al. 2007; Bortolus 2008). Further, the problem of how to delimit species, particularly, but not exclusively, with genetic data, has generally received little attention until recently (Sites Jr. & Marshall 2003, 2004; Wiens 2007; Carstens et al. 2013). Interest in this non-trivial task has increased substantially, and numerous new methods for species delimitation have become available (e.g. Pons et al. 2006; Hausdorf & Hennig 2010; O’Meara 2010; Yang & Rannala 2010; Reid & Carstens 2012; Grummer et al. 2014). While some studies have aimed to investigate the accuracy of these methods, including examining the sensitivity to phylogenetic uncertainty and sampling strategy, these investigations, as well as most applications of these methods, have generally involved small numbers of species (<~10, Hausdorf & Hennig 2010; O’Meara 2010; Rittmeyer & Austin 2012; Carstens et al. 2013; Edwards & Knowles 2014). The few studies that have involved larger numbers of species have generally applied the generalized mixed Yule coalescent model, which requires a single topology (or, in the case of the Bayesian implementation, a distribution of topologies, Reid & Carstens 2012) and as such cannot account for heterogeneity among gene genealogies (Pons et al. 2006; Monaghan et al. 2009; Carstens et al. 2013; Fujisawa & Barraclough 2013). Further, the generalized mixed Yule coalescent model assumes the monophyly of species and thus is likely to fail to delimit recently divergent species that retain ancestral polymorphisms and have not completed lineage sorting. The broader utility of other methods for species delimitations in more species-rich systems remains unclear.

4.1.2. Biogeography of New Guinea, Wallacea, and the Sahul Shelf Region

Southeast Asia and Oceania, where the oriental biota of Sundaland meets the Australian biota of the Sahul region (New Guinea and Australia), and the intervening islands of Wallacea (the Banda Arc, Halmahera, and Sulawesi; Fig. 4.1), has long been an area of particular interest and importance in biogeography and evolutionary biology (Wallace 1860; Simpson 1977; Brown et al. 2013; Bacon et al. 2013). Indeed, the marked similarity between the fauna of the Aru Islands to that of New Guinea and Australia, combined with its dissimilarity to that of the geographically more proximate Kei Islands, played a key role in foundation of biogeography (Wallace 1857, 1860). This long history of biogeographic research has yielded numerous proposed lines to describe the biogeographic breaks in the biota of the region (Simpson 1977), a trend that began with the description of what is now known as Wallace’s line (Wallace 1860). Later work has resulted in modifications of this line (e.g. Huxley’s line, Huxley 1868) and other proposals for the boundary between these biotas, including Weber’s (Weber 1902) and Lydekker’s lines (Lydekker 1896), among others (Simpson 1977). While these described
boundaries represent important biogeographic breaks in the region, numerous taxa have distributions that span them (e.g. Clouse & Giribet 2007; Zug 2010; Linkem et al. 2013).

Within this region, the island of New Guinea is particularly biologically diverse and geologically complex. The island is home to a disproportionately large portion of the world’s biodiversity and has been identified as one of only five global high biodiversity wilderness areas (Mittermeier et al. 2003). However, New Guinea remains biologically poorly explored and houses numerous undiscovered species (Allison 1996, 2007; Austin et al. 2008). The complex geologic history of the island has likely played a critical role in the generation of this phenomenal diversity. While the southern portion of New Guinea formed as the northern edge of the Australian plate, much of the island is the result of the accretion of multiple island archipelagos, including the Outer Melanesian Island Arc, which was accreted some five to ten million years ago, and today forms much of northern New Guinea, as well as Halmahera and the Admiralty, Bismarck, and Solomon Archipelagos (Abbott et al. 1994; Hall 1997, 2002; Tregoning et al. 1999; Heads 2002; Polhemus 2007). Over its history, New Guinea has further experienced substantial fluctuations in climate that have altered its connectivity with Australia and driven habitat shifts (Bowler et al. 1976; Allison 1996, 2006; Voris 2000; Marshall &

Figure 4.1. Map of New Guinea and Southeast Asia showing the positions of Lydekker’s and Wallace’s lines; dark grey indicates currently subaerial land masses, light grey indicates shallow seas that were subaerial during periods of low sea levels. Reproduced from Jungers & Baab (2009) with permission from John Wiley and Sons Publishing.
Beehler 2007). Combined, these processes have resulted in an engine for diversification in New Guinea, yielding the phenomenal, yet underexplored, diversity of the region.

### 4.1.3. Systematics of Carla and the Complexity of the Carla fusca group

The distribution of skinks of the genus Carla spans the major biogeographic barriers of this region. The majority of the diversity in the genus, including six of the nine recognized species groups and over half the species, is restricted to Australia (Ingram & Covacevich 1989; Zug 2010). One species group, the C. peronnii group, includes three species that occur on Timor and other islands of Wallacea and a fourth recently described species restricted to a small offshore island near Java, on the opposite side of Wallace’s line relative to the rest of the genus (Zug 2010; Zug & Kaiser 2014). Of the two remaining species groups, the C. bicarinata group includes species in Australia and New Guinea, while the C. fusca group is most diverse in New Guinea, but also includes three Australian species and four species that occur in Wallacea, on the eastern side of Lydekker’s line (Zug 2004, 2010). Recent phylogenetic studies of the genus Carla have recovered a strongly supported sister relationship between these latter two species groups (Stuart-Fox et al. 2002; Dolman & Hugall 2008); however, sampling within these groups was sparse and neither study addressed the biogeographic implications of the results, such as the number of colonizations of New Guinea or of Wallacea, or if recolonization of Australia occurred.

As currently recognized, the C. fusca group consists of 18 species distributed largely parapatrically throughout the lowlands of New Guinea, as well as adjacent regions of northern Australia, the islands of the Admiralty and Bismarck Archipelagos, and parts of eastern Indonesia (Fig. 4.2; Zug 2004, 2010; Zug & Allison 2006; Donnellan et al. 2009). Although long recognized as a complex, species-rich group, the systematics of the group been comprehensively examined only recently (Zug 2004). In this monograph, which largely excludes the Australian members of the group, Zug described six new species and substantially clarified the complex taxonomic history of the group, including synonymizing several species and resurrecting a number other species. However, due to the low levels of morphological divergence among species in the group, species delimitation was difficult (G. Zug pers. comm.) and numerous populations could not be assigned to a species (Zug 2004). Two further species of C. fusca group skinks from western New Guinea were described later by Zug and Allison (2006) and Donnellan et al. (2009) subsequently revised the systematics of the Australian species in the group, including splitting C. longipes into three species based on morphological and molecular data. However, Austin et al. (2011) more recently used mitochondrial and nuclear sequence data to examine the origin of invasive C. fusca group populations in the islands of Palau, Guam, and the Northern Marianas. This study found extensive incongruence between the molecular data and currently recognized species, based largely on Zug’s morphological revision, including populations apparently assigned to the incorrect species, apparently synonymous species, and several cryptic species. Thus, while several studies have recently made extensive progress towards resolving the taxonomy of the group, extensive work remains necessary to clarify the systematics of the group.

Here, we use multi-locus sequence data to 1) assess the accuracy and utility of species delimitation methods in systems involving relatively large numbers of species, 2) test the monophyly of the C. fusca group, 3) examine the systematics and species boundaries within the
group, and 4) examine the biogeographic history of the group, specifically with reference to reference to the populations in Australia and in Indonesia, west of Lydekker’s line.

**4.2. Methods**

**4.2.1. Sampling and Molecular Data Collection**

We collected 208 tissue samples spanning much of the distribution of the *Carlia fusca* group, including 16 of the 18 currently recognized species (Fig. 4.3, Table D.1). We were unable to obtain samples for *C. babarensis*, from Babar and Tanimbar islands in Maluku Province, Indonesia, or from the nominate species, *C. fusca*, which was restricted to northwestern Indonesian New Guinea by Zug (2004). This ingroup sampling was combined with 56 outgroup samples, including representatives of all species in the putative sister species group (the *C. bicarinata* group; Dolman & Hugall, 2008; Stuart-Fox, Hugall, & Moritz, 2002; Zug, 2010), representatives of five of the seven remaining species groups within the genus *Carlia*, samples of two species in closely related genus *Lygisaurus*, of the closely related genus *Liburnascincus* (both of which were formerly included within *Carlia*; Dolman & Hugall, 2008; Stuart-Fox et al., 2002), and of the genus *Emoia* (Table D.1). We also obtained sequences for two species in

---

Figure 4.2. Distribution of currently recognized species in the *Carlia fusca* group, based on Zug (2004), Zug & Allison (2006), and Donnellan et al. (2009). Points represent collection localities of specimens used in these previous systematic studies on the species group. Black points indicate sampling localities not assigned to species by Zug (2004).
another Carlia species group, the *C. rhomboidalis* group, from GenBank; thus, our sampling includes representatives of eight of the nine species groups in Carlia, as well as samples representing all genera formerly included within this genus.

For all samples, whole genomic DNA was extracted using Qiagen DNeasy Blood and Tissue Kits (Valencia, CA, USA) as per manufacturer’s instructions, or using salt extractions (Fetzner 1999). Seven nuclear loci and one mitochondrial locus were then amplified via the polymerase chain reaction (PCR) as in Austin *et al.* (2010) using the primers and annealing temperatures in Table D.2. PCR amplicons were then sequenced in both directions via Sanger sequencing using the amplification primers by Beckman Coulter (Danvers, MA, USA). Resultant sequences were visually edited and complementary strands assembled in Geneious ver. 6.1.2. For nuclear sequences, heterozygous sites were identified by visual inspection of the chromatograms with the assistance of the heterozygotes plugin in Geneious.

MUSCLE ver. 3.8.31 (Edgar 2004) was then used to align sequences with a maximum of 1000 iterations. We then used a custom python script, phaser, to facilitate implementing PHASE ver. 2.1.1 (Stephens & Donnelly 2003) in order to identify nuclear alleles. PHASE analyses were run for 6000 iterations, sampling every fifth iteration, the first 200 samples of which were discarded as burnin. For any heterozygous sites that could not be phased with high posterior
probability (>0.95), standard IUPAC ambiguity codes were retained. We then used the corrected Akaike information criterion (cAIC) to select the best-fit model of sequence evolution for the mitochondrial locus, partitioned by codon position, and for each phased nuclear locus in jModelTest ver. 2.1.4 (Posada 2008).

4.2.2. Concatenated Phylogenetic Analyses

Mitochondrial and unphased nuclear sequences were concatenated, and phylogenetic relationships among samples were estimated using both Bayesian and maximum likelihood analyses, partitioned by mitochondrial codon position and locus. Bayesian analyses were conducted in MrBayes ver. 3.2.2. (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck 2003; Ronquist et al. 2012) with two replicate runs, each including four chains of 20 million iterations, sampling every 1000 iterations, with the first 25% of samples discarded as burn-in. For each partition, models of sequence evolution were estimated from general-time-reversible (GTR) model space simultaneously with estimation of the phylogeny (Huelsenbeck et al. 2004) with the model of among-site rate heterogeneity (i.e. no rate heterogeneity, a gamma distribution, or a proportion of invariant sites) set based on the best-fit model of sequence evolution as estimated in jModelTest. A compound Dirichlet prior was used on branch- and tree-lengths with an unconstrained gamma distribution on branch lengths ($\alpha = \beta = 1.0$) as Bayesian phylogenetic analyses have been shown to be prone to branch-length inflation under the default prior settings implemented in MrBayes (Marshall 2010; Brown et al. 2010), and a compound Dirichlet prior has been shown to be far more robust for estimating tree- and branch-lengths (Rannala et al. 2012; Zhang et al. 2012). Convergence of Bayesian analyses was assessed by inspecting the traces and effective sample sizes (ESSs) of all parameters in Tracer ver. 1.6, and by comparing the posterior probabilities of all splits between runs in AWTY (Nylander et al. 2008). Maximum likelihood analyses were implemented in GARLI ver. 2.01 (Zwickl 2006) using the best-fit models of sequence evolution from jModeltest; 48 search replicates were conducted to ensure the maximum likelihood solution had been found, and branch support was assessed via 1000 nonparametric bootstrap replicates.

4.2.3. Species Delimitation

Several powerful methods for species delimitation under the multi-species coalescent have been described (e.g. Bayesian species delimitation, Yang & Rannala 2010; Bayes factors delimitation, Grummer et al. 2014; maximum likelihood delimitations in spedeSTEM, Ence & Carstens 2011); however, these all represent species validation methods that require the a priori assignment of samples to putative species. We expected current taxonomy to be highly problematic in Carlia due to the low levels of morphological divergence (Zug 2004) and the previously documented incongruences between molecular and morphological data (Austin et al. 2011), and thus do not have reliable putative species a priori. We therefore focus our species delimitation analyses on two methods for species discovery that avoid this requirement: Gaussian clustering and genetic clustering under a Dirichlet process prior.

Species delimitation via Gaussian clustering involves calculating a single distance matrix for all samples, using non-metric multidimensional scaling (NMDS, such as the Kruskal method, Kruskal 1964), and finally implementing the Gaussian clustering to identify the number and composition of species (Hausdorf & Hennig 2010; Edwards & Knowles 2014). To calculate a single multilocus distance matrix, we used PAUP* ver. 4.1.0b (Swofford 2003) to calculate
maximum likelihood distances among samples for each locus and mitochondrial codon position using the best-fit model of sequence evolution as estimated in jModelTest. We used standardized distances in pofad ver. 1.03 (Joly & Bruneau 2006) to combine the distance matrices for each mitochondrial codon position into a single mitochondrial distance matrix, and then to combine this distance matrix with the seven phased nuclear distance matrices to obtain a single, multi-locus genetic distance matrix. The MASS package (Venables & Ripley 2002) in R ver. 3.0.2 was then used to implement Kruskal’s NMDS. To test the impact of dimensionality of the NMDS on species delimitation results, we used two different dimensionalities: three and six, which resulted in stress values of 13.75%, and 8.75%, respectively. The prabclus (Hausdorf 2012) and mclust (Fraley & Raftery 2006) packages in R were then used to implement Gaussian clustering. Previous work (see Chapter 3) has suggested that species delimitation via Gaussian clustering may more accurately delimit species without nearest-neighbor-based noise detection; however, others have suggested incorporating this into analyses (Hausdorf & Hennig 2010; Edwards & Knowles 2014). Therefore, we also investigate the impact of noise detection on species delimitations by implementing noise detection using a tuning constant of 6 (equal to the smallest integer greater or equal to the number of samples divided by 40, as suggested by Hausdorf & Hennig (2010)), a tuning constant of 1, and excluding noise detection. To compare these results based on the full dataset with results based on smaller, less species-rich clades, we also subdivided the data into nine major clades based on the concatenated phylogenetic analyses, and reanalyzed them under the same NMDS dimensionalities, but using only the noise detection tuning constant recommended by Hausdorf and Hennig (2010) or excluding noise detection.

Genetic clustering under a Dirichlet process prior was implemented in Structurama (Huelsenbeck & Andolfatto 2007; Huelsenbeck et al. 2011). Analyses were run for 1.5 million iterations, sampling every 100 iterations, the first 5000 samples of which were discarded as burn-in. Structurama requires allelic data, and assumes that loci are unlinked; therefore, we collapsed each locus into allele calls using a custom Python script, Seq2Struct. We conducted two analyses that differed in the prior distribution on the \( \alpha \) parameter that regulates the clustering of individuals in the Dirichlet process prior, and thus controls the prior distribution on the number of populations. In one analysis, \( \alpha \sim \text{G}(10,1) \), resulting in a prior mean number of populations of 30.41, and a variance of 20.59; and in the second analysis \( \alpha \sim \text{G}(10,2) \), resulting in a prior mean number of populations of 19.04, and a variance of 13.63. As in Gaussian clustering analyses, we then subdivided the dataset into nine clades based on the concatenated phylogenetic analyses to compare the results with those based on less diverse clades, and repeated the analyses, again sampling \( \alpha \) from a gamma distribution that resulted in a broad prior on the number of populations (\( \alpha \sim \text{G}(5,2) \)).

To further validate the species delimited using the above methods, we used Bayesian species delimitation (BSD, Yang & Rannala 2010; Rannala & Yang 2013) in bpp ver. 2.2. BSD analyses were conducted using species delimitation algorithm 0 with a fine-tuning parameter \( \epsilon = 15 \) following Leaché & Fujita (2010). For the guide tree, we used the species tree estimated via *BEAST (details below) using the species delimitations estimated using the subdivided Structurama analyses. BSD analyses were run with a 10,000 iteration burn-in, followed by a sampling period of 500,000 iterations, thinned to every 50 iterations. We used three different sets of prior distributions on population sizes (\( \theta \)) and divergence times (\( \tau \)), following Leaché & Fujita (2010): \( \theta \sim \text{Gamma}(1, 10) \), \( \tau \sim \text{Gamma}(1, 10) \), which sets large prior means on both the effective population sizes and divergence times, \( \theta \sim \text{Gamma}(2, 2000) \), \( \tau \sim \text{Gamma}(2, 2000) \), which sets small
prior means on both these parameters, and $\theta \sim \text{Gamma}(1, 10)$, $\tau \sim \text{Gamma}(2, 2000)$, which sets a large prior mean on the effective population sizes, and a small prior mean on the divergence times. BSD analyses including the complete dataset attempted to sample among over 1 million species delimitation models, which likely contributed to poor mixing, and, under some prior distributions, infinite likelihoods; therefore, we subdivided the analyses using strongly supported clades recovered in both the concatenated phylogenetic and species tree analyses, as in the species discovery methods above. In all BSD analyses, posterior probabilities greater than 0.95 were considered strong evidence for speciation.

4.2.4. Species Tree Inference

Species tree estimation under the multi-species coalescent was implemented in *BEAST ver. 2.1.0 (Heled & Drummond 2010; Bouckaert et al. 2013) to estimate the phylogeny of the $C. \text{fusca}$ group while accounting for heterogeneity among gene genealogies. Two analyses were run that differed in their species delimitation models: one with samples assigned to species based on the results of the within clade Structurama analyses to estimate a guide tree for BSD analyses, and a second with species delimitations set based on the results of the BSD analyses. The multi-species coalescent model assumes that heterogeneity among gene genealogies is due entirely to ancestral polymorphisms and incomplete lineage sorting (Edwards et al. 2007; Heled & Drummond 2010). Therefore, we excluded the sample from KarKar island from *BEAST analyses, as this sample showed evidence of hybridization or mitochondrial capture: in the mitochondrial gene tree and concatenated phylogenetic analyses, this sample grouped with $C. \text{mysi}$ from the Huon Peninsula (Fig. 4.4), but in within clade Structurama analyses, this sample clustered with the samples of $C. \text{pulla}$ from further west along the north coast of New Guinea (Fig. 4.5). Chains were run for 500 million iterations, sampling every 10,000 iterations, with the first 80% discarded as burnin. The species tree model was set to a Yule speciation model, and independent uncorrelated lognormal (UCLD) relaxed clocks were used for each locus and mitochondrial partition. Priors on the UCLD mean rates were set as diffuse gamma distributions with a shape of 0.001 and a scale of 1000. Models of sequence evolution for each nuclear locus and mitochondrial codon position were set based on the best-fit model as estimated in jModelTest. We assessed convergence by examining the effective sample sizes (ESS) and traces of all parameters in Tracer ver. 1.6.

4.2.5. Biogeographic Reconstructions

To examine the biogeographic history of the complex, and specifically to elucidate 1) the geographic origin of the $C. \text{fusca}$ group, 2) the number of colonization events between New Guinea and Australia, and 3) and the number of dispersals across Lydekker’s line (i.e. into Wallacea), we used two methods of biogeographic reconstruction: dispersal extinction cladogenesis (DEC, Ree et al. 2005; Ree & Smith 2008) and statistical dispersal vicariance analysis (S-DIVA, Yu et al. 2010). We defined five geographic areas for the analyses: Australia, New Guinea, Wallacea, Aru Islands, and Erub Island. DEC analyses were conducted using the maximum clade credibility tree from *BEAST species tree analyses in Lagrange ver. 2.0.1 (Ree et al. 2005; Ree & Smith 2008), with ancestral distributions including Wallacea restricted to two geographic areas, but no constraint on distributions not including Wallacea and no constraints on dispersal among regions. S-DIVA analyses were conducted in RASP ver. 2.1 (Yu et al. 2010, 2013) using the posterior sample of species trees from *BEAST, thinned to 1000 trees, to account for phylogenetic uncertainty, with ancestral species constrained as in the DEC analyses.
Figure 4.4. Maximum likelihood phylogeny of the *Carlia fusca* group from concatenated analysis. Numbers on branches refer to maximum likelihood bootstrap support Bayesian posterior probability, respectively; asterisks indicate bootstrap support of 100 or posterior probability of 1.0. Colors correspond to sampling localities in Fig. 4.3.
4.3. Results

4.3.1. Phylogenetic Analyses

The final aligned dataset consisted of 707 bp of mitochondrial sequence, and 3,648 bp of nuclear sequence, with nuclear loci ranging from 212 to 804 bp (Table D.2). In the concatenated phylogenetic analyses, we recovered strong support for the monophyly of the *C. fusca* group (maximum likelihood bootstrap, MLBS = 99; Bayesian posterior probability, PP = 1.0), and strong support for the *C. bicarinata* group (excluding the *C. schmeltzii* complex) as sister to the *C. fusca* group (Fig. 4.4; MLBS = 87, PP = 1.0). The *C. schmeltzii* complex (*C. schmeltzii* and *C. prava*) was recovered as sister to this *C. fusca* group + *C. bicarinata* group clade (MLBS = 86,
Deeper level relationships among the remaining Carlia species groups were generally poorly supported in one or both concatenated phylogenetic analyses; however, strong support was also recovered for the monophyly of the genera Carlia (MLBS = 93, PP = 1.0) and Lygisaurus (MLBS = 100, PP = 1.0).

Within the C. fusca group, we recovered nine strongly supported clades (MLBS > 98, PP = 1.0) that largely, with the exception of the Sahul Clade, correspond to major biogeographic provinces of New Guinea (e.g. the North Coast, the southern flanks of the mountainous central cordillera, the northern Milne Bay Islands, etc.; Fig. 4.4). Relationships among these clades, however, were unresolved (MLBS < 63, PP < 0.77), with one exception: the monophyly of the North, North Milne Bay, eothen, luctuosa, Rossel and South clades received moderate to strong support (MLBS = 84, PP = 1.0). Of these clades, three, the North Milne Bay, longipes, and Rossel Clades, are comprised of a single subclade, and a fourth, the Aru/Kei Clade, includes two subclades, strongly supported as reciprocally monophyletic (C. beccarii from the Kei Islands, and C. diguliensis from the Aru Islands; MLBS = 100, PP = 1.0). The remaining five major
clades with the *C. fusca* group, the North, *eothen*, *luctuosa*, South, and Sahul Clades, each include at least three strongly supported subclades, with the relationships among subclades largely well supported. However, these subclades show little concordance with the current taxonomy of the group. Multiple species (e.g. *C. mysi*, *C. eothen*, *C. aramia*) are recovered from multiple subclades, and even from multiple major clades, while we detect no apparent phylogenetic structure among *C. leucotaenia*, *C. ailenpalai*, *C. tutela*, and some populations of *C. mysi* (Fig. 4.4).

Figure 4.5. Species delimitation results from Structurama analyses. While separated into two rows for clarity and to facilitate comparing among analyses, the first and fourth rows and the second and fifth rows, labeled ‘Complete,’ are each based on analysis of the complete dataset, while the third and sixth rows, labeled ‘W/in,’ are from clade specific analyses. Colors correspond to identified clusters, and are only consistent within blocks for the within clade analyses (i.e. green in eothen Clade did not cluster with green in Aru/Kei Clade). Numbers correspond to population numbers in Table D.1.

Species tree estimations in *BEAST* largely corroborated the results of the concatenated phylogenetic analyses (Fig. 4.6). The monophyly of the *Carlia fusca* group was strongly supported (PP = 1.0), as was the monophyly of a clade comprised of the *C. fusca* group, the *C. schmetzii* complex, and the remaining members of the *C. bicarinata* group (PP = 1.0). However, the relationship among these three clades was unresolved (PP = 0.37). While the monophyly of the South Clade was only weakly supported (PP = 0.83), the remaining major clades were all
recovered with strong support (PP = 1.0). As in the concatenated analyses, the relationships among these clades was generally unsupported; however, as in concatenated analyses, we did recover strong support (PP = 1.0) for the monophyly of six of these major clades, with the exclusion of the *longipes*, Aru/Kei, and Sahul Clades.

### 4.3.2. Species Delimitation Analyses

Gaussian clustering analyses performed poorly, particularly based on the complete dataset (Table 4.1). Analyses were highly sensitive to the dimensionality of the NMDS, the inclusion of noise detection, and the tuning constant of noise detection (Table 4.1). Based on the complete dataset, Gaussian clustering identified between six and thirteen clusters, depending on the NMDS dimensions and the noise detection parameters; however, these results showed little concordance with the major clades of the concatenated phylogeny or Structurama analyses: clusters identified typically included samples from multiple major clades (mean 1.30 to 2.67 major clades per Gaussian cluster). Further, of 63 possible cases (i.e. six Gaussian clustering analyses x nine major clades), in only three cases did a cluster identified correspond perfectly with a major clade (all involving the *luctuosa* Clade with six NMDS dimensions), and in only two more cases were no samples of a major clade identified as belonging to a Gaussian cluster.
with samples from other major clades (both involving the Sahul clade with six NMDS dimensions and noise detection). Within clade analyses performed substantially better, however resultant clusters still showed limited concordance with other analyses: clusters identified typically included samples from multiple subclades in the concatenated phylogeny and identified as distinct clusters in the within clade Structurama analyses, and the clusters identified in the within clade Structurama analyses were frequently subdivided into multiple clades by Gaussian clustering (Table 4.1). Therefore, based on the sensitivity of Gaussian clustering to input parameter settings (i.e. NMDS dimensions and noise detection settings), we focus on Structurama results for subsequent species delimitation analyses.

Table 4.1. Results of Gaussian clustering analyses of complete dataset and restricted to within major clades, and a comparison with the assignments of samples from the same major clade or within clade Structurama analysis. GC refers to Gaussian clusters, SC refers to Structurama clusters, and MC refers to major phylogenetic clades.

<table>
<thead>
<tr>
<th></th>
<th>NMDS dimen.</th>
<th>Noise Tuning Constant</th>
<th>No. Clusters</th>
<th>Percent Unassigned</th>
<th>No. GCs per SC</th>
<th>No. SCs per GC</th>
<th>MCs per GC</th>
<th>GCs per MC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete Dataset</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>20.67%</td>
<td>1.214</td>
<td>7.286</td>
<td>2.429</td>
<td>2.429</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1</td>
<td>8</td>
<td>17.31%</td>
<td>1.262</td>
<td>6.625</td>
<td>2.250</td>
<td>2.571</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0</td>
<td>6</td>
<td>--</td>
<td>1.452</td>
<td>10.167</td>
<td>2.667</td>
<td>1.778</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>N/40</td>
<td>1-4</td>
<td>32.69%</td>
<td>0.952</td>
<td>2.167</td>
<td>(1-9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0</td>
<td>1-4</td>
<td>--</td>
<td>1.357</td>
<td>2.280</td>
<td>(1-9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>N/40</td>
<td>1-4</td>
<td>39.90%</td>
<td>0.714</td>
<td>2.143</td>
<td>(1-9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0</td>
<td>1-5</td>
<td>--</td>
<td>1.381</td>
<td>1.966</td>
<td>(1-9)</td>
<td></td>
</tr>
</tbody>
</table>

Despite the large prior means on numbers of populations, Structurama detected five ($\alpha \sim G(10,2)$), or eight ($\alpha \sim G(10,1)$) clusters, that largely, though not completely, corresponded to the major clades recovered in the phylogenetic analyses (Fig. 4.5). Within clade analyses resulted in a total of 41 clusters identified, ranging from one to nine clusters per major phylogenetic clade. While Structurama subdivided some of the smaller subclades from the concatenated phylogenetic analyses, in only case was sample recovered from divergent subclades assigned to
the same Structurama cluster. This case involved this single sample from KarKar Island, which was recovered in a clade with the *C. mysii* from the Huon Peninsula in the concatenated phylogeny, but was recovered with *C. pulla* from Madang and further west along the north coast of New Guinea in the Structurama analyses.

Bayesian species delimitations were largely concordant across tested prior distributions (Fig. 4.7); however, in each of two cases, the divergence of *C. bomberai* from *C. leucotaenia*, and the divergence of *C. sp. Kwatu* from *C. sp. Southern Highlands*, one prior (θ~Gamma(2, 2000), τ~Gamma(2, 2000) in each case) resulted in high posterior support (PP > 0.95) for speciation, while the other two priors supported lumping these species (PP < 0.90). Based on lumping the species in these two cases, and other cases in which species are lumped across all three priors, BSD analyses suggest the collapse of the 41 clusters identified via the within clade Structurama analyses to 28 species that largely reflect the major clades and subclades identified in the concatenated phylogenetic analyses (Fig. 4.7).

### 4.3.3. Biogeographic Reconstructions

DEC and S-DIVA analyses showed some differences in the reconstructed biogeographic history of the group (Fig. 4.8); however, the broad scale patterns estimated were largely concordant between the analyses. Both methods strongly support two independent dispersal events across Lydekker’s line and into Wallacea: dispersal from the Aru Islands to the Kei Islands in *C. beccarii*, and dispersal from New Guinea to Seram and Halmahera in *C. leucotaenia* (Fig. 4.8). DEC analyses suggest that the *C. fisca* group originated in Australia (probability = 0.57), though broader distributions spanning Australia and New Guinea (probability = 0.15) or Australia, New Guinea, and the Aru Islands (probability = 0.19) also received significant support. S-DIVA analyses, however, supported a broader distribution spanning Australia and New Guinea (probability = 0.79). The two methods also differ in the reconstruction of the biogeographic history of the two species currently assigned to *C. sexdentata*. S-DIVA suggests that the common ancestor of *C. fisca* group skinks exclusive of the Aru/Kei and *longipes* Clades was restricted to New Guinea (probability = 0.86), and thus suggests these species are the result of a recolonization of Australia by their common ancestor. DEC analyses support a broader distribution on the branches leading to *C. sexdentata* and the Sahul Clade, spanning both New Guinea and Australia, and thus do not clearly show the number of dispersal events between these regions.

### 4.4. Discussion

#### 4.4.1. The Influence of Phylogenetic Structure and Diversity on Species Delimitation

Species delimitation based on the complete datasets was difficult, but results were substantially improved by restricting analyses with major phylogenetic clades. While Gaussian clustering analyses generally performed poorly, results were substantially more congruent with other analyses when subdivided by clade (Table 4.1). Similarly, while Structurama analyses based on the complete dataset only identified a small number of clusters (five or eight) despite the large prior mean on number of populations, within clade analyses yielded clustering that was far more congruent with other analyses (Fig. 4.5). Finally, BSD analyses were unfeasible on the complete dataset: a single analysis would involve sampling among over one million species delimitation models, resulting in poor mixing and, under some prior distributions, infinite
likelihoods. However, when restricted to smaller, well-supported clades, these species validation analyses performed well. Most studies that investigated the accuracy of methods of species delimitation focused on a relatively small number of species (<~10; Hausdorf & Hennig 2010; Leaché & Fujita 2010; O’Meara 2010; Rittmeyer & Austin 2012; Grummer et al. 2014), and most empirical applications of these methods have similarly focused on groups with low diversity (Carstens et al. 2013). The few studies that have examined more species rich groups (more than approximately 12 species) have generally used the generalized mixed Yule coalescent model (Pons et al. 2006; Puillandre et al. 2009; Esselstyn et al. 2012; Carstens et al. 2013), which assumes monophyletic species and cannot account for heterogeneity among gene genealogies. The Carlia fusca group investigated here is much more diverse than the typical foci of species delimitation analyses. Therefore, it is unclear if the difficulties we experienced with the full dataset are peculiar to this particular dataset, such as the systematic complexity of group or the large number of species relative to the number of loci, or if these results suggest a broader
importance of deep phylogenetic structure and high species-level diversity in hindering species delimitation analyses.

4.4.2. Taxonomic Implications

While 18 species are currently recognized within the *C. fusca* group, our phylogenetic and species delimitation analyses identify 28 distinct species (Fig. 4.4-4.7), showing the diversity of the group has been underestimated. Further, our sampling in the western portion of New Guinea and in Wallacea was limited, and we were unable to obtain samples for two species, *C. fusca* and *C. babarensis*. These results also corroborate the findings of Austin et al. (2011) in showing pervasive incongruence between molecular data and species boundaries as currently recognized based largely on morphology. Samples of *C. mysi*, for example, were recovered in four different species and two different major clades (the North and Sahul Clades; Fig. 4.4). Similarly, samples of *C. aramia* were delimited to three distinct species in two major clades (the South and Sahul Clades; Fig. 4.4). *Carlia eothen* shows even more substantial discordance: samples currently assigned to this species were recovered from four major clades (South, *eothen*, North Milne Bay, and Rossel Clades, Fig. 4.4), and species delimitation analyses identified these samples as belonging to eight distinct species. Based on these results and the type localities for these species, we restrict *C. mysi* to the Huon Peninsula, and *C. eothen* to the Trobriand Islands.

Figure 4.8. Results of biogeographic analyses via DEC (A) and S-DIVA (B), pruned to the *C. fusca* group. Pie charts represent the reconstructed ancestral distributions; tips show extant distributions of each species. Regions abbreviated as: Au: Australia, NG: New Guinea, Ar: Aru Islands, Wa: Wallacea, Er: Erub Island.

66
We were unable to obtain topotypic tissues for *C. aramia*, and as such are unable to determine which, if any, of these clades represents this species; thus, we refrain from restricting this species until such time as further work can determine its taxonomic status.

This extensive discordance is not, however, limited to over lumping of species: our data also suggests several currently recognized species are synonymous. For example, *C. pulla* was restricted to the central portion of northern New Guinea by Zug (2004), yet we find this species to be much more broadly distributed throughout the remnants of the Outer Melanesian Island Arc that now form northern New Guinea, including the unassigned population from Wewak and some samples from the Madang area, just west of the Huon Peninsula. However, the grossest example of oversplitting is in *C. leucotaenia* in the Sahul Clade. Low levels of divergence and shared mitochondrial haplotypes and nuclear alleles were recovered among samples of *C. leucotaenia* from Seram, *C. tutela* from Halmahera, *C. bomberai* from the Vogelkopf region of western New Guinea, *C. ailanpalai* from the Admiralty Archipelago, and the populations assigned to *C. mysi* from the Bismarck Archipelago, and species delimitation analyses lumped these as a single species. We previously found this species to be the source for invasive populations of *Carlia* in the Pacific islands of Palau, Guam, and the Northern Marianas (Austin et al. 2011). Further, much of our sampling from Madang and all our sampling from Lae in northern New Guinea also belong to this species. Madang and Lae are both major port cities in Papua New Guinea, and samples collected nearby, or, in the case of Madang, syntopically, belong to distinct species in a different major clade; thus, these samples almost certainly represent recent, anthropogenic introductions, similar to the World War II era introductions of this species to the aforementioned Pacific islands. The conspecific status of *C. leucotaenia*, *C. tutela*, *C. ailanpalai*, and *C. bomberai* is not particularly surprising, as the latter three were all described in the recent morphological revisions of the group based on minor differences in body size and coloration (Zug 2004; Zug & Allison 2006). However, the broad distribution of this species, including populations from western New Guinea, Wallacea, and Northern Melanesian islands, is quite enigmatic. Further, the other members of the Sahul Clade are all restricted to Australia or south central New Guinea. Given the broad, unique distribution of *C. leucotaenia*, the distribution of closely related species, and the documented introduction of this species into multiple other areas, it is reasonable to question what the geographic origin for this species is, and whether the far-reaching populations of this species in the Admiralty and Bismarck Archipelagoes represent anthropogenic introductions or recent, but natural, waif dispersal. Denser sampling, particularly geographically, but also genomically, is necessary to better elucidate the history of this species, and resolve this question. Additionally, the nominate species for this group, *C. fusca*, occurs nearby in the northern Vogelkop region, the type locality is Waigeo Island, close to the northwestern tip of New Guinea, and is morphologically similar to this species (Zug 2004). Thus, it is plausible that this species is conspecific with *C. fusca* as well; however, until further morphological study tests this, or topotypic tissue samples of *C. fusca* are collected and analyzed, we prefer to retain *C. leucotaenia* for this biogeographically anomalous species.

### 4.4.3. Biogeography and Diversification in New Guinea

While we found some discordance between methods for reconstructing the biogeographic history of the group, these differences are likely due to the different approaches of the methods: DEC estimates the ranges of daughter branches following speciation events (Ree et al. 2005; Ree
& Smith 2008), while S-DIVA estimates ancestral distributions at nodes (Yu et al. 2010). Despite these differences, our biogeographic analyses suggest that the *C. fusca* group may have originated in Australia, subsequently colonized and diversified in New Guinea, and that *C. sexdentata* may represent a later colonization of Australia (Fig. 4.8). Our results also show evidence for two independent dispersal events across Lydekker’s line: colonization of Seram and the Northern Moluccan Islands by *C. leucotaenia* and colonization of the Kei Islands from the Aru Islands in the Aru/Kei Clade. We were unable to obtain samples for the fourth Wallacean species in the *C. fusca* group, *C. babarensis*, distributed in Tanimbar and Babar Islands. These islands are part of the Banda Arc islands, as are the Kei Islands, inhabited by *C. beccarii*; thus, the Kei Islands represent a plausible source population. However, further work is necessary to test whether *C. babarensis* reflects dispersal within Wallacea from the Kei Islands, or if this species represents a third dispersal across Lydekker’s line.

Within New Guinea, we find evidence for a complex mix of maintained divergence on fine geographic scales, possible hybridization, and low levels of genetic divergence over large geographic distances. In southern New Guinea, we recover the samples from Kiunga and Kwatu as highly divergent and in different major clades (the Sahul and South Clades, respectively), despite their close geographic proximity of only approximately 25 kilometers (Fig. 4.3). However, these populations are separated by the Fly River, the largest river by volume in New Guinea (Nilsson et al. 2005), suggesting that major rivers may play a role in driving and maintaining speciation in the group. Contrarily, we recover low levels of divergence among the populations of *C. pulla* from accreted portions of the Outer Melanesian Island Arc, including Vanimo, Wewak, and some samples from Madang, and delimit these populations as a single species (Fig. 4.3-4.5, 4.7). Yet another major river system, the Sepik River, separates Madang from these other populations. Recent work has shown the Sepik River is not an important barrier to dispersal in frogs in the region (Dahl et al. 2013), but the lack of divergence across this major river suggests a more limited role of rivers in driving diversification in this group. Alternatively, a recent shift in the course of this river could also explain its apparent lack of importance as a barrier to dispersal.

Our data also provides some evidence for hybridization in the group. In the concatenated phylogeny, we recover the sample from Karkar Island as sister to the samples of *C. mysi* from the Huon Peninsula to the southeast (Fig. 4.4); however, species delimitation analyses group this sample with the populations of *C. pulla* from Madang and further west (Fig. 4.5). The pattern in the concatenated phylogeny is likely being driven largely by the highly variable mitochondrial locus, while the Structurama result may reflect a broader pattern in the nuclear loci, as this method is less sensitive to a single, highly variable locus. These results suggest that both *C. pulla* and *C. mysi* colonized Karkar and subsequently hybridized, yielding the discordant results among analyses with respect to this population. Most species delimited in this group are apparently parapatrically distributed, with no obvious barriers to migration in many cases. Thus, detailed further sampling of the contact zones between species is necessary to fully elucidate the systematics in this group and the processes responsible for the generation and maintenance of this diversity.

Finally, we recover relatively low levels of divergence among geographically distantly distributed species in the South Clade, including species from southern Milne Bay Province and
Misima and Sudest islands of the Louisiade Archipelago in the east, species from the southern flanks of the central mountains, and *C. caesius* from southwestern New Guinea (Fig. 4.3, 4.4, 4.7), suggesting extensive, relatively recent dispersal throughout much of southern New Guinea.

These results contribute to a growing body of work on diversification in the terrestrial fauna of New Guinea that is increasingly showing the complexity of diversification in the region, the high frequency of cryptic species, and the overall vast underestimation of biodiversity. Oliver *et al.* (2013), for example, showed the broadly distributed microhylid frog species *Mantophryne lateralis* is comprised of at least nine distinct species, Kraus (2008) partitioned the widespread gecko *Cyrtodactylus louisiadensis* into five species, and Zug & Fisher (2012) found evidence for extensive unrecognized diversity within the gecko genus *Nactus*. These patterns of complex diversification and cryptic species in New Guinea are not restricted to herpetofauna, but have similarly been found in birds (Murphy *et al.* 2007; Benz 2011; Deiner *et al.* 2011), mammals (Malekian *et al.* 2010; Macqueen *et al.* 2011), and other taxa (De Bruyn *et al.* 2004; Craft *et al.* 2010). These previous results, combined with our data on the *C. fusca* group, highlight the evolutionary complexity of the New Guinea region, and are beginning to elucidate the roles of factors such as island accretions, habitat shifts, and montane uplift on diversification in this dynamic landscape.
CHAPTER 5
COMBINED NEXT-GENERATION SEQUENCING AND MORPHOLOGY REVEAL FINE-SCALE SPECIATION IN CROCODILE SKINKS (SQUAMATA: SCINCIDAE: TRIBOLONOTUS)

5.1. Introduction

With the rapidly expanding availability of next-generation sequencing (NGS) technologies capable of collecting genomic datasets of a scale previously restricted to model systems and the increasing availability of computation tools capable of analyzing the resultant datasets, these vastly superior new technologies have potential to revolutionize the studies of phylogenetics and population genetics (McCormack et al. 2012, 2013; Wagner et al. 2013). These massively multi-locus datasets can be leveraged to elucidate evolutionary patterns and processes with unprecedented power and precision, and thus to transform these fields. Several studies have shown that many evolutionarily important parameters, such as population size, divergence times, and migration rates, require large datasets of dozens or hundreds of loci, of a scale previously restricted to model systems (Carling & Brumfield 2007). The increasing application and availability of NGS technologies is already enabling researchers to disentangle the evolutionary histories of non-model systems with increasing accuracy and robustness, bettering our understanding of the processes responsible for the generation of biodiversity.

Despite the ability of NGS to generate genome-wide DNA sequence datasets, other types of data, such as morphological, ecological, or acoustic, remain critically important, particularly in studies endeavoring to clarify systematics and species limits in complexes of closely related species (O’Meara 2010; Yang & Rannala 2010; Rittmeyer & Austin 2012). Although species are a fundamental unit in biology, vital to a wide variety of disciplines and conservation efforts (Bickford et al. 2007), much contention remains over species concepts and how to identify and delimit species (Coyne & Orr 2004; de Queiroz 2005, 2007; Hausdorf 2011). This controversy is likely due largely to the complex and gradual nature of speciation and to differences in the processes driving speciation among systems (Coyne & Orr 2004). Dependent on the specific evolutionary forces acting on a system, morphological data can reveal extremely recent divergences that may be difficult to distinguish with molecular data (Lance et al. 2008; Rheindt et al. 2011; McCormack et al. 2012). Alternatively, molecular analyses frequently reveal previously overlooked lineages that are deeply divergent yet morphologically similar (Bickford et al. 2007; Arbogast & Kenagy 2008), although subsequent detailed morphological analyses often reveal subtle characters that can distinguish among these lineages identified using molecular data (Burbank 2001; Pyron & Burbank 2009). Identifying these characters facilitates field studies and conservation efforts through enabling researchers to identify museum vouchers and field specimens without laboratory analysis, thus enhancing the dissemination of biodiversity records to scientists and the general public.

Due to variation in the process of speciation, arguably the most generally applicable species concept is the general lineage species concept (but see Baum 2009; Hausdorf 2011; Naomi 2011), which defines species as “independently evolving meta-population lineages” (de Queiroz 2005; de Queiroz 2007). The general lineage species concept argues that rather than
defining species, other species concepts describe characteristics of species that evolve as these lineages diverge. For recently divergent lineages, multiple sources of data may be critical to the accurate delimitation of species. In more diverged systems, fewer types of data may be necessary. Regardless of the level of divergence, through combining genomic data collected via NGS with other datasets, the species limits, as well as the evolutionary patterns and processes responsible for their generation, can be elucidated more robustly than either dataset alone would accomplish. Here, we combine morphological data with a genomic scale dataset collected via NGS to examine the divergence between Crocodile Skinks (Squamata: Scincidae: Tribolonotus) on the islands of Buka and Bougainville.

The Crocodile Skinks of the genus Tribolonotus are a group of eight species distributed throughout northern New Guinea and the northern Melanesian islands of the Admiralty, Bismarck, and Solomon Archipelagos (Cogger 1972; McCoy 2006). The genus is united by the presence of two peculiar synapomorphies: abdominal glands, and palmar and plantar pores (Zweifel 1966; Greer & Parker 1968; McCoy 2006). Tribolonotus is also unusual among scincid lizards in having strongly keeled or spinose scales and, in at least two species (T. gracilis, T. ponceleti), the ability to vocalize (Hartdegen et al. 2001; McCoy 2006), an ability known from only one other scincid genus (Nannoscincus, Bauer et al. 2004). In a recent phylogenetic study of the genus, Austin et al. (2010) found that T. pseudoponceleti from the islands of Buka and Bougainville in the northwestern Solomon Archipelago are reciprocally monophyletic and deeply divergent (4.3% divergent at the mitochondrial cytochrome B and NADH dehydrogenase subunit 2 loci). This divergence is surprising given the geographic proximity and geologic history of the islands: Buka and Bougainville are currently separated by the Buka Passage, a narrow channel only approximately 300 meters wide. Further, the islands were historically connected multiple times in periods of lower sea levels during Pleistocene glaciations, including as recently as the last glacial maximum less than 20,000 years ago, forming a larger Greater Bougainville Island, along with the islands of Choiseul and Isabel further to the southeast in the Solomon Archipelago (Chappel & Shackleton 1986; Mayr & Diamond 2001). Here, we combine morphological data with genomic data collected via next-generation sequencing (NGS) to examine this divergence in more detail and 1) test if the Buka and Bougainville populations represent distinct species, 2) examine the demographic history and potential drivers of divergence, and 3) test for morphological divergence and diagnosibility of the two populations. While T. pseudoponceleti is also known from the northwestern tip of Choiseul Island, Solomon Archipelago, no genetic samples are available from this island; thus, we here focus exclusively on examining the divergence between the populations on the islands of Buka and Bougainville.

5.2. Methods

5.2.1. Morphological Data and Analyses

To test for morphological divergence between the Buka and Bougainville populations of Tribolonotus pseudoponceleti, we examined a total of 39 specimens from Buka, and 115 specimens from 11 populations spanning Bougainville (1-42 specimens per population, mean 10.5), including the holotype from Kunua, Bougainville (Fig. 5.1, Appendix F.1). For each specimen, eight mensural and eight meristic characters were scored. Mensural characters included first, third, and fifth finger lengths, first, fourth, and fifth toe lengths, forelimb length and hindlimb length. Meristic characters included first, third, and fifth finger subdigital lamellae,
Principal component analyses (PCA) were implemented on the mensural and meristic datasets in R ver. 2.15.1 to test for morphological divergences between Buka and Bougainville populations. Prior to PCAs, all mensural characters were scaled by snout-vent length (SVL) to correct for differences in body size among individuals. To further test for morphological
diagnosability, discriminant function analyses (DFA) were also implemented on the mensural and meristic datasets in R using the MASS package (Venables & Ripley 2002). As in PCAs, all mensural characters were scaled by SVL prior to DFAs.

5.2.2. Next-generation Sequencing

One of the biggest challenges with the application of NGS technology to population genetics studies is the need to reduce the genome such that numerous orthologous loci can be sequenced for many individuals (Hird et al. 2011; McCormack et al. 2012). This can be particularly challenging in organisms with large genomes, such as salamanders (McCormack et al. 2013; Gregory 2014). While the genome size of Tribolonotus is unknown, a closely related species, Tiliqua scincoides, has a genome of 1.780 Gb (De Smet 1981; MacCulloch et al. 1996; Gregory 2014) and other scincid lizards have genomes ranging from 1.027 Gb in Chalcides mionecton (De Smet 1981; Gregory 2014) to 3.130 Gb in Chalcides ocellatus (Capriglione et al. 1987; Gregory 2014). To accomplish genome reduction, we applied a double-digest approach (Vos et al. 1995; McCormack et al. 2012). Whole genomic DNA was extracted from liver taken from 12 individuals, including five individuals from Bougainville, five individuals from Buka (Fig. 5.1, Table F.1), and two individuals of the sister taxon, T. ponceleti, using a Qiagen DNeasy Blood and Tissue Kit (Qiagen, Inc., Valencia, CA, USA) as per manufacturer’s instructions. Library preparation protocols largely followed McCormack et al. (2012), but several modifications were made. First, digestion and ligation steps were separated: approximately 250 ng of DNA extracts were completely digested with EcoRI and MseI restriction enzymes (New England Biolabs, Ipswich, MA, USA) with 5 units of EcoRI and 1 unit of MseI in 1X T4 DNA Ligase buffer (New England Biolabs), 0.061 mg/mL BSA, and 0.056 M NaCl by incubation for 4 hours at 37ºC, followed by 20 minutes incubation at 65ºC to denature the restriction enzymes. AFLP adapters were then ligated to the DNA digests by adding 0.091 µM of each adapter (Table F.2) and 20 units of T4 DNA ligase (New England Biolabs) and incubating at 16ºC for 4 hours. Second, both amplification steps were conducted using Platinum PCR SuperMix High Fidelity (Invitrogen, Carlsbad, CA, USA) as per manufacturer’s instructions using the MseI-PreAmp primer and a biotinylated EcoRI-PreAmp primer (Table F.2). Third, due to the shorter read lengths of the Ion Torrent PGM, amplicons in the 100 to 160 bp range were excised from the gel, and extracted using a Qiagen Qiaquick gel extraction kit (Qiagen, Inc., Valencia, CA, USA) as per manufacturer’s instructions. Gel extracts were purified to isolate biotinylated fragments (i.e. those with EcoRI cut sites) using Dynabeads MyOne Strepavidin C1 (Invitrogen, Carlsbad, CA, USA) as per manufacturer’s instructions. Libraries for each sample were then purified using Agencourt AMPure XP beads (Agencourt Bioscience, Beverly, MA, USA) as per manufacturer’s instructions, quantified at two or more dilution points per sample, and six samples were pooled into each of two equimolar libraries. Emulsion PCR was then applied on each library to amplify library DNA onto Ion Particle Spheres using an Ion OneTouch Template Kit as per manufacturer’s instructions, enriched, and sequenced using 316 chips on an Ion Torrent PGM. All primer and adapter sequences are provided in the Table F.2.

5.2.3. Locus and Allele Calling

Sequence reads from each 316 chip were initially processed using the Ribosomal Database Project Pyrosequencing Initial Process (Cole et al. 2009) to sort reads by barcode and filter out reads < 50 bp or with a minimum quality score < Q20. The PRGmatic v1.6 pipeline (Hird et al. 2011) was then used to identify loci and call alleles. A minimum threshold of 10X
coverage was applied to call high confidence alleles and the PRGmatic default setting of 90% identity was used to call loci. Minimum coverage for calling consensus sequences in an individual was set to 10X, and the minimum coverage for calling a single nucleotide polymorphism (SNP) in an individual was 5X. Muscle ver. 3.8.31 (Edgar 2004) was then used to align the identified loci. As the threshold of 90% identity for calling loci may result in paralogous loci being combined as a single locus, we examined the all loci for multiple (>2 bp) SNP calls at a single site for an individual. Any loci where at least one individual had >2 bp (including gaps) at a single position in either >5% of the total reads for that locus, or in >3 reads total were discarded as potentially paralogous.

5.2.4. Genetic Structure Analyses
We examined the structure of the populations using genetic clustering algorithms implemented in Structure (Pritchard et al. 2000; Falush et al. 2003) and Structurama (Huelsenbeck & Andolfatto 2007; Huelsenbeck et al. 2011). While both programs apply the same algorithm to attempt to maximize Hardy-Weinberg equilibrium and minimize linkage disequilibrium by clustering the samples into populations, Structurama has the added benefit of implementing a Dirichlet process prior to simultaneously estimate the number of clusters and the assignments of individuals to clusters (Huelsenbeck & Andolfatto 2007). However, both Structure and Structurama assume that loci are unlinked allelic markers, and thus require the coding of each locus as alleles or the calling of a single SNP from each locus obtained from the NGS runs. We prefer the method of coding the alleles at each locus as integers because it results in less loss of data than calling a single SNP from each locus, although both approaches result in the loss of genetic distance information among alleles. We used a custom python script, seq2struct, to collapse each NGS locus to alleles coded as integers and to prepare the input files for both Structure and Structurama. This and all subsequent custom scripts are available via the Dryad repository, DOI doi:10.5061/dryad.87550. Three sets of analyses were run, all of which included all 12 individuals, but varied in their completeness and the number of loci: one including only those loci sequenced in all 12 individuals (i.e. no missing data), one including all loci sequenced in at least nine of the 12 individuals (i.e. including all 12 individuals and all loci included in the previous dataset, but allowing for some missing data to also include loci sequenced in nine, 10, or 11 individuals), and a third including all loci sequenced in at least six of the 12 individuals (i.e. including all 12 individuals and all loci included in the previous two datasets, but allowing for missing data to also include loci sequenced in six, seven, or eight individuals).

Structure analyses were implemented in Structure ver. 2.3.4 and consisted of 20 replicates at each value for K (the number of clusters) from one to ten. Each run consisted of a burn in of $1 \times 10^5$ iterations, followed by a sampling period of $1 \times 10^5$ iterations, sampling every 100 iterations, with correlated allele frequencies and all other parameters set to default values. To estimate the optimal number of clusters, the ΔK statistic (Evanno et al. 2005) was calculated using Structure Harvester (Earl & VonHoldt 2012), and the estimated ln probability of the data, ln Pr(D|K), was examined. Several additional, longer runs were also conducted with a burnin of $5 \times 10^5$ iterations and a sampling period of $1 \times 10^6$ iterations at a variety of values of K. These longer runs resulted in similar values of likelihood, ln Pr(D|K), and cluster assignments as the shorter runs, thus the full 20 replicates for each value of K were only conducted for the shorter runs.
To further examine population structure within *T. pseudoponceleti*, we used Structurama to implement the Dirichlet process prior in estimating the number of clusters and cluster membership. We tested a variety of prior distributions on the prior number of populations to test the sensitivity of the posterior number of populations to differing prior distribution. This prior distribution on the number of populations is indirectly set by the $\alpha$ parameter of the Dirichlet process prior, which controls the probability of two samples being assigned to the same cluster. In the first set of analyses, the $\alpha$ parameter was fixed to a value yielding an exponential distribution on the number of populations with a fixed prior mean value of one, two, three, four, or five. In a second set of analyses, the $\alpha$ parameter was sampled from one of three gamma distributions: G~(1.0,1.0), G~(2.5,0.5), and G~(0.5,2.5) selected based on a series of preliminary analyses and to provide a broad range of prior distributions on the number of populations. For all eight prior distributions on the number of populations, analyses were conducted for each of three datasets, as in the Structure analyses: loci sequenced from all twelve samples, loci sequenced in at least nine samples, and loci sequenced in at least six samples. Structurama analyses were run for $1.5 \times 10^6$ iterations, sampling every 100 iterations, with the first 5,000 samples discarded as burnin. All other priors were left at default values.

### 5.2.5. Species Delimitation using Bayes Factors

To test between the two competing species delimitation hypotheses (i.e. 2 species: *T. ponceleti*, Buka + Bougainville vs. 3 species: *ponceleti*, Buka, Bougainville) in a coalescent framework, we used Bayes factors to compare the species trees estimated under each of these models in SNAPP (Bryant *et al.* 2012; Leaché *et al.* 2013). SNAPP requires biallelic SNP data and assumes that all SNPs are unlinked (Bryant *et al.* 2012). We used a custom python script, biSNPcaller, to select the first biallelic SNP from each locus. We tested the impact of a variety of prior distributions on each parameter in a series of preliminary runs, selected based on previous studies (Leaché *et al.* 2013; Rheindt *et al.* 2014), and to provide a range of possible prior distributions. Mutation rate parameters were sampled from a Gamma (2, 2), Gamma (2, 10), or Gamma (2, 100) distribution. An improper infinite uniform (0, $\infty$), and a uniform (0, 1e5) distribution on the Yule speciation rate, and a Gamma (2, 100), Gamma (2, 1000), Gamma (2, 10000) on theta were also tested. Prior distributions on the mutation rate and Yule speciation rate parameters had no impact on the resultant posterior probability, likelihood, or parameter estimates (results not shown), so default settings we retained for each of these priors. Selection of prior distribution on theta, however, substantially impacted the results: smaller prior means resulted in smaller estimated population sizes and shallower divergences. Therefore, full analyses were run under each of the three aforementioned priors on theta. Similar to the Structurama analyses, SNAPP analyses were conducted on three datasets: one including only those loci sequenced in all 12 samples, one including all loci sequenced in at least nine of the 12 samples, and a third including all loci sequenced in at least six of the 12 samples. Marginal likelihoods were estimated via path sampling (Lartillot & Philippe 2006) in BEAST ver. 2.1.0 (Bouckaert *et al.* 2013) with the SNAPP ver. 1.1.4 (Bryant *et al.* 2012) and BEASTii ver. 1.1.0 add-ons, which implements the proportionality constant correction described by Leaché *et al.* (2013) to make marginal likelihoods estimated under different species delimitation models comparable. Each path sampling analysis included 48 steps, each consisting of $1 \times 10^5$ generations, the first ten percent of which were discarded as burn-in, which was sufficient to obtain suitable effective sample sizes (>200). Bayes factors were then calculated as $2 \times \ln BF$, where BF is the ratio of the marginal likelihoods for the two competing models (i.e. twice the
difference in the log marginal likelihood), and evaluated following the framework of Kass and Raftery (1995). We subsequently estimated the species trees and relevant evolutionary parameters (effective population size, divergence times) in SNAPP (i.e. without path sampling) for each of the datasets and priors under the preferred species delimitation model. Analyses were run for $2 \times 10^6$ iterations, sampling every 1,000 iterations, the first ten percent of which were discarded as burnin.

5.2.6. Demographic Model Selection

The fitting of demographic models, such as isolation with migration (IM) class models, represents a powerful method for comparing among evolutionary hypotheses and thus for testing for the importance of various evolutionary parameters, such as migration or population size changes (Hey & Nielsen 2007; Carstens et al. 2009; Hey 2010; Sousa & Hey 2013). The size of datasets obtained via NGS makes full Bayesian implementation of demographic models computationally unfeasible and the short read lengths obtained by some methods precludes gene genealogy based analyses. However, the large number of loci obtained also enable comparisons among more complex and potentially more realistic models using alternative methods that are more computationally efficient, such as diffusion approximation of allele frequency spectra (Gutenkunst et al. 2009) or approximate Bayesian computation (Beaumont et al. 2002; Csilléry et al. 2010).

During Pleistocene glacial cycles, the islands of Buka and Bougainville repeatedly enlarged and merged during glaciations and shrunk and separated during interglacials (Chappel & Shackleton 1986; Shackleton 1987; Chappell et al. 1996). As the mitochondrial divergence suggests a more ancient divergence between the Buka and Bougainville populations than the last glacial maximum (LGM), we hypothesize that these populations may have diverged during an early interglacial, with some migration potentially occurring between the populations during later glacial cycles. Due to the more recent decrease in island area and isolation of the islands since the LGM, we further hypothesize that these populations may have experienced a shift in population size and migration rate when glacial retreat drove sea level rise, isolating the two islands. We test among nine models of the demographic history for these populations, varying from the simplest model of no divergence between Buka and Bougainville, to the most complex model, including a shift in the population sizes of, and migration rates between, Buka and Bougainville subsequent to their divergence (Fig. 5.2, Fig. F.1). To compare among these models, we use two approaches: diffusion approximation of the allele frequency spectrum (AFS), and approximate Bayesian computation (ABC).

Diffusion approximation of the AFS was implemented in $\texttt{ca}\hat{a}$ i v. 1.6.3 (Gutenkunst et al. 2009). $\texttt{ca}\hat{a}$ i requires biallelic SNP data, thus, as for the SNAPP analyses, we used a custom python script, biSNPcaller, to identify the first biallelic SNP in each locus. Five search replicates were conducted for each model, and compared using the Akaike information criterion (AIC). We then estimated confidence intervals from 100 nonparametric bootstrap replicates of the best-fit model. All parameters are, by definition, positive; therefore, we used log-transformed parameters to estimate confidence intervals (i.e. $e^{(\ln(\theta^*) \pm 1.96*\ln(\sigma))}$, where $\theta^*$ is the maximum likelihood estimate, and $\sigma$ is the standard deviation of the bootstrap replicates), following Gutenkunst et al. (2009).
ABC analyses consist of simulating a number of datasets under the models of interest, and calculating a suite of summary statistics for each simulated dataset (Beaumont et al. 2002; Csilléry et al. 2010). A set of simulated datasets with summary statistics most similar to the observed summary statistics are then accepted as an approximation of the Bayesian posterior distribution, and can be used to assess model support and parameter values. To test among the nine demographic models using ABC, we used a custom python script to simulate two million datasets for each model (18 million total) and calculate a suite of summary statistics for each simulated dataset using msABC (Pavlidis et al. 2010). The number of loci and the lengths of

Figure 5.2. Schematic of the most complex demographic model (model 9) assessed in ∂a∂i and ABC analyses. Note that population sizes following the shift at time T3 are not constrained to be smaller than population sizes prior to the shift; similarly, migration rates were allowed to increase or decrease at the shift. For details of additional demographic models tested, see Fig. F.1.
generating similar summary statistics, we applied PCA in R to datasets with comparable summary statistics to the empirical values being tested, as well as on the selection of appropriate prior distributions that result in simulated statistics used with the ABC package in R. For each parameter, the geometric mean was estimated in R using the psych package. Multinomial logistic regression was implemented using the abc package (Csilléry et al. 2012) in R to perform the model selection step with a tolerance of $5.55 \times 10^{-5}$ to retain 1,000 simulated datasets. We then increased the number of simulations for the best-fit model to $5 \times 10^6$ total, and applied local linear regression using the abc package in R to estimate parameters using a tolerance of $2 \times 10^{-4}$ to retain 1,000 simulated datasets. For each parameter, the geometric mean was estimated in R using the psych package (Revelle 2012), and the 95% HPD intervals were calculated in R using the coda package (Plummer et al. 2005).

The accuracy of ABC analyses is highly dependent on the correlation of the summary statistics used with the parameters of interest and their utility in discerning among the models being tested, as well as on the selection of appropriate prior distributions that result in simulated datasets with comparable summary statistics to the empirical values (Beaumont et al. 2002; Csilléry et al. 2010). To assess the appropriateness of the selected prior distributions in generating similar summary statistics, we applied PCA in R and plotted the first two principal
components to visualize whether the empirical values fell within the cloud of values from the simulated datasets. To validate the accuracy of ABC model selection and parameter estimation, 100 simulated datasets were selected at random for each of the nine models and used as pseudo-observed data for model selection and parameter estimation rejection steps.

5.3. Results

5.3.1. Morphological Analyses

Principal components analysis of the mensural data reveals a trend towards differentiation between Buka and Bougainville, though the clusters do overlap broadly (Fig. 5.3A). Specimens from Buka exhibit slightly lower values for the first principal component (82.1% of variation), which corresponds to shorter limbs relative to body length, but the Buka and Bougainville populations show no substantial differentiation on any other principal components axes. Discriminant function analyses of the mensural data further substantiate this trend of differentiation of the Buka and Bougainville populations based on mensural data: 105 of 114 specimens (92.11%), examined from Bougainville are correctly classified and 37 of 39 specimens (94.87%) examined from Buka are correctly classified based on the mensural data. Meristic data reveal far greater differentiation and diagnosibility of the Buka and Bougainville populations. PCA of meristic data (Fig. 5.3B) shows the populations are primarily differentiated along the first principal component axis (68.9% of variation), with specimens from Bougainville exhibiting larger values, that corresponds both to a larger number of ventral scale rows, and a larger number of subdigital lamellae. DFA based on the meristic data also reveals the diagnosibility of the Buka and Bougainville populations: 111 of 114 specimens from Bougainville are correctly classified (97.37%), and 37 of 39 specimens from Buka are correctly classified (94.87%).

![Figure 5.3](image)

Figure 5.3. Scatterplots of the first two principal components of the morphological data based on A. mensural characters, and B. meristic characters. Dark grey squares indicate Buka specimens, light grey circles indicate Bougainville specimens. The holotype is shown as a light grey triangle.

5.3.2. Next-generation Sequencing

The two Ion Torrent PGM 316 chips resulted in a combined 6.45 million reads, or 648.13 MB of sequence data, of which 486.28 MB had an estimated quality score of Q20 or above. After parsing reads by barcoded sample and filtering out low quality reads and those with
ambiguos bases or errors in the barcode or forward primer sequences, the NGS experiments resulted in an average of 255,748 reads per individual (range 194,980 to 314,946). A total of 6,983 loci were identified using the PRGmatic pipeline, of which 1,526 loci were sequenced from all 12 individuals (mean coverage 62.27 ± 49.03 reads per locus per individual), 2,608 were sequenced in at least nine individuals (mean coverage 50.09 ± 46.61 reads per locus per individual), and 3,342 were sequenced in at least six individuals (mean coverage 43.21 ± 44.66 reads per locus per individual). The datasets including loci sequenced in at least nine individuals were 93.57% complete, whereas those with loci sequenced in at least six individuals were 85.80% complete. Of the 1,526 loci sequenced in all 12 individuals, 1002 (65.66%) were variable; 1,703 loci of the 2,608 sequenced in at least nine individuals (64.99%) were variable; and of 3,342 loci sequenced in at least six individuals, 2,135 (63.61%) were variable. Among the loci recovered in all 12 individuals, 941 included at least one biallelic SNP, and were thus useful for SNAPP and δαçı analyses; 1,590 loci recovered in at least nine individuals had at least one biallelic SNP, and 1,973 loci recovered in at least six individuals had biallelic SNPs. Among the loci recovered in at least six individuals, the mean number of segregating sites was 0.994 (range: 0 – 9).

5.3.3. Genetic clustering analyses

With the complete dataset of loci sequenced for all 12 individuals, the ln Pr(D|K) from Structure peaked at K=3 (Fig. F.3). ΔK peaked at K=2 (ΔK=582.54), with a slightly lower peak at K=3 (ΔK=487.60); for all other tested values of K, ΔK was less than 3.27. With K=2, the two clusters corresponded to *T. ponceleti* and *T. pseudoponceleti* with all samples assigned to the appropriate cluster with probability 1.0 (Fig. 5.4A). With K=3, the three clusters corresponded to Bougainville, Buka, and *T. ponceleti*, with 11 of the 12 samples assigned to the appropriate cluster with probability of 0.999 or greater (Fig. 5.4B). The twelfth sample, from Buka, had an admixed assignment, with probability 0.907 to the appropriate cluster, and probability 0.093 to the Bougainville cluster. Structure results were similar when including all loci sequenced for at least nine individuals: ln Pr(D|K) peaked at K=3, whereas ΔK peaked at K=2 (ΔK=474.54), with a slightly lower peak at K=3 (ΔK=321.50); for all other tested values of K, ΔK was less than 6.72. Population assignment probabilities were identical to those obtained with the dataset of loci sequenced for all 12 individuals. With the dataset of loci sequenced in at least six individuals, both the ln Pr(D|K) and ΔK peaked at K=3 (ΔK=420.93), with a slightly lower peak for ΔK at K=2 (ΔK=321.65); for all other tested values of K, ΔK was less than 3.57. As with the other, more complete datasets with fewer loci, the three clusters corresponded to the Buka, Bougainville and *T. ponceleti* populations; 11 of the 12 individuals were assigned to the appropriate cluster with assignment probability greater than 0.999, with the final sample, an individual from Buka, assigned to the appropriate cluster with probability 0.900, and with probability 0.100 to the Bougainville cluster.

When examining genetic clustering with the implementation of the Dirichlet process prior to estimate K in Structurama, the posterior probability was 1.00 for K=3, and 0.00 for all other values of K, regardless of the dataset used (i.e. loci sequenced in all 12 individuals, those sequenced in at least nine individuals, or those sequenced in at least six individuals), or the selection of prior distribution on the α parameter of the Dirichlet process prior. As with the Structure analyses, these three clusters corresponded to *T. ponceleti*, and the Buka and Bougainville populations, with all samples assigned appropriately.
5.3.4. Species Delimitation using Bayes Factors

Divergence times and effective population sizes were sensitive to the selection of prior on theta: the prior distributions with smaller means resulted in smaller estimated population sizes and shallower divergences (Fig. 5.5, Table 5.1, Table F.3). However, other parameters, likelihoods, and posterior probabilities were not sensitive to prior selection. Similarly, selection of the prior distribution on theta had a negligible impact on marginal likelihood estimates. Regardless of the dataset used (i.e. complete dataset, loci recovered from 9+ individuals, or loci recovered from 6+ individuals) the Bayes factors strongly favored the three species model (i.e. *ponceleti*, Buka, Bougainville; Table 5.2). This pattern was strongest in the largest dataset (i.e. loci recovered from 6+ individuals); however, for the smallest dataset (including only loci recovered from all 12 individuals), the Bayes factors were over 1480 in favor of the three species model (Table 5.2). Bayes factors in excess of 10 are typically considered decisive support for a model over the alternative, thus these results strongly support the treating of Buka and Bougainville as distinct species. Estimated population sizes and divergence times from the full analyses without path sampling are presented in Figure 5.5, Table 5.1, and Table F.4.

5.3.5. Demographic Model Selection

$c$ analyses of the AFS strongly support model 9, including population size changes and shift in migration rates subsequent to the divergence between the Buka and Bougainville
populations, as the best fit to the data, and encompassing 0.985 of the AIC weight (Table 5.3).

Model 5, including a population size change and a stop of migration subsequent to the divergence between the Buka and Bougainville populations, was the next best fit model; however, this model was a substantially worse fit to the data, including only 0.011 of the AIC weight, and with an evidence ratio of 88.871 relative to model 9. All remaining models had AIC weights of below 0.0016, and evidence ratios in excess of 645, suggesting the other tested models were a far worse fit to the data. Maximum likelihood parameter estimates suggest that migration rates were moderate between the Buka and Bougainville populations immediately after divergence (3.943 from Buka to Bougainville, 1.180 from Bougainville to Buka, in units of $2N_0m$, where $N_0$ is the ancestral effective population size, and $m$ is the proportion of each population made up of migrants each generation), with a substantial decrease in migration subsequently (0.251 and 0.613, respectively; Table 5.4). Confidence intervals, however, were extremely broad, particularly for migration rates prior to the shift (Table 5.4, Fig. 5.5).

The plot of the first two principal components of the summary statistics reveal that the observed summary statistics fall well within the cloud of summary statistics generated by the simulated data (Fig. F.2), indicating that the selected prior distributions are appropriate for producing summary statistic vectors comparable to those observed in the empirical data. Pseudo-rejection analyses using simulated summary statistics suggest that, with the exception of model 1 (no divergence between Buka and Bougainville), the summary statistics were generally
insufficient to distinguish among models (Table F.5). Model 1 was correctly identified with the highest posterior probability in 93% of cases, and in 76% of the replicates, model 1 was correctly identified with moderate support (posterior probability > 0.75). Among the remaining models, the correct model had the highest posterior probability in only 37.4% of replicates, and in only six of the 800 total replicates did the correct model receive moderate support. Replicates in which the model with the maximum posterior probability was incorrect typically involved difficulty in distinguishing between the models with no migration (i.e. model 2 or model 3), or among models involving migration (i.e. models 4-9), although models in which migration stopped subsequent to the divergence (model 5 or model 7) were difficult to distinguish from

Figure 5.5. A. Posterior probabilities of demographic models estimated via ABC. Dark grey indicates model 7, light grey indicates model 9. All other models had posterior probabilities <0.0001. B. Estimated population sizes, relative to the ancestral population size. C. Estimated timing of demographic events in coalescent units: T1 is the divergence of ponceleti and Buka+Bougainville, T2 is the divergence of Buka and Bougainville, and T3 is the timing of the shift in population size and migration rates. D. Estimated migration rates, in units of $4N_0m_{ij}$, from Bougainville to Buka (m12) and Buka to Bougainville (m21), before the shift in migration rates (‘a’) and, for $\partial a\partial i$ analyses, following the shift (‘b’). E. Key to the symbols used in B-D. For sections B-D, error bars indicate the limits of the 95% confidence intervals ($\partial a\partial i$ analyses) or the 95% HPD intervals (ABC, SNAPP analyses).
either isolation models or isolation with migration models (Table 5.4). However, when the model was not identified correctly, all models generally received poor support, and in only a single replicate did an incorrect model have posterior support greater than 0.75, thus suggesting that when the posterior probability strongly supported a single model, this was typically the correct model. Parameter estimation using pseudo-observed data performed better, with 97.8% of the replicates including the true value within the 95% HPD intervals, and 68.1% of replicates including the true value within a single standard deviation of the geometric mean (Table F.5). However, 95% HPD intervals for most parameters included the majority of the prior distribution, though with smaller variance.

Table 5.3. Statistics of demographic model selection via Akaike information criterion from δabc analyses of the AFS.

<table>
<thead>
<tr>
<th>Model</th>
<th>K</th>
<th>lnL</th>
<th>AIC</th>
<th>ΔAIC</th>
<th>$e^{(-ΔAIC/2)}$</th>
<th>$w_i$</th>
<th>$E_{min,i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 9</td>
<td>13</td>
<td>-449.368</td>
<td>924.736</td>
<td>0.000</td>
<td>1.000</td>
<td>0.985</td>
<td>1.000</td>
</tr>
<tr>
<td>Model 7</td>
<td>11</td>
<td>-455.855</td>
<td>933.711</td>
<td>8.974</td>
<td>0.011</td>
<td>0.011</td>
<td>88.871</td>
</tr>
<tr>
<td>Model 5</td>
<td>9</td>
<td>-459.840</td>
<td>937.679</td>
<td>12.943</td>
<td>1.547E-03</td>
<td>1.524E-03</td>
<td>646.433</td>
</tr>
<tr>
<td>Model 3</td>
<td>8</td>
<td>-460.933</td>
<td>937.866</td>
<td>13.130</td>
<td>1.409E-03</td>
<td>1.388E-03</td>
<td>709.695</td>
</tr>
<tr>
<td>Model 6</td>
<td>11</td>
<td>-459.634</td>
<td>941.269</td>
<td>16.533</td>
<td>2.570E-04</td>
<td>2.532E-04</td>
<td>3890.720</td>
</tr>
<tr>
<td>Model 10</td>
<td>9</td>
<td>-480.319</td>
<td>972.637</td>
<td>47.901</td>
<td>3.967E-11</td>
<td>3.908E-11</td>
<td>&gt;1E15</td>
</tr>
<tr>
<td>Model 2</td>
<td>6</td>
<td>-480.038</td>
<td>978.076</td>
<td>53.340</td>
<td>2.615E-12</td>
<td>2.576E-12</td>
<td>3.825E11</td>
</tr>
<tr>
<td>Model 1</td>
<td>3</td>
<td>-1320.474</td>
<td>2646.947</td>
<td>1722.211</td>
<td>&lt; 1E-15</td>
<td>&lt; 1E-15</td>
<td>&gt;1E15</td>
</tr>
</tbody>
</table>

ABC analyses of the empirical data resulted in a high posterior probability for model 7 (PP = 0.9043), which includes a shift in population size and cessation of migration subsequent to the divergence of Buka and Bougainville. Model 9, involving a shift in population size and in migration rate subsequent to the divergence of Buka and Bougainville, also received some limited posterior support (PP = 0.0957); the remaining models received little support, with posterior probabilities less than 0.0001. Similar to the pseudo-observed data, parameter estimation resulted in broad 95% HPD intervals, and posterior distributions similar in width to, but more tightly clustered than, then prior distributions (Fig. 5.5, D.4).

5.4. Discussion

Although species are the most basic unit of biological classification and as such are fundamental to all biological research, there remains much contention over species concepts and what criteria should be used in identifying species. This controversy likely stems at least in part from the variability and complexity in the underlying processes that lead to speciation (Coyne & Orr 2004). The general lineage species concept acknowledges this variation and leverages it to a more general definition of species as independently evolving metapopulation lineages, allowing researchers to apply the specific criterion or criteria most applicable to their system. While in many systems, species boundaries may be robustly inferred with less extensive data and analyses than we present, through combining multiple datasets and analyses, we are able to validate the speciation of the Buka and Bougainville populations of Tribolonotus pseudoponceleti under a variety of criteria, and thus provide a far more comprehensive assessment of the status of these
Table 5.4. Demographic parameter estimates and 95% HPD intervals (ABC) or 95% confidence intervals (\(\delta a \delta i\)) from ABC and \(\delta a \delta i\) analyses. Asterisks (*) indicate population sizes relative to the ancestral population size. See Figure 2 for details of parameter values.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>ABC</th>
<th>(\delta a \delta i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(T_1)</td>
<td>1.4207</td>
<td>1.3475</td>
</tr>
<tr>
<td></td>
<td>(0.5497-2.4177)</td>
<td>(0.7498-2.4215)</td>
</tr>
<tr>
<td></td>
<td>0.7312</td>
<td>0.9143</td>
</tr>
<tr>
<td>(T_2)</td>
<td>0.1072-2.1832</td>
<td>(0.5743-1.4556)</td>
</tr>
<tr>
<td></td>
<td>0.2919</td>
<td>0.7445</td>
</tr>
<tr>
<td>(T_3)</td>
<td>0.0405-0.6988</td>
<td>(0.2424-2.2871)</td>
</tr>
<tr>
<td></td>
<td>0.0017696</td>
<td>0.9551*</td>
</tr>
<tr>
<td></td>
<td>(0.0002429-0.0058747)</td>
<td>(0.0467-3.5791)</td>
</tr>
<tr>
<td></td>
<td>0.0034115</td>
<td>1.8413*</td>
</tr>
<tr>
<td></td>
<td>(0.000196-0.0112576)</td>
<td>(0.0268-8.5474)</td>
</tr>
<tr>
<td></td>
<td>0.0016104</td>
<td>0.8692*</td>
</tr>
<tr>
<td></td>
<td>(0.000034-0.0112666)</td>
<td>(0.1236-2.9613)</td>
</tr>
<tr>
<td></td>
<td>0.0031761</td>
<td>1.7142*</td>
</tr>
<tr>
<td>(\theta_{1b})</td>
<td>(0.0000071-0.0045353)</td>
<td>(0.0129-9.591)</td>
</tr>
<tr>
<td></td>
<td>0.0015089</td>
<td>0.8144*</td>
</tr>
<tr>
<td>(\theta_{1a})</td>
<td>(0.0000071-0.0045353)</td>
<td>(0.0058-2.3666)</td>
</tr>
<tr>
<td></td>
<td>0.0039872</td>
<td>2.152*</td>
</tr>
<tr>
<td>(\theta_{2b})</td>
<td>(0.0000071-0.0045353)</td>
<td>(0.0064-8.3025)</td>
</tr>
<tr>
<td></td>
<td>0.0018528</td>
<td>0.6126</td>
</tr>
<tr>
<td>(\theta_{2a})</td>
<td>(0.0000071-0.0045353)</td>
<td>(0.0058-2.3666)</td>
</tr>
<tr>
<td></td>
<td>0.0015089</td>
<td>0.6126</td>
</tr>
<tr>
<td>(\theta_3)</td>
<td>(0.0000071-0.0045353)</td>
<td>(0.0058-2.3666)</td>
</tr>
<tr>
<td></td>
<td>0.004595-0.0113485</td>
<td>(0.0064-8.3025)</td>
</tr>
<tr>
<td>(\theta_{12})</td>
<td>(0.0005345-0.0039068)</td>
<td>(0.0058-2.3666)</td>
</tr>
<tr>
<td>(m_{12b})</td>
<td>0</td>
<td>(0.0869-4.3169)</td>
</tr>
<tr>
<td>(m_{12a})</td>
<td>11.4474</td>
<td>1.1797</td>
</tr>
<tr>
<td>(m_{21b})</td>
<td>(1.5165-31.4366)</td>
<td>(0.0816-17.0565)</td>
</tr>
<tr>
<td>(m_{21a})</td>
<td>0</td>
<td>0.2512</td>
</tr>
<tr>
<td></td>
<td>(0.4857-29.6822)</td>
<td>(0.0155-4.0678)</td>
</tr>
</tbody>
</table>

populations. Our previous analyses (Austin et al. 2010) found the populations from Buka and Bougainville to be reciprocally monophyletic on the basis of concatenated analysis of 2,252 bp of mitochondrial and nuclear sequence data (2 mitochondrial and 3 nuclear gene regions). The morphological data we present here indicate the presence of diagnostic morphological characters distinguishing the two populations (body size, number of ventral scale rows, number of subdigital lamellae). DFA were accurate in distinguishing the Buka and Bougainville populations.
on the basis of the meristic data, and, to a lesser extent, on the basis of mensural data, further supporting the morphological diagnosability of the populations. PCAs based on meristic data also show the two populations generally form distinct clusters; again to a lesser extent, the populations largely cluster in principal components space on the basis of mensural data. Genetic clustering analyses further support the distinctness of the Buka and Bougainville populations. While ΔK peaks at K=2 for the smaller (but more complete) datasets, with the two clusters corresponding to *T. ponceleti* and *T. pseudoponceleti*, both show a slightly lower secondary peak at K=3, with the clusters corresponding to *T. ponceleti*, Buka, and Bougainville. In cases of hierarchical population structure, ΔK detects higher levels of clustering first (Evanno *et al.* 2005), thus this likely reflects the hierarchical nature of the data (i.e. the Buka and Bougainville populations are far more closely related to each other than either is to *T. ponceleti*), rather than evidence for a single *T. pseudoponceleti* cluster. Further, with the dataset consisting of loci sequenced in at least six samples, ΔK peaks at K=3. Structurama analyses always resulted in a posterior probability of 1.00 for K=3, regardless of the dataset used or the prior distribution. Coalescent estimates of the species tree in SNAPP strongly prefer the three species model over two species, regardless of the selection of prior on theta and the completeness of the dataset. Finally, demographic analyses strongly reject a model of no divergence between Buka and Bougainville, and suggest that, while gene flow has likely occurred between these populations following divergence, this migration has subsequently decreased substantially (ΔdΔ), or ceased completely (ABC). Further, other scincid lizards with comparable differences in body size (*Plestiodon skiltonianus* group) have been shown to be reproductively isolated as a direct result of body size differences preventing successful copulation (Richmond & Jockusch 2007). While the copulatory posture of *Tribolonotus* is unknown, it is plausible a similar mechanism acts as a partial prezygotic barrier to reproduction between the two island populations. These data thus corroborate each other and provide comprehensive evidence that the Buka and Bougainville populations represent distinct, evolutionarily independent species. These data further support the species level differentiation of these populations under a variety of other species concepts, including the genealogical species concept (Baum & Shaw 1995), the diagnostic or phylogenetic species concept (Cracraft 1989), the phenetic species concept (Sokal & Crovello 1970), the genotypic or genomic species concept (Mallet 1995, 2001), and, arguably, the biological species concept (Mayr 1942, 1970). While the description of this new species of *Tribolonotus* from Buka Island is beyond the scope of this paper, from a biodiversity and conservation standpoint it is critically important that the new species be described. Thus, we are currently preparing an additional manuscript to formally describe this new species.

This divergence between the Buka and Bougainville populations is particularly remarkable given their geographic context. The Buka Passage that separates the two islands is only approximately 300 meters in width, and the islands were repeatedly merged during Pleistocene glacial cycles, including as recently as 10,000 – 20,000 years ago (Chappel & Shackleton 1986; Mayr & Diamond 2001). The few studies have examined phylogeographic patterns of terrestrial species in the Solomon Archipelago have found varied levels of divergence among islands, with some taxa showing limited divergence and others showing deep divergence among islands (Filardi & Smith 2005; Smith & Filardi 2007; Pulvers & Colgan 2007; Hagen *et al.* 2012). However, these studies have not included samples from both Buka and Bougainville, and have focused on species that likely have relatively high dispersal capabilities (e.g. birds, bats, or large, arboreal lizards). In contrast, *Tribolonotus* are semi-fossorial, secretive species,
rarely found outside cover material (e.g. decaying logs), and likely susceptible to desiccation (Greer & Parker 1968; McCoy 2006). Thus, although no estimates of dispersal distances in *Tribolonotus* are available, it is reasonable to expect these species are extremely dispersal limited. Many species in the Solomon Archipelago are known from multiple island groups, including islands far more isolated than Buka and Bougainville, and with no history of connectivity during Pleistocene glacial cycles (Mayr & Diamond 2001; McCoy 2006; Menzies 2006). These results therefore suggest that some such species, particularly those with limited vagility, may instead represent complexes of multiple distinct taxa, and thus vastly underestimating biodiversity of the region.

Demographic analyses reveal further details on the underlying mechanisms responsible for the speciation of *Tribolonotus* between Buka and Bougainville. While ABC pseudo-observed analyses suggests model selection is difficult, the two models receiving posterior support were also identified as best fit models by $\partial a_i \partial i$ analyses, lending further credibility to this model. This model is also reasonable given the geological history of repeated glaciations, connecting the two islands and likely enabling moderate levels of gene flow, and interglacial cycles, isolating the islands and likely limiting or preventing migration. Demographic analyses estimate that the initial divergence between Buka and Bougainville occurred 0.9142 (0.5743-1.4556) $N_e$ generations ago (based on $\partial a_i \partial i$ analyses of the AFS) or 0.7425 (0.1429-2.2280) $N_e$ generations ago (based on ABC), with a shift in population sizes and migration rates occurring at 0.7445 (0.2424-2.2871) $N_e$ generations (based on $\partial a_i \partial i$ analyses of the AFS) or 0.2934 (0.0581-0.7596) $N_e$ generations (based on ABC analyses) before present. SNAPP analyses suggest a far more recent divergence of approximately 0.15 $N_e$ generations; however, this disparity is likely the result of from SNAPP analyses not accounting for migration, as even low levels of gene flow can result in underestimates of divergence time when no migration is assumed. This result further highlights the need to test for and, where necessary, account for migration in estimates of divergence time.

Unfortunately, with no useful fossil calibrations available, and no estimates of generation time in *Tribolonotus*, it is difficult to put the divergence in terms of years or the migration rates in terms of proportion of each population composed of migrants (Graur & Martin 2004; Donoghue & Benton 2007). However, if we estimate a mutation rate, $\mu$, of between 0.1 and 1.0 % per million years, and assume a generation time of between 1.7 years, as estimated for *Scincella lateralis* (a smaller, more active, but temperate lygosomine scincid lizard; Jackson & Austin 2010), and 6.53 years, as estimated for *Gnypetoscincus queenslandiae* (an ecologically similar, Australian tropical lygosomine lizard; Sumner *et al.* 2001), we can approximate a plausible estimate of divergence time and migration rates (Table 5.5). While these calculations result in broad intervals due to the uncertainty in mutation rate and generation time, compounded by the uncertainty in the parameter estimates, and as such represent, at best, general estimates of plausible values for these evolutionary parameters, they do highlight several important features. First, the initial divergence between Buka and Bougainville likely occurred in the Pliocene or early to mid-Pleistocene, and thus may predate Pleistocene glaciations, or correspond with early interglacial periods. Second, we estimate a shift in population size and migration rate occurring during the Pleistocene. While a more realistic model may include multiple such shifts, corresponding with glacial cycles, estimating such shifts along the branches of the phylogeny is
extremely difficult, as exhibited by the breadth of the confidence intervals on migration estimates prior to this shift (Table 5.4, Fig. 5.5). Thus, accurately testing such a model would likely require

Table 5.5. Estimated parameters values based the maximum likelihood (\(\partial a/\partial i\)) or geometric mean (ABC) parameter estimates. Estimates were scaled using the geometric mean estimate of ancestral theta (\(\theta_a\)), a mutation rate, \(\mu\), of 0.1% per million years or 1.0% per million years, and a generation time of 1.7 or 6.53 years. Confidence intervals calculated based on the limits of 95% HPD interval on ancestral theta (\(\theta_a\)) and the confidence intervals on the parameter estimates from \(\partial a/\partial i\), or the limits of the 95% HPD interval for ABC. See Figure 2 for details of parameters.

<table>
<thead>
<tr>
<th></th>
<th>(\mu=1% / \text{My}, \text{gen}=1.7 \text{ yrs})</th>
<th>(\mu=0.1% / \text{My}, \text{gen}=6.53 \text{ yrs})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Divergence Time - (\partial a/\partial i) (years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>143,997</td>
<td>5,531,184</td>
</tr>
<tr>
<td></td>
<td>(26,095, 483,385)</td>
<td>(1,002,362, 18,567,671)</td>
</tr>
<tr>
<td>T3</td>
<td>117,253</td>
<td>4,503,894</td>
</tr>
<tr>
<td></td>
<td>(11,012, 759,505)</td>
<td>(422,988, 29,173,913)</td>
</tr>
<tr>
<td>Divergence Time - ABC (years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>115,154</td>
<td>4,423,287</td>
</tr>
<tr>
<td></td>
<td>(4,871, 724,997)</td>
<td>(187,119, 27,848,400)</td>
</tr>
<tr>
<td>T3</td>
<td>45,964</td>
<td>1,765,555</td>
</tr>
<tr>
<td></td>
<td>(1,837, 232,039)</td>
<td>(70,561, 8,913,035)</td>
</tr>
<tr>
<td>Proportion New Migrants Per Generation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>m12a</td>
<td>1.27E-05</td>
<td>1.27E-06</td>
</tr>
<tr>
<td></td>
<td>(4.18E-07, 6.38E-04)</td>
<td>(4.18E-08, 6.38E-05)</td>
</tr>
<tr>
<td>m21a</td>
<td>4.26E-05</td>
<td>4.26E-06</td>
</tr>
<tr>
<td></td>
<td>(1.87E-07, 1.60E-02)</td>
<td>(1.87E-08, 1.60E-03)</td>
</tr>
<tr>
<td>m12b</td>
<td>6.61E-06</td>
<td>6.61E-07</td>
</tr>
<tr>
<td></td>
<td>(4.45E-07, 1.62E-04)</td>
<td>(4.45E-08, 1.62E-05)</td>
</tr>
<tr>
<td>m21b</td>
<td>2.71E-06</td>
<td>2.71E-07</td>
</tr>
<tr>
<td></td>
<td>(7.94E-08, 1.52E-04)</td>
<td>(7.94E-09, 1.52E-05)</td>
</tr>
<tr>
<td># New Migrants Per Generation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>m12a</td>
<td>0.59</td>
<td>5.899</td>
</tr>
<tr>
<td></td>
<td>(0.041, 8.528)</td>
<td>(0.408, 85.282)</td>
</tr>
<tr>
<td>m21a</td>
<td>1.972</td>
<td>19.717</td>
</tr>
<tr>
<td></td>
<td>(0.018, 213.421)</td>
<td>(0.182, 2134.209)</td>
</tr>
<tr>
<td>m12b</td>
<td>0.306</td>
<td>3.063</td>
</tr>
<tr>
<td></td>
<td>(0.043, 2.158)</td>
<td>(0.435, 21.585)</td>
</tr>
<tr>
<td>m21b</td>
<td>0.126</td>
<td>1.256</td>
</tr>
<tr>
<td></td>
<td>(0.008, 2.034)</td>
<td>(0.078, 20.339)</td>
</tr>
</tbody>
</table>
denser sampling, particularly with regards to the number of alleles sampled per species. Finally, even at the upper limits of the estimated migration rates between Buka and Bougainville, only a fraction of a percent of each population is composed of new migrants each generation (Table 5.5), suggesting that, while migration appears to have occurred subsequent to the divergence of these populations, it has not occurred at a rate sufficient to overwhelm divergence, and provides further evidence that these populations represent evolutionarily independent lineages.

The proposed scenario, as estimated via the demographic analyses, posits no necessity for ecological divergence between these species, although it does not preclude this possibility. Indeed, differences in limb lengths, digit lengths, and body size have all been shown to correlate with ecological divergence in lizards (Melville & Swain 2000; Kohlsdorf et al. 2001; Goodman 2007). However, the ecology of Tribolonotus is poorly known, and, beyond the morphological differences described here, no data is available to test if ecological divergence may accompany the speciation of the Buka and Bougainville populations. More detailed studies aimed specifically at examining the ecology of these species are necessary to assess if the divergence between Buka and Bougainville occurred solely as a result of allopatric divergence, or if ecological divergence plays an important role along with allopatry in driving and maintaining speciation in this group.

The shallow divergence between the Bougainville and Buka populations illustrates the power of NGS data to elucidate evolutionary processes in recently diverged species groups. In spite of this power, it is through combining this genomic data with morphological data that we are able to assess the validity of the Buka and Bougainville populations under a variety of species concepts, and thus provide robust evidence of the status of these populations as distinct species. While many systems do not require a genomic dataset of the scale collected in this study for the accurate and robust validation of putative species, this study provides a framework for how next-generation sequencing data can be combined with morphological data and leveraged for species validation, and how further analyses of genomic data can facilitate elucidating more detailed demographic processes associated with speciation.
6.1. Species Delimitation

Despite being among the most debated topics in biology (Sokal & Crovello 1970; Mallet 2001; Lee 2003; Coyne & Orr 2004; de Queiroz 2005, 2007; Hausdorf 2011), species are a fundamental unit of the life sciences, critical to a wide variety of studies (Agapow et al. 2004; Agapow 2005; de Queiroz 2007; Bickford et al. 2007; Bortolus 2008; Wheeler 2008). Yet species remain primarily delimited using relatively subjective means, and little work has focused on developing or testing more objective means for delimiting species, particularly from molecular data (Sites Jr. & Marshall 2003, 2004; Wiens 2007; Burbrik et al. 2012; Carstens et al. 2013). While the formal description of new species following the traditional framework laid out by such organizations as the International Code of Zoological Nomenclature (ICZN 1999) remains of utmost importance (Bauer et al. 2011), increasing the objectivity of species delimitation is an important avenue for research in systematic biology. A number of new methods for the more objective delimitation of species have been described, including phylogeny-based (e.g. nonparametric delimitation, O’Meara, 2010; Bayesian species delimitation, Yang & Rannala, 2010, generalized mixed Yule coalescent model, Monaghan et al., 2009; Pons et al., 2006; Reid & Carstens, 2012) and non-phylogeny-based (e.g. fields for recombination, Doyle, 1995; Structurama, Huelsenbeck, Andolfatto, & Huelsenbeck, 2011; Huelsenbeck & Andolfatto, 2007) methods, validation methods that require a priori hypotheses of putative species (e.g. Bayes factor delimitations, Grummer, Bryson, & Reeder, 2014, spedeSTEM, Ence & Carstens, 2011; Bayesian species delimitation Yang & Rannala, 2010) and discovery methods that do not (e.g. nonparametric delimitation, O’Meara, 2010, generalized mixed Yule coalescent model, Monaghan et al., 2009; Pons et al., 2006; Reid & Carstens, 2012; Structurama, Huelsenbeck et al., 2011; Huelsenbeck & Andolfatto, 2007); however, little work has focused on the utility of these methods, such as their accuracy, sensitivity to sampling strategy, or ability to accurately delimit species in with large, speciose groups.

In Chapter 2, I investigated the influence of sampling strategy, both in terms of number of samples per species and number of loci, on several methods for species delimitation that do not rely on a priori assignment of samples to putative species. These analyses reveal that with a moderate number of loci (ten or more), Structurama (Huelsenbeck & Andolfatto 2007; Huelsenbeck et al. 2011) can accurately delimit even recently divergent (greater than 1.5 Ne generations) species, and that Gaussian clustering (Hausdorf & Hennig 2010) is similarly accurate, though less sensitive to detecting recent divergences, accurately delimiting divergences greater than roughly 2.5 Ne generations. I also show that nonparametric delimitation (O’Meara 2010) can be highly accurate when the true gene genealogies are known. However, nonparametric delimitation, as currently implemented, uses point estimates of gene genealogies and cannot account for phylogenetic uncertainty, but is highly sensitive to errors in gene genealogy estimation. Thus, while theoretically accurate, the errors in gene genealogy estimation and phylogenetic uncertainty present in empirical data limit the utility of the method.

I apply these methods for species delimitation empirically using the *Carlia bicarinata* and *C. fusca* groups in Chapters 3 and 4, respectively. In the application to the *C. bicarinata*
I found Gaussian clustering to be prone to detecting only higher levels of clustering in this hierarchical dataset: based on the complete dataset, only major, deeply divergent clades were detected via Gaussian clustering. However, reanalyzing the data within the initially detected clades recovered finer levels of structure. Structurama apparently performed well, and delimited species that corresponded to the secondary, within cluster Gaussian clustering analyses, and the phylogenetic analyses with a single run under a variety of prior settings. In the more complex and speciose C. fusca complex examined in Chapter 4, however, both Gaussian clustering and Structurama were more problematic based on the complete dataset. Gaussian clustering analyses only identified a small number of clusters based on the complete datasets, and these clusters corresponded poorly to major clades identified in phylogenetic analyses. While this method performed better for delimiting species when limited to the major clades, corroboration between these results and other analyses was limited, suggesting an overall poor performance of this method in the C. fusca group. Structurama analyses based on the complete dataset similarly delimited only a small number of clusters despite the large prior mean on number of clusters. However, these clusters largely corresponded to the major clades recovered in the phylogenetic analyses. Further, analyses restricted to within each major clade resulted in clusters that largely corresponded to the finer level clustering recovered in the phylogenetic analyses. Subsequent analysis via Bayesian species delimitation collapsed several of these lower level clusters, highlighting the importance of further validation of species delimitations using other methods and other datasets.

Finally, in chapter five I reveal the potential for genomic scale data collected via next generation sequencing for species delimitation in both species discovery and species validation frameworks. With the large dataset of several thousand loci collected in this study, I found strong evidence for the recent speciation between the Tribolontus populations from the geographically proximate islands of Buka and Bougainville using genetic clustering algorithms (Pritchard et al. 2000; Falush et al. 2003; Huelsenbeck & Andolfatto 2007; Huelsenbeck et al. 2011) for species discovery and Bayes factors delimitations (Leaché et al. 2013; Grummer et al. 2014) for species validation; I also provided further evidence for this divergence through the analysis of morphological data.

Combined, these results show the potential of various methods for species delimitation, particularly Structurama and Gaussian clustering. These data also highlight the difficulties of these and other methods, including the strong potential for errors under nonparametric delimitation due to the failure of the method to account for phylogenetic uncertainty or errors in gene genealogy estimation. In particular, I also find that Structurama and Gaussian clustering, while generally accurate and useful for species delimitation, are prone to fail to delimit species in groups that are species rich or have high levels of deeper phylogenetic structure. However, further work is necessary to determine if these difficulties represent a peculiarity of the particular datasets, such as their complexity or the large number of species relative to the number of loci, or if these results reveal the importance of these features as a hindrance to species delimitation. Thus, with both Structurama and Gaussian clustering analyses, it is clearly important to carefully examine the results, and to reanalyze initially identified clusters independently to test for further potentially species-level clustering. However, it is also imperative that these results be tested via other methods, such as species validation methods like Bayesian species delimitation (Yang & Rannala 2010) or Bayes factors delimitations (Leaché et al. 2013; Grummer et al. 2014), and, as
Carstens et al. (2013) argued, studies endeavoring to delimit species should apply multiple methods and compare among the results in order to identify and validate lineages and to ensure the accuracy of species delimitation.

### 6.2. Implications for the Biodiversity and Biogeography of New Guinea and the Sahul Shelf Region

In Chapter 3, I found evidence for speciation among the populations previously assigned to *Carlia storri* from New Guinea, Australia, and the Aru Islands. Previous work on taxa whose distribution spans the Torres Strait separating New Guinea from Australia has been limited, and largely has focused on species with comparatively high vagilities, such as pythons (Rawlings & Donnellan 2003), elapid snakes (Wüster et al. 2005), or birds (Murphy et al. 2007). However, these studies have found little divergence among the Aru Islands, Cape York, Australia, and the Trans-Fly region of southern New Guinea. Only small decreases in sea level results in the formation of land bridges among these regions (Voris 2000). Thus, the divergence I observed among these regions is relatively surprising. It does, however, highlight the need for further phylogeographic study of groups with a variety of vagilities and with distributions that span the Torres Strait to elucidate the broader role of this transient barrier in driving diversification.

The *Carlia fusca* group examined in Chapter 4 yields extensive additional insight into the biogeography and diversity of the Sahul shelf region. While my sampling included 16 of the 18 currently recognized species in the group, species delimitation analyses recover a total of 28 distinct species. Further, these delimited species correspond poorly to currently recognized taxonomy of the group, including cryptic species, synonymous species, and mis-assigned populations, corroborating previous work that found extensive incongruence between molecular data and current taxonomy (Austin et al. 2011). These results also corroborate a growing number of recent studies that have found extensive cryptic diversity within broadly distributed species in the New Guinea region (De Bruyn et al. 2004; Murphy et al. 2007; Kraus 2008; Benz 2011; Deiner et al. 2011; Macqueen et al. 2011; Zug & Fisher 2012; Oliver et al. 2013), and suggest that broad distributions are atypical in the region. Biogeographic analyses suggest an Australian origin for the species group, with subsequent dispersal to and diversification within New Guinea, as well as two independent colonizations across Lydekker’s line into Wallacea, and a recolonization of Australia. On a finer geographic scale, I also find some evidence for the importance of riverine barriers in driving diversification in the region. Samples from Kiunga and Kwatu in south central New Guinea were collected only 25 kilometers apart, but are separated by the Fly River, the largest river in New Guinea by volume (Nilsson et al. 2005); these species were recovered as deeply divergent, and belonging to different major clades within the *C. fusca* group, suggesting the Fly River may play an important role in maintaining the divergence between these parapatrically distributed species. Samples from opposite sides of another major river in New Guinea, the Sepik River, however, were found to be conspecific. Finally, I also found a unique and particularly enigmatic pattern in the distribution of *C. leucotaenia*. This species was found to include samples from Seram and Halmahera in Wallacea, the Bomberai Peninsula in western New Guinea, and the Northern Melanesian islands of the Admiralty and Bismarck Archipelagos, while its closest relatives were restricted to Australia and south central New Guinea. I additionally found evidence for the recent, probably anthropogenic introduction of this species into the major port towns of Madang and Lae, and previous work found this
species to be the source for human introduced populations of *Carlia* in the Pacific islands of Guam, Palau, and the Northern Marianas (Austin *et al.* 2011). The extent to which the bizarre distribution of this species is natural, versus a result of recent, inadvertent introductions by humans is an area in need of further research.

In Chapter 5, I reveal speciation on a fine scale between *Tribolonotus* on the islands of Buka and Bougainville in the northwestern Solomon Archipelago, despite the historic connections between these islands during Pleistocene glaciations (Chappel & Shackleton 1986; Chappell *et al.* 1996; Mayr & Diamond 2001). Demographic analyses further suggest that subsequent to their divergence, gene flow between these species continued, possibly during periods of low sea levels when the islands were merged, but that migration has since decreased in rate or ceased entirely. These results corroborate other studies that have similarly found high levels of divergence among islands in the Solomon Archipelago (Filardi & Smith 2005; Smith & Filardi 2007; Pulvers & Colgan 2007; Hagen *et al.* 2012; Andersen *et al.* 2013), and suggest that the diversity in this imperiled biodiversity hotspot may be vastly underestimated.

6.3. Taxonomic Implications and Lingering Taxonomic Questions

At deeper phylogenetic scales, the results of Chapters 3 and 4 reveal a strongly supported, monophyletic clade comprised of the *Carlia fusca* and *C. bicornata* groups, including the *C. schmeltzii* complex. Within this clade, the *C. fusca* group, the *C. bicornata* group, and the *C. schmeltzii* complex are recovered as strongly supported monophyletic clades; however, while concatenated analyses support a sister relationship between the *C. fusca* group and the *C. bicornata* group (excluding the *C. schmeltzii* complex; maximum likelihood bootstrap support, MLBS > 79, Bayesian posterior probability, PP = 1.0), species tree analyses are unable to resolve this node (PP < 0.42). This difficulty likely reflects the short internode separating these three clades, reflecting the similar timing of divergence among them. The results of the concatenated phylogeny would result in a paraphyletic *C. bicornata* group with respect to the *C. fusca* group. Further, while the northern species in the *C. schmeltzii* complex is morphologically similar to having a relatively gracile body and bicornate scales, the southern species is more similar to the *C. fusca* group in having tricornate scales and a more robust body (Ingram & Covacevich 1989), and based on these southern populations, *C. schmeltzii* was previously synonymized with *Leiolopisma fusca fuscum* (=*C. fusca*), before being resurrected by Ingram and Covacevich (1989). Therefore, I suggest that treating the *C. schmeltzii* complex as a separate species group, distinct from the *C. bicornata* group, is a better reflection of our current understanding of their evolutionary history and diversity.

Within the *Carlia schmeltzii* group, my results in Chapter 3 provide strong support for the treatment of the northern and southern morphotypes as distinct species. The type locality for *C. schmeltzii* is Rockhampton, Queensland, within the range of the southern species. Covacevich & Ingram (1975) previously described the northern species as *C. prava* (type locality: Magnificent Creek, Kowanyama, North Queensland), which was subsequently synonymized with *C. schmeltzii* in Ingram and Covacevich (1989). I propose the resurrection of *C. prava* for the northern species in the *C. schmeltzii* group. However, both species, as well as intermediate individuals, are documented from the apparent contact zone between the species in the Townsville area suggesting the two species may form a hybrid zone in this area. Further work
should focus on examining this putative hybrid zone with denser sampling to determine the extent to which the two species may hybridize.

In Chapter 3, species delimitation analyses also reveal further species diversity within the *C. bicarinata* group. The new species from Amau village, southeastern Papua New Guinea is recovered as a distinct species, sister to *C. bicarinata* with strong support. Within what is currently recognized as *C. storri*, analyses strongly support the presence of three distinct and geographically partitioned species: one from the Trans-Fly region of southern New Guinea, one from the Aru Islands, and one from Cape York, Australia. The type locality of *C. storri* is in Cape York, Australia (Dulhunty River Crossing on Telegraph Road, 110 km S of Bamaga, Cape York, Queensland, Australia; -11.833º, 142.5º; Ingram & Covacevich, 1989). Therefore, I restrict *C. storri* to the Australian populations referred to this species. However, no names are available for the Aru Islands or New Guinea populations, or for the other new species, *C. sp. Amau*; formal descriptions of these species are underway.

My examination of the *Carlia fusca* group in Chapter 4 has particularly extensive taxonomic implications. While 18 species are currently recognized in the group, 16 of which were included in my sampling, I found evidence for 28 species of *C. fusca* group skink, and extensive discordance between these species and the currently recognized taxonomy. Several recognized species (e.g. *C. aramia*, *C. eothen*, *C. mysi*) were found to include multiple distinct species, and frequently included populations from deeply divergent clades; many of these represent previously unrecognized, cryptic diversity, but some also represent populations incorrectly assigned to species in previous studies (e.g. *Carlia* from the Bismarck Archipelago). I would argue that recognizing populations as an undescribed species is preferable to inaccurately including it within another described species; therefore, based on these results and the type localities for these species, I restrict *C. mysi* to the Huon Peninsula, and *C. eothen* to the Trobriand Islands. Unfortunately, I was unable to obtain topotypic samples of *C. aramia*, and thus cannot determine which, if any, of the species identified in my analyses this species epithet refers to. Further, I also find low divergence among some currently recognized species, particularly among *C. leucotaenia* from Seram, *C. tutela* from Halmahera, *C. bomberai* from the Vogelkopf region of western New Guinea, *C. ailanpalai* from the Admiralty Archipelago, and the populations assigned to *C. mysi* from the Bismarck Archipelago, and recover these as a single species in species delimitation analyses. The latter three of these species were recently described based on only minor morphological differences, thus the lack of divergence among these populations is unsurprising. Therefore, I suggest synonymizing *C. ailanpalai*, *C. bomberai*, and *C. tutela* with *C. leucotaenia*, which as priority, as well as assigning the populations of *Carlia* from the Bismarck Archipelago to this species. The morphological differences between this species and the nominate species of the group, *C. fusca*, are also slight, and *C. fusca* occurs in the northern Vogelkopf region, geographically close to much of the distribution of this species. Therefore, it is plausible that this species and *C. fusca* are also conspecific. However, until either further morphological study can test this, or topotypic tissue samples for *C. fusca* are collected and analyzed, I suggest the retention of *C. leucotaenia* for this biogeographically unique species. These changes are a small step towards reconciling the taxonomy of the group with the species identified in this study; however, numerous new species of *C. fusca* group skinks need to be described.
Finally, in Chapter 5, my analyses of genomic and morphological data provide evidence for a recent speciation between the *Tribolonotus pseudoponceleti* from Buka and Bougainville Islands in the northwestern Solomon Archipelago. The type locality for this species is Kunua, Bougainville (Greer & Parker 1968), and the holotype clearly clusters with other specimens from Bougainville, including those in the genomic analyses. Formal description of the Buka species is in progress. One important taxonomic question lingers with respect to the *T. pseudoponceleti* group: the status of the Choiseul populations. No tissue samples are available for this population, and so it could not be included in the genomic analyses; however, further work is needed to determine the status of this population, and determine if this population is a member of either the Bougainville or Buka populations, or if it represents another undescribed species in the group.

### 6.4. Conclusions

My primary goals in undertaking this research were to understand how methods for delimiting species from molecular data perform under various scenarios, such as different sampling strategies and in speciose clades, to examine evolution and diversification in New Guinea, and to clarify the systematics of the scincid genera *Carlia* and *Tribolonotus* in New Guinea. My results show that even with relatively small sample sizes of a few individuals per species and ten loci, some methods, such as Structurama and Gaussian clustering, can accurately delimit species, but phylogeny-based methods for species delimitation may be highly error prone if phylogenetic uncertainty is not accounted for. I also demonstrate that methods for species delimitation may tend to underestimate diversity in clades with deep phylogenetic structure or high species richness. Finally, I find evidence extensive cryptic diversity in New Guinea, elucidate patterns and potential drivers of diversification in the region, and make some strives towards reconciling the discordance between current taxonomy and the results of these studies. While much remains to be done, this work substantially improves our understanding of the strengths and caveats of species delimitation, and of biogeography and biodiversity in the megadiverse and dynamic landscape of New Guinea.


Bazinet AL, Neel MC, Shaw KL, Cummings MP The genealogical sorting index : software and web site for quantifying the exclusivity of lineages., 1–15.


Pyron RA, Burbrink FT (2009) Systematics of the Common Kingsnake (Lampropeltis getula; Serpentes: Colubridae) and the burden of heritage in taxonomy. Zootaxa, 2241, 22–32.


Reid NM, Carstens BC (2012) Phylogenetic estimation error can decrease the accuracy of species delimitation: a Bayesian implementation of the general mixed Yule-coalescent model. *BMC evolutionary biology, 12,* 196.


Zwickl DJ (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. The University of Texas at Austin.
APPENDIX A
PERMISSIONS FOR PORTIONS OF CHAPTER 1

OXFORD UNIVERSITY PRESS LICENSE
TERMS AND CONDITIONS

Jun 09, 2014

This is a License Agreement between Eric N Rittmeyer (“You”) and Oxford University Press (“Oxford University Press”) provided by Copyright Clearance Center (“CCC”). The license consists of your order details, the terms and conditions provided by Oxford University Press, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

<table>
<thead>
<tr>
<th>License Number</th>
<th>3405030018243</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Jun 09, 2014</td>
</tr>
<tr>
<td>Licensed content publisher</td>
<td>Oxford University Press</td>
</tr>
<tr>
<td>Licensed content publication</td>
<td>Systematic Biology</td>
</tr>
<tr>
<td>Licensed content title</td>
<td>Species Concepts and Species Delimitation:</td>
</tr>
<tr>
<td>Licensed content author</td>
<td>Kevin De Queiroz</td>
</tr>
<tr>
<td>Licensed content date</td>
<td>12/01/2007</td>
</tr>
<tr>
<td>Type of Use</td>
<td>Thesis/Dissertation</td>
</tr>
<tr>
<td>Institution name</td>
<td>None</td>
</tr>
<tr>
<td>Title of your work</td>
<td>SYSTEMATICS AND SPECIES DELIMITATION IN NEW GUINEA SKINK SPECIES COMPLEXES (SQUAMATA: SCINCIDAE)</td>
</tr>
<tr>
<td>Publisher of your work</td>
<td>n/a</td>
</tr>
<tr>
<td>Expected publication date</td>
<td>Jun 2014</td>
</tr>
<tr>
<td>Permissions cost</td>
<td>0.00 USD</td>
</tr>
<tr>
<td>Value added tax</td>
<td>0.00 USD</td>
</tr>
<tr>
<td>Total</td>
<td>0.00 USD</td>
</tr>
<tr>
<td>Total</td>
<td>0.00 USD</td>
</tr>
<tr>
<td>Terms and Conditions</td>
<td></td>
</tr>
</tbody>
</table>

STANDARD TERMS AND CONDITIONS FOR REPRODUCTION OF MATERIAL FROM AN OXFORD UNIVERSITY PRESS JOURNAL

1. Use of the material is restricted to the type of use specified in your order details.

2. This permission covers the use of the material in the English language in the following territory: world. If you have requested additional permission to translate this material, the
This is a License Agreement between Eric N Rittmeyer (“You”) and Elsevier (“Elsevier”) provided by Copyright Clearance Center (“CCC”). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier
Elsevier Limited
The Boulevard, Langford Lane
Kidlington, Oxford, OX5 1GB, UK

Registered Company Number
1982084

Customer name
Eric N Rittmeyer

Customer address
Museum of Natural Science
Baton Rouge, LA 70803

License number
3404981219324

License date
Jun 09, 2014

Licensed content publisher
Elsevier

Licensed content publication
Molecular Phylogenetics and Evolution

Licensed content title
The effects of sampling on delimiting species from multi-locus sequence data

Licensed content author
Eric N. Rittmeyer, Christopher C. Austin

Licensed content date
November 2012

Licensed content volume number
65

Licensed content issue number
2

Number of pages
13

Start Page
451

End Page
463

Type of Use
reuse in a thesis/dissertation

Portion
full article

Format
both print and electronic

Are you the author of this Elsevier article?
Yes
Table C.1. Collection localities for all included *C. bicarinata* group and outgroup samples, as well as, for *Carlia* samples. PNG refers to Papua New Guinea, AUS to Australia, and IND to Indonesia. States and provinces are abbreviated as: CP: Central Province, PNG; GP: Gulf Province, PNG; MBP: Milne Bay Province, PNG; MkP: Maluku Province, IND; MnP: Manus Province, PNG; NCD: National Capital District, PNG; NSW: New South Wales, Australia; NT: Northern Territory, Australia; QLD: Queensland, Australia; Vic.: Victoria, Australia; WP: Western Province, PNG.

<table>
<thead>
<tr>
<th>Cat.No.</th>
<th>Genus</th>
<th>Species</th>
<th>Locality</th>
<th>Latitude</th>
<th>Longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCA16357</td>
<td><em>Carlia</em></td>
<td>bicarinata</td>
<td>PNG: CP: Yule Island</td>
<td>-8.82865</td>
<td>146.52869</td>
</tr>
<tr>
<td>CCA16358</td>
<td><em>Carlia</em></td>
<td>bicarinata</td>
<td>PNG: CP: Yule Island</td>
<td>-8.82865</td>
<td>146.52869</td>
</tr>
<tr>
<td>CCA16359</td>
<td><em>Carlia</em></td>
<td>bicarinata</td>
<td>PNG: CP: Yule Island</td>
<td>-8.82865</td>
<td>146.52869</td>
</tr>
<tr>
<td>CCA16360</td>
<td><em>Carlia</em></td>
<td>bicarinata</td>
<td>PNG: CP: Yule Island</td>
<td>-8.82865</td>
<td>146.52869</td>
</tr>
<tr>
<td>CCA16436</td>
<td><em>Carlia</em></td>
<td>bicarinata</td>
<td>PNG: CP: Yule Island</td>
<td>-8.82865</td>
<td>146.52869</td>
</tr>
<tr>
<td>CCA16437</td>
<td><em>Carlia</em></td>
<td>bicarinata</td>
<td>PNG: CP: Yule Island</td>
<td>-8.82865</td>
<td>146.52869</td>
</tr>
<tr>
<td>CCA2292</td>
<td><em>Carlia</em></td>
<td>bicarinata</td>
<td>PNG: NCD: Port Moresby</td>
<td>-9.42642</td>
<td>147.18995</td>
</tr>
<tr>
<td>CCA22758</td>
<td><em>Carlia</em></td>
<td>bicarinata</td>
<td>PNG: NCD: Port Moresby</td>
<td>-9.47328</td>
<td>147.17760</td>
</tr>
<tr>
<td>CCA22763</td>
<td><em>Carlia</em></td>
<td>bicarinata</td>
<td>PNG: NCD: Port Moresby</td>
<td>-9.47328</td>
<td>147.17760</td>
</tr>
<tr>
<td>CCA3827</td>
<td><em>Carlia</em></td>
<td>bicarinata</td>
<td>PNG: CP: Bootless Bay</td>
<td>-9.50933</td>
<td>147.29018</td>
</tr>
<tr>
<td>CCA3828</td>
<td><em>Carlia</em></td>
<td>bicarinata</td>
<td>PNG: CP: Bootless Bay</td>
<td>-9.50933</td>
<td>147.29018</td>
</tr>
<tr>
<td>CCA5592</td>
<td><em>Carlia</em></td>
<td>sp. Amau</td>
<td>PNG: CP: Amau village</td>
<td>-10.03680</td>
<td>148.56463</td>
</tr>
<tr>
<td>CCA5869</td>
<td><em>Carlia</em></td>
<td>sp. Amau</td>
<td>PNG: CP: Amau village</td>
<td>-10.03680</td>
<td>148.56463</td>
</tr>
<tr>
<td>CCA5870</td>
<td><em>Carlia</em></td>
<td>sp. Amau</td>
<td>PNG: CP: Amau village</td>
<td>-10.03680</td>
<td>148.56463</td>
</tr>
<tr>
<td>CCA5879</td>
<td><em>Carlia</em></td>
<td>sp. Amau</td>
<td>PNG: CP: Amau village</td>
<td>-10.03680</td>
<td>148.56463</td>
</tr>
<tr>
<td>ABTC11026</td>
<td><em>Carlia</em></td>
<td>storri</td>
<td>AUS: QLD: near Mossman</td>
<td>-16.461</td>
<td>145.373</td>
</tr>
<tr>
<td>ABTC11027</td>
<td><em>Carlia</em></td>
<td>storri</td>
<td>AUS: QLD: near Mossman</td>
<td>-16.461</td>
<td>145.373</td>
</tr>
<tr>
<td>ABTC11028</td>
<td><em>Carlia</em></td>
<td>storri</td>
<td>AUS: QLD: near Mossman</td>
<td>-16.461</td>
<td>145.373</td>
</tr>
<tr>
<td>DSP05</td>
<td><em>Carlia</em></td>
<td>storri</td>
<td>AUS: QLD: Magnetic Island</td>
<td>-19.136</td>
<td>146.830</td>
</tr>
<tr>
<td>QMJ78178</td>
<td><em>Carlia</em></td>
<td>storri</td>
<td>AUS: QLD: Lakeland, Cooktown Rd</td>
<td>-15.862</td>
<td>144.855</td>
</tr>
<tr>
<td>ALS182</td>
<td><em>Carlia</em></td>
<td>&quot;storri&quot;</td>
<td>IND: MkP: Aru Islands</td>
<td>-6.88459</td>
<td>134.26018</td>
</tr>
<tr>
<td>ALS213</td>
<td><em>Carlia</em></td>
<td>&quot;storri&quot;</td>
<td>IND: MkP: Aru Islands</td>
<td>-6.85517</td>
<td>134.26065</td>
</tr>
<tr>
<td>ALS214</td>
<td><em>Carlia</em></td>
<td>&quot;storri&quot;</td>
<td>IND: MkP: Aru Islands</td>
<td>-6.85517</td>
<td>134.26065</td>
</tr>
<tr>
<td>ALS254</td>
<td><em>Carlia</em></td>
<td>&quot;storri&quot;</td>
<td>IND: MkP: Aru Islands</td>
<td>-6.81289</td>
<td>134.37794</td>
</tr>
<tr>
<td>CCA16071</td>
<td><em>Carlia</em></td>
<td>&quot;storri&quot;</td>
<td>PNG: WP: Daru</td>
<td>-9.07353</td>
<td>143.21118</td>
</tr>
<tr>
<td>CCA16078</td>
<td><em>Carlia</em></td>
<td>&quot;storri&quot;</td>
<td>PNG: WP: Daru</td>
<td>-9.07353</td>
<td>143.21118</td>
</tr>
<tr>
<td>CCA16102</td>
<td><em>Carlia</em></td>
<td>&quot;storri&quot;</td>
<td>PNG: WP: Daru</td>
<td>-9.07353</td>
<td>143.21118</td>
</tr>
<tr>
<td>CCA16103</td>
<td><em>Carlia</em></td>
<td>&quot;storri&quot;</td>
<td>PNG: WP: Daru</td>
<td>-9.07353</td>
<td>143.21118</td>
</tr>
<tr>
<td>CCA16156</td>
<td><em>Carlia</em></td>
<td>&quot;storri&quot;</td>
<td>PNG: WP: Daru</td>
<td>-9.07353</td>
<td>143.21118</td>
</tr>
<tr>
<td>ABTC16223</td>
<td><em>Carlia</em></td>
<td>schmeltzii - North</td>
<td>AUS: QLD: Townsville</td>
<td>-19.259</td>
<td>146.802</td>
</tr>
<tr>
<td>DSF06</td>
<td><em>Carlia</em></td>
<td>schmeltzii - North</td>
<td>AUS: QLD: Magnetic Island</td>
<td>-19.136</td>
<td>146.830</td>
</tr>
<tr>
<td>ABTC11024</td>
<td><em>Carlia</em></td>
<td>schmeltzii - South</td>
<td>AUS: QLD: Gladstone</td>
<td>-23.863</td>
<td>151.248</td>
</tr>
<tr>
<td>ABTC32058</td>
<td><em>Carlia</em></td>
<td>schmeltzii - South</td>
<td>AUS: QLD: Sarina</td>
<td>-21.424</td>
<td>149.218</td>
</tr>
<tr>
<td>CCA2033</td>
<td><em>Carlia</em></td>
<td>ailanpalai</td>
<td>PNG: MnP: Los Negros Is., Rio Rio Village</td>
<td>-2.04917</td>
<td>147.41867</td>
</tr>
<tr>
<td>CCA2039</td>
<td><em>Carlia</em></td>
<td>ailanpalai</td>
<td>PNG: MnP: Los Negros Is., Rio Rio Village</td>
<td>-2.04917</td>
<td>147.41867</td>
</tr>
<tr>
<td>Cat.No.</td>
<td>Genus</td>
<td>Species</td>
<td>Locality</td>
<td>Latitude</td>
<td>Longitude</td>
</tr>
<tr>
<td>----------</td>
<td>--------</td>
<td>-------------</td>
<td>-----------------------------------------------</td>
<td>-----------</td>
<td>------------</td>
</tr>
<tr>
<td>CCA2040</td>
<td>Carlia</td>
<td>ailanpalai</td>
<td>PNG: MnP: Los Negros Is., Rio Rio Village</td>
<td>-2.04917</td>
<td>147.41867</td>
</tr>
<tr>
<td>CCA2046</td>
<td>Carlia</td>
<td>ailanpalai</td>
<td>PNG: MnP: Los Negros Is., Peyton Village</td>
<td>-2.03267</td>
<td>147.43417</td>
</tr>
<tr>
<td>CCA15382</td>
<td>Carlia</td>
<td>eothen</td>
<td>PNG: MBP: Trobriand Islands, Kiriwina</td>
<td>-8.54221</td>
<td>151.07916</td>
</tr>
<tr>
<td>CCA15422</td>
<td>Carlia</td>
<td>eothen</td>
<td>PNG: MBP: Trobriand Islands, Kiriwina</td>
<td>-8.54221</td>
<td>151.07916</td>
</tr>
<tr>
<td>CCA15425</td>
<td>Carlia</td>
<td>eothen</td>
<td>PNG: MBP: Trobriand Islands, Kiriwina</td>
<td>-8.54221</td>
<td>151.07916</td>
</tr>
<tr>
<td>CCA15450</td>
<td>Carlia</td>
<td>eothen</td>
<td>PNG: MBP: Trobriand Islands, Kiriwina</td>
<td>-8.54221</td>
<td>151.07916</td>
</tr>
<tr>
<td>ABTC1107</td>
<td>Carlia</td>
<td>jarnoldae</td>
<td>AUS: QLD: Morehead River</td>
<td>-15.024</td>
<td>143.665</td>
</tr>
<tr>
<td>AMJ46155</td>
<td>Carlia</td>
<td>jarnoldae</td>
<td>AUS: QLD: Morehead River</td>
<td>-15.024</td>
<td>143.665</td>
</tr>
<tr>
<td>AMJ62732</td>
<td>Carlia</td>
<td>jarnoldae</td>
<td>AUS: QLD: Mt Abbot</td>
<td>-20.102</td>
<td>147.747</td>
</tr>
<tr>
<td>QMJ62695</td>
<td>Carlia</td>
<td>jarnoldae</td>
<td>AUS: QLD: Mt Aberdeen</td>
<td>-20.202</td>
<td>147.920</td>
</tr>
<tr>
<td>ABTC28872</td>
<td>Carlia</td>
<td>munda</td>
<td>AUS: NT: Bing Bong Station</td>
<td>-16.30</td>
<td>136.87</td>
</tr>
<tr>
<td>Brandy1*</td>
<td>Carlia</td>
<td>rhomboidalis</td>
<td>AUS: QLD: Conway Range</td>
<td>-20.335</td>
<td>153.80</td>
</tr>
<tr>
<td>Brandy2*</td>
<td>Carlia</td>
<td>rhomboidalis</td>
<td>AUS: QLD: Conway Range</td>
<td>-20.335</td>
<td>153.80</td>
</tr>
<tr>
<td>Thornt1*</td>
<td>Carlia</td>
<td>rubrigularis</td>
<td>AUS: QLD: Thornton Uplands</td>
<td>-15.04</td>
<td>145.25</td>
</tr>
<tr>
<td>Thornt2*</td>
<td>Carlia</td>
<td>rubrigularis</td>
<td>AUS: QLD: Thornton Uplands</td>
<td>-15.04</td>
<td>145.25</td>
</tr>
<tr>
<td>ABTC11044</td>
<td>Carlia</td>
<td>tetradoctyla</td>
<td>AUS: Vic.: Euroa</td>
<td>-36.754</td>
<td>145.573</td>
</tr>
<tr>
<td>CCA16072</td>
<td>Lygisaurus</td>
<td>macfarlani</td>
<td>PNG: WP: Daru</td>
<td>-9.07353</td>
<td>132.2118</td>
</tr>
<tr>
<td>CCA16072</td>
<td>Lygisaurus</td>
<td>macfarlani</td>
<td>PNG: WP: Daru</td>
<td>-9.07353</td>
<td>132.2118</td>
</tr>
</tbody>
</table>

* Sequences for *Carlia rhomboidalis* and *C. rubrigularis* were obtained from GenBank, accession numbers DQ349242-5, DQ349342-5, DQ349354-7, DQ349454-7, DQ349466-9, DQ349566-9, DQ349690-3, DQ349790-3, DQ349802-5, DQ349900-3, DQ349912-5, DQ350012-5, DQ350020-1, DQ350070-1.
Table C.2. Loci included in this study, along with primers used in amplification and sequencing, annealing temperature used in PCR, and final aligned locus length. Ta: annealing temperature for PCR.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Primers</th>
<th>Sequence</th>
<th>Ta (°C)</th>
<th>Length (bp)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ND4</td>
<td>ND4.light</td>
<td>CACCTATGACTACCAAAAGCTCATGTAGAAGC</td>
<td>55</td>
<td>707</td>
<td>(Arèvalo et al. 1994)</td>
</tr>
<tr>
<td></td>
<td>Leu3-heavy</td>
<td>GAATTAGCGATTCTTTTRTG</td>
<td></td>
<td></td>
<td>(Stuart-Fox et al. 2002)</td>
</tr>
<tr>
<td>Aldolase</td>
<td>Ald1CR</td>
<td>AAGAAGGATGGAGCTGACT TTGC</td>
<td>60</td>
<td>212</td>
<td>(Dolman &amp; Phillips 2004)</td>
</tr>
<tr>
<td></td>
<td>Ald2CR</td>
<td>GCCATCTGTAACACAACAG CCAA</td>
<td></td>
<td></td>
<td>(Dolman &amp; Phillips 2004)</td>
</tr>
<tr>
<td>Anonymous</td>
<td>Sk13</td>
<td>CTTTACATGCACTATTGAG AC</td>
<td>60</td>
<td>692</td>
<td>(Dolman &amp; Phillips 2004)</td>
</tr>
<tr>
<td></td>
<td>Sk14</td>
<td>ACATGTGCGCTGCTCACAT C</td>
<td></td>
<td></td>
<td>(Dolman &amp; Phillips 2004)</td>
</tr>
<tr>
<td>β-globin</td>
<td>Bglo1CR</td>
<td>GCIAACTGCACCTGYGACAA G</td>
<td>59</td>
<td>804</td>
<td>(Dolman &amp; Phillips 2004)</td>
</tr>
<tr>
<td></td>
<td>Bglo2CR</td>
<td>GCTGCCAAGCGGGGTGTTGA</td>
<td></td>
<td></td>
<td>(Dolman &amp; Phillips 2004)</td>
</tr>
<tr>
<td>GAPD</td>
<td>GAPD-L890</td>
<td>ACCTTTAAATGCGGGTGCTG GCATTGC</td>
<td>63</td>
<td>347</td>
<td>(Friesen et al. 1997)</td>
</tr>
<tr>
<td></td>
<td>GAPD-H950</td>
<td>CATCAAGTCCACAACACGGTGTGCTGA</td>
<td></td>
<td></td>
<td>(Friesen et al. 1997)</td>
</tr>
<tr>
<td>Myosin</td>
<td>MYH2-F</td>
<td>GAACACCAGGCTCATCAACC</td>
<td>60</td>
<td>409</td>
<td>(Lyons et al. 1997)</td>
</tr>
<tr>
<td></td>
<td>MYH2-R</td>
<td>TGGTGTCTCTTGTCTC</td>
<td></td>
<td></td>
<td>(Lyons et al. 1997)</td>
</tr>
<tr>
<td>R35</td>
<td>R35-F</td>
<td>GACTGTGGAYGAYCTGATCAGTGTTGTG</td>
<td>63</td>
<td>650</td>
<td>(Brandley et al. 2011)</td>
</tr>
<tr>
<td></td>
<td>R35-R</td>
<td>GCCAAAATGAGCGAGAARCAGCTCTTGAGC</td>
<td></td>
<td></td>
<td>(Brandley et al. 2011)</td>
</tr>
<tr>
<td>Rhodopsin</td>
<td>Rho3CR</td>
<td>CCTTGCCCTGACACCCCTATGCTG</td>
<td>59</td>
<td>513</td>
<td>(Dolman &amp; Phillips 2004)</td>
</tr>
<tr>
<td></td>
<td>Rho4CR</td>
<td>CAGGAGACCCCTCACCATG</td>
<td></td>
<td></td>
<td>(Dolman &amp; Phillips 2004)</td>
</tr>
<tr>
<td>Selenoprotein T</td>
<td>SelT-F</td>
<td>GTTATYAGCCACCGGTACCCACACATCCG</td>
<td>63</td>
<td>813</td>
<td>(Jackson &amp; Austin 2012)</td>
</tr>
<tr>
<td></td>
<td>SelT-R</td>
<td>GCCTATTAAYACTAGTTGGAAGACTGACAG</td>
<td></td>
<td></td>
<td>(Jackson &amp; Austin 2012)</td>
</tr>
</tbody>
</table>
APPENDIX D
SUPPLEMENTAL MATERIALS FOR CHAPTER 4
Table D.1. Collection localities samples included in this study. Species* refers to the currently recognized species, Species** refers to the species following the results of this study. States and provinces abbreviated as follows. Australia (AUS): NSW: New South Wales, Australia; NT: Northern Territory; QLD: Queensland; Vic.: Victoria. Indonesia (IND): MkP: Maluku Province; NMkP: North Maluku Province; PP: Papua Province; WPP: West Papua Province. Papua New Guinea (PNG): CP: Central Province; ENBP: East New Britain Province; ESP: East Sepik Province; GP: Gulf Province; MBP: Milne Bay Province; MdP: Madang Province; MoP: Morobe Province; NCD: National Capital District; NIP: New Ireland Province; OrP: Oro Province; SHP: Southern Highlands Province; SP: Sandaun Province; WP: Western Province.

<table>
<thead>
<tr>
<th>Cat.No.</th>
<th>Genus</th>
<th>Species*</th>
<th>Species**</th>
<th>Locality (Locality #)</th>
<th>Latitude</th>
<th>Longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA14481</td>
<td>Carla</td>
<td>sp. Lakekamu</td>
<td>sp. Lakekamu</td>
<td>PNG: GP: Lakekamu (45)</td>
<td>-7.735</td>
<td>146.496</td>
</tr>
<tr>
<td>AA14492</td>
<td>Carla</td>
<td>sp. Lakekamu</td>
<td>sp. Lakekamu</td>
<td>PNG: GP: Lakekamu (45)</td>
<td>-7.735</td>
<td>146.496</td>
</tr>
<tr>
<td>AA14518</td>
<td>Carla</td>
<td>sp. Lakekamu</td>
<td>sp. Lakekamu</td>
<td>PNG: GP: Lakekamu (45)</td>
<td>-7.735</td>
<td>146.496</td>
</tr>
<tr>
<td>AA14564</td>
<td>Carla</td>
<td>caesius</td>
<td>caesius</td>
<td>IND: PP: Timika (54)</td>
<td>-4.394</td>
<td>136.933</td>
</tr>
<tr>
<td>AA14842</td>
<td>Carla</td>
<td>caesius</td>
<td>caesius</td>
<td>IND: PP: Timika (54)</td>
<td>-4.394</td>
<td>136.933</td>
</tr>
<tr>
<td>AA14880</td>
<td>Carla</td>
<td>caesius</td>
<td>caesius</td>
<td>IND: PP: Timika (54)</td>
<td>-4.394</td>
<td>136.933</td>
</tr>
<tr>
<td>AA14931</td>
<td>Carla</td>
<td>caesius</td>
<td>caesius</td>
<td>IND: PP: Timika (54)</td>
<td>-4.394</td>
<td>136.933</td>
</tr>
<tr>
<td>AA15014</td>
<td>Carla</td>
<td>caesius</td>
<td>caesius</td>
<td>IND: PP: Timika (54)</td>
<td>-4.394</td>
<td>136.933</td>
</tr>
<tr>
<td>AA15893</td>
<td>Carla</td>
<td>caesius</td>
<td>caesius</td>
<td>IND: PP: Timika (54)</td>
<td>-4.394</td>
<td>136.933</td>
</tr>
<tr>
<td>AA16175</td>
<td>Carla</td>
<td>bomberai</td>
<td>leucotaenia</td>
<td>IND: WPP: Saengga (39)</td>
<td>-2.458</td>
<td>133.111</td>
</tr>
<tr>
<td>AA16178</td>
<td>Carla</td>
<td>bomberai</td>
<td>leucotaenia</td>
<td>IND: WPP: Saengga (39)</td>
<td>-2.458</td>
<td>133.111</td>
</tr>
<tr>
<td>AA16179</td>
<td>Carla</td>
<td>bomberai</td>
<td>leucotaenia</td>
<td>IND: WPP: Saengga (39)</td>
<td>-2.458</td>
<td>133.111</td>
</tr>
<tr>
<td>AA16253</td>
<td>Carla</td>
<td>bomberai</td>
<td>leucotaenia</td>
<td>IND: WPP: Saengga (39)</td>
<td>-2.443</td>
<td>133.14</td>
</tr>
<tr>
<td>AA16285</td>
<td>Carla</td>
<td>bomberai</td>
<td>leucotaenia</td>
<td>IND: WPP: Saengga (39)</td>
<td>-2.438</td>
<td>133.135</td>
</tr>
<tr>
<td>AA16398</td>
<td>Carla</td>
<td>eothen</td>
<td>sp. Woodlark</td>
<td>PNG: MBP: Woodlark Is., Guasopa (4)</td>
<td>-9.223</td>
<td>152.943</td>
</tr>
<tr>
<td>AA16399</td>
<td>Carla</td>
<td>eothen</td>
<td>sp. Woodlark</td>
<td>PNG: MBP: Woodlark Is., Guasopa (4)</td>
<td>-9.223</td>
<td>152.943</td>
</tr>
<tr>
<td>AA16453</td>
<td>Carla</td>
<td>eothen</td>
<td>sp. Woodlark</td>
<td>PNG: MBP: Woodlark Is., Guasopa (4)</td>
<td>-9.223</td>
<td>152.943</td>
</tr>
<tr>
<td>AA16494</td>
<td>Carla</td>
<td>eothen</td>
<td>sp. Woodlark</td>
<td>PNG: MBP: Woodlark Is., Guasopa (4)</td>
<td>-9.223</td>
<td>152.943</td>
</tr>
<tr>
<td>AA16540</td>
<td>Carla</td>
<td>eothen</td>
<td>sp. Woodlark</td>
<td>PNG: MBP: Woodlark Is., Guasopa (4)</td>
<td>-9.223</td>
<td>152.943</td>
</tr>
<tr>
<td>ABTC10993</td>
<td>Carla</td>
<td>longipes</td>
<td>longipes</td>
<td>AUS: QLD: near Cairns (7)</td>
<td>-16.7</td>
<td>145.6</td>
</tr>
<tr>
<td>ABTC10995</td>
<td>Carla</td>
<td>longipes</td>
<td>longipes</td>
<td>AUS: QLD: near Cairns (7)</td>
<td>-16.7</td>
<td>145.6</td>
</tr>
<tr>
<td>ABTC24813</td>
<td>Carla</td>
<td>longipes</td>
<td>longipes</td>
<td>AUS: QLD: Cape Flattery (8)</td>
<td>-14.918</td>
<td>145.218</td>
</tr>
<tr>
<td>ABTC28524</td>
<td>Carla</td>
<td>sexdentata</td>
<td>sexdentata NT</td>
<td>AUS: NT: Arnhem Land (41)</td>
<td>-12.35</td>
<td>134.683</td>
</tr>
<tr>
<td>ABTC32144</td>
<td>Carla</td>
<td>longipes</td>
<td>longipes</td>
<td>AUS: QLD: Mt Cook via Cooktown (8)</td>
<td>-15.49</td>
<td>145.265</td>
</tr>
<tr>
<td>ABTC44480</td>
<td>Carla</td>
<td>sp. Waro</td>
<td>sp. So. High.</td>
<td>PNG: SHP: Waro (56)</td>
<td>-6.533</td>
<td>143.183</td>
</tr>
<tr>
<td>ABTC44481</td>
<td>Carla</td>
<td>sp. Waro</td>
<td>sp. So. High.</td>
<td>PNG: SHP: Waro (56)</td>
<td>-6.533</td>
<td>143.183</td>
</tr>
<tr>
<td>ABTC44496</td>
<td>Carla</td>
<td>sp. Waro</td>
<td>sp. So. High.</td>
<td>PNG: SHP: Waro (56)</td>
<td>-6.533</td>
<td>143.183</td>
</tr>
<tr>
<td>ABTC44725</td>
<td>Carla</td>
<td>sp. Waro</td>
<td>sp. So. High.</td>
<td>PNG: SHP: Waro (56)</td>
<td>-6.533</td>
<td>143.183</td>
</tr>
<tr>
<td>ABTC44731</td>
<td>Carla</td>
<td>sp. Waro</td>
<td>sp. So. High.</td>
<td>PNG: SHP: Waro (56)</td>
<td>-6.533</td>
<td>143.183</td>
</tr>
<tr>
<td>ABTC44877</td>
<td>Carla</td>
<td>luctuosa</td>
<td>luctuosa</td>
<td>PNG: NCD: Port Moresby (13)</td>
<td>-9.407</td>
<td>147.166</td>
</tr>
<tr>
<td>ABTC44879</td>
<td>Carla</td>
<td>luctuosa</td>
<td>luctuosa</td>
<td>PNG: NCD: Port Moresby (13)</td>
<td>-9.407</td>
<td>147.166</td>
</tr>
<tr>
<td>Cat.No.</td>
<td>Genus</td>
<td>Species*</td>
<td>Species**</td>
<td>Locality</td>
<td>Latitude</td>
<td>Longitude</td>
</tr>
<tr>
<td>--------</td>
<td>-------</td>
<td>----------</td>
<td>-----------</td>
<td>----------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>ABTC44880</td>
<td>Carlia</td>
<td>luctuosa</td>
<td>luctuosa</td>
<td>PNG: NCD: Port Moresby (13)</td>
<td>-9.407</td>
<td>147.166</td>
</tr>
<tr>
<td>ABTC45393</td>
<td>Carlia</td>
<td>sp. Namosado</td>
<td>sp. So. High.</td>
<td>PNG: SHP: Namosado (49)</td>
<td>-6.25</td>
<td>142.783</td>
</tr>
<tr>
<td>ABTC45394</td>
<td>Carlia</td>
<td>sp. Namosado</td>
<td>sp. So. High.</td>
<td>PNG: SHP: Namosado (49)</td>
<td>-6.25</td>
<td>142.783</td>
</tr>
<tr>
<td>ABTC48165</td>
<td>Carlia</td>
<td>mysii</td>
<td>pulla</td>
<td>PNG: MdP: Nagada Harbour (21)</td>
<td>-5.159</td>
<td>145.788</td>
</tr>
<tr>
<td>ABTC48187</td>
<td>Carlia</td>
<td>mysii</td>
<td>pulla</td>
<td>PNG: MdP: Nagada Harbour (21)</td>
<td>-5.159</td>
<td>145.788</td>
</tr>
<tr>
<td>ABTC48701</td>
<td>Carlia</td>
<td>mysii</td>
<td>pulla cf.</td>
<td>PNG: MdP: KarKar Is. (20)</td>
<td>-4.64</td>
<td>146</td>
</tr>
<tr>
<td>ABTC49794</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. D’Entrecasteaux</td>
<td>PNG: MBP: Normanby Is., Guleguleu (2)</td>
<td>-9.989</td>
<td>151.284</td>
</tr>
<tr>
<td>ABTC49796</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. D’Entrecasteaux</td>
<td>PNG: MBP: Normanby Is., Guleguleu (2)</td>
<td>-9.989</td>
<td>151.284</td>
</tr>
<tr>
<td>ABTC49858</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. So. MB</td>
<td>PNG: MBP: Sideia Is. (52)</td>
<td>-10.581</td>
<td>150.852</td>
</tr>
<tr>
<td>ABTC49859</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. So. MB</td>
<td>PNG: MBP: Sideia Is. (52)</td>
<td>-10.581</td>
<td>150.852</td>
</tr>
<tr>
<td>ABTC49867</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. So. MB</td>
<td>PNG: MBP: Sideia Is. (52)</td>
<td>-10.581</td>
<td>150.852</td>
</tr>
<tr>
<td>ABTC49870</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. So. MB</td>
<td>PNG: MBP: Sideia Is. (52)</td>
<td>-10.581</td>
<td>150.852</td>
</tr>
<tr>
<td>ABTC49872</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. So. MB</td>
<td>PNG: MBP: Sideia Is. (52)</td>
<td>-10.581</td>
<td>150.852</td>
</tr>
<tr>
<td>ABTC49961</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. D’Entrecasteaux</td>
<td>PNG: MBP: Normanby Is., Guleguleu (2)</td>
<td>-9.989</td>
<td>151.284</td>
</tr>
<tr>
<td>ABTC49967</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. D’Entrecasteaux</td>
<td>PNG: MBP: Normanby Is., Guleguleu (2)</td>
<td>-9.989</td>
<td>151.284</td>
</tr>
<tr>
<td>ABTC51118</td>
<td>Carlia</td>
<td>longipes</td>
<td>longipes</td>
<td>AUS: QLD: Wangetti Beach (7)</td>
<td>-16.666</td>
<td>145.571</td>
</tr>
<tr>
<td>ABTC79201</td>
<td>Carlia</td>
<td>aramia</td>
<td>sp. Kiunga</td>
<td>PNG: WP: Wegamu (43)</td>
<td>-8.433</td>
<td>141.112</td>
</tr>
<tr>
<td>ABTC79202</td>
<td>Carlia</td>
<td>aramia</td>
<td>sp. Kiunga</td>
<td>PNG: WP: Wegamu (43)</td>
<td>-8.433</td>
<td>141.112</td>
</tr>
<tr>
<td>ABTC98459</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. Misima</td>
<td>PNG: MBP: Misima Is., North Dump (48)</td>
<td>-10.6645</td>
<td>152.7851</td>
</tr>
<tr>
<td>ABTC98931</td>
<td>Carlia</td>
<td>sp. Nogoli</td>
<td>sp. So. High.</td>
<td>PNG: SHP: Nogoli (50)</td>
<td>-6.302</td>
<td>143.943</td>
</tr>
<tr>
<td>ABTC98932</td>
<td>Carlia</td>
<td>sp. Nogoli</td>
<td>sp. So. High.</td>
<td>PNG: SHP: Nogoli (50)</td>
<td>-6.302</td>
<td>143.943</td>
</tr>
<tr>
<td>ABTC102373</td>
<td>Carlia</td>
<td>quinquecarinata</td>
<td>quinquecarinata</td>
<td>AUS: QLD: Erub Island (51)</td>
<td>-9.583</td>
<td>143.767</td>
</tr>
<tr>
<td>ABTC102374</td>
<td>Carlia</td>
<td>quinquecarinata</td>
<td>quinquecarinata</td>
<td>AUS: QLD: Erub Island (51)</td>
<td>-9.583</td>
<td>143.767</td>
</tr>
<tr>
<td>ABTC102375</td>
<td>Carlia</td>
<td>sexdentata</td>
<td>sexdentata Qld</td>
<td>AUS: QLD: Yam Island (42)</td>
<td>-9.9</td>
<td>142.767</td>
</tr>
<tr>
<td>ABTC102376</td>
<td>Carlia</td>
<td>sexdentata</td>
<td>sexdentata Qld</td>
<td>AUS: QLD: Yam Island (42)</td>
<td>-9.9</td>
<td>142.767</td>
</tr>
<tr>
<td>ABTC102377</td>
<td>Carlia</td>
<td>sexdentata</td>
<td>sexdentata Qld</td>
<td>AUS: QLD: Yam Island (42)</td>
<td>-9.9</td>
<td>142.767</td>
</tr>
<tr>
<td>ABTC103153</td>
<td>Carlia</td>
<td>sexdentata</td>
<td>sexdentata NT</td>
<td>AUS: NT: Maningrida (41)</td>
<td>-12.048</td>
<td>134.222</td>
</tr>
<tr>
<td>ABTC103154</td>
<td>Carlia</td>
<td>sexdentata</td>
<td>sexdentata NT</td>
<td>AUS: NT: Maningrida (41)</td>
<td>-12.048</td>
<td>134.222</td>
</tr>
<tr>
<td>ALS107</td>
<td>Carlia</td>
<td>beccarii</td>
<td>beccarii</td>
<td>IND: MKP: Kei Is. (6)</td>
<td>-5.647</td>
<td>132.638</td>
</tr>
<tr>
<td>ALS140</td>
<td>Carlia</td>
<td>beccarii</td>
<td>beccarii</td>
<td>IND: MKP: Kei Is. (6)</td>
<td>-5.647</td>
<td>132.638</td>
</tr>
<tr>
<td>ALS249</td>
<td>Carlia</td>
<td>diguliensis</td>
<td>diguliensis</td>
<td>IND: MKP: Aru Is. (5)</td>
<td>-6.813</td>
<td>134.378</td>
</tr>
<tr>
<td>ALS250</td>
<td>Carlia</td>
<td>diguliensis</td>
<td>diguliensis</td>
<td>IND: MKP: Aru Is. (5)</td>
<td>-6.503</td>
<td>134.391</td>
</tr>
<tr>
<td>ALS262</td>
<td>Carlia</td>
<td>diguliensis</td>
<td>diguliensis</td>
<td>IND: MKP: Aru Is. (5)</td>
<td>-6.503</td>
<td>134.391</td>
</tr>
<tr>
<td>ALS268</td>
<td>Carlia</td>
<td>diguliensis</td>
<td>diguliensis</td>
<td>IND: MKP: Aru Is. (5)</td>
<td>-6.503</td>
<td>134.391</td>
</tr>
<tr>
<td>ALS270</td>
<td>Carlia</td>
<td>diguliensis</td>
<td>diguliensis</td>
<td>IND: MKP: Aru Is. (5)</td>
<td>-6.813</td>
<td>134.378</td>
</tr>
<tr>
<td>ALS395</td>
<td>Carlia</td>
<td>beccarii</td>
<td>beccarii</td>
<td>IND: MKP: Kei Is. (6)</td>
<td>-5.647</td>
<td>132.638</td>
</tr>
<tr>
<td>Cat.No.</td>
<td>Genus</td>
<td>Species*</td>
<td>Species**</td>
<td>Locality</td>
<td>Latitude</td>
<td>Longitude</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
<td>----------</td>
<td>-----------</td>
<td>-------------------------------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>ALS440</td>
<td>Carlia</td>
<td>beccarii</td>
<td>beccarii</td>
<td>IND: MkP: Kei Is. (6)</td>
<td>-5.647</td>
<td>132.638</td>
</tr>
<tr>
<td>ALS523</td>
<td>Carlia</td>
<td>leucotaenia</td>
<td>leucotaenia</td>
<td>IND: MkP: Seram Is. (40)</td>
<td>-5.647</td>
<td>132.638</td>
</tr>
<tr>
<td>ALS573</td>
<td>Carlia</td>
<td>leucotaenia</td>
<td>leucotaenia</td>
<td>IND: MkP: Seram Is. (40)</td>
<td>-3.12</td>
<td>130.513</td>
</tr>
<tr>
<td>ALS612</td>
<td>Carlia</td>
<td>leucotaenia</td>
<td>leucotaenia</td>
<td>IND: MkP: Seram Is. (40)</td>
<td>-3.12</td>
<td>130.513</td>
</tr>
<tr>
<td>ALS619</td>
<td>Carlia</td>
<td>leucotaenia</td>
<td>leucotaenia</td>
<td>IND: MkP: Seram Is. (40)</td>
<td>-3.12</td>
<td>130.513</td>
</tr>
<tr>
<td>BJE1146</td>
<td>Carlia</td>
<td>tutela</td>
<td>leucotaenia</td>
<td>IND: NMkP: Halmahera (28)</td>
<td>1.003</td>
<td>128.002</td>
</tr>
<tr>
<td>BJE1156</td>
<td>Carlia</td>
<td>tutela</td>
<td>leucotaenia</td>
<td>IND: NMkP: Halmahera (28)</td>
<td>1.003</td>
<td>128.002</td>
</tr>
<tr>
<td>CCA0664</td>
<td>Carlia</td>
<td>mysi</td>
<td>leucotaenia</td>
<td>PNG: ENBP: Rabaul (36)</td>
<td>-4.196</td>
<td>152.173</td>
</tr>
<tr>
<td>CCA0665</td>
<td>Carlia</td>
<td>mysi</td>
<td>leucotaenia</td>
<td>PNG: ENBP: Rabaul (36)</td>
<td>-4.196</td>
<td>152.173</td>
</tr>
<tr>
<td>CCA0682</td>
<td>Carlia</td>
<td>mysi</td>
<td>leucotaenia</td>
<td>PNG: NIP: Lelet (37)</td>
<td>-3.43</td>
<td>151.998</td>
</tr>
<tr>
<td>CCA0703</td>
<td>Carlia</td>
<td>mysi</td>
<td>leucotaenia</td>
<td>PNG: NIP: Lelet (37)</td>
<td>-3.43</td>
<td>151.998</td>
</tr>
<tr>
<td>CCA0704</td>
<td>Carlia</td>
<td>mysi</td>
<td>leucotaenia</td>
<td>PNG: NIP: Lelet (37)</td>
<td>-3.43</td>
<td>151.998</td>
</tr>
<tr>
<td>CCA0752</td>
<td>Carlia</td>
<td>mysi</td>
<td>leucotaenia</td>
<td>PNG: MoP: Kavieng (38)</td>
<td>-2.566</td>
<td>150.799</td>
</tr>
<tr>
<td>CCA0756</td>
<td>Carlia</td>
<td>mysi</td>
<td>leucotaenia</td>
<td>PNG: MoP: Kavieng (38)</td>
<td>-2.566</td>
<td>150.799</td>
</tr>
<tr>
<td>CCA0780</td>
<td>Carlia</td>
<td>mysi</td>
<td>mysi</td>
<td>PNG: MoP: Huon Penin., Oligadu (18)</td>
<td>-6.709</td>
<td>147.573</td>
</tr>
<tr>
<td>CCA0817</td>
<td>Carlia</td>
<td>mysi</td>
<td>mysi</td>
<td>PNG: MoP: Huon Penin., Oligadu (18)</td>
<td>-6.709</td>
<td>147.573</td>
</tr>
<tr>
<td>CCA0830</td>
<td>Carlia</td>
<td>mysi</td>
<td>mysi</td>
<td>PNG: MoP: Huon Penin., Oligadu (18)</td>
<td>-6.709</td>
<td>147.573</td>
</tr>
<tr>
<td>CCA0853</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. D'Entrecasteaux</td>
<td>PNG: MBP: Fergusson Is., Salamo (1)</td>
<td>-9.665</td>
<td>150.792</td>
</tr>
<tr>
<td>CCA0854</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. D'Entrecasteaux</td>
<td>PNG: MBP: Fergusson Is., Salamo (1)</td>
<td>-9.665</td>
<td>150.792</td>
</tr>
<tr>
<td>CCA0878</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. D'Entrecasteaux</td>
<td>PNG: MBP: Fergusson Is., Salamo (1)</td>
<td>-9.665</td>
<td>150.792</td>
</tr>
<tr>
<td>CCA1520</td>
<td>Carlia</td>
<td>ailanpalai</td>
<td>leucotaenia</td>
<td>PNG: MnP: Manus Is., Kawaliap (31)</td>
<td>-2.11</td>
<td>147.068</td>
</tr>
<tr>
<td>CCA1521</td>
<td>Carlia</td>
<td>ailanpalai</td>
<td>leucotaenia</td>
<td>PNG: MnP: Manus Is., Kawaliap (31)</td>
<td>-2.11</td>
<td>147.068</td>
</tr>
<tr>
<td>CCA1522</td>
<td>Carlia</td>
<td>ailanpalai</td>
<td>leucotaenia</td>
<td>PNG: MnP: Manus Is., Kawaliap (31)</td>
<td>-2.11</td>
<td>147.068</td>
</tr>
<tr>
<td>CCA1545</td>
<td>Carlia</td>
<td>ailanpalai</td>
<td>leucotaenia</td>
<td>PNG: MnP: Los Negros Is., Loniu (32)</td>
<td>-2.07</td>
<td>147.341</td>
</tr>
<tr>
<td>CCA1546</td>
<td>Carlia</td>
<td>ailanpalai</td>
<td>leucotaenia</td>
<td>PNG: MnP: Los Negros Is., Loniu (32)</td>
<td>-2.07</td>
<td>147.341</td>
</tr>
<tr>
<td>CCA2033</td>
<td>Carlia</td>
<td>ailanpalai</td>
<td>leucotaenia</td>
<td>PNG: MnP: Los Negros Is., Rio Rio (34)</td>
<td>-2.049</td>
<td>147.419</td>
</tr>
<tr>
<td>CCA2039</td>
<td>Carlia</td>
<td>ailanpalai</td>
<td>leucotaenia</td>
<td>PNG: MnP: Los Negros Is., Rio Rio (34)</td>
<td>-2.049</td>
<td>147.419</td>
</tr>
<tr>
<td>CCA2040</td>
<td>Carlia</td>
<td>ailanpalai</td>
<td>leucotaenia</td>
<td>PNG: MnP: Los Negros Is., Rio Rio (34)</td>
<td>-2.049</td>
<td>147.419</td>
</tr>
<tr>
<td>CCA2046</td>
<td>Carlia</td>
<td>ailanpalai</td>
<td>leucotaenia</td>
<td>PNG: MnP: Los Negros Is., Peyon (32)</td>
<td>-2.033</td>
<td>147.434</td>
</tr>
<tr>
<td>CCA2157</td>
<td>Carlia</td>
<td>mysi</td>
<td>sp. Oro</td>
<td>PNG: OrP: Koreaf (22)</td>
<td>-9.338</td>
<td>149.14</td>
</tr>
<tr>
<td>CCA2168</td>
<td>Carlia</td>
<td>mysi</td>
<td>sp. Oro</td>
<td>PNG: OrP: Koreaf (22)</td>
<td>-9.338</td>
<td>149.14</td>
</tr>
<tr>
<td>CCA2190</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. No. MB</td>
<td>PNG: MBP: Lapoa #2 (16)</td>
<td>-10.259</td>
<td>150.574</td>
</tr>
<tr>
<td>CCA2192</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. No. MB</td>
<td>PNG: MBP: Lapoa #2 (16)</td>
<td>-10.259</td>
<td>150.574</td>
</tr>
<tr>
<td>CCA2245</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. So. MB</td>
<td>PNG: MBP: Saga Aho (47)</td>
<td>-10.544</td>
<td>150.115</td>
</tr>
<tr>
<td>CCA2246</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. So. MB</td>
<td>PNG: MBP: Saga Aho (47)</td>
<td>-10.544</td>
<td>150.115</td>
</tr>
<tr>
<td>CCA2252</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. So. MB</td>
<td>PNG: MBP: Saga Aho (47)</td>
<td>-10.544</td>
<td>150.115</td>
</tr>
<tr>
<td>Cat.No.</td>
<td>Genus</td>
<td>Species*</td>
<td>Species**</td>
<td>Locality</td>
<td>Latitude</td>
<td>Longitude</td>
</tr>
<tr>
<td>---------</td>
<td>---------</td>
<td>----------</td>
<td>-----------</td>
<td>--------------------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>CCA2274</td>
<td>Carlia</td>
<td>mysie</td>
<td>sp. Lababia</td>
<td>PNG: MoP: Lababia (19)</td>
<td>-7.301</td>
<td>147.134</td>
</tr>
<tr>
<td>CCA2275</td>
<td>Carlia</td>
<td>mysie</td>
<td>sp. Lababia</td>
<td>PNG: MoP: Lababia (19)</td>
<td>-7.301</td>
<td>147.134</td>
</tr>
<tr>
<td>CCA2282</td>
<td>Carlia</td>
<td>mysie</td>
<td>sp. Lababia</td>
<td>PNG: MoP: Lababia (19)</td>
<td>-7.301</td>
<td>147.134</td>
</tr>
<tr>
<td>CCA2289</td>
<td>Carlia</td>
<td>mysie</td>
<td>sp. Lababia</td>
<td>PNG: MoP: Lababia (19)</td>
<td>-7.301</td>
<td>147.134</td>
</tr>
<tr>
<td>CCA2307</td>
<td>Carlia</td>
<td>mysie</td>
<td>sp. Oro</td>
<td>PNG: OrP: Mt Victory (22)</td>
<td>-9.239</td>
<td>149.054</td>
</tr>
<tr>
<td>CCA2309</td>
<td>Carlia</td>
<td>mysie</td>
<td>sp. Oro</td>
<td>PNG: OrP: Mt Victory (22)</td>
<td>-9.239</td>
<td>149.054</td>
</tr>
<tr>
<td>CCA2323</td>
<td>Carlia</td>
<td>mysie</td>
<td>sp. Oro</td>
<td>PNG: OrP: Wanigela (22)</td>
<td>-9.339</td>
<td>149.159</td>
</tr>
<tr>
<td>CCA2512</td>
<td>Carlia</td>
<td>pulla</td>
<td>pulla</td>
<td>PNG: SP: Waromo (24)</td>
<td>-2.666</td>
<td>141.239</td>
</tr>
<tr>
<td>CCA2527</td>
<td>Carlia</td>
<td>pulla</td>
<td>pulla</td>
<td>PNG: SP: Waromo (24)</td>
<td>-2.666</td>
<td>141.239</td>
</tr>
<tr>
<td>CCA2528</td>
<td>Carlia</td>
<td>pulla</td>
<td>pulla</td>
<td>PNG: SP: Waromo (24)</td>
<td>-2.666</td>
<td>141.239</td>
</tr>
<tr>
<td>CCA2556</td>
<td>Carlia</td>
<td>luctuosa</td>
<td>luctuosa</td>
<td>PNG: CP: Loloipa (11)</td>
<td>-8.415</td>
<td>146.959</td>
</tr>
<tr>
<td>CCA2557</td>
<td>Carlia</td>
<td>luctuosa</td>
<td>luctuosa</td>
<td>PNG: CP: Loloipa (11)</td>
<td>-8.415</td>
<td>146.959</td>
</tr>
<tr>
<td>CCA3541</td>
<td>Carlia</td>
<td>pulla</td>
<td>pulla</td>
<td>PNG: SP: Vanimo (23)</td>
<td>-2.684</td>
<td>141.305</td>
</tr>
<tr>
<td>CCA3726</td>
<td>Carlia</td>
<td>pulla</td>
<td>pulla</td>
<td>PNG: SP: Vanimo (23)</td>
<td>-2.684</td>
<td>141.305</td>
</tr>
<tr>
<td>CCA3727</td>
<td>Carlia</td>
<td>pulla</td>
<td>pulla</td>
<td>PNG: SP: Vanimo (23)</td>
<td>-2.684</td>
<td>141.305</td>
</tr>
<tr>
<td>CCA3729</td>
<td>Carlia</td>
<td>pulla</td>
<td>pulla</td>
<td>PNG: SP: Vanimo (23)</td>
<td>-2.684</td>
<td>141.305</td>
</tr>
<tr>
<td>CCA3950</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. No. MB</td>
<td>PNG: MBP: Halowia (15)</td>
<td>-10.336</td>
<td>150.573</td>
</tr>
<tr>
<td>CCA3951</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. No. MB</td>
<td>PNG: MBP: Halowia (15)</td>
<td>-10.336</td>
<td>150.573</td>
</tr>
<tr>
<td>CCA4071</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. No. MB</td>
<td>PNG: MBP: Halowia (15)</td>
<td>-10.336</td>
<td>150.573</td>
</tr>
<tr>
<td>CCA4072</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. No. MB</td>
<td>PNG: MBP: Halowia (15)</td>
<td>-10.336</td>
<td>150.573</td>
</tr>
<tr>
<td>CCA4073</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. No. MB</td>
<td>PNG: MBP: Halowia (15)</td>
<td>-10.336</td>
<td>150.573</td>
</tr>
<tr>
<td>CCA4186</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. No. MB</td>
<td>PNG: MBP: Wamawamana (17)</td>
<td>-10.292</td>
<td>150.354</td>
</tr>
<tr>
<td>CCA4187</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. No. MB</td>
<td>PNG: MBP: Wamawamana (17)</td>
<td>-10.292</td>
<td>150.354</td>
</tr>
<tr>
<td>CCA4188</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. No. MB</td>
<td>PNG: MBP: Wamawamana (17)</td>
<td>-10.292</td>
<td>150.354</td>
</tr>
<tr>
<td>CCA4191</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. No. MB</td>
<td>PNG: MBP: Wamawamana (17)</td>
<td>-10.292</td>
<td>150.354</td>
</tr>
<tr>
<td>CCA4192</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. No. MB</td>
<td>PNG: MBP: Wamawamana (17)</td>
<td>-10.292</td>
<td>150.354</td>
</tr>
<tr>
<td>CCA4223</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. No. MB</td>
<td>PNG: MBP: East Cape (14)</td>
<td>-10.23</td>
<td>150.875</td>
</tr>
<tr>
<td>CCA4224</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. No. MB</td>
<td>PNG: MBP: East Cape (14)</td>
<td>-10.23</td>
<td>150.875</td>
</tr>
<tr>
<td>CCA4304</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. So. MB</td>
<td>PNG: MBP: Oil Palm Plantation (46)</td>
<td>-10.399</td>
<td>150.092</td>
</tr>
<tr>
<td>CCA4305</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. So. MB</td>
<td>PNG: MBP: Oil Palm Plantation (46)</td>
<td>-10.399</td>
<td>150.092</td>
</tr>
<tr>
<td>CCA4306</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. So. MB</td>
<td>PNG: MBP: Oil Palm Plantation (46)</td>
<td>-10.399</td>
<td>150.092</td>
</tr>
<tr>
<td>CCA5535</td>
<td>Carlia</td>
<td>sp. Amau</td>
<td>sp. Amau</td>
<td>PNG: CP: Amau (9)</td>
<td>-10.037</td>
<td>148.565</td>
</tr>
<tr>
<td>CCA5538</td>
<td>Carlia</td>
<td>sp. Amau</td>
<td>sp. Amau</td>
<td>PNG: CP: Amau (9)</td>
<td>-10.037</td>
<td>148.565</td>
</tr>
<tr>
<td>CCA5539</td>
<td>Carlia</td>
<td>sp. Amau</td>
<td>sp. Amau</td>
<td>PNG: CP: Amau (9)</td>
<td>-10.037</td>
<td>148.565</td>
</tr>
<tr>
<td>CCA5583</td>
<td>Carlia</td>
<td>sp. Amau</td>
<td>sp. Amau</td>
<td>PNG: CP: Amau (9)</td>
<td>-10.037</td>
<td>148.565</td>
</tr>
<tr>
<td>CCA5891</td>
<td>Carlia</td>
<td>aenigma</td>
<td>aenigma</td>
<td>PNG: GP: Wabo (55)</td>
<td>-6.98</td>
<td>145.069</td>
</tr>
<tr>
<td>Cat.No.</td>
<td>Genus</td>
<td>Species*</td>
<td>Species**</td>
<td>Locality</td>
<td>Latitude</td>
<td>Longitude</td>
</tr>
<tr>
<td>---------</td>
<td>---------</td>
<td>----------</td>
<td>-----------</td>
<td>-----------------------------------------------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>CCA5979</td>
<td>Carlia</td>
<td>aenigma</td>
<td>aenigma</td>
<td>PNG: GP: Wabo (55)</td>
<td>-6.98</td>
<td>145.069</td>
</tr>
<tr>
<td>CCA15161</td>
<td>Carlia</td>
<td>aenigma</td>
<td>aenigma</td>
<td>PNG: GP: Wabo (55)</td>
<td>-6.98</td>
<td>145.069</td>
</tr>
<tr>
<td>CCA15168</td>
<td>Carlia</td>
<td>aenigma</td>
<td>aenigma</td>
<td>PNG: GP: Wabo (55)</td>
<td>-6.98</td>
<td>145.069</td>
</tr>
<tr>
<td>CCA15382</td>
<td>Carlia</td>
<td>eothen</td>
<td>eothen</td>
<td>PNG: MBP: Trobiand Is., Kiriwina (3)</td>
<td>-8.542</td>
<td>151.079</td>
</tr>
<tr>
<td>CCA15422</td>
<td>Carlia</td>
<td>eothen</td>
<td>eothen</td>
<td>PNG: MBP: Trobiand Is., Kiriwina (3)</td>
<td>-8.542</td>
<td>151.079</td>
</tr>
<tr>
<td>CCA15425</td>
<td>Carlia</td>
<td>eothen</td>
<td>eothen</td>
<td>PNG: MBP: Trobiand Is., Kiriwina (3)</td>
<td>-8.542</td>
<td>151.079</td>
</tr>
<tr>
<td>CCA16068</td>
<td>Carlia</td>
<td>aramia</td>
<td>sp. Daru</td>
<td>PNG: WP: Daru (27)</td>
<td>-9.074</td>
<td>143.211</td>
</tr>
<tr>
<td>CCA16069</td>
<td>Carlia</td>
<td>aramia</td>
<td>sp. Daru</td>
<td>PNG: WP: Daru (27)</td>
<td>-9.074</td>
<td>143.211</td>
</tr>
<tr>
<td>CCA16104</td>
<td>Carlia</td>
<td>aramia</td>
<td>sp. Daru</td>
<td>PNG: WP: Daru (27)</td>
<td>-9.074</td>
<td>143.211</td>
</tr>
<tr>
<td>CCA16170</td>
<td>Carlia</td>
<td>aramia</td>
<td>sp. Daru</td>
<td>PNG: WP: Daru (27)</td>
<td>-9.074</td>
<td>143.211</td>
</tr>
<tr>
<td>CCA16191</td>
<td>Carlia</td>
<td>aramia</td>
<td>sp. Daru</td>
<td>PNG: WP: Daru (27)</td>
<td>-9.074</td>
<td>143.211</td>
</tr>
<tr>
<td>CCA16214</td>
<td>Carlia</td>
<td>aramia</td>
<td>sp. Kiunga</td>
<td>PNG: WP: Kiunga (29)</td>
<td>-6.119</td>
<td>141.302</td>
</tr>
<tr>
<td>CCA16215</td>
<td>Carlia</td>
<td>aramia</td>
<td>sp. Kiunga</td>
<td>PNG: WP: Kiunga (29)</td>
<td>-6.119</td>
<td>141.302</td>
</tr>
<tr>
<td>CCA16220</td>
<td>Carlia</td>
<td>aramia</td>
<td>sp. Kiunga</td>
<td>PNG: WP: Kiunga (29)</td>
<td>-6.119</td>
<td>141.302</td>
</tr>
<tr>
<td>CCA16222</td>
<td>Carlia</td>
<td>aramia</td>
<td>sp. Kiunga</td>
<td>PNG: WP: Kiunga (29)</td>
<td>-6.119</td>
<td>141.302</td>
</tr>
<tr>
<td>CCA16223</td>
<td>Carlia</td>
<td>aramia</td>
<td>sp. Kiunga</td>
<td>PNG: WP: Kiunga (29)</td>
<td>-6.119</td>
<td>141.302</td>
</tr>
<tr>
<td>CCA16236</td>
<td>Carlia</td>
<td>aramia</td>
<td>sp. So. High.</td>
<td>PNG: WP: Kwatu (44)</td>
<td>-6.075</td>
<td>141.523</td>
</tr>
<tr>
<td>CCA16272</td>
<td>Carlia</td>
<td>aramia</td>
<td>sp. So. High.</td>
<td>PNG: WP: Kwatu (44)</td>
<td>-6.075</td>
<td>141.523</td>
</tr>
<tr>
<td>CCA16273</td>
<td>Carlia</td>
<td>aramia</td>
<td>sp. So. High.</td>
<td>PNG: WP: Kwatu (44)</td>
<td>-6.075</td>
<td>141.523</td>
</tr>
<tr>
<td>CCA16274</td>
<td>Carlia</td>
<td>aramia</td>
<td>sp. So. High.</td>
<td>PNG: WP: Kwatu (44)</td>
<td>-6.075</td>
<td>141.523</td>
</tr>
<tr>
<td>CCA16275</td>
<td>Carlia</td>
<td>aramia</td>
<td>sp. So. High.</td>
<td>PNG: WP: Kwatu (44)</td>
<td>-6.075</td>
<td>141.523</td>
</tr>
<tr>
<td>CCA16563</td>
<td>Carlia</td>
<td>sp.</td>
<td>leucotaenia</td>
<td>PNG: MdP: Madang (21)</td>
<td>-5.21</td>
<td>145.807</td>
</tr>
<tr>
<td>CCA16565</td>
<td>Carlia</td>
<td>sp.</td>
<td>leucotaenia</td>
<td>PNG: MdP: Madang (21)</td>
<td>-5.21</td>
<td>145.807</td>
</tr>
<tr>
<td>CCA16566</td>
<td>Carlia</td>
<td>sp.</td>
<td>leucotaenia</td>
<td>PNG: MdP: Madang (21)</td>
<td>-5.21</td>
<td>145.807</td>
</tr>
<tr>
<td>CCA16579</td>
<td>Carlia</td>
<td>sp.</td>
<td>leucotaenia</td>
<td>PNG: MdP: Madang (21)</td>
<td>-5.21</td>
<td>145.807</td>
</tr>
<tr>
<td>CCA16618</td>
<td>Carlia</td>
<td>mysi</td>
<td>pulla</td>
<td>PNG: MdP: Madang (21)</td>
<td>-5.21</td>
<td>145.807</td>
</tr>
<tr>
<td>CCA16667</td>
<td>Carlia</td>
<td>sp.</td>
<td>leucotaenia</td>
<td>PNG: MdP: Madang (21)</td>
<td>-5.21</td>
<td>145.807</td>
</tr>
<tr>
<td>CCA16772</td>
<td>Carlia</td>
<td>sp. Wewak</td>
<td>pulla</td>
<td>PNG: ESP: Wewak (25)</td>
<td>-3.586</td>
<td>143.642</td>
</tr>
<tr>
<td>CCA16781</td>
<td>Carlia</td>
<td>sp. Wewak</td>
<td>pulla</td>
<td>PNG: ESP: Wewak (25)</td>
<td>-3.586</td>
<td>143.642</td>
</tr>
<tr>
<td>CCA16822</td>
<td>Carlia</td>
<td>sp. Wewak</td>
<td>pulla</td>
<td>PNG: ESP: Wewak (25)</td>
<td>-3.586</td>
<td>143.642</td>
</tr>
<tr>
<td>CCA16877</td>
<td>Carlia</td>
<td>sp. Wewak</td>
<td>pulla</td>
<td>PNG: ESP: Wewak (25)</td>
<td>-3.586</td>
<td>143.642</td>
</tr>
<tr>
<td>CCA16961</td>
<td>Carlia</td>
<td>sp. Wewak</td>
<td>pulla</td>
<td>PNG: ESP: Wewak (25)</td>
<td>-3.586</td>
<td>143.642</td>
</tr>
<tr>
<td>FK6744</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. D'Entrecasteaux</td>
<td>PNG: MBP: Normanby Is. (2)</td>
<td>-9.964</td>
<td>150.955</td>
</tr>
<tr>
<td>FK6745</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. D'Entrecasteaux</td>
<td>PNG: MBP: Normanby Is. (2)</td>
<td>-9.964</td>
<td>150.955</td>
</tr>
<tr>
<td>FK6850</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. Misima</td>
<td>PNG: MBP: Misima Is., Liak (48)</td>
<td>-10.661</td>
<td>152.685</td>
</tr>
<tr>
<td>Cat.No.</td>
<td>Genus</td>
<td>Species*</td>
<td>Species**</td>
<td>Locality</td>
<td>Latitude</td>
<td>Longitude</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
<td>-----------</td>
<td>---------------</td>
<td>-------------------------------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>FK6851</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. Misima</td>
<td>PNG: MBP: Misima Is., Liak (48)</td>
<td>-10.661</td>
<td>152.685</td>
</tr>
<tr>
<td>FK6852</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. Misima</td>
<td>PNG: MBP: Misima Is., Liak (48)</td>
<td>-10.661</td>
<td>152.685</td>
</tr>
<tr>
<td>FK7033</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. Misima</td>
<td>PNG: MBP: Misima Is., Liak (48)</td>
<td>-10.661</td>
<td>152.685</td>
</tr>
<tr>
<td>FK7034</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. Misima</td>
<td>PNG: MBP: Misima Is., Liak (48)</td>
<td>-10.661</td>
<td>152.685</td>
</tr>
<tr>
<td>FK8128</td>
<td>Carlia</td>
<td>luctuosa cf.</td>
<td>sp. Dorobisoro</td>
<td>PNG: CP: Dorobisoro (10)</td>
<td>-9.463</td>
<td>147.922</td>
</tr>
<tr>
<td>FK8129</td>
<td>Carlia</td>
<td>luctuosa cf.</td>
<td>sp. Dorobisoro</td>
<td>PNG: CP: Dorobisoro (10)</td>
<td>-9.463</td>
<td>147.922</td>
</tr>
<tr>
<td>FK8130</td>
<td>Carlia</td>
<td>luctuosa cf.</td>
<td>sp. Dorobisoro</td>
<td>PNG: CP: Dorobisoro (10)</td>
<td>-9.463</td>
<td>147.922</td>
</tr>
<tr>
<td>FK8131</td>
<td>Carlia</td>
<td>luctuosa cf.</td>
<td>sp. Dorobisoro</td>
<td>PNG: CP: Dorobisoro (10)</td>
<td>-9.463</td>
<td>147.922</td>
</tr>
<tr>
<td>FK8132</td>
<td>Carlia</td>
<td>luctuosa cf.</td>
<td>sp. Dorobisoro</td>
<td>PNG: CP: Dorobisoro (10)</td>
<td>-9.463</td>
<td>147.922</td>
</tr>
<tr>
<td>FK8283</td>
<td>Carlia</td>
<td>mysi</td>
<td>leucotaenia</td>
<td>PNG: MoP: Lae (30)</td>
<td>-6.733</td>
<td>147.002</td>
</tr>
<tr>
<td>FK8284</td>
<td>Carlia</td>
<td>mysi</td>
<td>leucotaenia</td>
<td>PNG: MoP: Lae (30)</td>
<td>-6.733</td>
<td>147.002</td>
</tr>
<tr>
<td>FK8285</td>
<td>Carlia</td>
<td>mysi</td>
<td>leucotaenia</td>
<td>PNG: MoP: Lae (30)</td>
<td>-6.733</td>
<td>147.002</td>
</tr>
<tr>
<td>FK8286</td>
<td>Carlia</td>
<td>mysi</td>
<td>leucotaenia</td>
<td>PNG: MoP: Lae (30)</td>
<td>-6.733</td>
<td>147.002</td>
</tr>
<tr>
<td>FK8287</td>
<td>Carlia</td>
<td>mysi</td>
<td>leucotaenia</td>
<td>PNG: MoP: Lae (30)</td>
<td>-6.733</td>
<td>147.002</td>
</tr>
<tr>
<td>FK9419</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. Sudest</td>
<td>PNG: MBP: Sudest Is. (53)</td>
<td>-11.435</td>
<td>153.431</td>
</tr>
<tr>
<td>FK9430</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. Sudest</td>
<td>PNG: MBP: Sudest Is. (53)</td>
<td>-11.435</td>
<td>153.431</td>
</tr>
<tr>
<td>FK9431</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. Sudest</td>
<td>PNG: MBP: Sudest Is. (53)</td>
<td>-11.435</td>
<td>153.431</td>
</tr>
<tr>
<td>FK9432</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. Sudest</td>
<td>PNG: MBP: Sudest Is. (53)</td>
<td>-11.435</td>
<td>153.431</td>
</tr>
<tr>
<td>FK10081</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. Rossel</td>
<td>PNG: MBP: Rossel Is. (26)</td>
<td>-11.322</td>
<td>154.244</td>
</tr>
<tr>
<td>FK10085</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. Rossel</td>
<td>PNG: MBP: Rossel Is. (26)</td>
<td>-11.322</td>
<td>154.244</td>
</tr>
<tr>
<td>FK10405</td>
<td>Carlia</td>
<td>eothen</td>
<td>sp. Rossel</td>
<td>PNG: MBP: Rossel Is. (26)</td>
<td>-11.335</td>
<td>154.225</td>
</tr>
<tr>
<td>FK11062</td>
<td>Carlia</td>
<td>mysi</td>
<td>leucotaenia</td>
<td>PNG: ENBP: Mamar (35)</td>
<td>-5.519</td>
<td>151.502</td>
</tr>
<tr>
<td>FK11064</td>
<td>Carlia</td>
<td>mysi</td>
<td>leucotaenia</td>
<td>PNG: ENBP: Mamar (35)</td>
<td>-5.519</td>
<td>151.502</td>
</tr>
<tr>
<td>FK11114</td>
<td>Carlia</td>
<td>mysi</td>
<td>leucotaenia</td>
<td>PNG: ENBP: Mamar (35)</td>
<td>-5.519</td>
<td>151.502</td>
</tr>
<tr>
<td>FK11115</td>
<td>Carlia</td>
<td>mysi</td>
<td>leucotaenia</td>
<td>PNG: ENBP: Mamar (35)</td>
<td>-5.519</td>
<td>151.502</td>
</tr>
<tr>
<td>FK11352</td>
<td>Carlia</td>
<td>luctuosa cf.</td>
<td>sp. Dorobisoro</td>
<td>PNG: CP: Moroka (12)</td>
<td>-9.438</td>
<td>147.598</td>
</tr>
<tr>
<td>FK11354</td>
<td>Carlia</td>
<td>luctuosa cf.</td>
<td>sp. Dorobisoro</td>
<td>PNG: CP: Moroka (12)</td>
<td>-9.438</td>
<td>147.598</td>
</tr>
<tr>
<td>FK11358</td>
<td>Carlia</td>
<td>luctuosa cf.</td>
<td>sp. Dorobisoro</td>
<td>PNG: CP: Moroka (12)</td>
<td>-9.438</td>
<td>147.598</td>
</tr>
<tr>
<td>FK11359</td>
<td>Carlia</td>
<td>luctuosa cf.</td>
<td>sp. Dorobisoro</td>
<td>PNG: CP: Moroka (12)</td>
<td>-9.438</td>
<td>147.598</td>
</tr>
<tr>
<td>FK11360</td>
<td>Carlia</td>
<td>luctuosa cf.</td>
<td>sp. Dorobisoro</td>
<td>PNG: CP: Moroka (12)</td>
<td>-9.438</td>
<td>147.598</td>
</tr>
</tbody>
</table>

**Outgroups**

<table>
<thead>
<tr>
<th>Cat.No.</th>
<th>Genus</th>
<th>Species*</th>
<th>Locality</th>
<th>Latitude</th>
<th>Longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCA16357</td>
<td>Carlia</td>
<td>bicarinata</td>
<td>PNG: CP: Yule Island</td>
<td>-8.829</td>
<td>146.529</td>
</tr>
<tr>
<td>CCA16358</td>
<td>Carlia</td>
<td>bicarinata</td>
<td>PNG: CP: Yule Island</td>
<td>-8.829</td>
<td>146.529</td>
</tr>
<tr>
<td>Cat.No.</td>
<td>Genus</td>
<td>Species*</td>
<td>Species**</td>
<td>Locality</td>
<td>Latitude</td>
</tr>
<tr>
<td>-----------</td>
<td>-------</td>
<td>----------</td>
<td>-----------</td>
<td>----------------------</td>
<td>----------</td>
</tr>
<tr>
<td>CCA16359</td>
<td>Carlia</td>
<td>bicarinata</td>
<td></td>
<td>PNG: CP: Yule Island</td>
<td>-8.829</td>
</tr>
<tr>
<td>CCA16360</td>
<td>Carlia</td>
<td>bicarinata</td>
<td></td>
<td>PNG: CP: Yule Island</td>
<td>-8.829</td>
</tr>
<tr>
<td>CCA16436</td>
<td>Carlia</td>
<td>bicarinata</td>
<td></td>
<td>PNG: CP: Yule Island</td>
<td>-8.829</td>
</tr>
<tr>
<td>CCA16437</td>
<td>Carlia</td>
<td>bicarinata</td>
<td></td>
<td>PNG: CP: Yule Island</td>
<td>-8.829</td>
</tr>
<tr>
<td>CCA2292</td>
<td>Carlia</td>
<td>bicarinata</td>
<td></td>
<td>PNG: NCD: Port Moresby</td>
<td>-9.426</td>
</tr>
<tr>
<td>CCA2758</td>
<td>Carlia</td>
<td>bicarinata</td>
<td></td>
<td>PNG: NCD: Port Moresby</td>
<td>-9.473</td>
</tr>
<tr>
<td>CCA2763</td>
<td>Carlia</td>
<td>bicarinata</td>
<td></td>
<td>PNG: NCD: Port Moresby</td>
<td>-9.473</td>
</tr>
<tr>
<td>CCA3827</td>
<td>Carlia</td>
<td>bicarinata</td>
<td></td>
<td>PNG: CP: Bootless Bay</td>
<td>-9.509</td>
</tr>
<tr>
<td>CCA3828</td>
<td>Carlia</td>
<td>bicarinata</td>
<td></td>
<td>PNG: CP: Bootless Bay</td>
<td>-9.509</td>
</tr>
<tr>
<td>CCA5592</td>
<td>Carlia</td>
<td>sp. Amau bicarinate</td>
<td></td>
<td>PNG: CP: Amau village</td>
<td>-10.037</td>
</tr>
<tr>
<td>CCA5869</td>
<td>Carlia</td>
<td>sp. Amau bicarinate</td>
<td></td>
<td>PNG: CP: Amau village</td>
<td>-10.037</td>
</tr>
<tr>
<td>CCA5870</td>
<td>Carlia</td>
<td>sp. Amau bicarinate</td>
<td></td>
<td>PNG: CP: Amau village</td>
<td>-10.037</td>
</tr>
<tr>
<td>CCA5879</td>
<td>Carlia</td>
<td>sp. Amau bicarinate</td>
<td></td>
<td>PNG: CP: Amau village</td>
<td>-10.037</td>
</tr>
<tr>
<td>ABTC11026</td>
<td>Carlia</td>
<td>storri</td>
<td></td>
<td>AUS: QLD: near Mossman</td>
<td>-16.461</td>
</tr>
<tr>
<td>ABTC11027</td>
<td>Carlia</td>
<td>storri</td>
<td></td>
<td>AUS: QLD: near Mossman</td>
<td>-16.461</td>
</tr>
<tr>
<td>ABTC11028</td>
<td>Carlia</td>
<td>storri</td>
<td></td>
<td>AUS: QLD: near Mossman</td>
<td>-16.461</td>
</tr>
<tr>
<td>DSF05</td>
<td>Carlia</td>
<td>storri</td>
<td></td>
<td>AUS: QLD: Magnetic Island</td>
<td>-19.136</td>
</tr>
<tr>
<td>QMJ78178</td>
<td>Carlia</td>
<td>storri</td>
<td></td>
<td>AUS: QLD: Lakeland, Cooktown Rd</td>
<td>-15.862</td>
</tr>
<tr>
<td>ALS182</td>
<td>Carlia</td>
<td>&quot;storri&quot;</td>
<td></td>
<td>IND: MKP: Aru Islands</td>
<td>-6.885</td>
</tr>
<tr>
<td>ALS213</td>
<td>Carlia</td>
<td>&quot;storri&quot;</td>
<td></td>
<td>IND: MKP: Aru Islands</td>
<td>-6.855</td>
</tr>
<tr>
<td>ALS214</td>
<td>Carlia</td>
<td>&quot;storri&quot;</td>
<td></td>
<td>IND: MKP: Aru Islands</td>
<td>-6.855</td>
</tr>
<tr>
<td>ALS254</td>
<td>Carlia</td>
<td>&quot;storri&quot;</td>
<td></td>
<td>IND: MKP: Aru Islands</td>
<td>-6.813</td>
</tr>
<tr>
<td>CCA16071</td>
<td>Carlia</td>
<td>&quot;storri&quot;</td>
<td></td>
<td>PNG: WP: Daru</td>
<td>-9.074</td>
</tr>
<tr>
<td>CCA16078</td>
<td>Carlia</td>
<td>&quot;storri&quot;</td>
<td></td>
<td>PNG: WP: Daru</td>
<td>-9.074</td>
</tr>
<tr>
<td>CCA16102</td>
<td>Carlia</td>
<td>&quot;storri&quot;</td>
<td></td>
<td>PNG: WP: Daru</td>
<td>-9.074</td>
</tr>
<tr>
<td>CCA16103</td>
<td>Carlia</td>
<td>&quot;storri&quot;</td>
<td></td>
<td>PNG: WP: Daru</td>
<td>-9.074</td>
</tr>
<tr>
<td>CCA16156</td>
<td>Carlia</td>
<td>&quot;storri&quot;</td>
<td></td>
<td>PNG: WP: Daru</td>
<td>-9.074</td>
</tr>
<tr>
<td>ABTC16223</td>
<td>Carlia</td>
<td>prava</td>
<td></td>
<td>AUS: QLD: Townsville</td>
<td>-19.259</td>
</tr>
<tr>
<td>DSF06</td>
<td>Carlia</td>
<td>prava</td>
<td></td>
<td>AUS: QLD: Magnetic Island</td>
<td>-19.136</td>
</tr>
<tr>
<td>ABTC11024</td>
<td>Carlia</td>
<td>schmeltzii</td>
<td></td>
<td>AUS: QLD: Gladstone</td>
<td>-23.863</td>
</tr>
<tr>
<td>ABTC32058</td>
<td>Carlia</td>
<td>schmeltzii</td>
<td></td>
<td>AUS: QLD: Sarina</td>
<td>-21.424</td>
</tr>
<tr>
<td>ABTC1107</td>
<td>Carlia</td>
<td>jarnoldae</td>
<td></td>
<td>AUS: QLD: Morehead River</td>
<td>-15.024</td>
</tr>
<tr>
<td>AMJ46155</td>
<td>Carlia</td>
<td>jarnoldae</td>
<td></td>
<td>AUS: QLD: Morehead River</td>
<td>-15.024</td>
</tr>
<tr>
<td>AMJ62732</td>
<td>Carlia</td>
<td>jarnoldae</td>
<td></td>
<td>AUS: QLD: Mt Abbot</td>
<td>-20.102</td>
</tr>
<tr>
<td>QMJ62695</td>
<td>Carlia</td>
<td>jarnoldae</td>
<td></td>
<td>AUS: QLD: Mt Aberdeen</td>
<td>-20.202</td>
</tr>
<tr>
<td>ABTC28872</td>
<td>Carlia</td>
<td>munda</td>
<td></td>
<td>AUS: NT: Bing Bong Station</td>
<td>-16.088</td>
</tr>
</tbody>
</table>
Table D.2. Continued.

<table>
<thead>
<tr>
<th>Cat.No.</th>
<th>Genus</th>
<th>Species*</th>
<th>Species**</th>
<th>Locality</th>
<th>Latitude</th>
<th>Longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brandy1</td>
<td>Carla</td>
<td>rhomboidalis</td>
<td></td>
<td>AUS: QLD: Conway Range</td>
<td>-20.335</td>
<td>148.681</td>
</tr>
<tr>
<td>Brandy2</td>
<td>Carla</td>
<td>rhomboidalis</td>
<td></td>
<td>AUS: QLD: Conway Range</td>
<td>-20.335</td>
<td>148.681</td>
</tr>
<tr>
<td>Thornt1</td>
<td>Carla</td>
<td>rubrigularis</td>
<td></td>
<td>AUS: QLD: Thornton Uplands</td>
<td>-16.176</td>
<td>145.385</td>
</tr>
<tr>
<td>ABTC29093</td>
<td>Carla</td>
<td>triacantha</td>
<td></td>
<td>AUS: NT: Jabiru Airstrip</td>
<td>-12.660</td>
<td>132.893</td>
</tr>
<tr>
<td>NTMR20879</td>
<td>Carla</td>
<td>triacantha</td>
<td></td>
<td>AUS: NT: Jabiru Airstrip</td>
<td>-12.660</td>
<td>132.893</td>
</tr>
<tr>
<td>ABTC11045</td>
<td>Carla</td>
<td>vivax</td>
<td></td>
<td>AUS: NSW: Yamba</td>
<td>-29.438</td>
<td>153.360</td>
</tr>
<tr>
<td>ABTC1123</td>
<td>Carla</td>
<td>vivax</td>
<td></td>
<td>AUS: QLD: Mt Mulgrave</td>
<td>-16.311</td>
<td>143.997</td>
</tr>
<tr>
<td>CCA15431</td>
<td>Emoia</td>
<td>caeruleacauda</td>
<td></td>
<td>PNG: MBP: Trobriand Is., Kiriwina</td>
<td>-8.542</td>
<td>151.079</td>
</tr>
<tr>
<td>CJS921</td>
<td>Liburnascincus</td>
<td>coensis</td>
<td></td>
<td>AUS: QLD: Mallwraith, Peach Ck</td>
<td>-13.638</td>
<td>143.099</td>
</tr>
<tr>
<td>CCA15079</td>
<td>Lygisaurus</td>
<td>curtus</td>
<td></td>
<td>PNG: CP: Wabo Village</td>
<td>-6.980</td>
<td>145.069</td>
</tr>
<tr>
<td>CCA5896</td>
<td>Lygisaurus</td>
<td>curtus</td>
<td></td>
<td>PNG: CP: Wabo Village</td>
<td>-6.980</td>
<td>145.069</td>
</tr>
<tr>
<td>CCA5897</td>
<td>Lygisaurus</td>
<td>curtus</td>
<td></td>
<td>PNG: CP: Wabo Village</td>
<td>-6.980</td>
<td>145.069</td>
</tr>
<tr>
<td>CCA16019</td>
<td>Lygisaurus</td>
<td>macfarlani</td>
<td></td>
<td>PNG: WP: Daru</td>
<td>-9.074</td>
<td>143.211</td>
</tr>
<tr>
<td>CCA16020</td>
<td>Lygisaurus</td>
<td>macfarlani</td>
<td></td>
<td>PNG: WP: Daru</td>
<td>-9.074</td>
<td>143.211</td>
</tr>
<tr>
<td>CCA16072</td>
<td>Lygisaurus</td>
<td>macfarlani</td>
<td></td>
<td>PNG: WP: Daru</td>
<td>-9.074</td>
<td>143.211</td>
</tr>
</tbody>
</table>
Table D.2. Loci included in this study, along with primers used in amplification and sequencing, annealing temperature used in PCR, and final aligned locus length. Ta: annealing temperature for PCR.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Primers</th>
<th>Sequence</th>
<th>Ta (°C)</th>
<th>Length (bp)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ND4</td>
<td>ND4.light</td>
<td>CACCTATGACTACCAAAAGCTCATGTAGAAGC</td>
<td>55</td>
<td>707</td>
<td>(Arèvalo et al. 1994)</td>
</tr>
<tr>
<td></td>
<td>Leu3-heavy</td>
<td>GAAATTAGCAGTTCTTTTRTG</td>
<td></td>
<td></td>
<td>(Stuart-Fox et al. 2002)</td>
</tr>
<tr>
<td>Aldolase</td>
<td>Ald1CR</td>
<td>AAGAAGGATGGAGCTGACT TTGC</td>
<td>60</td>
<td>212</td>
<td>(Dolman &amp; Phillips 2004)</td>
</tr>
<tr>
<td></td>
<td>Ald2CR</td>
<td>GCCATTCTGTAACACAACAG CCAA</td>
<td></td>
<td></td>
<td>(Dolman &amp; Phillips 2004)</td>
</tr>
<tr>
<td>Anonymous</td>
<td>Sk13</td>
<td>CTTTCATGCATCAATGGAC</td>
<td></td>
<td>698</td>
<td>(Dolman &amp; Phillips 2004)</td>
</tr>
<tr>
<td></td>
<td>Sk14</td>
<td>CATGTGCCTGCTCACAT C</td>
<td></td>
<td></td>
<td>(Dolman &amp; Phillips 2004)</td>
</tr>
<tr>
<td>β-globin</td>
<td>Bglo1CR</td>
<td>GCAGACTGCACGTACAAAG</td>
<td>59</td>
<td>804</td>
<td>(Dolman &amp; Phillips 2004)</td>
</tr>
<tr>
<td></td>
<td>Bglo2CR</td>
<td>GCTGCCAAGCGGGTGTGA</td>
<td></td>
<td></td>
<td>(Dolman &amp; Phillips 2004)</td>
</tr>
<tr>
<td>GAPD</td>
<td>GAPD-L890</td>
<td>ACCTTTAATGCGGCTGTGC CATTTGC</td>
<td>63</td>
<td>359</td>
<td>(Friesen et al. 1997)</td>
</tr>
<tr>
<td></td>
<td>GAPD-H950</td>
<td>CATCAAAGCTCCCAACACGTTGTGCTGA</td>
<td></td>
<td></td>
<td>(Friesen et al. 1997)</td>
</tr>
<tr>
<td>Myosin</td>
<td>MYH2-F</td>
<td>GAAACCAAGCCTCATCAACC</td>
<td>60</td>
<td>412</td>
<td>(Lyons et al. 1997)</td>
</tr>
<tr>
<td></td>
<td>MYH2-R</td>
<td>TGTTGTTCCTGCTCCTTCTC</td>
<td></td>
<td></td>
<td>(Lyons et al. 1997)</td>
</tr>
<tr>
<td>R35</td>
<td>R35-F</td>
<td>GACTGTGGAYGAYCTGATCAGTGTGGTGCTCC</td>
<td>63</td>
<td>650</td>
<td>(Brandley et al. 2011)</td>
</tr>
<tr>
<td></td>
<td>R35-R</td>
<td>GCCAAAATGAGCGAAACCGCCTTCTGAGC</td>
<td></td>
<td></td>
<td>(Brandley et al. 2011)</td>
</tr>
<tr>
<td>Rhodopsin</td>
<td>Rho3CR</td>
<td>CCTTGCTGACACCCCTATGCTG</td>
<td>59</td>
<td>513</td>
<td>(Dolman &amp; Phillips 2004)</td>
</tr>
<tr>
<td></td>
<td>Rho4CR</td>
<td>CAGGAGAGACCCCTACATTG</td>
<td></td>
<td></td>
<td>(Dolman &amp; Phillips 2004)</td>
</tr>
</tbody>
</table>
APPENDIX E
PERMISSIONS FOR PORTIONS OF CHAPTER 4

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS

Jun 17, 2014

This is a License Agreement between Eric N Rittmeyer ("You") and John Wiley and Sons ("John Wiley and Sons") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by John Wiley and Sons, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

<table>
<thead>
<tr>
<th>License Number</th>
<th>3410501399522</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Jun 15, 2014</td>
</tr>
<tr>
<td>Licensed content publisher</td>
<td>John Wiley and Sons</td>
</tr>
<tr>
<td>Licensed content publication</td>
<td>Significance</td>
</tr>
<tr>
<td>Licensed content title</td>
<td>The geometry of hobbits: Homo floresiensis and human evolution</td>
</tr>
<tr>
<td>Licensed copyright line</td>
<td>© 2009 The Royal Statistical Society</td>
</tr>
<tr>
<td>Licensed content author</td>
<td>William Jungers,Karen Baab</td>
</tr>
<tr>
<td>Licensed content date</td>
<td>Nov 19, 2009</td>
</tr>
<tr>
<td>Start page</td>
<td>159</td>
</tr>
<tr>
<td>End page</td>
<td>164</td>
</tr>
<tr>
<td>Type of use</td>
<td>Dissertation/Thesis</td>
</tr>
<tr>
<td>Requestor type</td>
<td>University/Academic</td>
</tr>
<tr>
<td>Format</td>
<td>Print and electronic</td>
</tr>
<tr>
<td>Portion</td>
<td>Figure/table</td>
</tr>
<tr>
<td>Number of figures/tables</td>
<td>1</td>
</tr>
<tr>
<td>Original Wiley figure/table number(s)</td>
<td>Figure 1</td>
</tr>
<tr>
<td>Will you be translating?</td>
<td>No</td>
</tr>
<tr>
<td>Title of your thesis / dissertation</td>
<td>SYSTEMATICS AND SPECIES DELIMITATION IN NEW GUINEA SKINK SPECIES COMPLEXES (SQUAMATA: SCINCIDAE)</td>
</tr>
<tr>
<td>Expected completion date</td>
<td>Jun 2014</td>
</tr>
<tr>
<td>Expected size (number of pages)</td>
<td>125</td>
</tr>
<tr>
<td>Total</td>
<td>0.00 USD</td>
</tr>
</tbody>
</table>

Terms and Conditions
F.1. Specimens Examined
List of specimens examined for morphological analyses; specimens are housed in the Australian Museum (AM), the American Museum of Natural History (AMNH), the Bernice P. Bishop Museum (BPBM), the Louisiana State University Museum of Natural Science (LSUMZ), the Museum of Comparative Zoology, Harvard University (MCZ), or the South Australian Museum (SAMA).

*Tribolonotus pseudoponceleti*
Papua New Guinea: Bougainville: Kunua, -5.767º, 154.717º – n=41 – AMNH 117707-17, MCZ 72839, 72871, 72873, 72878-80, 72885-6, 72899, 72910, 72914 (holotype), 76193-4, 76422, 76432-3, 76439, 76442, 76449, 76455, 78094, 78099-100, 78105, 78108, 78119, 78131, 78250, 78252, 78265.


*Tribolonotus ponceleti*
Figure F.1. Schematic of the nine demographic models tested. A. Model 1, no divergence between Buka and Bougainville. B. Model 2, divergence between Buka and Bougainville with no subsequent gene flow. C. Model 3, divergence between Buka and Bougainville with a shift in population size and no gene flow. D. Model 4, divergence between Buka and Bougainville with a constant rate of migration. E. Model 5, migration between Buka and Bougainville stopping at time $t_3$ subsequent to the divergence. F. Model 6, migration between Buka and Bougainville changing in rate at time $t_3$. G. Model 7, Buka and Bougainville population sizes shift and migration stops at time $t_3$. H. Model 8, Buka and Bougainville population sizes shift at time $t_3$ with no change in migration rate. I. Model 9, the most complex demographic model, both migration rates between and population sizes of Buka and Bougainville change at time $t_3$. 

135
Figure F.2. Plot of first two principal components (PC) from the summary statistics of 900 simulated datasets (100 for each of the three models), and the observed summary statistics. Inset indicates the source of each point (i.e. observed dataset, or the model under which the data was simulated). The first PC explained 81.52% of the variation; the second PC explained 10.18% of the variation. On the first PC axis, the total mean and variance in segregating sites had strong positive loadings, and the mean and variance of within population segregating sites, and total mean and variance in $\theta_w$ and $\pi$; other variables were also included in PC1, but with a weaker loadings. On the second PC axis, the mean and variance in segregating sites within the Buka and Bougainville populations, and the mean within population $\theta_w$ had strong negative loadings, while the mean and variance in total segregating sites had a strong positive loading; other variables were also included in PC2, but with a weaker loadings.
Figure F.3. A. Plot of $\Delta K$ for dataset of loci recovered from all twelve individuals. B. Plot of $\ln \Pr(D|K)$ for dataset of loci recovered from all twelve individuals. Blue circles indicate the mean value for each $K$, and grey lines show the standard deviation. C. Plot of $\Delta K$ for dataset of loci sequenced in at least nine individuals. D. Plot of $\ln \Pr(D|K)$ for dataset of loci sequenced in at least nine individuals. Blue circles indicate the mean value for each $K$, and grey lines show the standard deviation. E. Plot of $\Delta K$ for dataset of loci sequenced in at least six individuals. F. Plot of $\ln \Pr(D|K)$ for dataset of loci sequenced in at least six individuals. Blue circles indicate the mean value for each $K$, and grey lines show the standard deviation.
Figure F.4. Density plots of the posterior distributions of all parameters from the ABC analyses. A. $\theta_{1b}$, population size of the Bougainville population, following the shift in population size and cessation of migration. B. $\theta_{2b}$, population size of the Buka population, following the shift in population size and cessation of migration. C. $\theta_3$, population size of *Tribolonotus ponceleti*. D. $\theta_{1a}$, population size of the Bougainville population, prior to the shift in population size and cessation of migration. E. $\theta_{2a}$, population size of the Buka population, prior to the shift in population size and cessation of migration. F. $\theta_{12}$, population size of common ancestor of Buka + Bougainville. G. $\theta_a$, ancestral population size of common ancestor of Buka + Bougainville + *T. ponceleti*. H. $t_1$, divergence time of *T. ponceleti* from Bougainville + Buka. I. $t_2$, divergence time between Buka and Bougainville. J. $t_3$, timing of the shift in population size in, and cessation of migration between, the Bougainville and Buka populations. K. $m_{12a}$, migration from Bougainville to Buka prior to the shift in population size and cessation of migration. L. $m_{21a}$, migration from Buka to Bougainville prior to the shift in population size and cessation of migration.
F.3. Supplemental Tables

Table E.1. Specimens included in next-generation sequencing (NGS) analyses.

<table>
<thead>
<tr>
<th>Catalogue #</th>
<th>Island</th>
<th>Locality</th>
<th>Latitude</th>
<th>Longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tribolonotus pseudoponceleti</td>
<td>LSUMZ 93494</td>
<td>Bougainville</td>
<td>Mt. Balbi, SE Slope</td>
<td>-5.94018</td>
</tr>
<tr>
<td></td>
<td>LSUMZ 93495</td>
<td>Bougainville</td>
<td>Mt. Balbi, SE Slope</td>
<td>-5.94018</td>
</tr>
<tr>
<td></td>
<td>LSUMZ 93498</td>
<td>Bougainville</td>
<td>Mt. Balbi, SE Slope</td>
<td>-5.94018</td>
</tr>
<tr>
<td></td>
<td>LSUMZ 93492</td>
<td>Bougainville</td>
<td>Togarau Two Village, SE slope of Mt. Balbi</td>
<td>-5.96038</td>
</tr>
<tr>
<td></td>
<td>LSUMZ 93493</td>
<td>Bougainville</td>
<td>Togarau Two Village, SE slope of Mt. Balbi</td>
<td>-5.96038</td>
</tr>
<tr>
<td></td>
<td>LSUMZ 93503</td>
<td>Buka</td>
<td>Nova Area, Chi Chi Hav Village</td>
<td>-5.39075</td>
</tr>
<tr>
<td></td>
<td>LSUMZ 93505</td>
<td>Buka</td>
<td>Nova Area, SE of Bridge at Ramunfun River</td>
<td>-5.38925</td>
</tr>
<tr>
<td></td>
<td>LSUMZ 93507</td>
<td>Buka</td>
<td>Nova Area, SE of Bridge at Ramunfun River</td>
<td>-5.38925</td>
</tr>
<tr>
<td></td>
<td>LSUMZ 93509</td>
<td>Buka</td>
<td>Nova Area, SE of Bridge at Ramunfun River</td>
<td>-5.38925</td>
</tr>
<tr>
<td></td>
<td>LSUMZ 93513</td>
<td>Buka</td>
<td>Nova Area, SE of Bridge at Ramunfun River</td>
<td>-5.38925</td>
</tr>
<tr>
<td>Tribolonotus ponceleti</td>
<td>SJR 5431</td>
<td>Isabel Island</td>
<td>Northwestern Isabel Island</td>
<td>-7.54444</td>
</tr>
<tr>
<td></td>
<td>SJR 5572</td>
<td>Choiseul Island</td>
<td>Southeastern Choiseul Island</td>
<td>-7.34416</td>
</tr>
</tbody>
</table>

Table F.2. Sequences of adapters and primers used in next-generation library preparation. Eco-F and Eco-R, and Mse-F and Mse-F are the AFLP adapters for EcoRI and MseI, respectively, from Vos et al. (1995). EcoRI-PreAmp and MseI-PreAmp are the preamplification primers from McCormack et al. (2012); ‘b’ represents biotinylation of the EcoRI-PreAmp primer, and the NN are selective bases, in this study, TA were used. EcoRI-A-MIDX is the EcoRI fusion primer used in the final selective amplification, and includes the Ion Torrent A adapter (CCATCTCATCCTCGGTGTTTCTCGGAC), the key sequence (TCAG), an individual-specific MID barcode (indicated by X’s), and the portion after the barcode (GACTGCGTACCAATTC) is the EcoRI adapter. Finally, MseRI-P1 is the MseI fusion primer used in the final selective amplification, and includes the Ion Torrent P1 adapter sequence (CCTCTCATATGGGCAGTCGGTGAT), and the MseI adapter (GATGAGTCCTGAGTAA), and the selective bases (NN, in this study, we used TA).

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eco-F</td>
<td>CTCGTAAGACTGCGTACC</td>
<td>(Vos et al. 1995)</td>
</tr>
<tr>
<td>Eco-R</td>
<td>AATTTGAGCAGTCTTAC</td>
<td>(Vos et al. 1995)</td>
</tr>
<tr>
<td>Mse-F</td>
<td>GACGATGAGTCTGAG</td>
<td>(Vos et al. 1995)</td>
</tr>
<tr>
<td>Mse-R</td>
<td>TACTCAGGACTCAT</td>
<td>(Vos et al. 1995)</td>
</tr>
<tr>
<td>EcoRI-PreAmp</td>
<td>b-GACTGCGTACCAATTC</td>
<td>(McCormack et al. 2012)</td>
</tr>
<tr>
<td>MseI-PreAmp</td>
<td>GATGAGTCCTGAGTAANN</td>
<td>(McCormack et al. 2012)</td>
</tr>
<tr>
<td>EcoRI-A-MIDX</td>
<td>CCATCTCATCCTCGGTGTTTCTCGGACTCA</td>
<td>This Study</td>
</tr>
<tr>
<td>MseRI-P1</td>
<td>CCTCTCATATGGGCAGTCGGTGATGATGA</td>
<td>This Study</td>
</tr>
<tr>
<td></td>
<td>GTCCTGAGTAANN</td>
<td></td>
</tr>
</tbody>
</table>
Table F.3. Estimated divergence times (in coalescent units) and thetas from SNAPP analyses under different priors on theta for each dataset of varying completeness. Values in parentheses represent the 95% HPD interval. T₁ is the divergence time of Buka+Bougainville from *T. ponceleti*, T₂ is the divergence time of Buka and Bougainville. θ₁ represents the Bougainville population, θ₂ represents Buka, θ₃ represents *T. ponceleti*, θ₁₂ is the common ancestor of Buka+Bougainville, and θₐ is the ancestral population size.

<table>
<thead>
<tr>
<th></th>
<th>12 Inds</th>
<th></th>
<th>9+ Inds</th>
<th></th>
<th>6+ Inds</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>θ~G(2,100)</td>
<td>θ~G(2,1000)</td>
<td>θ~G(2,100)</td>
<td>θ~G(2,1000)</td>
<td>θ~G(2,100)</td>
<td>θ~G(2,1000)</td>
</tr>
<tr>
<td>T₁</td>
<td>1.636 (0.704-2.5259)</td>
<td>1.476 (0.4752-2.8639)</td>
<td>1.2618 (0.5698-2.0907)</td>
<td>1.5578 (0.6107-2.603)</td>
<td>1.355 (0.4868-2.5019)</td>
<td>1.2581 (0.6083-2.0918)</td>
</tr>
<tr>
<td>θ₁</td>
<td>0.0081078 (0.0048028-0.012013)</td>
<td>0.0030024 (0.0014012-0.0048345)</td>
<td>0.0002705 (0.0001259-0.0004366)</td>
<td>0.0061401 (0.003513-0.0086217)</td>
<td>0.0023739 (0.0010189-0.00368)</td>
<td>0.0002399 (0.0001069-0.0003804)</td>
</tr>
<tr>
<td>θ₂</td>
<td>0.008271 (0.004776-0.011949)</td>
<td>0.0030828 (0.0014095-0.0049205)</td>
<td>0.0002799 (0.0001308-0.0004467)</td>
<td>0.0069499 (0.003713-0.0099636)</td>
<td>0.002693 (0.0012093-0.0041847)</td>
<td>0.0002731 (0.0001243-0.0004264)</td>
</tr>
<tr>
<td>θ₃</td>
<td>0.0047094 (0.0026811-0.0065761)</td>
<td>0.0018105 (0.0006955-0.0028514)</td>
<td>0.0001578 (0.000684-0.0021075)</td>
<td>0.0035954 (0.0006383-0.00052471)</td>
<td>0.0013895 (0.0006006-0.00022693)</td>
<td>0.0001362 (0.00021581-0.00049581)</td>
</tr>
<tr>
<td>θ₁₂</td>
<td>0.0054355 (0.0024854-0.0079584)</td>
<td>0.0021111 (0.0008057-0.0037287)</td>
<td>0.0001642 (0.0000663-0.0002716)</td>
<td>0.0048802 (0.0005398-0.0075383)</td>
<td>0.0017319 (0.00009717-0.0031547)</td>
<td>0.0001541 (0.0002681-0.007634)</td>
</tr>
<tr>
<td>θₐ</td>
<td>0.0060652 (0.0010893-0.011336)</td>
<td>0.0028306 (0.0002678-0.0047145)</td>
<td>0.0003795 (0.0001553-0.0006027)</td>
<td>0.0069561 (0.001476-0.01185)</td>
<td>0.0036838 (0.0019622-0.0058033)</td>
<td>0.0004421 (0.0005916-0.0010106)</td>
</tr>
</tbody>
</table>
Table F.4. Accuracy of approximate Bayesian computation model prediction, based on performing rejection steps on 100 randomly selected suites of summary statistics under each of the models. Values shown represent the proportion of tests in which the correct model had the maximum posterior probability (PP), in which the correct model had a posterior probability > 0.75, and in which an incorrect model had a posterior probability > 0.75. Max PP = Isolation indicate the percent of replicates in which an isolation model (i.e. no migration, models 2,3) had the highest posterior probability, Max PP = IM indicates the replicates in which an isolation with migration model (models 4–9) had the highest posterior probability, and Max PP = No Divergence indicates the replicates in which the model with no divergence between Buka and Bougainville (model 1) had the highest posterior probability.

<table>
<thead>
<tr>
<th>True Model</th>
<th>Max PP = Correct Model</th>
<th>Correct Model PP &gt; 0.75</th>
<th>Incorrect Model PP &gt; 0.75</th>
<th>Max PP = Isolation</th>
<th>Max PP = IM</th>
<th>Max PP = No Divergence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>93%</td>
<td>76%</td>
<td>0%</td>
<td>0%</td>
<td>7%</td>
<td>93%</td>
</tr>
<tr>
<td>2</td>
<td>54%</td>
<td>3%</td>
<td>0%</td>
<td>71%</td>
<td>29%</td>
<td>0%</td>
</tr>
<tr>
<td>3</td>
<td>44%</td>
<td>1%</td>
<td>0%</td>
<td>72%</td>
<td>28%</td>
<td>0%</td>
</tr>
<tr>
<td>4</td>
<td>42%</td>
<td>0%</td>
<td>0%</td>
<td>5%</td>
<td>93%</td>
<td>2%</td>
</tr>
<tr>
<td>5</td>
<td>32%</td>
<td>0%</td>
<td>1%</td>
<td>41%</td>
<td>59%</td>
<td>0%</td>
</tr>
<tr>
<td>6</td>
<td>39%</td>
<td>0%</td>
<td>0%</td>
<td>5%</td>
<td>94%</td>
<td>1%</td>
</tr>
<tr>
<td>7</td>
<td>25%</td>
<td>0%</td>
<td>0%</td>
<td>33%</td>
<td>67%</td>
<td>0%</td>
</tr>
<tr>
<td>8</td>
<td>27%</td>
<td>0%</td>
<td>0%</td>
<td>6%</td>
<td>93%</td>
<td>1%</td>
</tr>
<tr>
<td>9</td>
<td>36%</td>
<td>2%</td>
<td>0%</td>
<td>3%</td>
<td>96%</td>
<td>1%</td>
</tr>
</tbody>
</table>
Table F.5. Accuracy of parameter estimation from approximate Bayesian computation, based on performing rejection steps on 100 simulated suites of summary statistics.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Proportion within 95% HPD interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>m_{12a}</td>
<td>96%</td>
</tr>
<tr>
<td>m_{21a}</td>
<td>96%</td>
</tr>
<tr>
<td>t_1</td>
<td>96%</td>
</tr>
<tr>
<td>t_2</td>
<td>98%</td>
</tr>
<tr>
<td>t_3</td>
<td>98%</td>
</tr>
<tr>
<td>\theta_{1a}</td>
<td>98%</td>
</tr>
<tr>
<td>\theta_{2a}</td>
<td>100%</td>
</tr>
<tr>
<td>\theta_{1b}</td>
<td>98%</td>
</tr>
<tr>
<td>\theta_{2b}</td>
<td>100%</td>
</tr>
<tr>
<td>\theta_3</td>
<td>100%</td>
</tr>
<tr>
<td>\theta_{12}</td>
<td>96%</td>
</tr>
<tr>
<td>\theta_a</td>
<td>100%</td>
</tr>
<tr>
<td>All Parameters</td>
<td>97.80%</td>
</tr>
</tbody>
</table>
VITA

Eric Nikolaus Rittmeyer was born in Rochester, New York in November 1986 to Deborah Jane Rittmeyer (nee Gleason) and Thomas Hans Rittmeyer. He was raised in the nearby rural town of Byron, New York, where from a young age he developed an appreciation and enthusiasm for nature and the outdoors, and particularly for searching for and catching reptiles and amphibians. Eric attended Byron-Bergen High School, graduating in 2004, where he was involved in a number of extracurricular activities including varsity swimming, cross country, scholastic bowl, science olympiad, math league, and the marching, concert, and jazz bands.

In 2004, he moved to Ithaca, New York to attend Cornell University for his undergraduate studies. As a freshman, he joined the Cornell Herpetological Society, where his interest in reptiles and amphibians and their evolution developed under the guidance of other club members and the club’s advisor, Kraig Adler, and in which he soon took on a leadership role, including serving as the club president in his senior year. Early in his junior year, he joined the laboratory of Kelly Zamudio, where his research on reptile evolution began under the guidance of both Kelly Zamudio and Harry Greene. Later in his junior year, while working as an undergraduate teaching assistant for a course on vertebrate biology, Eric, along with two fellow members of the Cornell Herpetological Society, Mike Gründler and Derrick Thompson, were inspired to organize an expedition to New Guinea though a guest lecture by Ed Scholes on his experiences studying the birds of paradise in Papua New Guinea. Shortly thereafter, the three began applying for grants to fund an expedition to Papua New Guinea. With substantial support from Allen Allison of the Bishop Museum, Eric, Mike, and Derrick embarked on a six-week expedition to the north coast of Papua New Guinea in June 2008, shortly after Eric had graduated from Cornell, a trip that would solidify Eric’s interest in the evolution of the herpetofauna of the region.

Upon his return to the United States, Eric joined the laboratory of Chris Austin at Louisiana State University in August 2008 as a doctoral student. After a few years of graduate work and two more expeditions to Papua New Guinea, Eric moved within Baton Rouge, and met his new neighbor and future fiancée, Zeynep Altinay, soon before his fourth trip to New Guinea. In early 2014, Eric was awarded a National Science Foundation Postdoctoral Fellowship in Biology, which will support his continued research on the evolution of reptiles and amphibians in New Guinea and the biogeography of the Sahul Shelf region of New Guinea and Australia as a postdoctoral researcher in the laboratory of Craig Moritz at the Australian National University in Canberra, Australia.