COMPARATIVE PHYLOGENOMICS, CRYPTIC SPECIES AND EVOLUTION OF LIZARDS IN THE CERRADO BIODIVERSITY HOTSPOT

FABRICIUS MAIA CHAVES BICALHO DOMINGOS, BSC, MSC

Thesis submitted to fulfil

the requirement for the degree of

Doctor of Philosophy

School of Biological Sciences

Faculty of Science and Engineering

Flinders University

Adelaide, Australia

June 2015

“I cannot live without brainwork. What else is there to live for?”

Arthur Conan Doyle – The Sign of the Four

“Shame for the summit of creation, let sanity prevail, defending Darwin's throne. Where blind belief and science meet, we're defending Darwin's throne”

Heaven Shall Burn – Deaf to our prayers

Table of contents

Figure list

Table list

Appendix list

Summary

The evolution and diversification patterns of the Neotropical biota are a matter of extensive scientific debate. Despite centuries of research interest in this region, the levels of biodiversity in the Neotropics are still largely underestimated. This is particularly true for the Cerrado, the largest Neotropical savannah and a formally recognized biodiversity hotspot. The Cerrado landscape is dominated by ancient plateaus and younger valleys that were excavated by river catchments. Throughout the Quaternary climatic fluctuations, during moister periods, the savannah vegetation was restricted to refugia in the plateaus, while the valleys were invaded by forest-like vegetation. Both the landscape compartmentalisation caused by the uplift of the Central Brazilian Plateau and the Quaternary climatic fluctuations have been proposed as drivers of diversification of the Cerrado biota. However, these hypotheses have not been properly tested in a phylogeographic perspective, and the understanding of processes that shaped the distribution of biological diversity within the Cerrado is still incipient. Lizards have for long been used as model organisms in evolutionary studies. They are generally poor dispersers and thus can be used as indicators of fine-scale biogeographic history. The study of endemic Cerrado lizards has the potential to elucidate the influence of historical changes in the landscape on the ecological characteristics of lineages, and to clarify the resulting patterns of biodiversity. In this thesis, I employed a comparative phylogeography approach and used species delimitation methods to address knowledge gaps of Cerrado endemic lizards, and to clarify diversification patterns in the Cerrado. Three codistributed endemic lizard species were targeted: *Gymnodactylus amarali* (Phyllodactylidae), *Micrablepharus atticolus* (Gymnophtalmidae) and *Tropidurus itambere* (Tropiduridae). For each species, I used a combination of phylogenetic analyses, Bayesian species delimitation analyses, coalescent statistical phylogeography and population genetic estimates based on sequences of one mitochondrial DNA gene and of ~400 nuclear loci obtained using an anchored phylogenomics

protocol. Morphological data for one of the three taxa were also integrated into the analyses. The results suggested the existence of several cryptic species within each taxon. Statistical phylogeographic analyses coupled with species distribution modeling indicated that endemic lizards exhibit a degree of concordant phylogeographic history but also taxon-specific evolutionary patterns within the Cerrado. The two species groups that use similar habitats, *G. amarali* and *T. itambere*, displayed similar geographic distribution of basal clades, and similar estimated ancestral distributions. On the other hand, results also indicated that landscape compartmentalisation probably played different roles in the evolution of each taxon. The ecologically distinct *M. atticolus* and *T. itambere* had very similar palaeodistributional shifts throughout the Quaternary, while *G. amarali* presented a different refugia pattern. Overall patterns of diversification are associated with geologic processes during the Neogene and with a complex history of colonisation of plateaus and valleys during the Quaternary. This thesis pioneered the investigation of several competing diversification hypotheses about the Cerrado in a comparative context, and is the first example of species delimitations methods using next-generation sequencing for Cerrado organisms. The results will guide the description of several new species for the biome and directly contribute to conservation planning. Future research on the evolution of Cerrado biota should focus on linking patterns of genetic diversity and speciation with climatic and geomorphological processes of the biome.

Certificate of candidate

I certify that this thesis ‘Comparative Phylogenomics, Cryptic Species and Evolution of Lizards in the Cerrado Biodiversity Hotspot’ does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text

Fabricius Maia Chaves Bicalho Domingos

June 2015

Acknowledgements

This thesis was supported by a scholarship from 'Coordenação de Aperfeiçoamento de Pessoal de Nível Superior' (Capes) from the Brazilian Ministry of Education (MEC), and funded by Capes, 'Conselho Nacional de Desenvolvimento Científico e Tecnológico' (CNPq), 'Fundação de Apoio à Pesquisa do Distrito Federal' (FAPDF), 'GENPAC15' and 'Rede de Pesquisa Biota do Cerrado'.

I am greatly thankful to my supervisor Luciano Beheregaray for the support, both personal and academic, during my whole time in Australia. His broad scientific view, enthusiasm, and commitment to make good science are constant examples to follow. Thanks for being a great supervisor, for the friendship, for amplifying my scientific view and for making this project possible. Up the Irons!

To my supervisor from Brazil, Guarino Colli, for the trust and respect, and for all the support during my entire academic life. His academic integrity and constant dedication to science are examples that I will carry for the rest of my life. Eleven year have passed since I first started at CHUNB, and I never stoped learning from his intellectual sagacity and extensive field experience. Also, he owes me a beer.

Many people directly or indirectly contributed to this thesis. Of all of them, only one shared every single step with me, my partner Marina Scalon. I cannot express in words the support and love she dedicated to me during my whole candidature. Thank you for sharing your dreams with me, and for the love for science and life we developed together! Te amo muito!

To all my friends at the Molecular Ecology Lab at Flinders (and associated to it). Thank you all for the friendship, and for the awesome moments we spent together: Chris and Astrid, Cat,

Steve and Julie, Minami and Oliver, Pedro and Ju, Jimena, Peter, Shannon, Nikki, Daniel and Mari, Yuma, Tess, Hilary, Luciana, Edu and Rafa, and the lovely kids Moana, Enzo, Pipe, Luiza, Petria and Madlen who always brought happiness to our lives. In particular, to Chris Brauer for the long conversations about science, life, beer, and the wrongness of religion. Hail Sagan!

To all my friends at CHUNB (or that were there at some point), and that somehow helped by either going to field, providing samples, shape files, pictures, etc: Marcella e Renan, Roger, Chelinha, Fê Werneck, Jéssica e Léo, Ísis, Lalá, Bernardo, Guth, Joseana, Davi, Pedro, Mari Mira, Almir, Cazé, Gabriel Costa, Daniel Velho, Gabriel Horta, Reuber, and Rodrigo. Also to all other CHUNB members for the friendship and support throughout the years. I'm particularly thankful to Universidade de Brasília staff not only for going to the field, but also for sharing their field experiences with me during so many expeditions: Santos Balbino, Antônio Mendes, and Wandélio Mendes.

To my friends in Adelaide and Sydney for making life happier: Clarissa, Passa, Mayra, Felipe, Flávio and Alê (and Luquinhas), Tom, Leslie, Dave, Emma, Julieta Martinelli, James, Vashi, Anthony, Francesca, Saskia, Julieta and Mark, Julia and Matt, Vincent, Allyson, Raquel, Kasia, Yu Na Kim, and Denitsa.

To Instituto Chico Mendes de Conservação da Biodiversidade and their staff members, specially Tainah Guimarães and Isaías Reis, for providing collection permits. And to all administration staff, rangers and field assistants of several conservation parks around Cerrado: Laerte Marques, Divino, Dona Aurora, Gabriel, Joãozinho, Adelton, Bufinha, Cotonett, among others. I am also thankful to Dra. Christine Strussmann and Dr. Renato Farias for important information regarding collection sites.

I am particularly grateful to my friends and family from Brazil, for the constant presence in my life. Galera do Power Metal, and Macakongs 2099. To my sisters Dora and Didiça, Victor, Hegel, and my crazy little nephews, Lulu and Miguel, for being a source of constant happiness. To my other family, Luciana, Marcela, Marcelo e Bete. To my father Luiz, for all the love and support, and for the shared love about scepticism, science fiction and comic books. To my mother, Rose, that, above all, always shared the good and bad parts of every aspect of my life, and taught me how to be happy. That was the most important lesson I ever learned.

Finally, from the uncountable extreme music artists that inspired me and will continue to do so, three were of utmost important during the writing of this thesis: Mick Kenney, Kim Bendix Petersen and Jonne Järvelä. Thank you very much!

Chapter 1

General Introduction

1. Background

1.1 Evolutionary context and Cerrado biogeography

The distribution and geographic variation of organisms in nature may be influenced by several variables, such as spatial and environmental differences (Borcard *et al.*, 1992; Borcard & Legendre, 2002), biophysical processes (Austin, 2007), aggregation patterns (Condit *et al.*, 2000), and biotic interactions (Guisan & Thuiller, 2005). Different approaches have been used to unravel the relative importance of these variables on the biogeography of species, populations and communities (Cracraft, 2002; Guisan & Thuiller, 2005; Jablonski *et al.*, 2006). Although species distribution and assemblage composition depend on aspects of ecology (MacArthur & Wilson, 1963; Lomolino, 2000; Thorpe *et al.*, 2008) and on evolutionary and phylogenetic history (Thorpe *et al.*, 1991; Posadas *et al.*, 2006), the bridge between ecological and historical biogeography is still hardly explored (Wiens & Donoghue, 2004). The integration of historical and ecological approaches can elucidate complex evolutionary patterns related to species richness and diversification (Smith *et al.*, 2005; Werneck *et al.*, 2009). Moreover, some studies linked ecological and historical information to further investigate the evolution of phylogenetically related organisms in different taxonomic levels: within the same genus (Hickerson & Cunningham, 2005; Parra *et al.*, 2010), family (Lima *et al.*, 2012), and even within the same order (Parra *et al.*, 2010). Similar approaches have been used to propose conservation strategies (Carnaval *et al.*, 2009), which can be important when dealing with the conservation of biomes under great human pressure (Carnaval *et al.*, 2014).

The Brazilian savannah, locally known as Cerrado, only recently became the focus of conservation efforts (Klink & Machado, 2005). Indeed, more than 50% of its ~2 million km^2 have been transformed for cattle raising and agriculture in the last 45 years (Klink & Machado, 2005). The Cerrado is the second largest biome in South America, has the richest flora among all savannahs in the world (Mendonça *et al.*, 1998; Castro *et al.*, 1999; Furley,

1999), and each year many new animals and plants are described, indicating that a significant part of its biodiversity is still unknown (Mendonça *et al.*, 1998; Colli *et al.*, 2002; Marinho-Filho *et al.*, 2002; Oliveira-Filho & Ratter, 2002). The combination of these characteristics with the current high rates of deforestation, and the fact that only 2.2% of its area is under legal protection (Klink & Machado, 2005) makes the Cerrado one of the formally identified global biodiversity hotspots for conservation (Myers *et al.*, 2000; Mittermeier *et al.*, 2005).

The Cerrado is characterized by a savannah-like landscape with a complex and diverse vegetation structure (Goodland, 1971; Oliveira Filho & Ratter, 2000). The diversity of habitats probably had an important role in determining the local diversity of the Cerrado herpetofauna (Colli *et al.*, 2002; Nogueira *et al.*, 2005), as already suggested for birds (Silva, 1997; Silva & Bates, 2002) and mammals (Redford & da Fonseca, 1986; Mares & Ernest, 1995; Johnson *et al.*, 1999). This diversity is likely a reflection of geomorphological, geographical and historical factors that influenced the evolution of the biome (Nogueira *et al.*, 2011). The landscape is dominated by ancient plateaus, which underwent an epeirogenic uplift in the Miocene-Pliocene transition (King, 1956) and have their surfaces covered by a layer of soils dating from the Neogene (Radambrasil, 1982-83). An erosion cycle through the Neogene exposed a second surface covered by clay soils (King, 1956; Gomes *et al.*, 2004), and a third surface, considerably variable across the Cerrado, was formed when this erosion cut through the first two surfaces exposing Precambrian rocks in some areas (Geraldes *et al.*, 2001). Several soil types are found over the biome distribution, but most are young tropical soils (cambisols), red and red-yellow oxidated soils with high organic content (oxisols), and some basically unaltered soils (entisols) (Gomes *et al.*, 2004). This geomorphological mosaic combined with the great geographical variation of soil types, results in a complex landscape (Oliveira Filho & Ratter, 2000; Bridgewater *et al.*, 2004) and leads to the formation of isolated habitats and ecotones that could partially explain the high biodiversity of the biome (Beheregaray *et al.*, 2015).

The spatial heterogeneity hypothesis holds that an increase in the number of habitats leads to a growth in species number due to a greater number of niche dimensions (MacArthur & MacArthur, 1961; Pianka, 1966). Habitat heterogeneity appears to play a key role in the generation and maintenance of high species richness in 'open areas' (e.g., deserts and savannahs) around the world, and has provided a fertile ground for testing hypotheses about speciation and evolution of biodiversity (Schall & Pianka, 1978; Pianka, 1989). However, the high diversity found in heterogeneous areas can occur for distinct reasons. On one hand, these areas may act as biodiversity museums, which have lower extinction rates, and may accumulate species that have arrived there through dispersal over evolutionary time (Haffer, 1997; Moritz *et al.*, 2000; Sanmartín *et al.*, 2008). On the other hand, these areas can provide lineage diversification serving as evolutionary factories (Fjeldsaå *et al.*, 1997). Finally, a recent study showed that habitat heterogeneity due to geological history, the ability to disperse over the landscape, and the amount of time a lineage has persisted in the area are strong predictors of speciation in Amazonian birds (Smith *et al.*, 2014b).

The isolation provided by the mosaic of different habitats was appointed as a key element for the speciation of the central Brazilian herpetofauna (Vanzolini, 1997). In the Cerrado region, this mosaic is usually represented by open formations dominated by grasses and shrubs, with few trees (grasslands and cerrado *sensu stricto*), and forested formations with closed canopy and sparse grasses and shrubs ('cerradão' and gallery forests) (Oliveira-Filho & Ratter, 2002). Nogueira and collaborators (2009) showed that there is little overlap between species of lizards in forests and open formations, apparently because these habitats act as natural barriers to the distribution and dispersal of lizards. These small-scale boundaries between different habitats in the Cerrado may lead to the formation of new species by acting as barriers and leading to isolation (Ogden & Thorpe, 2002), while the high habitat heterogeneity may also allow for the coexistence of many species in local scales (Colli *et al.*, 2002).

There are very few vertebrate fossil localities in central Brazil (Báez & de Gasparini, 1979; Estes & Báez, 1985), and most well preserved squamate fossils date from the Quaternary (Camolez & Zaher, 2010). Only eight squamate fossils are known from the Brazilian Cretaceous, four of them being lizards, but only one represents a modern family (Candeiro, 2007; Candeiro *et al.*, 2009). Thus, there is an evident gap of fossil information from the Palaeogene and Neogene, when most modern families and species very likely originated (Townsend *et al.*, 2011; Mulcahy *et al.*, 2012), hindering phylogenetic studies of lower taxonomic levels (i.e., among genera and species). Studies on the herpetofauna of the open landscapes of South America are based almost entirely in the much better quality fossil record of the southern portion of the continent (Webb, 1978; Duellman, 1979). Phylogenetic and phylogeographic analyses of extant Cerrado reptiles are still scarce (Gamble *et al.*, 2012; Werneck *et al.*, 2012a; Giugliano *et al.*, 2013; Santos *et al.*, 2014), and significant interpretations of historical events that shaped the distribution and evolutionary patterns of organisms can only be obtained in light of their phylogenetic relationships (Brown & Lomolino, 1998).

In a preliminary analysis, Colli (2005) suggested that the main events contributing to the origins and speciation of the Cerrado herpetofauna consisted of: (1) a climatic gradient associated with the formation of three floristic provinces at the beginning of the Palaeogene, (2) the Miocene marine introgressions, (3) the epeirogenic uplift of the Central Brazilian Plateau uplift, (4) the arrival of immigrants from North and Central America at the end of the Neogene, and (5) the Quaternary climatic fluctuations. After this first publication on biogeography of the Cerrado herpetofauna, new insights about the evolution of squamate diversity in the Cerrado were brought by Nogueira and collaborators (2011). These authors described ten putative areas of endemism that prevailed in open and elevated plateaus whereas, in peripheral depressions, faunal interchange associated with forested habitats was more common. Moreover, they argued that vicariant speciation was probably the major

process shaping Cerrado squamate diversity, and that such pattern is strongly correlated with the Central Brazilian Plateau uplift (Nogueira *et al.*, 2011).

Since many endemic reptile species from the Cerrado are found in elevated plateaus, they may have evolved in isolation after the uplift of the Central Brazilian Plateau (Colli, 2005; Werneck *et al.*, 2009; Nogueira *et al.*, 2011). The abundance of these endemic species in isolated plateaus indicates an older origin for the Central Brazilian savannahs, confronting the initially proposed interpretations that the Cerrado has an anthropogenic origin due to human generated fire regime (Salgado-Laboriau, 2005). There are suggestions that the Cerrado is a Cretaceous formation, present before the final separation of South America and Africa (Ratter *et al.*, 1997), but more conservative estimates place its origin in the Eocene (Werneck, 2011). Nonetheless, the Cerrado vegetation has not been static in its distribution. Palynological data indicates the expansion of the Cerrado vegetation during dry periods and retraction in wet periods, specially of the savannah woodlands currently found in the valleys (Ledru, 1993; Sifeddine *et al.*, 2003). Although the origin of the Cerrado vegetation is still controversial (Salgado-Laboriau, 2005), palaeopalynological data indicate that typical Cerrado vegetation existed at least 32,000 years ago (Ledru *et al.*, 2006). Therefore, it should be expected that the herpetofauna is older than some of the extant vegetation patterns, and the explanations for its distribution patterns is probably deeply embedded in the geological and geomorphological history of the Cerrado.

The separation of the Cerrado in high altitude (plateaus) and low altitude areas (valleys) also determines dominant soil composition (Motta *et al.*, 2002) and vegetation mosaics (Oliveira-Filho & Ratter, 2002). Indeed, evidence based on distribution models from the Pleistocene Last Interglacial (~120 thousand years ago) until the present suggest that the Cerrado vegetation was probably restricted to refugia in plateaus during the Quaternary climatic fluctuations (Werneck *et al.*, 2012b). The valleys were, then, probably occupied by forests (Ab'Sáber, 1998; Werneck *et al.*, 2012b), which could have extinguished Cerrado-

adapted organisms from the valleys (Werneck, 2011). Thus, expected effects of Quaternary climatic fluctuations would broadly depend on these elevation and topographical conditions (Bush, 1994). Based on these assertions, it is expected that populations on plateaus should have higher genealogical structure and genetic diversity, correlated with their older origin, whereas populations in the valleys should show the opposite pattern, consistent with their younger age (Werneck, 2011). Moreover, if the species were present in the landscape before the uplift of the Central Brazilian Plateau, this process could have promoted vicariance between plateau *versus* valley populations, resulting in reciprocally monophyletic groups of populations between these two areas (Werneck, 2011). However, most of these hypotheses were not properly tested in a phylogeographic perspective, and the understanding of processes that shaped the distribution of herpetofauna diversity within the Cerrado is still incipient. Phylogenetic and phylogeographic analyses of lineages predominantly distributed in the biome, especially of endemic taxa, are further required to evaluate such proposals.

The Cerrado herpetofauna was initially described as being impoverished and lacking a substantial endemic lizard fauna when compared to other Brazilian biomes (Vanzolini, 1948, 1976; Vitt, 1991). However, recent studies have shown that the Cerrado bears a very rich and locally diverse herpetofauna, with high levels of endemism (Colli *et al.*, 2002; Nogueira *et al.*, 2005; Mesquita *et al.*, 2006; Nogueira *et al.*, 2009; Nogueira *et al.*, 2011). Several new lizard species are still being described for the region (e.g., Colli *et al.*, 2003c; Colli *et al.*, 2003b; Nogueira & Rodrigues, 2006; Rodrigues *et al.*, 2007; Rodrigues *et al.*, 2008; Colli *et al.*, 2009; Ribeiro *et al.*, 2009; Strüssmann & Mott, 2009; Pinna *et al.*, 2010; de Freitas *et al.*, 2011; Giugliano *et al.*, 2013; Recoder *et al.*, 2014) stressing the fact that the biome is still poorly sampled. The most recently published account of the Cerrado squamate fauna recorded 267 squamate species, of which 103 (39%) are endemics, including 20 amphisbaenians (61% endemism), 32 lizards (42%) and 52 snakes (32%) (Nogueira *et al.*, 2011). Nevertheless, data obtained for Squamata species in major collections and museums in Brazil indicates extensive

regions, mainly in the northern portion of the Cerrado, where there is a lack of basic information on diversity (Costa *et al.*, 2007).

Despite the fact that the glamour of taxonomy diminished a lot in the last century, the description of new species is a central activity in biology (Wagele *et al.*, 2011; Tancoigne & Dubois, 2013). The task of knowing all species in the world is maybe unachievable, and a recent review points that we are far from solving this problem, with around 86% of species on Earth still awaiting description (Mora *et al.*, 2011). Nonetheless, the first task on describing species is actually finding those species, an activity that is commonly hindered in high diverse regions such as the Neotropics (Brito, 2010). Not only we lack information about the presence or absence (distribution) of the species (Costa *et al.*, 2010), but the available information from biological collections might be insufficient to enable proper comparisons across specimens and the recognition of discrete entities. Moreover, new species might arise with no clear morphological differentiation (Pfenninger & Schwenk, 2007) and, in this case, recognising discrete species might be impossible based solely on morphological data (Beheregaray & Caccone, 2007). The discovery of new species has been greatly influenced by the development of molecular tools, the use of new molecular markers, and because of the development of molecular species delimitation (Carstens *et al.*, 2013) and species-tree methods (Degnan & Rosenberg, 2009) in recent years.

Considering the fast rate with which the Cerrado is being destroyed, and the lack of financial and human resources, it becomes impractical to conduct a detailed sampling of the biome to cover gaps of distribution within a short timeframe. The use of species distribution modelling (SDM) can be an alternative to minimize this problem (Guisan & Zimmermann, 2000; Rodríguez *et al.*, 2007). The tool of SDM is also powerful in disentangling the potential roles of ecology and historical factors influencing species distributions (Costa *et al.*, 2008), helping in conservation planning (Rodríguez *et al.*, 2007; Costa *et al.*, 2010), understanding species richness and their changes over time (Blach-Overgaard *et al.*, 2013), and even

modelling the distribution of whole biomes (Carnaval & Moritz, 2008; Werneck *et al.*, 2011; Werneck *et al.*, 2012b). There are limitations to the use of SDM, especially when the environmental space is poorly sampled (Elith *et al.*, 2006), and this problem could bias the models built for large and hyper-diverse regions such as the Cerrado. Nonetheless, detailed insights on the evolution of organisms can be drawn using models of recent distribution and palaeodistribution in conjunction with phylogeographic analyses (Carstens & Richards, 2007; Knowles *et al.*, 2007; Richards *et al.*, 2007).

Despite recent advances in the field, phylogeographic studies have traditionally separated the 'phylo-' from the '-geographic' component (Kidd & Ritchie, 2006). This is a particularly critical limitation considering that conservation efforts are primarily based on geographic information to assess the conservation status of taxa and to select areas to be preserved. In this context, using SDMs together with molecular phylogeography brings a strong reconciliation of the two components (Richards *et al.*, 2007; Swenson, 2008). Moreover, the hypothesis testing accuracy and power of comparative phylogeography studies can be considerably improved when combining the use of SDM data (Carstens & Richards, 2007; Richards *et al.*, 2007; Smith *et al.*, 2011), as well as specifically helping in asserting conservation strategies (Carnaval *et al.*, 2009; Provan & Maggs, 2012).

Comparative phylogeograpy is a powerful framework to investigate patterns of shared evolutionary history (Bermingham & Moritz, 1998; Bernatchez & Wilson, 1998). The incorporation of historical and ecological information about the species, allied to coalescent analyses of genetic data, can help identifying important evolutionary patterns for a particular geographic region (Carstens & Richards, 2007; Fouquet *et al.*, 2012; Bagley & Johnson, 2014b). In the context of the Cerrado, this framework provides the means to test the above-mentioned diversification hypotheses, and to investigate patterns of genetic diversity within the Cerrado landscape. If the tectonic uplift of the Central Brazilian Plateau acted as a strong vicariant mechanism (Nogueira *et al.*, 2011), geographic discontinuities and similar

phylogenetic patterns would potentially be found among codistributed endemic species. In the same line of thought, the influence of the Quaternary climatic fluctuations on population structure and genetic diversity would potentially be replicated among different species (Werneck, 2011).

Lizards have for long been used as model organisms in evolutionary studies (Vitt & Pianka, 2005; Losos, 2009), probably because they are easily collected and manipulated (Camargo *et al.*, 2010). Also, they are generally poor dispersers and thus can be used as indicators of fine-scale biogeographic history. As such, phylogeographic studies using lizards have exponentially increased since 1997, but less then 6% of them used South American species (Camargo *et al.*, 2010). The Cerrado region still lacks consistent research assessments of genetic diversity, especially concerning biogeographic patterns aimed at reconstructing the evolutionary history of the biome. Using a comparative phylogeographic framework can help explore intrinsic evolutionary patterns within the biome, and potentially inform on conservation strategies (Carnaval *et al.*, 2009).

1.2 PhD justification and aims

Since the evolution of the Cerrado endemic fauna is still substantially unknown, it is of paramount importance that new efforts are focused on broadening the range of available information about biogeographic patterns and generating knowledge about the processes that influenced the diversification, genetic structure and speciation of these organisms. In my PhD project I investigated the biogeographic history in the Brazilian Cerrado using endemic codistributed lizards as a research subject. My initial and primary aim was to conduct a comparative phylogeographic study of three Cerrado endemic lizards, namely *Gymnodactylus amarali* Barbour, 1925, *Tropidurus itambere* Rodrigues, 1987, and *Micrablepharus atticolus* Rodrigues, 1996. However, at the start of data collection, preliminary analyses of mtDNA for all species disclosed deep intraspecific divergences, with genetic distances among populations

indicative of the presence of cryptic species. Therefore, the need for clarifying the status of those putative cryptic species became evident before a more in-depth phylogeographic investigation could be conducted.

Cryptic species are discrete lineages that, because of morphological similarity, have been incorrectly assigned to a single formal taxon (Beheregaray & Caccone, 2007; Pfenninger & Schwenk, 2007). Many methodological approaches using genetic data for the delimitation of species have been developed in recent years (Carstens *et al.*, 2013), with the potential to bridge the gap between evolutionary studies and formal taxonomy (Fujita *et al.*, 2012). Therefore, I initially used coalescent species delimitation analyses and morphological data (the latter for *G. amarali*) to investigate species limits in all three taxa. In all cases, I adopted the Generalized Lineage Concept (GLC; de Queiroz, 2007) as the species concept, under which species are viewed as evolving entities over time, and molecular or morphological differences can be understood as properties of the evolutionary divergence of such entities. Subsequently, I used coalescent phylogeographic analyses and SDMs (recent and past climate) to investigate the evolutionary history and genetic diversity of these three widespread and codistributed Cerrado lizards.

1.3 Sampling strategy

I initially Sanger-sequenced one mtDNA locus (cytochrome-*b*, cytb), and also started collecting three nuclear loci (KIF24, *β*-fibrinogen, and MYH2) for a large subset of individuals. The strategy laid out back in 2010 was to Sanger-sequence ~10 loci for each species. In late 2013, my primary supervisor contemplated the opportunity to collaborate with Alan and Emily Lemmon from Florida State University, who had recently developed an anchored hybrid enrichment protocol for the collection of nuclear data (Lemmon *et al.*, 2012). With the field of molecular ecology actively moving towards the use of next-generation sequencing (NGS) technologies (Carstens *et al.*, 2012; McCormack *et al.*, 2013), and the fact

that our laboratory (the Molecular Ecology Lab at Flinders) had practically shifted entirely to using NGS, we decided this was an important opportunity to increase our nuclear loci sampling. In addition, this would allow the development of my PhD project in line with recent advances in the field. In mid 2014 we obtained a ~400 loci anchored phylogenomics (AP) dataset using the above-mentioned protocol (Lemmon *et al.*, 2012). Due to financial and computational limitations, the AP datasets were generated for a subset of our total sample sequenced for cytb. Although it is common practice to sub-sample individuals when Sanger-sequencing nuclear loci for phylogeographic studies (e.g., Morando *et al.*, 2003; Camargo *et al.*, 2012; Werneck *et al.*, 2012a), most studies would still aim for a reasonable number of sequenced individuals per population. Given the high costs of acquiring our dataset, in many cases it was not possible to sequence more than one individual per population. Nevertheless, recent work has shown that even sampling only one individual per population can be enough for strong phylogeographic estimation, as long as enough loci are sequenced (Lohse *et al.*, 2012; Leaché *et al.*, 2013a; Smith *et al.*, 2014a).

The individual lizard samples used in my project were previously available at the 'Coleção Herpetológica da Universidade de Brasília' (CHUNB), obtained through a collaboration with Miguel Rodrigues and José Cassimiro from 'Universidade de São Paulo' (for *G. amarali*), and donated by other institutions ('Museu de Zoologia da Universidade de São Paulo' – MZUSP, and 'Universidade Federal do Mato Grosso' – UFMT). In addition, a large number of specimens were obtained as a result of an eight months intensive fieldwork in Brazil. For that field expedition we travelled ~45.000 Km by car, sampling areas that were not previously visited and filling key gaps in the sampling design of this PhD project.

1.4 Focal species

All three of our focal species are Cerrado endemic lizards widely distributed in the biome. Below, I present a short introduction of each species in terms of its taxonomy, systematics and ecology. More detailed information is given in the subsequent chapters when applicable.

1.4.1 Gymnodactylus amarali (Phyllodactylidae)

The Neotropical gecko *Gymnodactylus amarali* Barbour, 1925 has a wide distribution in the central and northern portions of the Cerrado biome (Vanzolini, 2005). It was synonymized with *Gymnodactylus geckoides* Spix, 1825 by Vanzolini (1953a) who, at that point, believed that no ecological barriers existed between the Caatinga and the Cerrado. Vanzolini himself later recognized morphological and ecological dissimilarities between the two species, but restricted *G. amarali* to the 'Alto Parnaíba' region (close to the type locality of Barbour) based only on one specimen, and described a new species, *Gymnodactylus carvalhoi* Vanzolini, 2005, as the widespread form in the Cerrado (Vanzolini, 2005). However, Cassimiro and Rodrigues (2009) synonymised *G. carvalhoi* with *G. amarali* after rechecking the type specimen described by Vanzolini, arguing that his diagnostic characters were highly variable within *Gymnodactylus* specimens sampled in the Cerrado. Currently, there are five described species in the genus *Gymnodactylus*, all within the Brazilian territory: *G. amarali,* endemic to the Cerrado; *G. darwinii* (Gray, 1845), endemic to the Atlantic Rainforest; *G. geckoides*, endemic to the Caatinga; and two other species restricted to the 'Espinhaço' mountain range, known only from the surroundings of their type localities: *G. guttulatus* Vanzolini, 1982, from 'Diamantina', 'Minas Gerais', in the southernmost segment of the 'Espinhaço', and *G. vanzolinii* Cassimiro and Rodrigues, 2009, from 'Mucugê', 'Bahia', in the northern portion of the Espinhaço.

As many other Neotropical geckos, *G. amarali* has crepuscular habits (Colli *et al.*, 2003a), and lives primarily in the rock crevices of rocky outcrops found in the Cerrado

landscape ('Cerrado rupestre') (Colli *et al.*, 2003a). This species feeds on termites and, indeed, when not in the rock crevices it can be found living inside the termite nests (Vitt *et al.*, 2007). Its reproduction cycle takes place during the dry season and, unlike most other geckos, clutch size is correlated with female body size (Colli *et al.*, 2003a).

1.4.2 Micrablepharus atticolus (Gymnophtalmidae)

The genus *Micrablepharus* Boettger, 1885 (Gymnophtalmidae) comprises only two species of eyelid-less lizards: *Micrablepharus atticolus* Rodrigues, 1996 and *Micrablepharus maximiliani* (Reinhardt and Lütken, 1861). While both species are distributed in the Cerrado (*M. maximiliani* is also found in the Caatinga and Pantanal), they are rarely found in sympatry (Santos *et al.*, 2014). Our focal species, *M. atticolus*, has a very wide distribution in the Cerrado, and is also found on a few isolated Cerrado enclaves inside the Amazon forest (Gainsbury & Colli, 2003; Santos *et al.*, 2014). Out of our three focal species, *M. atticolus* is the only that has been subject of previous genetic studies, with high levels of mtDNA genetic diversity found among its populations (Santos *et al.*, 2014). *Micrablepharus atticolus* is a cryptozoic species, usually found in the leaf-litter or inside ant nests (Vitt, 1991; Rodrigues, 1996), inhabits open physiognomies in the Cerrado (Cerrado *sensu stricto*) (Vitt, 1991; Vieira *et al.*, 2000), and reproduces mainly during the dry season (Vieira *et al.*, 2000).

1.4.3 Tropidurus itambere (Tropiduridae)

The species of *Tropidurus* Wied-Neuwied, 1825 are usually locally abundant (Wiederhecker *et al.*, 2003; Faria & Araujo, 2004), and generally found in many urban areas (Rodrigues, 1987; Martins *et al.*, 1999). *Tropidurus* lizards have been the subject of investigations in many fields, such as population and community ecology (Van-Sluys, 1993; Van-Sluys, 1997; Vitt & Zani, 1998; Van-Sluys *et al.*, 2004), physiology and performance (Kohlsdorf *et al.*, 2004; Kiefer *et al.*, 2005; Kohlsdorf & Navas, 2007), morphological evolution (Vitt *et al.*,

1997; Kohlsdorf *et al.*, 2001), and island biogeography (Schluter, 1984). Despite the great amount of published studies on *Tropidurus*, there are no temporal diversification studies that can inform on speciation patterns of the genus in South America (Carvalho *et al.*, 2013). Of all three study genera, *Tropidurus* is the one with more taxonomic problems, and a phylogeny and taxonomic review of the genus is currently in progress (Carvalho, *pers. comm.*). Our focal species, *Tropidurus itambere* Rodrigues, 1987 is found on rocky outcrops (similarly to *G. amarali*) (Van-Sluys, 1997), using the vertical surfaces of the rocks to forage and thermoregulate (Faria & Araujo, 2004), and reproduces mainly during the wet season (Van-Sluys, 1993; Ferreira *et al.*, 2009).

2. Thesis outline

This thesis is divided in five chapters: the present general introduction, three data chapters, and a general and intentionally brief discussion. In the first data chapter (Chapter 2) I used the genetic data available at that time (cytb and one nuclear locus, KIF24), together with morphological data, to investigate species limits in *G. amarali*. This chapter was published in the journal *Molecular Phylogenetics and Evolution* (Appendix 1) in 2014. From Chapter 3 onwards I already had access to the AP dataset and it became clear that a reassessment of phylogenetic relationships and species limits in *G. amarali* using a larger number of loci was warranted. As such, Chapter 3 was conceived in a more atypical way that we believe will bring benefits for the reading and understanding of the thesis: instead of dividing it in 'publication chapters' where ideally we would have one chapter per species, Chapter 3 deals with species delimitation using the AP dataset for all three species. Because the methods used for each species were identical, Chapter 3 provides a much concise way of presenting findings of this PhD thesis. Finally, Chapter 4 is a comparative phylogeographic analysis of the three species groups. Below, I summarize the main findings of each data chapter.

Chapter 2: *Out of the deep: Cryptic speciation in a Neotropical gecko (Squamata, Phyllodactylidae) revealed by species delimitation methods.*

In this chapter we employed mitochondrial and nuclear DNA, as well as morphological data, to assess the monophyly and cryptic speciation in *Gymnodactylus amarali*, an endemic lizard of the Brazilian savannah (Cerrado) biodiversity hotspot. Our study is the first to use samples that cover the entire distribution of this species and assess cryptic speciation of a widespread Cerrado endemic lizard using phylogenetic and species-trees methods, as well as a coalescent-based Bayesian species delimitation method. We recovered eight deeply divergent molecular clades within *G. amarali*, and two additional ones from seasonally dry tropical forest enclaves between the Cerrado and the Caatinga biome. Because of the low morphological sample size for each recovered species, applying the usual multivariate analyses to our morphological data was not feasible. Thus, we developed a new strategy to discriminate species based on a computer-learning algorithm (support vector machines), which does not require large sample sizes. To the best of my knowledge, this was the first example of such techniques being applied to morphological discrimination.

For the publication of this chapter, we used letters (A to H) to refer to delimited *G. amarali* species. To avoid confusion, and to allow for comparisons with the results from the species delimitation analyses employed in Chapter 3, the *G. amarali* putative species analysed in Chapter 3 were then referred to as numbers (1 to 12).

Chapter 3: *Cryptic species in the Neotropics: Coalescent species delimitation of Cerrado endemic lizards using anchored phylogenomics.*

In Chapter 3, I used mtDNA and the AP dataset to investigate phylogenetic relationships and cryptic speciation within three Neotropical lizards endemic to the Brazilian Cerrado. I applied a series of phylogenetic methods, and a coalescent species delimitation method (BPP) to investigate species limits within the three species. Our main results suggest that the existence

of cryptic lineages in the biome is more common than previously thought, highlighting the value of using NGS data and coalescent techniques to investigate patterns of diversity in this understudied Neotropical region. For *G. amarali* the number of cryptic species increased from 8 to 9 (compared to Chapter 2), with an overall high consistency between previous results and those obtained by analyses of AP data. The BPP results suggest that the widespread nominal taxon *M. atticolus* actually forms a complex of eight different cryptic species, with a very complicated biogeographic pattern and possible sympatry between some species. Finally, for *T. itambere*, there were five inferred cryptic lineages with high support values in every BPP run, and these lineages were also clearly geographically structured, which should facilitate their upcoming taxonomic descriptions.

Chapter 4: *Inner conflict: the roles of ecology and history on the evolution of a Neotropical biodiversity hotspot*

In Chapter 4, I tested several diversification hypotheses about the evolution of Cerrado organisms using a comparative phylogeography framework. Briefly, I tested two groups of hypotheses, those related to tectonic events that happened in the Neogene, and those related to Quaternary climatic fluctuations. Moreover, I also included more detailed diversification hypotheses to assess how endemic species with different ecologies evolved within the biome. Our statistical phylogeographic analyses and hypothesis-testing framework indicates that codistributed endemic lizard species exhibit taxon-specific evolutionary histories within the Cerrado biome, but some concordant phylogeographic patterns could be identified. The two species that use similar habitats, *G. amarali* and *T. itambere*, have similar geographic distribution of basal clades, and similar estimated ancestral distributions. On the other hand, results suggest that landscape compartmentalisation probably played different roles in the evolution of each taxon. Divergence times and ancestral effective population sizes were also more similar between *G. amarali* and *T. itambere*. However, there was no congruence in

temporal evolutionary patterns related to Neogene diversification among the three species complexes. Population genetic estimates suggest no differences in effective population size and population expansion between populations in valleys *versus* plateaus for any of the three taxa. Unexpectedly, the ecologically distinct *M. atticolus* and *T. itambere* had very similar palaeodistributional shifts throughout the Quaternary while *G. amarali* presented a different pattern.

Chapter 2

Out of the deep: Cryptic speciation in a Neotropical gecko (Squamata, Phyllodactylidae) revealed by species delimitation methods.

1. Introduction

Biodiversity in the Neotropical region has been a matter of great interest of biologists for centuries (Spix & Martius, 1824; Humboldt, 1849; Rull, 2011). The levels of biodiversity in this region remain relatively unknown (Fouquet *et al.*, 2007; Scheffers *et al.*, 2012; Fouquet *et al.*, 2013) and a large amount of species is still waiting to be discovered (Mora *et al.*, 2011; Wheeler *et al.*, 2012; Costello *et al.*, 2013). The Brazilian Cerrado is the largest Neotropical savannah (Eiten, 1972; Oliveira & Marquis, 2002) and one of the world's formally recognized biodiversity hotspots (Myers *et al.*, 2000). However, most of its area lacks adequate sampling efforts (Costa *et al.*, 2007; Costa *et al.*, 2010), which makes the discovery of new taxa not uncommon. Considering that only 2.2% of the Cerrado is under legal protection (Klink & Machado, 2005), one of the first steps towards the conservation of this biome is to investigate the taxonomic diversity and phylogenetic relationships of its endemic biota.

In early studies, the Cerrado herpetofauna was considered impoverished compared to surrounding biomes, such as the Caatinga (seasonally dry tropical forests - SDTF) and the Amazon (Vanzolini, 1948, 1976; Vitt, 1991). This paradigm has changed substantially with improved sampling efforts (Colli *et al.*, 2002), and currently 267 squamate species (39% endemics) are recognized to occur in the Cerrado (Nogueira *et al.*, 2011). Moreover, the number of species descriptions is still increasing (e.g. Nogueira & Rodrigues, 2006; Rodrigues *et al.*, 2007; Rodrigues *et al.*, 2008; Colli *et al.*, 2009; Giugliano *et al.*, 2013; Teixeira *et al.*, 2013), as well as the recognition of previously unknown cryptic lineages (Gamble *et al.*, 2012; Prado *et al.*, 2012; Giugliano *et al.*, 2013), mostly in the light of new data from populations previously assigned to the same species.

Cryptic lineage recognition can be severely impacted by morphological stasis (Pfenninger & Schwenk, 2007) and as such it is not surprising that the majority of recent cryptic species studies rely largely on genetic data (Beheregaray & Caccone, 2007; Bickford *et al.*, 2007). Coalescent-based methods have recently become popular to assist in species

delimitation (Knowles & Carstens, 2007; Yang & Rannala, 2010; Fujita *et al.*, 2012; Carstens *et al.*, 2013), especially regarding cryptic speciation in biodiversity hotspots (Nair *et al.*, 2012). Despite the unquestionable value of those methods in assessing cryptic diversity (Leaché & Fujita, 2010), it is advisable to use independent morphological or ecological data to corroborate molecular-based hypotheses of cryptic diversification (Bauer *et al.*, 2011; Burbrink *et al.*, 2011; Sistrom *et al.*, 2012). In this context, morphological data can be used to test the placement of individuals within the reconstructed molecular clades and evaluate the validity of such lineages (Hebert *et al.*, 2004; Tan *et al.*, 2010; Sistrom *et al.*, 2013). This integrative approach can provide valuable support when delimiting 'candidate species' for conservation management strategies (Morando *et al.*, 2003; Padial *et al.*, 2010).

The gecko *Gymnodactylus amarali* Barbour, 1925 is a Cerrado endemic with a wide distribution in the central and northern portions of the biome (Vanzolini, 2005). It was synonymized with *Gymnodactylus geckoides* Spix, 1825 by Vanzolini (1953a), but later resurrected and restricted to the 'Alto Parnaíba' region (close to the type locality of Barbour, Vanzolini, 2005). A new species, *Gymnodactylus carvalhoi* Vanzolini 2005, was described as the widespread form in the Cerrado (Vanzolini, 2005), but soon synonymized with *G. amarali* after Cassimiro and Rodrigues (2009) rechecked the type specimen and found that the diagnostic characters were highly variable within all *Gymnodactylus* sampled in the Cerrado. Thus, five species of *Gymnodactylus* are currently recognized, all within the Brazilian territory: *G. amarali,* endemic to the Cerrado; *Gymnodactylus darwinii* (Gray, 1845), endemic to the Atlantic Rainforest; *G. geckoides*, endemic to the Caatinga; and two other species restricted to the Espinhaço mountain range, known only from the surroundings of their type localities: *Gymnodactylus guttulatus* Vanzolini, 1982, in the southernmost segment of the Espinhaço, and *Gymnodactylus vanzolinii* Cassimiro and Rodrigues, 2009, in the northern portion. Nevertheless, only one study evaluated phylogenetic relationships within *Gymnodactylus*, addressing the phylogeography and cryptic speciation of *G. darwinii*

(Pellegrino *et al.*, 2005). The *G. darwinii* species group is monophyletic in relation to at least *G. geckoides* from Caatinga (Pellegrino *et al.*, 2005), and *G. vanzolinii* appears to be more closely related to *G. guttulatus* (Cassimiro & Rodrigues, 2009). Apart from these two assertions, no other evolutionary relationships among *Gymnodactylus* species are known.

The evolution of groups sharing a Caatinga-Cerrado distribution remains a poorly understood subject in South American biogeography (Werneck, 2011) and few studies have implemented molecular techniques to investigate the relationship between those biomes (Almeida *et al.*, 2007; Moraes *et al.*, 2009; Werneck *et al.*, 2012a; Faria *et al.*, 2013; Recoder *et al.*, 2014). Dissimilarities noted between *G. amarali* and *G. geckoides* include ecological differences such as clutch and egg sizes (Colli *et al.*, 2003a), karyological differences in chromosome number and type (Pellegrino *et al.*, 2009), and morphological differences in pholidosis and coloration patterns (Vanzolini, 1953a, 2005; Cassimiro & Rodrigues, 2009). Nonetheless, it remains unclear whether the widespread Cerrado populations of *G. amarali* form a monophyletic group in relation to its Caatinga counterpart. In fact, it was proposed that relict populations of *G. geckoides* might be present in the core of the Cerrado region (Pellegrino *et al.*, 2009), because the karyotype of one specimen was identical to the karyotype observed in *G. geckoides* populations. In addition, extensive chromosomal polymorphism has been observed within and between *G. amarali* populations (Pellegrino *et al.*, 2009), as well as great morphological variation among populations (Vanzolini, 1953a, 2005; Cassimiro & Rodrigues, 2009). As such, it is possible that populations treated under the name *G. amarali* are paraphyletic in relation to *G. geckoides* and, moreover, that differences among *G. amarali* populations reflect the existence of cryptic species. Thus, in order to investigate the evolution of *G. amarali* in the Cerrado it is also important to assess its contact zone with *G. geckoides*, accounting for the shared evolutionary history of the two biomes.

Here we investigate the potential cryptic diversity within *Gymnodactylus amarali* by implementing a framework that comprehends molecular and morphological data,

phylogenetic and 'species tree' methods, and coalescent-based Bayesian species delimitation approaches. Our hypotheses are that 1) *G. amarali* in the Cerrado is a monophyletic group; 2) there are relict *Gymnodactylus geckoides* populations inside the Cerrado biome, and 3) several cryptic species exist within *G. amarali*. Evolutionary relationships are reconstructed using samples spanning the entire distribution of *G. amarali* in the Cerrado, *Gymnodactylus* populations from SDTF enclaves in the contact zone between Cerrado and Caatinga, and populations of *G. geckoides* as outgroups. This framework enabled testing for an important contact zone between two understudied biomes and allowed us to conduct the first assessment of how molecular diagnosis predicts morphology-based cryptic divergence in a vertebrate endemic to the Cerrado.

2. Material and Methods

2.1 Taxon sampling

We obtained samples of *Gymnodactylus amarali* from 24 sites in the Cerrado, as well as two populations inhabiting SDTF enclaves (Fig. 1). Three New World gecko species (Phyllodactylidae) were used as outgroups: *Gymnodactylus darwinii*, *Gymnodactylus geckoides* and *Phyllopezus pollicaris* (Spix, 1825). Specimens were curated at the 'Coleção Herpetológica da Universidade de Brasília' (CHUNB) and 'Museu de Zoologia da Universidade de São Paulo' (MZUSP). Our final dataset consisted of 83 *G. amarali*, 4 *G. geckoides*, 4 *G. darwinii* and 3 *P. pollicaris*. Voucher numbers, localities, and GenBank accession numbers are in Appendix 2.

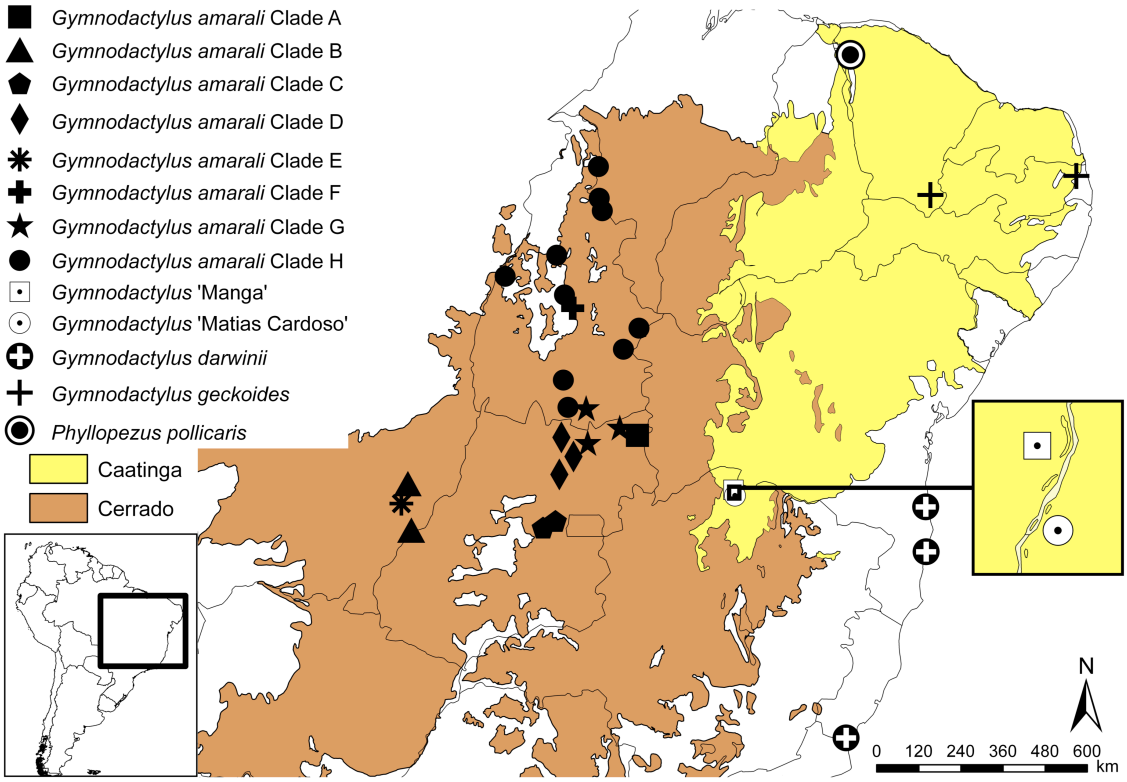


Fig. 1: Partial map of Brazil with *Gymnodactylus* and outgroup samples, in the context of the distribution of the Cerrado and Caatinga biomes. Symbols indicate clades selected by the GMYC analysis. Inset map detail show the populations of 'Manga' and 'Matias Cardoso' separated by the São Francisco River.

2.2 Molecular methods and analyses

We extracted genomic DNA using a modified salting-out technique (Sunnucks & Hales, 1996) and used PCR to amplify fragments of the mitochondrial DNA (mtDNA) cytochrome b (cytb) and the exonic locus Kinesin Family Member 24 (KIF24). Primers and PCR cycle protocols are in Appendix 3. The PCR products were visualized on a 1.5% agarose gel and sequenced using Big Dye v3.1 on an ABI 3130xl at the Flinders Sequencing Facility, SA Pathology. We assembled and visually inspected chromatograms using SEQUENCHER 4.9 (Gene Codes Corporation, Ann Arbor, MI USA). Sequences were codon aligned using MUSCLE (Edgar, 2004) as implemented in MEGA 5.2.2 (Tamura *et al.*, 2011) applying a gap open penalty of 3 and a gap extension penalty of 1. Prior to analyses, we tested for third codon saturation using Xia et al. (2003) index of substitution saturation as implemented in DAMBE5 (Xia, 2013). The index suggested that saturation was negligible (cytb: Iss 0.165 < Issc 0.810, $p<0.001$; KIF24: Iss 0.36 < Issc 0.792, $p<0.001$) and we proceeded with analyses using the complete alignments.

Our molecular hypothesis-testing framework aimed to concomitantly test the monophyly of *G. amarali* and identify possible cryptic lineages within the species, based on the following approach: 1) reconstructing phylogenetic trees with the concatenated dataset using two methods; 2) building a 'species tree' that incorporates individual gene genealogies using a coalescent method; and 3) test the fit of the data from both genes to different evolutionary hypotheses generated by the previous analyses via a Bayesian coalescent species delimitation method.

2.2.1 Phylogenetic reconstructions

All downstream phylogenetic analyses used the partition strategies and models of sequence evolution selected based on the Bayesian Information Criterion (BIC) in PartitionFinder (Lanfear *et al.*, 2012; Lanfear *et al.*, 2014). Partition strategies and evolution models were

separately estimated for the concatenated and individual locus alignments, and for each phylogenetic software used (MrBayes, RAxML or Beast). Selected evolution models and partitions are in Appendix 4.

We used Bayesian inference implemented in MrBayes v3.2.2 (Ronquist *et al.*, 2012), to investigate phylogenetic relationships with the concatenated dataset and separately for each gene fragment. We conducted two independent runs using four parallel Markov Chain Monte Carlo (MCMC) chains for 5 million generations, sampling every 500th generation. Substitution rates, character state frequencies, gamma shape parameters and proportion of invariable sites were all unlinked. We used a minimum acceptable effective sample size (ESS) of 200 for each parameter, and assessed stationarity and convergence of Bayesian analysis respectively by plotting MCMC generations versus the log-likelihood values of the data and checking the potential scale reduction factor in MrBayes. Stationarity and convergence were also visually inspected by plotting likelihood values in Tracer v1.5 (Rambaut & Drummond, 2009). *Phyllopezus pollicaris* was used as outgroup.

We also implemented a maximum likelihood (ML) phylogenetic analysis on the concatenated dataset using RAxML (Stamatakis, 2014), with unlinked partitions as selected by PartitionFinder. We used 1000 bootstrap replicates in a rapid bootstrap analysis, and a thorough search for the best-scoring ML tree.

2.2.2 Species discovery methods and species tree

We explored the performance of two coalescent species discovery methods (sensu Carstens *et al.*, 2013). First, we used spedeSTEM discovery (Satler *et al.*, 2013), a method that uses information theory to compare models of lineage composition through Akaike Information Criteria (AIC) and returns the ranked 'species tree' models. The spedeSTEM software takes as input gene trees that we separately estimated for both genes using RAxML (as above), and converted to ultrametric trees using package *APE* (Paradis *et al.*, 2004) in R v3.0.1 (R Core

Team, 2013). It also requires a $\theta=4N_e\mu$ value that we estimated with Migrate-n v3.6.1 (Beerli & Felsenstein, 2001). We ran Migrate-n using a random starting tree and four multiple Markov chains for 1 x 10^7 generations sampled every 20th generation, discarding 10% as burn-in. Second, we used the Generalized Mixed Yule Coalescent (GMYC), a method especially developed for only one mitochondrial locus (Pons *et al.*, 2006). Using unique haplotypes of cytb (Appendix 2) we built an ultrametric phylogenetic tree in BEAST v1.7.5 (Drummond *et al.*, 2012) required to run the GMYC algorithm. This algorithm estimates the number of "species" by classifying the branching rates of a phylogram as being the result of interspecific or intraspecific lineage branching patterns (Pons *et al.*, 2006).

We implemented two versions of the GMYC algorithm: the originally proposed ML-based calculation in package *splits* (Pons *et al.*, 2006; Fujisawa & Barraclough, 2013), and a Bayesian implementation that accounts for uncertainty in phylogenetic estimation in package *bGMYC* (Reid & Carstens, 2012), using R v3.0.1 (R Core Team, 2013). In BEAST, we ran phylogenetic analysis under a strict molecular clock set to an evolutionary rate of 1.0 (i.e., no attempt to estimate divergence time) considering a coalescent tree with constant population size, using an UPGMA starting tree, and with 1 x 10^7 Markov Chain Monte Carlo (MCMC) generations sampled every 1,000th generation. We implemented three independent runs and combined results using LogCombiner v1.7.5 (Drummond *et al.*, 2012), burning the first 10% of the samples and subsequently used Tracer v1.5 (Rambaut & Drummond, 2009) to check for minimum adequate ESS (200) and visually inspect stationarity and convergence by plotting likelihood values. We summarized the resulting trees into a target maximum clade credibility tree to use in the ML implementation, and alternatively kept 100 random trees for the Bayesian implementation, using TreeAnnotator v1.7.5 (Drummond *et al.*, 2012). For the ML-GMYC we also performed a log-likelihood ratio test of the fitted model against a null model of no distinct species clusters, and calculated AIC-based support values for the species clusters with a $p < 0.05$ (Fujisawa & Barraclough, 2013).

To investigate the phylogenetic relationship between the species retrieved by the GMYC analyses in a multilocus perspective, and also estimate divergence times between the putative species, we ran a *Beast analysis (Heled & Drummond, 2010) as implemented in BEAST v.1.7.5. The lack of *Gymnodactylus* fossils prevented a more robust calibration, and we estimated divergence times based in the putative substitution rate of 2% changes million/year (Johns & Avise, 1998). We used the evolutionary models selected for each locus under a relaxed lognormal molecular clock set for cytb, and the KIF24 evolution rate dependent on cytb. We selected a Yule process prior for the tree using an UPGMA starting tree and performed the analysis with 5 x 10^7 MCMC generations sampled every 1,000th generation. We implemented three independent runs and combined results using LogCombiner v1.7.5 (Drummond *et al.*, 2012), burning the first 10% of the samples. We summarized resulting trees into a target tree using TreeAnnotator v1.7.5 (Drummond *et al.*, 2012), and used Tracer v1.5 (Rambaut & Drummond, 2009) to check for minimum adequate ESS (200) and visually inspect stationarity and convergence.

We also calculated cytb and KIF24 net between-group distances using lineages selected by the GMYC analysis with MEGA 5.2.2 (Tamura *et al.*, 2011). We computed both uncorrected *p*-distances and ML corrected distances with standard error estimates calculated using 1,000 bootstrap replicates.

2.2.3 Bayesian coalescent species delimitation

We used a coalescent approach implemented in the software Bayesian Phylogenetics and Phylogeography (BPP) (Yang & Rannala, 2010) to test the performance of different 'species trees' by assessing their posterior probabilities considering both loci. This method accommodates the species phylogeny, as well as lineage sorting due to ancestral polymorphism, by comparing the posterior probability of an *a priori* user-specified phylogenetic ('species') tree with the posterior probability of all possible variations of the

same tree when branches of a particular node are collapsed (Yang & Rannala, 2010). After initial trials testing different parameters (Appendix 5), we used a gamma prior of G(2,1000) for population size (θs) and the age of the root in the species tree (τ_0), and the Dirichlet prior (Yang and Rannala, 2010: equation 2) for other divergence time parameters. We ran analyses for 5 x 10^5 MCMC generations, taking samples every five and using 1 x 10^4 burn-in generations. To check for consistency of results, we conducted three independent runs, starting at two random tree models, and the fully resolved tree model, using both available reversible-jump MCMC species delimitation algorithms (Yang & Rannala, 2010). We repeated this process for three different 'species trees': 1) the one generated by spedeSTEM, and two considering the GMYC groups – 2) with the *Beast topology and, 3) the tree topology generated by the concatenated dataset (ML and Bayesian analysis).

2.3 Morphological analyses

We performed analyses to evaluate whether divergence patterns based on morphology were concordant with the retrieved molecular lineages. From the 94 samples used in the molecular analyses we had access to 81 preserved museum specimens. Because we did not have access to the same *Gymnodactylus darwinii* specimens, we generated data from other three available specimens (Appendix 2). Thus, our total morphological dataset comprised 84 specimens from all the cryptic and described species. With the aid of a stereomicroscope, a single person (FMCBD) processed all specimens and generated the data to avoid bias. Morphological characters were selected in order to maximise variation among *G. amarali* morphotypes. The data consisted of 21 pholidosis variables and 8 categorical variables (see Appendix 6 for a detailed description of characters). From a total of 2,436 observations (29 characters of 84 specimens), 255 (10.5%) were missing values because of damaged specimens. In multivariate approaches, missing value usually means that the whole case should be omitted, resulting in loss of information (Rubin, 2003) and biased evolutionary estimations (Nakagawa &

Freckleton, 2008). To avoid such problems, we imputed missing values through chained equations using a predictive mean matching algorithm implemented in R package *mice* (Buuren & Groothuis-Oudshoorn, 2011). All morphological analyses were carried out in R v3.0.1 (R Core Team, 2013).

Because specimen-lineage affiliation retrieved by the GMYC and all phylogenetic analyses were exactly the same (see Results 3.2), we could assign each individual to a unique 'species' in the following analyses. In a multivariate space, to statistically classify and predict cases belonging to different groups, one would generally employ a Discriminant Function Analysis (DFA) (Quinn & Keough, 2002). However, the DFA linear analytical process assumes normality, no collinearity, and homoscedasticity; in addition, it cannot be applied when the number of cases is smaller than the number of variables (Quinn & Keough, 2002). Meristic characters are known not to be normally distributed (Houle, 1992), and for some clades we had a maximum number of three individuals. To overcome those limitations, we employed a Support Vector Machine (SVM), which is a sophisticated model-training approach for classifying and predicting sample-affiliation based on learning theory (Schölkopf *et al.*, 2000). The SVM builds a kernel function that maps cases into a high-dimensional space, subsequently finding a "margin" in the hyperspace that maximizes the separation between the groups (Cortes & Vapnik, 1995). Although successfully used in computational biology (Ben-Hur *et al.*, 2008), some areas of molecular biology (Park & Kanehisa, 2003; Xue *et al.*, 2005), and ecological distribution modeling (Kelly *et al.*, 2007; Giovanelli *et al.*, 2010), to the best of our knowledge, this is the first time that SVM is used to investigate morphological segregation in animals.

We performed the SVM analysis using R package *e1071* (Meyer *et al.*, 2014). Initially, we implemented a manual search for the best fine tune parameters for the model, i.e. the ones that minimized the error-rate estimated via cross-validation (Chang & Lin, 2011). We then trained the model using the fine tuned *C*-classification SVM algorithm on the whole

morphological dataset, setting 'species' to be explained by the 29 morphological characters. Lastly, we tested the predictive power of the generated model using the default *predict.svm* function of the package, which predicts case affiliation to groups (individual to 'species') based in the model trained by the SVM. More details about the SVM analysis and implementation are in Appendix 7.

3. Results

3.1 Taxon sampling and molecular data

We sequenced both fragments for all *Gymnodactylus amarali* (n=83) and *G. geckoides* (4) specimens, and downloaded GenBank sequences for *Phyllopezus pollicaris* (3). We did not have access to *G. darwinii* tissue samples and available cytb sequences were obtained from GenBank (n = 4). The aligned cytb fragment was 749 bp long from which 369 were variable sites, and KIF24 was 486 bp with 123 variable sites (i.e. 1,235 aligned base pairs in the concatenated dataset). A few contiguous deletions comprising different numbers of codons were found in KIF24: *P. pollicaris* presented two gaps, one with two codons and another with three codons; the two *Gymnodactylus* populations from SDTF enclaves ('Manga' and 'Matias Cardoso') presented different non-shared patterns of deletions, where 'Matias Cardoso' had two gaps of two codons each, and 'Manga' had only one gap of four codons in another position, the latter shared by *G. geckoides*. All specimens of *G. amarali* presented no deletions.

3.2 Monophyly of Gymnodactylus amarali and cryptic species recognition in the G. amarali species group

All phylogenetic analyses (using both the concatenated dataset and the two genes separately) supported the monophyly of *Gymnodactylus amarali* from the Cerrado region, excluding the two populations from SDTF enclaves ('Manga' and 'Matias Cardoso') (Fig. 2, Appendix 8).

Bayesian and ML phylogenetic analyses of the concatenated dataset returned the same topology (Fig. 2).

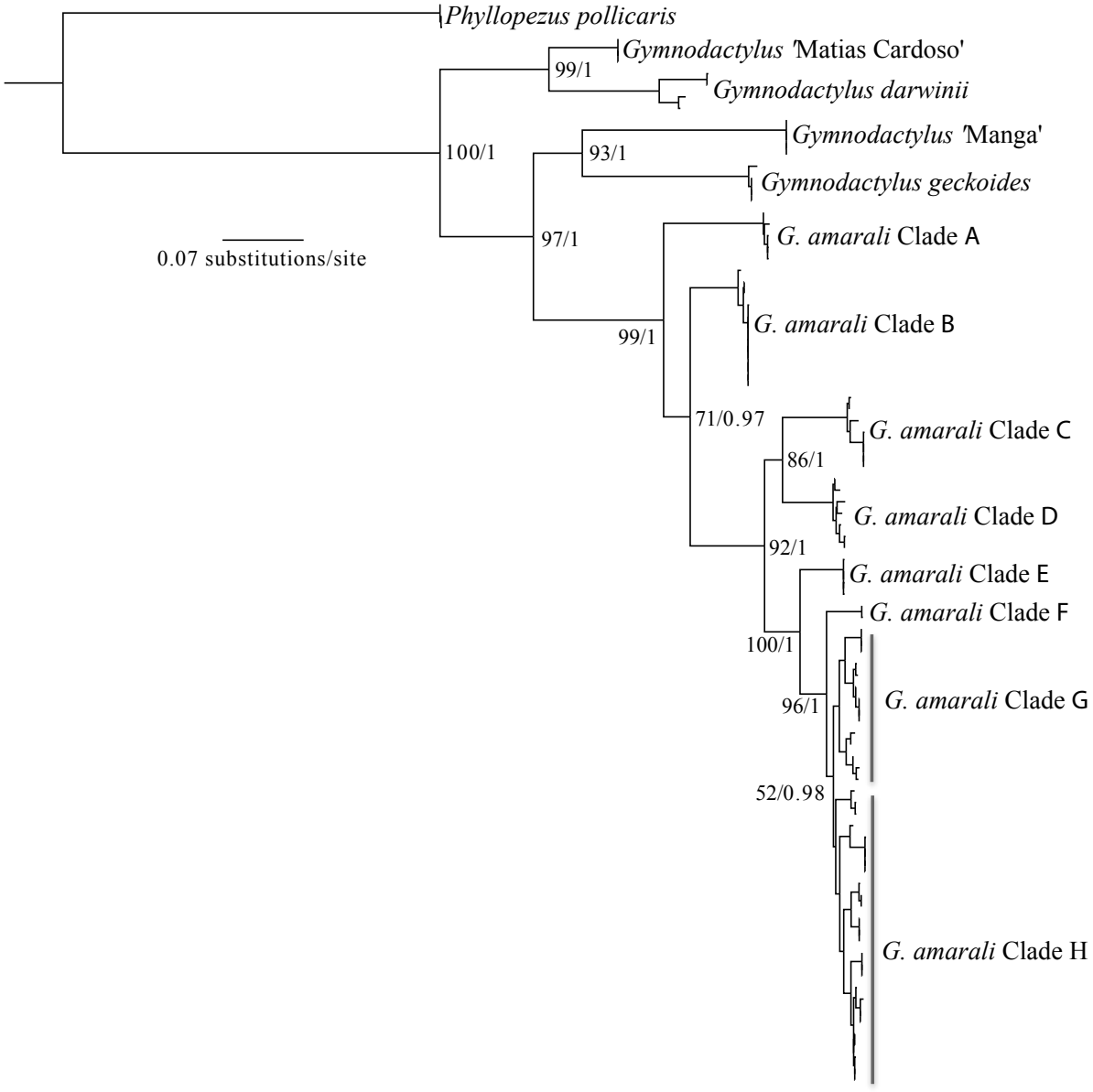


Fig. 2: Maximum likelihood tree of the concatenated dataset for all samples. Bayesian analysis returned the same topology. Numbers in nodes are ML bootstrapping scores/ Bayesian posterior probabilities. Clades A to H refer to *Gymnodactylus amarali* clades identified by GMYC analysis.

The ML-GMYC analysis returned 14 ML entities ('species'), including outgroups, with a confidence interval from three to 36. The log-likelihood ratio test was significant (p = 0.037), i.e. the null hypothesis of a single species was rejected. Most 'species' nodes had $p <$ 0.05 in the AIC based support value of the ML-GMYC analysis, and high posterior probability in the bGMYC (Fig. 3). The 14 entities were: *Phyllopezus pollicaris*, two *Gymnodactylus darwinii* clusters, 'Matias Cardoso', *G. geckoides,* 'Manga', and eight clusters for *G. amarali*, which were named A to H (Fig. 3).

'Matias Cardoso' and 'Manga' were very divergent from the *Gymnodactylus amarali* clusters: cytb uncorrected distances ranged from 19.1% to 21.4% and ML corrected distances from 27.4% to 32.1% for the former and 18.3% to 21.5% and from 25.7% to 32.8%, respectively, for the latter. Cytb levels of uncorrected sequence divergence were lower among *G. amarali* clusters, ranging from 1.8% to 17.5% and ML corrected distances from 2.1% to 23.7% (Table 1). Levels of divergence were lower for KIF24: uncorrected and ML corrected distances respectively ranged from 5.8% to 8.0% and from 6.7% to 9.1% for 'Matias Cardoso', from 5.9% to 8.0% and from 6.7% to 9.1% for 'Manga', and from 0.2% to 4.2% and from 0.2% to 4.7% among *G. amarali* clusters. Interestingly, 'Manga' is more related to *G. geckoides* of the Caatinga, and 'Matias Cardoso' to *G. darwinii* of the Atlantic Rainforest (Fig. 2-4 and Appendix 9).

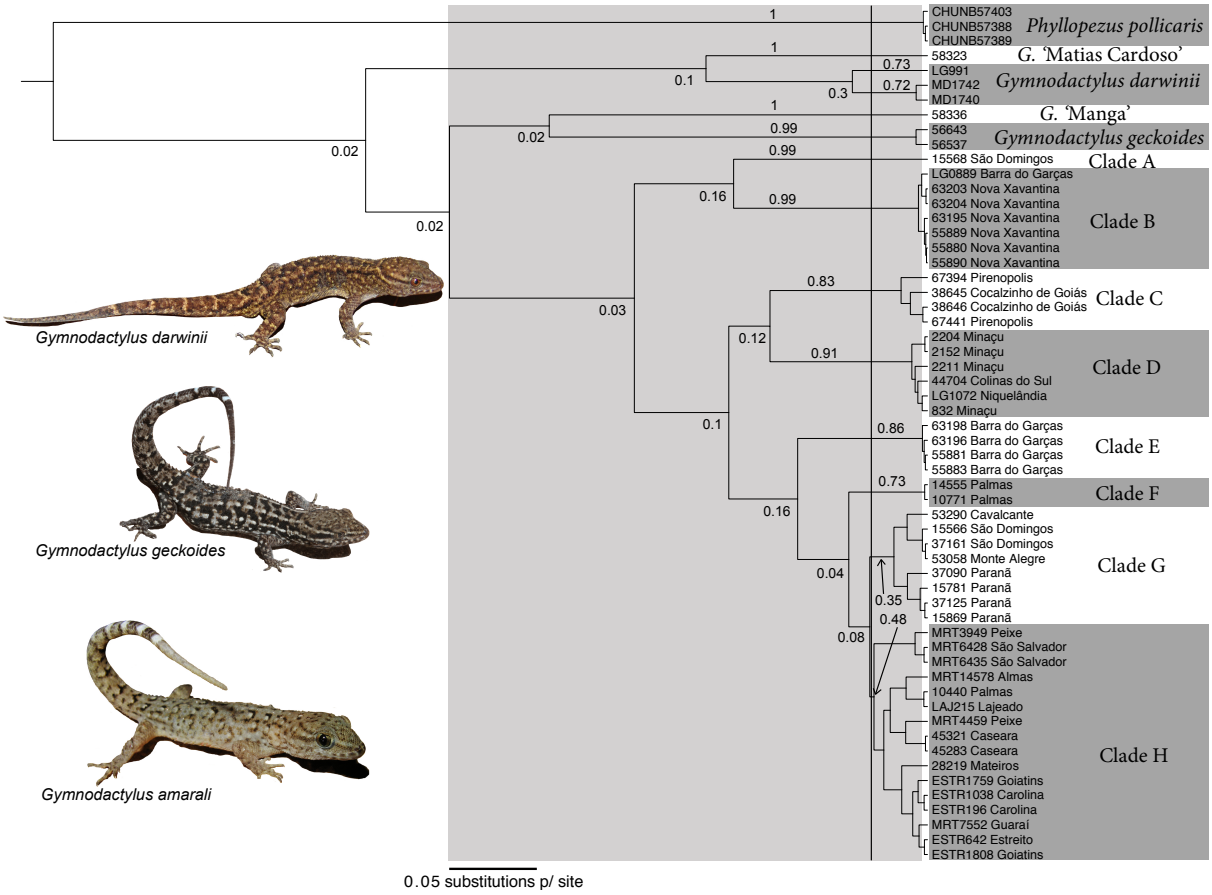


Fig. 3: Ultrametric tree of unique cytb haplotypes. Vertical line represents the limit value for ML species clusters identified by ML-GMYC analysis, and the large grey box represents the confidence interval of species level clusters (3 -16). Numbers below nodes are *p* values of the AIC based support of the ML-GMYC analysis. Posterior probability for each species from bGMYC analysis is shown above branches, or indicated by arrows. Species-level clusters are enclosed by boxes, and *Gymnodactylus amarali* 'candidate species' are named clusters A to H. Photos of *G. darwinii* and *G. geckoides* by Miguel Rodrigues, and *G. amarali* by Daniel Velho.

Table 1: Net among-group distances between GMYC clades for cytb data. ML corrected distances using the Tamura-Nei model are above the diagonal, and uncorrected p-distances below. Standard error estimates, calculated using 1000 bootstrap replicates, are shown in parentheses.

Taxa	*G. amarali* Clade A	*G. amarali* Clade B	*G. amarali* Clade C	*G. amarali* Clade D	*G. amarali* Clade E	*G. amarali* Clade F	*G. amarali* Clade G
G. amarali Clade A	-	0.158 (0.019)	0.237 (0.027)	0.230 (0.026)	0.255 (0.028)	0.233 (0.026)	0.211 (0.024)
G. amarali Clade B	0.122 (0.011)	-	0.220 (0.026)	0.174 (0.021)	0.203 (0.023)	0.192 (0.022)	0.178 (0.022)
G. amarali Clade C	0.163 (0.012)	0.153 (0.013)	-	0.131 (0.017)	0.162 (0.020)	0.157 (0.019)	0.141 (0.018)
G. amarali Clade D	0.161 (0.012)	0.131 (0.012)	0.102 (0.010)	-	0.154 (0.018)	0.150 (0.019)	0.127 (0.016)
G. amarali Clade E	0.175 (0.013)	0.149 (0.013)	0.122 (0.011)	0.119 (0.011)	-	0.122 (0.015)	0.092 (0.012)
G. amarali Clade F	0.166 (0.013)	0.142 (0.012)	0.119 (0.011)	0.115 (0.011)	0.097 (0.010)	-	0.054 (0.010)
G. amarali Clade G	0.150 (0.012)	0.131 (0.011)	0.106 (0.010)	0.099 (0.010)	0.075 (0.009)	0.047 (0.007)	-
G. amarali Clade H	0.144 (0.012)	0.133 (0.011)	0.103 (0.010)	0.100 (0.010)	0.079 (0.009)	0.044 (0.006)	0.018 (0.003)
Manga	0.211 (0.014)	0.188 (0.013)	0.215 (0.014)	0.183 (0.014)	0.215 (0.013)	0.211 (0.014)	0.189 (0.013)
Matias Cardoso	0.194 (0.014)	0.204 (0.014)	0.212 (0.014)	0.202 (0.015)	0.212 (0.015)	0.214 (0.015)	0.191 (0.014)
G. geckoides	0.195 (0.012)	0.183 (0.013)	0.187 (0.013)	0.173 (0.012)	0.199 (0.013)	0.189 (0.013)	0.182 (0.013)
G. darwinii A	0.212 (0.015)	0.208 (0.015)	0.222 (0.015)	0.200 (0.014)	0.223 (0.014)	0.214 (0.015)	0.183 (0.014)
G. darwinii B	0.211 (0.014)	0.204 (0.014)	0.226 (0.014)	0.197 (0.014)	0.222 (0.014)	0.203 (0.014)	0.185 (0.014)
Phyllopezus pollicaris	0.261 (0.015)	0.253 (0.015)	0.239 (0.014)	0.243 (0.015)	0.265 (0.015)	0.256 (0.015)	0.246 (0.015)

Taxa	*G. amarali* Clade H	Manga	Matias Cardoso	*G. geckoides*	*G. darwinii* A	*G. darwinii* B	*Phyllopezus pollicaris*
G. amarali Clade A	0.203 (0.023)	0.310 (0.030)	0.274 (0.029)	0.280 (0.028)	0.315 (0.032)	0.312 (0.033)	0.417 (0.041)
G. amarali Clade B	0.187 (0.022)	0.263 (0.026)	0.299 (0.030)	0.257 (0.026)	0.308 (0.030)	0.294 (0.029)	0.397 (0.040)
G. amarali Clade C	0.139 (0.018)	0.328 (0.032)	0.321 (0.033)	0.270 (0.027)	0.351 (0.037)	0.358 (0.038)	0.369 (0.036)
G. amarali Clade D	0.132 (0.016)	0.257 (0.027)	0.307 (0.034)	0.246 (0.026)	0.300 (0.032)	0.291 (0.031)	0.381 (0.040)
G. amarali Clade E	0.099 (0.013)	0.319 (0.029)	0.312 (0.031)	0.290 (0.029)	0.344 (0.033)	0.335 (0.032)	0.429 (0.041)
G. amarali Clade F	0.050 (0.009)	0.310 (0.030)	0.319 (0.032)	0.272 (0.028)	0.323 (0.033)	0.293 (0.030)	0.403 (0.040)
G. amarali Clade G	0.021 (0.004)	0.269 (0.026)	0.274 (0.028)	0.264 (0.027)	0.265 (0.027)	0.265 (0.027)	0.389 (0.039)
G. amarali Clade H	-	0.281 (0.027)	0.286 (0.029)	0.266 (0.027)	0.285 (0.030)	0.282 (0.029)	0.417 (0.042)
Manga	0.192 (0.013)	-	0.275 (0.028)	0.248 (0.026)	0.278 (0.028)	0.284 (0.029)	0.430 (0.043)
Matias Cardoso	0.192 (0.014)	0.194 (0.014)	-	0.262 (0.026)	0.163 (0.018)	0.172 (0.020)	0.368 (0.035)
G. geckoides	0.180 (0.012)	0.178 (0.013)	0.186 (0.013)	-	0.273 (0.027)	0.270 (0.027)	0.408 (0.041)
G. darwinii A	0.189 (0.014)	0.195 (0.014)	0.126 (0.012)	0.190 (0.014)	-	0.061 (0.009)	0.393 (0.038)
G. darwinii B	0.190 (0.013)	0.199 (0.014)	0.132 (0.012)	0.190 (0.013)	0.055 (0.008)	-	0.403 (0.038)
Phyllopezus pollicaris	0.252 (0.015)	0.266 (0.016)	0.241 (0.016)	0.254 (0.015)	0.248 (0.015)	0.255 (0.015)	-

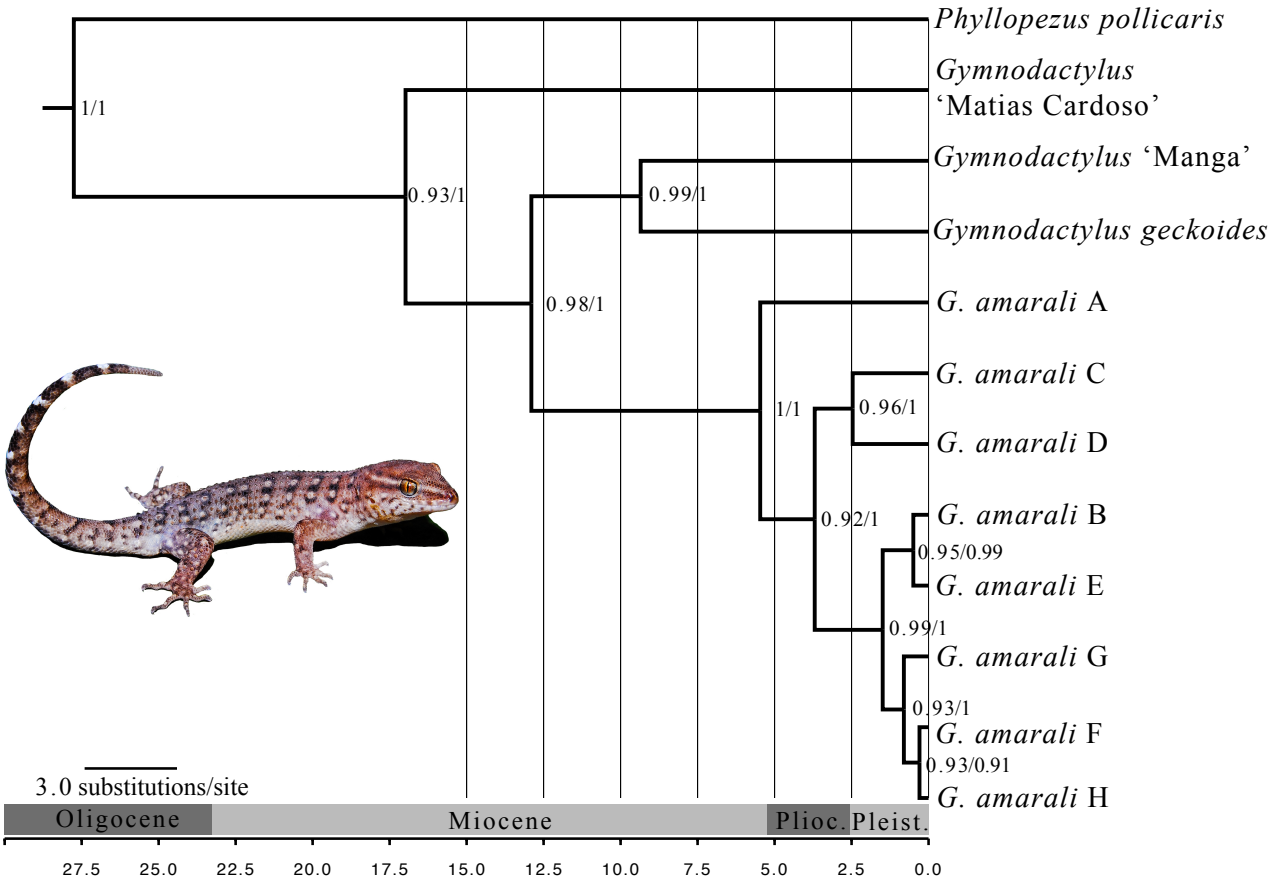


Fig. 4: Phylogenetic relationships and divergence times between *Gymnodactylus* clades estimated using a Bayesian 'species tree' coalescent analysis with *Beast. Numbers inside nodes indicates Bayesian posterior probabilities/ posterior probabilities of the species splits estimated by BPP. Clades A to H refer to *Gymnodactylus amarali* clades identified by GMYC analysis. Photo of *G. amarali* by Daniel Velho.

The *Beast 'species tree' also supports the monophyly of *Gymnodactylus amarali* with 'Matias Cardoso' and 'Manga' nested outside the *G. amarali* clade (Fig. 4). The position of some *G. amarali* clades differed between the *Beast and the concatenated dataset phylogenetic tree (Table 2) because clades that shared specimens from geographically close localities were recovered as sister species (B and E, F and H; Fig.1). Divergence time between the most basal clade of *G. amarali* and remaining clades was ~5 million years (MY) ago, while most other clades diverged within the last 2 MY (Fig. 4). Support was high (>0.95) for most nodes in the *Beast consensus tree. Considering that our main interest was in the relationships among *Gymnodactylus amarali* clades, although we could have used a *G. darwinii* KIF 24 alignment with nothing but missing data, because we only had two loci and the position of *G. darwinii* as the sister group of '*Gymnodactylus* Matias Cardoso' was unlikely to change, we adopted a conservative approach and omitted *G. darwini* from this analysis.

Table 2. Different 'species trees' used in the Bayesian species delimitation analysis (BPP), based on the groups defined by the GMYC analysis. The lowest and highest posterior probability of the model for different BPP runs is shown. Phy = *Phyllopezus pollicaris*, Dar = *Gymnodactylus darwinii*, MaC = 'Matias Cardoso', Gec = *Gymnodactylus geckoides*, Ma = 'Manga', A to H = *G. amarali* clades A to H.

Analysis with concatenated dataset	Resulting species tree	Posterior probabilities of full model
Bayesian and ML	(Phy,((MaC,(DarA,DarB)),((Ma,Gec),(A,(B,((C,D),(E,(F,(G,H))))))))	0.963 - 0.982
*Beast	(Phy,(MaC,((Ma,Gec),(A,((C,D),((E,B),(G,(F,H))))))))	0.891 - 0.914

The spedeSTEM analysis returned 14 groups as the most likely model, but lineage composition was different from previous analyses. Only three of the eight *G. amarali* groups identified by the GMYC were also retrieved (Appendix 10). The posterior probability of the full model calculated by BPP was very low, and no other model showed higher probability (Appendix 11). As such, we considered the GMYC results as the best lineage diversification hypothesis (Fig. 3) and proceeded with morphological analyses and discussion without considering the spedeSTEM results.

The other two lineage relationship hypotheses tested using BPP returned slightly different results considering the posterior probabilities of the fully resolved tree model (Table 2). These results support the placement of individuals within the defined GMYC lineages, considering that models with collapsed versions of the tree had very small posterior probabilities. Thus, combining those lineages in the same species would not reflect the best evolutionary hypothesis from a coalescent perspective.

In summary, the results support the monophyly *of Gymnodactylus amarali* distributed within the Cerrado, and point to the existence of eight well-defined clades that could represent different cryptic species. Furthermore, the two populations in the contact zone between Cerrado and Caatinga (SDTF enclaves), 'Matias Cardoso' and 'Manga', belong respectively to *G. darwinii* and *G. geckoides* species groups, and likely represent cryptic lineages of those two groups.

3.3 Morphological support of lineages

The analysis of morphological data corroborated the retrieved evolutionary lineages, with a very low rate of specimens misidentification (3%) returned by the model prediction. Assignment errors were observed only between *G. amarali* clades B, D and H, where one specimen was incorrectly assigned in each group. All other *G. amarali* clades (A, C, E, F and G), 'Matias Cardoso', 'Manga', and the outgroup species were correctly assigned. From all *G.*

amarali specimens sequenced for cytb and KIF24, only nine (9) were not available for pholidosis and could not be included in the morphological analysis (Appendix 2). The morphological characters can therefore reliably discriminate the seven *G. amarali* lineages, as well as 'Manga' and 'Matias Cardoso', and can be used to diagnose those clades (Appendix 12).

4. Discussion

Biologists have used morphological data for centuries to describe and infer relationships among species. The advent of molecular tools has drastically changed this activity (Wiens, 2007) and molecular data overcame the use of traditional characters to reconstruct lineage relationships. Not surprisingly, the use of molecular tools also became the major approach to recognize cryptic species (Bickford *et al.*, 2007). On the other hand, morphological data are still essential for species description and play an important role to make uncovered cryptic lineages identifiable and available for scientific and conservation purposes (Beheregaray & Caccone, 2007).

Under the Generalized Lineage Concept (de Queiroz, 2007) we presented species hypotheses using two lines of evidence, genetic and morphological, and provided the first example of cryptic species identified by coalescent lineage delimitation analyses in a Cerrado endemic vertebrate. Moreover, no 'species' are shared between the Cerrado and the Caatinga, reinforcing a still understudied evolutionary pattern between these two biomes (Werneck & Colli, 2006; Werneck, 2011).

4.1 Monophyly of Gymnodactylus amarali in the Cerrado

Molecular and morphological results corroborate the monophyly of *Gymnodactylus amarali* within the Cerrado region and the existence of multiple cryptic lineages within this taxon. The two populations sampled in the contact zone between Cerrado and Caatinga, SDTF enclaves,

do not belong to the *G. amarali* species group but are recovered as sister groups to *G. geckoides* ('Manga') and *G. darwinii* ('Matias Cardoso') respectively.

Our samples cover the whole distribution of the species, which suggests that the herein recognized *G. amarali* species complex is the only *Gymnodactylus* lineage to inhabit the Cerrado biome. We found no evidence for the existence of *G. geckoides* populations in the Cerrado as previously suggested by Pellegrino and collaborators (2009). These authors found one specimen in 'Barra do Garças' (one of our sampled locations in central Cerrado) to have an identical karyotype to *G. geckoides*, and suggested it was a potential relict population of *G. geckoides*. Considering the heterogeneous landscape that characterises the region of 'Barra do Garças' and that two different clades of *G. amarali* (B and E) inhabit the area, the presence of an additional species remains to be tested. Additional sampling and chromosome data from 'Barra do Garças' as well as a complete phylogeny of the genus are critical to understand geographical patterns of karyotypic evolution in *Gymnodactylus*. Testing phylogenetic hypotheses for the genus would require a multilocus dataset for all five currently described taxa, as well as for different cryptic species recognized for *G. darwinii* (Pellegrino *et al.*, 2005) and *G. amarali* (this study). Nonetheless, our results suggest a (*G. darwinii*, (*G. geckoides*, *G. amarali*)) topology.

4.2 Cryptic species in the Gymnodactylus amarali species complex

We uncovered eight cryptic lineages within *Gymnodactylus amarali*. Levels of mtDNA divergence between recovered clades (2.1% to 23.7%, Table 1) were higher than usually observed between species of lizards or other vertebrate groups (Avise *et al.*, 1998; Fouquet *et al.*, 2007; Oliver *et al.*, 2009) and consistent with those recently reported for cryptic lineages of New World geckos (Gamble *et al.*, 2012; Werneck *et al.*, 2012a).

Regarding lineage relationships, the phylogenetic and 'species tree' methods resulted in different placements of some clades (Table 2). This is a common issue comparing 'gene

trees' and 'species trees' (Pamilo & Nei, 1988; Degnan & Rosenberg, 2009) and can probably be suppressed by the use of more loci (but see Degnan & Rosenberg, 2006).

Despite this topological disagreement between the results of the two methods, the Bayesian species delimitation analysis performed by BPP resulted in similar posterior probabilities for both hypotheses (Table 2). From a coalescent perspective, these results imply that every option where different clades are collapsed into one would be a worse diversification scenario. A similar result was found for geckos of the *Hemidactylus fasciatus* complex, where BPP also returned very limited differences between different phylogenetic hypotheses (Leaché & Fujita, 2010). Based on a series of simulations and different *a priori* phylogenetic trees, the authors concluded that when divergent populations are placed as sister taxa, large divergences among the species are artificially created, and the algorithm interprets those divergences as speciation events. This suggestion likely reflects the trend of our results, and reinforces the placement of the eight different clades as 'candidate species' in the *G. amarali* complex (Fig. 4, Appendix 13). Another simulation study showed that even when only one individual is sampled, the accuracy of BPP using two loci is almost as good as using 10 loci (Camargo *et al.*, 2012). Divergence times and migration rates also did not substantially influence the performance of the algorithm (Camargo *et al.*, 2012), and we believe our results depict a real trend in the evolution of *G. amarali*, in spite of our limited number of loci and the fact that we had as few as 2 individuals for at least one 'species'.

Incomplete lineage sorting (Degnan & Rosenberg, 2009) and gene flow among lineages (Leaché *et al.*, 2014) are also known to affect 'species tree' reconstruction. These two processes would have an effect on the input phylogenetic tree to be used in BPP, interfering with the species delimitation algorithm (Leaché & Fujita, 2010; Yang & Rannala, 2010). On the other hand, concordant reciprocal monophyly between lineages in different gene trees is not essential when delimiting species (Knowles & Carstens, 2007) and the morphological analyses supported the placement of individuals in the clades using a different

dataset. This suggests that our hypotheses testing framework was strong enough to support the recovered clades as distinct evolutionary lineages. Finally, given that the multi-species coalescent is more likely to recover a pattern of diversification than gene genealogies (McVay & Carstens, 2013), we suggest that the *Beast topology is a better provisionary arrangement for the *G. amarali* 'candidate species', and discuss the evolution of the group considering this phylogenetic hypothesis below.

4.3 Evolution of Gymnodactylus amarali in the Cerrado

This study was not aimed at reconstructing the biogeographic history of *Gymnodactylus amarali* but it has enabled a number of inferences about the evolution of the species in the Cerrado. The fact that *G. amarali* inhabiting the Cerrado form a monophyletic group suggests that they diversified within this biome, likely influenced by landscape evolution of the Cerrado (Prado *et al.*, 2012; Werneck *et al.*, 2012a). Moreover, the two populations from SDTF enclaves in the border of Cerrado are clearly distinct lineages, supporting the view that *G. amarali* does not occur in SDTF physiognomies. The transition between the Caatinga and Cerrado is not marked by topographical barriers (Ab'Sáber, 1974, 1998), indicating that environmental filters are probably responsible for the absence of *G. geckoides* from the Cerrado and the absence of *G. amarali* from the Caatinga (Colli *et al.*, 2003a).

Traditional hypotheses for the origins of the high Neotropical biodiversity include vegetation refugia created by Pleistocene climatic fluctuations (Williams & Vanzolini, 1966; Vanzolini, 1968a), a scenario suggested to account for the diversification of forest animals (Haffer, 1969; Moraes-Barros *et al.*, 2006; Fouquet *et al.*, 2012) and SDTF endemic *Drosophila* species (Moraes *et al.*, 2009; Franco & Manfrin, 2013). However, recent studies point towards ancient events of lineage diversification for Cerrado vertebrates, dating back to the Neogene (Prado *et al.*, 2012; Werneck *et al.*, 2012a; Giugliano *et al.*, 2013). Three main Neogene vicariant events were proposed to influence the diversification of endemic

herpetofauna in the Cerrado: the formation of a latitudinal temperature gradient in the early Palaeogene, the Miocene marine transgression, and the final uplift of the Central Brazilian Plateau in the Miocene-Pliocene transition (Colli, 2005). The latter event is responsible for the major compartmentalization currently observed in the Cerrado landscape: a mosaic of plateaus separated by valleys excavated by river drainages (King, 1956; Ab'Sáber, 1998). It is possible that an ancestral *G. amarali* lineage was distributed over the landscape before the compartmentalization, which is consistent with our estimated divergence times starting at approximately five MY ago (Fig. 4). This assumption is also corroborated by estimated divergence times for other Cerrado vertebrates (Prado *et al.*, 2012; Giugliano *et al.*, 2013) and Neotropical geckos (Werneck *et al.*, 2012a).

Clades A, D and C are distributed in different plateaus and show deep divergences (Fig. 1 and 4). Clade E lizards were collected in a plateau ~630 m above sea level that is only ~50 km apart from the ~300 m valley inhabited by lizards from a sister clade (clade B). These two groups showed a cytb genetic distance of 20% and only 0.02% for the nuclear gene KIF24. Similarly, sister clades F and H, (5% divergent at cytb and 0.08% at KIF24), were also distributed across different elevations (650 m and mostly 150-350 m, respectively). The above results might reflect ancient events of gene flow during early stages of landscape compartmentalization, a pattern still apparent in the slower evolving nuclear gene (Appendix 8). Gene flow estimation using statistical phylogeography are beyond the aims of this study and would be an ideal tool to evaluate such a pattern (Knowles & Maddison, 2002).

4.4 Status of Gymnodactylus amarali species group and conservation in the Brazilian Cerrado

Using 'species tree' reconstructions based on molecular data and a Bayesian species delimitation method we identified ten novel clades in the genus *Gymnodactylus* in a pattern concordant with morphology. In addition, the low assignment error (3%) of the SVM analysis

shows that these lineages are morphologically distinguishable. We acknowledge that prompt descriptions of identified cryptic species are needed to avoid delays of taxonomic availability (Schlick-Steiner *et al.*, 2007), but assessing morphological diagnostic characters is essential when proposing taxonomic revisions (Bauer *et al.*, 2011). Because species' descriptions can be time consuming and laborious, we argue that the uncovered clades should be referred to as 'candidate species' for conservation delineation and management purposes (Whittaker *et al.*, 2005; Bickford *et al.*, 2007). Moreover, knowledge on the evolutionary relationship between newly discovered lineages can efficiently improve potential conservation initiatives (Diniz-Filho *et al.*, 2013). To our knowledge, only two other studies (Giugliano *et al.*, 2013; Recoder *et al.*, 2014) focused on squamate cryptic species recognition in the Cerrado using both molecular and morphological datasets. We suggest that using both types of data should be a priority in studies on squamate diversity in the Cerrado.

The rate of species description in the Brazilian Cerrado is biased by unequal distribution of sampling efforts across the biome (Diniz-Filho *et al.*, 2005; Diniz-Filho *et al.*, 2008). Even large-bodied cryptic squamate species were recently described following expeditions to previously unsampled regions (Nogueira & Rodrigues, 2006; Giugliano *et al.*, 2013). Sampling in remote areas is an expensive activity (Costa *et al.*, 2010) and we suggest that funding should be direct towards research projects that combine faunal inventories with collection of data useful for assessing putative cryptic diversification. This is especially important if we seek to understand the evolution of the endemic biota and to inform conservation management strategies.

Chapter 3

Cryptic species in the Neotropics: Coalescent species delimitation of Cerrado endemic lizards using anchored phylogenomics

1. Introduction

The so-called Linnean shortfall was initially envisioned to describe the lack of taxonomic knowledge and the associated impediments it brings to biological studies (Brown & Lomolino, 1998), but it soon developed to acknowledge the problem in terms of conservation practices (Possingham *et al.*, 2007). Similarly, there is a lack of knowledge concerning the distribution of species and its associated issues, the Wallacean shortfall (Whittaker *et al.*, 2005). Both shortfalls are certainly correlated and can strongly restrict conservation actions, especially in biodiversity hotspots (Bini *et al.*, 2006). Nonetheless, even if both issues can be overcomed, information concerning phylogenetic relationships would still be lacking for many species in the world, and conservation actions hindered by this Darwinian shortfall (Diniz-Filho *et al.*, 2013). Not surprisingly, the three shortfalls are highly accentuated in the Neotropics, where high biodiversity and remoteness of several areas make it hard to surpass these knowledge gaps (Kier *et al.*, 2005; Balian *et al.*, 2007; Schipper *et al.*, 2008; Silva *et al.*, 2014).

1.1 Recent molecular approaches to uncover cryptic species

Overcoming the above-mentioned shortfalls is not a trivial task. Recent attention has been devoted to the discovery and description of cryptic species which, given the correct technical and analytical tools, has the potential to confront the three shortfalls at once (Beheregaray & Caccone, 2007; Bickford *et al.*, 2007; Pfenninger & Schwenk, 2007). Species descriptions have traditionally relied on morphological information and on data about the geographical distribution of closely related species. This practice changed substantially with the advent of automated DNA sequencing technology (Scheffers *et al.*, 2012). For metazoans, it initiated with basic assumptions of discrimination based on mtDNA phylogenies and associated barcoding data (Pereira *et al.*, 2008). The field has largely advanced in the last few years due to increases in computational power to run more demanding analyses (O'Meara, 2010; Ence &

Carstens, 2011; Rittmeyer & Austin, 2012). This technological advance allowed the development of new analyses based on different parametric and non-parametric approaches. In a recent review, Carstens *et al.* (2013) listed most of these procedures and associated software, as well as their use and advantages. Above all, coalescent species delimitation seems to be the most suitable approach to provide clear information for taxonomy and assist conservation practices (Fujita *et al.*, 2012).

Despite those methodological improvements, delimitating and describing new cryptic species is a laborious and slow process (Winston, 1999), demanding not only the use of genetic data but also access to several populations and closely related species to allow reliable comparisons. While traditional taxonomy can be subjective and highly dependent on traditions of certain groups of scientists (Isaac *et al.*, 2004; Mace, 2004), the use of genetic data associated with modern coalescent-based species delimitation methods provides objectivity to the practice of taxonomy. These methods can provide insights about the underlying evolutionary patterns associated with the speciation process, since they take into consideration the species phylogeny itself, uncertainties in gene tree topology and branch lengths, and random fluctuations in the coalescent process (Zhang *et al.*, 2011). Nonetheless, the use of coalescent species delimitation is still in its infancy compared to many other phylogenetic methods, with only a few examples of its application in Neotropical studies (e.g., Crawford *et al.*, 2010; Pinzon-Navarro *et al.*, 2010; Camargo *et al.*, 2012; Ceccarelli *et al.*, 2012; Gamble *et al.*, 2012; Domingos *et al.*, 2014; Gehara *et al.*, 2014; Smith *et al.*, 2014a). Moreover, to the best of my knowledge, there is only one study of Neotropical organisms (on rainforest birds; Smith *et al.*, 2014a) that combined Next-Generation Sequencing (NGS) data with coalescent species delimitation methods.

1.2 Cerrado cryptic squamate species

The Cerrado, arguably the richest savannah in the world (Castro *et al.*, 1999; Furley, 1999; Oliveira & Marquis, 2002), has been the focus of an increasing number of squamate species descriptions in recent years (e.g., Colli *et al.*, 2003c; Colli *et al.*, 2003b; Rodrigues *et al.*, 2007; Rodrigues *et al.*, 2008; Colli *et al.*, 2009; Ribeiro *et al.*, 2009; Strüssmann & Mott, 2009; Pinna *et al.*, 2010; de Freitas *et al.*, 2011; Teixeira *et al.*, 2013). Similarly, several studies have recently recognised cryptic lineages among previously described Cerrado endemics (Gamble *et al.*, 2012; Giugliano *et al.*, 2013; Domingos *et al.*, 2014; Recoder *et al.*, 2014). In fact, the simple activity of visiting and collecting biological specimens in previously unexplored regions of the Cerrado might result in the discovery of new species (Diniz-Filho *et al.*, 2006; Diniz-Filho *et al.*, 2008), with even large-bodied lizard species recently found and described (Nogueira & Rodrigues, 2006; Giugliano *et al.*, 2013). The above mentioned advances in DNA sequencing technology and analytical methods can substantially help to overcome this hidden diversity problem, as they might enable the recognition of new entities even if a single individual per population (putative species) is available, provided that enough loci from this individual are sequenced (Yang & Rannala, 2010).

Although there has been a recent increase in the number of studies describing biogeographic (Nogueira *et al.*, 2011), community ecology (Nogueira *et al.*, 2009) and distribution patterns of Cerrado squamates (Costa *et al.*, 2007), there is still a noticeable knowledge gap in terms of evolutionary patterns of Cerrado endemic squamates (Werneck, 2011). In a previous effort, we described the existence of eight cryptic lineages within *Gymnodactylus amarali* using two loci together with morphological data (Chapter 2), but relationships between a few lineages were not clear in terms of their geographic distribution (Chapter 2, Table 2). Increasing the number of loci used for species delimitation can assist in the detection of different lineages in non-model organisms (Pante *et al.*, 2015), avoid problems associated with not sampling the possibly different demographic histories retrieved

by different loci (Garrick *et al.*, 2015), and reduce the probability of errors when using coalescent species delimitation methods (Zhang *et al.*, 2011; Zhang *et al.*, 2014). This improved efficacy of using larger sets of nuclear loci has been shown applying both simulated and empirical data to coalescent species delimitation (Camargo *et al.*, 2012; Rannala & Yang, 2013).

As such, sampling a larger number of nuclear loci could notably improve the power of species delimitation analyses when investigating highly diverse lineages, such as is the case of *G. amarali*, and potentially of other Cerrado endemic lizards. Below I describe patterns and lines of evidence that suggest the presence of cryptic species on each of our focal species.

1.3.1 Cryptic species in the Gymnodactylus amarali complex

The genus *Gymnodactylus* Spix, 1825 (Phyllodactylidae) is restricted to the Caatinga, Cerrado and Atlantic Forest domains in Brazil (Vanzolini, 1953a, b, 2004, 2005; Cassimiro & Rodrigues, 2009). As the generic name suggests, species on this genus are characterised by gymnodactily, i.e. free fingers without dilatations, and wide undivided subdigital lamellae (Vanzolini, 1968b, 1982). Currently, there are five described species in the genus, but this number is certainly underestimated because of the presence of cryptic species, already reported for *G. darwinii* (Pellegrino *et al.*, 2005) and *G. amarali* (Chapter 2) groups. Cryptic species are likely present in the type-species of the genus as well, *G. geckoides* Spix 1825 (Vanzolini, 2004), and at least one new species was already reported (Chapter 2).

Gymnodactylus amarali Barbour, 1925 is mainly distributed in the northern portion of the Cerrado (Vanzolini, 2005; Chapter 2), and high morphological variation (Vanzolini, 1953a, 2005; Cassimiro & Rodrigues, 2009) as well as extensive karyotypical variation has been reported among its populations (Pellegrino *et al.*, 2009).

1.3.2 Potential cryptic species in Micrablepharus atticolus

The genus *Micrablepharus* Boettger, 1885 (Gymnophtalmidae) comprises two species of eyelid-less lizards: *M. atticolus* Rodrigues, 1996 and *M. maximiliani* (Reinhardt and Lütken, 1861). While *M. maximiliani* ranges in the Caatinga, Pantanal and Cerrado, *M. atticolus* is a Cerrado endemic but the two are rarely found in sympatry (Santos *et al.*, 2014). The study species *M. atticolus* is also found on peripheral Cerrado enclaves within the Amazon forest (Gainsbury & Colli, 2003) and whether these are relict populations isolated during Pleistocene climatic fluctuations or represent recent colonisations is still controversial (Santos *et al.*, 2014). High levels of mtDNA genetic diversity have been found within this taxon, including divergence times of ~3 My estimated between the most basal population and others (Santos *et al.*, 2014). This suggests the possibility of cryptic species, although the authors did not explore this idea. In addition, variation on chromosome diploid number was found among five populations of *M. atticolus* (2n = 50-53) and these different karyotypes appear to be geographically structured (Yonenaga-Yassuda & Rodrigues, 1999).

1.3.3 Potential cryptic species in Tropidurus itambere

Of all three genera in this study, *Tropidurus* Wied-Neuwied, 1825 (Tropiduridae) is the one with more taxonomic problems. Probably because of its relatively high abundance (Wiederhecker *et al.*, 2003; Faria & Araujo, 2004), and for being conspicuously found in many urban areas (Rodrigues, 1987; Martins *et al.*, 1999), *Tropidurus* lizards have been the subject of investigations in many fields, such as population and community ecology (Van-Sluys, 1993; Van-Sluys, 1997; Vitt & Zani, 1998; Van-Sluys *et al.*, 2004), physiology and performance (Kohlsdorf *et al.*, 2004; Kiefer *et al.*, 2005; Kohlsdorf & Navas, 2007), morphological evolution (Vitt *et al.*, 1997; Kohlsdorf *et al.*, 2001), and island biogeography (Schluter, 1984). Many species of *Tropidurus* are well-studied in terms of ecology (see Carvalho, 2013 for a comprehensive literature review) and have also been the subject of

systematics (Harvey & Gutberlet, 2000; Frost *et al.*, 2001), biogeography and conservation studies (Carvalho, 2013), making them a broadly studied group of organisms in South America.

Nonetheless, an updated taxonomic review of the genus *Tropidurus* is urgently warranted. There are presently 23 described species of *Tropidurus* distributed across different South American biomes (Carvalho, 2013). The last and still the most comprehensive taxonomic review of the genus was made by Rodrigues (1987), and accounted only for the species in the *T. torquatus* group south of the Amazon (11 species). In this study, Rodrigues used "mite pockets" as one of the most important character for diagnosing species. This character has been since extensively used in the taxonomy of the genus and is probably one of the main sources of taxonomic confusion. Studies regarding the evolution of mite pockets in *Tropidurus* have not been carried out and such structures might represent plesiomorphic characters not suitable for discrimination of species-level taxa. Given the high species diversity of the genus, the taxonomy of *Tropidurus* is still meagre considering the amount of other published studies for this iconic group of South American lizards.

The taxon targeted in this thesis, *Tropidurus itambere* Rodrigues, 1987, is usually found on rock outcrops (cerrado rupestre) (Van-Sluys, 1997), and is diagnosed by the presence of a deep mite pocket in the inguinal region, and another on the side of the neck (Rodrigues, 1987). However, there is great variation in the depth and format of mite pockets between populations, as well as on anterior and posterior limb lengths that might be associated with different degrees of specialization for the use of rock crevices. Although there is clear variation among localities, the underlying pattern of variation does not seem to be geographically clustered (Domingos and Colli, pers. obs.). Therefore, it is currently unknown if all populations that can be morphologically assigned to *T. itambere* actually belong to the same species.

1.4 Aims and hypotheses of Chapter 3

We previously reported the existence of cryptic species for *G. amarali* (Chapter 2), and highlighted that some degree of morphological differentiation exists among the different lineages. Additionally, considering the above-mentioned patterns of genetic and karyological variation observed for *M. atticolus*, and the differences in morphology observed for *T. itambere*, a few general expectations about patterns of diversification of Cerrado lizards can be drawn: (1) the observed variation in different biological attributes between populations should reflect species-level divergences; (2) assuming that lizards are, generally, organisms with restricted dispersal capacities, and given the complex geomorphological history of the Cerrado (discussed in Chapter 1), these divergences might be associated with different allopatric speciation events caused by Neotectonic processes (Werneck, 2011; Werneck *et al.*, 2012a); (3) morphological variation might be associated with local adaptation to environmental variables, and the genetic divergence among populations might reflect this pattern (Glor *et al.*, 2003; Nosil *et al.*, 2005; Thorpe *et al.*, 2008); and (4) patterns of genetic divergence might simply be more influenced by geographical distance among populations.

Based on these expectations, I applied a coalescent species delimitation method and a series of phylogenetic analyses to test whether the morphological and geographical variation observed among Cerrado lizards populations reflect genome-wide divergences and, as such, indicate the existence of species-level cryptic diversity. I clarify phylogenetic relationships among lineages, tested and discussed the possible advantages and shortcomings of applying coalescent species delimitation methods to a robust dataset of ~400 loci obtained by an anchored hybrid enrichment phylogenomic approach (Lemmon *et al.*, 2012). I used concatenated Maximum Likelihood and Bayesian phylogenetic analyses to infer relationships among lineages within each species. While species trees based on the multispecies coalescent may yield better accuracy than traditional concatenated approaches (Heled & Drummond, 2010; Xi *et al.*, 2014), they are computationally demanding and most methods can not be

applied to phylogenomic datasets (O'Neill *et al.*, 2013). Thus, I used two multispecies coalescent approaches that incorporate information from previously estimated gene trees in a coalescent framework (Song *et al.*, 2012) and compared their results to the concatenated estimations. To the best of my knowledge, this is the first evolutionary study applying next-generation sequencing data to investigate such questions in Cerrado vertebrates.

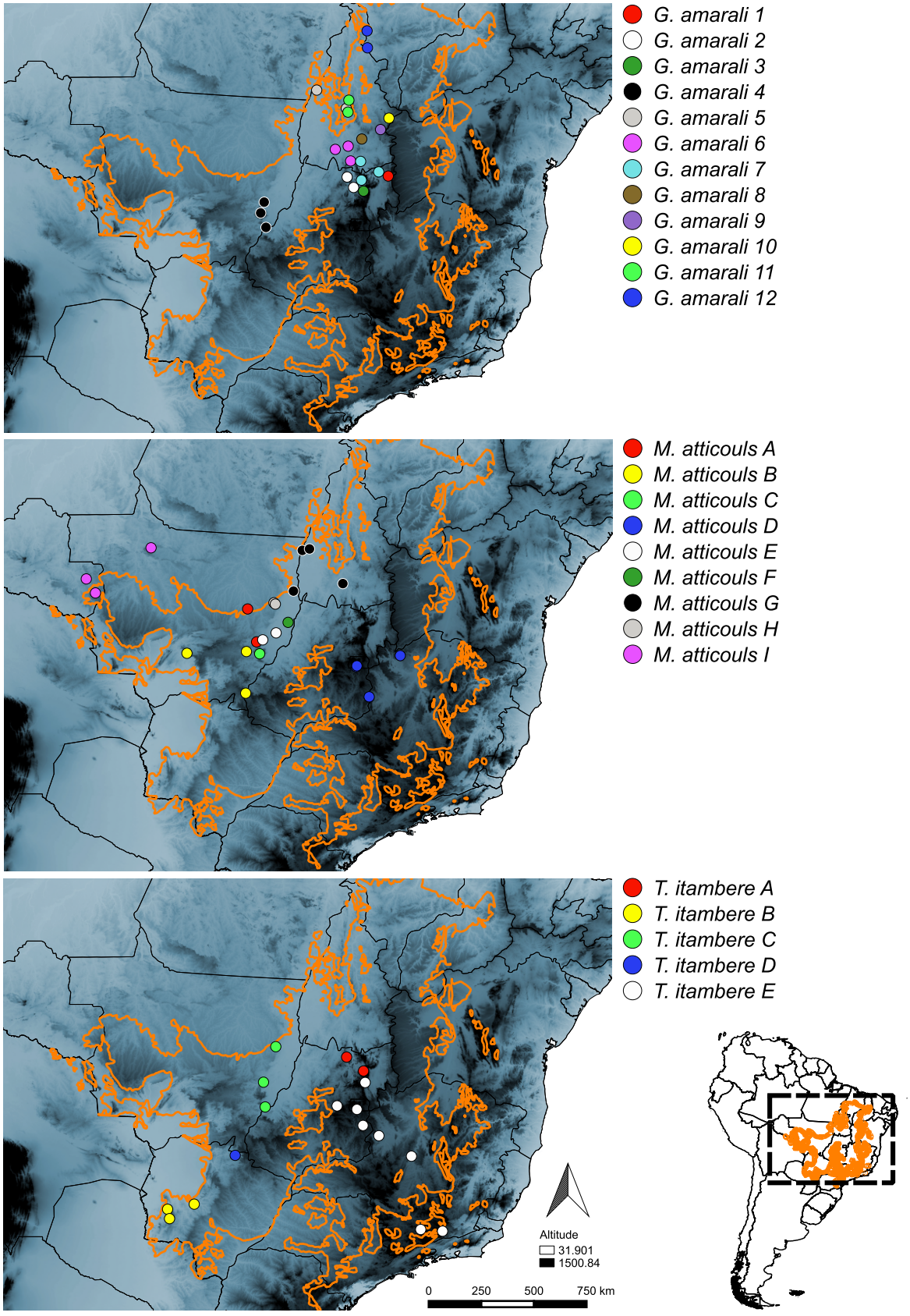
2. Material and Methods

2.1 Sampling and genetic protocols

We obtained mtDNA (cytb) data for 170 *Gymnodactylus amarali* and outgroup individuals from 34 localities, 139 *Micrablepharus atticolus* and outgroups from 35 localities, and 103 *Tropidurus itambere* and outgroups from 29 localities, using the methods and primers described in Chapter 2. Our choice of outgroup taxa was based on Chapter 2 for *G. amarali*, Castoe and colleagues (2004) for *M. atticolus*, and Frost and colleagues (2001) for *T. itambere*. As such, we used *G. darwinii* 'Matias Cardoso' as outgroup for *G. amarali*, *M. maximiliani* for *M. atticolus*, and *Uranoscodon superciliosus* for *T. itambere*. Additionally, we included specimens of other three *Tropidurus* species (*T. hispidus*, *T. oreadicus* and *T. torquatus*) in the *Tropidurus* analyses, in an attempt to identify *T. itambere*-like individuals that could actually be more closely related to other *Tropidurus* species.

A subset of individuals from different clades (putative cryptic species) was selected based on geographic criteria and Neighbour-Joining phylogenetic trees (Appendices 14, 15 and 16), in an attempt to maximise the sampling of divergent clades and to cover as much of the known range of nominal species (Fig. 1).

Fig. 1: Partial map of Brazil with *Gymnodactylus amarali* (upper), *Micrablepharus atticolus* (middle) and *Tropidurus itambere* (bottom) sample sites in the context of the distribution of the Cerrado (orange outline). Different colours indicate clades (species hypotheses) used in BPP analyses.


G. amarali 1
G. amarali 2
G. amarali 3
G. amarali 4
G. amarali 5
G. amarali 6
G. amarali 7
G. amarali 8
G. amarali 9
G. amarali 10
G. amarali 11
G. amarali 12
M. atticouls A
M. atticouls B
M. atticouls C
M. atticouls D
M. atticouls E
M. atticouls F
M. atticouls G
M. atticouls H
M. atticouls I
T. itambere A
T. itambere B
T. itambere C
T. itambere D
T. itambere E
Altitude
31.901
1500.84
0
250
500
750 km

These individuals (n = 94) were used for the phylogenomic section and paired-end sequenced after being used in an anchored hybrid enrichment protocol (Lemmon *et al.*, 2012) on an Illumina HiSeq platform. Briefly, the probe kit implemented in the anchored phylogenomics (AP) consists of 512 loci and was designed to capture 1500 bp; 240 bp relate to the probe itself, and 25 tiles of 120 bp overlapping every 5 bp covers ~700 bp in the flanks of either side of the probe. This translates into 56,664 probes targeting 122,800 bp of the genome to capture ~800,000 bp per individual. Not all these data will be captured, enriched and sequenced in the same way for every individual, and the final dataset will be as such smaller than the theoretical one. The kit was developed from the genomes of five model species from different classes of vertebrates (*Danio*, *Xenopus*, *Homo*, *Gallus*, and *Anolis*). Information about individuals sampled for cytb and AP, and their collection sites is provided in Appendices 17, 18 and 19.

The bioinformatics of the protocol started with a quasi-*de novo* approach to assemble the sequences, by matching the reads to probe region sequences. Loci with assemblies containing fewer than 100 reads per loci (minimum coverage) were excluded (Appendix 20). Afterwards, sites with <10 called bases were marked as N in the consensus. For each unambiguous site the probability of sequencing error was calculated, and sites with more than 1% probability were converted to the consensus base. True polymorphism was expressed using IUPAC notation. Additional information about the genetic protocol, assembling, data filtering, base calling and alignment can be found in Lemmon *et al.* (2012).

2.2 Phylogenetic relationships

One of the difficulties when using the AP approach for diploid organisms is that the complementary versions of the nuclear genomes are sequenced at the same time. Heterozygous positions can be easily identified during bioinformatics workflow (Lemmon *et al.*, 2012), but assembling two genomes from one diploid sample (phasing) can be challenging

(Sousa & Hey, 2013). Although branch length estimation can be influenced if heterozygous sites are ignored during analysis (Lischer *et al.*, 2014), it is unlikely that the estimated topology of the trees will be influenced (Wiens & Morrill, 2011; Lischer *et al.*, 2014; Pyron *et al.*, 2014). Thus, we inferred relationships within each species using IUPAC notation of ambiguous codes for the heterozygous sites. All analyses were run using the high performance computer facilities (HPCF) *Colossus*, a centralised supercomputer at Flinders University, or using *Phoenix*, a HPCF based at the Molecular Ecology Lab at Flinders University. *Colossus* HPC has 1,160 cores and 4.25TB of RAM, whereas *Phoenix* has 40 cores and 512GB of RAM.

While software that estimate models of sequence evolution have been around for quite some time (Nylander, 2004; Posada, 2008), estimating partitions of the data has been substantially based on researchers decision and not on statistical approaches. In the genomics era, partitioning the data is not a trivial task, and PartitionFinder (Lanfear *et al.*, 2012; Lanfear *et al.*, 2014) probably provides the best available option to overcome this issue. PartitionFinder selects the best-fit substitution model for each possible combination of user-defined partitions (each locus in our case), calculates the log-likelihood sum of each possible combination of these partitions, and ranks them using information-theoretic metrics (i.e. Bayesian Information Criterion or Akaike Information Criterion). Users have the option to choose PhyML (Guindon *et al.*, 2010) or RAxML for the likelihood calculations, using exhaustive or greedy heuristic searches (PhyML, Lanfear *et al.*, 2012) or two types of hierarchical clustering (RAxML, Lanfear *et al.*, 2014). Using the first two can be computationally unfeasible for large data sets, and the authors suggest using the latter when the dataset comprises hundreds of loci (Lanfear *et al.*, 2014). As such, for each species the partition strategy was selected by PartitionFinder v1.1.1 (Lanfear *et al.*, 2012) using the RAxML relaxed clustering algorithm under the 10% search default condition (Lanfear *et al.*, 2014), and all downstream phylogenetic analyses used the estimated partition schemes.

We implemented Maximum Likelihood (ML) phylogenetic analyses in RAxML v8.1.1 (Stamatakis, 2014) using rapid hill-climbing searches, and estimated bootstrap support values using 1000 replicates with the RELL bootstrap option (Minh *et al.*, 2013). We also ran phylogenetic analyses using Bayesian inference implemented in Exabayes v1.2.1 (Aberer *et al.*, 2014). Starting from a parsimony tree, we conducted two independent runs with four parallel Markov Chain Monte Carlo (MCMC) chains for at least 1 million generations (sampled every 500th), and set to automatically stop when the average standard deviation of split frequencies was below 0.05 (indicating good convergence). We used a minimum acceptable effective sample size (ESS) of 200 for each parameter and checked the potential scale reduction factor (PSRF, ~1.0) using the "postProcParam" and "extractBips" programs distributed with Exabayes v1.2.1. Branch lengths were linked across partitions, while substitution rates, character state frequencies, gamma shape parameters and proportion of invariable sites were all unlinked. An extended majority-rule consensus tree was obtained using the "consense" program distributed with Exabayes v1.2.1, discarding 25% of the initial samples as burn-in.

The GTR model with gamma shape distribution and invariant sites (GTRGAMMAI) was used on all partitions for the Bayesian analyses. This was carried out because over parameterising (over-fitting) the evolution model on Bayesian analyses has little influence in the resulting topology (Huelsenbeck & Rannala, 2004), especially when numerous and long loci are used (Lemmon & Moriarty, 2004), and to avoid highly intense computations. As RAxML can only implement the GTR model, the same strategy was adopted for the ML analyses.

Species tree reconstructions based on the coalescent are very computationally intensive, and such methods are apparently unable to deal with phylogenomic datasets (Leache & Rannala, 2011; O'Neill *et al.*, 2013; Pyron *et al.*, 2014). These species-tree approaches may have difficulties to identify the correct topology over competing hypotheses (Lischer *et al.*,

2014), and even show decreasing resolution and lineage support as more loci are included (O'Neill *et al.*, 2013). Nonetheless, a few methods may overcome this limitation (Kubatko *et al.*, 2009; Liu *et al.*, 2009), mainly because they incorporate already estimated gene trees and treat them under coalescent models. Indeed, the coalescent species-tree methods STAR (Liu *et al.*, 2009) and NJst (Liu & Yu, 2011) have performed well using phylogenomic datasets (Pyron *et al.*, 2014). We used the web-server STRAW (Shaw *et al.*, 2013) to estimate both STAR and NJst species-trees for the three study taxa. Individual gene trees were generated using RAxML v8.1.1 (Stamatakis, 2014) performing 100 rapid bootstrap inferences and a thorough ML search, under a GTR evolution model with gamma shape distribution.

For the sake of comparability we used exactly the same data for all phylogenetic analyses. Accordingly, all loci for which the selected outgroup was not captured (sequenced) by AP were excluded from analyses, so that individual gene trees used for STAR and NJst could be generated (even thought they could have been used as missing data for the ML and Bayesian concatenated analyses). Thus, the final dataset for the phylogenetic analyses differed slightly from that used for the coalescent species delimitation described below, for which all loci were used (Appendices 21, 22 and 23).

Finally, it was not the aim of this Chapter to estimate divergence times between lineages, but rather to delimit them, so we could use this information in Chapter 4 and for future taxonomic purposes. Divergence times are estimated and discussed in Chapter 4.

2.3 Coalescent species delimitation

BPP (Yang & Rannala, 2010) rapidly became one of the most used species delimitation software (Carstens *et al.*, 2013), especially because of its powerful coalescent approach to delimit species when a phylogenetic hypothesis is presented. In brief, it collapses the branches of the phylogenetic hypothesis inputted by the user and compares the posterior probability of the full (tree) hypothesis with all other versions of the tree with collapsed branches.

Moreover, the most recent version of the software, BPP v3 (Yang & Rannala, 2014), implements a new species tree estimation algorithm: a MCMC proposal based on the nearest-neighbor interchange (NNI) algorithm to change the species tree topology. Although the program still requires *a priori* phylogenetic hypothesis, BPP v3 can simultaneously change the topology, estimating a species tree, and run the reversible-jump MCMC species delimitation algorithm. Consequently, the full analyses would deliver the posterior probabilities of the species tree hypothesis, and of the delimited coalescent species hypotheses.

Carstens *et al.* (2013) defined BPP as a species "validation" method because it could not assign individuals to unknown *a priori* groups, and was restricted to the user-inputted tree. Because of this new feature that allows BPP to change the topology of the user-inputted species tree, the *a priori* phylogenetic hypothesis would have no influence in the species delimitation, since different species trees are being tested during the run (Yang & Rannala, 2014). In these terms, it is reasonable to assume that BPP now presents a "mixed" approach in terms of species "discovery" and "validation". While it will not specifically separate individuals into different species, these individuals can be assigned to many exclusive species hypotheses (clades), making the use of a species discovery method redundant. Being so, we believe BPP provides a strong coalescent species delimitation approach, which can handle the amount of data used in this study. We separated individuals into monophyletic groups (species hypotheses) based on the estimated phylogenetic trees and on the geographic distribution of clades (Fig. 1). Specimen affiliation to these clades was the same among all four estimated phylogenies for all species, i.e. specimen assignment to a species hypothesis agreed among all phylogenetic analyses.

After initial trials testing different parameters (as in Chapter 2), we used a gamma prior of ~G (2,1000) for population size (θs) and the age of the root in the species tree (τ_0), and the Dirichlet prior (Yang & Rannala, 2010: Equation 2) for other divergence time parameters. We

ran analyses for 5 x 10^5 MCMC generations, taking samples every five and using 1 x 10^4 burn-in generations. Because we had a considerable amount of gaps and ambiguous sites, and to make sure we were getting consistent results, we ran analyses using a few different options: 1) using both available reversible-jump MCMC species delimitation algorithms (Algorithms 0 and 1, Yang & Rannala, 2010), and 2) using or not the "cleandata" option. In BPP cleandata = 1 means the program will remove all columns in the alignment which have gaps or ambiguity characters, and cleandata = 0 means that those will be used in the likelihood calculation. To check for consistency of results, for each analyses type we conducted at lest two independent runs starting at random tree models.

BPP can handle any number of missing individuals per loci; thus, we excluded all individuals with more than 30% missing data for a given locus from this locus alignment. We empirically decided on this threshold after visually inspecting the data and running initial BPP trials: most individuals with more than 30% missing data also had a very high number of ambiguous and undetermined sites, i.e. they were mainly low quality captures. Excluding them from the loci alignments was then necessary to avoid an undue influence of these individuals on the likelihood calculations.

BPP and spedeSTEM are the only two coalescent-based species "validation" methods (*sensu* Carstens *et al.,* 2013) where the user inputs their species hypothesis and the software tests them. While there is a "spedeSTEM discovery" option as well, it is not possible to use it with our data because it requires at least two individuals assigned per species (which is not our case), otherwise spedeSTEM will return positive log-likelihood values and the analysis will simply not work. Following similar procedures to Chapter 2, we tried using spedeSTEM in the same trees used to generate the STAR and NJst species trees, by transforming them in ultrametric trees using the package *ape* (Paradis *et al.*, 2004) in R v3.0.1 (R Core Team, 2013). For unknown reasons spedeSTEM was not able to read those trees, even if different smaller subsets were used. The only other option to try and test spedeSTEM with our data

would be using Beast or Exabayes to estimate time-calibrated ultrametric trees of every single loci, which would take an immense amount of time.

2.4 Summary statistics

We calculated population genetics summary statistics for each AP locus (Appendices 21, 22 and 23) using Arlequin v3.5 (Excoffier & Lischer, 2010) and, for comparative reasons, obtained the same statistics for cytb from individuals used in the AP protocol (Table 1). Watterson's θ provides an unbiased way of estimating the population mutation rate ($\theta = 4N_e\mu$) from the infinite-site equilibrium relationship between the number of segregating sites and the sample size for non-recombining DNA (Watterson, 1975), while the population pairwise nucleotide diversity (θ_π) is estimated from the infinite-site equilibrium relationship between the mean number of pairwise differences (Tajima, 1983). Tajima's *D* is a statistical test calculated as the difference between the mean number of pairwise differences and the number of segregating sites (Tajima, 1989), which are expected to be the same when scaled and in a neutrally evolving population, hence large deviations from zero may be caused by changes in population size or natural selection (Tajima, 1989).

We also calculated cytb net between-group distances using the delimited species with MEGA 5.2.2 (Tamura *et al.*, 2011). We used all individuals for which cytb was available (Appendices 17, 18 and 19) and computed both uncorrected *p*-distances and ML corrected distances with standard error estimates calculated using 1,000 bootstrap replicates.

3. Results

3.1 Sampling and population genetic summary statistics

Out of 170 *Gymnodactylus amarali* and outgroup specimens sequenced for cytb, we chose and performed AP sequencing for 32, but only 26 were actually captured (Appendix 17). For *Micrablepharus atticolus* and outgroups those numbers were 139 cytb samples, 28 used for

AP, and 27 captured (Appendix 18); whereas *Tropidurus itambere* and outgroups had 103 cytb samples, 34 used for AP, and 30 captured (Appendix 19). The final AP alignment after cleaning and pruning contained 415 loci and it was 590,398 bp long for *G. amarali*, 394 loci with 575,495 bp for *M. atticolus*, and 383 loci with 538,171 bp for *T. itambere*. The amounts of gaps or undetermined (missing) sites in the final alignments were 1.8% for *G. amarali*, 3.1% for *M. atticolus*, and 4.5% for *T. itambere*. With the exception of a few individuals, the number of reads per loci (coverage) and the number of loci above the coverage threshold did not vary substantially (Appendix 20). Among species, *M. atticolus* had the highest average coverage compared to the more similar coverage obtained for *G. amarali* and *T. itambere* (Table 1). Cytb alignments were 749 bp long for *G. amarali*, 692 bp for *M. atticolus*, and 801 bp for *T. itambere*.

All loci were polymorphic, although number of polymorphic sites per locus varied extensively (Appendices 21, 22 and 23). Population genetics summary estimates did not vary substantially among species, but *M. atticolus* Watterson's θ and θ_π estimates were slightly lower compared to the other two species. As expected, estimates were generally much lower for the AP dataset compared to the faster-evolving mitochondrial cytb (Table 1). Tajima's D estimates were not significantly different from zero (see Table 2 in Tajima, 1989).

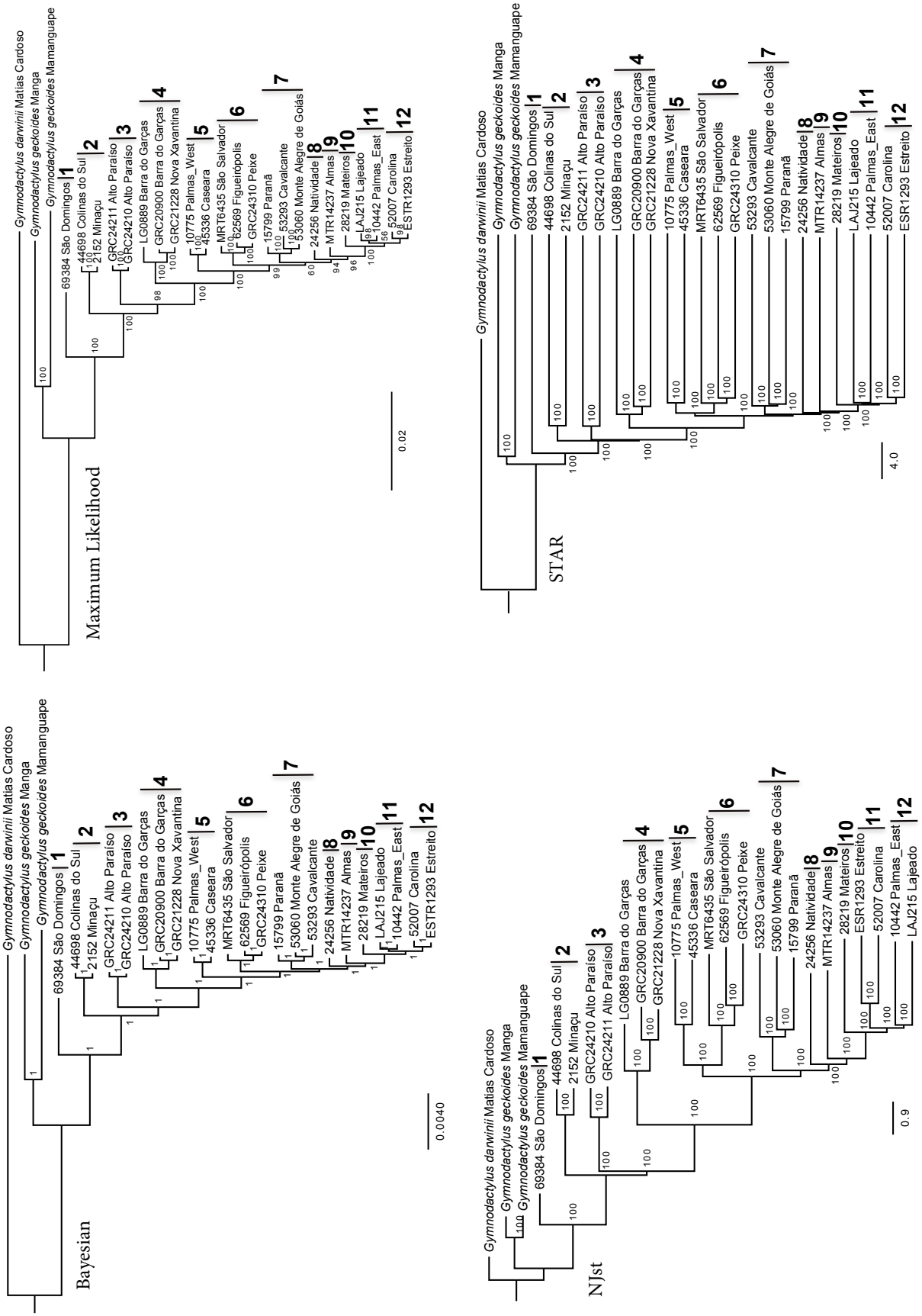
Table 1: Mean values (±SD) of population genetics summary statistics (Watterson's θ, pairwise nucleotide diversity (θ_π), and Tajima's *D)* from AP and cytb alignments. Cytb statistics were calculated from the same individuals used for AP. Also shown are the coverage statistics for the AP dataset. All statistics were calculated after excluding outgroups.

Species (AP loci number)	AP					Cytb		
	Watterson's θ	Nucleotide diversity θ_π	Tajima's *D*	Average coverage across loci	Loci passing coverage threshold	Watterson's θ	Nucleotide diversity θ_π	Tajima's *D*
Gymnodactylus amarali (415)	7.91 (4.03)	5.26 (3.30)	-1.31 (0.58)	2498.19	388.57	70.44	66.18	-0.24
Micrablepharus atticolus (394)	6.16 (2.82)	3.58 (1.69)	-1.48 (0.50)	3525.95	381.46	37.74	33.65	-0.43
Tropidurus itambere (383)	7.09 (3.63)	5.19 (3.03)	-1.02 (0.54)	2705.58	374.81	55.14	59.12	0.29

3.2 Phylogenetic relationships

The major inferred clades were all strongly supported (Bayesian posterior probability = 1, Bootstrap values > 70) by all approaches for all three taxa. Nonetheless, differences between the concatenated approaches (Bayesian and ML) and the coalescent approaches are expected, especially for short internodes (Pyron *et al.*, 2014), and a few topological differences were observed among the four estimated trees for the three species. For *G. amarali*, a main topological difference was observed among the concatenated and the coalescent trees between clades 5 and 6, which are recovered as sister species in the coalescent trees but not in the concatenated trees (Fig. 2).

Fig. 2: Phylogenetic relationships among *Gymnodactylus amarali* lineages estimated by Bayesian, Maximum Likelihood, and coalescent methods NJst and STAR. Numbers in nodes denote posterior probabilities for the Bayesian analyses, and bootstrap scores for all others. Grouping numbers refer to clades used on BPP coalescent species delimitation analyses.

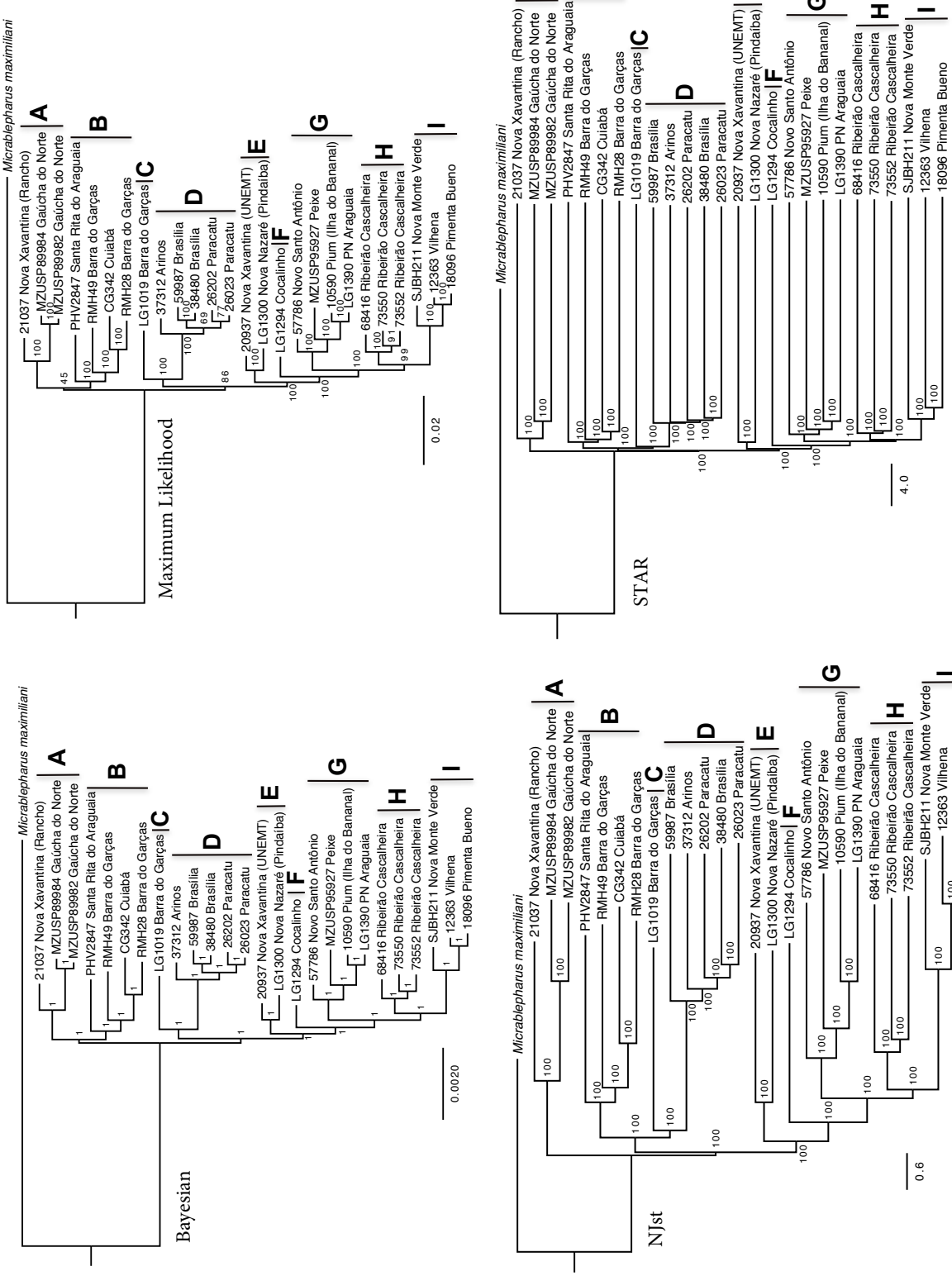

Gymnodactylus darwinii Matias Cardoso
Gymnodactylus geckoides Manga
Gymnodactylus geckoides Mamanguape
69384 São Domingos 1
44698 Colinas do Sul 2
2152 Minaçu
GRC24211 Alto Paraíso 3
GRC24210 Alto Paraíso
LG0889 Barra do Garças 4
GRC20900 Barra do Garças
GRC21228 Nova Xavantina
10775 Palmas_West 5
45336 Caseara
MRT6435 São Salvador 6
62569 Figueirópolis
GRC24310 Peixe
15799 Paraná 7
53060 Monte Alegre de Goiás
53293 Cavalcante
24256 Natividade 8
MTR14237 Almas 9
28219 Mateiros 10
LAJ215 Lajeado
10442 Palmas_East 11
52007 Carolina
ESTR1293 Estreito 12
Bayesian
0.0040
Maximum Likelihood
0.02
STAR
4.0
NJst
0.9

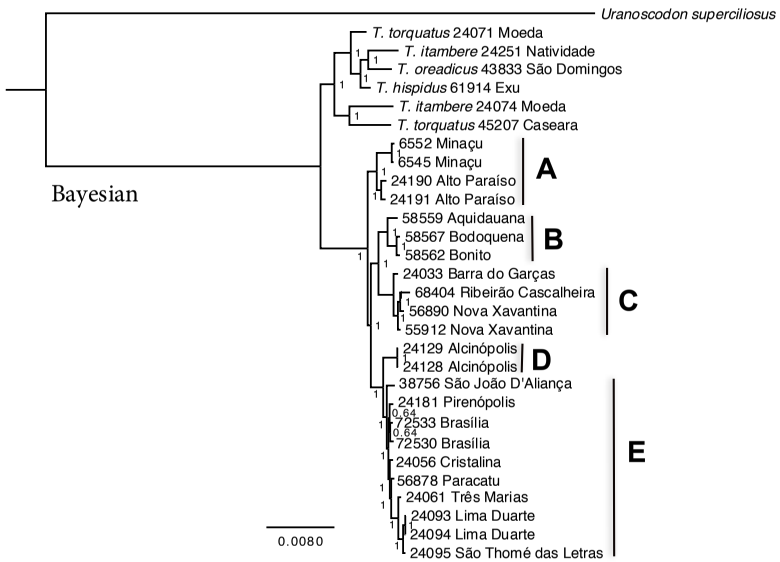
Another main topological difference between the two approaches is seen for *M. atticolus* regarding the position of clade B, which is sister to clade A in the concatenated trees, and sister to clades C-D in the coalescent trees (Fig. 3). Similarly, for *T. itambere*, clade A is sister to all other species in the concatenated trees, whereas clades B-C are sister to A-E in the coalescent trees (Fig. 4).

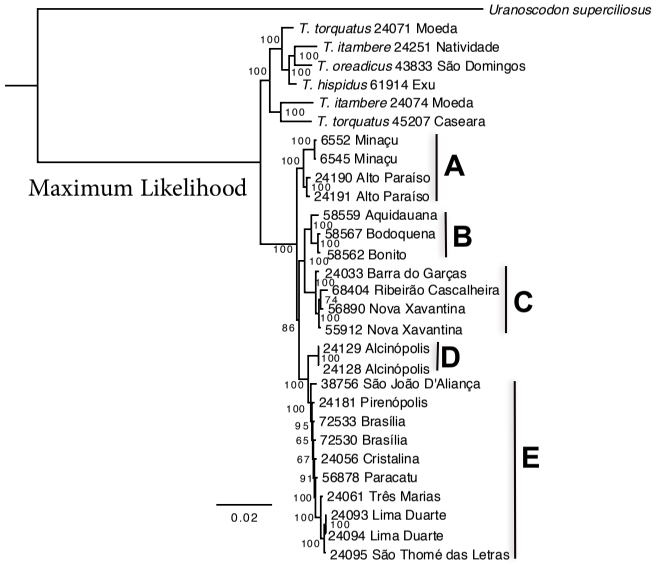
All Bayesian analyses converged before 1 million generations, usually around 2 x 10^5 iterations for *Gymnodactylus* and *Micrablepharus*, but as high as 5 x 10^5 for *Tropidurus*, a result probably attributed to the larger number of outgroup species used in the latter analysis.

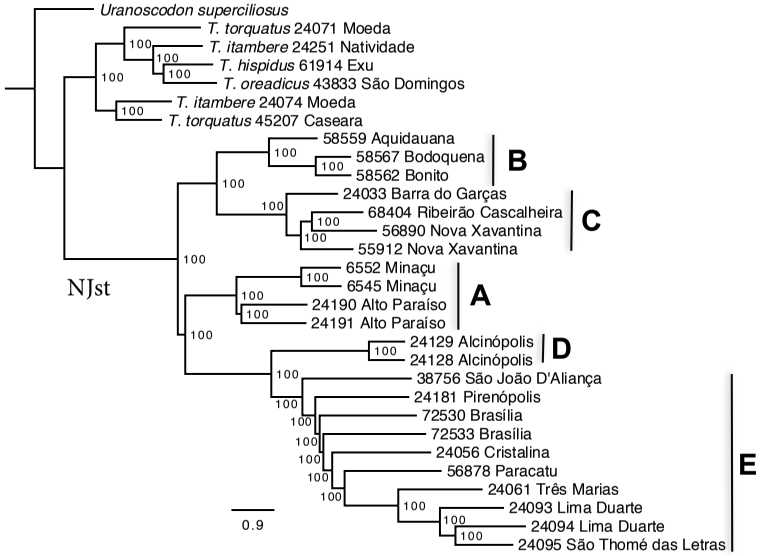
Fig. 3: Phylogenetic relationships among *Micrablepharus atticolus* lineages estimated by Bayesian, Maximum Likelihood, and coalescent methods NJst and STAR. Numbers in nodes denote posterior probabilities for the Bayesian analyses, and bootstrap scores for all others. Grouping letters refer to clades used on BPP coalescent species delimitation analyses.

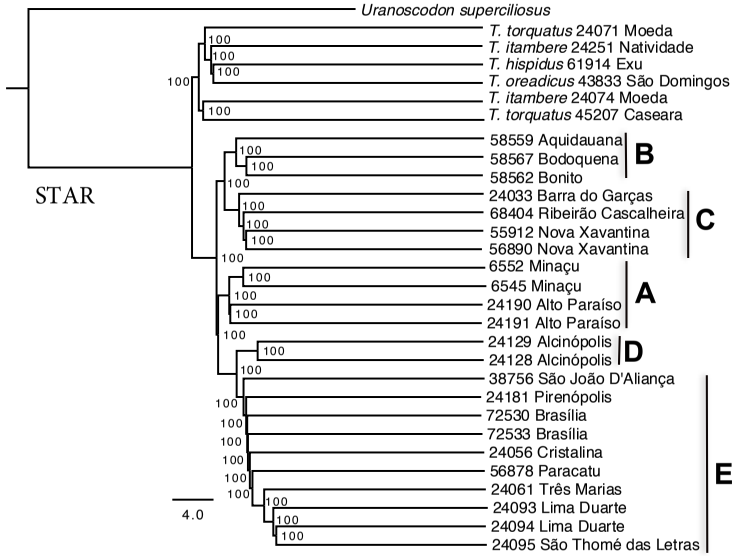
Fig. 4: Phylogenetic relationships among *Tropidurus itambere* lineages estimated by Bayesian, Maximum Likelihood, and coalescent methods NJst and STAR. Numbers in nodes denote posterior probabilities for the Bayesian analyses, and bootstrap scores for all others. Grouping letters refer to clades used on BPP coalescent species delimitation analyses.


Micrablepharus maximiliani
21037 Nova Xavantina (Rancho)
MZUSP89984 Gaúcha do Norte
MZUSP89982 Gaúcha do Norte
PHV2847 Santa Rita do Araguaia
RMH49 Barra do Garças
CG342 Cuiabá
RMH28 Barra do Garças
LG1019 Barra do Garças
37312 Arinos
59987 Brasília
38480 Brasília
26202 Paracatu
26023 Paracatu
20937 Nova Xavantina (UNEMT)
LG1300 Nova Nazaré (Pindaíba)
LG1294 Cocalinho
57786 Novo Santo Antônio
MZUSP95927 Peixe
10590 Pium (Ilha do Bananal)
LG1390 PN Araguaia
68416 Ribeirão Cascalheira
73550 Ribeirão Cascalheira
73552 Ribeirão Cascalheira
SJBH211 Nova Monte Verde
12363 Vilhena
18096 Pimenta Bueno
A
B
C
D
E
F
G
H
I
Bayesian
0.0020
Maximum Likelihood
0.02
STAR
4.0
NJst
0.6


Uranoscodon superciliosus
T. torquatus 24071 Moeda
T. itambere 24251 Natividade
T. oreadicus 43833 São Domingos
T. hispidus 61914 Exu
T. itambere 24074 Moeda
T. torquatus 45207 Caseara
6552 Minaçu
6545 Minaçu
24190 Alto Paraíso
24191 Alto Paraíso
A
58559 Aquidauana
58567 Bodoquena
58562 Bonito
B
24033 Barra do Garças
68404 Ribeirão Cascalheira
56890 Nova Xavantina
55912 Nova Xavantina
C
24129 Alcinópolis
24128 Alcinópolis
D
38756 São João D'Aliança
24181 Pirenópolis
72533 Brasília
72530 Brasília
24056 Cristalina
56878 Paracatu
24061 Três Marias
24093 Lima Duarte
24094 Lima Duarte
24095 São Thomé das Letras
E
Bayesian
0.0080


Uranoscodon superciliosus
T. torquatus 24071 Moeda
T. itambere 24251 Natividade
T. oreadicus 43833 São Domingos
T. hispidus 61914 Exu
T. itambere 24074 Moeda
T. torquatus 45207 Caseara
6552 Minaçu
6545 Minaçu
24190 Alto Paraíso
24191 Alto Paraíso
A
58559 Aquidauana
58567 Bodoquena
58562 Bonito
B
24033 Barra do Garças
68404 Ribeirão Cascalheira
56890 Nova Xavantina
55912 Nova Xavantina
C
24129 Alcinópolis
24128 Alcinópolis
D
38756 São João D'Aliança
24181 Pirenópolis
72533 Brasília
72530 Brasília
24056 Cristalina
56878 Paracatu
24061 Três Marias
24093 Lima Duarte
24094 Lima Duarte
24095 São Thomé das Letras
E
Maximum Likelihood
0.02


Uranoscodon superciliosus
T. torquatus 24071 Moeda
T. itambere 24251 Natividade
T. hispidus 61914 Exu
T. oreadicus 43833 São Domingos
T. itambere 24074 Moeda
T. torquatus 45207 Caseara
58559 Aquidauana
58567 Bodoquena
58562 Bonito
B
24033 Barra do Garças
68404 Ribeirão Cascalheira
56890 Nova Xavantina
55912 Nova Xavantina
C
6552 Minaçu
6545 Minaçu
24190 Alto Paraíso
24191 Alto Paraíso
A
24129 Alcinópolis
24128 Alcinópolis
D
38756 São João D'Aliança
24181 Pirenópolis
72530 Brasília
72533 Brasília
24056 Cristalina
56878 Paracatu
24061 Três Marias
24093 Lima Duarte
24094 Lima Duarte
24095 São Thomé das Letras
E
NJst
0.9


Uranoscodon superciliosus
T. torquatus 24071 Moeda
T. itambere 24251 Natividade
T. hispidus 61914 Exu
T. oreadicus 43833 São Domingos
T. itambere 24074 Moeda
T. torquatus 45207 Caseara
58559 Aquidauana
58567 Bodoquena
58562 Bonito
B
24033 Barra do Garças
68404 Ribeirão Cascalheira
55912 Nova Xavantina
56890 Nova Xavantina
C
6552 Minaçu
6545 Minaçu
24190 Alto Paraíso
24191 Alto Paraíso
A
24129 Alcinópolis
24128 Alcinópolis
D
38756 São João D'Aliança
24181 Pirenópolis
72530 Brasília
72533 Brasília
24056 Cristalina
56878 Paracatu
24061 Três Marias
24093 Lima Duarte
24094 Lima Duarte
24095 São Thomé das Letras
E
STAR
4.0

3.3 Coalescent species delimitation

All BPP runs included outgroup taxa, which could also be recovered as different species or not. Accordingly, the number of species in the (user inputted) full model (Table 2) includes the outgroup species, and thus is higher than the number of ingroup clades shown in the phylogenetic trees (Figs. 2, 3 and 4). BPP results indicated the existence of several cryptic lineages within all three taxa, using both the complete alignment (no cleandata) and the one excluding sites with missing and ambiguous characters (cleandata) (Table 2).

Table 2: Delimited number of species and the posterior probability of the best species delimitation model as estimated by BPP v3.0 on different runs for *Gymnodactylus amarali*, *Micrablepharus atticolus*, and *Tropidurus itambere* including outgroup sequences. No cleandata is when ambiguous and missing sites were included in the likelihood calculations, whereas for cleandata they were removed.

	Number of species and posterior probability of the best model			
Species (number of species in full model)	Algorithm 0 No cleandata	Algorithm 1 No cleandata	Algorithm 0 Cleandata	Algorithm 1 Cleandata
Gymnodactylus amarali (15)	14 – 0.95 - 0.990	14 – 0.82 - 0.989	13 – 0.52-0.70	14 – 0.42 13 – 0.96
Micrablepharus atticolus (10)	10 – 1.00	10 – 1.00	10 – 0.85-0.92	9 – 0.70 10 – 0.99
Tropidurus itambere (12)	7 – 1.00 10 – 0.947	7 – 1.00 10 – 0.909	9 – 0.60 10 – 0.84	9 – 0.62-0.67

Because BPP outputs posterior probabilities for the delimited species, the threshold of whether a species should be considered a different entity or not depends on the empirical system (Satler *et al.*, 2013), and will therefore rely on the authors interpretation of this system. Considering that over-splitting can be worse than not separating true species (Carstens *et al.*, 2013), a feasible conservative approach would be assuming the lineages to be "true species" when a posterior probability of 1 was consistently found across different runs. On the other hand, if being conservative means not splitting species that might be the same, a

similar idea must apply in the opposite case, i.e., separating species that have a very small probability of being the same should also be viewed as a conservative approach.

Not all clades (Figs. 2, 3 and 4) were fully supported by the BPP species delimitation algorithm (Fig. 5). Hence, following the above-mentioned conservative approach, the minimum number of cryptic species (excluding outgroups) that would be recognised are: nine within *G. amarali*, including clades 1 to 7 as separate species, 8–9 and 10–11–12 as the remaining (Fig. 5, Fig. 2); eight within *M. atticolus*, being clades A to E, F–G, H and I (even though F–G had a high probability of being a single species on only one run) (Fig. 5, Fig. 3); and five for *T. itambere* since all clades were consistently recovered as distinct species with a posterior probability of 1 (Fig. 5, Fig. 4).

.

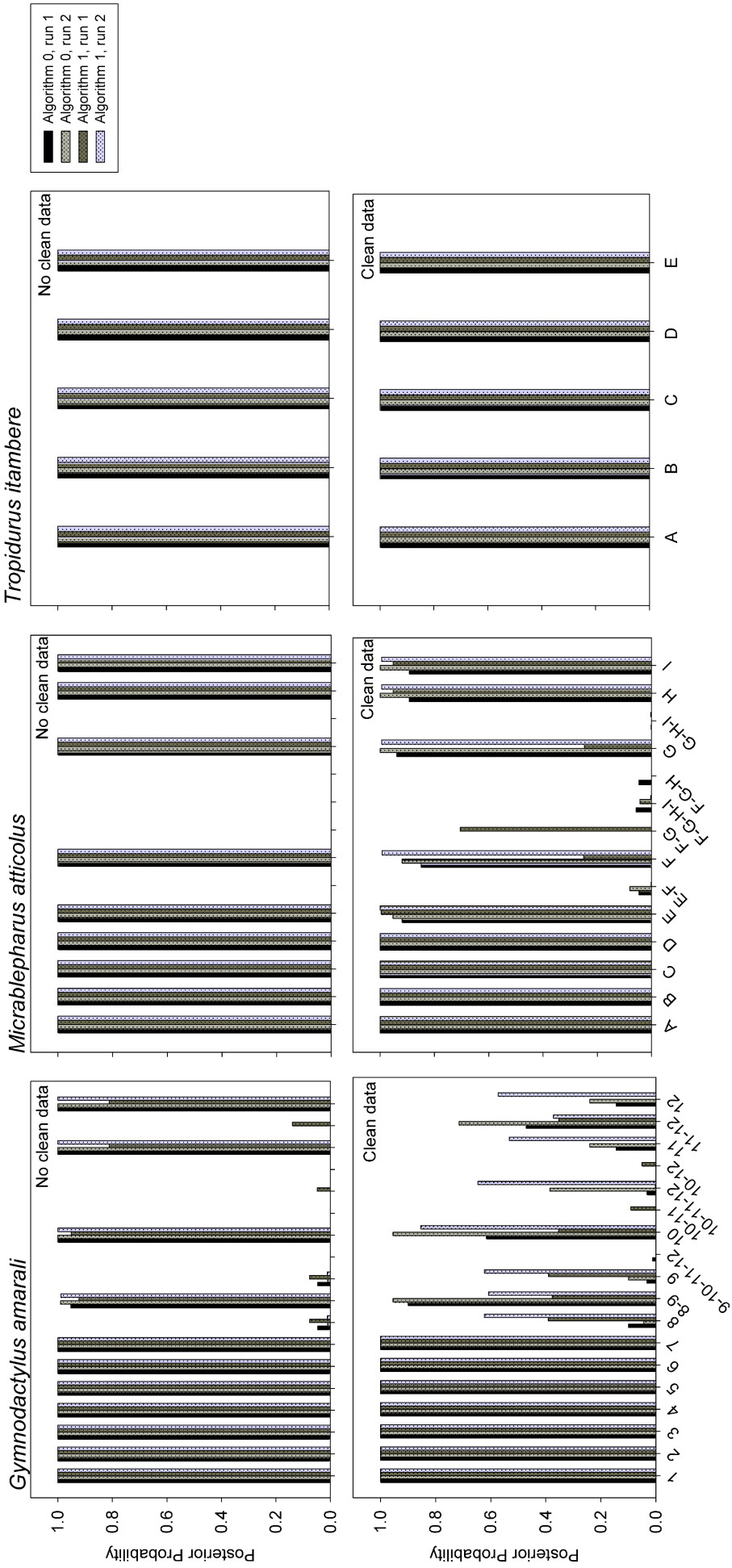
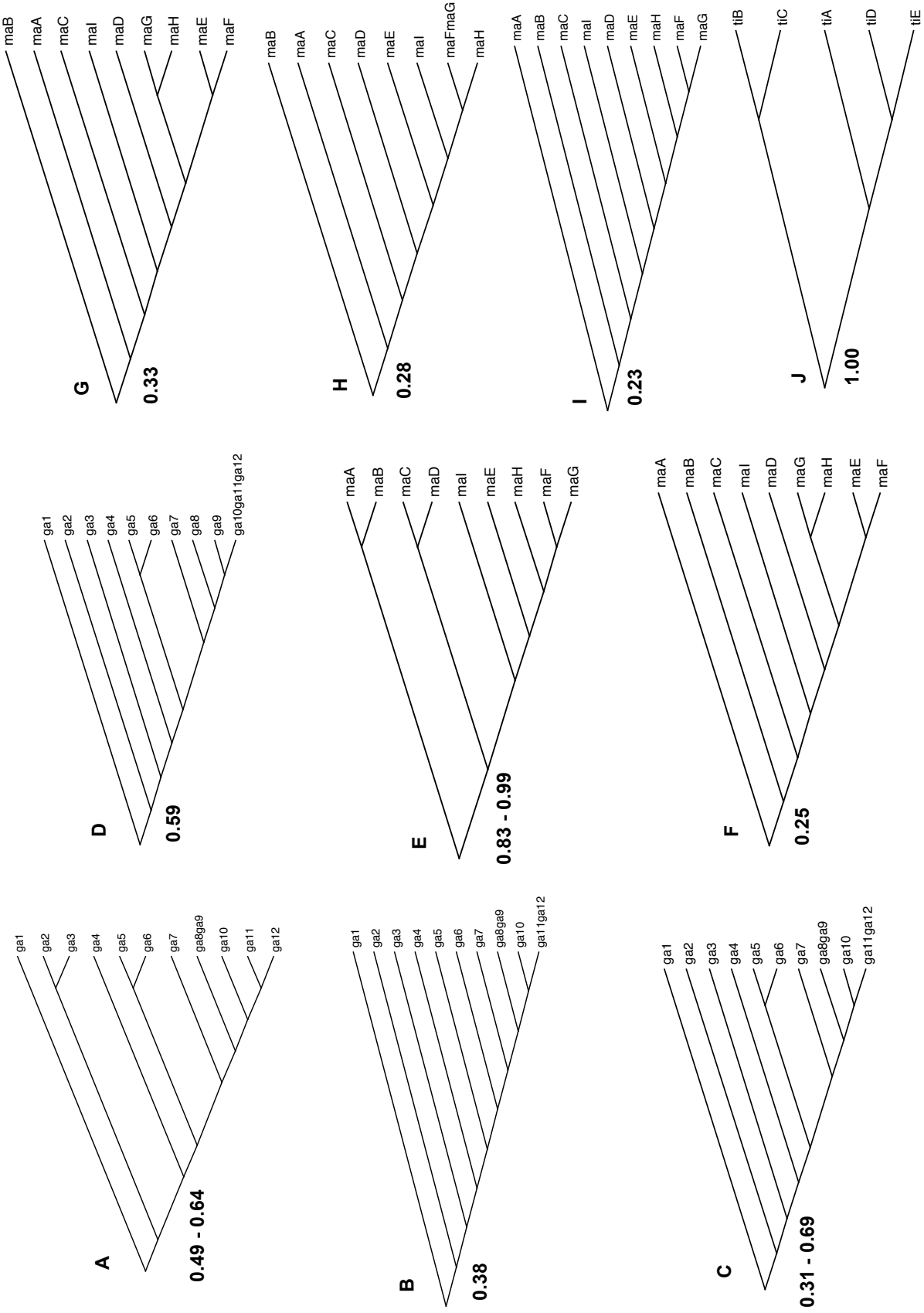


Fig. 5: Delimited species and their posterior probability as estimated by BPP v3 on different runs for *Gymnodactylus amarali*, *Micrablepharus atticolus* and *Tropidurus itambere*. Species with less than 0.01 posterior probabilities in every run were omitted for clarity. No cleandata is when ambiguous and missing sites were included in the likelihood calculations, whereas for cleandata they were removed

The main problem with this approach would be recognising the paraphyletic species 8–9 for *G. amarali*, and F–G for *M. atticolus*. In no phylogeny or species tree were these clades recognised as monophyletic groups (Figs. 2 and 3), except for the BPP species tree estimation itself (Fig. 6). The BPP species tree NNI algorithm is the only fully multi-coalescent species tree estimation we used that is based directly on the genetic data (i.e., STAR and NJst are based on gene trees). Unexpectedly, species trees estimated by BPP for *G. amarali* and *M. atticolus* were different from the topologies estimated by other species tree approaches. BPP ranks the different species trees by their posterior probabilities and, hence, only the results of the best tree per run is presented for each species (Fig. 6). For *G. amarali*, while NJst and STAR retrieved the same topology, with clades 5 and 6 as sister clades (Fig. 2), BPP retrieved four different best models in different runs (Fig. 6A-D), and one of them did not have clades 5 and 6 as sisters (Fig. 6B). For *M. atticolus*, clades B and G were consistently found in different positions in BPP trees (Fig. 6E-I), and never with the same relationships retrieved by NJst and STAR where clade B is sister to C and D, and clade G sister to H and I (Fig. 3). For *T. itambere* all three species tree approaches retrieved exactly the same topology (Fig. 4, Fig. 6J).

Fig. 6: Best estimated species trees by all 8 BPP runs for each taxon. Outgroups were excluded for clarity. Posterior probabilities of model are shown below each tree. ga= *G. amarali*, ma= *M. atticolus*, ti= *T. itambere*. (A) No cleandata - Algorithms 0 and 1; (B) Cleandata - Algorithm 0; (C) Cleandata - Algorithms 0 and 1; (D) Cleandata - Algorithm 1; (E) No cleandata - Algorithms 0 and 1; (F) Cleandata - Algorithm 0; (G) Cleandata - Algorithm 0; (H) Cleandata - Algorithm 1; (I) Cleandata - Algorithm 1; (J) All runs.


maB
maA
maC
maI
maD
maG
maH
maE
maF
G
0.33
maB
maA
maC
maD
maE
maI
maFmaG
maH
H
0.28
maA
maB
maC
maI
maD
maE
maH
maF
maG
I
0.23
tiB
tiC
tiA
tiD
tiE
J
1.00
ga1
ga2
ga3
ga4
ga5
ga6
ga7
ga8
ga9
ga10ga11ga12
D
0.59
maA
maB
maC
maD
maI
maE
maH
maF
maG
E
0.83 - 0.99
maA
maB
maC
maI
maD
maG
maH
maE
maF
F
0.25
ga1
ga2
ga3
ga4
ga5
ga6
ga7
ga8ga9
ga10
ga12
A
0.49 - 0.64
ga1
ga2
ga3
ga4
ga5
ga6
ga7
ga8ga9
ga10
ga11ga12
B
0.38
ga1
ga2
ga3
ga4
ga5
ga6
ga7
ga8ga9
ga10
ga11ga12
C
0.31 - 0.69

3.4 Cytb net between-group distances

Cytb levels of uncorrected sequence divergence among *G. amarali* cryptic species retrieved by BPP ranged from 1% to 15.2% and ML corrected distances from 1.1% to 20% (Table 3). These estimates were smaller for both other taxa: uncorrected divergence among *M. atticolus* cryptic species ranged from 1% to 5% and ML corrected distances from 1% to 5.5% (Table 4); while uncorrected divergence among *T. itambere* cryptic species ranged from 3.1% to 8.3% and ML corrected distances from 3.3% to 9.7% (Table 5)

Table 3: Net among group distances between *Gymnodactylus amarali* cryptic species for cytb data. ML corrected distances using the Tamura-Nei model are above the diagonal, and uncorrected p-distances below. Standard error estimates, calculated using 1000 bootstrap replicates, are shown in parentheses.

	Clade 1	Clade 2	Clade 3	Clade 4	Clade 5	Clade 6	Clade 7	Clade 8-9	Clade 10-11-12
Clade 1	–	0.200 [0.020]	0.195 [0.020]	0.119 [0.013]	0.180 [0.018]	0.181 [0.018]	0.183 [0.019]	0.176 [0.018]	0.180 [0.019]
Clade 2	0.152 [0.012]	–	0.112 [0.014]	0.112 [0.012]	0.122 [0.014]	0.118 [0.014]	0.119 [0.014]	0.131 [0.016]	0.127 [0.015]
Clade 3	0.147 [0.012]	0.094 [0.010]	–	0.129 [0.014]	0.117 [0.013]	0.119 [0.014]	0.118 [0.014]	0.123 [0.014]	0.120 [0.014]
Clade 4	0.095 [0.009]	0.092 [0.008]	0.100 [0.009]	–	0.098 [0.011]	0.095 [0.010]	0.094 [0.010]	0.105 [0.012]	0.105 [0.012]
Clade 5	0.136 [0.011]	0.099 [0.009]	0.094 [0.009]	0.078 [0.007]	–	0.019 [0.003]	0.029 [0.005]	0.028 [0.004]	0.023 [0.003]
Clade 6	0.137 [0.011]	0.097 [0.010]	0.096 [0.009]	0.077 [0.007]	0.017 [0.002]	–	0.019 [0.004]	0.021 [0.004]	0.015 [0.003]
Clade 7	0.141 [0.011]	0.099 [0.010]	0.096 [0.010]	0.078 [0.007]	0.027 [0.004]	0.018 [0.003]	–	0.030 [0.005]	0.025 [0.005]
Clade 8-9	0.137 [0.011]	0.108 [0.011]	0.101 [0.010]	0.085 [0.007]	0.026 [0.004]	0.020 [0.004]	0.028 [0.005]	–	0.011 [0.002]
Clade 10-11-12	0.139 [0.011]	0.104 [0.010]	0.098 [0.010]	0.085 [0.007]	0.022 [0.003]	0.014 [0.003]	0.023 [0.004]	0.010 [0.002]	–

Table 4: Net among group distances between *Micrablepharus atticolus* cryptic species for cytb data. ML corrected distances using the Tamura-Nei model are above the diagonal, and uncorrected p-distances below. Standard error estimates, calculated using 1000 bootstrap replicates, are shown in parentheses.

	Clade A	Clade B	Clade C	Clade D	Clade E	Clade F-G	Clade H	Clade I
Clade A	–	0.024 [0.004]	0.032 [0.006]	0.023 [0.005]	0.020 [0.004]	0.031 [0.006]	0.026 [0.005]	0.033 [0.006]
Clade B	0.022 [0.004]	–	0.042 [0.007]	0.042 [0.007]	0.037 [0.006]	0.050 [0.008]	0.035 [0.006]	0.036 [0.006]
Clade C	0.030 [0.006]	0.038 [0.006]	–	0.049 [0.008]	0.042 [0.008]	0.055 [0.009]	0.041 [0.007]	0.049 [0.009]
Clade D	0.021 [0.004]	0.037 [0.005]	0.046 [0.007]	–	0.030 [0.006]	0.010 [0.003]	0.043 [0.007]	0.042 [0.007]
Clade E	0.019 [0.004]	0.033 [0.005]	0.039 [0.007]	0.028 [0.005]	–	0.038 [0.007]	0.039 [0.007]	0.043 [0.008]
Clade F-G	0.028 [0.005]	0.044 [0.006]	0.050 [0.008]	0.010 [0.003]	0.035 [0.006]	–	0.046 [0.008]	0.047 [0.008]
Clade H	0.025 [0.004]	0.031 [0.005]	0.038 [0.006]	0.039 [0.006]	0.036 [0.006]	0.042 [0.007]	–	0.021 [0.005]
Clade I	0.031 [0.005]	0.032 [0.005]	0.046 [0.007]	0.039 [0.007]	0.040 [0.007]	0.043 [0.007]	0.020 [0.004]	–

Table 5: Net among group distances between *Tropidurus itambere* cryptic species for cytb data. ML corrected distances using the Tamura-Nei model are above the diagonal, and uncorrected p-distances below. Standard error estimates, calculated using 1000 bootstrap replicates, are shown in parentheses.

	Clade A	Clade B	Clade C	Clade D	Clade E
Clade A	–	0.078 [0.009]	0.077 [0.009]	0.092 [0.011]	0.085 [0.011]
Clade B	0.065 [0.007]	–	0.059 [0.008]	0.093 [0.011]	0.078 [0.010]
Clade C	0.065 [0.006]	0.050 [0.006]	–	0.097 [0.012]	0.087 [0.011]
Clade D	0.079 [0.008]	0.080 [0.009]	0.083 [0.009]	–	0.033 [0.006]
Clade E	0.073 [0.008]	0.067 [0.008]	0.074 [0.008]	0.031 [0.006]	–

4. Discussion

The Darwinian shortfall represents a problem not only for conservation biology (Diniz-Filho *et al.*, 2013; Redding *et al.*, 2014; Forest *et al.*, 2015), but it also prevents adequate assessments of evolutionary and biogeographic hypotheses (Monnet *et al.*, 2014; Rangel *et al.*, 2015). Information about geological and ecological history and associated biogeographic patterns is relatively abundant for northern-hemisphere biomes, which provides support for the implementation of detailed phylogeographic analyses, particularly in vertebrates (Bernatchez & Wilson, 1998; Burbrink *et al.*, 2011; Leaché *et al.*, 2013b; Pelletier & Carstens, 2014). On the other hand, deficient geomorphological and ecological information in the Neotropics has hindered the understanding of its biogeographic and evolutionary history (Beheregaray, 2008). Exceptions to this trend are sometimes found in regions within Amazonia and other rainforest habitats, such as Brazil's Atlantic forest (Carnaval *et al.*, 2009; Fernandes *et al.*, 2012; Lougheed *et al.*, 2013; Cooke *et al.*, 2014; Leite *et al.*, 2014; Beheregaray *et al.*, 2015). However, investigations about the Neotropical dry biomes have to rely on much coarser geomorphological information (Prado *et al.*, 2012; Werneck *et al.*, 2012a; Novaes *et al.*, 2013; de Lima *et al.*, 2014b; Santos *et al.*, 2014). A better understanding of the phylogenetic relationships of Neotropical organisms and the disclosure of cryptic species are critically important initial steps towards in-depth investigations of biogeography and evolution in the region.

Here, we used a powerful anchored phylogenomics dataset to investigate phylogenetic relationships and cryptic speciation within three Neotropical lizards endemic to the Brazilian Cerrado. Our main results suggest that the existence of cryptic lineages in the biome is more common than previously thought, highlighting the value of using NGS data and coalescent techniques to investigate patterns of diversity in the understudied Neotropical region. While it is not possible, at this stage, to weight the relative influence of ecology and geography as drivers of speciation in our study system (Losos & Glor, 2003), it seems clear that

interpopulation morphological and genetic (cytb) variation can be an indicator of species-level divergence within our species. However, the degrees of cytb sequence divergence between a few BPP delimited species is not substantially high (*G. amarali* clades 8-9 and 10–11–12 (Table 3), and *M. atticollus* clades D and F–G (Table 4)), and higher mtDNA divergence thresholds have been used for delimiting vertebrate species (e.g., Fujita *et al.*, 2010; Fouquet *et al.*, 2013). For *G. amarali*, not treating these clades as distinct species might be a more conservative approach (Johns & Avise, 1998, but see discussion on G. amarali species below), but for *M. atticollus* it seems like a case of mtDNA and nuclear DNA discordance (Fig. 3, Appendix 15), which will require further attention when dealing with species descriptions in the future (Toews & Brelsford, 2012). The applications of the general species delimitation results for clarifying biogeographic history are explored in Chapter 4, and proposals for future in-depth investigations of the drivers of speciation are discussed in Chapter 5.

4.1 BPP species tree hypotheses

The topologies of the two concatenated phylogenetic approaches and those estimated by NJst and STAR were reasonably similar for all three taxa, but the species trees estimated by BPP were actually very different for *G. amarali* and *M. atticolus* (Fig. 6). There is but a single published study using the BPP v.3 NNI algorithm for species tree reconstruction (Leaché *et al.*, 2015). The authors used a 471 loci (358,363 bp) dataset to estimate species trees using six different phylogenetic or species tree methods, and the same topology was supported by all methods. However, Leaché et al. (2015) investigated relationships at the family level (i.e. among 9 genera), which probably explains the high support for a single topology. While BPP has been demonstrably useful as a powerful species delimitation tool (Leaché & Fujita, 2010; Carstens *et al.*, 2013; Satler *et al.*, 2013; Sistrom *et al.*, 2013), the usefulness of the newly implemented NNI algorithm for species tree estimation among closely related species has yet

to be assessed by empirical tests. During Bayesian phylogenetic estimation, it is well known that over-parameterization (i.e., using a more complex model than necessary) is less problematic than the opposite situation (i.e., using a more simplistic model than necessary) (Huelsenbeck & Rannala, 2004; Lemmon & Moriarty, 2004). BPP only implements the very simple JC69 evolution model that, although being a fair assumption for closely related species, would still not be the best evolution model for all our loci. Therefore, we regard the concatenated trees more accurate; at least until a full multispecies coalescent species tree analysis that can handle such data becomes available.

4.2.1 Cryptic species in the Gymnodactylus amarali complex

Most cryptic species previously reported for *G. amarali* based on cytb, a nuclear gene (KIF24) and morphological data (Chapter 2) correspond to the same species retrieved by BPP using the AP data. The differences are that clades B and E in Chapter 2 are now recovered as a single species (clade 4), and clade H was here divided into three species (clades 6, 8–9 and 10–11–12). These were not completely unexpected outcomes: clades B and E were recovered as sister clades before (Chapter 2, Fig. 4), and were mainly retrieved as different species in Chapter 2 because of differences between cytb and KIF24 trees in the placement of clade E (Appendix 8, Chapter 2). In relation to clade H in Chapter 2, individuals from Peixe and São Salvador (clade 6 here) were in the threshold of being considered a different species by the GMYC analysis (Chapter 2, Fig.3), which would separate them from clades 10–11–12 (as we found here). Taken together, and also considering the separation of clade 8–9 (differently from Chapter 2), these new outcomes are due to the better resolution provided by the AP dataset. Nonetheless, it is interesting to note that these clades were exactly the same where the few morphological misidentifications by the SVM model took place (Chapter 2). Otherwise, clade A in Chapter 2 corresponds to clade 1 here; clade C to clade 3; clade D to clade 2; clade F to clade 5; and clade G to clade 7. In summary, we increased from 8 to 9 the number of

cryptic species in the *G. amarali* complex with an overall high consistency between previous results and those obtained by analyses of AP data.

4.2.2 Cryptic species in the Micrablepharus atticolus complex

The BPP results suggest that the widespread nominal taxon *M. atticolus* actually forms a complex of eight different cryptic species: clades A to E, F–G, H and I (Fig. 5), most of them with cytb genetic distances that are above 2% (Table 4). The first distinguishing pattern within the *M. atticolus* cryptic lineages is the high diversity in eastern Mato Grosso (Fig. 1). Samples from the geographically close municipalities of Barra do Garças, Nova Xavantina, Ribeirão Cascalheira and Cocalinho were retrieved in six different cryptic lineages (A, B, C, E, F–G and H; Fig. 5). The fact that clades from Nova Xavantina (A and E), Barra do Garças (B and C), Cocalinho (F) and Ribeirão Cascalheira (H) are not even retrieved as sister clades in the concatenated phylogenies (Fig. 3) apparently reflects the very intricate geological history of the region (Ab' Sáber, 1954; Ab'Sáber, 1998). The reduced number of specimens in our analysis from a region with high levels of genetic diversity (Table 4) and geological distinctiveness calls for additional work in eastern Mato Grosso state. Fine-scale sampling for genetic analyses aimed at increasing the number of individuals sampled per locality and the number of localities covering the region, along with the inclusion of morphological data, is needed to shed light into the origins of the observed pattern. Elsewhere, there is a distinct cryptic lineage in the more central region of the Cerrado (clade D, Fig. 1), another in a Cerrado peripheral area (clade F–G, Fig. 1), and the last one is found in Cerrado enclaves within the Amazon Forest (clade I, Fig. 1).

4.2.3 Cryptic species in the Tropidurus itambere complex

The results of species delimitation approaches for *T. itambere* returned simpler results compared to the other two species groups. The five inferred cryptic lineages had high support

values in every BPP run. Importantly, the cryptic lineages were also clearly geographically structured, which will facilitate their future taxonomic descriptions. Based on its geographical distribution (Rodrigues, 1987), clade E probably corresponds to the described nominal species, and has the larger distribution across the southeast part of the Cerrado. The topological pattern recovered within clade E (our best sampled lineage) suggests the influence of isolation by distance, with populations structured in a NW-SE direction, similarly to what was observed for the Cerrado frog *Hypsiboas punctatus* (Prado *et al.*, 2012). Clade A is at the northernmost distribution of the species, and the remaining ones are located in the western portion of the Cerrado. In addition, clades B and C were always retrieved as sister species, even though clade D is distributed between them (Fig. 1, Fig. 4).

Individuals from two populations (Moeda and Natividade), morphologically diagnosable as *T. itambere*, actually belong to different species (Fig. 3). This reinforces the limitations of current morphological diagnoses for species of *Tropidurus*, often based on just a few characters (e.g., mite pockets). Because of the highly variable number of species identified by different BPP runs for outgroup species (Appendix 24), and the paraphyly found for *T. torquatus* and the two "itambere-like" populations, we refrain from making additional remarks on the status of the outgroup species until more samples are available. A detailed multi-locus phylogenetic analysis that includes described and cryptic lineages (this Chapter) of the genus *Tropidurus* is highly warranted, since it could potentially clarify such issues.

4.3 Delimitation of paraphyletic species by BPP v.3

The BPP algorithm has assumptions that might not be completely fulfilled by our data, namely no recombination within loci, free recombination between loci, and no gene flow between species (Rannala & Yang, 2003; Yang & Rannala, 2010). The samples from *G. amarali* clade 8 (Natividade) and clade 9 (Almas), recovered as a single species by BPP (Fig. 5), are geographically close (ca. 100 km), and there is some (little) gene flow between them

(Chapter 4). This might be the reason why they were estimated as one species, despite being paraphyletic in regards to the phylogenetic reconstructions (Yang, 2015). The same issue is present between *M. atticolus* clades F and G (Fig. 5), which are also relatively close geographically, and also show signs of gene flow (Chapter 4). There were no similar problems in *T. itambere*, since all species had a posterior probability of 1 for all BPP runs. Although it has been shown, under simulated scenarios, that BPP is sensitive to migration (Zhang *et al.*, 2011), the method also performed well under empirical systems with gene flow (Camargo *et al.*, 2012). Therefore, the main issue here seems to be the ability of BPP to estimate its own species trees, and the differences between this estimation and those based on other phylogenetic hypotheses.

The recognition that speciation under scenarios of gene flow happens more often than previously thought has gained much attention in recent years (Hey, 2010; Sousa & Hey, 2013). With the increase in our ability to sequence more loci, it was just a matter of time until paraphyletic species started appearing in non-model species delimitation studies, i.e. that discordance between methods would become apparent. Unfortunately, our ability to generate huge amounts of sequence data has not been matched by our ability to analyse them (Gronau *et al.*, 2011; Jarvis *et al.*, 2014), which is hindered by the computational resources required to estimate parameters under very complex models (Lemmon & Lemmon, 2013). For example, it was recently recognised that the widespread use of *Beast for the generation of species-trees might not be a feasible method for the genomic era (O'Neill *et al.*, 2013; Pyron *et al.*, 2014). A recent avian phylogeny study that used a very high number of loci had to rely on custom-developed analytical strategies and software to estimate phylogenies and species trees (Jarvis *et al.*, 2014). Until those shortfalls can be surpassed, we predict that the appearance of paraphyletic species will probably not be uncommon in species delimitation studies.

Considering the short period of time since BPP v.3 was released, this is perhaps the first example of species delimitation applying a large number of loci to its NNI algorithm. To date,

there are only two published studies that used BPP v.3 (curiously enough, both on lizards): the first one used four loci to investigate species limits on Mexican geckos of the genus *Phyllodactylus*, and also retrieved paraphyletic species compared to the *Beast tree, but the authors did not discuss this finding (Blair *et al.*, 2015). The second study used the NNI algorithm for species tree estimation only, and not species delimitation (Leaché *et al.*, 2015).

Nevertheless, our paraphyletic species "problem" could be surpassed using the "fixed tree" option of BPP (equivalent to using BPP v.2), which only enables estimation of monophyletic species. To investigate this possibility, we repeated the same analyses using the "fixed-tree" option on Algorithm 0 (results not shown). Indeed we found lower posterior probabilities for the same clades for *G. amarali*, but not for *M. atticolus* and *T. itambere*. Within *G. amarali*, all species had a posterior probability of 1 with "no cleandata", but clades 9 to 12 became collapsed into one species using "cleandata". We attribute this result to the low posterior probability found for clade 9, which cannot be estimated together with clade 8 in this case. Clade 8 had posterior probabilities between 0.6 and 1 in the same analyses. We once again highlight the very low posterior probability of joining clades 9 to 12 estimated by the runs where the species tree was free to vary (Fig. 5). All clades were recovered with a posterior probability of 1 for *M. atticolus* and *T. itambere* using the "fixed-tree" option.

Although we would be reluctant to recognize such paraphyletic clades as one species, this would only be a concern under the phylogenetic species concept, but not under the Generalized Lineage Concept (GLC; de Queiroz, 2007). Under the GLC species are viewed as evolving entities over time. Here, monophyly would be only one of the aspects to be taken in consideration when delimiting species because it could still be gained over the course of time (de Queiroz, 1998). Finally, the cryptic species are paraphyletic when we compare results of the concatenated and coalescent species trees (NJst and STAR) with BPP species delimitation results, but they are certainly monophyletic when the species tree is estimated by the BPP multispecies coalescent model. This apparent dilemma might be solved when BPP

implements additional evolutionary models to its algorithms (Yang, 2015), but at this stage more empirical studies are needed to better evaluate the BPP v.3 NNI algorithm when dealing with closely related species.

4.4 Cryptic speciation in the Neotropics

It would be feasible to consider, given the high number of cryptic lineages revealed by our analyses, that BPP is over-splitting species. However, there is also an increasing number of cases where Neotropical cryptic lineages are recognised using different approaches (and not BPP) (Fouquet *et al.*, 2007; Gamble *et al.*, 2012; Prado *et al.*, 2012; Werneck *et al.*, 2012a; Fouquet *et al.*, 2014; Gehara *et al.*, 2014). The Brazilian lizard fauna is one of the most diverse in the world (Costa & Bérnils, 2014) but, with few exceptions (Giugliano *et al.*, 2013; Recoder *et al.*, 2014), recent species descriptions rely mostly on morphological information (e.g. Teixeira *et al.*, 2013; Arias *et al.*, 2014a; Arias *et al.*, 2014b; Teixeira *et al.*, 2014). In hyperdiverse regions such as the Neotropics, the sole use of morphological data on the recognition of new species can be problematic, since taxonomists might be confounded by morphological stasis when trying to separate those different entities (Bickford *et al.*, 2007).

In other lizard study systems where detailed phylogenies are available, it is not uncommon to recognise an enormous diversity of closely related species with relatively restricted, and even overlapping distributions. Examples exist for Australian groups such as *Gehyra* (Sistrom *et al.*, 2013; Hutchinson *et al.*, 2014), *Heteronotia binoei* (Fujita *et al.*, 2010), *H. spelea* (Pepper *et al.*, 2013), and *Diplodactylus* (Pepper *et al.*, 2006); for Melanesian lizards *Cyrtodactylus* (Oliver *et al.*, 2012); Brazilian *Coleodactylus* (Geurgas *et al.*, 2008); and for west African *Hemidactylus* (Leaché & Fujita, 2010), among others. Expanding on Brazilian studies, it was reported that the lizard *Gymnodactylus darwinii* from the Atlantic Forest shows a strong clinal morphological variation (Freire, 1998), and indeed several cryptic species were later reported to occur within that species group (Pellegrino *et al.*,

2005). The same is true for the *Phyllopezus pollicaris* complex (Gamble *et al.*, 2012; Werneck *et al.*, 2012a), although no morphological study was done for this species so far. In addition, cryptic lineages with strong morphological support were described for *G. amarali* (Chapter 2). The Cerrado comprises a massive area of ~2 million km^2, and evidences suggest that much of its diversity is still to be uncovered (Diniz-Filho *et al.*, 2006). That taxonomists have not recognised different lineages among our study species in the past, strengthens the importance of applying modern analytical tools for the recognition of cryptic biodiversity. Hence, an increasing number of cryptic species in the Neotropics should be revealed through the use of modern NGS data and coalescent species delimitation analyses. Wether all of them should be considered "true" species is likely to generate great debate in the literature in the near future. Likewise, although much more difficult to implement, studies and new techniques that deal with the speciation process itself will probably prove to be a very strong approach to help and clarify those patterns (Andrew *et al.*, 2013; Arnegard *et al.*, 2014; Faria *et al.*, 2014; Smouse *et al.*, 2015).

5. Conclusion

The problem of how many species are there in the world is substantially augmented by the eminent biodiversity loss we currently face (Costello *et al.*, 2013), and by accompanying losses of ecosystem functions (May, 2011). In this study we applied new coalescent methods of species delimitation using a powerful anchored phylogenomics dataset, compared phylogenetic reconstruction methods, and provide indications of the usefulness of such data and analytical approaches for species delimitation hypothesis testing. Whether or not the patterns of morphological and/or karyological variation observed within the three taxa corresponds to the here proposed species boundaries, and what is the relative role of geography and ecology in the generation of diversity in our system is yet to be investigated (see Chapter 5).

Our results indicate that the amount of lizard cryptic species for the Cerrado is unexpectedly high, and the results from other recent studies with amphibians (Funk *et al.*, 2012a; Fouquet *et al.*, 2014) and birds (Smith *et al.*, 2014b) suggest that this elevated hidden vertebrate diversity might turn out to be a pattern for Neotropical biomes. Furthermore, the uncovering of these lizard cryptic species is essential for future species descriptions, which is a source of invaluable information for the concise formulation of conservation strategies in the Cerrado (Silva *et al.*, 2014). At this stage, even though cryptic species are yet to be described, the evolutionary lineages are available for management purposes, and can already be used for direct conservation planning (Niemiller *et al.*, 2013).

Chapter 4

Inner conflict: the roles of ecology and history on the evolution of a Neotropical biodiversity hotspot

1. Introduction

1.1 History and ecology in comparative evolutionary studies

Understanding the relative influence of historical versus ecological processes on the distribution and diversity of organisms has been a central theme of evolutionary biology for centuries (Buffon, 1761; Candolle, 1855; Nelson, 1978; Wiens & Donoghue, 2004). This causal dichotomy was long debated in biogeographic terms, probably dating back to the arguments between dispersalists and extensionists in the nineteenth century (Brown & Lomolino, 1998), which were later replaced by twentieth century debates between dispersalist and vicariant biogeographers (Stace, 1989). Modern discussions on the importance of historical and ecological processes in population divergence and speciation can be viewed as another branch of a similar controversy (Gavrilets, 2003; Schluter, 2009). These are focused on the relative influence of historical (usually geological) processes versus the various biological processes related to the cessation of gene flow (Schluter, 2001; Rundle & Nosil, 2005). Spatial-genetic patterns of population structure in multiple codistributed species provide a way of illuminating this problem. This is because congruent spatial subdivisions may indicate a shared history of population isolation owing to historical contingencies (Bernatchez & Wilson, 1998; Carnaval *et al.*, 2009), or to selective pressures (Cooke *et al.*, 2014; Beheregaray *et al.*, 2015). Another way of looking into this apparent dichotomy is through comparing temporal diversification patterns, where ecology and history are viewed as acting at different time scales ('shallow' versus 'deep') to shape evolutionary processes (Wiens, 2004; Heads, 2015). Comparative phylogeography studies can shed light on our understanding of the relative roles of historical and ecological processes by allowing evolutionary studies to test processes operating at different spatial scales and timeframes, while allowing for comparisons between organisms with contrasting ecologies (Avise, 1998; Bermingham & Moritz, 1998; Lapointe & Rissler, 2005).

A central question in comparative phylogeography is whether codistributed species have experienced congruent spatial and temporal diversification, or not (Arbogast & Kenagy, 2001). Incongruent patterns appear to be about as common as tightly coupled diversification histories across species (Soltis *et al.*, 2006; Leache *et al.*, 2007; Moritz *et al.*, 2009; Fouquet *et al.*, 2012; Bagley & Johnson, 2014a), although an updated literature review compiling information from a large number of comparative phylogeographies is lacking (Beheregaray, 2008). In comparative phylogeography, ecological differences among species are usually invoked as *post-hoc* explanations for incongruent patterns (Taberlet *et al.*, 1998; Feldman & Spicer, 2006). In fact, comparative phylogeographic studies rarely test hypotheses that incorporate temporal assessments of the relative influence of ecological and historical factors. There are, however, few studies where explicit *a priori* hypotheses have been used to test for congruent evolutionary responses of ecologically similar taxa (e.g., Jezkova *et al.*, 2009; Morgan *et al.*, 2011; Topp *et al.*, 2013) or incongruent diversification patterns among ecologically divergent taxa (e.g., Carstens & Richards, 2007; Papadopoulou *et al.*, 2009; Smith *et al.*, 2014a).

1.2 Evolution and diversification of the Cerrado biota

The Brazilian Cerrado biome has only recently become the focus of conservation efforts (Myers *et al.*, 2000). Biodiversity loss in the Cerrado region is mainly attributed to habitat destruction due to rapid and uncontrolled agricultural development (Oliveira & Marquis, 2002). These characteristics, combined with the current high rates of deforestation, as well as with the small representation of the biome in protected areas (2.2%) (Klink & Machado, 2005), makes the Cerrado one of the 34 global biodiversity hotspots for conservation (Myers *et al.*, 2000; Mittermeier *et al.*, 2005).

The Cerrado is the biologically richest savannah in the world (Castro *et al.*, 1999; Oliveira & Marquis, 2002). The high biodiversity in the region may have been promoted by its complex and diverse vegetation structure (Furley, 1999; Oliveira Filho & Ratter, 2000), with plant physiognomies differing in the predominance and size of woody elements (Goodland, 1971). The biome contains a complex mosaic of landscapes, including high plateaus dating back to the epeirogenic uplift of the Central Brazilian Plateau during the Miocene–Pliocene transition (King, 1956), and low valleys excavated by major river drainages (Ab'Sáber, 1998). This geomorphological compartmentalisation, combined with the high geographical variation of regional soil types (Gomes *et al.*, 2004), yields a heterogeneous vegetation landscape (Oliveira Filho & Ratter, 2000) with many isolated habitats and small-scale ecotones that could potentially promote ecological isolation or speciation (reviewed in Beheregaray *et al.*, 2015). Indeed, community data show that habitat diversity and landscape compartmentalisation have probably played an important role in influencing local diversity of Cerrado amphibians and reptiles (Colli *et al.*, 2002; Nogueira *et al.*, 2005; Nogueira *et al.*, 2011), birds (Silva, 1997; Silva & Bates, 2002) and mammals (Redford & da Fonseca, 1986; Mares & Ernest, 1995; Johnson *et al.*, 1999). Nonetheless, specific tests of how these characteristics may have influenced biological diversification within the biome are still scarce (Werneck *et al.*, 2012a; Santos *et al.*, 2014), and no studies have investigated this topic in the Cerrado using comparative phylogeography (Werneck, 2011; Collevatti *et al.*, 2015).

At a deeper temporal scale, the diversification of Cerrado biotas might have been influenced by population processes associated with climatic fluctuations during the Quaternary glacial cycles (Haffer, 1969; Vanzolini & Williams, 1981). Palaeoclimatic vegetation distribution models supported by palaeopalynological evidence indicate that Cerrado valleys were climatically unstable compared to plateaus during Quaternary climatic fluctuations (Werneck *et al.*, 2012b). Hence, populations inhabiting valleys might have gone

extinct during periods of low habitat suitability during glaciations. In this context, species distribution modelling (SDM) has been used as a surrogate to investigate ecological divergence between species (Rissler & Apodaca, 2007). By adding information about past climate into SDMs, one can also propose and test spatially explicit biogeographic hypotheses about species responses to Quaternary climatic fluctuations (Richards *et al.*, 2007), e.g. concerning Pleistocene refugia (Knowles & Carstens, 2007), range contractions during the last glacial maximum (Bagley *et al.*, 2013), and even assess patterns of common vicariant history among different species (Hugall *et al.*, 2002).

Although Neotropical diversification patterns cannot be simply attributed to a few mechanisms acting over particular time intervals (Rull, 2011; Beheregaray *et al.*, 2015), the general hypotheses potentially accounting for the evolution of extant Neotropical biotas have been divided in two broad geological timeframes: the climatic fluctuations during Quaternary glaciation cycles, and periods of major tectonic events of the Neogene (Rull, 2008, 2011). The sets of events that probably influenced diversification of endemic Cerrado species and coincided with the above timeframes are: (1) the Quaternary climatic fluctuations and their impacts on the distribution of Cerrado vegetation (Werneck *et al.*, 2012b), and (2) the uplift of the Central Brazilian Plateau in the Neogene (Colli, 2005; Werneck, 2011 and references therein). Testing specific hypotheses related to these sets of events using comparative phylogeographic analyses and SDMs would provide a powerful framework to investigate diversification of the Cerrado endemic biota (Collevatti *et al.*, 2015) while accounting for ecological dissimilarities among species that might have influenced their spatial patterns of genetic diversity (Carstens & Richards, 2007).

1.3 Lizards comparative evolutionary studies

Lizards serve as key model organisms for addressing questions in evolutionary investigations. They display a range of evolutionary and ecological patterns, making them ideal candidates for comparative studies (Vitt & Pianka, 2005; Camargo *et al.*, 2010). For instance, comparative studies using lizard species have been used to test the theory of density dependent natural selection based on their different reproductive strategies (Tinkle *et al.*, 1970), and to investigate how physiological and morphological parameters are linked to differences in ecology and habitat use (Kohlsdorf *et al.*, 2001; Kohlsdorf *et al.*, 2004; Kohlsdorf & Navas, 2007). Lizards have also been study subjects to comparatively test models of morphological evolution (Skinner & Lee, 2010), and investigate hypotheses on the evolutionary reversibility of morphological characters (Kohlsdorf & Wagner, 2006). The availability of detailed physiological data has also allowed for the development of thermal requirement models and their influence on extinction risks (Sinervo *et al.*, 2010), and also to understand how ecophysiological parameters can be used to predict general patterns of geographic distribution (Navas, 2002).

Ecological evidence based on SDM coupled with phylogeographic analyses (Moussalli *et al.*, 2009), and niche similarity coupled with phylogenetic information (Schulte *et al.*, 2012) suggests that lizards maintain their ecological and climatic preferences through evolutionary time. The latter is an important attribute when testing diversification hypotheses based on possible range expansions and contractions through time (Wiens & Graham, 2005). The three species complexes used as model organisms in this study (*Gymnodactylus amarali*, *Micrablepharus atticolus* and *Tropidurus itambere*) are endemic to the Cerrado, and hence ideal candidates to assess diversification hypotheses in this biome. The sister species of both the *Gymnodactylus amarali* complex (Phyllodactylidae) and the *Tropidurus itambere* complex (Tropiduridae) are distributed in the adjacent Caatinga biome (Rodrigues *et al.*,

1988; Frost *et al.*, 2001, Chapter 2), while the sister species of the *Micrablepharus atticolus* species complex (Gymnophtalmidae) is widely distributed in the open South American biomes (Rodrigues, 1996; Santos *et al.*, 2014). If niche conservatism holds as the 'true' pattern for our focal taxa (Losos, 2008), then the restricted distributions of our study species within the Cerrado, compared to their sister species distribution, suggest that their ecology is intrinsically related to unique features of this biome. Indeed, the three species complexes reproductive cycles are intimately associated with the seasonality of the biome: *G. amarali* reproduces exclusively during the dry season (Colli *et al.*, 2003a), as does *M. atticollus* (Vieira *et al.*, 2000), whereas *T. itambere* reproduces only during the wet season (Van-Sluys, 1993; Ferreira *et al.*, 2009). As a result, our theoretical expectation is that evolutionary patterns inferred among the three focal species complexes are associated with regional landscape evolution and climatic fluctuations that occurred in the Cerrado. In particular, we predict that patterns of genetic divergence within the three species complexes will track past vegetation shifts inferred from SDMs, as described for endemic vertebrates in other species-rich biomes (Carnaval & Moritz, 2008; Carnaval *et al.*, 2009; Moritz *et al.*, 2009).

1.4 Comparative phylogeography and diversification hypotheses in the Cerrado

The field of phylogeography was developed based on using mtDNA to investigate intraspecific diversification patterns (Avise *et al.*, 1987). Reliance on the single mtDNA locus as a genetic marker has remained widespread for a long time, particularly for animals (Soltis *et al.*, 2006). This scenario has recently started to change with the development of cheaper and more efficient sequencing technologies, notably the advent of next-generation sequencing (NGS) technologies and genomic datasets (Garrick *et al.*, 2015). Sampling a larger number of nuclear loci may improve the accuracy of coalescent estimates of phylogeographic divergence (Huang *et al.*, 2011; Robinson *et al.*, 2014), alleviate gene tree discordances (Leache, 2009;

Sistrom *et al.*, 2014), and avoid evolutionary misinterpretations due to incomplete lineage sorting (Maddison & Knowles, 2006; Heled & Drummond, 2010). Statistical phylogeographic methods can also be used to disentangle complex biogeographic and temporal scenarios (Knowles & Maddison, 2002; Hickerson *et al.*, 2006) which, in conjunction with the use of many genetic markers (Rannala & Yang, 2003), provides an ideal framework to investigate evolutionary patterns in the Cerrado biome.

The development of novel NGS methods is promoting new opportunities for phylogeographic studies (Carstens *et al.*, 2012; McCormack *et al.*, 2013). Although coalescent models have been applied to NGS data in recent comparative studies (e.g., Leaché *et al.*, 2013a; Brandley *et al.*, 2015), to the best of my knowledge there is only one published comparative phylogeography of non-model organisms using NGS (Smith *et al.*, 2014a). In that study, the authors classified clades of five widespread Neotropical bird species *a priori* into distinct geographic units used to group genetic samples when estimating coalescent population parameters, species trees, and hypotheses tests for cryptic species using BPP (Yang & Rannala, 2010). Their results indicate that the bird population-lineages are monophyletic and geographically isolated in proposed geographic units based on geology and other factors (Smith *et al.*, 2014a, Fig. 3). For this thesis, grouping lineages of the three lizard species complexes into distinct *a priori* geographic units within the Cerrado is not appropriate because we would be creating paraphyletic groups for analysis (Chapter 3). On the other hand, the cryptic species previously identified in this thesis could be used as a proxy for delineating clades from which we could draw hypotheses and estimate population parameters based on the coalescent.

Lizards from open areas of the Cerrado are predominantly found in two structurally and functionally different habitats: those associated with rocky outcrops (Cerrado 'rupestre'), and those associated with the typical savannah vegetation (Cerrado '*sensu stricto*'). Both *G.*

amarali and *T. itambere* species complexes are strongly associated with rocky outcrops (Van-Sluys, 1997; Colli *et al.*, 2003a), whereas the *M. atticollus* complex is found in the leaf litter (Vieira *et al.*, 2000) and occasionally within ant nests (Rodrigues, 1996). Thus, the differing ecologies of our focal species complexes allow for identifying ecological contrasts to be incorporated into explicit tests of historical hypotheses. This framework can help elucidating the influence of ecological attributes on diversification of Cerrado lizards (Marske *et al.*, 2013).

To understand mechanisms that influenced diversification in species-rich biomes, diversification hypotheses should be tested within and among different codistributed species (Lexer *et al.*, 2013). For this endeavor, I used two datasets (AP and cytb) and SDM methods to test the hypothesis that species complexes living on rocky outcrops (*G. amarali* and *T. itambere*) have experienced similar spatial and temporal histories of divergence. This framework was also used to formulate a suite of specific hypotheses to investigate how Neogene geological events and Quaternary climatic fluctuations influenced the evolution of Cerrado lizards (summarised in Table 1).

Given the lack of detailed geological information for the Cerrado, and considering that historical evolutionary patterns of its endemic biota have not been investigated in a comparative fashion, it was decided that a series of specific hypotheses of diversification should be tested. This strategy was adopted to understand the responses of species to historical events, and to elucidate the relative influence of demographic processes in shaping current biogeographic patterns. Below, I introduce these hypotheses and link them to background information about Cerrado biogeography.

Assuming that endemic lizards were distributed in the Cerrado before the uplift of the Central Brazilian Plateau, the landscape compartmentalisation that followed its origin created a scenario for vicariant events in which gene flow is restricted among populations inhabiting

different plateaus (hypothesis 1), but gene flow should be higher among valley populations (2). Considering that rivers excavated the landscape during and especially after the uplift process, I expect that older (basal) lineages should be located in plateaus (3), and that the ancestral distribution of each species was located in a plateau area (4). In addition, reciprocally monophyletic clades should be found between plateaus and adjacent valleys (5) (Werneck, 2011). Likewise, if these Neogene tectonic events influenced the evolution of endemic groups in a similar way, it is also expected that the deepest population divergence within each species complexes occurred synchronically among the three taxa (6), and that divergence in clades that share similar plateau-valley distribution occurred as clusters of divergences around the same time (7).

Given that stable areas of the Cerrado were mainly located in plateaus during the Quaternary climatic fluctuations (Werneck *et al.*, 2012b), I expect that endemic lizards have tracked the biome distribution and that their stable areas (i.e. refugia) have also been located in plateaus throughout glaciation-interglaciation cycles (8). Besides, if populations in valleys have more recent colonisation histories they should show signs of population expansion (9), have lower genetic diversity (10) and smaller effective population size compared to plateau populations (11). Furthermore, if there are coincident Quaternary refugia among species, I expect that divergence between separated refugia occurred synchronically among different species (12) (Carnaval *et al.*, 2009).

Finally, I also expect that species with similar habitat ecologies (*G. amarali* and *T. itambere*) experienced similar palaeodistributional shifts in response to climatic changes (13), and that clades within these two species complexes will present similar divergence times and effective population sizes (14).

To address the above suite of *a priori* hypotheses, I used the AP data to infer evolutionary relationships among clades/populations using phylogenetic analyses, and used

coalescent approaches to estimate demographic population parameters within each species complex. Because of the limited number of individuals and sites sequenced for AP data (n = 72 and 53 sites across the three groups), I took advantage of the better sampling available for the mtDNA cytb data (n = 356 and 67 sites across the three groups). The latter enabled more robust inferences of ancestral geographic locations and calculation of several population genetic statistics for the three study taxa. Finally, I also used mtDNA to investigate whether the three species complexes showed patterns of simultaneous diversification (vicariance) consistent with those predicted by two hypotheses (6 and 12). The latter was achieved with a hierarchical approximate Bayesian computation (ABC) model recently developed for comparative multi-locus phylogeography (Huang *et al.*, 2011).

In general, the results indicate congruent phylogeographic patterns related to Neogene diversification hypotheses among the ecologically similar *G. amarali* and *T. itambere* complexes. However, it appears that landscape compartmentalisation played different roles in the evolution of each taxon. Population genetic estimates suggest no genetic diversity or effective population size differences between populations in valleys and plateaus for any of the three taxa. Unexpectedly, the *M. atticollus* and *T. itambere* species complexes had very similar palaeodistributional shifts throughout the Quaternary, while *G. amarali* presented a different pattern. Accordingly, species complexes with more similar geographic ranges (*M. atticolus* and *T. itambere*) had concordant demographic responses to Quaternary climatic fluctuations. Overall, our results suggest that unveiling biogeographic patterns of endemic lizards in the Cerrado requires accounting for historical and climatic events of the biome in light of the ecological characteristics of each particular taxon.

Table 1: Diversification hypotheses for endemic lizards in the Cerrado biome. Expectations were draw based on Neogene and Quaternary events and differences in ecology among species. The metrics used for hypothesis testing and respective software used for estimations are listed (see Material and Methods for details).

	Hypothesis	Metric	Software
1	Restricted gene flow among populations found in different plateaus	Migration estimates	G-PhoCS
2	Higher gene flow among populations in valleys	Migration estimates	G-PhoCS
3	Older (basal) lineages are located in plateaus	Phylogenetic relationships	RAxML
4	Ancestral distribution of each species located in a plateau	Ancestor geographic location estimates	Phyllomaper
5	Reciprocally monophyletic clades between plateaus and adjacent valleys	Phylogenetic relationships	RAxML
6	Deepest population divergence within each species complex (i.e. population structure) formed synchronously among the three taxa	Test for simultaneous diversification	MTML-msBayes
7	Divergence in clades that share similar plateau-valley distribution occurred as clusters of divergences around the same time	Divergence time estimates	G-PhoCS
8	Stable areas (refugia) located in plateaus throughout Quaternary	Palaeoclimatic SDM-inferred refugia	Maxent
9	Populations in valleys show signs of population expansion	F_s and R_2 statistics	DNASP
10	Populations in valleys have lower genetic diversity	Number of haplotypes, haplotype and nucleotide diversity	DNASP
11	Populations in valleys have smaller effective population size compared to plateau populations	Effective population size (N_e) estimates	G-PhoCS
12	Divergence between refugia (identified using SDMs) occurred synchronously among different species complexes	Test for simultaneous diversification	MTML-msBayes
13	Ecologically similar species experienced similar palaeodistributional shifts in response to Quaternary climatic changes	Palaeoclimatic SDM	Maxent
14	Ecologically similar species present similar divergence times and effective population sizes	Divergence times and effective population size (N_e) estimates	G-PhoCS

2. Material and Methods

2.1 Sampling and genetic protocols

Sampling of species and individuals, as well as genetic data collection, followed the methods described in Chapter 3 for cytb and AP. The final ingroup dataset for the *G. amarali* complex consisted of 155 individuals from 27 localities sequenced for cytb and 23 individuals from 22 localities sequenced using AP (Appendix 17). The dataset for the *M. atticolus* complex consisted of 126 individuals from 28 localities sequenced for cytb and 26 individuals from 20 localities assessed using AP (Appendix 18). Finally, the dataset for the *T. itambere* complex dataset consisted of 75 individuals from 18 localities sequenced for cytb and 23 individuals from 17 localities sequenced using AP (Appendix 19).

2.2 Phylogenetic relationships (Hypotheses 3 and 5)

One of the difficulties when using the AP approach for diploid organisms is that the complementary versions of the nuclear genome are sequenced at the same time. Heterozygous positions can easily be identified during bioinformatics workflows used to assess AP data (Lemmon *et al.*, 2012), but assembling two haplotypes from one diploid sample of sequences (i.e. phasing heterozygous sites) can be challenging (Sousa & Hey, 2013). Ambiguous heterozygous sites are not expected to have a strong impact on tree topology (Lemmon *et al.*, 2009; Wiens & Morrill, 2011), but absolute and relative branch length estimations can be substantially influenced if heterozygous sites are ignored during phylogenetic analyses (Sota & Vogler, 2003; Lischer *et al.*, 2014). To deal with phase uncertainty and obtain more realistic branch length estimates when reconstructing phylogenetic relationships within *G. amarali*, *M. atticolus*, and *T. itambere* complexes (Chapter 3), we used a Repeated Random Haplotype Sampling (RRHS) approach (Lischer *et al.*, 2014) that integrates information from all alleles into tree reconstructions. In short, the method involves generating thousands of

alignments in which the phases of heterozygous sites have been randomly resolved, running phylogenetic analyses on each one of those alignments, and then combining results into a single phylogenetic tree by calculating a majority rule consensus (MRC) tree with mean branch lengths (Lischer *et al.*, 2014).

For each species complex, 3000 simulated phase alignments of the AP dataset were generated using the software RRHS v1.0.0.2 (Lischer *et al.*, 2014). Maximum Likelihood (ML) phylogenetic analyses were run on each alignment using RAxML v8.1.1 (Stamatakis, 2014). All runs specified a GTRGAMMAI model using rapid hill-climbing searches and estimated bootstrap support values using 1000 pseudoreplicates of the data with the RELL bootstrap option (Minh *et al.*, 2013). We applied the same data partition strategy used in Chapter 3, which involved selecting independently evolving partitions using PartitionFinder v1.1.1 (Lanfear *et al.*, 2012), with the RAxML relaxed clustering algorithm under the 10% search default condition (Lanfear *et al.*, 2014). The 3000 trees resulting from the runs were summarised by calculating a MRC tree using the 'consense' tool of Exabayes v1.2.1 (Aberer *et al.*, 2014). Bootstrap values for all 3000 runs were combined into a single file, and visualized over the MRC tree using RAxML v8.1.1. We used *G. darwinii* 'Matias Cardoso' as the outgroup for the *G. amarali* complex analyses; *M. maximiliani* was the outgroup for the *M. atticolus* complex; and *Uranoscodon superciliosus* was used as the outgroup for the analyses of the *T. itambere* complex (Chapter 3).

We opted for an ML approach because of computational time constraints, which would be much higher if we had used Bayesian phylogenetic analyses: it took ~12 minutes for each ML phylogenetic analysis to finish using 8 cores (Intel Xeon CPU E5-2680) with the RAxML PTHREADS-AVX flag on our HPCF *Phoenix*. Under similar conditions (8 cores using the AVX capable Exabayes version on Phoenix), it took over 24 hours to compute a Bayesian analysis and achieve acceptable convergence (Chapter 3). Nonetheless, it is very unlikely that

the final topologies change when using ML versus Bayesian approaches (Chapter 3), and branch length estimates would probably not be significantly different (Lischer *et al.*, 2014). Related to that, ML branch length estimates have been shown to be more accurate than Bayesian estimates in empirical datasets (Schwartz & Mueller, 2010). All analyses were run using the HPCF *Colossus*, a centralised supercomputer at Flinders University, or using *Phoenix*, a HPCF based at the Molecular Ecology Lab at Flinders University.

Given the discordance observed between results of STAR and NJst species trees compared to the BPP species tree (Chapter 3), and the lack of a method that can accurately infer a species tree directly from AP data (Leache & Rannala, 2011; O'Neill *et al.*, 2013; Pyron *et al.*, 2014), we used the RRHS ML result for each species as our 'preferred' topology for guiding subsequent analyses. Despite possible errors associated with incomplete lineage sorting (ILS) when estimating trees from concatenated datasets (Heled & Drummond, 2010; Mirarab *et al.*, 2014), a recent simulation study suggested that concatenation has similar accuracy to that of species tree methods even when loci differ in coalescence rates (Tonini *et al.*, 2015).

Finally, branch lengths of final trees estimated by the RRHS ML approach were compared with branch lengths of the ML trees estimated with ambiguous IUPAC notation in Chapter 3, using the R package *ape* (Paradis *et al.*, 2004). Outgroups were removed from rooted trees and relative branch length differences calculated between the two trees. Additionally, to allow for a proportional comparison between the trees, branch lengths were scaled to 1 before calculating overall absolute branch lengths differences.

2.3 Demographic history inference (Hypotheses 1, 2, 7, 11, 14)

We used the software G-PhoCS (Gronau *et al.*, 2011) to estimate population demographic parameters for our study taxa: divergence times (τ , tau; coalescent units in generations),

population size (θ, theta) and migration. The G-PhoCS computational code was developed by upgrading the code of MCMCcoal (Rannala & Yang, 2003) to calculate the parameters tau and theta, while also introducing a migration parameter into the model allowing gene flow between population lineages. The improvements in relation to MCMCcoal (apart from the obvious improvement wrought by addition of migration estimation) are that G-PhoCS was designed to work on genomic datasets in a way that is computationally reasonably fast and efficient, and it also implements an algorithm to account for phase uncertainty when calculating each locus likelihood (Gronau *et al.*, 2011). Estimating the phases of unknown haplotypes is hindered by relatively large error rates (Scheet & Stephens, 2006), and by the fact that those errors scale up when large genomic datasets are being analysed (Browning & Browning, 2011). The G-PhoCS phasing algorithm integrates all possible phases during the run. Previous analyses of simulated datasets suggest that its estimations are as accurate as using the true haplotype phases (Gronau *et al.*, 2011). G-PhoCS also has a clear advantage over other available software that can estimate similar parameters, such as the programs 3s (Zhu & Yang, 2012) and *dadi* (Gutenkunst *et al.*, 2009), because it does not suffer from the limitation of using a population triplet. Although still not a widely used software by phylogeographers, G-PhoCS has been used to successfully model phylogeographies of non-model organisms (Leaché *et al.*, 2013b; Smith *et al.*, 2014a), and it has produced realistic estimates of population demographic parameters. While the raw estimates will probably vary because of code implementation, G-PhoCS thus represents a novel permutation of the MCMCcoal model aimed at facilitating applications of the model to scaled up genomic data.

To estimate the demographic parameters, we separated *G. amarali*, *T. itambere* and *M. atticolus* clades based on strongly supported monophyletic groups (bootstrap values > 75) inferred in the RRHS ML analyses, which were the same ones used for species delimitation tests in Chapter 3. Similar to MCMCcoal (and, for that matter, also BPP (Yang & Rannala,

2010)), G-PhoCS uses a gamma (α, β) distribution prior for the population standardized mutation rate parameter ($\theta = 4N_e\mu$ for a diploid locus, where μ is the per nucleotide site per generation mutation rate) and for the divergence time parameter ($\tau = T\mu$; T is absolute divergence time in millions of years), as well as for migration bands ($m_{sx} \times \theta_x/4 = M_{sx}$, or migration rate per generation). Migration is, thus, the proportion of individuals in population x that arrived by migration from population s per generation. Using the same approach of Chapter 2 (Appendix 5), we performed several preliminary analyses using different priors. Because the results were very similar, we used a gamma prior of ~G (2,1000) for the population size and divergence time parameters, and ~G (1,10) for the migration bands. Due to code limitations, G-PhoCS migration estimates may be affected if one tries to estimate several migration bands at the same time (see Gronau *et al.*, 2011 supplementary material). Thus, bidirectional migration was inferred separately for each population pair in different runs. Because the migration estimates take much longer to converge than the other two parameters, we did not estimate migration between every population pair, but only between the ones that were important for our hypotheses testing framework (see Results 3.2).

G-PhoCS analyses were set to automatically find the best fine tuning parameters for MCMC updates, and to run for 1 x 10^6 iterations taking samples every 100 iterations. Tracer v1.5 (Rambaut & Drummond, 2009) was used to check for minimum adequate ESS, and to extract raw demographic estimates using a 10% burn-in. At least one run without migration bands was performed to check for possible differences in parameter estimation. As suggested by the G-PhoCS manual, we used our initial run trials to get an idea of MCMC convergence and only start sampling migration after convergence. Migration estimates were then set to start sampling after 1 x 10^4 MCMC iterations. Similar to BPP, G-PhoCS can handle any number of missing individuals per locus; thus, we used exactly the same datasets used for BPP analyses in Chapter 3, where individuals with more than 30% missing data for a given

locus were excluded from the locus alignment. Because G-PhoCS cannot handle indels (-), all indels were coded as missing data (N) prior to analyses.

Unfortunately, no fossils are available for our focal species or outgroups that would allow calibrating a molecular clock and calculating divergence times among clades using the RRHS ML tree. Also, there are no clear geologic events in the Cerrado that could be associated to split events and then used as calibration points to estimate rates of molecular evolution (Werneck, 2011). As previously discussed (Chapter 3), using a species tree allied to a relaxed molecular clock (e.g., with *BEAST) is also not feasible. To overcome the above problems, we estimated rates of DNA substitution using the strategy of Smith and collaborators (2014a) whereby G-PhoCS raw divergence time estimates (τ) are converted using relative substitution rates (μ, Appendix 25) to obtain absolute divergence times in millions of years (T) (see MCMCcoal manual). Within each species, we calculated the average pairwise genetic distance (π) for each AP locus and for cytb (using sequences from the same individuals for which we had AP sequences). We then averaged π across AP loci and calculated a relative substitution ratio (AP π / cytb π) that was scaled to a cytb rate of 0.0065 substitutions/site/million years (Macey *et al.*, 1998). The latter is a per-lineage mutation rate that is widely used for dating squamate phylogenies (Hugall & Lee, 2004; Torres-Carvajal & de Queiroz, 2009; Werneck *et al.*, 2012a; Morando *et al.*, 2014).

We used the relative substitution rates calculated above to convert G-PhoCS raw theta estimates ($\theta = 4N_e\mu$) to effective population sizes (N_e) (Yang, 2002), assuming a generation time of one year, which is realistic for all of our focal species (Van-Sluys, 1993; Vieira *et al.*, 2000; Colli *et al.*, 2003a).

While the method to obtain our AP dataset has only recently been develop (Lemmon *et al.*, 2012), and thus few studies have explored its performance to estimate population-level parameters (Lemmon & Lemmon, 2013; Brandley *et al.*, 2015), there is an immense literature

on mtDNA inter- and intra-specific divergence estimates (Johns & Avise, 1998; Weir & Schluter, 2008; Freeland *et al.*, 2011). To allow for comparisons between AP divergence times and N_e to those estimated from mtDNA, coalescent demographic parameters were also estimated using cytb. We attempted running G-PhoCS for this purpose, but runs would not converge (or even start) with only one locus, perhaps because G-PhoCS was designed to draw estimates from genomic datasets. For this reason, both parameters (τ and θ) were estimated using MCMCcoal (Rannala & Yang, 2003), a program whose algorithm is very similar to that of G-PhoCS, as discussed above. We ran MCMCcoal for 1×10^6 MCMC generations, sampling every 5 generations, with a burn-in of 1×10^4. To ensure that the runs converged, each run was repeated three times for all species complexes.

2.4 Species distribution modelling (Hypotheses 8 and 13)

We compiled geographic distribution records of *G. amarali*, *M. atticollus* and *T. itambere* from Nogueira (2006), a review of the published literature on each species, and from data from major Brazilian collections containing extensive representations of Cerrado species (Coleção Herpetológica da Universidade de Brasília – CHUNB, Museu de Zoologia da Universidade de São Paulo – MZUSP, and Universidade Federal do Mato Grosso – UFMT, Appendices 26, 27 and 28). We built SDMs with Maxent v3.3.3 (Phillips *et al.*, 2006; Phillips & Dudík, 2008) using the same environmental variables used by Werneck and collaborators (2012b) to allow for comparison with the modelled distribution of the Cerrado biome vegetation. These variables include elevation (Alt) and nine bioclimatic variables: precipitation of wettest quarter (BIO16), temperature seasonality (BIO4), mean temperature of coldest quarter (BIO11), precipitation seasonality (BIO15), temperature annual range (BIO7), isothermality (BIO3), mean temperature of warmest quarter (BIO10), precipitation of driest quarter (BIO17), and precipitation of driest month (BIO14). We used four time

projections to build SDMs: present (obtained from WorldClim (Hijmans *et al.*, 2005)), mid-Holocene (6 thousand years ago (ka), ECHAM3 atmospheric General Circulation Model (DKRZ, 1992)), Last Glacial Maximum (LGM, 21 ka, ECHAM3) and Last Interglacial (LIG, 120 ka, obtained from Otto-Bliesner and collaborators (2006)).

To reconstruct stable areas of suitable habitat through the Quaternary (refugia) for each species, the four SDMs were transformed into presence-absence rasters by using threshold values where sensitivity equals specificity in the model. This process maximizes the agreement between modeled and observed distributions, alleviating problems associated with incorrect predictions while also incorporating the benefits of correct predictions (Pearson *et al.*, 2006). During model building, *M. atticolus* records from outside the Cerrado area (Cerrado enclaves in Amazonia) that would behave as outliers were excluded. There were no such cases for *G. amarali* or *T. itambere*.

2.5 Ancestor geographic location estimation (Hypothesis 4)

To estimate the geographic location of ancestral populations for each species complex, we used a likelihood method implemented in the program PhyloMapper v1 (Lemmon & Lemmon, 2008). This method reconstructs phylogeographic history by estimating the geographic locations of ancestors based on an ultrametric gene tree topology and the geographic locations of the corresponding tip individuals. The statistical framework employed by PhyloMapper was developed to estimate the history of a single locus and hence benefits from large sample sizes and well-resolved trees (Lemmon & Lemmon, 2008). Thus, for each species, all unique cytb haplotypes were used to run an ML phylogenetic analysis in RAxML v8.1.1 (Stamatakis, 2014), and the resulting ML trees used as inputs for PhyloMapper. For all the RAxML runs, a GTRGAMMAI model was assigned, and nodal support assessed using 1000 rapid bootstraps followed by a thorough ML search. In PhyloMapper, exact geographic

locations georeferenced in decimal degrees were assigned to each tip (Appendices 17, 18 and 19). The trees were rate-smoothed using 1000 replicates to ensure a global optimum was found, applying the default (1.0) age for the root of the trees (i.e., obtained an ultrametric genealogy). Finally, the ancestor (ingroup root) geographic locations were estimated by optimising the parameters using 1000 ML search replicates. The altitude of the estimated ancestral location was extracted for each taxa from the elevation raster used to model their current distribution (section 2.4) using the R package *raster* (Hijmans, 2014).

2.6 Population genetics summary statistics (Hypotheses 9 and 10)

Population genetics summary statistics were calculated for each species from the cytb data using DnaSP v5.0 (Librado & Rozas, 2009). These descriptive statistics were calculated to specifically test hypotheses 9 and 10 (Table 1), thus we grouped cytb samples into two groups based on their location in either plateau or valley areas.

The number of haplotypes (h), haplotype diversity (H_d) and nucleotide diversity (π) were estimated for each group within the three species complexes. To detect signs of population expansion into plateaus or valleys, we calculated F_s (Fu, 1997) and R_2 statistics (Ramos-Onsins & Rozas, 2002) and tested their departures from neutrality (statistical significance) using 1×10^4 coalescent simulations. We used cytb data only for these estimations because of the higher number of available samples and geographic locations in both plateaus and valleys.

Samples were tentatively separated between plateaus and valleys based on the elevation at their collection site (Appendices 17, 18 and 19): samples where elevation was >500 m above sea level (asl) were considered to belong to plateaus, and samples <500 m asl to belong to valleys. This threshold has been successfully used to investigate broad community biogeographic patterns in the Cerrado before (Silva, 1996; Silva & Bates, 2002; Nogueira *et*

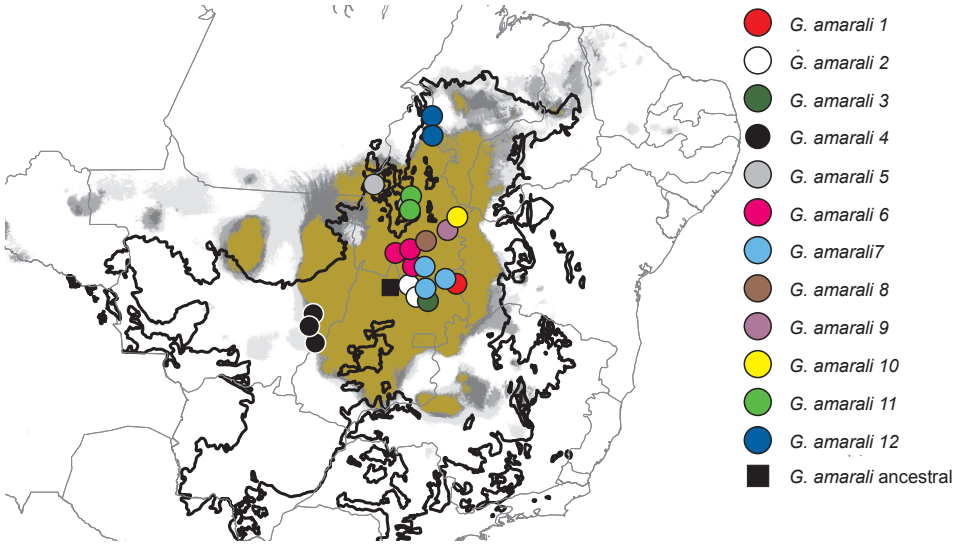
al., 2011), and areas found below 500 m asl appear to be correlated with regions eroded by major river drainages (Silva *et al.*, 2006).

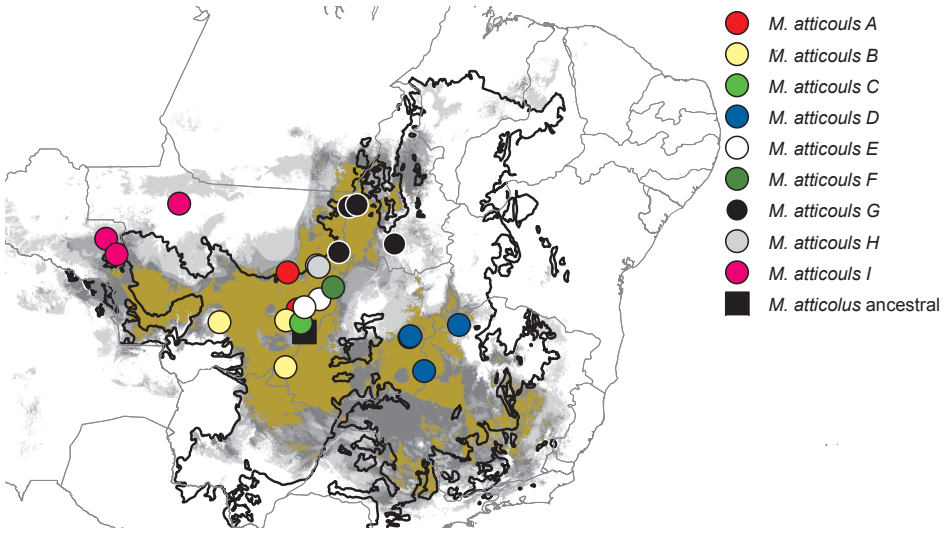
2.7 Temporal diversification congruence tests (Hypotheses 6 and 12)

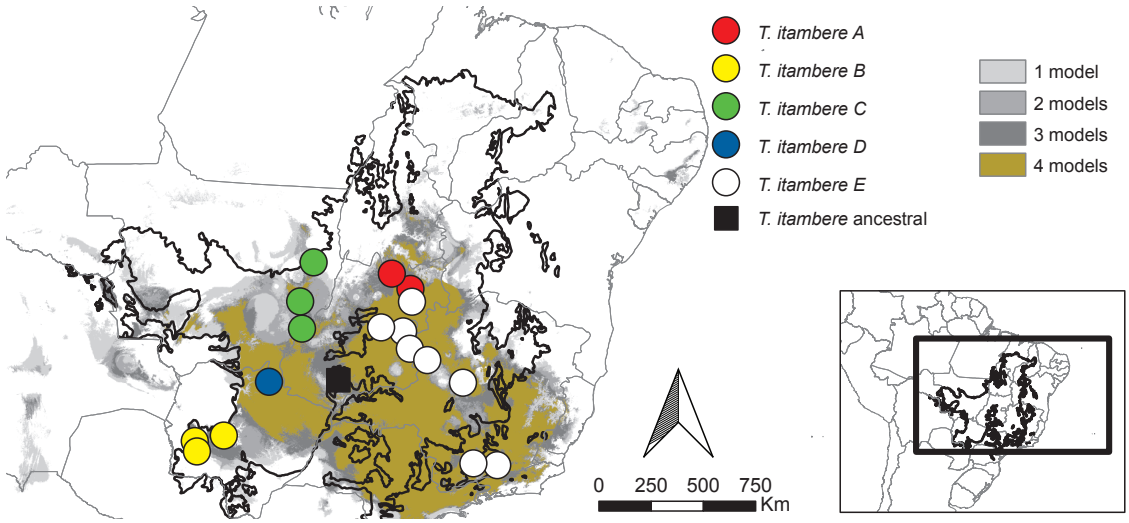
2.7.1 Selection of clades

The selection of clades to test for synchronous diversification of taxa among putative refugial populations (hypothesis 12) was straightforward: *M. atticollus* and *T. itambere* had similar stable areas with a clear gap separating populations in the east from those in the west (Fig. 1). Thus, monophyletic sister clades were selected on each side, and the tests conducted using *T. itambere* clades A *versus* B+C and D *versus* E. Because we wanted to test for the role of Quaternary environmental-climatic fluctuations as drivers of vicariant events, we included the separation between *M. atticolus* populations in stable (refugia within Cerrado) and non-stable areas outside Cerrado (putative refugia within Amazonia), so that divergence between *M. atticolus* clades H *versus* I was included in this analysis. Unfortunately, we could not include *M. atticolus* clades C *versus* D, because clade C had a sample size of only one individual. Samples of *G. amarali* were not included in tests of this hypothesis because its refugia was inferred to be located in the centre of the Cerrado biome, making it difficult to identify any genetic breaks that might be associated with Quaternary climatic dynamics (Fig. 1).

Fig. 1: Partial map of Brazil with the *Gymnodactylus amarali* species complex (upper), the *Micrablepharus atticolus* complex (middle) and the *Tropidurus itambere* complex (bottom) sample sites in the context of the distribution of the Cerrado (black outline). Different colours indicate clades used as extant populations in G-PhoCS analyses. Brown area depicts Quaternary refugia, i.e. overlap between four SDMs inferred from current, mid-Holocene (6 ka), LGM (21 ka) and LIG (120 ka) climatic data. Ancestral geographic location for each species complex as inferred by PhyloMapper is also shown (black squares).


G. amarali 1
G. amarali 2
G. amarali 3
G. amarali 4
G. amarali 5
G. amarali 6
G. amarali7
G. amarali 8
G. amarali 9
G. amarali 10
G. amarali 11
G. amarali 12
G. amarali ancestral


M. atticouls A
M. atticouls B
M. atticouls C
M. atticouls D
M. atticouls E
M. atticouls F
M. atticouls G
M. atticouls H
M. atticouls I
M. atticolus ancestral


T. itambere A
T. itambere B
T. itambere C
T. itambere D
T. itambere E
T. itambere ancestral
1 model
2 models
3 models
4 models
0 250 500 750
Km

Selecting clades for testing the hypothesis that the deepest population divergence within each species occurred simultaneously was not so straightforward. This test was envisioned to investigate congruent patterns of diversification related to Neogene geological events, which potentially played a similar role in initiating population divergence for all three of our codistributed focal species. In other words, if early geological evolution in the Cerrado influenced these species in a similar fashion, we would expect a synchronous pattern of divergence. We devised two approaches for sampling design: 1) selecting the earliest clade (as estimated by ML RRHS) in comparison to all the rest; and 2) selecting populations based on the oldest estimated divergence time (as estimated by G-PhoCS, section 3.2) and high genetically divergent populations (based on cytb genetic distances among them, Chapter 3, Tables 3, 4 and 5). Results from both approaches were very similar, and our runs suggested that the second approach provides better confidence intervals on the hyper-parameters of the model (see below). While both experimental designs appear to result in models with similar amounts of evolutionary information, we believe the second approach to be statistically more robust and report only the results from this approach. Thus, the clade pairs selected for this test were: *G. amarali* clade 1 versus 2, *M. atticolus* clade B versus G, and *T. itambere* clade C versus D.

2.7.2 Testing for simultaneous diversification/vicariance

The use of multiple loci can improve statistical phylogeographic inferences of synchronous divergence among species (e.g. by using the program MTML-msbayes; Huang *et al.*, 2011). On the other hand, MTML-msbayes requires phased data to run, which is not the case of our present AP dataset. As discussed above (section 2.3), phasing a large number of loci can result in substantially high error estimates (Browning & Browning, 2011), unless the procedure is done for a large number of closely related individuals (O'Connell *et al.*, 2014).

Considering the fact that we have multiple cryptic species within our focal taxa (Chapter 3), we refrained from using statistical phasing algorithms, but expect to be able to use new implementations for empirically phasing the AP data in the future (Brandley *et al.*, 2015). At this moment, all the analyses presented below were run using mtDNA data only, which has been successfully used for such types of investigation (Leache *et al.*, 2007; Carnaval *et al.*, 2009).

We tested the hypotheses of simultaneous diversification within the three taxa consistent with Cerrado diversification predictions (Table 1) using MTML-msBayes (Hickerson *et al.*, 2006; Huang *et al.*, 2011). The MTML-msBayes pipeline implements a hierarchical approximate Bayesian computation (ABC) model in which population-pairs from multiple species/lineages diverge from their ancestral populations, while allowing inter-population variation at demographic parameters as well as varying coalescence times and mutation rates among multiple unlinked loci (Huang *et al.*, 2011). We ran MTML-msBayes using a three-step procedure similar to that employed in recently published studies using the method (e.g., Bell *et al.*, 2012; Bagley & Johnson, 2014b; Hickerson *et al.*, 2014). Specifically, for each modelling analysis, we 1) compared candidate priors (i.e., model classes) with and without migration using ABC model choice; 2) estimated the number and temporal congruence of discrete co-divergence events (Ψ) from the best-fit model and using model averaging; and 3) estimated the timing of community divergences at each break where simultaneous diversification was supported.

In order to compare different models using ABC model choice, we first developed two model classes for the analysis for each shared genetic break that we identified from our hypotheses (section 2.7.1). The first model (M_1; 'isolation model') specified vicariance followed by complete population isolation and placed several uniform prior bounds on parameters (see below). The second model (M_2; 'low-migration model') had the same prior

values as the first, except it allowed a limited degree of migration between daughter lineages by setting the upper bound of the migration rate (m) parameter to 1 individual per generation. Following the authors' instructions (see MTML-msbayes manual) and previously published msBayes analyses (e.g., Barber & Klicka, 2010), we set priors for the upper bounds of current population sizes (θ_D) and the ancestral population size (θ_A) based on empirical estimates of nucleotide diversity (π) returned by msbayes. We used twice the within-population π (π_w) estimate for current populations as our standard maximum population size (θ_{max}) prior. To set priors on the ancestral theta multiplier representing the ratio of ancestral population size ($\theta_{anc\text{-}max}$) to θ_{max}, we used the ratio of empirical estimate of the ancestral population size (θ_{anc}) for each of the three species divided by the sum of the empirical θ estimates for all tip clades from our G-PhoCS analyses. Finally, we identified upper bounds for the population-pair divergence time parameters (τ, in units of $4N_{ave}$, where N_{ave} is average population size) in the model from empirical estimates of mean population divergence times (T) output by G-PhoCS. Preliminary analyses suggested that adjusting G-PhoCS τ estimates to units appropriate for msBayes priors gave prior bounds that were too narrow; so we set the upper limit of τ priors to the largest estimated T value across population-pairs.

In the first step of our analysis, we approximated the posterior probabilities of M_1 and M_2 by randomly simulating 5 million samples from each model class with equal probability. Next, we obtained the ABC joint posterior distribution using the default summary statistic vector (D) from MTML-msBayes and rejection sampling to identify the 1000 closest Euclidean distances between the observed summary statistics (D*) for the data and D_i calculated from 10 million random draws across both priors. This procedure outputs the approximate posterior probabilities $[P(M_K/D)^{1000}]$ of the prior model classes, allowing ABC model choice (Hickerson *et al.*, 2014). We compared the approximate posterior probabilities of the model classes to identify the best-supported model with the highest posterior support.

In the second step of the analysis, we estimated the number and temporal congruence of discrete co-divergence events (Ψ) by examining the hyper-posterior probability distributions of Ψ (number of possible assignments of Y taxon-pairs [population-pairs] across Ψ events) and the dispersion index of population divergence times ($\Omega = Var[\tau]/E[\tau]$; the ratio of variance to the mean of the divergence times, where $E[\tau]$ is mean divergence time) from independent runs of the 'best-fit' models. We also used the results from the above procedures to estimate Ψ and Ω that were weighted, by ABC model averaging, on the posterior probability of the two prior model classes (Huang *et al.*, 2011; Hickerson *et al.*, 2014). Following previous studies (e.g., Leache *et al.*, 2007; Bagley & Johnson, 2014b), we conducted hypotheses testing by comparing the posterior probabilities for the expected values of the hyper-parameters under a 'null' scenario of asynchronous diversification (H_0: $\Psi > 1$, and $\Omega > 0.05$) against the alternative of simultaneous diversification (H_A: $\Psi = 1$, and $\Omega < 0.05$). We also evaluated support for these hypotheses by comparing B_{10} Bayes factors calculated under the parameter thresholds above while accounting for prior support for the hypotheses, using established criteria for B_{10} "weight of evidence" (Kass & Raftery 1995). During interpretation, we placed our confidence in Ω because it has been shown to outperform Ψ in correctly rejecting simultaneous divergence, even over very recent coalescent timescales (Hickerson *et al.*, 2014). Ω also correctly rejects simultaneous divergences over a range of conditions, including with large or small sample sizes (Hickerson *et al.*, 2007).

In the third step of our analysis, we estimated the timing of community divergences at each break where simultaneous diversification was supported by our modelling results (i.e. the synchronous diversification among taxa). We converted best-fit-model and model-averaged $E[\tau]$ estimates (which are in coalescent units of $4N_{ave}$ generations, where N is mean N_e) to absolute time (T_{div}) using the equation $T_{div} = E[\tau] \times (\theta_{ave}/\mu) \times g$, where μ is the mutation rate per gene per generation, θ_{ave} is the midpoint of the θ prior, and g is generation time

(assumed to be 1). In these conversions, θ_{ave}/μ is an estimate of $4N_{ave}$, and we used μ values calculated as mean per-gene mutation rates averaged across the species/lineages included in the analysis.

3. Results

3.1 Phylogenetic relationships (Hypotheses 3 and 5)

The topologies of the RRHS ML trees were very similar to ML and Bayesian trees topologies estimated with IUPAC notations of ambiguous characters in Chapter 3, a result found for the three species complexes (Fig. 2, 3 and 4). The topologies were identical among all samples for the *G. amarali* complex (Fig. 2, Chapter 3 – Fig.2) and, for the other two taxa, the few differences in topology took place among samples within clades. For instance, in the ML tree of *M. atticolus* (Chapter 3 – Fig. 3) the clade D individual from Arinos is sister to all other samples, while in the RRHS ML tree an individual from Paracatu is sister to all other samples (Fig. 3). For *T. itambere*, RRHS ML clade E individuals from Brasília were sister to individuals from Cristalina and Pirenópolis (Fig. 4), while they were found in an increasing order in Chapter 3 (Chapter 3 – Fig. 4). These are not unexpected results, since relationships among shorter branches are more difficult to resolve (Wiens *et al.*, 2008).

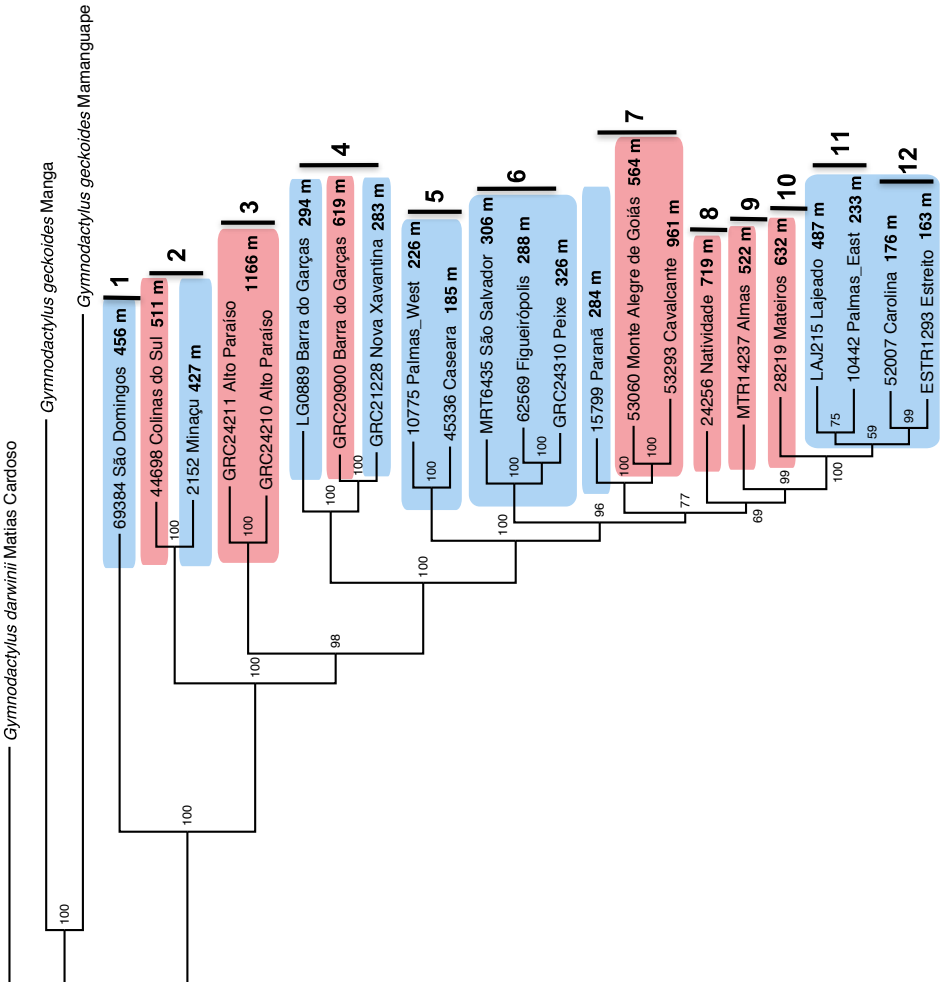


Fig. 2: Phylogenetic relationships for the *Gymnodactylus amarali* species complex. The phylogenetic tree is a majority rule consensus of 3000 ML trees estimated using Repeated Random Haplotype Sampling (RRHS, see section 2.2 for details). Numbers in nodes are Bootstrap scores, and elevation at sample site is indicated in bold besides each sample. Samples in valleys are highlighted in blue and samples in plateaus are highlighted in red. Grouping numbers indicate clades used as extant populations in G-PhoCS analyses.

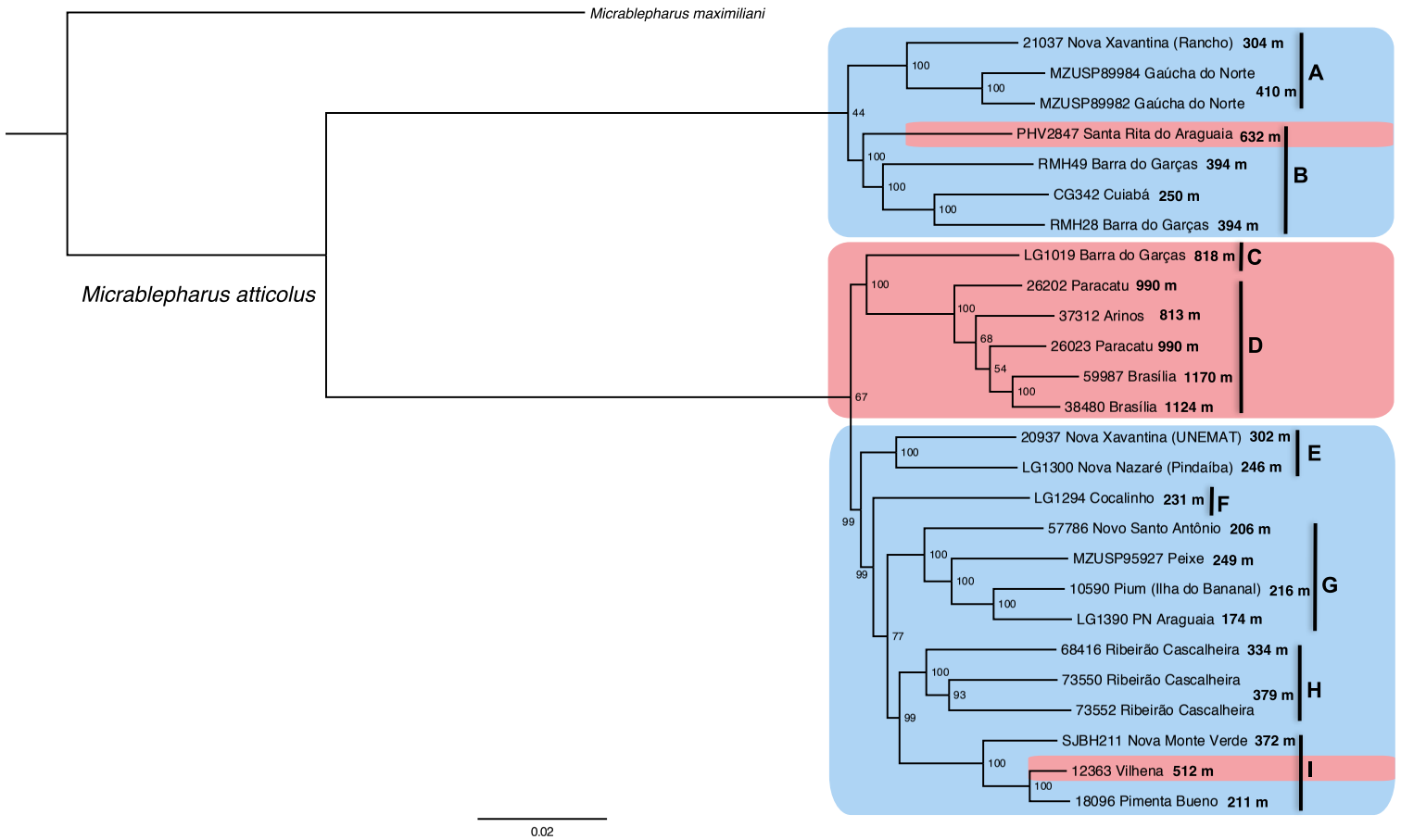


Fig. 3: Phylogenetic relationships for the *Micrablepharus atticolus* species complex. The phylogenetic tree is a majority rule consensus of 3000 ML trees estimated using Repeated Random Haplotype Sampling (RRHS, see section 2.2 for details). Numbers in nodes are Bootstrap scores, and elevation at sample site is indicated in bold besides each sample. Samples in valleys are highlighted in blue and samples in plateaus are highlighted in red. Grouping numbers indicate clades used as extant populations in G-PhoCS analyses.

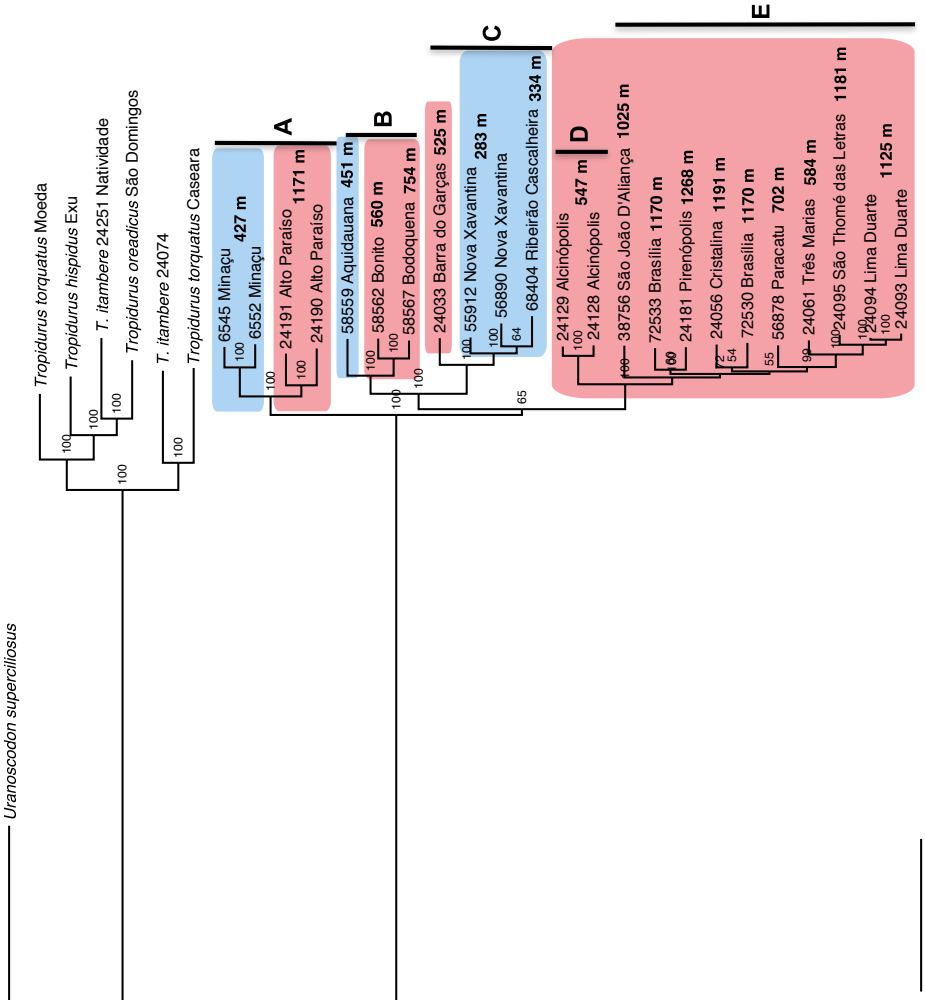


Fig. 4: Phylogenetic relationships for the *Tropidurus itambere* species complex. The phylogenetic tree is a majority rule consensus of 3000 ML trees estimated using Repeated Random Haplotype Sampling (RRHS, see section 2.2 for details). Numbers in nodes are Bootstrap scores, and elevation at sample site is indicated in bold besides each sample. Samples in valleys are highlighted in blue and samples in plateaus are highlighted in red. Grouping numbers indicate clades used as extant

Branch length estimates based on RRHS ML were relatively different compared to the ML tree estimation in Chapter 3. Mean absolute (scaled) branch length differences between the RRHS ML and ML trees were: 0.003 for the *G. amarali* complex, 0.009 for the *M. atticolus* complex, and 0.002 for the *T. itambere* complex. Mean relative (non-scaled) differences were: 0.722 for *G. amarali*, 0.687 for *M. atticolus*, and 0.727 for *T. itambere*. Notably, RRHS ML branches between *M. atticolus* clades A-B and the other clades were much longer than those estimated with ambiguous codes (Fig. 3, Chapter 3 – Fig. 3). However, the independently estimated divergence times (section 3.2) do not seem to reflect such a deep divergence between these clades (Fig. 4).

To investigate predictions of hypotheses 3 and 5, the phylogenetic trees depict the elevation at the site were each sample was collected (Fig. 2, 3 and 4). Our results do not corroborate hypothesis 3 for all species using the 500 m asl threshold to separate plateaus and valleys (section 2.6), i.e. the basal lineages are not necessarily located in plateaus. For the *M. atticolus* complex a separation between plateaus and valleys can be seen since clades C-D are found in different plateaus and their sister group, formed by clades E to I, are mostly found in valleys, corroborating hypothesis 5. In addition, clades A and B are almost exclusively found in valleys. Although not exactly as predicted by hypothesis 5, there is a geographic pattern related to plateaus and valleys for the *G. amarali* and *T. itambere* complexes: most clades are restrict to either a valley or a plateau, and a few have samples distributed both in valleys and plateaus (*G. amarali* clades 4 and 7, and *T. itambere* clades A, B and C).

3.2 Inferences of demographic history (Hypotheses 1, 2, 7, 11, 14)

The complete G-PhoCS runs with 1 x 10^6 iterations took ~2,400 hours for the *G. amarali* complex, ~450 hours for the *M. atticolus* complex, and ~740 hours for the *T. itambere* complex. Because final parameter estimates were very similar among runs (within a species

complex), most runs were aborted after ESS reached values >200 for migration bands, and >1000 for all other parameters. As migration sampling started after 1 x 10^4 MCMC iterations, the migration parameter for all runs had smaller ESS compared to other parameters. In a complete 1 x 10^6 iterations run, migration ESS values were between ~300 and ~450. Accordingly, all G-PhoCS runs converged and had ESS values >1000 for all parameters, and migration ESS values >200. MCMCcoal runs took 40 hours for the *G. amarali* complex, 12 hours for the *M. atticolus* complex and 4 hours for the *T. itambere* complex, and had ESS values >1000 for all parameters.

G-PhoCS runs with and without migration bands returned highly congruent results. The divergence times (*T*, Fig. 5a) and effective population size (N_e, Fig. 5b and 5c) results are reported for runs without migration. Cytb divergence time estimates were always older and had larger confidence intervals compared to AP estimates (Fig. 5a; standard errors for AP estimates were between 2 and 4 orders of magnitude smaller than mean values and do not appear at the presented scale). Species complexes that use rocky outcrops (*G. amarali* and *T. itambere*) had older AP estimated basal divergences (>2 Ma) compared to the leaf litter species (*M. atticolus*, <2 Ma) (Fig. 5a), corroborating hypothesis 14 that predicted similar divergence times for ecologically similar species.

The *G. amarali* AP effective population size estimates for extant populations where generally smaller (<1 x 10^6) than N_e estimates for the other two complexes (Fig. 5b). Estimates of ancestral N_e were generally larger (between 1 x 10^6 and 2 x 10^6) than extant populations estimates for all species, with some *M. atticollus* populations returning the larger estimates (Fig. 5c). Thus, concerning extant N_e among the three taxa, *M. atticollus* and *T. itambere* have more similar estimates, while *G. amarali* and *T. itambere* have more similar ancestral N_e estimates, results that are consistent with predictions from hypothesis 14.

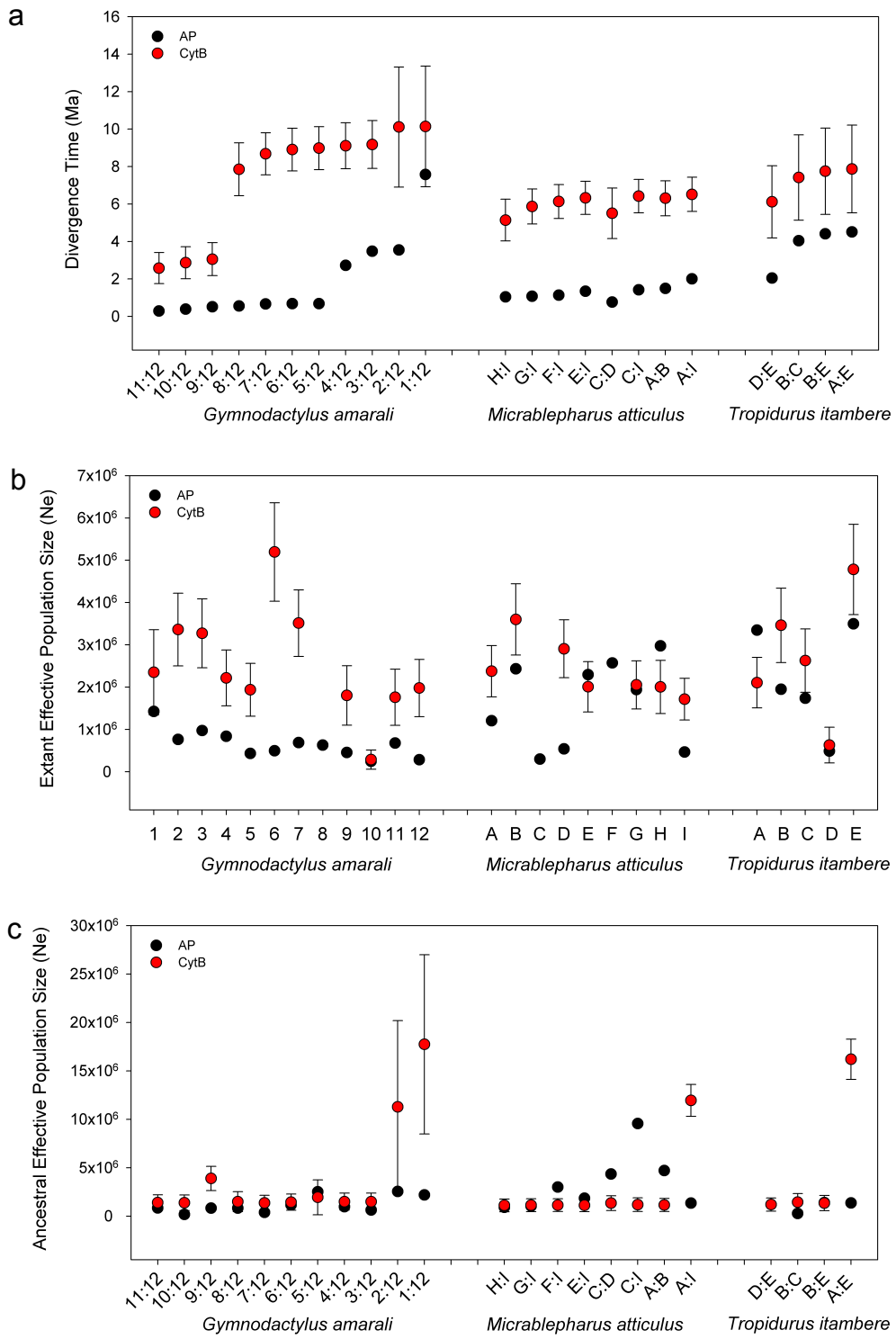


Fig. 5: Results from G-PhoCS (AP) and MCMCcoal (cytb) demographic estimates. a) Divergence times (*T*) for each modelled divergence event for the three species complexes. b) Extant effective population size (N_e), and c) ancestral N_e. Results are from runs without migration. Standard Errors for AP estimates were always between 2 and 4 orders of magnitude smaller than mean values and therefore do not appear at the presented scale. Clades divergence times and ancestral population sizes are identified as the common ancestor between the clades represented in the tick labels.

Results suggest that no general vicariant effects associated with the uplift of the Central Brazilian plateau influenced the three different taxa in a similar fashion (section 3.1). That prevented an adequate test of hypothesis 7 (divergence in clades that share similar plateau-valley distribution occurred as clusters of divergences around the same time).

We did not find support for hypothesis 11, which predicts that populations in valleys have smaller N_e. For the *G. amarali* complex, most clades had very similar estimates, and clades that clearly belong to plateaus (e.g., clades 3, 8, 9 and 10) did not have larger estimates. Differences in N_e estimates were larger among *M. atticolus* clades, but clades belonging to plateaus (C and D) had two of the smallest estimates, whereas clades distributed in valleys (E, F, G and H) had some of the largest estimates. Finally, there was no different pattern between plateaus and valleys in *T. itambere*: clades D and E are both located in plateaus and had, respectively, the smallest and largest N_e estimates.

Estimates of migration (proportion of individuals that arrived by migration per generation) among clades were as low as 0.8%, with zero as the lowest credible interval (lower 95% highest posterior density (HPD)) (Fig. 6). The highest migration rates were found among *G. amarali* clades, ranging from 1% to 10% (Fig. 6). Migration estimates among *M. atticolus* clades ranged from 0.8% to 1.7%, and among *T. itambere* clades from 0.9% to 1.8% (Fig. 6). We found no support for hypotheses 1, which suggests restricted gene flow among populations in different plateaus, or hypothesis 2 suggesting higher gene flow among populations in valleys. For *M. atticolus* and *T. itambere*, migration estimates were very similar between all tested population pairs. For *G. amarali*, contrary to hypothesis 1, migration between populations in plateaus were as high as 10% (e.g., between clades 3 and 2), and ~7% between clades 7 and 8. Furthermore, migration rates between populations in valleys ranged from 1% to 10%, contrary to hypothesis 2 expectations. Although we did not have specific predictions about migration rates between populations in plateaus and valleys, the

results indicate no difference in relation to other estimates. Migration between geographically close clades C and E of *M. atticolus* ranged from 0.8% to 1%, and between clades C and H from 1.4% to 1.3%. For *G. amarali*, migration between the geographically close clades 6 and 7 was ~5%, and between clades 1 and 7 was ~6%, similar to other migration estimates obtained within plateaus or within valleys. All *T. itambere* populations in valleys were within clades already found in plateaus (Fig. 4, section 3.1).

Interestingly, we found an effect of the 'Tocantins' river on gene flow: samples from Palmas (clades 5 and 11) are only ~20 Km apart but separated by the Tocantins river, and the estimated migration between them was 3% (from clade 5 to 11) and 1% (11 to 5). On the other hand, migration between clades 11 and 12, both at the east side of the river but ~450 Km apart, were between 5% and 10%.

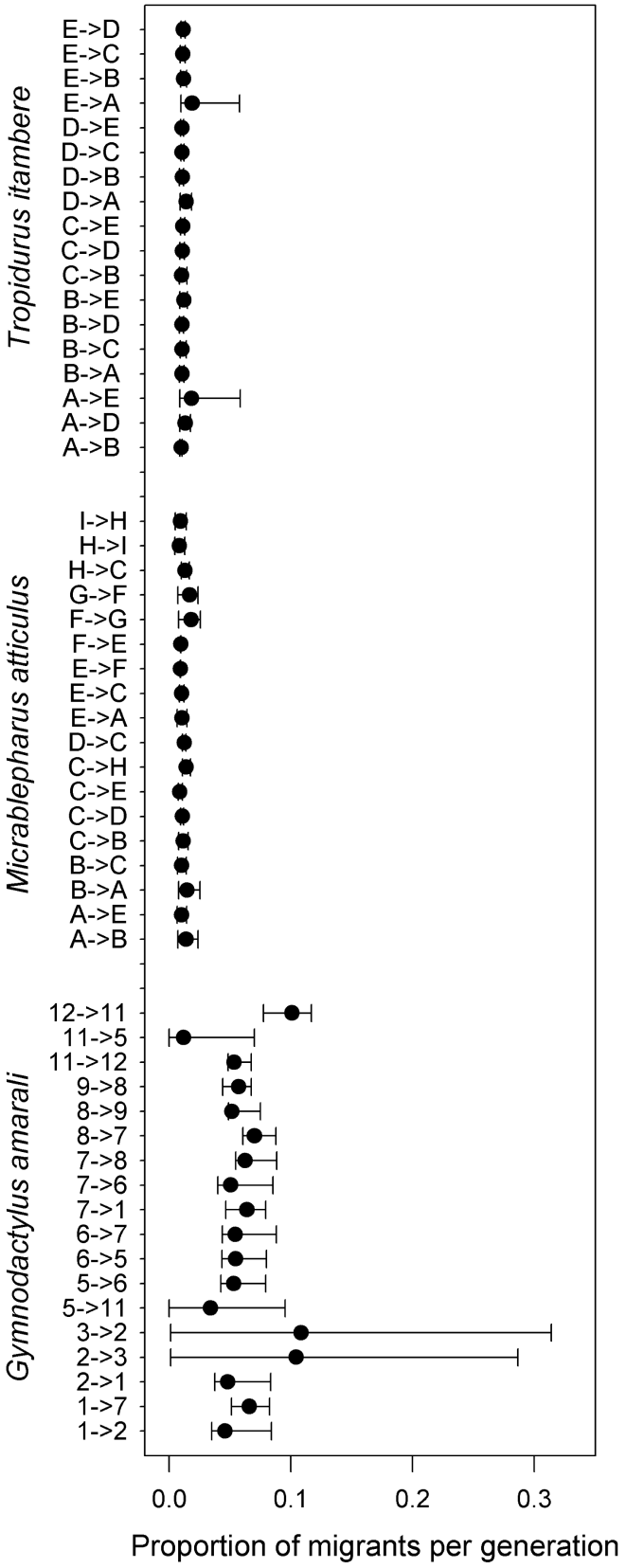


Fig. 6: Migration rates (proportion of migrants per generation (and 95% HPD) between selected clades for each species complex. Arrows in tick labels represent the direction of migration estimate.

3.3 Species distribution modelling (Hypotheses 8 and 13)

In agreement with hypothesis 8, Quaternary stable areas during the four modelled times (refugia) were predominantly located in plateaus for the three taxa (Fig. 1). However, contrary to hypothesis 13 suggesting that *G. amarali* and *T. itambere* palaeodistributional shifts would be similar because of their similar habitat, *M. atticolus* and *T. itambere* recovered refugia were more similar (Fig. 1). For the latter two complexes, a clear separation exists between an eastern and a western Cerrado plateaus (Silva *et al.*, 2006), whereas *G. amarali* refugia includes part of these two plateaus and the valley between them. The *G. amarali* complex distribution over the four modelled moments in the Quaternary (present, mid-Holocene, LGM and LGI) were relatively similar (Appendix 29), whereas for *M. atticolus* and *T. itambere* complexes the LGM model recovered a more restrict distribution (Appendices 30 and 31).

3.4 Ancestor geographic location estimation (Hypothesis 4)

Ancestral location estimated using PhyloMapper for all three taxa appear to be in a central position in relation to the current geographic location of clades, and does not correspond to any sampled site (Fig. 1). Elevation at the estimated ancestral location of the *G. amarali* and *M. atticolus* complexes suggest they were located in valleys rather than in plateaus (Table 2). Hence, hypothesis 4 was only corroborated for the *T. itambere* complex, for which the ancestral location was estimated to be in a plateau (Table 2).

Table 2: PhyloMapper estimated ancestral location and correspondent elevation for the *Gymnodactylus amarali*, *Micrablepharus atticolus* and *Tropidurus itambere* species complexes.

Species complex	Latitude	Longitude	Altitude
G. amarali	-13.59	-49.17	353
M. atticolus	-15.73	-52.37	466
T. itambere	-18.07	-50.71	836

3.5 Population genetics summary statistics (Hypotheses 9 and 10)

We could not reject the null hypothesis of constant population size for any plateau or valley groups using F_s and R_2 statistics for the three taxa (Table 3). Hence, the prediction of hypothesis 9 that signs of population expansion would be found in valleys was not corroborated. Although the number of haplotypes (*h*) is variable between plateaus and valleys, this measure is probably more related to sample size (n), and no expressive difference between haplotype (H_d) or nucleotide diversity (π) was found between population in plateaus and valleys (Table 3). As such, the idea that populations in valleys would have lower genetic diversity (hypothesis 10) was not corroborated.

Table 3: Genetic diversity estimates and population size neutrality tests for the *Gymnodactylus amarali*, *Micrablepharus atticolus* and *Tropidurus itambere* species complexes. Shown are number of populations (n Pop.), cytb sequences sample size (n), number of haplotypes (*h*), haplotype diversity (H_d), nucleotide diversity (π), Fu's (1997) F_s statistic, and Ramos-Onsins & Rozas' (2002) R_2 statistic.

Species complex	n Pop.	n	*h*	H_d	π	F_s (*P*-value)	R_2 (*P*-value)
G. amarali							
Plateaus	10	60	22	0.931	0.080	−0.459 (0.99)	0.101 (0.95)
Valleys	17	91	51	0.977	0.093	−0.751 (0.83)	0.093 (0.92)
M. atticolus							
Plateaus	11	47	29	0.961	0.0362	−0.359 (0.42)	0.106 (0.63)
Valleys	17	79	43	0.963	0.043	−0.526 (0.38)	0.096 (0.64)
T. itambere							
Plateaus	13	48	26	0.976	0.068	−0.218 (0.87)	0.110 (0.97)
Valleys	4	27	17	0.957	0.067	0.023 (0.97)	0.120 (0.84)

3.6 Temporal diversification congruence tests (Hypotheses 6 and 12)

Our ABC model choice results provided substantial support for the low-migration models (M2) as the best-supported models over the isolation models (M1) in both analyses (Table 4). Indeed, posterior support was approximately two to three times greater for the low-migration than for the isolation models.

Table 4: Model comparisons and parameter estimates from ABC model choice and model averaging analyses using MTML-msBayes. Results are presented from two prior model classes ran for each of two analyses of Y population-pairs used to test Cerrado diversification hypotheses (6 and 12, Table 1). Prior models had identical τ, θD, and θA prior distributions P(x), but varied in having either zero migration (M1) versus a low level of post-divergence migration (M2). Approximate posterior probabilities P(MK|D)1000 of each model are given based on 1000 accepted simulated draws from 10 million random draws from both prior models. Results of the best-supported model for each analysis are given in bold face with its posterior probability underlined. Mode and mean Ψ and Ω estimates, and 95% highest posterior densities (HPDs) of Ω, are also shown (see text for hyper-parameter details). Hyper-parameter estimates from model averaging over both prior models are given in the first row of each section.

Prior	$P(\tau)$	$P(\theta_D)$	$P(\theta_A)$	$P(m)$	$P(M_K\|D)^{1000}$	Ψ mode	Ψ mean	Ω mode	Ω mean [95% HPDs]
Deepest population divergence ($Y = 3$)						3	2.504	0.235	0.305 [0.023, 0.595]
M_1	~U(0, 7.57)	~U(0, 0.023)	~U(0, 0.28)	~U(0, 0)	0.2100	–	–	–	–
***M*$_2$**	**~U(0, 7.57)**	**~U(0, 0.023)**	**~U(0, 0.28)**	**~U(0, 1)**	**0.7089**	**2**	**2.056**	**0.170**	**0.345 [0.000, 0.991]**
Quaternary SDM split ($Y = 3$)						1	1.762	0.0109	0.187 [0.000, 0.8468]
M_1	~U(0, 4.51)	~U(0, 0.25)	~U(0, 0.28)	~U(0, 0)	0.3386	–	–	–	–
***M*$_2$**	**~U(0, 4.51)**	**~U(0, 0.25)**	**~U(0, 0.28)**	**~U(0, 1)**	**0.6614**	**2**	**1.911**	**0.00011**	**0.514 [0.000, 2.390]**

In our tests of hypothesis 6 (synchronous deepest population divergence among taxa), the mtDNA data were consistent with the null hypothesis of asynchronous diversification, supporting multiple origins of the deepest population divergences within the three species complexes in our analysis (Fig. 7; Table 4). In the best-fit model, the modal posterior number of co-divergences, Ψ, was 2 (with a similar mean value), and the posterior distribution and Bayes factors of Ω derived from local linear regression (Beaumont *et al.*, 2002) also did not support simultaneous diversification (modal $\Omega = 1.7$; M_2 $P(\Omega < 0.05|D) = 0.14$, and $B_{10} = 0.16$ for $\Omega < 0.05$ versus $\Omega > 0.05$). Likewise, posterior Ψ derived from multinomial logistical regression (e.g., Fagundes *et al.*, 2007) lent little support for simultaneous diversification (M_2 $P(\Psi = 1|D) = 0.35$, and $B_{10} = 1.09$ for $\Psi = 1$ versus $\Psi > 1$). We also calculated Bayes factors for parameter ranges opposite to those above, and these lent definitive support to a scenario of asynchronous diversification (M_2 $P(\Psi > 1|D) = 0.70$, and $B_{10} = 3.67$ for $\Psi > 1$ versus $\Psi = 1$; M_2 $P(\Omega > 0.05|D) = 0.86$, and $B_{10} = 6.41$ for $\Omega > 0.05$ versus $\Omega < 0.05$). The model-averaged hyper-parameter estimates were also indicative of asynchronous diversification ($P(\Omega < 0.05|D) = 0.02$, and $B_{10} = 0.02$ for $\Omega < 0.05$ versus $\Omega > 0.05$; $P(\Psi = 1|D) = 0.041$, and $B_{10} = 0.09$ for $\Psi = 1$ versus $\Psi > 1$), but were more consistent with 3 divergence events, with population structure forming in each species/lineage at different times (modal $\Psi = 3$; Table 4).

In contrast to our findings above for hypothesis 6, the mtDNA results from our tests of hypothesis 12 (divergence between refugia occurred synchronously among different species complexes) were mostly consistent with an inference of simultaneous diversification of the focal clades in Quaternary refugia (Fig. 7; Table 4). However, support for simultaneous diversification here was weaker than that for asynchronous diversification in our tests of hypothesis 6. This was due to conflicts among results, particularly because point estimates of the hyper-parameters from the best-fit model in our test of hypothesis 12 conflicted with one

another (Table 4). Therefore, we based our interpretations of the Quaternary refugia model choice results on Ω. Despite posterior modal $\Psi = 2$, the posterior of Ω of the best-supported model was very close to zero, being much smaller than the 0.05 threshold value accepted as indicating simultaneous divergence in our tests (Fig. 7). Still, Ω posterior values lent only slightly greater support to simultaneous diversification (M_2 $P(\Omega < 0.05|D) = 0.52$, and $B_{10} = 1.07$ for $\Omega < 0.05$ versus $\Omega > 0.05$) than asynchronous diversification (M_2 $P(\Omega > 0.05|D) = 0.48$, and $B_{10} = 0.93$ for $\Omega > 0.05$ versus $\Omega < 0.05$). An inference that population-pairs diverged simultaneously in refugia was more strongly supported by the model-averaged estimates of both hyper-parameters, which agreed in supporting a single divergence event. Specifically, posterior modal Ψ from multinomial logistical regression was 1, and posterior Ω peaked at values less than the 0.05 threshold value, with a modal value of 0.011 (Table 4). However, model-averaged posterior distributions and Bayes factor hypotheses tests again gave relatively weak support for simultaneous diversification ($P(\Psi = 1|D) = 0.40$, and $B_{10} = 1.76$ for $\Psi = 1$ versus $\Psi < 1$; $P(\Omega < 0.05|D) = 0.57$, and $B_{10} = 1.35$ for $\Omega < 0.05$ versus $\Omega > 0.05$).

The best-fit model in the deepest population divergence analysis revealed wide variance in divergence times across species/lineages, with populations diverging over τ values of around 4–8, or the equivalent of approximately 1.3–2.5 Ma (using the θ_A prior to estimate θ_{AVE}). The best model in the Quaternary refugia analysis indicated even wider variance in divergence times, with τ values ranging from around 0–6, which is equivalent to approximately 0–1.69 Ma. We used the modal $E[\tau]$ value from the best-fit model and its 95% HPDs to estimate the timing of community divergences at the only level of our analysis where simultaneous diversification was supported, the Quaternary refugia model. Here, the timing of synchronous divergence in refugia is placed at around 299 ka, with 95% HPDs of 75.2 ka to 1.23 Ma.

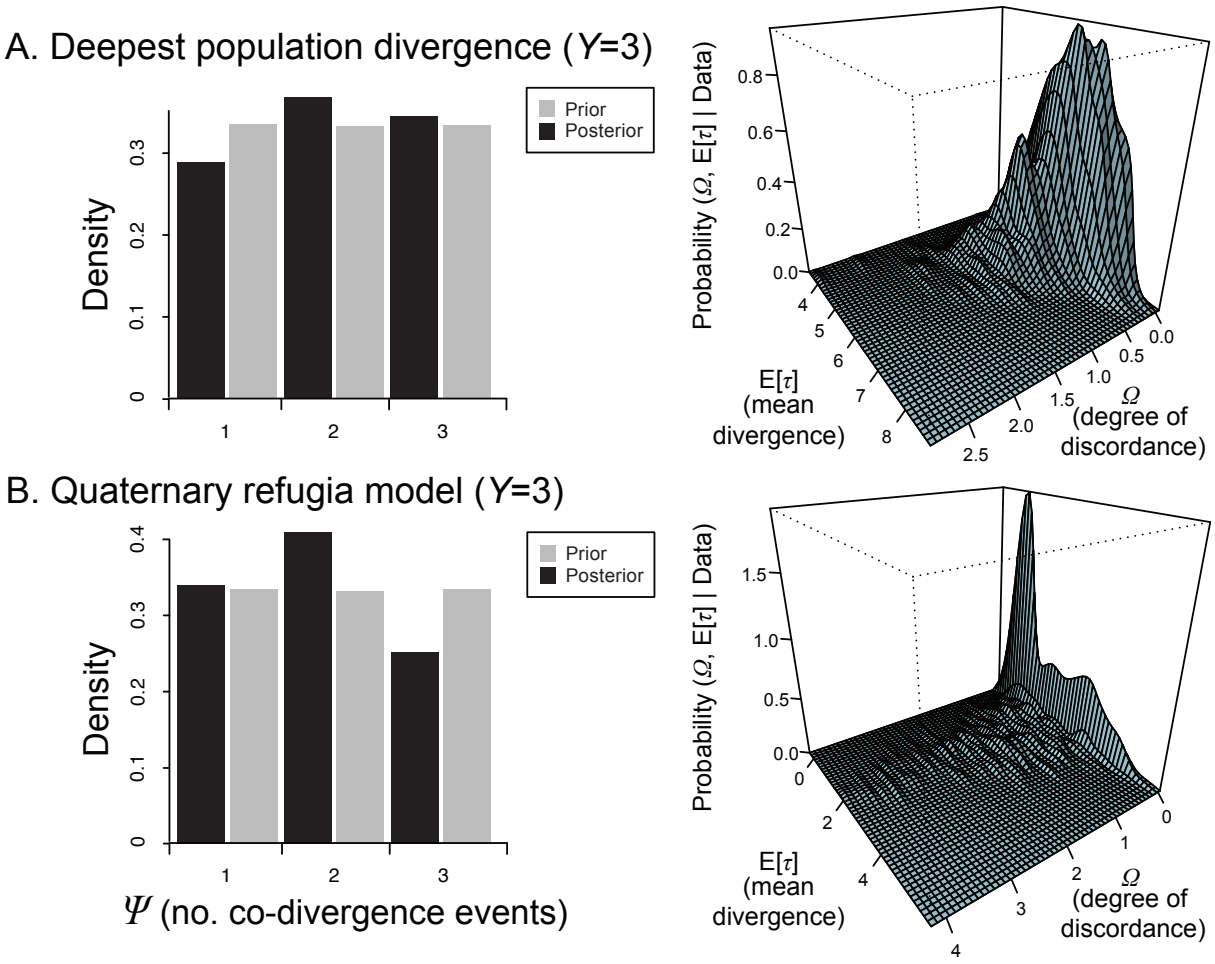


Fig. 7: Results of tests for simultaneous diversification in MTML-msBayes. Left panels show comparisons of the prior versus posterior densities of hyper-parameter Ψ (number of co-divergence events), while surface plots at right show the joint posterior probability densities of Ω and $E[\tau]$. Results are presented only for the best-fit models in each analysis, (A) the deepest population divergence analysis, and (B) the Quaternary refugia model analysis.

4. Discussion

Unravelling the geographical and temporal patterns of diversification of biological entities is an essential step towards understanding how ecology and history affected their demographic processes. The field of comparative phylogeography emerged as an approach to describe concordant historical patterns that could be used to explain evolutionary patterns of a whole biotic region (Avise, 1992). Nonetheless, more recent developments suggest that discordant historical patterns bear invaluable information about complex evolutionary processes (Bernatchez & Wilson, 1998; Leache *et al.*, 2007). Comparing codistributed organisms with different ecological characteristics in a statistical phylogeography framework provides a powerful approach to investigate evolutionary patterns in highly diverse Neotropical biomes (Smith *et al.*, 2014b).

Our statistical phylogeographic analyses and hypothesis-testing framework indicate that codistributed species of Cerrado lizards exhibit some concordant evolutionary patterns. Moreover, the discordant patterns appear to be related to the biology of each species, since ecologically similar species had different Quaternary refugia patterns, but more similar ancient divergences probably related to Neogene geological events. Because the relative impacts of Neogene and Quaternary events in the evolution of the Cerrado endemic biota is yet unclear (Colli, 2005; Werneck, 2011), and to facilitate the interpretation of the many specific hypotheses assessed, we divided our discussion in these two timeframes.

4.1 Neogene tectonic events and their influence on the diversification of Cerrado endemic lizards (hypotheses 1 to 7)

The main geological event proposed to account for diversification in the Cerrado is the epeirogenic uplift of the Central Brazilian Plateau (Colli, 2005; Werneck, 2011). This event was used to explain general patterns of squamate species diversity in the Cerrado (Nogueira *et*

al., 2011), and to assess patterns of bird community biogeography (Silva, 1995; Silva, 1996; Silva & Bates, 2002). In addition, studies have used it as *post-hoc* interpretations to account for the diversification of lizards (Giugliano *et al.*, 2013; Chapter 2). Due to the very broad spatial scale of the Cerrado, we used several specific hypotheses to explore diversification patterns in a clearer but more in-depth investigation. This framework provided us with the opportunity to replicate tests within and across species complexes using two scales of investigation: a full comparative test of synchronous divergence among the three species complexes (hypothesis 6), and several tests of predicted patterns of divergence considering the plateaus and valleys generated by the uplift, which were replicated across the three taxa (hypothesis 1-5).

For the comparative test we used the deepest divergence between two clades within each species in an attempt to attain information on wether their onset of diversification occurred simultaneously in the Cerrado. The main assumption behind this test is that the tectonic uplift would be the first vicariant event to account for the diversification of endemic species, regardless of their past distribution within the biome. The latter was based on the fact that the uplift created a landscape compartmentalisation that can be observed throughout the whole distribution of the biome (Ab'Sáber, 1998; Silva *et al.*, 2006). We found no support for this hypothesis, as we recovered multiple temporal origins of the deepest population divergences within the three species complexes (Fig. 7).

Examining the phylogenetic relationships recovered for the three species complexes (Fig. 2, 3 and 4), it seems unlikely that the tectonic uplift played a central role as a vicariant process in the speciation patterns within these widespread Cerrado endemics, at least as described by Werneck (2011). Werneck (2011) proposed a biogeographic vicariant process where the uplift would have separated populations between valleys and plateaus, generating reciprocally monophyletic clades in these two areas. We found no clear support for this

hypothesis for the *G. amarali* (Fig. 2) or *T. itambere* complexes (Fig. 4). For the *M. atticolus* complex this picture is clearer, and clades C-D are each found in a different plateau, and their sister group (clades E to I) are mostly found in valleys (Fig. 3).

Furthermore, considering that plateaus are geologically older, and were subsequently excavated by the river courses (Ab'Sáber, 1998), we also expected that older lineages would be found in plateaus, and that the ancestral distribution of endemic species should be located in plateaus. Again, we found no support for neither hypotheses (Fig. 2, 3 and 4; Table 2), except for the *T. itambere* complex in which the ancestral location was estimated to be in a plateau. With few exceptions, most of the *T. itambere* complex distribution is within plateaus (Appendix 19), and the few populations located in valleys are part of cryptic species that also have populations in plateaus (Chapter 3, Clades A, B and C), which suggests that the uplift did not drove most of the speciation in the *T. itambere* complex.

Expanding on the compartmentalisation scenario, we proposed that populations inhabiting the more interconnected valleys (Silva *et al.*, 2006) would show higher levels of gene flow, while populations in isolated plateaus would have restricted gene flow among them. None of these hypotheses were supported by our results. Migration estimates within the *M. atticolus* and *T. itambere* complexes had very little variation (Fig. 6), while migration between the *G. amarali* complex clades in valleys varied from 1% to 10%, and between clades in plateaus from 7% to 10% (Fig. 6).

Because clades inhabiting valleys and plateaus were generally not reciprocally monophyletic (Fig. 2, 3 and 4), the distribution of our clades prevented a test for simultaneous diversification between plateaus and valleys using MTML-msbayes, or even an interpretation of this scenario using the estimated divergence times (hypothesis 7, Table 1). The final stage of the uplift in the late Pliocene (Werneck, 2011) pre-dates all *M. atticolus* divergence time estimates (Fig. 5a), and no shared splits can be found in the *G. amarali* and *T. itambere*

phylogenies when they are found in the same plateaus and valleys (e.g., 'Alto Paraíso' and 'Minaçu', and 'Barra do Garças' and 'Nova Xavantina', Fig. 2 and 4). Hence, testing such prediction would require a within clade test using the *T. itambere* complex samples of 'Alto Paraíso' and 'Minaçu' compared to the *G. amarali* complex samples from the same localities. Nonetheless, given the highly different evolutionary distances between these localities for *T. itambere* compared to *G. amarali* (Fig. 2 and 4), it is unlikely that a synchronous diversification pattern would be found. Other option would require using samples from eastern 'Mato Grosso' ('Barra do Garças, 'Nova Xavantina' and 'Ribeirão Cascalheira') for the three species but, once more, the samples from the *G. amarali* and *T. itambere* complexes are all within a clade (Fig. 2 and 4), highly contrasting with the pattern observed for the *M. atticolus* complex (Fig. 4). Thus, considering the distribution of cryptic species of the three taxa (Chapter 3), the uplift of the Central Brazilian Plateau does not seem to have created one or a few major vicariant events that could have influenced speciation within the three species complexes.

Notwithstanding, our results indicate an intricate history of several colonisations of plateaus and valleys, especially within the *G. amarali* complex (Fig. 2) for which the highest migration rates were estimated (Fig. 6). Interestingly, a history of several colonisations of plateaus and valleys was previously proposed for *M. atticolus* using a mtDNA phylogeny (Santos *et al.*, 2014). However, from our three focal taxa, *M. atticolus* is the only which appears to have a relatively concordant distribution of monophyletic groups in either plateaus or valleys (Fig. 3): clades C-D are found in different plateaus, whereas their sister group clades E to I are mostly found in valleys, and clades A and B are almost exclusively found in valleys. Additionally, all the *M. atticolus* complex divergence time estimates (Fig. 5) are placed after the final stages of the tectonic uplift in the late Pliocene (Werneck, 2011), suggesting that its dispersal in the landscape happened after the compartmentalisation. Given

the high diversity of cryptic *M. atticolus* species found in eastern 'Mato Grosso' (Chapter 3), a deeper investigation of its dispersal patterns in the landscape is warranted.

In summary, if the compartmentalisation created by the uplift of the Central Brazilian plateau generated vicariant events between plateaus and valleys (Werneck, 2011), or between different plateaus (Nogueira *et al.*, 2011), these events clearly influenced our species in different ways. An appropriate strategy to better investigate such predictions will be using a large suite of distinct and more complex diversification models (Pelletier & Carstens, 2014; Robinson *et al.*, 2014), associated with further sampling of individuals and localities, and test them for each species complex individually.

4.2 Quaternary climatic fluctuations and their influence on the diversification of Cerrado endemic lizards (hypotheses 8 to 13)

Traditional diversification hypothesis concerning the Quaternary climatic fluctuations in the Neotropics suggests that during extremely dry periods the forests (i.e., wet-adapted vegetation) would retract and be restricted to refugia of suitable climatic conditions. This would create vicariance among these presumably geographically discrete areas (Vanzolini & Williams, 1981). A similar process would happen with dry-adapted vegetation, like the Cerrado and Caatinga, during extremely wet periods (Vanzolini & Williams, 1981). In its early form, this hypothesis was used to explain speciation and distribution patterns of lizards (Williams & Vanzolini, 1966; Vanzolini, 1968a) and birds (Haffer, 1969) in Amazonia. Likewise, many recent studies in the Cerrado and Caatinga used this refugia hypothesis as *post-hoc* explanations for phylogeographic patterns found among species of trees (Collevatti *et al.*, 2003; Ramos *et al.*, 2007; Collevatti *et al.*, 2012a), herbaceous plants (Collevatti *et al.*, 2009; Barbosa *et al.*, 2012) and *Drosophila* (Moraes *et al.*, 2009; Franco & Manfrin, 2013). However, studies on orchids (Pinheiro *et al.*, 2014), cacti (Bonatelli *et al.*, 2014), *Drosophila*

(de Ré *et al.*, 2014) and lizards (Werneck *et al.*, 2012a; Santos *et al.*, 2014) have benefited from the use of SDMs that, allied with genetic information, provide a more robust framework to investigate the effects of Quaternary climatic fluctuations in Neotropical biomes.

Here, we used the similar SDM patterns retrieved for the *M. atticolus* and *T. itambere* complexes (Fig. 1) to test the hypothesis that divergence among recovered refugia occurred synchronously. Results corroborated this hypothesis (#12, Table 1), and indicated that divergence among the tested clades-pair occurred synchronously (Fig. 7, Table 4) at around 299 ka. While this MTML-msbayes time estimate differ from our estimated divergence times using G-PhoCS (Fig. 5), this result is likely due to the different models of each program, and the different datasets used. Furthermore, the fact that this ~299 ka estimate pre-dates the available climatic datasets used to infer SDMs (current to 120 ka) should not be viewed as a contradiction, since climatic fluctuations probably happened in the whole of the Pleistocene (Pennington *et al.*, 2004; Graham, 2011). The effects of such Pleistocene fluctuations on the distribution of both the *M. atticolus* and *T. itambere* complexes would probably generate similar refugia, as indicated by the similar models recovered for all four temporal windows (Appendices 30 and 31).

A single more restricted refugia was recovered for the *G. amarali* complex (Fig. 1). While the reasons for this pattern are not entirely clear, it was very similar to that found for *Phyllopezus pollicaris* (Werneck *et al.*, 2012a), another Neotropical gecko also found in rocky outcrops. The distribution models of *G. amarali* did not change much during the Quaternary fluctuations (Appendix 29) and, while *P. pollicaris* is also distributed in the Caatinga and Chaco biomes, its recovered refugia within the Cerrado was also restricted to a single area (Werneck *et al.*, 2012a). That area was actually geographically similar to the *G. amarali* refugia in the centre of the Cerrado. Interestingly, levels of genetic divergence between *P. pollicaris* Cerrado populations (Werneck *et al.*, 2012a) are similar to those within *G. amarali*,

which are generally higher than those estimated for *M. atticolus* and *T. itambere* (Chapter 3). It is feasible that the habitat use patterns and the more cryptic and crepuscular behaviour of these gecko species (Colli *et al.*, 2003a; Recoder *et al.*, 2012), in contrast with the more active and diurnal *T. itambere* (which also uses rocky outcrops) (Faria & Araujo, 2004), influences their distribution and migration patterns in the Cerrado (Fig. 1, Fig. 6). Because SDMs are generated from data of the actual sampling sites, these different behaviours and distribution patterns might explain why our two species complexes that use the same habitat (rocky outcrops), *T. itambere* and *G. amarali*, had such different Quaternary refugia, contrary to our initial expectations (hypothesis 13, Table 1).

The two main refugia areas recovered for the *M. atticollus* and *T. itambere* complexes are highly coincident with the two eastern and western plateaus recognised as landscape units in the Cerrado (Silva *et al.*, 2006), whereas the *G. amarali* complex refugia mainly include the eastern plateau, a small part of the western plateau and part of the valley between them (Fig. 1). Although the modelled refugia for our three focal taxa appear to be restricted to plateaus, corroborating our initial prediction (hypothesis 8, Table 1), they do not match the refugia distribution recovered for the Cerrado biome using the same environmental variables, which were also mainly restricted to plateaus (Werneck *et al.*, 2012b). Criticism have been raised about the biogeographical significance of building whole biomes distribution models because there might not be a match between species and biome distribution (Collevatti *et al.*, 2013). Nonetheless, our SDMs indicate that the most suitable Cerrado habitats during Quaternary climatic fluctuations were probably located in plateaus, at least for endemic lizard species.

Considering that valleys were mostly unstable during Quaternary fluctuations, populations inhabiting valleys could potentially be subject to different extinction or bottleneck events, and extant populations should therefore show signs of population

expansion, less genetic diversity and smaller N_e (hypotheses 9 -11, Table 1). Nevertheless, population genetic summary statistics (Table 3) and coalescent demographic estimates (Fig. 5) indicated no such differences between populations in valleys and plateaus. Altogether, these results suggest that populations in valleys were not recently colonised. Climatic unsuitability as predicted by SDMs probably does not represent an ideal surrogate for population dynamics in tropical habitats, as opposed to patterns observed for temperate organisms (Soltis *et al.*, 2006; Knowles *et al.*, 2007; Rodriguez-Sanchez *et al.*, 2010). Instead, it is possible that micro-habitat characteristics in the Cerrado might have sustained populations and genetic diversity through adverse climatic periods (Scheffers *et al.*, 2014).

4.3 Do ecologically similar endemic lizards present similar evolutionary patterns in the Cerrado? (hypothesis 13 and 14)

Because *G. amarali* and the *T. itambere* complexes use the same habitat (rocky outcrops), and are found in full sympatry in some sampled sites, we expected that they would have experienced similar palaeodistributional shifts during the Quaternary climatic fluctuations (hypothesis 13, Table 1). As discussed above (section 4.2), that was not the case. The *T. itambere* complex presented very similar responses to Quaternary fluctuations to the *M. atticolus* complex (Fig. 1), which has cryptozoic habits (Vieira *et al.*, 2000) that are very different from both *T. itambere* and *G. amarali*. On the other hand, both the *M. atticollus* and *T. itambere* complexes have wider distributions compared to the *G. amarali* complex, which is mostly found in a central position in the Cerrado (Fig. 1). This distribution pattern is probably the reason why *G. amarali* recovered refugia differed from the other two species complexes.

We also predicted that ecologically similar species would present similar divergence times and N_e (hypothesis 14, Table 1). This hypothesis was corroborated, since basal splits

among *G. amarali* clades and *T. itambere* clades were similarly older (between 2 and 8 Ma, Fig. 5a) than divergence time estimates among *M. atticolus* clades (<2 Ma, Fig. 5a). Also, ancestral N_e estimates were similar between the *G. amarali* and *T. itambere* complexes (Fig. 5c), whereas the *M. atticolus* complex had much higher ancestral N_e (Fig. 5). However, the *M. atticolus* and *T. itambere* complexes presented similarly higher extant N_e estimates (Fig. 5), compared to the smaller estimates for the *G. amarali* complex (Fig. 5b). Perhaps, these similar extant N_e estimates between the *M. atticolus* and *T. itambere* complexes reflect similar demographic responses to Quaternary climatic fluctuations, as suggested by their comparable recovered refugia (Fig. 1).

Similar patterns of phylogenetic relationships were found between clades of the *G. amarali* and *T. itambere* complexes: clades in a central plateau in the Cerrado ('Alto Paraíso' and 'Minaçu') were found to be basal lineages, and western clades (in 'Mato Grosso') have a more recent common ancestor with the remaining clades of central Cerrado (Fig. 2 and 4). Although the ancestral geographic distribution of *G. amarali* and *T. itambere* complexes, as inferred by PhyloMapper, were not particularly similar (Fig. 1), they were both located in this central region in the Cerrado biome. These concordant phylogenetic and phylogeographic patterns might partially explain the more similar divergence times and ancestral N_e estimates of *G. amarali* and *T. itambere*. Both *G. amarali* and *T. itambere* sister species are found in the neighbouring biome of Caatinga (Frost *et al.*, 2001; Chapter 2), whereas *M. atticolus* sister species is widespread in the Caatinga, Cerrado and Chaco (Santos *et al.*, 2014). Wether group-specific phylogenetic relationships and biogeography can help understanding and predicting the possible centre of origin of endemic lizard species in the Cerrado deserves further attention.

In summary, species with similar habitats (the *G. amarali* and *T. itambere* complexes) showed concordant phylogeographic patterns in terms of divergence times (Fig. 5a), ancestral

N_e estimates (Fig. 5c) and geographic ancestral distribution (Fig. 1). In addition, they showed more similar diversification histories within the Cerrado biome (Fig. 2 and 4). On the other hand, the similar SDMs (Fig. 1) and extant N_e estimates (Fig. 5b) for the *M. atticolus* and *T. itambere* complexes suggest that the current geographic range, and not the habitat of the species, might better explain their demographic responses to Quaternary climatic fluctuations. The latter is also supported by the synchronous diversification inferred among clades of *M. atticolus* and *T. itambere* in response to Quaternary cycles (section 4.3).

4.4 Evolution and speciation of endemic lizards in the Cerrado

In general, concordant evolutionary patterns were observed for endemic Cerrado lizard species sharing similar habitats (the *G. amarali* and *T. itambere* complexes) when patterns were temporally related to Neogene events (Fig. 1, 5a and 5c). Additionally, species that share similarly wide distributions in the Cerrado (the *M. atticolus* and *T. itambere* complexes) presented more concordant evolutionary patterns related to Quaternary climatic fluctuations (Fig. 1, 5b and 7). Despite these similarities, many predictions on how Neogene tectonic events (section 4.1) and Quaternary climatic fluctuations (section 4.2) influenced the evolution of endemic species were not corroborated. The patterns observed for our three focal taxa indicate that some broad Neogene and Quaternary diversification hypotheses proposed for the Cerrado (Werneck, 2011; and see Table 1) are generalisations that do not necessarily uphold when simultaneously tested for different species.

Taking into account the hypotheses tested for the Neogene, and the distribution of the cryptic species identified for each taxon (Chapter 3, Fig. 1), the uplift of the Central Brazilian Plateau did not have a shared vicariant speciation effect on our focal taxa. Nonetheless, this tectonic uplift clearly generated a compartmentalisation of the Cerrado landscape (Ab'Sáber, 1998; Silva & Bates, 2002), which might have influenced the diversification of our species in

some way that we could not predict and test. The lack of detailed geological information on Cerrado neotectonics makes it difficult to even propose diversification mechanisms in some cases. In the eastern region of the 'Mato Grosso' state, for example, *M. atticolus* lineages from three very divergent groups are found almost in sympatry (Fig. 1 and 3), whereas the same is not true for the *G. amarali* and *T. itambere* complexes (Fig. 2 and 4). The ancestral geographic distribution of *M. atticolus* was also recovered in this region (Fig. 1). Two basic alternative explanations for this pattern are presented: *M. atticolus* may have experienced several speciation events within the area, followed by dispersal to other regions, or lineages that diversified elsewhere have recently colonized the region. Similarly deep divergence times (~2 Ma) and strong population structure were found for a Cerrado frog, *Hypsiboas albopunctatus*, in the same eastern 'Mato Grosso' region (Prado *et al.*, 2012). A phylogeographic study with a detailed sampling of the region for *M. atticolus*, together with morphological data, would probably shed light in the origins of observed patterns. Such investigation would nonetheless be severely hindered by the lack of detailed regional geologic information.

Related to the possible role played by plateaus and valleys in the diversification of Cerrado organisms (Table1, and see section 4.1), the patterns found here for endemic lizards lend little support for a vicariant process generated by the excavation of valleys (Werneck, 2011). Taking the *G. amarali* complex as an example, almost no sister lineages show a plateau-valley relationship. A few cryptic species are specific to valleys (e.g., 5 and 6), and others to plateaus (3), but the majority show a mixed pattern (4, 7, 8-9, 10-11-12). Considering this scenario, the distribution (Fig. 1) and phylogenetic relationships among the clades (Fig. 2), it is difficult to precise if species restricted to plateaus or valleys simply present restrict distributions, or if the tectonic uplift had an actual influence on speciation in the *G. amarali* complex. The *M. atticolus* complex has an apparent pattern or altitudinal

segregation (section 4.1) and, surprisingly, the sister species located in plateaus (C and D) are distributed geographically distant (Fig. 1), whereas other cryptic *M. atticolus* species are mostly located in valleys. Nevertheless, considering the divergence time estimates in the *M. atticolus* complex (Fig. 5a), the final stages of the tectonic uplift in the late Pliocene cannot be attributed as the main event driving speciation in this group.

A possible speciation scenario that is usually not investigated in the Cerrado is the effect of rivers as vicariant events for terrestrial organisms. Rivers are thought to be responsible for speciation and population structure in terrestrial animals in Amazonia (Fouquet *et al.*, 2012; Upham & Patterson, 2012), and there is evidence for an effect of rivers in the cryptic speciation of *Gymnodactylus darwinii* in the Atlantic Forest (Pellegrino *et al.*, 2005). Here, we found an apparent effect of the Tocantins River on gene flow among *G. amarali* clades: samples from 'Palmas', which are only ~20 Km apart but separated by the Tocantins River, showed lower migration rates than those within the same river bank but ~450 Km apart. Although the Cerrado is not characterized by the presence of large rivers compared to the scenario in Amazonia, the effect of rivers in population structure and speciation in the Cerrado clearly deserves further attention.

In agreement with our earlier findings (Chapter 2, Fig. 4), several species-level divergences in the *G. amarali* complex happened in the last 2 Ma (Fig. 5a), which suggests an influence of Quaternary events in the diversification of this group. Similarly, divergence times among the *M. atticolus* complex were all within the last 2 Ma (Fig. 5a). Only for the *T. itambere* complex were all divergences >2 Ma (Fig. 5a), suggesting a stronger influence of Neogene events in the speciation within this group. At this stage, however, the relationship between the identified cryptic species for each complex (Chapter 3) and the inferred demographic and phylogeographic patterns is still unclear. This relationship can be better explored by investigating speciation patterns using alternative hypotheses of demographic

model evolution (Sousa & Hey, 2013; Robinson *et al.*, 2014). Furthermore, detailed species-specific phylogeographies should shed light about biogeographic events that influenced speciation within each species complex (Pelletier & Carstens, 2014; Rittmeyer & Austin, 2015).

5. Conclusion

Although geographical variation in nature has been intuitively correlated with historical processes (Noble, 1927; Nelson, 1978), ecological aspects of the species seem to extraordinarily contribute to the observed biogeographic patterns. Recent studies assessing representatives of the entire animal tree of life showed that life-history reproductive strategies (and not geography) are the main determinants of animal genetic diversity (Romiguier *et al.*, 2014). Accordingly, the ability to move through the landscape seems as the major factor driving speciation in Amazonian birds (Smith *et al.*, 2014b). Phylogeographic studies can serve as a bridge between ecological and historical biogeography (Riddle, 1996; Marske *et al.*, 2013), and comparative approaches can help elucidating the relative roles of ecology and history on the intrinsic evolutionary patterns of a biogeographic region (Bermingham & Moritz, 1998; Moritz & Faith, 1998).

Using comparative phylogeographic approaches, coalescent demographic estimates, and SDMs, we tested several hypotheses accounting for both historical and ecological patterns in the Cerrado biodiversity hotspot. While ecologically similar species showed similar estimates of divergence times and ancestral N_e, species with similar extant distributions had both concordant SDMs and current N_e estimates. Our overall results suggest that the evolution of endemic Cerrado lizards is better understood when taking into consideration historical events recognised for the biome allied with the ecological characteristics of each particular taxon.

Future research should be focused towards the understanding of more fine-scale diversification patterns, such as in the eastern part of the 'Mato Grosso' state, and in testing alternative demographic hypotheses that could link historical events or taxon-specific ecological characteristics to speciation in the Cerrado.

Chapter 5

General Discussion

General Discussion

Since the early days of biogeography, organisms have been grouped and documented in order to understand spatial patterns of biodiversity and to delineate biogeographic units or regions (Sclater, 1858; Wallace, 1869). Wallace himself called the division of the Malay Archipelago "two parts of the primary divisions of the earth" (Wallace, 1869). Biomes, as the modern definition stands, could hardly be considered "primary divisions of the earth". Rather, biome divisions are based on the architecture and community composition of plants, which are correlated with soil and climatic variables (Prentice *et al.*, 1992) and, hence, reflect the adaptation of these plants to such variables. Thus, biomes are natural sub-divisions of terrestrial habitats, and bear characteristic life forms subjected to similar historical and climatic processes for long periods of time (Crisp *et al.*, 2009). This combination of ecological (climatic) and historical (phylogenetic) factors certainly influences the distribution of different organisms on Earth (Holt *et al.*, 2013). Therefore, the study of a biome's endemic biota has the potential to elucidate the action of historical changes in the landscape on the ecological characteristics of lineages, and clarify the resulting patterns of biodiversity (Donoghue, 2008).

Understanding the processes that generated replicated evolutionary patterns among different groups is a topic of great interest to the evolutionary community (Arbogast & Kenagy, 2001). This can provide key insights into central themes in evolutionary biology such as the origin of species (Wiens, 2004; Emerson & Gillespie, 2008), and the demographic processes behind it (Schluter, 2009; van Doorn *et al.*, 2009). In the Neotropics, biomes such as Amazonia and Cerrado have large continental distributions which, allied to their high biological diversity, hinders comprehensive studies that can link historical and ecological factors to diversification patterns (Antonelli & Sanmartín, 2011). Comparative phylogeography presents, thus, an ideal framework for the investigation of such evolutionary patterns (Hugall *et al.*, 2002; Lapointe & Rissler, 2005).

Prior to this thesis, much of our understanding about the origins and evolution of the Cerrado endemic biota was centred in the delineation of mechanisms that generated biodiversity (Colli, 2005; Werneck, 2011). In fact, only a handful of tests about how those mechanisms influenced the evolution of the endemic biota have been published (Collevatti *et al.*, 2009; Collevatti *et al.*, 2012b; Santos *et al.*, 2014). Many hypotheses remained untested and, to date, they have not been tested in a comparative approach aimed at identifying concordant patterns among species. In this thesis I employed a comparative phylogeographic approach and used species delimitation methods to address knowledge gaps about the evolution and diversification of Cerrado endemic lizards.

In this chapter I summarise and integrate key results from the three data chapters and discuss their major implications. Specifically, I discuss the biogeography and evolution of the Cerrado in light of the comparative phylogeography results of this PhD (1), the conservation of Cerrado reptiles (2), and perspectives on cryptic biodiversity of Cerrado lizards (3). I finish the chapter by presenting suggestions for future research directions (4).

1. Biogeography and evolution in the Neotropics: the case of the Cerrado biodiversity hotspot

Despite recent reviews (Rull, 2008, 2011), the discussion about Neotropical diversification timing is still in its infancy because the papers assessing this subject mainly deal with Amazonian diversification patterns (e.g., Moritz *et al.*, 2000; Hoorn *et al.*, 2010; Smith *et al.*, 2014b). In the Cerrado, an apparent dichotomy can be observed: while species-level diversification seems to be very recent for plants (Simon *et al.*, 2009), the opposite pattern is observed for many animal groups, particularly frogs and lizards (Geurgas *et al.*, 2008; Maciel *et al.*, 2010; Prado *et al.*, 2012; Werneck *et al.*, 2012a; Giugliano *et al.*, 2013; Chapter 4). Plant phylogeographies also indicate very recent population divergences compared to animals (de Lima *et al.*, 2014a), even among disjunct and geographically distant populations

(Collevatti *et al.*, 2009; Collevatti *et al.*, 2012b). There are exceptions to this trend, with divergence times of ~3.3 Ma estimated for an endemic Cerrado tree, *Caryocar brasiliense* (Collevatti *et al.*, 2012b). Importantly, the study of Simon and collaborators (2009) was primarily focused on small species of Leguminosae, a group known for showing high diversification rates (Lavin *et al.*, 2005; Delgado-Salinas *et al.*, 2006).

In Chapter 4, most of our divergence time estimates lie within the last ~2 Ma, but some are older than ~4 Ma. The oldest palaeopalynological evidence for the existence of a typical Cerrado vegetation dates from 32,000 years ago (Ledru *et al.*, 2006), which is very recent in evolutionary terms. If strong vegetation shifts were indeed recently observed, as proposed by Simon and collaborators (2009), the deep diversification found for the *Gymnodactylus amarali* and *Tropidurus itambere* complexes mean they were present in the region before the current vegetation appeared (Crisp *et al.*, 2009). This temporal dichotomy among different groups might be an artefact of the reduced number of study organisms, and additional dated phylogenies are needed before a more general temporal pattern can be attributed to the Cerrado.

Using papers published until 2006, Rull (2008) found that ~43% out of 42 Neotropical reptile evolutionarily significant units (ESUs) originated in the Neogene, while ~57% originated in the Quaternary. A detailed examination of his list of papers reveals that only one study, on *Crotalus durissus* (Wuster *et al.*, 2005), included samples from the Cerrado region – these are clustered with all other samples south of the Amazon. This trend, apart from indicating an impressive lack of molecular studies on Cerrado reptiles, suggests that Rull's temporal patterns of diversification are generally based on organisms from forested habitats (the exception, *C. durissus*, uses open-habitats, but invaded South America recently). Using Web of Science searches carried out in June 2015, I conducted a compilation of papers on Cerrado reptiles (endemic or not) that included divergence time estimates, and found that all papers were published from 2008 onwards. My compilation retrieved 21 ESUs, with 81% of

them originated in the Neogene. If we add to this list the results from Chapter 4, the trend changes to 58% of the ESUs originating in the Neogene, substantially approximating it to Rull's general inferred pattern for the Neotropics.

Another recent review, using papers published until 2011, suggests that herpetofauna intraspecific divergence splits are usually much older than estimations made for other taxa in South America (Turchetto-Zolet *et al.*, 2013). However, it also indicates that only 15% of them used a combination of nuclear and mtDNA markers (Turchetto-Zolet *et al.*, 2013). As our divergence time estimates suggest, mtDNA alone retrieves much older splits (Chapter 4, Fig. 5). The advances on DNA sequencing technology (Carstens *et al.*, 2012), and the fact that sequencing large amounts of nuclear DNA is becoming increasingly cheaper (Lemmon & Lemmon, 2013), will probably have an impact on the trends of divergence time estimates for South American reptiles in future publications.

Comparative phylogeographic approaches seek to reconstruct the history of ecological associations between codistributed species, and to compare individual genealogies to understand how the evolution of the landscape can influence the natural history of populations and community structure (Avise, 1998; Bermingham & Moritz, 1998; Arbogast & Kenagy, 2001). The three species complexes studied here constitute appropriate models to investigate Cerrado evolution since they belong to distantly related reptile groups (Gekkota, Lacertoidea and Iguania; Reeder *et al.*, 2015). As such, the probability of obtaining concordant results due to inherent lineage evolutionary responses (i.e., concordant historical patterns because related taxa similarly respond to environmental changes) is practically negligible. Our results indicate no congruent evolutionary patterns related to Neogene tectonic events among the three species complexes, and the landscape compartmentalisation caused by the uplift of the Central Brazilian Plateau probably played different roles in the evolution of each taxon. To the best of my knowledge, the only study to test *a priori* hypotheses about the influence of the uplift in a Cerrado endemic vertebrate was the phylogeography of *Micrablepharus atticolus* (Santos *et*

al., 2014). Our results are in line with their findings: population genetic estimates suggest no genetic diversity or effective population size differences between populations in valleys *versus* populations in plateaus for any of the three taxa.

The vicariant events created by the uplift of the Central Brazilian Plateau in the Neogene (Nogueira *et al.*, 2011) appear to have had an effect in the distribution of the cryptic species, but only *M. atticolus* showed monophyletic clades distributed either in plateaus or valleys, as hypothesized by Werneck (2011). In the case of *G. amarali*, landscape compartmentalisation seems to have created isolated species in different plateaus, following the patterns suggested by Nogueira and collaborators (2011), where endemism is attributed to restrict distribution in plateaus. The distribution of *G. amarali* cryptic species suggests independent colonisations of plateaus and valleys, with cryptic species mostly distributed in both the plateaus and valleys. Similarly, landscape compartmentalisation appears to have affected the distribution of the *T. itambere* cryptic species since different species are found in different plateaus. On the other hand, this complex is mainly distributed on plateaus, and the few populations in valleys are part of cryptic species also found in plateaus. It was not possible to infer wether populations of *T. itambere* went extinct in valleys during Quaternary climatic fluctuations since detecting lineage extinction is not a trivial task (Rabosky & Lovette, 2008; Bokma, 2009; Rabosky & Lovette, 2009; Stadler & Bokma, 2013).

The Quaternary climatic fluctuations also had different effects on palaeodistributional shifts observed for the three species: while the ecologically distinct *M. atticollus* and *T. itambere* had similar palaeodistributional shifts throughout the Quaternary, *G. amarali* refugia were concentrated in a single region in the centre of the biome. Indeed, our comparative test for synchronous divergence caused by Quaternary climatic fluctuations among *M. atticolus* and *T. itambere* clades was corroborated. For Cerrado plants, multiple refugia were inferred for a palm species (de Lima *et al.*, 2014a), while a single central refugia was inferred for an endemic tree (Collevatti *et al.*, 2012b). Moreover, different population genetic patterns owing

to Quaternary climatic fluctuations were recovered for different plant species of the Cerrado (Ramos *et al.*, 2007; Collevatti *et al.*, 2009; Collevatti *et al.*, 2012a; Collevatti *et al.*, 2012c; de Lima *et al.*, 2014a). It appears that vicariant effects and patterns of population expansion and retraction due to Quaternary climatic fluctuations in the Cerrado must be interpreted in light of biological and ecological aspects of each species (Collevatti *et al.*, 2015), since even closely related species with similar ecology can depict strikingly different phylogeographical patterns (Michaux *et al.*, 2005).

2. Conservation of reptiles in the Cerrado

Linking evolutionary knowledge with conservation practices is a major challenge for both scientists and decision makers (Mace & Purvis, 2008). Nonetheless, the field of conservation biology can benefit from detailed genetic assessments (Moritz, 1994; Pearse & Crandall, 2004), from knowledge about biogeography and phylogenetic relationships (Whittaker *et al.*, 2005; Diniz-Filho *et al.*, 2013) and, of course, from detailed taxonomic information (Hey *et al.*, 2003; Mace, 2004). Phylogeographic analysis can establish a relationship between species diversity and intraspecific variation (Avise, 2001), which are fundamental indexes to surpass the Darwinian and Wallacean shortfalls (Diniz-Filho *et al.*, 2013). Moreover, the first crucial step in conservation planning is to have good taxonomy so that the identity and distribution of the organisms to be managed are known (Margules & Pressey, 2000). Species delimitation methods therefore have the potential to play a central role in facing the modern biodiversity crisis (Fujita *et al.*, 2012).

Only one endemic lizard species from the Brazilian Cerrado (*Bachia bresslaui*) has so far been included in the IUCN Red List (IUCN, 2014). A recent species gap analysis to identify conservation targets in Cerrado lizards indicated that out of the 30 endemic lizards, our three focal species are within the three (*T. itambere*), eight (*M. atticolus*), and thirteen (*G. amarali*) most vulnerable species (Silva *et al.*, 2014). Furthermore, 94% of the endemic

lizards have either total or major conservation gaps, lacking adequate protection within legally protected areas (Silva *et al.*, 2014). This disturbing scenario can deteriorate even more in light of the cryptic species identified in Chapter 3. After proper taxonomic description most inferred species will display restrict distributions, probably not covered by legally protected areas.

Our results also suggest very high cryptic diversity in the eastern region of the 'Mato Grosso' state (Chapter 3). This is a highly threatened region due to the expansion of soy crops, being part of the Cerrado-Amazonia ecotone known as the 'arc of deforestation' (Fearnside, 2005; Aldrich *et al.*, 2012). In the more degraded areas of Cerrado, between 50% and 92% of the landscape has already been modified (Cavalcanti & Joly, 2002). The uptake of research outcomes and translation into management policies and conservation actions in Brazil can be a lengthy process (Cavalcanti & Joly, 2002; Klink & Machado, 2005). As such, the rapid description of cryptic species identified in this thesis so they become available for conservation actions (Mace, 2004) should be an essential component of conservation programs of reptiles from the Brazilian Cerrado.

Finally, the use of genomic datasets to clarify patterns of neutral and putatively adaptive (i.e. functional) genetic diversity and intraspecific gene flow will allow for better informed conservation practices in the genomic era (Funk *et al.*, 2012b).

3. Cryptic biodiversity in Cerrado lizards

The field of species delimitation is flourishing, and empirical tests of the available methods are being constantly conducted (e.g., Camargo *et al.*, 2012; Esselstyn *et al.*, 2012; Reid & Carstens, 2012; Rittmeyer & Austin, 2012; Satler *et al.*, 2013; Olave *et al.*, 2014; Zhang *et al.*, 2014). During the final stages of writing this thesis, two assignment-free Bayesian species delimitation methods were described, namely DISSECT (Jones *et al.*, 2015) and STACEY (Jones, 2015), which are implemented within modified *BEAST and BEAST 2 (Bouckaert *et*

al., 2014) packages, respectively. A Bayesian species delimitation method that integrates morphological data into the analysis was also recently published (Solis-Lemus *et al.*, 2015). The latter is a parallel version of the algorithm implemented by BPP v2 (named iBPP for integrated-BPP). How well this myriad of new analyses will perform in relation to one another, and wether they will be able to incorporate genomic datasets, will be crucial characteristics towards the development of the field.

The size of datasets used in species delimitation studies was once described as a limitation because some methods could only correctly identify species boundaries when using a large number of loci (Rittmeyer & Austin, 2012). There is currently no assessment of how an increasing number of loci, possibly presenting very different coalescences, might influence the results of Bayesian species delimitation. A possible shortfall of the species delimitation analyses used in Chapter 3 relies on the possibility that BPP might be over-splitting the lineages (Satler *et al.*, 2013), calling for a more conservative approach when formally describing cryptic species (Carstens *et al.*, 2013). We acknowledge this problem and intend to use other lines of evidence before taking taxonomic decisions (see section 4.2). Nevertheless, our large nuclear dataset will allow for empirical tests and comparisons between different species delimitation methods, such as the new above-mentioned approaches. It will also allow testing for the influence of the number and characteristics of loci used in the analyses.

The identification of cryptic species in the Neotropics has increased substantially with the advent of molecular techniques (Fouquet *et al.*, 2007; Condon *et al.*, 2008; Ceccarelli *et al.*, 2012; Funk *et al.*, 2012a; Fouquet *et al.*, 2014), and this trend will probably continue with intensified research of Neotropical organisms (Scheffers *et al.*, 2012). Results from Chapter 3 suggest that the number of species in three Cerrado endemic lizards is highly underestimated, with morphological data supporting this conclusion for at least one of the groups (Chapter 2). As discussed in Chapter 3 (section 4.4), high diversity levels within previously widespread nominal taxa is not uncommon for other lizard groups in the Neotropics (Pellegrino *et al.*,

2005; Geurgas *et al.*, 2008; Gamble *et al.*, 2011; Gamble *et al.*, 2012) and elsewhere (Fujita *et al.*, 2010; Pepper *et al.*, 2013; Sistrom *et al.*, 2013). Additional research in the Cerrado is expected to reveal many more cryptic species that could not be identified using morphological data alone.

4. Future research directions

Two main topics were investigated in my PhD thesis: Cerrado biogeography and species delimitation. Evidently, as discussed above, there are still many open questions regarding the evolution of the Cerrado biota that deserve further attention. Some of these questions will be investigated using data from my PhD as part of a large ongoing research program on Cerrado herpetofauna, under the projects 'Herpetofauna do Cerrado: Origens, Evolução e Conservação' (FAP-DF/ CNPq grant 193.000.292), 'Filogeografia Comparada da Herpetofauna do Cerrado (CNPq grant 479026), and 'Conservação e análise dos padrões e processos associados à diversidade genética da herpetofauna do Cerrado (GENPAC 15)' (CNPq/ FAP-DF/ Capes grant 031/2010) coordinated by my co-supervisor, Guarino Colli. Nonetheless, they should also constitute key research directions for other evolutionary investigations performed on the Cerrado biota.

4.1 Cerrado biogeography

Certainly, the most important aspect on investigating Cerrado biogeography will be linking patterns of genetic diversity to geological and climatic processes (He *et al.*, 2013). This activity should be focused not only on understanding the effects of landscape evolution and Quaternary climatic fluctuations on evolutionary diversification in the biome, but also on intrinsic demographic patterns related to the evolution of each taxon. For this enterprise, phylogeographic studies will benefit from explicitly comparing many demographic models (Pelletier & Carstens, 2014), and modelling different demographic scenarios using

phylogenomics datasets (Robinson *et al.*, 2014). Additionally, approaches in landscape genomics (Cushman & Landguth, 2010), preferably implemented using comparative evolutionary simulations (Beheregaray *et al.*, 2015), may shed light on fine-scale processes that influence genetic divergence.

Furthermore, a closer dialogue between evolutionary biologists and geologists is urgently warranted. As exposed in Chapter 4, the lack of information about Cerrado Neotectonics hinders more in-depth investigations regarding the influences of the uplift of the Central Brazilian Plateau on the evolution of endemic biota. It is not uncommon for valuable geological data to be found in the so-called grey literature (Bichteler, 1991), but evolutionary research would, nonetheless, benefit from more detailed geological surveys of the Cerrado landscape. Specifically, precise information on the chronology of different compartmentalisation events through the uplift process, and on the possible relationship among different plateaus and valleys would be crucial to develop detailed diversification hypotheses. These are not unreal expectations since similar detailed geological evolution research has been done, for example, for extensive parts of the Andes (Baby *et al.*, 1997; Lara *et al.*, 2004).

Finally, future research should also seek to identify the causes for genetic structure and their relationships with cryptic speciation (Wang & Summers, 2010), and possible morphological differentiation (Robertson & Vega, 2011). For example, whether or not patterns of morphological and/or karyological variation observed within our three taxa (Chapter 3) correspond to proposed species boundaries, and what is the relative role of geography and ecology in the generation of this diversity in our system is yet to be properly investigated.

4.2 Species delimitation methods and cryptic Cerrado lizard species

One of the shortfalls of species delimitation methods is that most cannot incorporate morphological data in the analyses process, which are essential for future taxonomic identification. A few approaches have been proposed before (Guillot *et al.*, 2012), but the recently described iBPP appears to be the best available option for integrated molecular-morphological species delimitation (Solis-Lemus *et al.*, 2015). Although more empirical tests are necessary to assess the performance of iBPP, one important factor is the selection of morphological characters: having more characters contributes to the power of the analyses, but might also add noise to the results. Building on the analyses developed in Chapter 2, which used support-vector machines to investigate morphological discrimination, we are already developing a more complex framework. Additional developments include training the data using several machine learning approaches (e.g., SVM, Random Forest, Neural Networks, etc), ranking the training performance of these different algorithms using Akaike Information Criteria and, finally, selecting the morphological variables that had greater influence in the discrimination model. One possible application of this framework is running iBPP with only the informative morphological data selected by our machine learning algorithms, hence avoiding the use of 'noisy' morphological characters.

Although species delimitation methods are increasingly contributing to the discovery of cryptic biodiversity across the globe (Fujita *et al.*, 2012), many species delimitation studies unfortunately fail in effectively describing the species. This takes place either due to the lack of clear morphological distinction (which can potentially be surpassed by the methods exposed above), or by the lack of other informative biological properties that can account for speciation in the study system (de Queiroz, 2007). Having in mind that species description is one of the main goals behind delimiting species, I believe that incorporating ecological and demographic information on species delimitation studies in very species-rich biomes, such as the Cerrado, might be the next step to obtain more holistic views about speciation patterns in

the Neotropics. One unexplored option is to estimate the niche overlap between the delimited species (Schulte *et al.*, 2012) using the software ENMTools (Warren *et al.*, 2010). Another possibility is to evaluate how climatic and other ecological variables (e.g., body temperature, use of substrate, etc) are correlated with the genetic variation among cryptic species using multivariate analyses. Genomic data can shed light on how natural selection might be acting on the different cryptic species (Vincent *et al.*, 2013), and wether those different lineages are subject to distinct selective pressures.

4.3 Near-future expected publication outcomes

Considering the two main topics above, we expect another seven publications to arise from the data presented here: (1) an updated publication on species delimitation and species descriptions of the *Gymnodactylus amarali* complex, (2) species delimitation in the *Micrablepharus atticolus* complex, (3) species delimitation in the *Tropidurus itambere* complex, (4) a phylogeography of the *G, amarali* complex, (5) a phylogeography of the *M. atticolus* complex, (6) a phylogeography of the *T. itambere* complex, (7) a comparative phylogeography of the three species complexes (Chapter 4). At this stage, *G. amarali* is the only complex that appears appropriate for species to be described within a species delimitation manuscript. This is due to a number of reasons: the familiarity that I acquired with the species complex when collecting morphological data (i.e., the complete morphological dataset includes 1,200 individuals from 47 sampling localities), and because we have collaborators already working on a taxonomic review of the genus (José Cassimiro and Miguel Rodrigues). For the other two species complexes, taxonomic descriptions will appear in future publications in herpetological/taxonomic oriented journals.

Appendices

Appendix 1: Title page of Chapter 2, as published in the journal *Molecular Phylogenetics and Evolution.*

Full citation: Domingos, F.M.C.B., Bosque, R.J., Cassimiro, J., Colli, G.R., Rodrigues, M.T., Santos, M.G. & Beheregaray, L.B. (2014) Out of the deep: Cryptic speciation in a Neotropical gecko (Squamata, Phyllodactylidae) revealed by species delimitation methods. *Molecular Phylogenetics and Evolution*, **80**, 113–124.

DOI: 10.1016/j.ympev.2014.07.022

Molecular Phylogenetics and Evolution 80 (2014) 113–124

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Out of the deep: Cryptic speciation in a Neotropical gecko (Squamata, Phyllodactylidae) revealed by species delimitation methods

Fabricius M.C.B. Domingos [a,*], Renan J. Bosque [b], José Cassimiro [c], Guarino R. Colli [d], Miguel T. Rodrigues [c], Marcella G. Santos [b], Luciano B. Beheregaray [a]

[a] *Molecular Ecology Laboratory, School of Biological Sciences, Flinders University, Adelaide, SA 5001, Australia*
[b] *Department of Biology, University of Mississippi, Box 1848, University, MS 38677-1848, USA*
[c] *Departamento de Zoologia, Instituto de Biociências, Universidade de São Paulo, Cidade Universitária, 05508-090 São Paulo, SP, Brazil*
[d] *Departamento de Zoologia, Universidade de Brasília, 70910-900 Brasília, DF, Brazil*

ARTICLE INFO

Article history:
Received 14 March 2014
Revised 21 July 2014
Accepted 29 July 2014
Available online 8 August 2014

Keywords:
Bayesian species delimitation
Coalescent analyses
Cerrado
Gymnodactylus
Morphology
Phylogeography

ABSTRACT

Levels of biodiversity in the Neotropics are largely underestimated despite centuries of research interest in this region. This is particularly true for the Cerrado, the largest Neotropical savanna and a formally recognized biodiversity hotspot. Molecular species delimitation methods have become essential tools to uncover cryptic species and can be notably robust when coupled with morphological information. We present the first evaluation of the monophyly and cryptic speciation of a widespread Cerrado endemic lizard, *Gymnodactylus amarali*, using phylogenetic and species-trees methods, as well as a coalescent-based Bayesian species delimitation method. We tested whether lineages resulting from the analyses of molecular data are morphologically diagnosed by traditional meristic scale characters. We recovered eight deeply divergent molecular clades within *G. amarali*, and two additional ones from seasonally dry tropical forest enclaves between the Cerrado and the Caatinga biomes. Analysis of morphological data statistically corroborated the molecular delimitation for all groups, in a pioneering example of the use of support vector machines to investigate morphological differences in animals. The eight *G. amarali* clades appear monophyletic and endemic to the Cerrado. They display several different properties used by biologists to delineate species and are therefore considered here as candidates for formal taxonomic description. We also present a preliminary account of the biogeographic history of these lineages in the Cerrado, evidence for speciation of sister lineages in the Cerrado–Caatinga contact, and highlight the need for further morphological and genetic studies to assess cryptic diversity in this biodiversity hotspot.

Appendix 2: Sampled lizard specimens used in Chapter 2. Cytb haplotypes refer to duplicates used in the GMYC analyses (Fig. 3). Asterisk indicates specimens with no morphology available. Brazilian states: TO = Tocantins, MT = Mato Grosso, MA= Maranhão, GO = Goiás, MG = Minas Gerais, CE = Ceará, PB = Paraíba. CHUNB: Coleção Herpetológica da Universidade de Brasília; ESTR: collection code Estreito; GRC:collection code Guarino Rinaldi Colli; LAJ: collection code Lajeado; LG: experiment code "lagartos do Laboratório de Citogenética de Vertebrados do Departamento de Genética do Instituto de Biociências da USP"; MD: colection code Mariana Dixo; MTR or MRT: collection code Miguel Trefaut Rodrigues; MZUSP: Museu de Zoologia da Universidade de São Paulo; OMNH: Oklahoma Museum of Natural History. *Phyllopezus pollicaris* sequences originally from Werneck and colleagues (2012), and *Gymnodactylus darwinii* sequences from Pellegrino and colleagues (2005).

Cytb Haplotype	Species	ID	Locality	Brazilian State	Genbank accession number (cytb/ KIF24)
1	*G. amarali Clade H*	MTR14578*	Almas	TO	KM283297/ KM283408
1	*G. amarali Clade H*	MTR14609*	Almas	TO	KM283298/ KM283415
1	*G. amarali Clade H*	MTR14808*	Almas	TO	KM283299/ KM283416
2	*G. amarali Clade E*	CHUNB55881	Barra do Garças	MT	KM283243/ KM283367
3	*G. amarali Clade E*	CHUNB55883	Barra do Garças	MT	KM283244/ KM283368
4	*G. amarali Clade E*	CHUNB63196	Barra do Garças	MT	KM283245/ KM283383
5	*G. amarali Clade E*	CHUNB63198	Barra do Garças	MT	KM283246/ KM283384
6	*G. amarali Clade B*	LG0889	Barra do Garças	TO	KM283291/ KM283407
7	*G. amarali Clade H*	ESTR1038	Carolina	MA	KM283247/ KM283393
8	*G. amarali Clade H*	ESTR00196*	Carolina	MA	KM283248/ KM283397
8	*G. amarali Clade H*	ESTR00197*	Carolina	MA	KM283249/ KM283398
9	*G. amarali Clade H*	CHUNB45283*	Caseara	TO	KM283250/ KM283355
9	*G. amarali Clade H*	CHUNB45318	Caseara	TO	KM283251/ KM283356
10	*G. amarali Clade H*	CHUNB45321	Caseara	TO	KM283252/ KM283357
9	*G. amarali Clade H*	CHUNB45336	Caseara	TO	KM283253/ KM283358
11	*G. amarali Clade G*	CHUNB53290	Cavalcante	GO	KM283255/ KM283363
11	*G. amarali Clade G*	CHUNB53292	Cavalcante	GO	KM283256/ KM283364
11	*G. amarali Clade G*	CHUNB53293	Cavalcante	GO	KM283257/ KM283365
12	*G. amarali Clade C*	CHUNB38645	Cocalzinho de Goiás	GO	KM283305/ KM283352
13	*G. amarali Clade C*	CHUNB38646	Cocalzinho de Goiás	GO	KM283306/ KM283353
14	*G. amarali Clade D*	CHUNB44704	Colinas do Sul	GO	KM283280/ KM283354
15	*G. amarali Clade H*	ESTR1293	Estreito	MA	KM283258/ KM283394
15	*G. amarali Clade H*	ESTR0642	Estreito	MA	KM283261/ KM283399
16	*G. amarali Clade H*	ESTR1759	Goiatins	TO	KM283259/ KM283395
17	*G. amarali Clade H*	ESTR1808	Goiatins	TO	KM283260/ KM283396
18	*G. amarali Clade H*	MRT7552*	Guarai	TO	KM283262/ KM283413
18	*G. amarali Clade H*	MRT7598*	Guarai	TO	KM283263/ KM283414
19	*G. amarali Clade H*	LAJ215	Lajeado	TO	KM283296/ KM283404
20	Manga	CHUNB58336	Manga	MG	KM283267/ KM283377
20	Manga	CHUNB58337	Manga	MG	KM283268/ KM283378
20	Manga	CHUNB58338	Manga	MG	KM283269/ KM283379
20	Manga	CHUNB58341	Manga	MG	KM283270/ KM283380

Cytb Haplotype	Species	ID	Locality	Brazilian State	Genbank accession number (cytb/ KIF24)
21	*G. amarali Clade H*	CHUNB28219	Mateiros	TO	KM283271/ KM283345
21	*G. amarali Clade H*	CHUNB28220	Mateiros	TO	KM283272/ KM283346
21	*G. amarali Clade H*	CHUNB28222	Mateiros	TO	KM283273/ KM283347
22	Matias Cardoso	CHUNB58323	Matias Cardoso	MG	KM283274/ KM283375
22	Matias Cardoso	CHUNB58335	Matias Cardoso	MG	KM283275/ KM283376
22	Matias Cardoso	CHUNB58348	Matias Cardoso	MG	KM283276/ KM283381
23	*G. amarali Clade D*	GRC2152	Minaçu	GO	KM283277/ KM283342
24	*G. amarali Clade D*	GRC2204	Minaçu	GO	KM283278/ KM283343
25	*G. amarali Clade D*	GRC2211	Minaçu	GO	KM283279/ KM283344
26	*G. amarali Clade D*	CHUNB832	Minaçu	GO	KM283281/ KM283392
27	*G. amarali Clade G*	CHUNB53054	Monte Alegre de Goiás	GO	KM283319/ KM283359
28	*G. amarali Clade G*	CHUNB53058	Monte Alegre de Goiás	GO	KM283320/ KM283360
28	*G. amarali Clade G*	CHUNB53059	Monte Alegre de Goiás	GO	KM283321/ KM283361
28	*G. amarali Clade G*	CHUNB53061	Monte Alegre de Goiás	GO	KM283322/ KM283362
29	*G. amarali Clade D*	LG1072	Niquelândia	TO	KM283282/ KM283405
29	*G. amarali Clade D*	LG1075	Niquelândia	GO	KM283283/ KM283406
30	*G. amarali Clade B*	CHUNB55880	Nova Xavantina	MT	KM283284/ KM283366
30	*G. amarali Clade B*	CHUNB55884	Nova Xavantina	MT	KM283285/ KM283369
30	*G. amarali Clade B*	CHUNB55885	Nova Xavantina	MT	KM283286/ KM283370
30	*G. amarali Clade B*	CHUNB55887	Nova Xavantina	MT	KM283287/ KM283371
30	*G. amarali Clade B*	CHUNB55888	Nova Xavantina	MT	KM283288/ KM283372
31	*G. amarali Clade B*	CHUNB55889	Nova Xavantina	MT	KM283289/ KM283373
32	*G. amarali Clade B*	CHUNB55890	Nova Xavantina	MT	KM283290/ KM283374
33	*G. amarali Clade B*	CHUNB63195	Nova Xavantina	MT	KM283327/ KM283382
34	*G. amarali Clade B*	CHUNB63203	Nova Xavantina	MT	KM283328/ KM283385
35	*G. amarali Clade B*	CHUNB63204	Nova Xavantina	MT	KM283329/ KM283386
36	*G. amarali Clade H*	GRC10440	Palmas	TO	KM283294/ KM283330
36	*G. amarali Clade H*	GRC10442	Palmas	TO	KM283295/ KM283331
37	*G. amarali Clade F*	GRC10771	Palmas	TO	KM283292/ KM283332
38	*G. amarali Clade F*	CHUNB14555	Palmas	TO	KM283293/ KM283333
39	*G. amarali Clade G*	GRC15781	Paranã	TO	KM283300/ KM283340
40	*G. amarali Clade G*	GRC15869	Paranã	TO	KM283301/ KM283341
41	*G. amarali Clade G*	CHUNB37090	Paranã	TO	KM283302/ KM283348
42	*G. amarali Clade G*	CHUNB37125	Paranã	TO	KM283303/ KM283349
41	*G. amarali Clade G*	CHUNB37128	Paranã	TO	KM283304/ KM283350
43	*G. amarali Clade H*	MRT3949	Peixe	TO	KM283323/ KM283409
44	*G. amarali Clade H*	MRT4459	Peixe	TO	KM283254/ KM283410
45	*G. amarali Clade C*	CHUNB67394	Pirenópolis	GO	KM283307/ KM283387
45	*G. amarali Clade C*	CHUNB67395	Pirenópolis	GO	KM283308/ KM283388
45	*G. amarali Clade C*	CHUNB67396	Pirenópolis	GO	KM283309/ KM283389
46	*G. amarali Clade C*	CHUNB67441	Pirenópolis	GO	KM283310/ KM283390
45	*G. amarali Clade C*	CHUNB67443*	Pirenópolis	GO	KM283311/ KM283391
47	*G. amarali Clade A*	GRC15533	São Domingos	GO	KM283312/ KM283334
47	*G. amarali Clade A*	GRC15558	São Domingos	GO	KM283313/ KM283335
47	*G. amarali Clade A*	GRC15560	São Domingos	GO	KM283314/ KM283336
27	*G. amarali Clade G*	GRC15566	São Domingos	GO	KM283317/ KM283337

Cytb Haplotype	Species	ID	Locality	Brazilian State	Genbank accession number (cytb/ KIF24)
47	*G. amarali Clade A*	GRC15568	São Domingos	GO	KM283315/ KM283338
47	*G. amarali Clade A*	GRC15569	São Domingos	GO	KM283316/ KM283339
48	*G. amarali Clade G*	GRC37161	São Domingos	GO	KM283318/ KM283351
49	*G. amarali Clade H*	MRT6428	São Salvador	TO	KM283324/ KM283411
50	*G. amarali Clade H*	MRT6435	São Salvador	TO	KM283325/ KM283412
51	*G. darwinii A*	LG958*	Porto Seguro	BA	AY630388.1/ NA
51	*G. darwinii A*	LG991*	Porto Seguro	BA	AY630392.1/ NA
NA	*G. darwinii*	CHUNB09443	Presidente Kennedy	ES	NA/ NA
NA	*G. darwinii*	CHUNB09453	Presidente Kennedy	ES	NA/ NA
NA	*G. darwinii*	CHUNB13546	Presidente Kennedy	ES	NA/ NA
52	*G. darwinii B*	MD1740*	Una	BA	AY630367.1/ NA
53	*G. darwinii B*	MD1742*	Una	BA	AY630368.1/ NA
54	*G. geckoides*	CHUNB56643	Mamanguape	PB	KM283264/ KM283400
54	*G. geckoides*	CHUNB56644	Mamanguape	PB	KM283265/ KM283401
54	*G. geckoides*	CHUNB56645	Mamanguape	PB	KM283266/ KM283402
55	*G. geckoides*	CHUNB56537	Milagres	CE	KM283326/ KM283403
56	*P. pollicaris*	CHUNB57388	Tianguá	CE	JQ827177.1/ JQ827663.1
57	*P. pollicaris*	CHUNB57389	Tianguá	CE	JQ827190.1/ JQ827664.1
58	*P. pollicaris*	CHUNB57403	Tianguá	CE	JQ827186.1/ JQ827667.1

Appendix 3: Details of primers and PCR protocols. NPCL = nuclear protein coding locus.

Marker	Primer	Primer sequence (5'-3')	Source	PCR profile	PCR reaction
Cytochrome b (cytb). mtDNA	CB3 WWF	GGCAAATAGGAARTATCATTC AAAYCAYCGTTGTWATTCAACTAC	(Palumbi, 1996) (Broadley et al., 2006)	94°C - 0:45, 55°C - 0:45, 72°C - 1:00 (30x) Decreasing 1°C until 49°C in annealing step	10mL reaction: 1.2mM of each primer, 3 mM MgCl2, 0.4 mM each dNTP, 2x Buffer (for Mango Taq - Bioline) and 1 U Taq polymerase (Mango - Bioline)
Kinesin Family Member 24 (KIF24). NPCL	KIF24 F1 KIF24 R1	SAAACGTRTCTCCMAAACGCATCC WGGCTGCTGRAAYTGCTGGTG	(Portik et al., 2010)	95°C - 0:35, 65.1°C - 0:35, 72°C - 1:35 (35x) Increasing 0:04 p/ cycle in elongation step	12.5mL reaction: 1mM of each primer, 4 mM MgCl2, 0.4 mM each dNTP, 2x Buffer (for Mango Taq - Bioline) and 1 U Taq polymerase (Mango - Bioline)

Appendix 4: Evolution models and partitioning strategy selected by PartitionFinder (Lanfear *et al.*, 2012; Lanfear *et al.*, 2014). As suggested by PartitionFinder's user manual, selection was done using the available models for each software used (MrBayes, RAxML or Beast). Cyt1, cyt2 and cyt3 are, respectively, the three codons of cytochrome b (cytb), and kif1, kif2 and kif3 the codons of Kinesin Family Member 24 (KIF24).

MrBayes concatenated		**MrBayes cytb**		**MrBayes KIF24**	
Subset Partitions	Best Model	Subset Partitions	Best Model	Subset Partitions	Best Model
cyt1, kif1, kif2	K80+I+G	cyt1	HKY+I+G	kif1, kif2	HKY+G
cyt2	K80+G	cyt2	K80+G	kif3	HKY+G
cyt3	GTR+G	cyt3	GTR+G		
kif3	HKY+I+G				

RAxML concatenated		**Beast cytb**		**Beast concatenated (*Beast)**	
Subset Partitions	Best Model	Subset Partitions	Best Model	Subset Partitions	Best Model
cyt1, cyt2	GTR+I+G	cyt1	HKY+I+G	cyt1	HKY+I+G
cyt3	GTR+G	cyt2	K80+G	cyt2	K80+G
kif1, kif2, kif3	GTR+I+G	cyt3	TrN+G	cyt3	TrN+G
				kif1, kif3	TrN+I+G
				kif2	K80+G

Appendix 5: BPP trials.

We tried different numbers of gamma priors on the population size parameters (θs), and in the age of the root in the species tree (τ_0) in the BPP runs, to represent different speciation histories: 1) large population size and deep divergence – we ran G(2, 2000) and also G(2, 1000) for both priors; 2) small population size and shallow divergence – G(1,10) for both priors; and 3) large population size and shallow divergence – G(2,2000) for θ prior, G(1,10) for τ prior. All different trials returned extremely similar results.

Appendix 6: Morphological characters of *Gymnodactylus*.

For every specimen used in the morphological analyses the following meristic (1 - 21) and qualitative (22 -29) variables were recorded:

1. Number of scales in canthus rostralis, counted from post nasal to the eye.
2. Number of scales above and in contact with the supralabials, counted from frontonasal to last supralabial.
3. Number of scales below and in contact with the infralabials, counted from mental to last infralabial.
4. Number of supralabials (sum of both sides).
5. Number of infralabials (sum of both sides).
6. Number of enlarged supraciliary scales.
7. Number of dorsal scales, counted from rostral scale to posterior margin of thigh (before tail).
8. Number of keeled scale rows in tail.
9. Number of keeled scales in one row in tail, counted in the third keeled scales row.
10. Number of paramedian tubercles, counted from tympanum to posterior margin of thigh (before tail).
11. Number of paramedian ocelli, counted in one row from rostral to posterior margin of thigh (before tail).
12. Number of longitudinal ocelli at midbody.
13. Number of longitudinal tubercles rows at midbody.
14. Number of longitudinal rows of ventral scales at midbody.
15. Number of scales between enlarged post mentals, in contact with mental.
16. Number of subdigital lamellae on fourth finger.
17. Number of transverse rows of ventral scales, counted from mental to cloaca.
18. Number of femoral and tibial ventral scale rows, counted from cloaca (start of thigh) to foot at mid part of the limb.
19. Number of subdigital lamellae on fourth toe.
20. Number of granule like scales from cloaca to first enlarged subcaudal.
21. Number of white bands in tail.
22. Relative size of post nasals in relation to supranasal – (0): both post nasals smaller than supranasal; (1): second post nasal as large as supranasal.
23. Contact between supranasals – (0): in full contact; (1): in partial contact, with distal indentation; (2): no contact, with scales in the space between them.
24. Alignment between frontonasals division and the incomplete suture of rostral – (0): aligned; (1) not aligned.
25. Ear opening shape – (0) circular; (1) sagitally elliptic; (2) dorsally elliptic.
26. Ear opening position– (0) aligned with supralabials; (1) aligned with eye.
27. Dorsal ocelli – (0): present; (1) absent.
28. Ocelli in limbs – (0): present; (1) band pattern (non-round ocelli); (2) absent.
29. Bands in tail – (0): present; (1) absent.

Appendix 7: Support Vector Machine (SVM) analysis.

When employing multivariate analyses to statistically classify and predict cases belonging to different groups, the usual choices are linear Discriminant Function Analysis (DFA) and its variants (Quinn & Keough, 2002). However, despite the great statistical power of such approaches, there are several assumptions that must be satisfied; namely normality, no collinearity, and homoscedasticity (Tabachnick & Fidell, 1996). The classification part of a DFA is quite sensitive to heterogeneous variance–covariance matrices between groups, and the aforementioned assumptions are violated when the number of cases is smaller than the number of variables (Quinn & Keough, 2002). In our case, we had a maximum number of two individuals for some clades (and 29 variables). Also, meristic characters are hardly normally distributed, and transforming the data to normalise them can be virtually impossible when group sample sizes are small.

The use of a Support Vector Machine (SVM) overcomes the limitations of linear multivariate approaches, especially because the above-mentioned assumptions are not required to create a SVM model (Cortes & Vapnik, 1995). The SVM builds a kernel function that maps the cases into a high-dimensional space, subsequently finding a "margin" in the hyperspace that maximizes the separation between the groups (Cortes & Vapnik, 1995; Schölkopf *et al.*, 2000). Instead of simple points in a statistical space, the data points represent the objects (cases) using a set of features derived from measurements performed in each object, and the relative position between these objects is more important for the model than the exact position of the objects (Ben-Hur *et al.*, 2008). Being so, SVM models are particularly appropriate in cases where the sample size of the groups are small, and can also be applied to datasets that do not conform to assumptions of traditional classification methods (Schölkopf *et al.*, 2000).

We performed a Radial SVM analysis using R package *e1071* (Meyer *et al.*, 2014). We implemented a manual search for the best fine tune parameters for the model (cost and

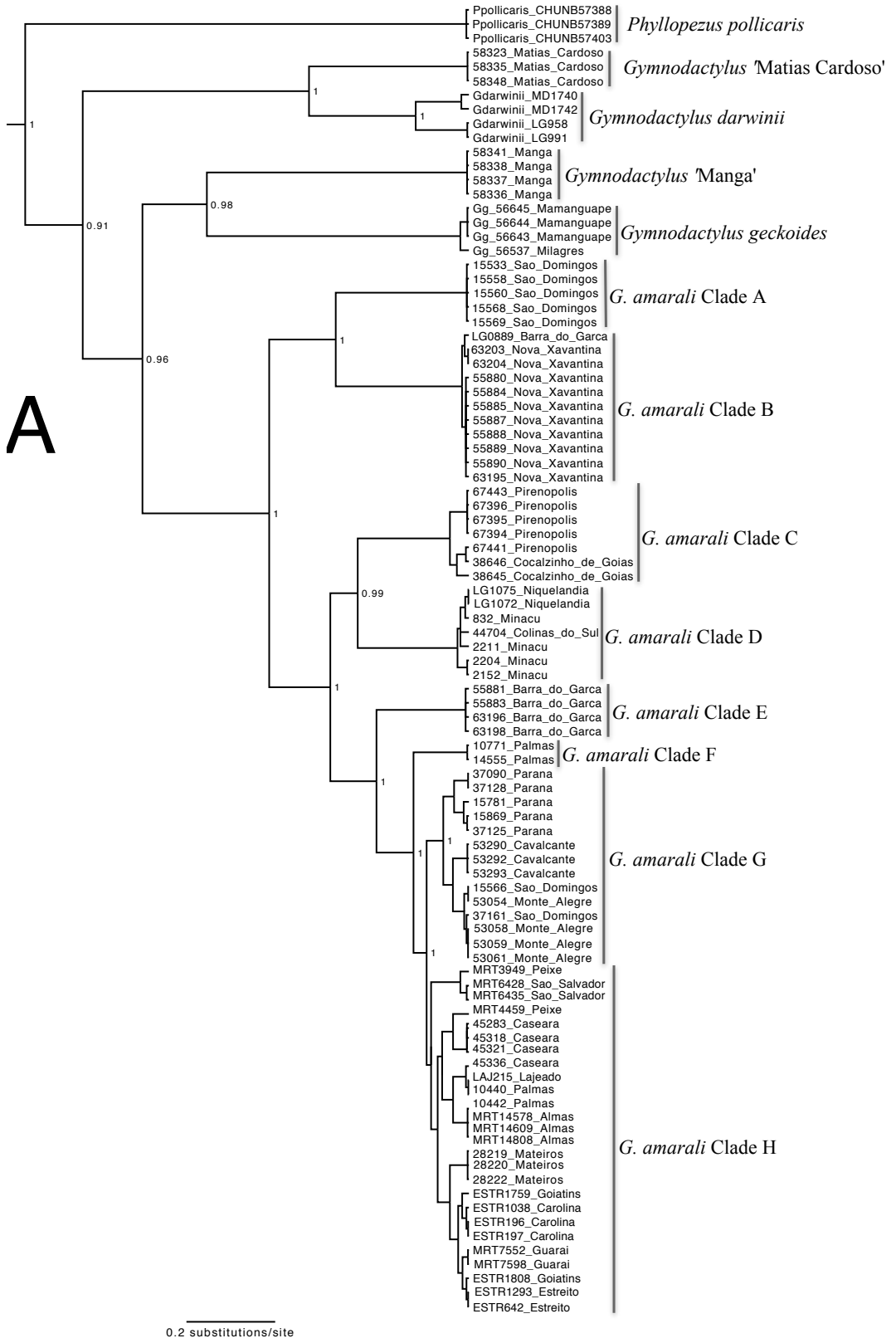
gamma) estimated via cross-validation as suggested by Chang and Lin (2011). This manual search was done using the function *tune.svm* of the package *e1071*. We then trained the model using the fine tuned *C*-classification SVM algorithm on the whole morphological dataset, setting 'species' to be explained by the 29 morphological characters. We tested the predictive power of the generated model using the *predict.svm* function of the package, which predicts case affiliation to groups (individual to 'species') based in the model trained by the SVM. Error rate of the model was calculated as the percentage of individuals that were incorrectly assigned to the species (GMYC group) it belongs. The full dataset used for this analysis is available upon request to the first author.

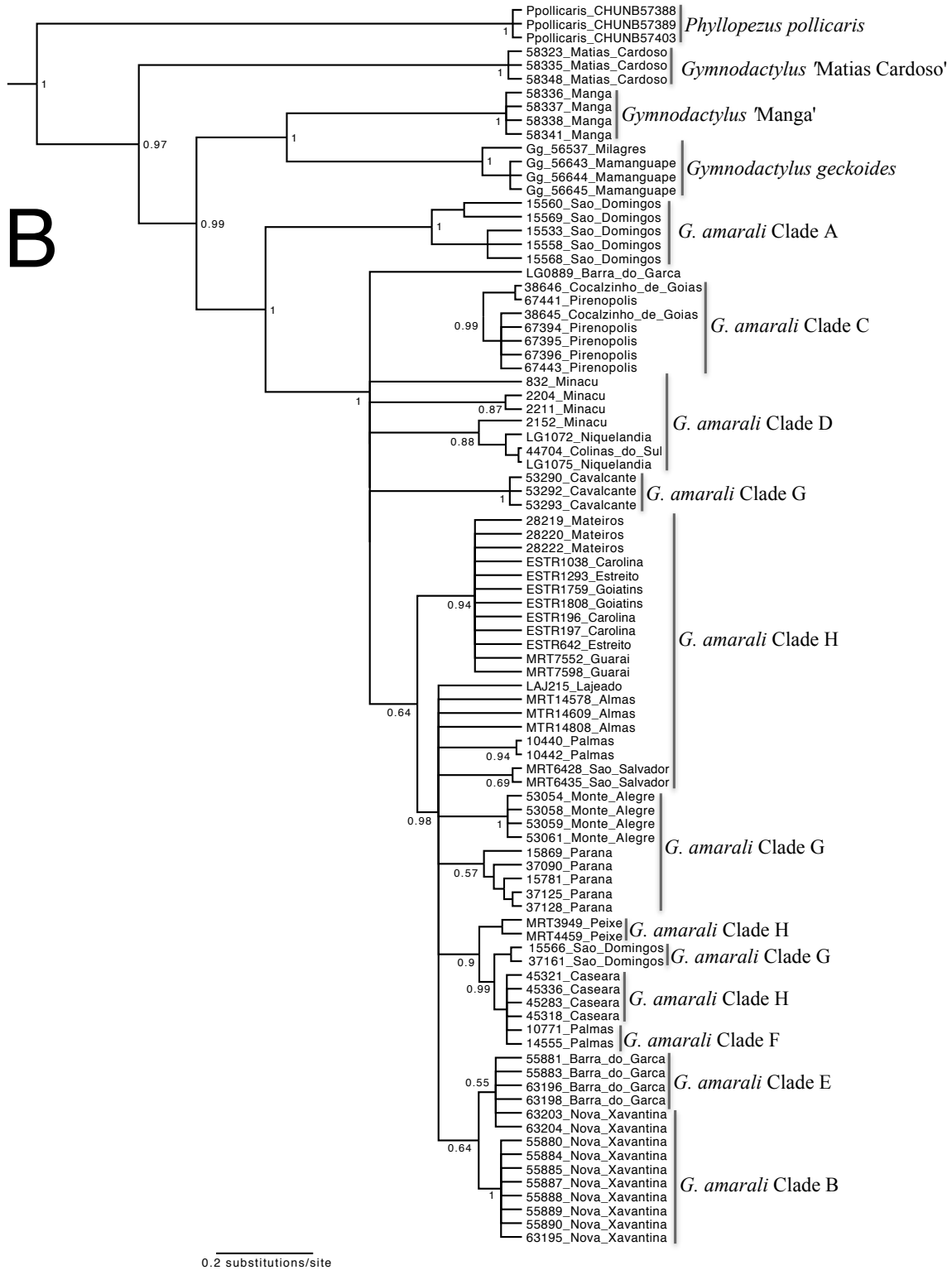
For comparison purposes we tested the performance of a DFA analysis with the performance of the employed SVM using R package *caret* (Kuhn, 2008). Even thought the results are extremely similar, it is important to recall that the DFA assumptions are violated and there would be little confirmatory value in the results if used in practice (Byron, 1983), so the following methods should be considered with caution. We investigated the performance of each model (DFA and SVM) using a 10 fold resampling cross-validation to calculate the 'Accuracy' (overall agreement rate averaged over cross-validation iterations) and 'Kappa' (Cohen's Kappa statistic averaged across the resampling results) using the *train* function in *caret*. This 10 fold resampling means the data was randomly divided into training and testing datasets in 10 different ways, and the statistics calculated based on the average results of all 10 iterations. The best SVM model has Accuracy= 0.686 and Kappa= 0.617, while the DFA model has Accuracy= 0.683 and Kappa= 0.628. If a penalised discriminant analysis model is used (that controls for collinearity but not for the other assumptions' violations), then Accuracy= 0.631 and Kappa= 0.573. The DFA and its penalised version had, respectively, classification errors of 2% and 7%, against our reported 3%. All these results suggest that our SVM model has similar predictive power when compared to traditional DFA analysis,

therefore supporting the use of a SVM when the available data violates the assumptions of a DFA.

Lastly, one could have concerns about the supposedly circular nature of the SVM model, considering that the same data is used to generate the model and then to test it. However, it is important to notice that using the data to build the model and, in our case, predict species affiliation, is philosophically identical to how a normal DFA works, where the classification function is calculated and then used to classify the same observations. In this sense, every classification method is circular unless a dataset is used to generate the model (or function) and a different dataset is used during the classification (Quinn & Keough, 2002). Here, when using one third of the data to train the model and then using the created model to classify the other two thirds, the classification error increases to 9%. This is an expected pattern is any classification statistical method (Quinn & Keough, 2002), and classification errors may substantially vary depending on the study scenario. As SVM was never used for classification using morphological characters before, misidentification errors cannot be directly compared with other published studies. Nonetheless, a non-exhaustive list of examples where error rates were reported include: 1) between 6.7% and 11% for microRNA precursors classification (Xue *et al.*, 2005); 2) between 20.9% and 46.6% for prediction of protein subcellular location (Park & Kanehisa, 2003); 3) respectively 3% and 5% when evaluating the distribution model of a Neotropical frog under two different calibration areas (Giovanelli *et al.*, 2010); and 4) 1.2% when modeling the distribution of a forest disease in North America (Kelly *et al.*, 2007). Note that the aforementioned errors might have been calculated using specific accuracy estimates (and not simple misclassification rates) and one should refer to the original publications for more details.

Appendix 8: Single locus Bayesian phylogenetic trees of cytb (A) and KIF24 (B) for *G. amarali* samples used in Chapter 2. Numbers in nodes are Bayesian posterior probabilities. Clades A to H refer to *Gymnodactylus amarali* clades identified by GMYC analysis.




B
Ppollicaris_CHUNB57388
Ppollicaris_CHUNB57389
Ppollicaris_CHUNB57403
Phyllopezus pollicaris
58323_Matias_Cardoso
58335_Matias_Cardoso
58348_Matias_Cardoso
Gymnodactylus 'Matias Cardoso'
58336_Manga
58337_Manga
58338_Manga
58341_Manga
Gymnodactylus 'Manga'
Gg_56537_Milagres
Gg_56643_Mamanguape
Gg_56644_Mamanguape
Gg_56645_Mamanguape
Gymnodactylus geckoides
15560_Sao_Domingos
15569_Sao_Domingos
15533_Sao_Domingos
15558_Sao_Domingos
15568_Sao_Domingos
G. amarali Clade A
LG0889_Barra_do_Garca
38646_Cocalzinho_de_Goias
67441_Pirenopolis
38645_Cocalzinho_de_Goias
67394_Pirenopolis
67395_Pirenopolis
67396_Pirenopolis
67443_Pirenopolis
G. amarali Clade C
832_Minacu
2204_Minacu
2211_Minacu
2152_Minacu
LG1072_Niquelandia
44704_Colinas_do_Sul
LG1075_Niquelandia
G. amarali Clade D
53290_Cavalcante
53292_Cavalcante
53293_Cavalcante
G. amarali Clade G
28219_Mateiros
28220_Mateiros
28222_Mateiros
ESTR1038_Carolina
ESTR1293_Estreito
ESTR1759_Goiatins
ESTR1808_Goiatins
ESTR196_Carolina
ESTR197_Carolina
ESTR642_Estreito
MRT7552_Guarai
MRT7598_Guarai
LAJ215_Lajeado
MRT14578_Almas
MTR14609_Almas
MTR14808_Almas
10440_Palmas
10442_Palmas
MRT6428_Sao_Salvador
MRT6435_Sao_Salvador
G. amarali Clade H
53054_Monte_Alegre
53058_Monte_Alegre
53059_Monte_Alegre
53061_Monte_Alegre
15869_Parana
37090_Parana
15781_Parana
37125_Parana
37128_Parana
G. amarali Clade G
MRT3949_Peixe
MRT4459_Peixe
G. amarali Clade H
15566_Sao_Domingos
37161_Sao_Domingos
G. amarali Clade G
45321_Caseara
45336_Caseara
45283_Caseara
45318_Caseara
G. amarali Clade H
10771_Palmas
14555_Palmas
G. amarali Clade F
55881_Barra_do_Garca
55883_Barra_do_Garca
63196_Barra_do_Garca
63198_Barra_do_Garca
G. amarali Clade E
63203_Nova_Xavantina
63204_Nova_Xavantina
55880_Nova_Xavantina
55884_Nova_Xavantina
55885_Nova_Xavantina
55887_Nova_Xavantina
55888_Nova_Xavantina
55889_Nova_Xavantina
55890_Nova_Xavantina
63195_Nova_Xavantina
G. amarali Clade B
0.2 substitutions/site

Appendix 9: Placement of Matias Cardoso and Manga populations within *Gymnodactylus* species.

Despite being less than 15 km apart, the populations of 'Manga' and 'Matias Cardoso' are separated by the São Francisco River. The degrees of sequence divergence between those two populations (Table 1, Chapter 1), as well as topology of the phylogenetic analyses (Fig. 2-4, Appendix 8, Chapter 1), showed that they belong to different groups, as similarly found for *Phyllopezus pollicaris* populations collected in the same two locations (Werneck *et al.*, 2012a). Also, 'Matias Cardoso' was more closely related to *G. darwinii* populations, whereas 'Manga' was more related to *G. geckoides*. Furthermore, based in the fact that deletions in KIF24 in 'Matias Cardoso' specimens are different from the ones in 'Manga' and *G. geckoides* specimens, and no deletions are found in *G. amarali* specimens, we believe that is little doubt that those two populations are not part of the *G. amarali* species complex. Moreover, there are clear morphological differences between those two populations and populations of *G. amarali* (Results section 3.3, Chapter 1). Only considering the main morphological character traditionally used to separate the species *G. geckoides*, *G. darwinii* and *G. amarali*, "number of tubercle rows" (Vanzolini, 1953a, 1982, 2005; Cassimiro & Rodrigues, 2009), 'Matias Cardoso' has coincident counts with *G. darwini* (14) while 'Manga' is clearly similar to *G. geckoides* (12), both considering the published data (Vanzolini, 1953a, 1982, 2005; Cassimiro & Rodrigues, 2009) and the data presented here (Appendix 12). Therefore, in a biogeographic perspective, 'Manga', located at the western side of the São Francisco River (Fig. 1, Chapter 1), is related to *Gymnodactylus geckoides* of the Caatinga, and 'Matias Cardoso', in the eastern side of the river, is related to *G. darwini* from the Atlantic Rainforest. In order to evaluate the possibility that 'Matias Cardoso' and 'Manga' are also undescribed species, future studies would need to sample more populations of *G. darwinii* and *G. geckoides*. Nevertheless, cytb sequence divergence between 'Matias Cardoso' and *G. darwinii* is higher than those reported for any *G. darwinii* cryptic species by

Pellegrino et al. (2005) (Table 1, Chapter 1). In addition, the divergence between 'Manga" and *G. geckoides* is higher than the ones reported here among *G. amarali* cryptic species (Table1, Chapter 1).

Appendix 10: SpedeSTEM results from Chapter 2.

The most likely group retrieved by spedeSTEM shared only three of the eight clades identified by the GMYC analysis. The groups were not geographically structured, and the BPP species delimitation algorithm failed to retrieve any informative result in terms of better grouping the clades into tentative species (Appendix 11). SpedeSTEM performance substantially decreases when only a few loci are included (Ence & Carstens, 2011), and we believe our result was heavily influenced by the discordances between the two gene trees (Knowles & Carstens, 2007). A study that compared several species delimitation and validation methods by Satler and collaborators (2013) found that GMYC was more prone to oversplitting compared to spedeSTEM. We found the opposite trend in our data, and the number of loci is most likely the reason behind it. Being so, in terms of the geographical location of the species, and in the actual number of delimited species, we believe that GMYC results are not only more conservative (Carstens *et al.*, 2013), but also a better hypothesis on the evolution of *Gymnodactylus amarali*.

The following topology was the best model selected by spedeSTEM, with numbers grouping the localities described in Appendix 2:

(Phy,(MaC,((Gec,Man),(A,(1,((2,((3,(4,5)),(6,7))),(C,D)))))))

Phy: *Phyllopezus pollicaris*; Mac: Matias Cardoso; Gec: *Gymodactylus geckoides*; Man: Manga; A: GMYC *G. amarali* clade A; 1: Cavalcante; 2: Nova Xavantina, Barra do Garças, Peixe, Caseara, Palmas, Monte Alegre, São Salvador, Paranã, Almas; 3: Estreito; 4: Mateiros; 5: Goiatins; 6: Guaraí; 7: Carolina; C: GMYC *G. amarali* clade C; D: GMYC *G. amarali* clade D.

Appendix 11: Results of BPP trials using the spedeSTEM recovered tree as input. In the tree model "1" means a branch that was kept as the original input and "0" a branch that was collapsed (please refer to BPP documentation for more details).

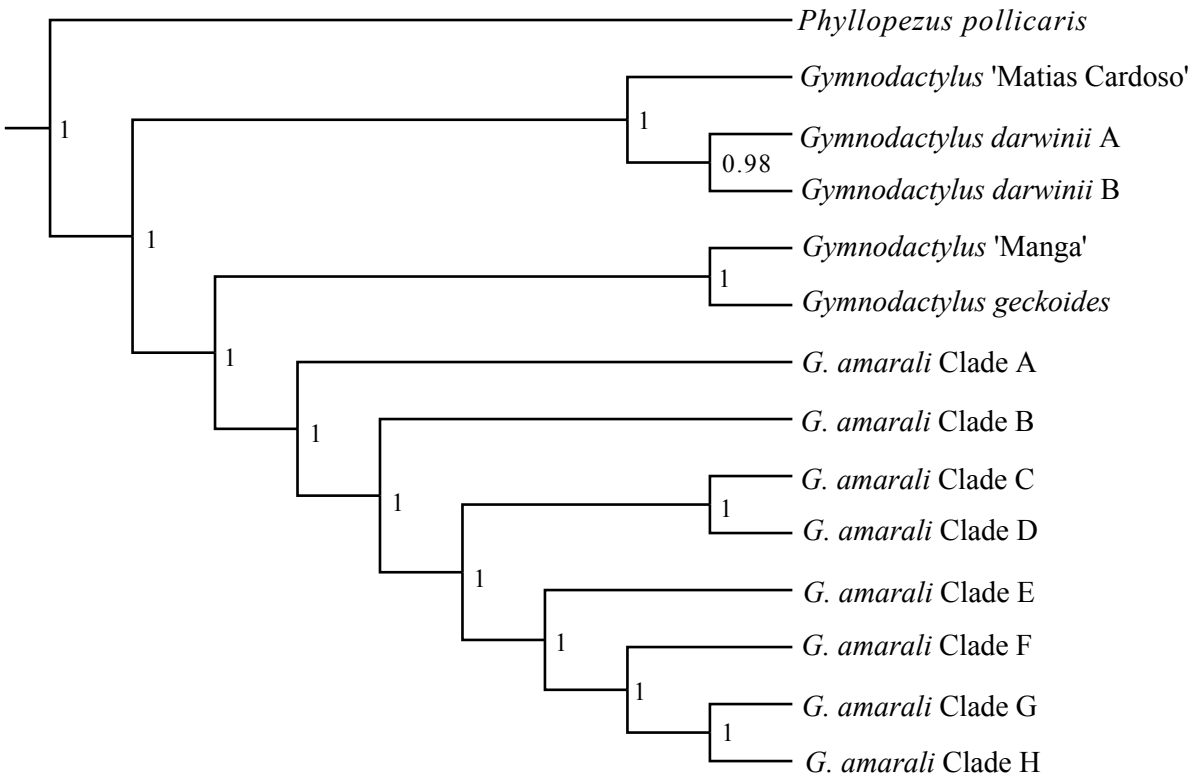
Tree Model	Posterior probability range
1111111100001	0.00779 - 0.02946
1111111110001	0.17529 - 0.31249
1111111110011	0.27461 - 0.28828
1111111111001	0.10485 - 0.12101
1111111111011	0.10031 - 0.13078
1111111111101	0.13626 - 0.08235
1111111111111	0.0661 - 0.16373

Appendix 12: Means (SD) of meristic (1 - 21) and mode of qualitative (22 -29) morphological characters comparing described *Gymnodactylus* species with *G. amarali* cryptic lineages, 'Manga' and 'Matias Cardoso'. Character numbers indicate morphological variable ID in Appendix 6.

Characters	*Gymnodactylus* lineages (n)								
	Clade A (5)	Clade B (11)	Clade C (6)	Clade D (7)	Clade E (4)	Clade F (2)	Clade G (14)	Clade H (18)	Manga (4)
1. Canthus rostralis	8 (0.7)	7.8 (0.4)	8 (0.6)	7.4 (1.1)	7.2 (0.5)	8 (0.0)	8.1 (0.8)	7.8 (0.7)	7 (0.0)
2. Above supralabials	15.8 (1.0)	14.9 (0.7)	16.5 (1.2)	15.7 (0.9)	15.7 (0.5)	16 (0.0)	15.5 (1.7)	14.4 (1.1)	14.7 (1.5)
3. Below infralabials	10.6 (1.1)	10.8 (1.4)	12 (1.0)	10 (0.8)	10 (0.8)	10 (0.0)	10 (0.7)	10.6 (1.1)	10.7 (0.5)
4. Supralabials	12.4 (1.1)	11.2 (0.7)	12.1 (0.4)	11.8 (0.6)	12 (0.0)	12 (0.0)	10.4 (0.9)	11.6 (1.0)	12.5 (1.0)
5. Infralabials	10.2 (0.4)	9.7 (0.6)	10 (0.0)	10 (0.0)	9 (0.8)	10 (0.0)	9.4 (0.8)	9.5 (0.7)	10.2 (0.5)
6. Supraciliary	14.2 (1.3)	13.9 (0.9)	14 (0.8)	13.2 (1.4)	13.5 (1.9)	15 (0.0)	15.5 (0.7)	14.3 (1.4)	13.5 (1.0)
7. Dorsal	240.2 (13.9)	208.6 (21.6)	248.6 (17.2)	229.4 (37.1)	220.5 (41.1)	224 (12.7)	242.7 (31.9)	211.2 (40.8)	177.5 (14.0)
8. Keeled rows in tail	4 (1.2)	6.3 (1.2)	4.5 (1.5)	5.4 (2.2)	5.7 (0.9)	7 (1.4)	5.5 (1.4)	6.4 (0.9)	9.7 (4.8)
9. Keeled in one row	4 (0.7)	4 (0.6)	3.1 (0.7)	3.5 (0.9)	3.5 (1.0	5 (0.0)	4.5 (0.8)	3.9 (0.8)	6.2 (1.2)
10. Paramedian tubercles	31.6 (2.9)	36.7 (1.8)	28.6 (6.2)	34.4 (1.2)	38.7 (4.5)	39 (0.0)	40.7 (3.7)	40.2 (3.3)	31 (3.5)
11. Paramedian ocelli	6.8 (4.3)	8.6 (1.4)	11.5 (1.3)	8.8 (1.0)	8 (0.8)	9.5 (0.7)	10.3 (2.0)	9.8 (1.7)	9.7 (1.2)
12. Longitudinal ocelli	1.8 (1.0)	3.7 (0.6)	4.6 (1.2)	3.5 (0.7)	3 (0.8)	4.5 (0.7)	3.4 (1.1)	3.5 (0.9)	4.5 (0.5)

Characters	*Gymnodactylus* lineages (n)								
	Clade A (5)	Clade B (11)	Clade C (6)	Clade D (7)	Clade E (4)	Clade F (2)	Clade G (14)	Clade H (18)	Manga (4)
13. Longitudinal tubercles	14.8 (1.0)	13.9 (0.3)	12.3 (1.2)	13.8 (0.6)	13.2 (1.5)	14 (0.0)	14 (0.6)	14.1 (0.6)	11.7 (0.5)
14. Longitudinal ventral rows	18.8 (0.8)	18.9 (1.2)	19.6 (1.0)	18.5 (1.3)	18 (1.4)	19 (1.4)	19.5 (1.4)	19.5 (1.1)	18.5 (0.5)
15. Between mental	5 (2.1)	4.9 (1.3)	3.5 (0.8)	5.2 (1.3)	5.5 (1.0)	4 (0.0)	6.5 (1.2)	6.1 (1.7)	3.2 (2.2)
16. Lamellae fouth finger	12.6 (0.5)	12 (0.7)	12.3 (1.0)	12 (1.1)	12.5 (1.2)	11.5 (0.7)	12.2 (0.6)	11.4 (0.7)	13 (1.1)
17. Ventrals	73 (4.9)	69.1 (5.0)	77.5 (5.6)	69 (8.4)	70.5 (9.4)	73.5 (4.9)	79.7 (6.8)	68.5 (9.2)	72.2 (3.8)
18. Femorals and tibials	19 (0.7)	21 (1.1)	21 (1.5)	20.1 (0.6)	20.7 (0.5)	23 (0.0)	19.5 (0.9)	19.7 (1.2)	19 (1.1)
19. Lamellae fourth toe	15.6 (1.1)	14.5 (0.8)	15.5 (1.7)	15.5 (1.3)	14.7 (0.5)	14.5 (0.7)	14.9 (1.2)	14.5 (0.8)	17 (0.8)
20. Before subcaudal	6.2 (0.8)	5.7 (1.5)	6.8 (0.9)	5.2 (0.9)	4.7 (1.7)	7 (2.8)	7.2 (1.9)	5.9 (1.4)	5.5 (2.0)
21. Bands in tail	13.6 (0.8)	12.2 (2.0)	14.3 (2.4)	11.5 (1.7)	11.5 (2.5)	9 (1.4)	12.5 (1.7)	13.1 (2.2)	8.7 (3.2)
22. Postnasal size	0	1	1	1	1	0	0	1	0
23. Supranasals contact	2	1	2	1	0	1	1	1	2
24. Frontonasals alignment	1	1	0	1	1	0	1	0	0
25. Ear opening	0	0	0	0	0	2	2	0	0
26. Ear position	0	0	0	0	0	0	0	0	0
27. Dorsal ocelli	0	0	0	0	0	0	0	0	0
28. Limbs ocelli	1	0	0	0	1	0	1	0	0
29. Tail bands	0	0	0	0	0	0	0	0	0

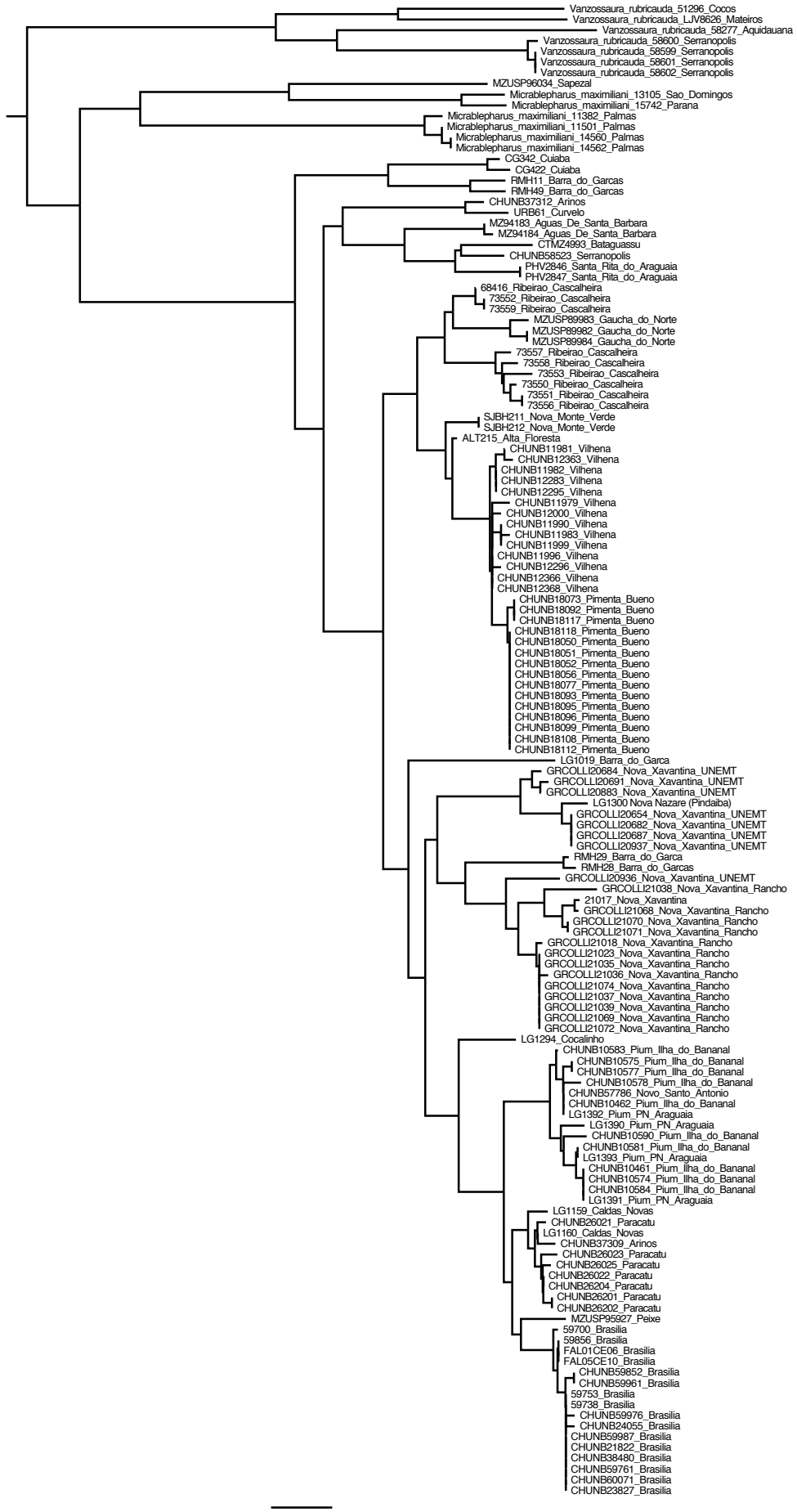
Appendix 13: Cladrogram depicting the 'species tree' hypothesis based on the concatenated phylogenies (Bayesian and ML – Fig.1, Table 1, Chapter 1) used as input in BPP. Numbers in nodes are posterior probabilities of the species splits estimated by BPP. Clades A to H refer to *Gymnodactylus amarali* clades identified by GMYC analysis.



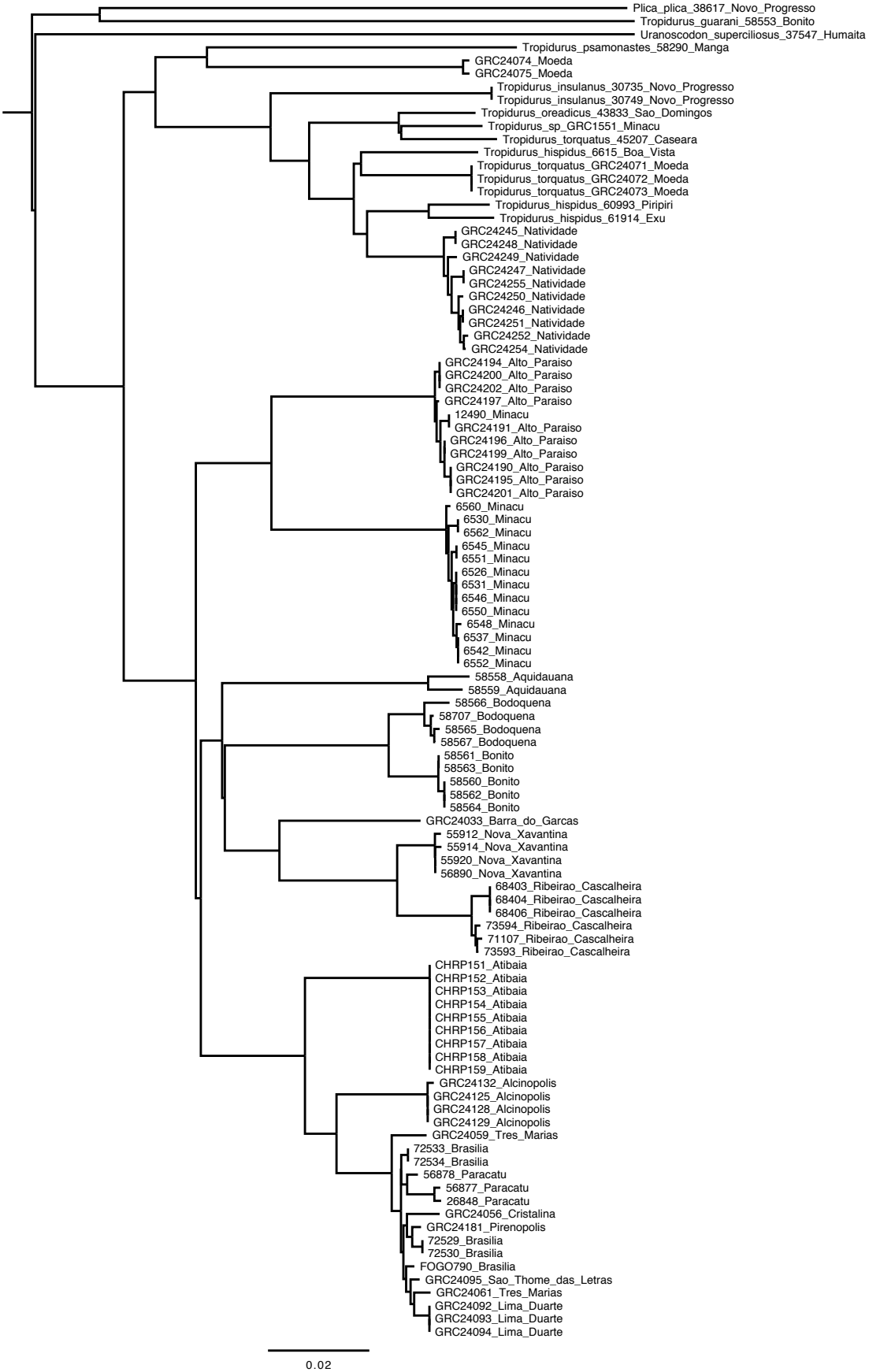
Appendix 14: *Gymnodactylus amarali* Neighbour-Joining phylogenetic tree based on p-distance estimated using MEGA5.2.2 (Tamura *et al.*, 2011).


Gdarwinii_new_58323_Matias_Cardoso
Gdarwinii_new_58335_Matias_Cardoso
Gdarwinii_new_58348_Matias_Cardoso
Ggeckoides_new_58336_Manga
Ggeckoides_new_58337_Manga
Ggeckoides_new_58338_Manga
Ggeckoides_new_58341_Manga
Ggeckoides_56643_Mamanguape
Ggeckoides_56644_Mamanguape
Ggeckoides_56645_Mamanguape
Ggeckoides_56697_Mamanguape
Ggeckoides_61905_Exu
Ggeckoides_61946_Junco_do_Serido
Ggeckoides_56537_Milagres
Ggeckoides_61904_Santana_do_Cariri
69384_Sao_Domingos
15533_Sao_Domingos
15534_Sao_Domingos
15558_Sao_Domingos
15560_Sao_Domingos
15564_Sao_Domingos
15568_Sao_Domingos
15569_Sao_Domingos
GRC21228_Nova_Xavantina
63204_Nova_Xavantina
LG0889_Barra_do_Garca
63195_Nova_Xavantina
55880_Nova_Xavantina
55884_Nova_Xavantina
55885_Nova_Xavantina
55886_Nova_Xavantina
55887_Nova_Xavantina
55888_Nova_Xavantina
55889_Nova_Xavantina
55890_Nova_Xavantina
2173_Minacu
2204_Minacu
2199_Minacu
2207_Minacu
2152_Minacu
2153_Minacu
2154_Minacu
831_Minacu
LG1072_Niquelandia
LG1075_Niquelandia
LG1083_Niquelandia
2202_Minacu
2211_Minacu
832_Minacu
44698_Colinas_do_Sul
44704_Colinas_do_Sul
GRC24210_Alto_Paraiso
GRC24211_Alto_Paraiso
38645_Cocalzinho_de_Goias
67441_Pirenopolis
67438_Pirenopolis
38646_Cocalzinho_de_Goias
67437_Pirenopolis
67440_Pirenopolis
67393_Pirenopolis
67399_Pirenopolis
67392_Pirenopolis
67394_Pirenopolis
67395_Pirenopolis
67396_Pirenopolis
67397_Pirenopolis
67398_Pirenopolis
67443_Pirenopolis
63198_Barra_do_Garca
55883_Barra_do_Garca
63197_Barra_do_garca
63196_Barra_do_Garca
55881_Barra_do_Garca
55882_Barra_do_Garca
10527_Palmas
14555_Palmas
10488_Palmas
10495_Palmas
10771_Palmas
10773_Palmas
10775_Palmas
37090_Parana
37128_Parana
15781_Parana
37125_Parana
15743_Parana
15780_Parana
15782_Parana
15797_Parana
15767_Parana
15795_Parana
15799_Parana
15869_Parana
53290_Cavalcante
53291_Cavalcante
53292_Cavalcante
53293_Cavalcante
53294_Cavalcante
62616_Combinado
53052_Monte_Alegre
15484_Sao_Domingos
15566_Sao_Domingos
33569_Parana
53051_Monte_Alegre
53054_Monte_Alegre
53060_Monte_Alegre
37161_Sao_Domingos
53055_Monte_Alegre
53056_Monte_Alegre
53058_Monte_Alegre
53059_Monte_Alegre
53061_Monte_Alegre
53062_Monte_Alegre
MRT3949_Peixe
52612_Peixe
MRT6428_Sao_Salvador
52609_Peixe
MRT6435_Sao_Salvador
62666_Peixe
GRC24256_Natividade
MRT4459_Peixe
62569_Figueiropolis
GRC24310_Peixe
GRC24311_Peixe
45278_Caseara
45321_Caseara
45283_Caseara
45318_Caseara
45324_Caseara
45336_Caseara
10433_Palmas
MTR14510_Almas
MRT14241_Lajeado
MRT14578_Almas
MRT14609_Almas
MRT14808_Almas
MTR14237_Almas
MTR14552_Almas
MRT14517_Almas
MRT6889_Lajeado
MTR8918_Lajeado
10440_Palmas
10442_Palmas
LAJ215_Lajeado
MRT8991_Lajeado
MRT9034_Lajeado
28219_Mateiros
28220_Mateiros
28222_Mateiros
28224_Mateiros
8674_Mateiros
8675_Mateiros
8743_Mateiros
8808_Mateiros
8841_Mateiros
8928_Mateiros
8980_Mateiros
ESTR1759_Goiatins
ESTR1038_Carolina
ESTR196_Carolina
ESTR197_Carolina
MRT7479_Guarai
MRT7542_Guarai
MRT7552_Guarai
MRT7598_Guarai
52006_Carolina
52007_Carolina
ESTR1808_Goiatins
ESTR1293_Estreito
ESTR642_Estreito
0.02

Appendix 15: *Micrablepharus atticolus* Neighbour-Joining phylogenetic tree based on p-distance estimated using MEGA5.2.2 (Tamura et al., 2011).


Vanzossaura_rubricauda_51296_Cocos
Vanzossaura_rubricauda_LJV8626_Mateiros
Vanzossaura_rubricauda_58277_Aquidauana
Vanzossaura_rubricauda_58600_Serranopolis
Vanzossaura_rubricauda_58599_Serranopolis
Vanzossaura_rubricauda_58601_Serranopolis
Vanzossaura_rubricauda_58602_Serranopolis
MZUSP96034_Sapezal
Micrablepharus_maximiliani_13105_Sao_Domingos
Micrablepharus_maximiliani_15742_Parana
Micrablepharus_maximiliani_11382_Palmas
Micrablepharus_maximiliani_11501_Palmas
Micrablepharus_maximiliani_14560_Palmas
Micrablepharus_maximiliani_14562_Palmas
CG342_Cuiaba
CG422_Cuiaba
RMH11_Barra_do_Garcas
RMH49_Barra_do_Garcas
CHUNB37312_Arinos
URB61_Curvelo
MZ94183_Aguas_De_Santa_Barbara
MZ94184_Aguas_De_Santa_Barbara
CTMZ4993_Bataguassu
CHUNB58523_Serranopolis
PHV2846_Santa_Rita_do_Araguaia
PHV2847_Santa_Rita_do_Araguaia
68416_Ribeirao_Cascalheira
73552_Ribeirao_Cascalheira
73559_Ribeirao_Cascalheira
MZUSP89983_Gaucha_do_Norte
MZUSP89982_Gaucha_do_Norte
MZUSP89984_Gaucha_do_Norte
73557_Ribeirao_Cascalheira
73558_Ribeirao_Cascalheira
73553_Ribeirao_Cascalheira
73550_Ribeirao_Cascalheira
73551_Ribeirao_Cascalheira
73556_Ribeirao_Cascalheira
SJBH211_Nova_Monte_Verde
SJBH212_Nova_Monte_Verde
ALT215_Alta_Floresta
CHUNB11981_Vilhena
CHUNB12363_Vilhena
CHUNB11982_Vilhena
CHUNB12283_Vilhena
CHUNB12295_Vilhena
CHUNB11979_Vilhena
CHUNB12000_Vilhena
CHUNB11990_Vilhena
CHUNB11983_Vilhena
CHUNB11999_Vilhena
CHUNB11996_Vilhena
CHUNB12296_Vilhena
CHUNB12366_Vilhena
CHUNB12369_Vilhena
CHUNB18073_Pimenta_Bueno
CHUNB18092_Pimenta_Bueno
CHUNB18117_Pimenta_Bueno
CHUNB18118_Pimenta_Bueno
CHUNB18050_Pimenta_Bueno
CHUNB18051_Pimenta_Bueno
CHUNB18052_Pimenta_Bueno
CHUNB18056_Pimenta_Bueno
CHUNB18077_Pimenta_Bueno
CHUNB18093_Pimenta_Bueno
CHUNB18095_Pimenta_Bueno
CHUNB18096_Pimenta_Bueno
CHUNB18099_Pimenta_Bueno
CHUNB18108_Pimenta_Bueno
CHUNB18112_Pimenta_Bueno
LG1019_Barra_do_Garca
GRCOLLI20684_Nova_Xavantina_UNEMT
GRCOLLI20691_Nova_Xavantina_UNEMT
GRCOLLI20883_Nova_Xavantina_UNEMT
LG1300 Nova Nazare (Pindaiba)
GRCOLLI20654_Nova_Xavantina_UNEMT
GRCOLLI20682_Nova_Xavantina_UNEMT
GRCOLLI20687_Nova_Xavantina_UNEMT
GRCOLLI20937_Nova_Xavantina_UNEMT
RMH29_Barra_do_Garca
RMH28_Barra_do_Garcas
GRCOLLI20936_Nova_Xavantina_UNEMT
GRCOLLI21038_Nova_Xavantina_Rancho
21017_Nova_Xavantina
GRCOLLI21068_Nova_Xavantina_Rancho
GRCOLLI21070_Nova_Xavantina_Rancho
GRCOLLI21071_Nova_Xavantina_Rancho
GRCOLLI21018_Nova_Xavantina_Rancho
GRCOLLI21023_Nova_Xavantina_Rancho
GRCOLLI21035_Nova_Xavantina_Rancho
GRCOLLI21036_Nova_Xavantina_Rancho
GRCOLLI21074_Nova_Xavantina_Rancho
GRCOLLI21037_Nova_Xavantina_Rancho
GRCOLLI21039_Nova_Xavantina_Rancho
GRCOLLI21069_Nova_Xavantina_Rancho
GRCOLLI21072_Nova_Xavantina_Rancho
LG1294_Cocalinho
CHUNB10583_Pium_Ilha_do_Bananal
CHUNB10575_Pium_Ilha_do_Bananal
CHUNB10577_Pium_Ilha_do_Bananal
CHUNB10578_Pium_Ilha_do_Bananal
CHUNB57786_Novo_Santo_Antonio
CHUNB10462_Pium_Ilha_do_Bananal
LG1392_Pium_PN_Araguaia
LG1390_Pium_PN_Araguaia
CHUNB10590_Pium_Ilha_do_Bananal
CHUNB10581_Pium_Ilha_do_Bananal
LG1393_Pium_PN_Araguaia
CHUNB10461_Pium_Ilha_do_Bananal
CHUNB10574_Pium_Ilha_do_Bananal
CHUNB10584_Pium_Ilha_do_Bananal
LG1391_Pium_PN_Araguaia
LG1159_Caldas_Novas
CHUNB26021_Paracatu
LG1160_Caldas_Novas
CHUNB37309_Arinos
CHUNB26023_Paracatu
CHUNB26025_Paracatu
CHUNB26022_Paracatu
CHUNB26204_Paracatu
CHUNB26201_Paracatu
CHUNB26202_Paracatu
MZUSP95927_Peixe
59700_Brasilia
59856_Brasilia
FAL01CE06_Brasilia
FAL05CE10_Brasilia
CHUNB59852_Brasilia
CHUNB59961_Brasilia
59753_Brasilia
59738_Brasilia
CHUNB59976_Brasilia
CHUNB24055_Brasilia
CHUNB59987_Brasilia
CHUNB21822_Brasilia
CHUNB38480_Brasilia
CHUNB59761_Brasilia
CHUNB60071_Brasilia
CHUNB23827_Brasilia
0.01

Appendix 16: *Tropidurus itambere* Neighbour-Joining phylogenetic tree based on p-distance estimated using MEGA5.2.2 (Tamura *et al.*, 2011).



Appendix 17: *Gymnodactylus amarali* and outgroup specimens sequenced for cytochrome b. Individuals in **bold** are those chosen for AP sequencing. The ones marked with asterisks (***) denote samples sent for AP sequencing, but failed to be captured.

Species	ID	Locality	Brazilian State	Latitude	Longitude	Altitude
G. amarali	**GRC24211**	**Alto Paraíso**	**GO**	**-14.16**	**-47.62**	1166
G. amarali	**GRC24210**	**Alto Paraíso**	**GO**	**-14.16**	**-47.62**	1166
G. amarali	53290	Cavalcante	GO	-13.64	-47.72	961
G. amarali	53292	Cavalcante	GO	-13.64	-47.72	961
G. amarali	53294	Cavalcante	GO	-13.64	-47.72	961
G. amarali	**53293**	**Cavalcante**	**GO**	**-13.64**	**-47.72**	961
G. amarali	53291	Cavalcante	GO	-13.64	-47.72	961
G. amarali	38645	Cocalzinho de Goiás	GO	-15.64	-48.55	738
G. amarali	38646	Cocalzinho de Goiás	GO	-15.64	-48.55	738
G. amarali	44704	Colinas do Sul	GO	-13.99	-48.09	511
G. amarali	**44698**	**Colinas do Sul**	**GO**	**-13.99**	**-48.09**	511
G. amarali	2202	Minaçu	GO	-13.50	-48.40	427
G. amarali	2211	Minaçu	GO	-13.50	-48.40	427
G. amarali	2153	Minaçu	GO	-13.50	-48.40	427
G. amarali	2154	Minaçu	GO	-13.50	-48.40	427
G. amarali	2207	Minaçu	GO	-13.50	-48.40	427
G. amarali	2199	Minaçu	GO	-13.50	-48.40	427
G. amarali	831	Minaçu	GO	-13.50	-48.40	427
G. amarali	832	Minaçu	GO	-13.50	-48.40	427
G. amarali	2204	Minaçu	GO	-13.50	-48.40	427
G. amarali	2173	Minaçu	GO	-13.50	-48.40	427
G. amarali	**2152**	**Minaçu**	**GO**	**-13.50**	**-48.40**	427
G. amarali	53059	Monte Alegre de Goiás	GO	-13.25	-46.90	564
G. amarali	53062	Monte Alegre de Goiás	GO	-13.25	-46.90	564
G. amarali	53061	Monte Alegre de Goiás	GO	-13.25	-46.90	564
G. amarali	53052	Monte Alegre de Goiás	GO	-13.25	-46.90	564
G. amarali	53054	Monte Alegre de Goiás	GO	-13.25	-46.90	564
G. amarali	53058	Monte Alegre de Goiás	GO	-13.25	-46.90	564
G. amarali	53056	Monte Alegre de Goiás	GO	-13.25	-46.90	564
G. amarali	**53060**	**Monte Alegre de Goiás**	**GO**	**-13.25**	**-46.90**	564
G. amarali	53055	Monte Alegre de Goiás	GO	-13.25	-46.90	564
G. amarali	53051	Monte Alegre de Goiás	GO	-13.25	-46.90	564
G. amarali	LG1075	Niquelândia	GO	-14.45	-48.45	594
G. amarali	LG1083	Niquelândia	GO	-14.45	-48.45	594
G. amarali	**LG1072*** **	**Niquelândia**	**GO**	**-14.45**	**-48.45**	594
G. amarali	67395	Pirenópolis	GO	-15.81	-48.87	1324
G. amarali	67440	Pirenópolis	GO	-15.81	-48.87	1324
G. amarali	67396	Pirenópolis	GO	-15.81	-48.87	1324

Species	ID	Locality	Brazilian State	Latitude	Longitude	Altitude
G. amarali	67399	Pirenópolis	GO	-15.81	-48.87	1324
G. amarali	67392	Pirenópolis	GO	-15.81	-48.87	1324
G. amarali	67394	Pirenópolis	GO	-15.81	-48.87	1324
G. amarali	67397	Pirenópolis	GO	-15.81	-48.87	1324
G. amarali	67437	Pirenópolis	GO	-15.81	-48.87	1324
G. amarali	67441	Pirenópolis	GO	-15.81	-48.87	1324
G. amarali	67438	Pirenópolis	GO	-15.81	-48.87	1324
G. amarali	67393	Pirenópolis	GO	-15.81	-48.87	1324
G. amarali	**67398*** **	**Pirenópolis**	**GO**	**-15.81**	**-48.87**	1324
G. amarali	**67443*** **	**Pirenópolis**	**GO**	**-15.81**	**-48.87**	1324
G. amarali	15564	São Domingos	GO	-13.45	-46.45	456
G. amarali	15560	São Domingos	GO	-13.45	-46.45	456
G. amarali	15558	São Domingos	GO	-13.45	-46.45	456
G. amarali	15568	São Domingos	GO	-13.45	-46.45	456
G. amarali	37161	São Domingos	GO	-13.45	-46.45	456
G. amarali	**15533*** **	**São Domingos**	**GO**	**-13.45**	**-46.45**	456
G. amarali	15566	São Domingos	GO	-13.45	-46.45	456
G. amarali	15569	São Domingos	GO	-13.45	-46.45	456
G. amarali	15484	São Domingos	GO	-13.45	-46.45	456
G. amarali	**69384**	**São Domingos**	**GO**	**-13.45**	**-46.45**	456
G. amarali	15534	São Domingos	GO	-13.45	-46.45	456
G. amarali	ESTR00197	Carolina	MA	-7.37	-47.43	176
G. amarali	52006	Carolina	MA	-7.37	-47.43	176
G. amarali	ESTR00196	Carolina	MA	-7.37	-47.43	176
G. amarali	ESTR1038	Carolina	MA	-7.37	-47.43	176
G. amarali	**52007**	**Carolina**	**MA**	**-7.37**	**-47.43**	176
G. amarali	**ESTR1293**	**Estreito**	**MA**	**-6.56**	**-47.45**	163
G. amarali	ESTR0642	Estreito	MA	-6.56	-47.45	163
G. amarali	**LG0889**	**Barra do Garça**	**MT**	**-15.88**	**-52.25**	294
G. amarali	55883	Barra do Garça	MT	-15.20	-52.50	619
G. amarali	55882	Barra do Garça	MT	-15.20	-52.50	619
G. amarali	55881	Barra do Garça	MT	-15.20	-52.50	619
G. amarali	CHUNB63196	Barra do Garças	MT	-15.20	-52.50	619
G. amarali	**CHUNB63198**	**Barra do Garças**	**MT**	**-15.20**	**-52.50**	619
G. amarali	CHUNB63197	Barra do Garças	MT	-15.20	-52.50	619
G. amarali	55884	Nova Xavantina	MT	-14.69	-52.34	283
G. amarali	63195	Nova Xavantina	MT	-14.69	-52.34	283
G. amarali	55885	Nova Xavantina	MT	-14.69	-52.34	283
G. amarali	55886	Nova Xavantina	MT	-14.69	-52.34	283
G. amarali	63204	Nova Xavantina	MT	-14.69	-52.34	283
G. amarali	**55880*** **	**Nova Xavantina**	**MT**	**-14.69**	**-52.34**	283
G. amarali	55887	Nova Xavantina	MT	-14.69	-52.34	283
G. amarali	55888	Nova Xavantina	MT	-14.69	-52.34	283

Species	ID	Locality	Brazilian State	Latitude	Longitude	Altitude
G. amarali	55890	Nova Xavantina	MT	-14.69	-52.34	283
G. amarali	55889	Nova Xavantina	MT	-14.69	-52.34	283
G. amarali	**GRC21228**	**Nova Xavantina**	**MT**	**-14.69**	**-52.34**	283
G. amarali	MTR14517	Almas (EESGT)	TO	-11.24	-46.81	522
G. amarali	MTR14808	Almas (EESGT)	TO	-11.24	-46.81	522
G. amarali	MTR14609	Almas (EESGT)	TO	-11.24	-46.81	522
G. amarali	**MTR14237**	**Almas (EESGT)**	**TO**	**-11.24**	**-46.81**	522
G. amarali	MTR14552	Almas (EESGT)	TO	-11.24	-46.81	522
G. amarali	MTR14578	Almas (EESGT)	TO	-11.24	-46.81	522
G. amarali	MTR14510	Almas (EESGT)	TO	-11.24	-46.81	522
G. amarali	45278	Caseara	TO	-9.37	-49.84	185
G. amarali	45318	Caseara	TO	-9.37	-49.84	185
G. amarali	45324	Caseara	TO	-9.37	-49.84	185
G. amarali	**45336**	**Caseara**	**TO**	**-9.37**	**-49.84**	185
G. amarali	45321	Caseara	TO	-9.37	-49.84	185
G. amarali	45283	Caseara	TO	-9.37	-49.84	185
G. amarali	62616	Combinado	TO	-12.81	-46.48	404
G. amarali	**CHUNB62569**	**Figueirópolis**	**TO**	**-12.18**	**-48.96**	288
G. amarali	ESTR1759	Goiatins	TO	-7.69	-47.35	187
G. amarali	ESTR1808	Goiatins	TO	-7.69	-47.35	187
G. amarali	MRT7479	Guarai	TO	-8.83	-48.52	265
G. amarali	**MRT7542*** **	**Guarai**	**TO**	**-8.83**	**-48.52**	265
G. amarali	MRT7598	Guarai	TO	-8.83	-48.52	265
G. amarali	MRT7552	Guarai	TO	-8.83	-48.52	265
G. amarali	MRT8991	Lajeado (UHE)	TO	-9.85	-48.32	487
G. amarali	MRT6918	Lajeado (UHE)	TO	-9.85	-48.32	487
G. amarali	MRT14241	Lajeado (UHE)	TO	-9.85	-48.32	487
G. amarali	MRT6889	Lajeado (UHE)	TO	-9.85	-48.32	487
G. amarali	**LAJ215**	**Lajeado (UHE)**	**TO**	**-9.85**	**-48.32**	487
G. amarali	MRT9034	Lajeado (UHE)	TO	-9.85	-48.32	487
G. amarali	8674	Mateiros	TO	-10.70	-46.41	632
G. amarali	8743	Mateiros	TO	-10.70	-46.41	632
G. amarali	8675	Mateiros	TO	-10.70	-46.41	632
G. amarali	8841	Mateiros	TO	-10.70	-46.41	632
G. amarali	8928	Mateiros	TO	-10.70	-46.41	632
G. amarali	28222	Mateiros	TO	-10.70	-46.41	632
G. amarali	8808	Mateiros	TO	-10.70	-46.41	632
G. amarali	8980	Mateiros	TO	-10.70	-46.41	632
G. amarali	28224	Mateiros	TO	-10.70	-46.41	632
G. amarali	**28219**	**Mateiros**	**TO**	**-10.70**	**-46.41**	632
G. amarali	28220	Mateiros	TO	-10.70	-46.41	632
G. amarali	**GRC24256**	**Natividade**	**TO**	**-11.69**	**-47.70**	719
G. amarali	10433	Palmas_East	TO	-10.42	-48.36	233

Species	ID	Locality	Brazilian State	Latitude	Longitude	Altitude
G. amarali	10440	Palmas_East	TO	-10.42	-48.36	233
G. amarali	**10442**	**Palmas_East**	**TO**	**-10.42**	**-48.36**	233
G. amarali	10488	Palmas_West	TO	-10.24	-48.44	226
G. amarali	10495	Palmas_West	TO	-10.24	-48.44	226
G. amarali	10527	Palmas_West	TO	-10.24	-48.44	226
G. amarali	10771	Palmas_West	TO	-10.24	-48.44	226
G. amarali	10773	Palmas_West	TO	-10.24	-48.44	226
G. amarali	**10775**	**Palmas_West**	**TO**	**-10.24**	**-48.44**	226
G. amarali	14555	Palmas_West	TO	-10.24	-48.44	226
G. amarali	15767	Paranã	TO	-12.75	-47.76	284
G. amarali	15743	Paranã	TO	-12.75	-47.76	284
G. amarali	33569	Paranã	TO	-12.75	-47.76	284
G. amarali	15782	Paranã	TO	-12.75	-47.76	284
G. amarali	15780	Paranã	TO	-12.75	-47.76	284
G. amarali	37125	Paranã	TO	-12.75	-47.76	284
G. amarali	15869	Paranã	TO	-12.75	-47.76	284
G. amarali	**15799**	**Paranã**	**TO**	**-12.75**	**-47.76**	284
G. amarali	37128	Paranã	TO	-12.75	-47.76	284
G. amarali	15795	Paranã	TO	-12.75	-47.76	284
G. amarali	15797	Paranã	TO	-12.75	-47.76	284
G. amarali	15781	Paranã	TO	-12.75	-47.76	284
G. amarali	37090	Paranã	TO	-12.75	-47.76	284
G. amarali	GRC24311	Peixe	TO	-11.88	-48.77	293
G. amarali	MRT3949	Peixe	TO	-12.03	-48.35	326
G. amarali	52612	Peixe	TO	-12.03	-48.35	326
G. amarali	52609	Peixe	TO	-12.03	-48.35	326
G. amarali	MRT4459	Peixe	TO	-12.03	-48.35	326
G. amarali	62666	Peixe	TO	-12.03	-48.35	326
G. amarali	**GRC24310**	**Peixe**	**TO**	**-12.03**	**-48.35**	326
G. amarali	**MRT6435**	**São Salvador**	**TO**	**-12.73**	**-48.23**	306
G. amarali	MRT6428	São Salvador	TO	-12.73	-48.23	306
Outgroups						
G. darwinii	58348	Matias Cardoso	MG	-14.99	-43.95	464
G. darwinii	58323	Matias Cardoso	MG	-14.99	-43.95	464
G. darwinii	**58335**	**Matias Cardoso**	**MG**	**-14.99**	**-43.95**	464
G. geckoides	CHUNB56537	Milagres	CE	-7.29	-38.94	406
G. geckoides	CHUNB 61904	Santana do Cariri	CE	-7.21	-39.73	755
G. geckoides	58341	Manga	MG	-14.84	-43.99	477
G. geckoides	58337	Manga	MG	-14.84	-43.99	477
G. geckoides	**58336**	**Manga**	**MG**	**-14.84**	**-43.99**	477
G. geckoides	58338	Manga	MG	-14.84	-43.99	477

Species	**ID**	**Locality**	**Brazilian State**	**Latitude**	**Longitude**	**Altitude**
G. geckoides	CHUNB 61946	Junco do Seridó	PA	-7.00	-36.71	627
G. geckoides	CHUNB56643	Mamanguape	PB	-6.80	-35.20	90
G. geckoides	**CHUNB56644**	**Mamanguape**	**PB**	**-6.80**	**-35.20**	90
G. geckoides	CHUNB56645	Mamanguape	PB	-6.80	-35.20	90
G. geckoides	CHUNB56697	Mamanguape	PB	-6.80	-35.20	90
G. geckoides	CHUNB 61905	Exu	PE	-7.43	-39.75	631

Appendix 18: *Micrablepharus atticolus* and outgroup specimens sequenced for cytochrome b. Individuals in **bold** are those chosen for AP sequencing. The ones marked with asterisks (***) denote samples sent for AP sequencing, but failed to be captured.

Species	ID	Locality	Brazilian State	Latitude	Longitude	Altitude
M. atticolus	CHUNB59976	Brasilia	DF	-15.97	-47.91	1170
M. atticolus	CHUNB59961	Brasilia	DF	-15.97	-47.91	1170
M. atticolus	CHUNB24055	Brasilia	DF	-15.78	-47.80	1101
M. atticolus	CHUNB21822	Brasilia	DF	-16.01	-47.94	1157
M. atticolus	CHUNB59753	Brasilia	DF	-15.70	-47.92	1063
M. atticolus	CHUNB59761	Brasilia	DF	-15.70	-47.92	1063
M. atticolus	CHUNB59738	Brasilia	DF	-15.70	-47.92	1063
M. atticolus	CHUNB59856	Brasilia	DF	-15.97	-47.91	1170
M. atticolus	CHUNB59700	Brasilia	DF	-15.70	-47.92	1063
M. atticolus	CHUNB60071	Brasilia	DF	-15.70	-47.92	1063
M. atticolus	CHUNB23827	Brasilia	DF	-16.01	-47.96	1186
M. atticolus	CHUNB59852	Brasilia	DF	-15.97	-47.91	1170
M. atticolus	**CHUNB59987**	**Brasilia**	**DF**	**-15.97**	**-47.91**	1170
M. atticolus	**CHUNB38480**	**Brasilia**	**DF**	**-15.93**	**-47.88**	1124
M. atticolus	FAL01CE06	Brasilia	DF	-15.97	-47.91	1170
M. atticolus	FAL05CE10	Brasilia	DF	-15.97	-47.91	1170
M. atticolus	LG1159	Caldas Novas	GO	-17.73	-48.62	669
M. atticolus	LG1160	Caldas Novas	GO	-17.73	-48.62	669
M. atticolus	PHV2846	Santa Rita do Araguaia	GO	-17.23	-53.16	632
M. atticolus	**PHV2847**	**Santa Rita do Araguaia**	**GO**	**-17.23**	**-53.16**	632
M. atticolus	CHUNB58523	Serranópolis	GO	-18.33	-51.97	648
M. atticolus	CHUNB37309	Arinos	MG	-15.45	-45.83	813
M. atticolus	**CHUNB37312**	**Arinos**	**MG**	**-15.45**	**-45.83**	813
M. atticolus	URB61	Curvelo/Pompéu	MG	-19.03	-44.71	652
M. atticolus	CHUNB26025	Paracatu	MG	-17.40	-47.30	990
M. atticolus	CHUNB26201	Paracatu	MG	-17.40	-47.30	990
M. atticolus	CHUNB26204	Paracatu	MG	-17.40	-47.30	990
M. atticolus	CHUNB26022	Paracatu	MG	-17.40	-47.30	990
M. atticolus	**CHUNB26023**	**Paracatu**	**MG**	**-17.40**	**-47.30**	990
M. atticolus	**CHUNB26202**	**Paracatu**	**MG**	**-17.40**	**-47.30**	990
M. atticolus	CHUNB26021	Paracatu	MG	-17.40	-47.30	990
M. atticolus	CTMZ04993	Bataguassu	MS	-21.81	-52.57	375
M. atticolus	ALT215	Alta Floresta	MT	-10.35	-56.98	329
M. atticolus	RMH11	Barra do Garça	MT	-15.25	-53.12	394
M. atticolus	RMH29	Barra do Garça	MT	-15.25	-53.12	394
M. atticolus	**RMH49**	**Barra do Garça**	**MT**	**-15.25**	**-53.12**	394
M. atticolus	**RMH28**	**Barra do Garça**	**MT**	**-15.25**	**-53.12**	394
M. atticolus	**LG1019**	**Barra do Garça**	**MT**	**-15.36**	**-52.50**	818
M. atticolus	CG422	Chapada dos Guimarães	MT	-15.33	-55.95	250
M. atticolus	**LG1294**	**Cocalinho**	**MT**	**-13.87**	**-51.15**	231
M. atticolus	**CG342**	**Cuiaba**	**MT**	**-15.33**	**-55.95**	250
M. atticolus	MZUSP89983	Gaúcha do Norte	MT	-13.23	-53.07	410
M. atticolus	**MZUSP89982**	**Gaúcha do Norte**	**MT**	**-13.23**	**-53.07**	410

Species	ID	Locality	Brazilian State	Latitude	Longitude	Altitude
M. atticolus	**MZUSP89984**	**Gaúcha do Norte**	**MT**	**-13.23**	**-53.07**	410
M. atticolus	SJBH212	Juará	MT	-10.33	-57.65	372
M. atticolus	**SJBH211**	**Nova Monte Verde**	**MT**	**-10.33**	**-57.65**	372
M. atticolus	**LG1300**	**Nova Nazaré (Pindaíba)**	**MT**	**-14.37**	**-51.72**	246
M. atticolus	GRCOLLI21072	Nova Xavantina (Rancho)	MT	-14.80	-52.64	304
M. atticolus	GRCOLLI21068	Nova Xavantina (Rancho)	MT	-14.80	-52.64	304
M. atticolus	GRCOLLI21018	Nova Xavantina (Rancho)	MT	-14.80	-52.64	304
M. atticolus	GRCOLLI21039	Nova Xavantina (Rancho)	MT	-14.80	-52.64	304
M. atticolus	GRCOLLI21038	Nova Xavantina (Rancho)	MT	-14.80	-52.64	304
M. atticolus	GRCOLLI21017	Nova Xavantina (Rancho)	MT	-14.80	-52.64	304
M. atticolus	GRCOLLI21070	Nova Xavantina (Rancho)	MT	-14.80	-52.64	304
M. atticolus	GRCOLLI21035	Nova Xavantina (Rancho)	MT	-14.80	-52.64	304
M. atticolus	GRCOLLI21023	Nova Xavantina (Rancho)	MT	-14.80	-52.64	304
M. atticolus	GRCOLLI21071	Nova Xavantina (Rancho)	MT	-14.80	-52.64	304
M. atticolus	GRCOLLI21069	Nova Xavantina (Rancho)	MT	-14.80	-52.64	304
M. atticolus	GRCOLLI21036	Nova Xavantina (Rancho)	MT	-14.80	-52.64	304
M. atticolus	GRCOLLI21074	Nova Xavantina (Rancho)	MT	-14.80	-52.64	304
M. atticolus	**GRCOLLI21037**	**Nova Xavantina (Rancho)**	**MT**	**-14.80**	**-52.64**	304
M. atticolus	GRCOLLI20684	Nova Xavantina (UNEMT)	MT	-14.70	-52.35	302
M. atticolus	GRCOLLI20883	Nova Xavantina (UNEMT)	MT	-14.70	-52.35	302
M. atticolus	GRCOLLI20654	Nova Xavantina (UNEMT)	MT	-14.70	-52.35	302
M. atticolus	GRCOLLI20682	Nova Xavantina (UNEMT)	MT	-14.70	-52.35	302
M. atticolus	GRCOLLI20691	Nova Xavantina (UNEMT)	MT	-14.70	-52.35	302
M. atticolus	GRCOLLI20936	Nova Xavantina (UNEMT)	MT	-14.70	-52.35	302
M. atticolus	**GRCOLLI20937**	**Nova Xavantina (UNEMT)**	**MT**	**-14.70**	**-52.35**	302
M. atticolus	GRCOLLI20687	Nova Xavantina (UNEMT)	MT	-14.70	-52.35	302
M. atticolus	**CHUNB57786**	**Novo Santo Antônio**	**MT**	**-12.38**	**-50.89**	206
M. atticolus	**CHUNB73552**	**Ribeirão Cascalheira**	**MT**	**-12.94**	**-51.82**	379
M. atticolus	CHUNB73557	Ribeirão Cascalheira	MT	-12.94	-51.82	379
M. atticolus	CHUNB73551	Ribeirão Cascalheira	MT	-12.94	-51.82	379

Species	ID	Locality	Brazilian State	Latitude	Longitude	Altitude
M. atticolus	CHUNB73558	Ribeirão Cascalheira	MT	-12.94	-51.82	379
M. atticolus	CHUNB73559	Ribeirão Cascalheira	MT	-12.94	-51.82	379
M. atticolus	CHUNB73556	Ribeirão Cascalheira	MT	-12.94	-51.82	379
M. atticolus	CHUNB73553	Ribeirão Cascalheira	MT	-12.94	-51.82	379
M. atticolus	**CHUNB73550**	**Ribeirão Cascalheira**	**MT**	**-12.94**	**-51.82**	379
M. atticolus	**CHUNB68416**	**Ribeirão Cascalheira**	**MT**	**-13.00**	**-51.75**	334
M. atticolus	MZUSP96034	Sapezal	MT	-13.53	-58.80	572
M. atticolus	CHUNB18077	Pimenta Bueno	RO	-11.81	-60.72	211
M. atticolus	CHUNB18073	Pimenta Bueno	RO	-11.81	-60.72	211
M. atticolus	CHUNB18099	Pimenta Bueno	RO	-11.81	-60.72	211
M. atticolus	CHUNB18117	Pimenta Bueno	RO	-11.81	-60.72	211
M. atticolus	CHUNB18108	Pimenta Bueno	RO	-11.81	-60.72	211
M. atticolus	CHUNB18118	Pimenta Bueno	RO	-11.81	-60.72	211
M. atticolus	CHUNB18092	Pimenta Bueno	RO	-11.81	-60.72	211
M. atticolus	CHUNB18051	Pimenta Bueno	RO	-11.81	-60.72	211
M. atticolus	CHUNB18052	Pimenta Bueno	RO	-11.81	-60.72	211
M. atticolus	CHUNB18050	Pimenta Bueno	RO	-11.81	-60.72	211
M. atticolus	CHUNB18056	Pimenta Bueno	RO	-11.81	-60.72	211
M. atticolus	CHUNB18095	Pimenta Bueno	RO	-11.81	-60.72	211
M. atticolus	CHUNB18093	Pimenta Bueno	RO	-11.81	-60.72	211
M. atticolus	**CHUNB18096**	**Pimenta Bueno**	**RO**	**-11.81**	**-60.72**	211
M. atticolus	CHUNB18112	Pimenta Bueno	RO	-11.81	-60.72	211
M. atticolus	CHUNB11979	Vilhena	RO	-12.47	-60.29	512
M. atticolus	CHUNB11982	Vilhena	RO	-12.47	-60.29	512
M. atticolus	CHUNB11999	Vilhena	RO	-12.47	-60.29	512
M. atticolus	CHUNB11996	Vilhena	RO	-12.47	-60.29	512
M. atticolus	CHUNB11990	Vilhena	RO	-12.47	-60.29	512
M. atticolus	CHUNB11983	Vilhena	RO	-12.47	-60.29	512
M. atticolus	CHUNB11981	Vilhena	RO	-12.47	-60.29	512
M. atticolus	CHUNB12283	Vilhena	RO	-12.47	-60.29	512
M. atticolus	CHUNB12295	Vilhena	RO	-12.47	-60.29	512
M. atticolus	CHUNB12000	Vilhena	RO	-12.47	-60.29	512
M. atticolus	CHUNB12368	Vilhena	RO	-12.47	-60.29	512
M. atticolus	CHUNB12366	Vilhena	RO	-12.47	-60.29	512
M. atticolus	**CHUNB12363**	**Vilhena**	**RO**	**-12.47**	**-60.29**	512
M. atticolus	CHUNB12296	Vilhena	RO	-12.47	-60.29	512
M. atticolus	MZUSP94184	Águas de Santa Bárbara	SP	-22.79	-49.23	642
M. atticolus	MZUSP94183	Águas de Santa Bárbara	SP	-22.79	-49.23	642
M. atticolus	**MZUSP95927**	**Peixe (UHE)**	**TO**	**-12.03**	**-48.55**	249
M. atticolus	CHUNB10574	Pium (Ilha do Bananal)	TO	-10.45	-50.47	216
M. atticolus	CHUNB10584	Pium (Ilha do Bananal)	TO	-10.45	-50.47	216
M. atticolus	CHUNB10581	Pium (Ilha do Bananal)	TO	-10.45	-50.47	216
M. atticolus	CHUNB10583	Pium (Ilha do Bananal)	TO	-10.45	-50.47	216

Species	ID	Locality	Brazilian State	Latitude	Longitude	Altitude
M. atticolus	CHUNB10462	Pium (Ilha do Bananal)	TO	-10.45	-50.47	216
M. atticolus	**CHUNB10590**	**Pium (Ilha do Bananal)**	**TO**	**-10.45**	**-50.47**	216
M. atticolus	CHUNB10577	Pium (Ilha do Bananal)	TO	-10.45	-50.47	216
M. atticolus	CHUNB10578	Pium (Ilha do Bananal)	TO	-10.45	-50.47	216
M. atticolus	CHUNB10575	Pium (Ilha do Bananal)	TO	-10.45	-50.47	216
M. atticolus	CHUNB10461	Pium (Ilha do Bananal)	TO	-10.45	-50.47	216
M. atticolus	LG1391	Pium (PN Araguaia)	TO	-10.39	-50.13	174
M. atticolus	LG1392	Pium (PN Araguaia)	TO	-10.39	-50.13	174
M. atticolus	LG1393	Pium (PN Araguaia)	TO	-10.39	-50.13	174
M. atticolus	**LG1390**	**Pium (PN Araguaia)**	**TO**	**-10.39**	**-50.13**	174
Outgroups						
M. maximiliani	**13105**	**São Domingos**	**GO**	**-13.45**	**-46.45**	456
M. maximiliani	11382	Palmas	TO	-10.19	-48.11	617
M. maximiliani	11501	Palmas	TO	-10.19	-48.11	617
M. maximiliani	14560	Palmas	TO	-10.19	-48.11	617
M. maximiliani	14562	Palmas	TO	-10.19	-48.11	617
M. maximiliani	15742	Paranã	TO	-12.75	-47.76	284
V. rubricauda	**51296*** **	**Cocos**	**BA**	**-14.55**	**-45.24**	752
V. rubricauda	58599	Serranópolis	GO	-18.33	-51.97	648
V. rubricauda	58600	Serranópolis	GO	-18.33	-51.97	648
V. rubricauda	58601	Serranópolis	GO	-18.33	-51.97	648
V. rubricauda	58602	Serranópolis	GO	-18.33	-51.97	648
V. rubricauda	58277	Aquidauana	MS	-20.46	-55.82	173
V. rubricauda	LJV8626	Mateiros	TO	-10.70	-46.41	632

Appendix 19: *Tropidurus itambere* and outgroup specimens sequenced for cytochrome b. Individuals in **bold** are those chosen for AP sequencing. The ones marked with asterisks (***) denote samples sent for AP sequencing, but failed to be captured.

Species	ID	Locality	Brazilian State	Latitude	Longitude	Altitude
T. itambere	FOGO790	Brasília	DF	-15.97	-47.91	1170
T. itambere	72529	Brasília	DF	-15.97	-47.91	1170
T. itambere	**72530**	**Brasília**	**DF**	**-15.97**	**-47.91**	1170
T. itambere	**72533**	**Brasília**	**DF**	**-15.97**	**-47.91**	1170
T. itambere	72534	Brasília	DF	-15.97	-47.91	1170
T. itambere	24196	Alto Paraíso	GO	-14.16	-47.62	1166
T. itambere	24197	Alto Paraíso	GO	-14.16	-47.62	1166
T. itambere	24195	Alto Paraíso	GO	-14.16	-47.62	1166
T. itambere	**24191**	**Alto Paraíso**	**GO**	**-14.15**	**-47.60**	1171
T. itambere	**24190**	**Alto Paraíso**	**GO**	**-14.15**	**-47.60**	1171
T. itambere	24202	Alto Paraíso	GO	-14.16	-47.62	1166
T. itambere	24194	Alto Paraíso	GO	-14.16	-47.62	1166
T. itambere	24200	Alto Paraíso	GO	-14.16	-47.62	1166
T. itambere	24201	Alto Paraíso	GO	-14.16	-47.62	1166
T. itambere	24199	Alto Paraíso	GO	-14.16	-47.62	1166
T. itambere	**24056**	**Cristalina**	**GO**	**-16.73**	**-47.62**	1191
T. itambere	6531	Minaçu	GO	-13.50	-48.40	427
T. itambere	6550	Minaçu	GO	-13.50	-48.40	427
T. itambere	6526	Minaçu	GO	-13.50	-48.40	427
T. itambere	6562	Minaçu	GO	-13.50	-48.40	427
T. itambere	6551	Minaçu	GO	-13.50	-48.40	427
T. itambere	6542	Minaçu	GO	-13.50	-48.40	427
T. itambere	6530	Minaçu	GO	-13.50	-48.40	427
T. itambere	6560	Minaçu	GO	-13.50	-48.40	427
T. itambere	6548	Minaçu	GO	-13.50	-48.40	427
T. itambere	**6545**	**Minaçu**	**GO**	**-13.50**	**-48.40**	427
T. itambere	6546	Minaçu	GO	-13.50	-48.40	427
T. itambere	6537	Minaçu	GO	-13.50	-48.40	427
T. itambere	**6552**	**Minaçu**	**GO**	**-13.50**	**-48.40**	427
T. itambere	12490	Minaçu	GO	-13.50	-48.40	427
T. itambere	**24181**	**Pirenópolis**	**GO**	**-15.81**	**-48.85**	1268
T. itambere	**38756**	**São Jõao D'Aliança**	**GO**	**-14.70**	**-47.52**	1025
T. itambere	24092	Lima Duarte	MG	-21.74	-43.85	1025
T. itambere	**24093**	**Lima Duarte**	**MG**	**-21.74**	**-43.85**	1025
T. itambere	**24094**	**Lima Duarte**	**MG**	**-21.74**	**-43.85**	1025
T. itambere	**56878**	**Paracatu**	**MG**	**-17.22**	**-46.87**	702
T. itambere	26848	Paracatu	MG	-17.22	-46.87	702
T. itambere	56877	Paracatu	MG	-17.22	-46.87	702
T. itambere	**24095**	**São Thomé das Letras**	**MG**	**-21.66**	**-44.89**	1181
T. itambere	24059	Três Marias	MG	-18.19	-45.32	584
T. itambere	**24061**	**Três Marias**	**MG**	**-18.19**	**-45.32**	584
T. itambere	24132	Alcinópolis	MS	-18.15	-53.68	547
T. itambere	**24128**	**Alcinópolis**	**MS**	**-18.15**	**-53.68**	547
T. itambere	24125	Alcinópolis	MS	-18.15	-53.68	547
T. itambere	**24129**	**Alcinópolis**	**MS**	**-18.15**	**-53.68**	547
T. itambere	58558	Aquidauana	MS	-20.46	-55.62	451

Species	ID	Locality	Brazilian State	Latitude	Longitude	Altitude
T. itambere	**58559**	**Aquidauana**	**MS**	**-20.46**	**-55.62**	451
T. itambere	58565	Bodoquena	MS	-20.70	-56.88	754
T. itambere	58566	Bodoquena	MS	-20.70	-56.88	754
T. itambere	**58567**	**Bodoquena**	**MS**	**-20.70**	**-56.88**	754
T. itambere	**58707*** **	**Bodoquena**	**MS**	**-20.70**	**-56.88**	754
T. itambere	**58564*** **	**Bonito**	**MS**	**-21.15**	**-56.79**	560
T. itambere	58560	Bonito	MS	-21.15	-56.79	560
T. itambere	58563	Bonito	MS	-21.15	-56.79	560
T. itambere	58561	Bonito	MS	-21.15	-56.79	560
T. itambere	**58562**	**Bonito**	**MS**	**-21.15**	**-56.79**	560
T. itambere	**24033**	**Barra do Garças**	**MT**	**-15.86**	**-52.25**	525
T. itambere	55914	Nova Xavantina	MT	-14.69	-52.34	283
T. itambere	**55912**	**Nova Xavantina**	**MT**	**-14.69**	**-52.34**	283
T. itambere	**56890**	**Nova Xavantina**	**MT**	**-14.69**	**-52.34**	283
T. itambere	55920	Nova Xavantina	MT	-14.69	-52.34	283
T. itambere	68403	Ribeirão Cascalheira	MT	-13.00	-51.75	334
T. itambere	68406	Ribeirão Cascalheira	MT	-13.00	-51.75	334
T. itambere	**68404**	**Ribeirão Cascalheira**	**MT**	**-13.00**	**-51.75**	334
T. itambere	71107	Ribeirão Cascalheira	MT	-12.47	-52.37	328
T. itambere	73594	Ribeirão Cascalheira	MT	-12.94	-51.82	379
T. itambere	73593	Ribeirão Cascalheira	MT	-12.94	-51.82	379
T. itambere	CHRP151	Atibaia	SP	-23.16	-46.53	1099
T. itambere	CHRP152	Atibaia	SP	-23.16	-46.53	1099
T. itambere	CHRP159	Atibaia	SP	-23.16	-46.53	1099
T. itambere	CHRP157	Atibaia	SP	-23.16	-46.53	1099
T. itambere	CHRP153	Atibaia	SP	-23.16	-46.53	1099
T. itambere	CHRP156	Atibaia	SP	-23.16	-46.53	1099
T. itambere	CHRP155	Atibaia	SP	-23.16	-46.53	1099
T. itambere	CHRP158	Atibaia	SP	-23.16	-46.53	1099
T. itambere	CHRP154	Atibaia	SP	-23.16	-46.53	1099
Outgroups						
"T. itambere"	**24074**	**Moeda**	**MG**	**-20.29**	**-43.96**	1310
"T. itambere"	24075	Moeda	MG	-20.29	-43.96	1310
"T. itambere"	24246	Natividade	TO	-11.69	-47.70	719
"T. itambere"	24247	Natividade	TO	-11.69	-47.70	719
"T. itambere"	24252	Natividade	TO	-11.69	-47.70	719
"T. itambere"	24248	Natividade	TO	-11.69	-47.70	719
"T. itambere"	**24251**	**Natividade**	**TO**	**-11.69**	**-47.70**	719
"T. itambere"	24250	Natividade	TO	-11.69	-47.70	719
"T. itambere"	24245	Natividade	TO	-11.69	-47.70	719
"T. itambere"	24255	Natividade	TO	-11.69	-47.70	719
"T. itambere"	24254	Natividade	TO	-11.69	-47.70	719
"T. itambere"	24249	Natividade	TO	-11.69	-47.70	719

Species	ID	Locality	Brazilian State	Latitude	Longitude	Altitude
Plica plica	**38617*** **	**Novo Progresso**	**PA**	**-8.60**	**-55.50**	354
T. guarani	**58553*** **	**Bonito**	**MS**	**-21.45**	**-56.79**	485
T. hispidus	**61914**	**Exu**	**PE**	**-7.43**	**-39.75**	631
T. hispidus	60993	Piripiri	PI	-4.10	-41.71	200
T. hispidus	6615	Boa Vista	RO	3.30	-60.80	85
T. insulanus	30735	Novo Progresso	PA	-8.60	-55.50	354
T. insulanus	30749	Novo Progresso	PA	-8.60	-55.50	354
T. oreadicus	**43833**	**São Domingos**	**GO**	**-13.45**	**-46.45**	456
T. psamonastes	58290	Manga	MG	-14.84	-43.99	477
T. sp.	GRCOLLI1551	Minaçu	GO	-13.50	-48.40	427
T. torquatus	**24071**	**Moeda**	**MG**	**-20.29**	**-43.96**	1310
T. torquatus	24072	Moeda	MG	-20.29	-43.96	1310
T. torquatus	24073	Moeda	MG	-20.29	-43.96	1310
T. torquatus	**45207**	**Caseara**	**TO**	**-9.37**	**-49.84**	185
U. superciliosus	**37547**	**Humaitá**	**AM**	**-7.20**	**-62.90**	53

Appendix 20: Average number of reads (coverage) across loci and average number of loci above the coverage threshold (100 reads) of the AP dataset for *Gymnodactylus amarali* (Table A), *Micrablepharus atticolus* (Table B), *Tropidurus itambere* (Table C), and their respective outgroups.

Table A:

ID	Species	Average coverage across loci	Loci passing coverage threshold
MTR14237	*G. amarali*	2402.90	391
53060	*G. amarali*	2665.26	391
GRC21228	*G. amarali*	1942.29	391
10775	*G. amarali*	1978.32	391
GRC24310	*G. amarali*	4633.62	391
53293	*G. amarali*	2992.61	390
44698	*G. amarali*	2265.45	390
ESTR1293	*G. amarali*	2742.30	390
62569	*G. amarali*	2125.05	390
GRC24210	*G. amarali*	2187.93	390
GRC24211	*G. amarali*	2741.00	390
52007	*G. amarali*	3060.78	390
69384	*G. amarali*	4618.60	390
10442	*G. amarali*	2008.26	390
15799	*G. amarali*	5219.43	390
MRT6435	*G. amarali*	2034.43	390
GRC20900	*G. amarali*	1790.44	389
45336	*G. amarali*	2733.41	389
24256	*G. amarali*	2376.59	388
LG0889	*G. amarali*	1571.32	385
2152	*G. amarali*	1058.54	384
28219	*G. amarali*	1546.24	379
LAJ215	*G. amarali*	763.63	378
58335	*G. darwini*	2319.71	390
58336	*G. geckoides*	2496.56	389
56644	*G. geckoides*	2029.24	385

Table B:

ID	Species	Average coverage across loci	Loci passing coverage threshold
21037	*M. atticolus*	3947.57	390
37312	*M. atticolus*	2986.00	389
26202	*M. atticolus*	4275.47	389
RMH28	*M. atticolus*	5667.98	388
LG1019	*M. atticolus*	5429.72	388
38480	*M. atticolus*	3408.07	388
89982	*M. atticolus*	4386.87	388
57786	*M. atticolus*	3600.74	388
26023	*M. atticolus*	4415.24	388
95927	*M. atticolus*	2945.73	388
LG1300	*M. atticolus*	4152.75	388
LG1294	*M. atticolus*	3250.78	387
SJBH211	*M. atticolus*	5084.85	387
18096	*M. atticolus*	5058.77	387
LG1390	*M. atticolus*	3247.01	387
73550	*M. atticolus*	5264.00	387
PHV2847	*M. atticolus*	3449.28	387
10590	*M. atticolus*	3223.66	387
12363	*M. atticolus*	2728.80	387
20937	*M. atticolus*	5071.17	386
73552	*M. atticolus*	2919.59	386
89984	*M. atticolus*	1903.33	384
68416	*M. atticolus*	2281.71	384
RMH49	*M. atticolus*	1678.99	381
CG342	*M. atticolus*	1129.63	375
59987	*M. atticolus*	167.09	254
13105	*M. maximiliani*	5426.50	388

Table C:

ID	Species	Average coverage across loci	Loci passing coverage threshold
24129	*T. itambere*	3005.64	390
58559	*T. itambere*	3293.81	390
58567	*T. itambere*	3108.13	390
6552	*T. itambere*	2501.77	390
24071	*T. itambere*	4313.42	390
24074	*T. itambere*	3601.15	390
24251	*T. itambere*	2551.15	390
56878	*T. itambere*	3140.55	390
24061	*T. itambere*	3214.66	390
24128	*T. itambere*	3308.00	390
24191	*T. itambere*	5325.36	390
24033	*T. itambere*	2823.39	390
72533	*T. itambere*	3687.65	390
24056	*T. itambere*	3031.99	390
6545	*T. itambere*	2588.09	390
24181	*T. itambere*	3601.52	390
55912	*T. itambere*	2903.15	389
24094	*T. itambere*	2125.03	389
24095	*T. itambere*	2238.69	389
72530	*T. itambere*	1851.04	388
24190	*T. itambere*	2320.94	383
58562	*T. itambere*	2100.75	382
38756	*T. itambere*	1037.33	366
24093	*T. itambere*	1730.18	363
68404	*T. itambere*	658.26	273
56890	*T. itambere*	283.34	183
61914	*T. hispidus*	2478.93	390
43833	*T. oreadicus*	3375.33	391
45207	*T. torquatus*	3216.73	391
37547	*U. superciliosus*	2945.09	391

Appendix 21: *Gymnodactylus amarali* loci summary statistics from the final Anchored Phylogenomics dataset used on BPP and phylogenetic analyses. The ones marked with an asterisk (*) were not used in the phylogenetic analyses because the outgroup specimen was not captured. Loci are ordered from highest to lowest molecular diversity (Watterson's θ). Outgroups and cryptic species outside of the monophyletic *G. amarali* group were not included in the calculations.

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L340	23	23	1982	79	3.99	70	11	48	0.11	0.0111	21.4045	21.9605	0.1032
L275	23	23	1913	77	4.03	64	14	160	0.36	0.0103	20.8626	19.7036	-0.2207
L341	23	23	1543	75	4.86	58	20	113	0.32	0.0120	20.3207	18.4506	-0.3653
L287	23	23	1541	73	4.74	44	30	29	0.08	0.0079	19.7788	12.1739	-1.5248
L313	23	23	1960	71	3.62	57	16	89	0.20	0.0070	19.2370	13.6640	-1.1479
L391	23	23	1589	71	4.47	55	16	142	0.39	0.0066	19.2370	10.4783	-1.8041
L7	23	23	1504	71	4.72	57	15	172	0.50	0.0097	19.2370	14.5178	-0.9720
L248	23	23	2018	69	3.42	56	14	119	0.26	0.0057	18.6951	11.5336	-1.5165
L321	23	23	1250	64	5.12	46	18	168	0.58	0.0093	17.3404	11.6443	-1.2972
L358	22	22	2084	63	3.02	52	11	112	0.24	0.0058	17.2823	12.1775	-1.1758
L357	23	23	1705	63	3.70	60	3	60	0.15	0.0065	17.0694	11.0791	-1.3851
L310	21	21	1588	61	3.84	45	16	119	0.36	0.0051	16.9551	8.1143	-2.0927
L9	22	22	1668	61	3.66	53	7	105	0.29	0.0080	16.7336	13.2987	-0.8162
L129	23	23	1370	61	4.45	50	12	74	0.23	0.0068	16.5275	9.3518	-1.7116
L87	23	23	1236	61	4.94	37	26	82	0.29	0.0096	16.5275	11.8814	-1.1082
L277	23	23	1588	60	3.78	48	12	81	0.22	0.0085	16.2566	13.4506	-0.6800
L3	23	23	1672	57	3.41	50	8	59	0.15	0.0070	15.4438	11.6285	-0.9715
L84	23	23	1577	57	3.61	38	19	115	0.32	0.0070	15.4438	11.0277	-1.1244
L94	23	23	964	57	5.91	42	15	176	0.79	0.0101	15.4438	9.7194	-1.4576
L354	22	22	1539	56	3.64	36	20	92	0.27	0.0065	15.3620	9.9957	-1.3847
L400	21	21	1179	55	4.66	42	14	105	0.42	0.0088	15.2874	10.4190	-1.2733
L100	23	23	1575	56	3.56	46	10	69	0.19	0.0063	15.1728	9.9249	-1.3592
L235	23	23	1798	56	3.11	40	18	137	0.33	0.0105	15.1728	18.9486	0.9779
L282	23	23	1288	56	4.35	46	11	56	0.19	0.0076	15.1728	9.7589	-1.4022

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L290	23	23	1348	56	4.15	43	13	55	0.18	0.0083	15.1728	11.2292	-1.0214
L59	23	23	1349	56	4.15	44	12	59	0.19	0.0068	15.1728	9.1818	-1.5516
L86	23	23	1651	56	3.39	42	15	43	0.11	0.0055	15.1728	9.0553	-1.5844
L250	23	23	1171	55	4.70	39	16	94	0.35	0.0086	14.9019	10.0791	-1.2709
L311	23	23	1802	55	3.05	29	25	178	0.43	0.0046	14.9019	8.3399	-1.7292
L319	23	23	1554	55	3.54	43	12	115	0.32	0.0075	14.9019	11.6206	-0.8647
L58	23	23	1682	55	3.27	41	15	58	0.15	0.0056	14.9019	9.4427	-1.4386
L88	23	23	1220	55	4.51	48	10	116	0.41	0.0102	14.9019	12.4743	-0.6397
L201	21	21	1507	53	3.52	44	11	95	0.30	0.0085	14.7315	12.8238	-0.5170
L251	23	23	1366	54	3.95	43	11	107	0.34	0.0075	14.6309	10.2925	-1.1636
L101	23	23	1441	53	3.68	47	6	86	0.26	0.0063	14.3600	9.0711	-1.4441
L128	23	23	1720	53	3.08	44	10	22	0.06	0.0086	14.3600	14.7589	0.1089
L274	23	23	1450	53	3.66	45	8	90	0.27	0.0072	14.3600	10.4387	-1.0707
L407	23	23	2011	53	2.64	42	13	150	0.32	0.0066	14.3600	13.2727	-0.2969
L195	23	23	1522	52	3.42	43	9	74	0.21	0.0052	14.0890	7.9802	-1.6988
L291	23	23	1791	52	2.90	36	17	70	0.17	0.0053	14.0890	9.5534	-1.2613
L390	23	23	1727	52	3.01	42	10	129	0.32	0.0049	14.0890	8.4585	-1.5658
L417	23	23	1723	51	2.96	43	8	54	0.14	0.0042	13.8181	7.2727	-1.8543
L54	23	23	1613	51	3.16	46	5	217	0.58	0.0049	13.8181	7.9289	-1.6685
L110	22	22	1575	50	3.17	36	16	22	0.06	0.0051	13.7161	8.0476	-1.6307
L36	23	23	1686	50	2.97	40	11	59	0.15	0.0046	13.5472	7.7154	-1.6838
L367	23	23	1454	50	3.44	37	13	76	0.23	0.0074	13.5472	10.7431	-0.8096
L147	22	22	1649	49	2.97	42	7	44	0.12	0.0039	13.4418	6.4892	-2.0392
L238	23	23	1747	49	2.80	38	11	39	0.10	0.0041	13.2762	7.2490	-1.7742
L356	23	23	1419	49	3.45	40	11	87	0.27	0.0086	13.2762	12.2174	-0.3117
L5	23	23	1162	49	4.22	29	20	28	0.10	0.0065	13.2762	7.5613	-1.6822
L276	23	23	1458	48	3.29	40	9	86	0.26	0.0092	13.0053	13.4071	0.1206
L329	23	23	1505	48	3.19	38	10	91	0.26	0.0091	13.0053	13.7075	0.2108
L347	23	23	1701	48	2.82	44	4	150	0.38	0.0055	13.0053	9.3043	-1.1111

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L11	23	23	1691	47	2.78	38	10	36	0.09	0.0050	12.7343	8.4427	-1.3146
L166	23	23	1609	47	2.92	39	10	44	0.12	0.0064	12.7343	10.2688	-0.7552
L339	23	23	1733	47	2.71	39	9	28	0.07	0.0071	12.7343	12.2332	-0.1535
L342	23	23	1228	47	3.83	33	14	77	0.27	0.0073	12.7343	8.9051	-1.1729
L1	23	23	1648	46	2.79	36	10	77	0.20	0.0062	12.4634	10.2727	-0.6849
L224	23	23	1679	46	2.74	31	14	38	0.10	0.0035	12.4634	5.8696	-2.0616
L256	23	23	1634	46	2.82	41	4	100	0.27	0.0053	12.4634	8.6957	-1.1780
L283	23	23	1708	46	2.69	37	10	53	0.13	0.0041	12.4634	6.9723	-1.7168
L171	23	23	1705	45	2.64	27	18	21	0.05	0.0041	12.1924	7.0316	-1.6477
L285	20	20	691	43	6.22	32	11	54	0.39	0.0115	12.1204	7.9368	-1.3799
L163	23	23	1759	44	2.50	36	9	76	0.19	0.0048	11.9215	8.3874	-1.1528
L308	23	23	1679	44	2.62	32	12	47	0.12	0.0051	11.9215	8.5929	-1.0857
L348	23	23	1665	44	2.64	28	17	32	0.08	0.0040	11.9215	6.7312	-1.6930
L413	23	23	2211	44	1.99	27	17	17	0.03	0.0026	11.9215	5.6680	-2.0398
L69	23	23	1508	44	2.92	41	4	95	0.27	0.0053	11.9215	7.9960	-1.2804
L10	23	23	1705	43	2.52	38	6	38	0.10	0.0044	11.6505	7.4743	-1.3923
L108	23	23	1622	43	2.65	33	10	49	0.13	0.0027	11.6505	4.4506	-2.4004
L218	23	23	1611	43	2.67	40	3	81	0.22	0.0048	11.6505	7.7945	-1.2856
L312	23	23	1276	43	3.37	33	11	127	0.43	0.0063	11.6505	8.0158	-1.2118
L401	23	23	1585	43	2.71	33	10	22	0.06	0.0075	11.6505	11.9605	0.1033
L57	23	23	1764	43	2.44	36	8	115	0.28	0.0053	11.6505	9.3439	-0.7690
L92	23	23	1589	43	2.71	37	6	45	0.12	0.0041	11.6505	6.5415	-1.7033
L346	20	20	767	41	5.35	33	9	25	0.16	0.0135	11.5567	10.3632	-0.4119
L131	23	23	863	42	4.87	26	16	50	0.25	0.0096	11.3796	8.2964	-1.0511
L242	23	23	1797	42	2.34	34	9	119	0.29	0.0037	11.3796	6.6917	-1.5982
L269	23	23	599	42	7.01	31	12	14	0.10	0.0130	11.3796	7.7747	-1.2290
L280	23	23	1517	42	2.77	34	8	53	0.15	0.0052	11.3796	7.8379	-1.2074
L289	23	23	1709	42	2.46	29	12	97	0.25	0.0049	11.3796	8.4229	-1.0080
L297	23	23	1819	42	2.31	27	16	20	0.05	0.0041	11.3796	7.4625	-1.3355

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L307	23	23	1357	42	3.10	33	10	30	0.10	0.0049	11.3796	6.5968	-1.6306
L384	23	23	1672	42	2.51	29	14	42	0.11	0.0041	11.3796	6.7826	-1.5672
L77	23	23	1731	42	2.43	35	7	49	0.12	0.0052	11.3796	8.9802	-0.8180
L90	23	23	1031	42	4.07	23	19	106	0.45	0.0060	11.3796	6.1581	-1.7802
L226	21	21	929	40	4.31	30	11	19	0.10	0.0074	11.1181	6.8429	-1.5161
L181	23	23	1745	41	2.35	32	9	26	0.06	0.0039	11.1087	6.7312	-1.5269
L222	23	23	1940	41	2.11	29	12	163	0.37	0.0023	11.1087	4.5415	-2.2907
L234	23	23	1545	41	2.65	27	15	32	0.09	0.0036	11.1087	5.5375	-1.9433
L327	23	23	1715	41	2.39	28	15	38	0.10	0.0044	11.1087	7.5771	-1.2319
L343	23	23	1605	41	2.55	34	7	66	0.18	0.0041	11.1087	6.5731	-1.5821
L350	23	23	1744	41	2.35	26	15	75	0.19	0.0043	11.1087	7.5613	-1.2374
L104	23	23	1734	40	2.31	27	13	32	0.08	0.0046	10.8377	7.9684	-1.0246
L2	23	23	1728	40	2.31	35	5	34	0.09	0.0046	10.8377	7.9881	-1.0175
L219	23	23	1667	40	2.40	26	14	65	0.17	0.0039	10.8377	6.5652	-1.5256
L56	23	23	1572	40	2.54	25	14	61	0.17	0.0041	10.8377	6.4783	-1.5566
L111	22	22	1879	39	2.08	28	12	23	0.06	0.0032	10.6985	5.9221	-1.7410
L109	23	23	1793	39	2.18	31	8	21	0.05	0.0038	10.5668	6.7905	-1.3811
L126	23	23	1655	39	2.36	21	18	31	0.08	0.0045	10.5668	7.4664	-1.1339
L158	23	23	1742	39	2.24	30	8	67	0.17	0.0061	10.5668	10.6601	0.0341
L231	23	23	1380	39	2.83	34	5	70	0.22	0.0052	10.5668	7.1581	-1.2467
L383	23	23	1430	39	2.73	26	14	32	0.10	0.0042	10.5668	5.9644	-1.6832
L50	23	23	1598	39	2.44	34	6	23	0.06	0.0052	10.5668	8.3360	-0.8159
L74	23	23	1790	39	2.18	34	5	32	0.08	0.0033	10.5668	5.9289	-1.6962
L85	23	23	1696	39	2.30	29	11	66	0.17	0.0048	10.5668	8.1542	-0.8824
L41	21	21	1347	38	2.82	28	10	84	0.30	0.0044	10.5622	5.8714	-1.7463
L330	18	18	1152	36	3.13	27	8	99	0.48	0.0061	10.4665	7.0458	-1.3263
L221	23	23	1873	38	2.03	28	10	23	0.05	0.0040	10.2958	7.5771	-1.0191
L355	23	23	1378	38	2.76	29	9	35	0.11	0.0044	10.2958	6.0870	-1.5776
L371	23	23	1225	38	3.10	22	16	88	0.31	0.0064	10.2958	7.8814	-0.9050

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L6	21	21	1028	37	3.60	30	7	24	0.11	0.0075	10.2842	7.7095	-0.9830
L121	23	23	1703	37	2.17	27	10	37	0.09	0.0048	10.0249	8.2055	-0.6993
L328	23	23	1680	37	2.20	25	12	38	0.10	0.0036	10.0249	6.0277	-1.5364
L332	23	23	1977	37	1.87	29	8	61	0.13	0.0036	10.0249	7.0395	-1.1475
L35	23	23	1749	37	2.12	32	5	162	0.40	0.0054	10.0249	9.3953	-0.2420
L366	23	23	1303	37	2.84	27	10	96	0.32	0.0048	10.0249	6.1976	-1.4711
L192	22	22	1841	36	1.96	32	4	62	0.15	0.0035	9.8756	6.3680	-1.3790
L259	22	22	2267	36	1.59	23	13	41	0.08	0.0020	9.8756	4.5195	-2.1057
L301	22	22	1228	36	2.93	26	10	88	0.33	0.0049	9.8756	6.0563	-1.5015
L370	23	23	1834	36	1.96	22	14	35	0.08	0.0028	9.7539	5.1462	-1.8174
L380	23	23	1715	36	2.10	29	7	71	0.18	0.0028	9.7539	4.8814	-1.9219
L82	18	18	1484	33	2.22	28	7	41	0.15	0.0055	9.5943	8.1765	-0.5969
L138	23	23	1761	35	1.99	25	10	43	0.11	0.0060	9.4830	10.5889	0.4479
L174	23	23	1663	35	2.10	23	12	71	0.19	0.0042	9.4830	7.0119	-1.0009
L196	23	23	1705	35	2.05	24	11	24	0.06	0.0032	9.4830	5.4071	-1.6509
L286	23	23	1368	35	2.56	20	15	18	0.06	0.0056	9.4830	7.7194	-0.7143
L303	23	23	1434	35	2.44	22	14	135	0.41	0.0061	9.4830	8.7668	-0.2901
L306	23	23	1713	35	2.04	28	7	55	0.14	0.0033	9.4830	5.7273	-1.5212
L47	23	23	1666	35	2.10	30	5	33	0.09	0.0047	9.4830	7.7945	-0.6839
L83	23	23	971	35	3.60	24	11	26	0.12	0.0076	9.4830	7.3320	-0.8712
L43	22	22	1644	34	2.07	25	9	26	0.07	0.0027	9.3269	4.4892	-2.0071
L106	23	23	1990	34	1.71	23	11	28	0.06	0.0023	9.2121	4.5929	-1.9226
L399	23	23	1110	34	3.06	26	8	51	0.20	0.0064	9.2121	7.0553	-0.8977
L279	21	21	1009	33	3.27	28	5	58	0.27	0.0071	9.1724	7.1714	-0.8510
L78	21	21	1598	33	2.07	22	11	25	0.07	0.0045	9.1724	7.1524	-0.8591
L246	22	22	1338	33	2.47	5	28	61	0.21	0.0044	9.0526	5.9004	-1.3450
L323	22	22	553	33	5.97	21	11	19	0.16	0.0057	9.0526	3.1558	-2.5160
L134	23	23	1784	33	1.85	23	11	30	0.07	0.0022	8.9411	3.8617	-2.1742
L143	23	23	1893	33	1.74	27	6	57	0.13	0.0032	8.9411	6.0909	-1.2200

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L155	23	23	1285	33	2.57	24	9	53	0.18	0.0030	8.9411	3.9012	-2.1573
L223	23	23	1432	33	2.30	26	7	50	0.15	0.0032	8.9411	4.5652	-1.8731
L229	23	23	1861	33	1.77	23	10	14	0.03	0.0024	8.9411	4.3755	-1.9543
L240	23	23	1722	33	1.92	21	12	28	0.07	0.0033	8.9411	5.6522	-1.4078
L281	23	23	1729	33	1.91	30	3	42	0.11	0.0027	8.9411	4.7352	-1.8003
L331	23	23	1610	33	2.05	31	2	68	0.18	0.0036	8.9411	5.7312	-1.3740
L278	21	21	832	32	3.85	29	3	38	0.22	0.0072	8.8945	6.0000	-1.2671
L188	22	22	1853	32	1.73	18	15	9	0.02	0.0022	8.7783	4.1255	-2.0433
L22	22	22	1626	32	1.97	29	3	38	0.11	0.0022	8.7783	3.5238	-2.3076
L349	22	22	1604	32	2.00	24	8	58	0.16	0.0046	8.7783	7.3550	-0.6251
L103	23	23	1394	32	2.30	22	11	19	0.06	0.0045	8.6702	6.2095	-1.0841
L130	23	23	1335	32	2.40	26	7	24	0.08	0.0058	8.6702	7.7787	-0.3928
L217	23	23	1698	32	1.88	28	4	40	0.10	0.0030	8.6702	5.1344	-1.5577
L398	23	23	1653	32	1.94	25	8	67	0.18	0.0037	8.6702	6.0751	-1.1433
L412	23	23	1756	32	1.82	24	8	19	0.05	0.0032	8.6702	5.5613	-1.3696
L8*	20	20	835	30	3.59	24	7	39	0.23	0.0083	8.4561	6.9316	-0.7067
L125	23	23	1452	31	2.13	27	4	43	0.13	0.0053	8.3992	7.6324	-0.3480
L157	23	23	1763	31	1.76	22	9	85	0.21	0.0041	8.3992	7.2530	-0.5202
L338	23	23	1533	31	2.02	26	5	63	0.18	0.0052	8.3992	7.9051	-0.2242
L364	23	23	1665	31	1.86	23	8	41	0.11	0.0029	8.3992	4.7747	-1.6449
L64	23	23	1856	31	1.67	25	6	50	0.12	0.0030	8.3992	5.5534	-1.2915
L187	22	22	1834	30	1.64	19	11	33	0.08	0.0029	8.2296	5.2727	-1.3793
L227	23	23	1787	30	1.68	22	8	20	0.05	0.0020	8.1283	3.5059	-2.1629
L265	23	23	1413	30	2.12	23	7	22	0.07	0.0035	8.1283	4.9407	-1.4915
L302	23	23	1116	30	2.69	24	6	48	0.19	0.0039	8.1283	4.3557	-1.7653
L71	23	23	1532	30	1.96	22	8	36	0.10	0.0033	8.1283	5.0830	-1.4250
L76	23	23	1696	30	1.77	25	5	40	0.10	0.0036	8.1283	6.0711	-0.9626
L203	22	22	1452	29	2.00	21	8	84	0.26	0.0039	7.9553	5.6494	-1.1102
L137	23	23	1690	29	1.72	21	8	15	0.04	0.0035	7.8573	5.8419	-0.9733

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L165	23	23	1803	29	1.61	25	4	53	0.13	0.0031	7.8573	5.5375	-1.1203
L19	23	23	1543	29	1.88	18	11	52	0.15	0.0028	7.8573	4.2846	-1.7254
L194	23	23	1414	29	2.05	22	7	50	0.15	0.0044	7.8573	6.1897	-0.8053
L202	23	23	1499	29	1.93	17	12	61	0.18	0.0046	7.8573	6.9368	-0.4446
L263	23	23	1652	29	1.76	21	8	47	0.12	0.0020	7.8573	3.3715	-2.1663
L368	23	23	1711	29	1.69	21	8	8	0.02	0.0032	7.8573	5.4783	-1.1489
L404	23	23	1640	29	1.77	24	5	25	0.07	0.0033	7.8573	5.4585	-1.1585
L410	23	23	1755	29	1.65	22	7	98	0.24	0.0027	7.8573	4.7945	-1.4791
L44	23	23	1584	29	1.83	25	4	40	0.11	0.0028	7.8573	4.4506	-1.6452
L91	23	23	1451	29	2.00	17	12	20	0.06	0.0030	7.8573	4.2846	-1.7254
L175	21	21	875	28	3.20	22	7	75	0.41	0.0066	7.7827	5.7714	-0.9974
L295	22	22	854	28	3.28	19	9	46	0.24	0.0081	7.6810	6.9134	-0.3818
L53	22	22	1809	28	1.55	18	10	29	0.07	0.0028	7.6810	5.0606	-1.3035
L152	23	23	1955	28	1.43	21	8	20	0.04	0.0022	7.5864	4.2332	-1.6730
L159	23	23	1727	28	1.62	21	7	41	0.10	0.0023	7.5864	3.9723	-1.8031
L191	23	23	1471	28	1.90	22	6	15	0.04	0.0023	7.5864	3.3755	-2.1009
L216	23	23	2049	28	1.37	22	8	20	0.04	0.0022	7.5864	4.5968	-1.4916
L237	23	23	1761	28	1.59	27	1	46	0.11	0.0037	7.5864	6.4348	-0.5746
L369	23	23	1295	28	2.16	19	9	19	0.06	0.0026	7.5864	3.3399	-2.1187
L386	23	23	1574	28	1.78	21	7	74	0.20	0.0025	7.5864	3.9289	-1.8248
L394	23	23	1570	28	1.78	19	10	36	0.10	0.0037	7.5864	5.8024	-0.8901
L4	23	23	675	28	4.15	20	10	23	0.15	0.0072	7.5864	4.8854	-1.3476
L81	21	21	844	27	3.20	21	6	25	0.14	0.0047	7.5047	3.9810	-1.8076
L136	22	22	1098	27	2.46	16	11	31	0.13	0.0050	7.4067	5.5281	-0.9665
L381	22	22	2008	27	1.34	20	5	44	0.10	0.0020	7.4067	4.0303	-1.7372
L167	23	23	1690	27	1.60	20	7	44	0.11	0.0028	7.3155	4.6798	-1.3600
L220	23	23	1737	27	1.55	23	4	20	0.05	0.0034	7.3155	5.9763	-0.6910
L408	23	23	1384	27	1.95	20	8	23	0.07	0.0032	7.3155	4.4783	-1.4641
L45	23	23	1491	27	1.81	19	8	21	0.06	0.0033	7.3155	4.8498	-1.2723

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L156	22	22	759	26	3.43	21	5	29	0.17	0.0059	7.1324	4.5022	-1.4014
L178	22	22	1220	26	2.13	18	9	18	0.07	0.0041	7.1324	5.0476	-1.1107
L124	23	23	983	26	2.64	23	7	22	0.10	0.0046	7.0445	4.4783	-1.3713
L145	23	23	1284	26	2.02	16	10	79	0.27	0.0046	7.0445	5.8814	-0.6215
L160	23	23	1545	26	1.68	16	10	33	0.09	0.0029	7.0445	4.4348	-1.3945
L172	23	23	1993	26	1.30	17	9	10	0.02	0.0018	7.0445	3.5375	-1.8739
L176	23	23	1338	26	1.94	22	4	32	0.10	0.0027	7.0445	3.5692	-1.8570
L387	23	23	1783	26	1.46	25	3	47	0.11	0.0030	7.0445	5.4111	-0.8728
L79	23	23	815	26	3.19	21	5	41	0.22	0.0078	7.0445	6.3281	-0.3828
L142	22	22	1229	25	2.03	17	8	43	0.16	0.0027	6.8580	3.2814	-1.9759
L385	22	22	1786	25	1.40	16	9	37	0.09	0.0025	6.8580	4.5325	-1.2847
L99	22	22	911	25	2.74	16	9	14	0.07	0.0036	6.8580	3.2987	-1.9663
L115	23	23	1437	25	1.74	17	8	32	0.10	0.0028	6.7736	4.0553	-1.5059
L12	23	23	1145	25	2.18	17	8	10	0.04	0.0031	6.7736	3.5415	-1.7906
L135	23	23	1796	25	1.39	21	4	49	0.12	0.0023	6.7736	4.1897	-1.4315
L154	23	23	2051	25	1.22	20	5	42	0.09	0.0023	6.7736	4.7905	-1.0986
L199	23	23	1470	25	1.70	12	13	83	0.25	0.0031	6.7736	4.4901	-1.2651
L230	23	23	1670	25	1.50	17	8	27	0.07	0.0019	6.7736	3.2530	-1.9505
L254	23	23	1079	25	2.32	18	7	22	0.09	0.0046	6.7736	4.9328	-1.0198
L255	23	23	1027	25	2.43	20	6	77	0.33	0.0058	6.7736	5.9921	-0.4329
L266	23	23	1723	25	1.45	15	10	6	0.02	0.0021	6.7736	3.5810	-1.7687
L305	23	23	1652	25	1.51	21	4	35	0.09	0.0023	6.7736	3.7945	-1.6505
L336	23	23	1393	25	1.79	18	8	5	0.02	0.0028	6.7736	3.8656	-1.6110
L359	23	23	1771	25	1.41	17	8	52	0.13	0.0031	6.7736	5.4269	-0.7461
L55	23	23	1631	25	1.53	18	9	30	0.08	0.0036	6.7736	5.8814	-0.4943
L89	23	23	589	25	4.24	20	5	68	0.50	0.0102	6.7736	6.0040	-0.4264
L419	21	21	1137	24	2.11	18	6	28	0.12	0.0038	6.6709	4.2810	-1.3670
L304	22	22	879	24	2.73	18	7	32	0.17	0.0034	6.5837	3.0130	-2.0482
L173	23	23	1431	24	1.68	19	6	34	0.10	0.0032	6.5026	4.5336	-1.1326

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L309	23	23	1818	24	1.32	21	3	48	0.11	0.0028	6.5026	5.0237	-0.8507
L325	23	23	2189	24	1.10	16	8	33	0.07	0.0013	6.5026	2.9368	-2.0511
L360	23	23	1246	24	1.93	22	3	48	0.17	0.0033	6.5026	4.0988	-1.3827
L395	23	23	1590	24	1.51	14	10	32	0.09	0.0029	6.5026	4.5336	-1.1326
L51	23	23	1650	24	1.45	15	9	20	0.05	0.0025	6.5026	4.1107	-1.3758
L316	21	21	857	23	2.68	18	5	66	0.37	0.0053	6.3929	4.5667	-1.0863
L421	21	21	516	23	4.46	17	7	33	0.30	0.0075	6.3929	3.8476	-1.5139
L184	22	22	1705	23	1.35	17	6	43	0.11	0.0020	6.3094	3.3896	-1.7415
L197	22	22	1660	23	1.39	17	6	29	0.08	0.0023	6.3094	3.7489	-1.5272
L252	22	22	756	23	3.04	17	6	19	0.11	0.0086	6.3094	6.4805	0.1021
L382	22	22	1405	23	1.64	19	4	16	0.05	0.0025	6.3094	3.4978	-1.6769
L133	23	23	1655	23	1.39	18	5	26	0.07	0.0020	6.2317	3.2727	-1.7697
L164	23	23	1269	23	1.81	19	5	36	0.12	0.0035	6.2317	4.3874	-1.1030
L193	23	23	1769	23	1.30	14	9	23	0.06	0.0015	6.2317	2.7194	-2.1006
L200	23	23	1936	23	1.19	21	2	99	0.22	0.0020	6.2317	3.8735	-1.4103
L206	23	23	1214	23	1.89	14	9	12	0.04	0.0022	6.2317	2.6482	-2.1431
L241	23	23	1158	23	1.99	18	5	42	0.16	0.0035	6.2317	4.0672	-1.2945
L257	23	23	1528	23	1.51	18	4	31	0.09	0.0021	6.2317	3.2174	-1.8027
L294	23	23	696	23	3.30	10	15	14	0.09	0.0067	6.2317	4.6443	-0.9494
L31	23	23	1511	23	1.52	16	8	12	0.03	0.0019	6.2317	2.8221	-2.0391
L334	23	23	1776	23	1.30	18	5	13	0.03	0.0016	6.2317	2.8300	-2.0344
L37	23	23	1479	23	1.56	20	4	23	0.07	0.0019	6.2317	2.8103	-2.0462
L396	23	23	1812	23	1.27	20	3	53	0.13	0.0019	6.2317	3.4704	-1.6515
L80	23	23	1704	23	1.35	16	7	38	0.10	0.0023	6.2317	3.8379	-1.4316
L60	20	20	1378	22	1.60	12	11	34	0.12	0.0028	6.2011	3.8421	-1.4575
L144	21	21	627	22	3.51	13	10	13	0.10	0.0045	6.1150	2.8095	-2.0478
L320	21	21	576	22	3.82	20	2	19	0.16	0.0062	6.1150	3.5667	-1.5787
L105	23	23	1428	22	1.54	16	6	27	0.08	0.0025	5.9607	3.6087	-1.4650
L233	23	23	1837	22	1.20	17	5	38	0.09	0.0017	5.9607	3.1344	-1.7604

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L298	23	23	1758	22	1.25	13	9	22	0.05	0.0032	5.9607	5.5375	-0.2636
L30	23	23	1364	22	1.61	19	3	102	0.33	0.0021	5.9607	2.8577	-1.9327
L351	23	23	1718	22	1.28	15	7	50	0.13	0.0027	5.9607	4.6996	-0.7855
L353	23	23	1332	22	1.65	15	7	165	0.54	0.0031	5.9607	4.0711	-1.1769
L40	23	23	1649	22	1.33	13	9	13	0.03	0.0020	5.9607	3.3636	-1.6176
L411	23	23	1622	22	1.36	15	7	23	0.06	0.0016	5.9607	2.6680	-2.0509
L414	23	23	1496	22	1.47	18	4	58	0.17	0.0021	5.9607	3.2055	-1.7161
L112	20	20	512	21	4.10	16	6	58	0.57	0.0123	5.9193	6.3000	0.2455
L409	21	21	594	21	3.54	19	2	18	0.14	0.0055	5.8370	3.2476	-1.6738
L122	23	23	909	21	2.31	13	8	18	0.09	0.0038	5.6898	3.4625	-1.4473
L182	23	23	1373	21	1.53	18	3	15	0.05	0.0028	5.6898	3.8656	-1.1853
L25	23	23	955	21	2.20	17	4	33	0.15	0.0028	5.6898	2.6443	-1.9789
L260	23	23	2006	21	1.05	17	4	35	0.08	0.0022	5.6898	4.3874	-0.8463
L393	23	23	1718	21	1.22	15	6	32	0.08	0.0017	5.6898	2.8617	-1.8376
L403	23	23	1314	21	1.60	18	3	10	0.03	0.0016	5.6898	2.0435	-2.3693
L67	23	23	758	21	2.77	14	7	24	0.14	0.0049	5.6898	3.6917	-1.2983
L70	23	23	1523	21	1.38	15	6	29	0.08	0.0021	5.6898	3.1304	-1.6630
L189	22	22	1666	20	1.20	16	4	19	0.05	0.0018	5.4864	3.0390	-1.6580
L208	22	22	1238	20	1.62	14	6	32	0.12	0.0029	5.4864	3.5411	-1.3179
L402	22	22	1672	20	1.20	11	8	35	0.10	0.0019	5.4864	3.2078	-1.5437
L114	23	23	1512	20	1.32	15	5	44	0.13	0.0027	5.4189	4.1107	-0.8884
L13	23	23	1573	20	1.27	16	4	22	0.06	0.0018	5.4189	2.9051	-1.7072
L151	23	23	1704	20	1.17	13	7	28	0.07	0.0014	5.4189	2.3636	-2.0749
L153	23	23	1229	20	1.63	15	5	20	0.07	0.0025	5.4189	3.0119	-1.6347
L198	23	23	1155	20	1.73	17	5	24	0.09	0.0037	5.4189	4.3083	-0.7542
L271	23	23	1481	20	1.35	14	6	9	0.03	0.0019	5.4189	2.8340	-1.7555
L326	23	23	1244	20	1.61	15	5	23	0.08	0.0027	5.4189	3.3043	-1.4361
L361	23	23	1705	20	1.17	16	5	100	0.26	0.0026	5.4189	4.3478	-0.7274
L365	23	23	1738	20	1.15	16	4	104	0.26	0.0016	5.4189	2.8182	-1.7662

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L405	23	23	1396	20	1.43	17	3	10	0.03	0.0024	5.4189	3.2925	-1.4441
L65	23	23	715	20	2.80	13	7	13	0.08	0.0037	5.4189	2.6798	-1.8602
L66	23	23	1401	20	1.43	19	1	36	0.11	0.0018	5.4189	2.4704	-2.0025
L95	23	23	1711	20	1.17	14	6	23	0.06	0.0016	5.4189	2.7510	-1.8119
L97	23	23	979	20	2.04	11	9	25	0.11	0.0037	5.4189	3.6640	-1.1918
L324	20	20	379	19	5.01	12	7	6	0.08	0.0095	5.3555	3.6053	-1.2360
L18	21	21	564	19	3.37	15	4	26	0.22	0.0073	5.2811	4.1238	-0.8192
L186	22	22	1000	19	1.90	15	4	47	0.21	0.0032	5.2121	3.2251	-1.4100
L214	22	22	916	19	2.07	15	4	25	0.12	0.0034	5.2121	3.1039	-1.4960
L376	22	22	1082	19	1.76	19	0	29	0.12	0.0024	5.2121	2.6320	-1.8308
L102	23	23	1433	19	1.33	16	3	11	0.03	0.0016	5.1479	2.2253	-2.0789
L210	23	23	1838	19	1.03	12	7	18	0.04	0.0013	5.1479	2.4545	-1.9158
L239	23	23	1213	19	1.57	16	3	54	0.19	0.0030	5.1479	3.6364	-1.0752
L247	23	23	1829	19	1.04	13	6	15	0.04	0.0016	5.1479	2.8538	-1.6319
L273	23	23	1830	19	1.04	13	6	19	0.05	0.0022	5.1479	4.0435	-0.7856
L284	23	23	1770	19	1.07	14	5	66	0.16	0.0028	5.1479	4.8814	-0.1896
L315	23	23	1142	19	1.66	15	4	36	0.14	0.0038	5.1479	4.3123	-0.5944
L374	23	23	1665	19	1.14	15	5	17	0.04	0.0017	5.1479	2.8340	-1.6459
L392	23	23	648	19	2.93	18	2	45	0.30	0.0053	5.1479	3.4071	-1.2383
L61	23	23	1691	19	1.12	12	7	12	0.03	0.0012	5.1479	2.1107	-2.1604
L75	23	23	1694	19	1.12	17	2	40	0.10	0.0016	5.1479	2.6917	-1.7471
L314	20	20	537	18	3.35	18	2	40	0.37	0.0102	5.0737	5.4684	0.2927
L232	22	22	829	18	2.17	16	2	11	0.06	0.0038	4.9378	3.1126	-1.3598
L63	22	22	1647	18	1.09	11	7	39	0.11	0.0014	4.9378	2.3377	-1.9371
L117	23	23	1152	18	1.56	11	7	29	0.11	0.0026	4.8770	2.9881	-1.4104
L141	23	23	1342	18	1.34	16	2	17	0.06	0.0017	4.8770	2.3360	-1.8974
L179	23	23	1649	18	1.09	15	3	24	0.06	0.0012	4.8770	2.0000	-2.1483
L243	23	23	960	18	1.88	13	5	50	0.23	0.0046	4.8770	4.4545	-0.3154
L270	23	23	1274	18	1.41	13	5	21	0.07	0.0020	4.8770	2.5573	-1.7321

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L299	23	23	1534	18	1.17	15	3	22	0.06	0.0017	4.8770	2.5336	-1.7498
L362	23	23	996	18	1.81	14	4	87	0.38	0.0034	4.8770	3.3557	-1.1359
L375	23	23	1797	18	1.00	15	3	15	0.04	0.0021	4.8770	3.6838	-0.8910
L48	23	23	1732	18	1.04	12	6	17	0.04	0.0014	4.8770	2.4229	-1.8325
L62	23	23	1682	18	1.07	14	4	11	0.03	0.0018	4.8770	3.0198	-1.3868
L170	19	19	785	17	2.17	15	2	21	0.14	0.0043	4.8639	3.3684	-1.1649
L116	21	21	1714	17	0.99	11	6	15	0.04	0.0017	4.7252	2.8381	-1.4764
L32	21	21	767	17	2.22	14	3	17	0.11	0.0040	4.7252	3.0571	-1.3050
L93	22	22	1496	17	1.14	12	5	16	0.05	0.0016	4.6635	2.4589	-1.7286
L14	23	23	1689	17	1.01	7	10	23	0.06	0.0020	4.6060	3.3597	-0.9794
L183	23	23	1752	17	0.97	13	4	18	0.04	0.0017	4.6060	2.9802	-1.2776
L21	23	23	1787	17	0.95	11	6	24	0.06	0.0013	4.6060	2.3360	-1.7839
L228	23	23	1525	17	1.11	13	4	34	0.10	0.0023	4.6060	3.4862	-0.8800
L335	23	23	1100	17	1.55	11	6	18	0.07	0.0027	4.6060	2.9447	-1.3056
L39	23	23	1061	17	1.60	11	6	24	0.10	0.0017	4.6060	1.8261	-2.1846
L52	23	23	1766	17	0.96	15	2	18	0.04	0.0013	4.6060	2.2609	-1.8429
L148	20	20	1101	16	1.45	12	5	26	0.12	0.0036	4.5099	3.9421	-0.4679
L292	22	22	1746	16	0.92	13	3	14	0.04	0.0014	4.3891	2.5238	-1.5437
L107	23	23	548	16	2.92	16	0	26	0.21	0.0055	4.3351	3.0198	-1.0908
L113	23	23	1489	16	1.07	12	4	16	0.05	0.0022	4.3351	3.2767	-0.8778
L139	23	23	1989	16	0.80	14	2	18	0.04	0.0010	4.3351	2.0435	-1.9005
L204	23	23	901	16	1.78	12	4	15	0.07	0.0027	4.3351	2.4150	-1.5924
L293	23	23	1335	16	1.20	11	5	32	0.10	0.0034	4.3351	4.5929	0.2138
L33	23	23	1063	16	1.51	10	7	14	0.06	0.0029	4.3351	3.0632	-1.0548
L345	23	23	491	16	3.26	11	5	35	0.31	0.0081	4.3351	3.9802	-0.2943
L46	23	23	1727	16	0.93	11	5	13	0.03	0.0014	4.3351	2.4743	-1.5432
L127	22	22	1790	15	0.84	11	5	15	0.04	0.0016	4.1148	2.8398	-1.1172
L149	22	22	1471	15	1.02	9	6	35	0.11	0.0015	4.1148	2.1991	-1.6786
L236	22	22	875	15	1.71	7	8	44	0.23	0.0027	4.1148	2.3810	-1.5193

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L68	22	22	1738	15	0.86	11	4	16	0.04	0.0016	4.1148	2.8225	-1.1324
L118	23	23	974	15	1.54	13	2	44	0.20	0.0037	4.0641	3.6364	-0.3756
L120	23	23	1240	15	1.21	11	4	18	0.06	0.0011	4.0641	1.3439	-2.3883
L207	23	23	895	15	1.68	14	1	34	0.17	0.0023	4.0641	2.0870	-1.7359
L373	23	23	1679	15	0.89	13	2	36	0.09	0.0013	4.0641	2.1660	-1.6665
L406	23	23	1700	15	0.88	10	5	20	0.05	0.0019	4.0641	3.2332	-0.7295
L211	20	20	1846	14	0.76	11	3	15	0.04	0.0012	3.9462	2.2895	-1.5368
L377	20	20	1605	14	0.87	13	1	27	0.08	0.0013	3.9462	2.0316	-1.7761
L300	22	22	1194	14	1.17	9	5	20	0.08	0.0019	3.8405	2.3247	-1.4113
L123	23	23	1334	14	1.05	8	6	26	0.08	0.0020	3.7932	2.6522	-1.0643
L132	23	23	1337	14	1.05	9	5	17	0.06	0.0010	3.7932	1.3913	-2.2403
L140	23	23	1273	14	1.10	12	2	54	0.18	0.0016	3.7932	1.9763	-1.6947
L180	23	23	726	14	1.93	11	3	5	0.03	0.0028	3.7932	2.0593	-1.6173
L213	23	23	1765	14	0.79	13	1	11	0.03	0.0010	3.7932	1.7312	-1.9233
L249	23	23	1583	14	0.88	11	3	11	0.03	0.0012	3.7932	1.9447	-1.7242
L268	23	23	1724	14	0.81	10	4	47	0.12	0.0013	3.7932	2.2609	-1.4293
L352	23	23	1405	14	1.00	10	4	90	0.28	0.0019	3.7932	2.7352	-0.9869
L397	23	23	1799	14	0.78	11	3	18	0.04	0.0011	3.7932	1.9763	-1.6947
L42	23	23	1665	14	0.84	12	2	17	0.04	0.0012	3.7932	1.9447	-1.7242
L49	23	23	1287	14	1.09	7	7	10	0.03	0.0015	3.7932	1.9881	-1.6836
L73	23	23	1697	14	0.82	9	5	17	0.04	0.0014	3.7932	2.3518	-1.3445
L96	23	23	1597	14	0.88	9	5	34	0.09	0.0013	3.7932	2.0237	-1.6505
L177*	20	20	936	13	1.39	11	2	12	0.06	0.0028	3.6643	2.5842	-1.0692
L253	20	20	540	13	2.41	10	3	6	0.06	0.0048	3.6643	2.5789	-1.0744
L388	21	21	701	13	1.85	9	4	23	0.16	0.0032	3.6134	2.2381	-1.3637
L209	22	22	1728	13	0.75	11	2	23	0.06	0.0014	3.5662	2.4372	-1.1213
L15	23	23	1214	13	1.07	12	1	12	0.04	0.0014	3.5223	1.6443	-1.8684
L161	23	23	1209	13	1.08	7	6	38	0.14	0.0028	3.5223	3.4229	-0.0988
L185	23	23	1289	13	1.01	8	5	14	0.05	0.0014	3.5223	1.7747	-1.7386

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L23	23	23	1790	13	0.73	12	1	39	0.09	0.0009	3.5223	1.5455	-1.9667
L244	23	23	1782	13	0.73	11	2	25	0.06	0.0010	3.5223	1.8459	-1.6678
L296	23	23	1044	13	1.25	8	5	31	0.13	0.0031	3.5223	3.2134	-0.3072
L322	23	23	1780	13	0.73	11	2	17	0.04	0.0009	3.5223	1.5375	-1.9745
L337	23	23	842	13	1.54	9	4	15	0.08	0.0021	3.5223	1.7352	-1.7779
L363	23	23	905	13	1.44	8	5	9	0.04	0.0016	3.5223	1.4743	-2.0374
L372	23	23	1191	13	1.09	10	3	17	0.06	0.0015	3.5223	1.7510	-1.7622
L379	23	23	1705	13	0.76	12	1	48	0.12	0.0020	3.5223	3.4585	-0.0634
L16	21	21	1785	12	0.67	10	2	9	0.02	0.0007	3.3354	1.2048	-2.2645
L267	22	22	1803	12	0.67	8	4	11	0.03	0.0010	3.2919	1.8485	-1.5363
L162	23	23	1686	12	0.71	9	3	25	0.06	0.0012	3.2513	2.0000	-1.3339
L17	23	23	1712	12	0.70	10	2	21	0.05	0.0010	3.2513	1.6364	-1.7215
L190	23	23	1381	12	0.87	12	0	20	0.06	0.0018	3.2513	2.4387	-0.8662
L258	23	23	1052	12	1.14	10	2	10	0.04	0.0014	3.2513	1.4506	-1.9196
L418	23	23	1551	12	0.77	9	3	23	0.06	0.0019	3.2513	2.9289	-0.3437
L72	23	23	830	12	1.45	9	3	42	0.22	0.0019	3.2513	1.6166	-1.7426
L150	23	23	1637	11	0.67	6	5	21	0.06	0.0010	2.9804	1.6838	-1.4888
L215	23	23	1732	11	0.64	9	2	17	0.04	0.0010	2.9804	1.7470	-1.4162
L27	23	23	1730	11	0.64	8	3	25	0.06	0.0008	2.9804	1.3992	-1.8156
L318	23	23	1534	11	0.72	9	2	18	0.05	0.0016	2.9804	2.4387	-0.6219
L245	21	21	831	10	1.20	7	3	24	0.14	0.0021	2.7795	1.7143	-1.3228
L288	23	23	765	10	1.31	7	3	35	0.20	0.0025	2.7094	1.9407	-0.9567
L416	23	23	1696	10	0.59	5	5	8	0.02	0.0012	2.7094	1.9526	-0.9419
L205*	21	21	549	9	1.64	7	3	6	0.05	0.0020	2.5016	1.1048	-1.8948
L333	23	23	821	9	1.10	8	1	9	0.05	0.0016	2.4385	1.3083	-1.5356
L389	21	21	927	8	0.86	8	0	19	0.10	0.0013	2.2236	1.2048	-1.5234
L415	22	22	1377	8	0.58	4	4	12	0.04	0.0007	2.1946	0.9567	-1.8517
L264	23	23	651	8	1.23	2	5	17	0.11	0.0013	2.1675	0.8459	-1.9782
L20*	21	21	792	7	0.88	6	1	379	2.28	0.0029	1.9457	2.2762	0.5508

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L26	21	21	742	7	0.94	4	3	3	0.02	0.0017	1.9457	1.2714	-1.1235
L420	21	21	493	7	1.42	3	4	23	0.22	0.0021	1.9457	1.0476	-1.4964
L119	22	22	652	7	1.07	7	1	9	0.06	0.0014	1.9203	0.9394	-1.6346
L38	23	23	705	7	0.99	2	5	19	0.12	0.0015	1.8966	1.0395	-1.4286
L378	20	20	478	6	1.26	3	3	58	0.61	0.0020	1.6912	0.9789	-1.3422
L98	21	21	478	6	1.26	3	3	2	0.02	0.0016	1.6677	0.7429	-1.7419
L212*	23	23	438	6	1.37	6	0	288	2.86	0.0027	1.6257	1.1818	-0.8355
L261	22	22	1772	5	0.28	2	3	17	0.04	0.0004	1.3716	0.6753	-1.5088
L262	22	22	1224	5	0.41	3	2	15	0.06	0.0005	1.3716	0.6450	-1.5745
L272	23	23	983	5	0.51	5	0	6	0.03	0.0015	1.3547	1.4862	0.2847
L29	20	20	953	4	0.42	2	2	7	0.04	0.0005	1.1275	0.4737	-1.6787
L317	23	23	1629	4	0.25	1	3	53	0.14	0.0003	1.0838	0.4071	-1.7296
L24	22	22	483	3	0.62	3	1	6	0.06	0.0023	0.8230	1.1039	0.8830
Mean (415 loci)	**22.54**	**22.54**	**1427.66**	**29.01**	**2.12**	**21.87**	**7.38**	**45.22**	**0.15**	**0.0038**	**7.8933**	**5.2406**	**-1.3137**
SD	**0.84**	**0.84**	**389.50**	**14.86**	**1.10**	**12.13**	**4.64**	**40.48**	**0.20**	**0.0024**	**4.0321**	**3.3055**	**0.5795**

Appendix 22: *Micrablepharus atticolus* loci summary statistics from the final Anchored Phylogenomics dataset used on BPP and phylogenetic analyses. The ones marked with an asterisk (*) were not used in the phylogenetic analyses because the outgroup specimen was not captured. Loci are ordered from highest to lowest molecular diversity (Watterson's θ). Outgroups and cryptic species outside of the monophyletic *M. atticolus* group were not included in the calculations.

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L74	25	25	1721	59	3.43	37	22	206	0.48	0.0047	15.6252	8.1433	-1.8561
L214	25	25	1799	57	3.17	38	19	130	0.29	0.0037	15.0955	6.5867	-2.1821
L171	26	26	1763	57	3.23	36	21	164	0.36	0.0052	14.9373	9.1138	-1.4985
L216	26	26	1811	54	2.98	36	18	118	0.25	0.0041	14.1511	7.4431	-1.8180
L182	26	26	1803	53	2.94	46	7	115	0.25	0.0038	13.8890	6.9015	-1.9279
L211	25	25	1399	52	3.72	40	12	107	0.31	0.0050	13.7713	6.9367	-1.9140
L295	25	25	1737	51	2.94	31	21	121	0.28	0.0036	13.5065	6.1700	-2.0930
L315	25	25	1668	50	3.00	39	11	75	0.18	0.0044	13.2417	7.3967	-1.6994
L309	26	26	2245	49	2.18	37	12	102	0.17	0.0026	12.8408	5.8831	-2.0691
L164	25	25	1602	47	2.93	36	12	110	0.27	0.0038	12.4472	6.1433	-1.9442
L266	25	25	1024	47	4.59	39	8	130	0.51	0.0057	12.4472	5.7967	-2.0511
L3	25	25	1744	47	2.69	39	9	97	0.22	0.0031	12.4472	5.3533	-2.1878
L326	25	25	1612	47	2.92	35	12	170	0.42	0.0044	12.4472	7.1133	-1.6450
L289	25	25	1590	46	2.89	40	6	205	0.52	0.0040	12.1823	6.4367	-1.8087
L328	25	25	1650	46	2.79	36	11	165	0.40	0.0034	12.1823	5.5433	-2.0899
L292	26	26	1707	45	2.64	26	19	85	0.19	0.0041	11.7926	6.9538	-1.5604
L339	26	26	1878	44	2.34	34	10	83	0.17	0.0058	11.5305	10.9046	-0.2062
L376	24	24	1789	43	2.40	33	11	120	0.28	0.0035	11.5149	6.1775	-1.7855
L247	25	25	2279	43	1.89	35	8	78	0.14	0.0030	11.3878	6.7700	-1.5499
L55	25	25	1410	43	3.05	32	11	106	0.30	0.0034	11.3878	4.8367	-2.1988
L268	26	26	1797	43	2.39	30	14	126	0.27	0.0028	11.2685	5.0215	-2.1034
L9	24	24	1591	42	2.64	21	21	14	0.04	0.0038	11.2471	6.0797	-1.7676
L176	23	23	1859	41	2.21	25	16	186	0.44	0.0038	11.1087	7.0237	-1.4249
L207	26	26	1624	42	2.59	28	14	84	0.20	0.0035	11.0064	5.7323	-1.8159

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L34	26	26	1430	42	2.94	27	15	110	0.30	0.0052	11.0064	7.4800	-1.2141
L271	25	25	1243	41	3.30	26	15	117	0.38	0.0056	10.8582	6.9867	-1.3594
L197	23	23	1395	40	2.87	27	14	74	0.23	0.0048	10.8377	6.7115	-1.4734
L288	25	25	1028	40	3.89	30	11	115	0.45	0.0058	10.5933	5.9200	-1.6797
L370	25	25	1748	40	2.29	35	6	94	0.22	0.0033	10.5933	5.7033	-1.7576
L210	24	24	2053	39	1.90	35	4	139	0.28	0.0022	10.4437	4.4384	-2.2036
L68	24	24	1685	39	2.31	27	13	88	0.22	0.0040	10.4437	6.7572	-1.3527
L265	26	26	1918	39	2.03	28	11	186	0.37	0.0041	10.2202	7.9200	-0.8494
L1	26	26	1828	38	2.08	30	8	152	0.32	0.0025	9.9582	4.4862	-2.0707
L319	26	26	1772	38	2.14	24	14	92	0.20	0.0035	9.9582	6.2308	-1.4105
L36	26	26	1626	38	2.34	23	16	94	0.22	0.0027	9.9582	4.3323	-2.1290
L80	26	26	1609	38	2.36	26	12	172	0.41	0.0038	9.9582	6.1538	-1.4396
L94	26	26	1812	38	2.10	26	14	81	0.17	0.0030	9.9582	5.4800	-1.6946
L112	25	25	1033	37	3.58	25	12	111	0.43	0.0061	9.7988	6.2867	-1.3587
L227	25	25	1668	37	2.22	34	3	236	0.57	0.0045	9.7988	7.5533	-0.8687
L267	25	25	1673	37	2.21	30	8	140	0.33	0.0032	9.7988	5.4100	-1.6978
L150	21	21	1420	35	2.46	22	13	52	0.17	0.0033	9.7283	4.6714	-2.0347
L17	26	26	1440	37	2.57	30	7	104	0.28	0.0041	9.6961	5.8800	-1.4808
L336	26	26	1728	37	2.14	30	10	125	0.28	0.0030	9.6961	5.1508	-1.7638
L114	24	24	1555	36	2.32	30	6	149	0.40	0.0033	9.6404	5.1667	-1.7703
L119	24	24	809	36	4.45	25	11	102	0.53	0.0099	9.6404	8.0362	-0.6348
L264	24	24	1554	36	2.32	30	6	69	0.19	0.0027	9.6404	4.1413	-2.1760
L212	25	25	1664	36	2.16	27	9	56	0.13	0.0034	9.5340	5.6800	-1.5298
L27	25	25	1729	36	2.08	30	7	136	0.31	0.0031	9.5340	5.3400	-1.6648
L404	25	25	1668	36	2.16	34	2	112	0.27	0.0044	9.5340	7.2967	-0.8881
L408	25	25	1575	36	2.29	27	9	135	0.34	0.0045	9.5340	7.1433	-0.9490
L123	26	26	1775	36	2.03	25	12	88	0.19	0.0027	9.4341	4.8154	-1.8390
L209	24	24	1875	35	1.87	25	10	92	0.20	0.0025	9.3726	4.6630	-1.9136
L343	24	24	2114	35	1.66	24	11	99	0.20	0.0027	9.3726	5.7645	-1.4661

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L226	25	25	1798	35	1.95	30	5	112	0.25	0.0029	9.2692	5.1867	-1.6640
L298	25	25	1175	35	2.98	24	11	56	0.19	0.0035	9.2692	4.1067	-2.1042
L303	25	25	1554	35	2.25	26	9	130	0.33	0.0026	9.2692	4.1100	-2.1028
L333	25	25	1604	35	2.18	26	9	95	0.24	0.0031	9.2692	4.9300	-1.7686
L342	25	25	1667	35	2.10	34	1	139	0.33	0.0039	9.2692	6.5100	-1.1246
L12	26	26	1698	35	2.06	21	14	81	0.18	0.0021	9.1720	3.5662	-2.2918
L6	26	26	1414	35	2.48	30	7	36	0.10	0.0035	9.1720	4.9938	-1.7081
L308	24	24	378	34	8.99	24	11	84	0.93	0.0115	9.1048	4.3333	-1.9923
L329	22	22	1468	33	2.25	18	15	115	0.36	0.0037	9.0526	5.3853	-1.5648
L300	25	25	1161	34	2.93	23	11	69	0.24	0.0031	9.0043	3.6067	-2.2606
L18	26	26	1808	34	1.88	26	8	86	0.18	0.0021	8.9100	3.8431	-2.1285
L337	26	26	1417	34	2.40	25	10	151	0.41	0.0034	8.9100	4.8277	-1.7149
L394	26	26	1951	34	1.74	26	9	117	0.23	0.0029	8.9100	5.5692	-1.4034
L323	24	24	1241	33	2.66	30	4	183	0.61	0.0051	8.8370	6.2754	-1.0999
L341	24	24	1561	33	2.11	27	6	138	0.37	0.0037	8.8370	5.7572	-1.3224
L234	25	25	1058	33	3.12	19	14	101	0.38	0.0042	8.7395	4.3967	-1.8704
L359	25	25	1701	33	1.94	26	7	140	0.33	0.0025	8.7395	4.3000	-1.9120
L387	25	25	1232	33	2.68	30	3	49	0.16	0.0028	8.7395	3.4933	-2.2595
L126	26	26	1799	33	1.83	26	8	91	0.19	0.0023	8.6479	4.0831	-1.9718
L147	26	26	1783	33	1.85	26	7	118	0.25	0.0020	8.6479	3.5015	-2.2231
L153*	26	26	1734	33	1.90	26	8	133	0.30	0.0026	8.6479	4.4369	-1.8190
L162	26	26	2087	33	1.58	26	7	73	0.13	0.0025	8.6479	5.2308	-1.4761
L28	26	26	1651	33	2.00	22	11	136	0.32	0.0031	8.6479	5.1169	-1.5253
L332	26	26	1785	33	1.85	23	10	145	0.31	0.0030	8.6479	5.4092	-1.3990
L390	26	26	1272	33	2.59	19	14	38	0.11	0.0038	8.6479	4.8215	-1.6529
L117	24	24	1363	32	2.35	18	14	88	0.27	0.0032	8.5692	4.3080	-1.8831
L42	24	24	1697	32	1.89	14	18	179	0.44	0.0021	8.5692	3.5797	-2.2050
L398	25	25	1663	32	1.92	29	4	88	0.21	0.0028	8.4747	4.5967	-1.7189
L109	26	26	1724	32	1.86	23	9	133	0.30	0.0028	8.3858	4.7908	-1.5982

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L168	26	26	1238	32	2.58	22	10	100	0.31	0.0039	8.3858	4.8154	-1.5872
L95	26	26	1470	32	2.18	25	7	67	0.18	0.0039	8.3858	5.6800	-1.2029
L148	25	25	1647	31	1.88	20	11	100	0.24	0.0027	8.2098	4.4567	-1.7136
L165	25	25	1725	31	1.80	22	9	86	0.20	0.0028	8.2098	4.8833	-1.5188
L321	25	25	1201	31	2.58	28	3	85	0.28	0.0032	8.2098	3.8800	-1.9769
L371	25	25	1573	31	1.97	23	8	157	0.40	0.0024	8.2098	3.8400	-1.9951
L240	26	26	1496	31	2.07	22	9	101	0.26	0.0024	8.1238	3.5600	-2.0897
L299	26	26	1938	31	1.60	22	10	127	0.25	0.0027	8.1238	5.3015	-1.2923
L96	26	26	1952	31	1.59	21	10	123	0.24	0.0020	8.1238	3.8954	-1.9362
L358	24	24	1243	30	2.41	23	7	101	0.34	0.0032	8.0337	3.9348	-1.9237
L203	25	25	1794	30	1.67	26	4	63	0.14	0.0031	7.9450	5.6333	-1.0881
L71	25	25	944	30	3.18	21	9	73	0.31	0.0054	7.9450	5.1433	-1.3188
L200	26	26	1969	30	1.52	24	7	85	0.17	0.0031	7.8617	6.1200	-0.8222
L254	26	26	1785	30	1.68	20	10	73	0.16	0.0024	7.8617	4.3446	-1.6603
L373	26	26	1706	30	1.76	28	2	109	0.25	0.0023	7.8617	3.9692	-1.8375
L400	26	26	2099	30	1.43	25	5	126	0.23	0.0025	7.8617	5.2615	-1.2275
L82	26	26	1415	30	2.12	22	8	97	0.26	0.0038	7.8617	5.4062	-1.1592
L33	24	24	1622	29	1.79	20	9	132	0.34	0.0019	7.7659	3.0942	-2.2628
L215	25	25	1448	29	2.00	25	4	73	0.20	0.0033	7.6802	4.8467	-1.3764
L224	25	25	1788	29	1.62	21	7	131	0.29	0.0030	7.6802	5.2933	-1.1594
L239	25	25	1639	29	1.77	18	11	179	0.44	0.0033	7.6802	5.4233	-1.0963
L245	25	25	1644	29	1.76	17	12	73	0.18	0.0027	7.6802	4.3733	-1.6063
L327	25	25	1305	29	2.22	19	10	121	0.37	0.0051	7.6802	6.6167	-0.5166
L382	25	25	1808	29	1.60	20	9	51	0.11	0.0019	7.6802	3.4467	-2.0565
L395	25	25	1257	29	2.31	21	8	91	0.29	0.0042	7.6802	5.2300	-1.1902
L57	25	25	1680	29	1.73	26	3	83	0.20	0.0026	7.6802	4.4133	-1.5869
L103	26	26	1525	29	1.90	21	9	93	0.23	0.0028	7.5997	4.2462	-1.6336
L131	26	26	1968	29	1.47	16	13	119	0.23	0.0025	7.5997	4.8738	-1.3279
L189	26	26	1595	29	1.82	23	7	133	0.32	0.0024	7.5997	3.7569	-1.8720

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L201	26	26	1842	29	1.57	24	6	58	0.12	0.0020	7.5997	3.7015	-1.8989
L218	26	26	2010	29	1.44	19	10	125	0.24	0.0021	7.5997	4.3169	-1.5992
L22	26	26	1855	29	1.56	28	1	34	0.07	0.0017	7.5997	3.2000	-2.1433
L273	26	26	1743	29	1.66	20	9	52	0.11	0.0026	7.5997	4.6154	-1.4538
L320	23	23	1095	28	2.56	24	4	72	0.29	0.0044	7.5864	4.7668	-1.4068
L155	25	25	1728	28	1.62	20	8	85	0.20	0.0023	7.4153	4.0500	-1.6888
L63	25	25	1649	28	1.70	22	7	61	0.15	0.0025	7.4153	4.0567	-1.6854
L149	26	26	1708	28	1.64	20	8	69	0.16	0.0019	7.3376	3.2923	-2.0357
L163	26	26	1602	28	1.75	22	7	72	0.17	0.0026	7.3376	4.1662	-1.5959
L178	26	26	1597	28	1.75	19	9	97	0.23	0.0020	7.3376	3.1446	-2.1100
L384	26	26	1796	28	1.56	18	10	106	0.23	0.0019	7.3376	3.4923	-1.9350
L391	26	26	1730	28	1.62	20	8	90	0.20	0.0020	7.3376	3.4215	-1.9706
L85	26	26	1633	28	1.71	22	6	48	0.11	0.0024	7.3376	3.9262	-1.7167
L89	26	26	1276	28	2.19	20	9	89	0.27	0.0026	7.3376	3.3692	-1.9970
L136	24	24	1037	27	2.60	21	6	101	0.41	0.0037	7.2303	3.8804	-1.7336
L45	24	24	1635	27	1.65	22	5	131	0.33	0.0024	7.2303	3.8623	-1.7430
L11	25	25	1635	27	1.65	15	12	125	0.31	0.0024	7.1505	4.0000	-1.6350
L237	25	25	2157	27	1.25	22	5	65	0.12	0.0014	7.1505	3.0067	-2.1505
L278	25	25	1269	27	2.13	19	8	96	0.30	0.0033	7.1505	4.2500	-1.5053
L330	25	25	1436	27	1.88	15	12	115	0.32	0.0029	7.1505	4.1967	-1.5330
L403	25	25	1769	27	1.53	18	9	59	0.13	0.0022	7.1505	3.8667	-1.7042
L5	25	25	1619	27	1.67	19	8	43	0.11	0.0036	7.1505	5.8967	-0.6507
L169	26	26	1698	27	1.59	19	9	117	0.27	0.0028	7.0755	4.8154	-1.1762
L231	26	26	1831	27	1.47	21	7	130	0.27	0.0019	7.0755	3.4523	-1.8855
L379	26	26	1770	27	1.53	17	10	107	0.23	0.0017	7.0755	3.0800	-2.0793
L73	26	26	1762	27	1.53	21	6	97	0.21	0.0026	7.0755	4.6092	-1.2835
L2	24	24	1625	26	1.60	20	6	138	0.35	0.0042	6.9625	6.8116	-0.0809
L23	24	24	1730	26	1.50	20	8	84	0.20	0.0025	6.9625	4.3913	-1.3778
L90	24	24	1564	26	1.66	18	8	71	0.19	0.0037	6.9625	5.8370	-0.6031

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L183	25	25	1699	26	1.53	18	8	89	0.21	0.0022	6.8857	3.6867	-1.7190
L185	25	25	1491	26	1.74	22	4	55	0.15	0.0030	6.8857	4.5133	-1.2748
L290	25	25	772	26	3.37	22	4	67	0.35	0.0046	6.8857	3.5200	-1.8085
L372	25	25	1641	26	1.58	21	7	87	0.21	0.0023	6.8857	3.8033	-1.6563
L54	25	25	723	26	3.60	16	10	71	0.39	0.0062	6.8857	4.4800	-1.2927
L141	26	26	2002	26	1.30	17	9	128	0.25	0.0027	6.8135	5.4492	-0.7351
L291	26	26	1712	26	1.52	18	8	72	0.16	0.0021	6.8135	3.6738	-1.6916
L363	26	26	1131	26	2.30	21	5	96	0.33	0.0033	6.8135	3.7138	-1.6701
L62	26	26	1919	26	1.35	16	10	68	0.14	0.0020	6.8135	3.8677	-1.5872
L65	26	26	1651	26	1.57	15	11	82	0.19	0.0033	6.8135	5.4831	-0.7168
L78	26	26	1074	26	2.42	19	7	72	0.26	0.0028	6.8135	3.0400	-2.0331
L106	23	23	1068	25	2.34	18	7	68	0.28	0.0038	6.7736	4.0751	-1.4950
L304	25	25	671	25	3.73	23	3	8	0.05	0.0055	6.6208	3.6600	-1.6494
L325	25	25	1957	25	1.28	15	10	22	0.04	0.0028	6.6208	5.5400	-0.6021
L353	25	25	1489	25	1.68	18	7	122	0.33	0.0024	6.6208	3.5633	-1.7032
L378	25	25	731	25	3.42	20	8	41	0.22	0.0051	6.6208	3.7367	-1.6067
L48	25	25	1082	25	2.31	19	6	89	0.33	0.0031	6.6208	3.3600	-1.8165
L116	26	26	1698	25	1.47	20	5	46	0.10	0.0017	6.5514	2.8831	-2.0489
L121	26	26	1734	25	1.44	18	7	109	0.24	0.0020	6.5514	3.5354	-1.6846
L137	26	26	1057	25	2.37	14	11	96	0.35	0.0033	6.5514	3.4677	-1.7224
L140	26	26	1724	25	1.45	19	6	57	0.13	0.0017	6.5514	3.0154	-1.9750
L297	26	26	1895	25	1.32	20	5	107	0.22	0.0023	6.5514	4.2954	-1.2601
L365	26	26	1714	25	1.46	20	5	63	0.14	0.0028	6.5514	4.7231	-1.0212
L383	26	26	1848	25	1.35	20	5	71	0.15	0.0019	6.5514	3.5169	-1.6949
L52	26	26	1528	25	1.64	19	6	83	0.21	0.0027	6.5514	4.0800	-1.3804
L64	26	26	1735	25	1.44	17	8	43	0.10	0.0024	6.5514	4.2246	-1.2996
L100	23	23	1844	24	1.30	15	9	143	0.34	0.0026	6.5026	4.7273	-1.0212
L146	25	25	1304	24	1.84	17	7	65	0.20	0.0020	6.3560	2.6700	-2.1316
L272	25	25	1716	24	1.40	19	5	111	0.26	0.0018	6.3560	3.0667	-1.9022

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L324	25	25	1686	24	1.42	21	4	85	0.20	0.0031	6.3560	5.2100	-0.6627
L37	25	25	1575	24	1.52	16	8	89	0.23	0.0025	6.3560	3.9533	-1.3895
L377	25	25	1206	24	1.99	22	2	94	0.31	0.0023	6.3560	2.8233	-2.0429
L407	25	25	614	24	3.91	21	3	49	0.32	0.0079	6.3560	4.8367	-0.8786
L44	25	25	1584	24	1.52	17	7	153	0.39	0.0021	6.3560	3.3433	-1.7422
L221	26	26	1406	24	1.71	12	12	117	0.32	0.0024	6.2894	3.3354	-1.7127
L232	26	26	955	24	2.51	16	8	72	0.29	0.0043	6.2894	4.1108	-1.2632
L312	26	26	1793	24	1.34	15	9	115	0.25	0.0022	6.2894	4.0031	-1.3256
L134	24	24	1295	23	1.78	16	7	111	0.36	0.0026	6.1591	3.3478	-1.6858
L280	24	24	1783	23	1.29	18	5	33	0.08	0.0022	6.1591	3.8913	-1.3599
L317	24	24	1960	23	1.17	19	4	117	0.25	0.0019	6.1591	3.7029	-1.4729
L7	24	24	584	23	3.94	21	2	32	0.23	0.0055	6.1591	3.2029	-1.7727
L184	25	25	1381	23	1.67	17	6	106	0.31	0.0017	6.0912	2.3167	-2.2693
L242	25	25	942	23	2.44	12	11	37	0.16	0.0056	6.0912	5.2467	-0.5077
L360	25	25	1721	23	1.34	18	5	53	0.12	0.0017	6.0912	2.9233	-1.9046
L59	25	25	1496	23	1.54	18	5	54	0.14	0.0028	6.0912	4.1733	-1.1530
L77	25	25	1286	23	1.79	18	5	127	0.40	0.0026	6.0912	3.3267	-1.6621
L91	25	25	1611	23	1.43	19	4	112	0.28	0.0024	6.0912	3.8633	-1.3394
L143	26	26	2062	23	1.12	15	8	39	0.07	0.0026	6.0273	5.3815	-0.3892
L194	26	26	1309	23	1.76	15	8	110	0.32	0.0036	6.0273	4.6769	-0.8140
L248	26	26	2045	23	1.12	18	5	44	0.08	0.0018	6.0273	3.5969	-1.4649
L25	26	26	968	23	2.38	17	7	50	0.20	0.0030	6.0273	2.8585	-1.9100
L313	26	26	1397	23	1.65	16	7	43	0.12	0.0025	6.0273	3.5200	-1.5113
L335	26	26	1690	23	1.36	19	4	128	0.29	0.0021	6.0273	3.5323	-1.5039
L369	26	26	1387	23	1.66	18	5	42	0.12	0.0026	6.0273	3.6215	-1.4501
L38	26	26	1835	23	1.25	13	9	64	0.13	0.0013	6.0273	2.3508	-2.2161
L40	26	26	1726	23	1.33	15	8	81	0.18	0.0023	6.0273	3.9077	-1.2776
L53	26	26	1870	23	1.23	20	3	41	0.08	0.0018	6.0273	3.3354	-1.6226
L14	21	21	1778	21	1.18	14	7	73	0.20	0.0024	5.8370	4.3333	-0.9720

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L205	25	25	1781	22	1.24	17	5	123	0.28	0.0019	5.8263	3.3100	-1.5754
L357	25	25	1212	22	1.82	15	7	85	0.28	0.0031	5.8263	3.7567	-1.2957
L124	26	26	1304	22	1.69	14	8	109	0.32	0.0029	5.7653	3.7231	-1.2817
L15	26	26	1735	22	1.27	14	8	113	0.25	0.0014	5.7653	2.3446	-2.1468
L181	26	26	1435	22	1.53	15	7	146	0.39	0.0031	5.7653	4.3877	-0.8646
L223	26	26	1606	22	1.37	17	5	68	0.16	0.0022	5.7653	3.5846	-1.3686
L26	26	26	1789	22	1.23	15	7	70	0.15	0.0019	5.7653	3.3785	-1.4980
L269	26	26	1575	22	1.40	17	5	79	0.19	0.0017	5.7653	2.7200	-1.9112
L285	26	26	1774	22	1.24	18	4	67	0.15	0.0016	5.7653	2.7569	-1.8881
L306	26	26	1742	22	1.26	17	5	68	0.15	0.0019	5.7653	3.2585	-1.5733
L366	26	26	1787	22	1.23	17	6	276	0.59	0.0030	5.7653	5.4092	-0.2234
L397	26	26	1763	22	1.25	19	3	152	0.33	0.0022	5.7653	3.8954	-1.1736
L399	26	26	1818	22	1.21	17	5	59	0.12	0.0020	5.7653	3.6646	-1.3184
L46	26	26	1684	22	1.31	17	5	111	0.25	0.0017	5.7653	2.8585	-1.8243
L67	26	26	1736	22	1.27	13	9	35	0.08	0.0022	5.7653	3.8615	-1.1948
L204	23	23	1852	21	1.13	14	7	78	0.18	0.0026	5.6898	4.8103	-0.5715
L41	25	25	1741	21	1.21	19	2	77	0.18	0.0024	5.5615	4.2067	-0.8847
L334	21	21	1572	20	1.27	17	3	88	0.27	0.0022	5.5590	3.4429	-1.4300
L125	26	26	1737	21	1.21	14	7	108	0.24	0.0020	5.5032	3.4677	-1.3325
L152	26	26	1668	21	1.26	14	7	21	0.05	0.0028	5.5032	4.7385	-0.5006
L251	26	26	1623	21	1.29	16	5	57	0.14	0.0018	5.5032	2.8954	-1.7071
L281	26	26	1446	21	1.45	12	10	43	0.11	0.0023	5.5032	3.3323	-1.4211
L345	26	26	1264	21	1.66	16	5	70	0.21	0.0020	5.5032	2.4954	-1.9690
L346	26	26	1756	21	1.20	15	7	103	0.23	0.0022	5.5032	3.8615	-1.0747
L43	26	26	1752	21	1.20	20	1	49	0.11	0.0018	5.5032	3.0677	-1.5943
L56	26	26	1804	21	1.16	18	3	40	0.09	0.0019	5.5032	3.4738	-1.3284
L70	22	22	1058	20	1.89	13	7	49	0.21	0.0041	5.4864	4.3420	-0.7753
L130	24	24	787	20	2.54	11	9	92	0.49	0.0033	5.3558	2.6268	-1.8579
L314	24	24	876	20	2.28	15	5	49	0.23	0.0028	5.3558	2.4529	-1.9763

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L217	25	25	1581	20	1.27	15	5	43	0.11	0.0016	5.2967	2.5500	-1.8745
L274	25	25	1506	20	1.33	9	11	66	0.18	0.0026	5.2967	3.9400	-0.9259
L386	25	25	1157	20	1.73	18	3	97	0.34	0.0032	5.2967	3.7333	-1.0669
L51	25	25	1592	20	1.26	16	4	76	0.19	0.0024	5.2967	3.8333	-0.9987
L120	26	26	1419	20	1.41	14	6	42	0.11	0.0030	5.2411	4.2862	-0.6533
L122	26	26	1811	20	1.10	10	10	82	0.17	0.0014	5.2411	2.6062	-1.8025
L219	26	26	1602	20	1.25	15	5	63	0.15	0.0018	5.2411	2.8892	-1.6088
L58	26	26	1344	20	1.49	11	10	69	0.20	0.0019	5.2411	2.5169	-1.8635
L79	26	26	1493	20	1.34	15	5	42	0.11	0.0020	5.2411	3.0277	-1.5141
L4	23	23	757	19	2.51	14	5	25	0.14	0.0060	5.1479	4.5059	-0.4567
L10	24	24	1115	19	1.70	14	5	49	0.18	0.0032	5.0880	3.5254	-1.1142
L30	25	25	1023	19	1.86	11	8	60	0.23	0.0034	5.0318	3.5233	-1.0781
L388	25	25	1601	19	1.19	14	5	48	0.12	0.0023	5.0318	3.6567	-0.9828
L406	25	25	1358	19	1.40	13	6	57	0.17	0.0020	5.0318	2.7500	-1.6307
L127	26	26	2155	19	0.88	10	9	31	0.06	0.0012	4.9791	2.4954	-1.7791
L175	26	26	1377	19	1.38	15	4	114	0.32	0.0014	4.9791	1.8862	-2.2155
L179	26	26	1465	19	1.30	11	8	26	0.07	0.0019	4.9791	2.7569	-1.5917
L206	26	26	1581	19	1.20	16	3	36	0.09	0.0013	4.9791	2.0862	-2.0722
L24	26	26	1536	19	1.24	12	7	75	0.19	0.0014	4.9791	2.1292	-2.0413
L355	26	26	1385	19	1.37	14	5	48	0.13	0.0023	4.9791	3.2215	-1.2589
L104	25	25	1353	18	1.33	12	6	75	0.22	0.0019	4.7670	2.5700	-1.6480
L118	25	25	1328	18	1.36	14	4	83	0.25	0.0018	4.7670	2.4000	-1.7755
L199	25	25	1707	18	1.05	17	1	62	0.15	0.0012	4.7670	2.0200	-2.0605
L349	25	25	1694	18	1.06	10	7	96	0.23	0.0017	4.7670	2.7967	-1.4779
L99	25	25	1829	18	0.98	13	5	134	0.29	0.0018	4.7670	3.2633	-1.1279
L135	26	26	1490	18	1.21	15	3	123	0.32	0.0024	4.7170	3.5754	-0.8582
L166	26	26	1418	18	1.27	13	6	104	0.28	0.0018	4.7170	2.5785	-1.6077
L190	26	26	1898	18	0.95	12	6	76	0.15	0.0015	4.7170	2.8831	-1.3787
L293	26	26	1494	18	1.20	11	7	50	0.13	0.0020	4.7170	2.9508	-1.3278

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L31	26	26	1657	18	1.09	11	7	100	0.23	0.0014	4.7170	2.2862	-1.8274
L381	26	26	1677	18	1.07	11	7	139	0.32	0.0021	4.7170	3.5046	-0.9114
L75	26	26	1731	18	1.04	18	0	92	0.20	0.0014	4.7170	2.4646	-1.6933
L19	24	24	1126	17	1.51	16	1	71	0.26	0.0026	4.5524	2.9710	-1.2454
L93	24	24	1352	17	1.26	12	5	64	0.20	0.0015	4.5524	2.0000	-2.0102
L105	25	25	1709	17	0.99	12	5	56	0.13	0.0016	4.5022	2.7467	-1.3855
L115	25	25	1760	17	0.97	14	3	12	0.03	0.0014	4.5022	2.4400	-1.6276
L138	25	25	1243	17	1.37	11	6	92	0.30	0.0028	4.5022	3.4933	-0.7962
L253	25	25	1487	17	1.14	11	6	67	0.18	0.0017	4.5022	2.5833	-1.5144
L307	25	25	584	17	2.91	10	8	49	0.34	0.0040	4.5022	2.3100	-1.7302
L340	25	25	1599	17	1.06	15	2	46	0.12	0.0025	4.5022	4.0100	-0.3884
L87	25	25	774	17	2.20	14	3	64	0.33	0.0042	4.5022	3.2733	-0.9699
L158	26	26	1775	17	0.96	9	8	115	0.25	0.0020	4.4550	3.5477	-0.7176
L172	26	26	1218	17	1.40	15	2	50	0.16	0.0017	4.4550	2.0585	-1.8955
L174	26	26	1589	17	1.07	17	0	66	0.16	0.0020	4.4550	3.1969	-0.9950
L186	26	26	1806	17	0.94	15	2	45	0.10	0.0014	4.4550	2.5385	-1.5158
L255	26	26	1878	17	0.91	14	3	12	0.02	0.0013	4.4550	2.4985	-1.5475
L260	26	26	1207	17	1.41	12	5	49	0.16	0.0026	4.4550	3.0892	-1.0802
L277	26	26	1835	17	0.93	15	2	71	0.15	0.0022	4.4550	4.0769	-0.2990
L316	26	26	1727	17	0.98	12	5	55	0.12	0.0019	4.4550	3.2031	-0.9902
L338	26	26	1454	17	1.17	12	5	61	0.16	0.0023	4.4550	3.2738	-0.9342
L350	26	26	1555	17	1.09	15	2	22	0.05	0.0021	4.4550	3.2092	-0.9853
L380	26	26	1452	17	1.17	13	4	70	0.19	0.0017	4.4550	2.4338	-1.5986
L60	26	26	823	17	2.07	12	5	60	0.28	0.0029	4.4550	2.3877	-1.6351
L49	22	22	1301	16	1.23	13	3	31	0.11	0.0031	4.3891	3.9957	-0.3256
L195	23	23	586	16	2.73	12	4	35	0.26	0.0037	4.3351	2.1502	-1.8120
L354	23	23	1629	16	0.98	12	4	66	0.18	0.0013	4.3351	2.1028	-1.8513
L364	23	23	1601	16	1.00	7	9	80	0.22	0.0026	4.3351	4.1858	-0.1238
L187	24	24	1611	16	0.99	15	1	78	0.20	0.0024	4.2846	3.8913	-0.3269

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L69*	24	24	982	16	1.63	14	2	53	0.22	0.0023	4.2846	2.2464	-1.6939
L177	25	25	1943	16	0.82	12	4	90	0.19	0.0023	4.2373	4.4433	0.1715
L191	25	25	1475	16	1.08	12	4	68	0.18	0.0013	4.2373	1.8867	-1.9575
L222	25	25	1800	16	0.89	12	4	55	0.12	0.0014	4.2373	2.4833	-1.4607
L35	25	25	604	16	2.65	8	9	48	0.32	0.0034	4.2373	2.0567	-1.8160
L101	26	26	1829	16	0.87	11	5	31	0.07	0.0017	4.1929	3.0831	-0.9261
L151	26	26	1343	16	1.19	13	4	45	0.13	0.0014	4.1929	1.8646	-1.9428
L157	26	26	1483	16	1.08	14	2	49	0.13	0.0012	4.1929	1.7262	-2.0584
L256	26	26	1679	16	0.95	8	9	78	0.18	0.0015	4.1929	2.4831	-1.4268
L276	26	26	824	16	1.94	11	5	26	0.12	0.0025	4.1929	2.0277	-1.8068
L389	26	26	1744	16	0.92	11	5	74	0.16	0.0010	4.1929	1.6923	-2.0866
L66	26	26	1786	16	0.90	11	5	46	0.10	0.0012	4.1929	2.2215	-1.6450
L142	23	23	1254	15	1.20	10	5	38	0.13	0.0015	4.0641	1.8854	-1.9129
L287	23	23	970	15	1.55	14	1	110	0.49	0.0034	4.0641	3.3043	-0.6671
L20	24	24	1216	15	1.23	14	1	61	0.21	0.0019	4.0168	2.2717	-1.5351
L108	25	25	1208	15	1.24	9	6	38	0.13	0.0017	3.9725	2.0267	-1.7151
L196	25	25	889	15	1.69	9	6	51	0.23	0.0022	3.9725	1.9433	-1.7885
L235	25	25	1755	15	0.85	9	6	84	0.19	0.0014	3.9725	2.4467	-1.3449
L13	26	26	1161	15	1.29	9	6	28	0.09	0.0019	3.9309	2.2462	-1.4877
L144	26	26	1613	15	0.93	8	7	48	0.11	0.0012	3.9309	1.9569	-1.7432
L198	26	26	1763	15	0.85	12	3	72	0.16	0.0011	3.9309	1.9262	-1.7703
L228	26	26	1313	15	1.14	14	1	65	0.19	0.0016	3.9309	2.0400	-1.6698
L393	26	26	1740	15	0.86	12	3	51	0.11	0.0008	3.9309	1.3200	-2.3056
L401	26	26	1670	15	0.90	9	6	41	0.09	0.0013	3.9309	2.1354	-1.5856
L154	22	22	1127	14	1.24	9	6	26	0.10	0.0020	3.8405	2.2944	-1.4395
L331	23	23	421	14	3.33	12	2	63	0.65	0.0054	3.7932	2.2569	-1.4329
L39	23	23	1174	14	1.19	10	4	14	0.05	0.0023	3.7932	2.6561	-1.0606
L193	24	24	689	14	2.03	9	5	39	0.24	0.0039	3.7490	2.6993	-0.9809
L202	25	25	846	14	1.65	11	3	36	0.17	0.0016	3.7077	1.3400	-2.2165

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L275	25	25	1763	14	0.79	9	5	51	0.12	0.0018	3.7077	3.0967	-0.5720
L347	25	25	1070	14	1.31	11	3	38	0.14	0.0010	3.7077	1.0633	-2.4755
L129	26	26	1324	14	1.06	8	6	78	0.23	0.0024	3.6688	3.1908	-0.4483
L188	26	26	1266	14	1.11	8	6	49	0.15	0.0017	3.6688	2.1569	-1.4179
L361	26	26	1873	14	0.75	10	4	79	0.16	0.0009	3.6688	1.7508	-1.7988
L92	26	26	1526	14	0.92	11	3	51	0.13	0.0012	3.6688	1.9015	-1.6574
L348	24	24	837	13	1.55	10	3	36	0.18	0.0021	3.4812	1.7500	-1.7253
L368	24	24	1539	13	0.84	10	3	52	0.14	0.0018	3.4812	2.7500	-0.7287
L396	24	24	1082	13	1.20	6	7	30	0.12	0.0017	3.4812	1.8478	-1.6278
L107	25	25	848	13	1.53	9	4	65	0.31	0.0032	3.4428	2.7167	-0.7249
L236	25	25	1931	13	0.67	8	5	59	0.12	0.0014	3.4428	2.6433	-0.7981
L244	25	25	1616	13	0.80	10	3	111	0.27	0.0018	3.4428	2.8400	-0.6018
L282	25	25	1391	13	0.93	8	5	100	0.29	0.0012	3.4428	1.7167	-1.7231
L294	25	25	1695	13	0.77	8	5	66	0.16	0.0010	3.4428	1.7367	-1.7031
L208	26	26	1786	13	0.73	12	1	69	0.15	0.0010	3.4067	1.7231	-1.6835
L250	26	26	1694	13	0.77	4	9	43	0.10	0.0009	3.4067	1.4892	-1.9173
L356	26	26	1936	13	0.67	10	3	28	0.06	0.0007	3.4067	1.2646	-2.1419
L72	23	23	641	12	1.87	9	3	36	0.24	0.0031	3.2513	1.9605	-1.3760
L84	24	24	1701	12	0.71	12	0	64	0.16	0.0015	3.2135	2.5000	-0.7617
L128	25	25	1353	12	0.89	9	3	112	0.33	0.0014	3.1780	1.9267	-1.3380
L230	25	25	1102	12	1.09	9	3	50	0.18	0.0019	3.1780	2.1100	-1.1420
L257	25	25	515	12	2.33	8	4	35	0.27	0.0028	3.1780	1.4267	-1.8726
L367	25	25	1688	12	0.71	8	6	47	0.11	0.0012	3.1780	2.0467	-1.2097
L97	25	25	621	12	1.93	10	2	42	0.27	0.0023	3.1780	1.4000	-1.9011
L233	26	26	1796	12	0.67	8	4	69	0.15	0.0018	3.1447	3.2092	0.0691
L279	26	26	1869	12	0.64	11	1	70	0.14	0.0009	3.1447	1.6769	-1.5718
L86	26	26	832	12	1.44	10	2	46	0.21	0.0026	3.1447	2.1723	-1.0413
L21	22	22	1280	11	0.86	8	3	41	0.15	0.0016	3.0175	2.0130	-1.1519
L238	23	23	1427	11	0.77	7	4	32	0.10	0.0015	2.9804	2.1502	-0.9532

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L76	23	23	1824	11	0.60	7	4	94	0.22	0.0012	2.9804	2.2095	-0.8852
L156	24	24	590	11	1.86	11	0	25	0.18	0.0018	2.9457	1.0688	-2.1579
L296	24	24	1246	11	0.88	8	3	16	0.05	0.0017	2.9457	2.1196	-0.9498
L229	25	25	1742	11	0.63	6	4	86	0.20	0.0011	2.9132	1.9833	-1.0705
L246	25	25	1088	11	1.01	6	5	41	0.15	0.0014	2.9132	1.5133	-1.6117
L318	25	25	786	11	1.40	5	6	19	0.10	0.0015	2.9132	1.2033	-1.9686
L139	26	26	1649	11	0.67	8	3	42	0.10	0.0010	2.8826	1.5692	-1.5142
L170	26	26	787	11	1.40	5	6	38	0.19	0.0029	2.8826	2.2677	-0.7090
L270	26	26	1853	11	0.59	9	2	100	0.21	0.0008	2.8826	1.4369	-1.6667
L50	26	26	1746	11	0.63	8	3	37	0.08	0.0009	2.8826	1.6308	-1.4433
L305	23	23	646	10	1.55	8	2	15	0.10	0.0017	2.7094	1.1186	-1.9798
L351	23	23	588	10	1.70	8	2	26	0.19	0.0032	2.7094	1.9012	-1.0059
L47	24	24	1611	10	0.62	7	3	76	0.20	0.0009	2.6779	1.3804	-1.6165
L220	25	25	648	10	1.54	6	4	52	0.32	0.0022	2.6483	1.4033	-1.5529
L310	25	25	1212	10	0.83	8	2	61	0.20	0.0013	2.6483	1.6167	-1.2868
L352	25	25	1352	10	0.74	6	4	41	0.12	0.0014	2.6483	1.8767	-0.9625
L61	25	25	1752	10	0.57	8	2	51	0.12	0.0006	2.6483	1.0300	-2.0186
L81	25	25	904	10	1.11	7	3	25	0.11	0.0020	2.6483	1.7900	-1.0706
L98	25	25	1659	10	0.60	6	4	68	0.16	0.0010	2.6483	1.7333	-1.1413
L111	26	26	1355	10	0.74	7	3	51	0.14	0.0011	2.6206	1.4431	-1.4705
L173	26	26	1844	10	0.54	10	0	77	0.16	0.0007	2.6206	1.2062	-1.7663
L263	26	26	1859	10	0.54	7	3	33	0.07	0.0007	2.6206	1.3415	-1.5973
L392	26	26	1357	10	0.74	4	6	41	0.12	0.0012	2.6206	1.6831	-1.1707
L302	24	24	1757	9	0.51	8	1	28	0.07	0.0009	2.4101	1.5000	-1.2377
L16*	25	25	531	9	1.69	8	1	28	0.21	0.0028	2.3835	1.4667	-1.2480
L262	25	25	1073	9	0.84	7	2	41	0.15	0.0007	2.3835	0.7100	-2.2780
L283	25	25	941	9	0.96	7	2	21	0.09	0.0008	2.3835	0.7667	-2.2009
L32	26	26	1679	9	0.54	4	5	24	0.05	0.0009	2.3585	1.5077	-1.1593
L132	22	22	652	8	1.23	7	1	35	0.24	0.0018	2.1946	1.1645	-1.5409

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L375	25	25	1036	8	0.77	4	5	29	0.11	0.0014	2.1187	1.4933	-0.9372
L110	26	26	939	8	0.85	7	1	33	0.14	0.0014	2.0965	1.2708	-1.2383
L402	26	26	1288	8	0.62	5	3	34	0.10	0.0008	2.0965	1.0831	-1.5198
L286	24	24	884	7	0.79	5	2	34	0.16	0.0011	1.8745	0.9783	-1.4944
L374	24	24	776	7	0.90	6	1	20	0.11	0.0009	1.8745	0.7319	-1.9052
L167*	25	25	1125	7	0.62	6	1	21	0.07	0.0008	1.8538	0.8633	-1.6521
L322	25	25	799	7	0.88	6	1	26	0.13	0.0024	1.8538	1.9133	0.0992
L243	26	26	1102	7	0.64	5	2	22	0.08	0.0009	1.8344	0.9846	-1.4180
L261	26	26	1529	7	0.46	7	0	39	0.10	0.0006	1.8344	0.9538	-1.4694
L344	26	26	1839	7	0.38	4	3	28	0.06	0.0004	1.8344	0.7938	-1.7364
L213	22	22	1202	6	0.50	6	0	36	0.14	0.0009	1.6459	1.0779	-1.0694
L145	24	24	451	6	1.33	4	2	39	0.36	0.0022	1.6067	0.9783	-1.1830
L362	24	24	715	6	0.84	3	3	20	0.12	0.0013	1.6067	0.8986	-1.3330
L102	25	25	1635	6	0.37	4	2	41	0.10	0.0006	1.5890	0.9133	-1.2718
L133	25	25	752	6	0.80	5	1	22	0.12	0.0008	1.5890	0.5900	-1.8804
L249	25	25	1835	6	0.33	6	0	26	0.06	0.0003	1.5890	0.6067	-1.8491
L301	25	25	1748	6	0.34	4	2	39	0.09	0.0007	1.5890	1.3033	-0.5377
L241	24	24	464	5	1.08	4	1	2	0.02	0.0009	1.3389	0.4094	-2.0118
L83	24	24	500	5	1.00	5	0	34	0.28	0.0031	1.3389	1.5399	0.4348
L252	26	26	867	5	0.58	2	3	24	0.11	0.0008	1.3103	0.6615	-1.4031
L192	25	25	546	4	0.73	4	0	10	0.07	0.0009	1.0593	0.5167	-1.3842
L385	25	25	506	0	0.00	0	0	9	0.07	0.0006	0.0000	0.2967	0.0000
Mean (394 loci)	**25.14**	**25.14**	**1471.74**	**23.30**	**1.65**	**17.05**	**6.40**	**77.73**	**0.22**	**0.0025**	**6.1605**	**3.5765**	**-1.4821**
SD	**1.01**	**1.01**	**386.99**	**10.67**	**0.83**	**8.28**	**3.95**	**40.86**	**0.11**	**0.0013**	**2.8188**	**1.6929**	**0.4968**

Appendix 23: *Tropidurus itambere* loci summary statistics from the final Anchored Phylogenomics dataset used on BPP and phylogenetic analyses. The ones marked with an asterisk (*) were not used in the phylogenetic analyses because the outgroup specimen was not captured. Loci are ordered from highest to lowest molecular diversity (Watterson's θ). Outgroups and cryptic species outside of the monophyletic *T. itambere* group were not included in the calculations.

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L255	23	23	1891	81	4.28	62	19	142	0.33	0.0078	21.9464	14.6640	-1.3199
L330	23	23	2052	80	3.90	64	16	201	0.43	0.0086	21.6754	17.7075	-0.7279
L206	23	23	1714	79	4.61	66	15	184	0.47	0.0063	21.4045	10.7747	-1.9740
L2	22	22	1857	71	3.82	66	5	38	0.09	0.0096	19.4768	17.7749	-0.3492
L42	23	23	1628	70	4.30	60	10	199	0.53	0.0090	18.9660	14.7115	-0.8885
L108	23	23	1509	65	4.31	58	9	182	0.52	0.0084	17.6113	12.6601	-1.1108
L313	23	23	1355	65	4.80	49	16	139	0.45	0.0116	17.6113	15.6719	-0.4351
L53	23	23	1383	64	4.63	50	13	112	0.35	0.0132	17.3404	18.2411	0.2051
L175	23	23	1452	62	4.27	50	12	126	0.38	0.0109	16.7985	15.8735	-0.2172
L220	23	23	1752	60	3.42	49	12	145	0.36	0.0090	16.2566	15.8221	-0.1053
L182	23	23	1520	59	3.88	41	18	168	0.48	0.0079	15.9856	11.9684	-0.9895
L315	23	23	1613	59	3.66	45	14	182	0.49	0.0060	15.9856	9.7115	-1.5454
L239	21	21	2184	56	2.56	42	14	47	0.10	0.0061	15.5653	13.2429	-0.5970
L202	23	23	1546	52	3.36	41	11	127	0.36	0.0056	14.0890	8.6838	-1.5031
L385	23	23	1763	52	2.95	45	8	117	0.29	0.0061	14.0890	10.8063	-0.9129
L267	23	23	1257	51	4.06	41	11	139	0.48	0.0076	13.8181	9.5494	-1.2093
L293	23	23	1371	51	3.72	38	13	136	0.43	0.0062	13.8181	8.5257	-1.4994
L51	22	22	1852	50	2.70	40	13	149	0.37	0.0044	13.7161	8.1861	-1.5909
L7	21	21	1457	48	3.29	34	15	145	0.47	0.0066	13.3417	9.5762	-1.1223
L207	23	23	1348	49	3.64	39	11	52	0.17	0.0071	13.2762	9.5810	-1.0877
L317	23	23	1428	49	3.43	38	11	156	0.47	0.0070	13.2762	10.0040	-0.9632
L328	23	23	1528	49	3.21	37	14	142	0.40	0.0062	13.2762	9.4743	-1.1191
L5	23	23	1074	49	4.56	41	10	141	0.57	0.0125	13.2762	13.4229	0.0432
L389	22	22	1682	48	2.85	40	8	120	0.32	0.0052	13.1674	8.7619	-1.3179

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L90	23	23	1620	48	2.96	40	10	62	0.17	0.0064	13.0053	10.3478	-0.7978
L86	23	23	1521	47	3.09	34	14	161	0.46	0.0054	12.7343	8.2767	-1.3654
L257	22	22	1727	46	2.66	32	14	76	0.20	0.0039	12.6188	6.7749	-1.8207
L96	21	21	1653	45	2.72	35	10	105	0.30	0.0069	12.5079	11.4762	-0.3270
L78	23	23	1574	46	2.92	32	15	134	0.37	0.0060	12.4634	9.3834	-0.9630
L128	23	23	1444	44	3.05	38	6	85	0.26	0.0070	11.9215	10.0514	-0.6100
L157	20	20	1459	42	2.88	33	9	147	0.50	0.0084	11.8385	12.1842	0.1166
L217	22	22	1651	43	2.60	35	8	193	0.53	0.0055	11.7958	9.0000	-0.9289
L140	21	21	1621	42	2.59	32	10	171	0.50	0.0062	11.6740	10.0905	-0.5361
L113	23	23	971	43	4.43	31	10	243	1.09	0.0155	11.6505	15.0237	1.1246
L258	23	23	1415	43	3.04	29	15	66	0.20	0.0057	11.6505	8.0435	-1.2026
L111	20	20	1366	41	3.00	31	12	118	0.43	0.0074	11.5567	10.0421	-0.5228
L169	20	20	1676	41	2.45	32	9	50	0.15	0.0057	11.5567	9.5737	-0.6844
L129	22	22	436	42	9.63	37	4	53	0.55	0.0198	11.5215	8.6407	-0.9788
L225	21	21	1149	41	3.57	24	17	79	0.33	0.0081	11.3960	9.3619	-0.7046
L198	20	20	1431	40	2.80	28	13	144	0.50	0.0059	11.2748	8.3737	-1.0251
L1	22	22	1699	41	2.41	27	13	128	0.34	0.0060	11.2472	10.1255	-0.3899
L70	17	17	1327	38	2.86	22	15	148	0.66	0.0060	11.2402	7.9191	-1.2188
L224	21	21	1679	40	2.38	35	6	117	0.33	0.0033	11.1181	5.6048	-1.9551
L213	23	23	1394	41	2.94	33	9	142	0.44	0.0060	11.1087	8.3162	-0.9740
L264	23	23	1753	41	2.34	36	5	55	0.14	0.0048	11.1087	8.3676	-0.9561
L266	23	23	1776	40	2.25	32	10	168	0.41	0.0050	10.8377	8.9170	-0.6858
L4	23	23	1530	40	2.61	33	7	84	0.24	0.0063	10.8377	9.5652	-0.4544
L94	23	23	1697	40	2.36	25	15	86	0.22	0.0034	10.8377	5.8024	-1.7980
L188	20	20	1537	38	2.47	33	6	112	0.36	0.0043	10.7110	6.6316	-1.5134
L335	22	22	1710	39	2.28	34	5	59	0.16	0.0066	10.6985	11.2900	0.2156
L167	19	19	1756	37	2.11	24	14	67	0.20	0.0042	10.5862	7.3743	-1.2178
L254	23	23	1544	39	2.53	29	10	147	0.41	0.0038	10.5668	5.9091	-1.7035
L260	23	23	1229	39	3.17	32	7	118	0.42	0.0073	10.5668	9.0198	-0.5658

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L326	23	23	1876	39	2.08	31	8	103	0.24	0.0043	10.5668	8.0711	-0.9127
L162	21	21	1728	38	2.20	29	9	115	0.32	0.0033	10.5622	5.7476	-1.7924
L127	20	20	1266	37	2.92	31	6	45	0.18	0.0062	10.4292	7.8000	-1.0003
L35	20	20	1562	37	2.37	28	9	54	0.17	0.0062	10.4292	9.7211	-0.2694
L329	22	22	1674	38	2.27	33	5	181	0.49	0.0050	10.4242	8.4026	-0.7552
L125	23	23	1961	38	1.94	31	7	81	0.18	0.0040	10.2958	7.8063	-0.9331
L33	23	23	1378	38	2.76	26	12	102	0.32	0.0042	10.2958	5.7628	-1.6991
L73	23	23	1669	38	2.28	31	6	98	0.26	0.0060	10.2958	9.9763	-0.1198
L357	21	21	1692	37	2.19	26	11	62	0.17	0.0037	10.2842	6.3095	-1.5176
L311	22	22	1739	37	2.13	32	5	88	0.23	0.0042	10.1499	7.3853	-1.0591
L88	22	22	1542	37	2.40	25	11	132	0.39	0.0047	10.1499	7.2424	-1.1139
L141	23	23	1683	37	2.20	29	9	86	0.22	0.0037	10.0249	6.2174	-1.4635
L365	21	21	1622	36	2.22	24	12	26	0.08	0.0047	10.0063	7.6905	-0.9074
L381	21	21	2021	36	1.78	26	10	84	0.20	0.0039	10.0063	7.8762	-0.8346
L43	21	21	1589	36	2.27	30	7	54	0.16	0.0040	10.0063	6.3048	-1.4503
L45	21	21	1600	36	2.25	30	7	59	0.18	0.0048	10.0063	7.6048	-0.9410
L181	22	22	1999	36	1.80	28	9	85	0.19	0.0025	9.8756	4.9351	-1.9423
L200	22	22	1842	36	1.95	31	5	73	0.18	0.0032	9.8756	5.8398	-1.5866
L261	22	22	1563	36	2.30	33	3	114	0.33	0.0041	9.8756	6.4632	-1.3415
L310	22	22	1307	36	2.75	30	7	34	0.12	0.0059	9.8756	7.7532	-0.8344
L367	22	22	1267	36	2.84	29	8	90	0.32	0.0040	9.8756	5.0346	-1.9032
L14	23	23	1674	36	2.15	29	8	121	0.31	0.0057	9.7539	9.5771	-0.0698
L23	23	23	1768	36	2.04	30	6	102	0.25	0.0027	9.7539	4.6957	-1.9952
L263	23	23	1479	36	2.43	30	7	50	0.15	0.0052	9.7539	7.6324	-0.8368
L74	23	23	1241	36	2.90	29	7	81	0.28	0.0073	9.7539	9.1186	-0.2506
L17	21	21	1473	35	2.38	31	5	129	0.42	0.0037	9.7283	5.4524	-1.7205
L295	21	21	569	35	6.15	23	12	44	0.37	0.0155	9.7283	8.8190	-0.3659
L196	22	22	1683	35	2.08	27	8	47	0.13	0.0038	9.6013	6.4545	-1.2704
L229	22	22	2099	35	1.67	29	6	64	0.14	0.0034	9.6013	7.1299	-0.9977

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L301	22	22	1629	35	2.15	28	7	77	0.21	0.0040	9.6013	6.5801	-1.2197
L38	22	22	1575	35	2.22	26	9	100	0.29	0.0045	9.6013	7.1558	-0.9872
L279	20	20	1444	34	2.35	23	11	18	0.06	0.0055	9.5836	7.9947	-0.6548
L142	23	23	1589	35	2.20	22	14	124	0.34	0.0042	9.4830	6.6719	-1.1386
L271	23	23	953	35	3.67	22	14	86	0.39	0.0075	9.4830	7.1028	-0.9641
L292	23	23	1603	35	2.18	30	5	166	0.45	0.0038	9.4830	6.1265	-1.3595
L3	23	23	1736	35	2.02	26	10	120	0.30	0.0048	9.4830	8.3557	-0.4566
L342	23	23	1174	35	2.98	31	4	91	0.34	0.0063	9.4830	7.4269	-0.8328
L306	21	21	1683	34	2.02	26	9	85	0.24	0.0043	9.4504	7.2333	-0.9168
L31	21	21	1680	34	2.02	26	7	49	0.14	0.0050	9.4504	8.4762	-0.4028
L124	20	20	1039	33	3.18	25	8	34	0.16	0.0072	9.3017	7.5316	-0.7503
L100	23	23	1380	34	2.46	28	6	69	0.22	0.0049	9.2121	6.7826	-1.0112
L296	21	21	2154	33	1.53	31	2	134	0.30	0.0026	9.1724	5.5857	-1.5254
L361	21	21	1488	33	2.22	24	9	51	0.16	0.0042	9.1724	6.2190	-1.2560
L368	21	21	1430	33	2.31	30	3	57	0.19	0.0050	9.1724	7.1667	-0.8530
L185	20	20	1615	32	1.98	22	10	44	0.14	0.0038	9.0198	6.1263	-1.2625
L321	18	18	1569	31	1.98	25	7	106	0.38	0.0045	9.0128	7.1307	-0.8404
L197	23	23	1507	33	2.19	24	10	96	0.28	0.0041	8.9411	6.1265	-1.2048
L214	23	23	917	33	3.60	19	14	97	0.46	0.0071	8.9411	6.4901	-1.0491
L240	23	23	1890	33	1.75	25	8	96	0.22	0.0029	8.9411	5.4506	-1.4941
L324	23	23	1391	33	2.37	24	9	44	0.14	0.0057	8.9411	7.8775	-0.4553
L371	23	23	1564	33	2.11	27	6	45	0.13	0.0037	8.9411	5.7431	-1.3689
L76	23	23	1017	33	3.24	22	12	37	0.16	0.0066	8.9411	6.6877	-0.9645
L288	21	21	1859	32	1.72	24	8	40	0.10	0.0033	8.8945	6.2190	-1.1712
L370	21	21	1213	32	2.64	24	8	66	0.26	0.0070	8.8945	8.4333	-0.2019
L312	22	22	1980	32	1.62	28	5	74	0.17	0.0045	8.7783	8.8355	0.0251
L32	22	22	1471	32	2.18	27	6	73	0.23	0.0045	8.7783	6.6277	-0.9445
L130	20	20	1596	31	1.94	27	4	93	0.29	0.0030	8.7380	4.8053	-1.7678
L210	20	20	1641	31	1.89	24	7	51	0.16	0.0033	8.7380	5.4684	-1.4697

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L85	20	20	1521	31	2.04	23	8	65	0.21	0.0043	8.7380	6.4737	-1.0178
L191	23	23	1962	32	1.63	23	9	69	0.15	0.0033	8.6702	6.3874	-1.0057
L133	21	21	1544	31	2.01	27	4	24	0.07	0.0055	8.6165	8.5095	-0.0483
L34	21	21	1495	31	2.07	23	8	82	0.26	0.0037	8.6165	5.5571	-1.3797
L374	21	21	1683	31	1.84	25	7	41	0.12	0.0035	8.6165	5.8857	-1.2315
L92	21	21	2034	31	1.52	28	3	85	0.20	0.0026	8.6165	5.3429	-1.4763
L95	21	21	1632	31	1.90	29	2	57	0.17	0.0047	8.6165	7.7000	-0.4133
L144	22	22	1203	31	2.58	20	11	69	0.26	0.0062	8.5040	7.4935	-0.4571
L327	22	22	1634	31	1.90	26	6	75	0.21	0.0038	8.5040	6.1515	-1.0643
L57	22	22	1381	31	2.24	23	8	65	0.21	0.0030	8.5040	4.1126	-1.9867
L80	22	22	1289	31	2.40	24	7	97	0.34	0.0035	8.5040	4.4892	-1.8163
L243	20	20	1609	30	1.86	25	5	81	0.25	0.0040	8.4561	6.4368	-0.9360
L275	20	20	1510	30	1.99	20	11	42	0.14	0.0042	8.4561	6.3947	-0.9555
L9	20	20	1577	30	1.90	21	10	91	0.29	0.0044	8.4561	6.9158	-0.7140
L117	23	23	1745	31	1.78	23	8	80	0.20	0.0027	8.3992	4.7470	-1.6575
L151	23	23	1698	31	1.83	21	11	42	0.11	0.0026	8.3992	4.3953	-1.8171
L184	23	23	1249	31	2.48	12	19	94	0.33	0.0040	8.3992	5.0119	-1.5373
L201	23	23	1237	31	2.51	23	8	77	0.27	0.0044	8.3992	5.4743	-1.3274
L356	23	23	1755	31	1.77	27	4	90	0.22	0.0032	8.3992	5.5455	-1.2951
L237	21	21	1586	30	1.89	18	12	49	0.15	0.0045	8.3386	7.1571	-0.5494
L339	21	21	1603	30	1.87	22	8	77	0.23	0.0035	8.3386	5.6619	-1.2447
L165	22	22	1642	30	1.83	19	11	64	0.18	0.0032	8.2296	5.2511	-1.3894
L218	23	23	878	30	3.42	20	10	130	0.64	0.0052	8.1283	4.5257	-1.6857
L375	21	21	1925	29	1.51	21	8	72	0.18	0.0036	8.0606	7.0238	-0.4976
L383	21	21	1361	29	2.13	21	9	60	0.21	0.0031	8.0606	4.1905	-1.8576
L39	21	21	1568	29	1.85	23	6	103	0.31	0.0033	8.0606	5.2381	-1.3547
L158	23	23	1313	29	2.21	21	8	64	0.21	0.0046	7.8573	6.0751	-0.8607
L173	23	23	1851	29	1.57	24	5	110	0.26	0.0031	7.8573	5.8221	-0.9829
L219	21	21	1613	28	1.74	21	7	16	0.05	0.0031	7.7827	5.0619	-1.3493

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L259	21	21	1725	28	1.62	22	6	71	0.20	0.0026	7.7827	4.4476	-1.6539
L289	21	21	1169	28	2.40	20	8	29	0.12	0.0045	7.7827	5.2762	-1.2430
L155	22	22	1518	28	1.84	24	4	75	0.22	0.0024	7.6810	3.5714	-2.0442
L16	22	22	1787	28	1.57	23	5	98	0.25	0.0026	7.6810	4.6017	-1.5317
L36	22	22	1680	28	1.67	21	7	43	0.12	0.0036	7.6810	5.9913	-0.8405
L50	20	20	1580	27	1.71	19	8	54	0.17	0.0031	7.6105	4.9105	-1.3808
L18	23	23	1782	28	1.57	21	8	62	0.15	0.0035	7.5864	6.2767	-0.6535
L232	23	23	1445	28	1.94	19	9	26	0.08	0.0038	7.5864	5.4308	-1.0755
L250	23	23	1218	28	2.30	22	6	108	0.39	0.0047	7.5864	5.6996	-0.9414
L282	23	23	1403	28	2.00	20	8	53	0.16	0.0039	7.5864	5.5178	-1.0321
L366	23	23	1100	28	2.55	21	7	55	0.22	0.0042	7.5864	4.6403	-1.4699
L131	21	21	1064	27	2.54	17	10	59	0.26	0.0044	7.5047	4.6333	-1.4729
L139	21	21	785	27	3.44	23	4	100	0.61	0.0059	7.5047	4.6381	-1.4705
L164	21	21	1744	27	1.55	18	9	66	0.18	0.0034	7.5047	5.8857	-0.8305
L216	21	21	1458	27	1.85	22	5	63	0.21	0.0030	7.5047	4.3810	-1.6024
L256	21	21	1355	27	1.99	23	4	104	0.37	0.0030	7.5047	4.0381	-1.7783
L227	19	19	1284	26	2.02	12	15	50	0.20	0.0047	7.4390	6.0468	-0.7351
L91	19	19	1353	26	1.92	19	7	81	0.32	0.0034	7.4390	4.5906	-1.5041
L106	22	22	985	27	2.74	19	9	48	0.22	0.0052	7.4067	5.1472	-1.1625
L349	22	22	1700	27	1.59	22	5	85	0.23	0.0030	7.4067	5.1385	-1.1670
L379	22	22	1679	27	1.61	19	8	53	0.14	0.0034	7.4067	5.6926	-0.8819
L120	23	23	1703	27	1.59	14	13	73	0.19	0.0024	7.3155	4.0474	-1.6864
L278	23	23	1056	27	2.56	21	7	16	0.07	0.0052	7.3155	5.4862	-0.9440
L283	18	18	1439	25	1.74	19	6	36	0.14	0.0041	7.2684	5.9542	-0.7175
L137	21	21	1947	26	1.34	22	4	42	0.10	0.0023	7.2268	4.5667	-1.4131
L172	21	21	1295	26	2.01	20	6	47	0.17	0.0048	7.2268	6.2048	-0.5429
L44	21	21	1649	26	1.58	20	6	58	0.17	0.0024	7.2268	3.9762	-1.7268
L47	22	22	1228	26	2.12	20	6	72	0.27	0.0041	7.1324	5.0130	-1.1292
L308	20	20	992	25	2.52	18	8	41	0.21	0.0054	7.0467	5.3737	-0.9189

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L305	23	23	698	26	3.72	19	7	61	0.38	0.0041	7.0445	2.8775	-2.2266
L136	21	21	1210	25	2.07	22	3	36	0.14	0.0040	6.9488	4.8714	-1.1443
L280	21	21	1325	25	1.89	17	7	31	0.11	0.0035	6.9488	4.6095	-1.2886
L303	21	21	1181	25	2.12	19	6	54	0.22	0.0040	6.9488	4.7190	-1.2283
L340	21	21	1129	25	2.21	20	5	66	0.28	0.0049	6.9488	5.5381	-0.7771
L347	19	19	1636	24	1.47	18	7	55	0.18	0.0027	6.8667	4.4386	-1.3809
L107	22	22	1386	25	1.80	15	10	62	0.20	0.0035	6.8580	4.8268	-1.1221
L138	22	22	993	25	2.52	19	6	44	0.20	0.0049	6.8580	4.9048	-1.0791
L160	22	22	1164	25	2.15	17	8	76	0.30	0.0041	6.8580	4.7403	-1.1700
L166	22	22	1178	25	2.12	16	9	69	0.27	0.0037	6.8580	4.3680	-1.3756
L203	22	22	759	25	3.29	17	8	35	0.21	0.0058	6.8580	4.4329	-1.3398
L223	22	22	1059	25	2.36	20	5	49	0.21	0.0049	6.8580	5.1472	-0.9451
L245	22	22	1416	25	1.77	19	6	47	0.15	0.0033	6.8580	4.7100	-1.1867
L332	22	22	1291	25	1.94	19	6	99	0.35	0.0025	6.8580	3.2857	-1.9735
L325	23	23	1449	25	1.73	18	7	61	0.18	0.0027	6.7736	3.8617	-1.6132
L333	23	23	1779	25	1.41	20	5	51	0.12	0.0033	6.7736	5.8340	-0.5205
L338	23	23	1252	25	2.00	23	2	66	0.23	0.0036	6.7736	4.4466	-1.2892
L345	23	23	1336	25	1.87	24	1	61	0.20	0.0050	6.7736	6.6206	-0.0848
L189	21	21	1624	24	1.48	21	3	51	0.15	0.0030	6.6709	4.7905	-1.0755
L320	21	21	1613	24	1.49	15	9	72	0.21	0.0029	6.6709	4.6143	-1.1763
L146	22	22	1717	24	1.40	21	3	65	0.17	0.0031	6.5837	5.2641	-0.7569
L19	22	22	1741	24	1.38	17	7	30	0.08	0.0031	6.5837	5.3117	-0.7296
L294	22	22	1711	24	1.40	15	9	39	0.10	0.0044	6.5837	7.5152	0.5343
L40	22	22	1401	24	1.71	20	5	35	0.11	0.0043	6.5837	6.0476	-0.3075
L360	23	23	1705	24	1.41	15	9	37	0.09	0.0027	6.5026	4.5217	-1.1394
L153	21	21	1677	23	1.37	17	6	48	0.14	0.0026	6.3929	4.4429	-1.1599
L236	21	21	1614	23	1.43	18	6	59	0.17	0.0022	6.3929	3.5762	-1.6754
L373	21	21	1304	23	1.76	16	8	67	0.24	0.0034	6.3929	4.4190	-1.1741
L380	21	21	1659	23	1.39	18	5	51	0.15	0.0026	6.3929	4.3667	-1.2052

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L384*	21	21	1758	23	1.31	16	7	67	0.18	0.0028	6.3929	4.8952	-0.8908
L41	21	21	1280	23	1.80	17	6	47	0.17	0.0027	6.3929	3.4048	-1.7774
L248	22	22	1638	23	1.40	17	6	73	0.20	0.0021	6.3094	3.3593	-1.7596
L336	22	22	1696	23	1.36	20	3	93	0.25	0.0027	6.3094	4.4978	-1.0805
L343	22	22	1620	23	1.42	15	8	74	0.21	0.0022	6.3094	3.6364	-1.5943
L8	22	22	1587	23	1.45	15	8	32	0.09	0.0036	6.3094	5.7359	-0.3420
L177	23	23	1736	23	1.32	18	4	47	0.12	0.0021	6.2317	3.6561	-1.5404
L99	23	23	1445	23	1.59	20	3	37	0.11	0.0033	6.2317	4.7984	-0.8572
L322	20	20	1713	22	1.28	19	3	202	0.59	0.0016	6.2011	2.7947	-2.1046
L156	21	21	1251	22	1.76	17	5	84	0.32	0.0033	6.1150	4.1048	-1.2454
L273	21	21	1682	22	1.31	17	5	55	0.16	0.0034	6.1150	5.6952	-0.2600
L304	21	21	1915	22	1.15	15	7	55	0.14	0.0020	6.1150	3.8190	-1.4224
L359	21	21	734	22	3.00	15	7	33	0.21	0.0054	6.1150	3.9952	-1.3132
L61	21	21	1615	22	1.36	16	5	58	0.17	0.0018	6.1150	2.9286	-1.9741
L231	22	22	1626	22	1.35	17	7	98	0.27	0.0020	6.0351	3.1818	-1.7724
L299	22	22	1311	22	1.68	17	5	62	0.21	0.0028	6.0351	3.6667	-1.4712
L323	22	22	1584	22	1.39	19	2	51	0.15	0.0032	6.0351	5.0823	-0.5919
L352	22	22	1221	22	1.80	20	2	33	0.12	0.0051	6.0351	6.2771	0.1503
L72	22	22	1559	22	1.41	13	9	26	0.08	0.0032	6.0351	4.9784	-0.6564
L358	19	19	1113	21	1.89	17	4	105	0.50	0.0050	6.0084	5.5497	-0.2949
L269	23	23	1784	22	1.23	14	8	44	0.11	0.0025	5.9607	4.4466	-0.9430
L334	23	23	997	22	2.21	17	5	41	0.18	0.0031	5.9607	3.0830	-1.7924
L178	21	21	1650	21	1.27	13	8	47	0.14	0.0022	5.8370	3.6952	-1.3845
L362	21	21	1387	21	1.51	14	7	48	0.16	0.0037	5.8370	5.1143	-0.4672
L48	21	21	1521	21	1.38	12	9	55	0.17	0.0033	5.8370	5.0667	-0.4980
L49	21	21	1504	21	1.40	18	3	29	0.09	0.0036	5.8370	5.3667	-0.3040
L123	22	22	946	21	2.22	16	5	37	0.18	0.0046	5.7608	4.3463	-0.9167
L286	22	22	2087	21	1.01	16	5	101	0.22	0.0024	5.7608	5.0260	-0.4762
L355	22	22	1499	21	1.40	19	2	44	0.13	0.0031	5.7608	4.5714	-0.7708

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L363	22	22	1759	21	1.19	16	5	82	0.21	0.0023	5.7608	3.9740	-1.1580
L386	22	22	1494	21	1.41	17	4	51	0.16	0.0031	5.7608	4.6840	-0.6979
L89	22	22	1663	21	1.26	17	5	93	0.25	0.0030	5.7608	4.9177	-0.5463
L116	23	23	1671	21	1.26	11	10	80	0.21	0.0028	5.6898	4.6522	-0.6742
L193	23	23	1796	21	1.17	17	4	48	0.12	0.0022	5.6898	4.0119	-1.0903
L221	23	23	1278	21	1.64	14	7	67	0.23	0.0034	5.6898	4.3715	-0.8566
L354	23	23	1660	21	1.27	17	4	57	0.15	0.0016	5.6898	2.7233	-1.9275
L97	23	23	1816	21	1.16	15	6	39	0.09	0.0024	5.6898	4.4229	-0.8232
L101	21	21	1306	20	1.53	17	3	44	0.16	0.0035	5.5590	4.5524	-0.6803
L103	21	21	1541	20	1.30	14	6	50	0.15	0.0026	5.5590	4.0190	-1.0406
L24	21	21	1672	20	1.20	15	5	66	0.19	0.0028	5.5590	4.6810	-0.5934
L119	22	22	1607	20	1.24	17	3	83	0.23	0.0023	5.4864	3.7749	-1.1595
L135	22	22	1809	20	1.11	14	6	95	0.24	0.0023	5.4864	4.1299	-0.9190
L348	22	22	1746	20	1.15	15	5	69	0.18	0.0019	5.4864	3.2727	-1.4997
L66	22	22	1625	20	1.23	18	2	59	0.17	0.0023	5.4864	3.7836	-1.1536
L11	23	23	1570	20	1.27	10	10	75	0.21	0.0027	5.4189	4.2134	-0.8187
L132	20	20	1236	19	1.54	17	2	53	0.21	0.0046	5.3555	5.6526	0.2098
L168	20	20	1789	19	1.06	11	8	96	0.27	0.0022	5.3555	3.8789	-1.0427
L15	21	21	1533	19	1.24	13	6	78	0.24	0.0022	5.2811	3.4333	-1.3080
L150	21	21	1545	19	1.23	14	5	45	0.14	0.0026	5.2811	4.0429	-0.8765
L20	21	21	1330	19	1.43	14	5	78	0.28	0.0028	5.2811	3.7000	-1.1192
L270	21	21	1334	19	1.42	15	4	48	0.17	0.0027	5.2811	3.6571	-1.1496
L276	21	21	1153	19	1.65	16	3	48	0.20	0.0022	5.2811	2.5048	-1.9653
L77	21	21	1340	19	1.42	16	3	39	0.14	0.0021	5.2811	2.8762	-1.7024
L186	22	22	1069	19	1.78	13	6	64	0.27	0.0029	5.2121	3.0476	-1.5359
L22	22	22	1675	19	1.13	18	1	44	0.12	0.0028	5.2121	4.7229	-0.3471
L252	22	22	933	19	2.04	13	6	37	0.18	0.0061	5.2121	5.7186	0.3594
L287	22	22	1179	19	1.61	13	6	36	0.14	0.0043	5.2121	5.0346	-0.1259
L30	22	22	1653	19	1.15	16	3	17	0.05	0.0022	5.2121	3.6017	-1.1427

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L122	19	19	1048	18	1.72	12	6	37	0.19	0.0033	5.1501	3.4678	-1.2445
L316	19	19	1460	18	1.23	14	4	57	0.21	0.0023	5.1501	3.3626	-1.3224
L12	23	23	1657	19	1.15	15	4	75	0.20	0.0021	5.1479	3.5573	-1.1314
L284	23	23	1839	19	1.03	12	7	128	0.30	0.0020	5.1479	3.6996	-1.0302
L376	23	23	632	19	3.01	12	7	19	0.13	0.0046	5.1479	2.8933	-1.6038
L54	23	23	1762	19	1.08	16	3	57	0.14	0.0020	5.1479	3.5455	-1.1399
L277	20	20	584	18	3.08	14	5	7	0.06	0.0055	5.0737	3.2316	-1.3659
L29	20	20	1539	18	1.17	14	5	102	0.33	0.0029	5.0737	4.4947	-0.4293
L134	21	21	1696	18	1.06	13	5	51	0.14	0.0025	5.0031	4.3190	-0.5085
L230	21	21	1158	18	1.55	14	4	26	0.11	0.0030	5.0031	3.4667	-1.1420
L25	21	21	1317	18	1.37	17	1	36	0.13	0.0033	5.0031	4.3905	-0.4554
L369	21	21	1660	18	1.08	15	3	42	0.12	0.0021	5.0031	3.4667	-1.1420
L56	21	21	1382	18	1.30	12	6	25	0.09	0.0020	5.0031	2.8143	-1.6269
L195	22	22	1914	18	0.94	16	3	41	0.10	0.0020	4.9378	3.7662	-0.8728
L55	22	22	1646	18	1.09	15	3	44	0.12	0.0023	4.9378	3.7229	-0.9050
L154	23	23	1981	18	0.91	16	2	67	0.15	0.0023	4.8770	4.5178	-0.2682
L314	23	23	1261	18	1.43	12	6	50	0.17	0.0022	4.8770	2.7510	-1.5875
L318	23	23	365	18	4.93	16	2	54	0.64	0.0094	4.8770	3.4190	-1.0887
L319	23	23	1788	18	1.01	17	1	136	0.33	0.0022	4.8770	3.8854	-0.7404
L378	23	23	1577	18	1.14	14	4	54	0.15	0.0018	4.8770	2.8063	-1.5462
L161	19	19	1438	17	1.18	12	5	37	0.14	0.0023	4.8639	3.2749	-1.2377
L65	19	19	1096	17	1.55	13	4	30	0.14	0.0026	4.8639	2.8421	-1.5748
L300	20	20	1205	17	1.41	13	4	47	0.20	0.0022	4.7918	2.6737	-1.6534
L126	21	21	619	17	2.75	14	3	29	0.22	0.0055	4.7252	3.4048	-1.0331
L149	21	21	1757	17	0.97	13	4	66	0.18	0.0019	4.7252	3.2619	-1.1448
L152*	21	21	927	17	1.83	14	3	47	0.24	0.0042	4.7252	3.9143	-0.6344
L180	21	21	1369	17	1.24	13	4	63	0.22	0.0019	4.7252	2.6048	-1.6590
L246	21	21	1629	17	1.04	14	3	65	0.19	0.0016	4.7252	2.6048	-1.6590
L26	21	21	587	17	2.90	12	5	67	0.54	0.0054	4.7252	3.1667	-1.2193

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L309	21	21	767	17	2.22	14	3	44	0.27	0.0036	4.7252	2.7905	-1.5137
L75	21	21	489	17	3.48	11	6	39	0.38	0.0075	4.7252	3.6667	-0.8282
L62	22	22	1640	17	1.04	12	5	69	0.19	0.0014	4.6635	2.2294	-1.9085
L268	23	23	1766	17	0.96	14	3	55	0.14	0.0019	4.6060	3.3439	-0.9919
L83	23	23	817	17	2.08	12	5	48	0.26	0.0047	4.6060	3.8459	-0.5974
L249	20	20	518	16	3.09	12	4	44	0.42	0.0070	4.5099	3.6474	-0.7108
L350	20	20	1936	16	0.83	11	5	53	0.14	0.0015	4.5099	2.9895	-1.2530
L81	20	20	1551	16	1.03	10	6	40	0.13	0.0018	4.5099	2.8105	-1.4005
L199	21	21	1606	16	1.00	12	4	34	0.10	0.0016	4.4472	2.5524	-1.5649
L353	21	21	1591	16	1.01	14	3	41	0.12	0.0021	4.4472	3.3286	-0.9239
L226	22	22	1611	16	0.99	10	6	33	0.09	0.0017	4.3891	2.7229	-1.3789
L307	22	22	986	16	1.62	13	3	63	0.29	0.0028	4.3891	2.7359	-1.3682
L159	20	20	1057	15	1.42	12	3	18	0.09	0.0026	4.2280	2.7842	-1.2602
L346	20	20	1215	15	1.23	13	3	65	0.27	0.0026	4.2280	3.1000	-0.9846
L228	21	21	1791	15	0.84	15	0	25	0.07	0.0019	4.1693	3.4476	-0.6311
L247	21	21	1743	15	0.86	9	6	18	0.05	0.0019	4.1693	3.2238	-0.8268
L297	21	21	1092	15	1.37	13	2	29	0.13	0.0026	4.1693	2.8667	-1.1392
L298	21	21	1611	15	0.93	11	4	40	0.12	0.0022	4.1693	3.5619	-0.5312
L46	21	21	1612	15	0.93	14	1	44	0.13	0.0016	4.1693	2.6095	-1.3640
L102	22	22	919	15	1.63	13	2	29	0.14	0.0051	4.1148	4.6667	0.4836
L114	22	22	1391	15	1.08	10	5	40	0.13	0.0024	4.1148	3.3723	-0.6506
L27	22	22	1037	15	1.45	10	5	7	0.03	0.0029	4.1148	3.0390	-0.9427
L37	22	22	1153	15	1.30	12	3	53	0.21	0.0032	4.1148	3.7056	-0.3586
L387	22	22	1223	15	1.23	7	8	53	0.20	0.0042	4.1148	5.1169	0.8781
L69	22	22	1686	15	0.89	12	3	52	0.14	0.0019	4.1148	3.2381	-0.7682
L337	23	23	1358	15	1.10	12	3	59	0.19	0.0021	4.0641	2.8656	-1.0523
L93	23	23	917	15	1.64	13	2	30	0.14	0.0031	4.0641	2.8142	-1.0974
L163	21	21	1278	14	1.10	8	6	21	0.08	0.0023	3.8913	2.9810	-0.8460
L176	21	21	1393	14	1.01	8	6	20	0.07	0.0019	3.8913	2.7048	-1.1027

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L209	21	21	1571	14	0.89	11	3	59	0.18	0.0018	3.8913	2.7905	-1.0231
L64	21	21	1530	14	0.92	12	2	34	0.11	0.0017	3.8913	2.5524	-1.2443
L82	21	21	1458	14	0.96	10	4	46	0.15	0.0016	3.8913	2.4000	-1.3859
L118	22	22	1095	14	1.28	10	4	44	0.18	0.0023	3.8405	2.5195	-1.2299
L190	22	22	1727	14	0.81	11	4	47	0.12	0.0015	3.8405	2.6364	-1.1211
L222	22	22	1638	14	0.85	12	2	59	0.16	0.0017	3.8405	2.7619	-1.0042
L291	23	23	1655	14	0.85	11	3	33	0.09	0.0013	3.7932	2.2095	-1.4772
L372	23	23	678	14	2.06	14	0	37	0.24	0.0040	3.7932	2.7075	-1.0127
L67	18	18	1115	13	1.17	7	6	77	0.38	0.0033	3.7796	3.7059	-0.0727
L204	20	20	1486	13	0.87	6	7	35	0.12	0.0021	3.6643	3.0632	-0.5951
L285	20	20	798	13	1.63	10	3	15	0.09	0.0031	3.6643	2.4684	-1.1838
L143	21	21	1669	13	0.78	12	0	115	0.33	0.0014	3.6134	2.3429	-1.2598
L208	21	21	1561	13	0.83	7	6	42	0.13	0.0013	3.6134	2.1000	-1.5006
L344	21	21	1653	13	0.79	11	2	63	0.18	0.0019	3.6134	3.1714	-0.4382
L364	21	21	1819	13	0.71	10	3	26	0.07	0.0011	3.6134	2.0143	-1.5856
L60	21	21	1776	13	0.73	9	4	39	0.10	0.0011	3.6134	2.0095	-1.5903
L112	22	22	1226	13	1.06	8	4	109	0.40	0.0022	3.5662	2.7532	-0.8074
L272	22	22	971	13	1.34	12	1	23	0.11	0.0026	3.5662	2.5195	-1.0396
L331	23	23	1786	13	0.73	12	1	35	0.09	0.0017	3.5223	2.9486	-0.5707
L115	21	21	1548	12	0.78	11	1	99	0.30	0.0027	3.3354	4.1333	0.8480
L21	21	21	1070	12	1.12	7	5	67	0.30	0.0024	3.3354	2.6190	-0.7614
L253	21	21	1821	12	0.66	8	4	38	0.10	0.0017	3.3354	3.1619	-0.1844
L28	21	21	1003	12	1.20	9	3	17	0.08	0.0020	3.3354	2.0524	-1.3636
L68	21	21	808	12	1.49	8	4	32	0.19	0.0024	3.3354	1.9381	-1.4851
L10	22	22	1005	12	1.19	10	2	54	0.24	0.0025	3.2919	2.4978	-0.8451
L13	22	22	983	12	1.22	10	2	21	0.10	0.0024	3.2919	2.3290	-1.0248
L211	22	22	1628	12	0.74	7	5	78	0.22	0.0013	3.2919	2.1126	-1.2552
L262	22	22	1727	12	0.69	9	3	35	0.09	0.0014	3.2919	2.4156	-0.9327
L274	22	22	1723	12	0.70	10	3	81	0.21	0.0015	3.2919	2.5628	-0.7760

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L265	23	23	765	12	1.57	9	3	40	0.23	0.0029	3.2513	2.2530	-1.0642
L79	23	23	929	12	1.29	10	2	10	0.05	0.0028	3.2513	2.6403	-0.6513
L233	20	20	481	11	2.29	8	3	15	0.16	0.0033	3.1006	1.5842	-1.7345
L351	20	20	1355	11	0.81	9	2	23	0.08	0.0017	3.1006	2.2947	-0.9218
L104	21	21	822	11	1.34	9	2	16	0.09	0.0029	3.0575	2.3429	-0.8184
L147	21	21	1148	11	0.96	7	4	46	0.19	0.0022	3.0575	2.5238	-0.6112
L194	21	21	1673	11	0.66	7	4	23	0.07	0.0013	3.0575	2.2238	-0.9548
L63	21	21	1614	11	0.68	8	3	23	0.07	0.0016	3.0575	2.6476	-0.4694
L105	22	22	1217	11	0.90	8	3	32	0.12	0.0015	3.0175	1.8701	-1.3158
L110	22	22	1547	11	0.71	8	3	45	0.13	0.0012	3.0175	1.7922	-1.4051
L179	22	22	934	11	1.18	8	3	34	0.17	0.0015	3.0175	1.3593	-1.9015
L251	22	22	1386	11	0.79	9	2	55	0.18	0.0011	3.0175	1.4935	-1.7476
L121	20	20	1906	10	0.52	7	3	51	0.13	0.0013	2.8187	2.5579	-0.3236
L238	20	20	931	10	1.07	9	1	33	0.18	0.0028	2.8187	2.6263	-0.2387
L52	20	20	646	10	1.55	7	3	28	0.22	0.0018	2.8187	1.1737	-2.0408
L187	21	21	695	10	1.44	9	1	48	0.33	0.0031	2.7795	2.1667	-0.7611
L215	21	21	1706	10	0.59	7	3	30	0.08	0.0011	2.7795	1.9095	-1.0804
L98	21	21	1652	10	0.61	6	4	38	0.11	0.0010	2.7795	1.5714	-1.5002
L109	22	22	847	9	1.06	5	3	29	0.16	0.0023	2.4689	1.9697	-0.6777
L71	22	22	1606	9	0.56	5	4	9	0.03	0.0011	2.4689	1.7619	-0.9598
L234	21	21	874	8	0.92	4	4	26	0.14	0.0017	2.2236	1.4476	-1.1603
L377	22	22	1329	7	0.53	6	1	84	0.29	0.0012	1.9203	1.6364	-0.4731
L205	20	20	378	6	1.59	5	1	3	0.04	0.0038	1.6912	1.4316	-0.4893
L235	20	20	937	6	0.64	4	2	13	0.07	0.0011	1.6912	0.9842	-1.3323
L87	20	20	1275	6	0.47	5	1	22	0.09	0.0006	1.6912	0.7737	-1.7290
L341	22	22	1877	6	0.32	5	1	43	0.10	0.0008	1.6459	1.4242	-0.4174
L183	23	23	471	6	1.27	3	3	21	0.19	0.0027	1.6257	1.2727	-0.6644
L192	20	20	622	5	0.80	4	1	18	0.14	0.0010	1.4093	0.6000	-1.7572
L244	21	21	768	5	0.65	4	1	15	0.09	0.0016	1.3898	1.1905	-0.4322

Locus	# individuals	# haplotypes	bp	# polymorphic sites	polymorphic sites (%)	# transitions	# transversions	# ambiguous bases in alignment	ambiguous bases (%)	Nucleotide diversity	θ (Watterson)	θπ (pairwise)	Tajima's D
L170*	22	22	1229	4	0.33	3	1	14	0.05	0.0006	1.0973	0.7489	-0.8916
L388*	22	22	384	4	1.04	1	3	11	0.13	0.0009	1.0973	0.3463	-1.9219
L290	23	23	1618	4	0.25	3	1	20	0.05	0.0002	1.0838	0.3399	-1.9013
L171	20	20	498	3	0.60	2	1	10	0.10	0.0014	0.8456	0.6842	-0.5098
L302	20	20	503	3	0.60	3	0	10	0.10	0.0012	0.8456	0.5789	-0.8423
L212	19	19	278	2	0.72	1	1	12	0.23	0.0014	0.5722	0.3801	-0.8024
L242	20	20	738	2	0.27	2	0	8	0.05	0.0010	0.5637	0.7368	0.7200
L84	20	20	757	2	0.26	1	1	15	0.10	0.0006	0.5637	0.4263	-0.5716
L241	20	20	1112	1	0.09	1	0	6	0.03	0.0001	0.2819	0.1000	-1.1644
L382*	20	20	335	1	0.30	1	0	32	0.48	0.0003	0.2819	0.1000	-1.1644
L58	23	23	403	0	0.00	0	0	458	4.94	0.0004	0.0000	0.1542	0.0000
Mean (383 loci)	**21.61**	**21.61**	**1417.24**	**25.77**	**1.87**	**19.72**	**6.20**	**67.42**	**0.23**	**0.0038**	**7.0924**	**5.1906**	**-1.0240**
SD	**1.16**	**1.16**	**370.09**	**13.34**	**1.00**	**10.97**	**3.72**	**44.95**	**0.27**	**0.0023**	**3.6306**	**3.0265**	**0.5360**

Appendix 24: Delimited species and their posterior probability as estimated by BPP v3.0 on different runs for *Gymnodactylus amarali* (Table A), *Micrablepharus atticolus* (Table B), *Tropidurus itambere* (Table C), and their respective outgroups. Species with less than 0.01 posterior probabilities in every run were omitted for clarity. No cleandata is when ambiguous and missing sites were included in the likelihood calculations, whereas for cleandata they were removed.

Table A:

Delimited species	Algorithm 1 No cleandata	Algorithm 2 No cleandata	Algorithm 1 Cleandata	Algorithm 2 Cleandata
G. darwinii - Matias Cardoso	1.000	1.000	1.000	1.000
G. geckoides - Manga	1.000	1.000	1.000	1.000
G. geckoides - Exu	1.000	1.000	1.000	1.000
G. amarali - clade 1	1.000	1.000	1.000	1.000
G. amarali - clade 2	1.000	1.000	1.000	1.000
G. amarali - clade 3	1.000	1.000	1.000	1.000
G. amarali - clade 4	1.000	1.000	1.000	1.000
G. amarali - clade 5	1.000	1.000	1.000	1.000
G. amarali - clade 6	1.000	1.000	1.000	1.000
G. amarali - clade 7	1.000	1.000	1.000	1.000
G. amarali - clade 8	0.046 - 0.010	0.076 - 0.011	0.099 - 0.044	0.391 - 0.623
G. amarali - clades 8–9	0.953 - 0.990	0.923 - 0.989	0.900 - 0.955	0.376 - 0.608
G. amarali - clade 9	0.046 - 0.010	0.076 - 0.011	0.032 - 0.099	0.390 - 0.623
G. amarali - clade 9–10–11–12	–	–	0.012	–
G. amarali - clade 10	1.000	0.951 - 1.000	0.615 - 0.955	0.353 - 0.854
G. amarali - clades 10–11	–	–	–	0.091
G. amarali - clades 10–11–12	–	0.048	0.032 - 0.384	0.002 - 0.646
G. amarali - clades 10–12	–	–	–	0.050
G. amarali - clades 11	1.000	0.812 - 1.000	0.144 - 0.240	0.001 - 0.532
G. amarali - clades 11–12	–	0.139	0.471 - 0.715	0.352 - 0.372
G. amarali - clade 12	1.000	0.812 - 1.000	0.144 - 0.240	0.001 - 0.573

Table B:

Delimited species	Algorithm 1 No cleandata	Algorithm 2 No cleandata	Algorithm 1 Cleandata	Algorithm 2 Cleandata
Micrablepharus maximiliani	1.000	1.000	1.000	1.000
M. atticolus - A	1.000	1.000	1.000	1.000
M. atticolus - B	1.000	1.000	1.000	1.000
M. atticolus - C	1.000	1.000	1.000	1.000
M. atticolus - D	1.000	1.000	1.000	1.000
M. atticolus - E	1.000	1.000	0.919 - 0.953	0.997 - 1.000
M. atticolus - E–F	–	–	0.046 - 0.080	0.002
M. atticolus - F	1.000	1.000	0.849 - 0.919	0.250 - 0.993
M. atticolus - F–G	–	–	–	0.705
M. atticolus - F–G–H–I	–	–	0.057	0.043 - 0.003
M. atticolus - F–H–I	–	–	0.046	–
M. atticolus - G	1.000	1.000	0.939 - 1.000	0.247 - 0.995
M. atticolus - G–H–I	–	–	0.002	0.001 - 0.003
M. atticolus - H	1.000	1.000	0.893 - 1.000	0.952 - 0.995
M. atticolus - I	1.000	1.000	0.893 - 1.000	0.952 - 0.995

Table C:

Delimited species	Algorithm 1 No cleandata	Algorithm 2 No cleandata	Algorithm 1 Cleandata	Algorithm 2 Cleandata
Uranoscodon superciliosus	1.000	1.000	1.000	1.000
torqMo–hisp–itaNat–ore–itaMo–torq	1.000	1.000		
T. torquatus - Moeda (torqMo)	0.027	0.082	0.001 - 0.003	0.001 - 0.006
torqMo–hisp	0.973	0.918	0.147	0.026 - 0.113
torqMo–itaMo	–		0.998	
torqMo–hisp–itaMo	–		0.849	0.879 - 0.972
T. hispidus - Exú (hisp)	0.027	0.068	0.003 - 1.000	0.001 - 0.006
hisp–itaNat–ore		0.014		
T. itambere - Natividade (itaNat)	0.027	0.041	0.161 - 0.269	0.282 - 0.311
itaNat–ore	0.973	0.945	0.730 - 0.838	0.688 - 0.717
T. oreadicus - São Domingos (ore)	0.027	0.041	0.161 - 0.269	0.282 - 0.311
T. itambere - Moeda (itaMo)	1.000	1.000	0.001 - 0.150	0.027 - 0.120
T. torquatus - Caseara (torq)	1.000	1.000	1.000	1.000
T. itambere - A	1.000	1.000	1.000	1.000
T. itambere - B	1.000	1.000	1.000	1.000
T. itambere - C	1.000	1.000	1.000	1.000
T. itambere - D	1.000	1.000	1.000	1.000
T. itambere - E	1.000	1.000	1.000	1.000

Appendix 25: Relative substitution rates for AP. Average pairwise genetic distance (π) for each AP locus and for cytb are shown. The latter was calculated using sequences from the same individuals for which we had AP sequences. The cytb substitution rate is a general per-lineage rate (sub/site/million years) calculated for lizards using geomorphological information (Macey *et al.*, 1998). Relative substitution ratios (AP π/ Cytb π) and relative substitution rates for AP (AP π/ cytB π x 0.0065 (sub/site/million years)) are provided.

Taxon	AP π	Cytb π	Cytb substitution rate (sub/site/million years)	AP relative substitution ratio	AP relative substitution rate (sub/site/million years)
Gymnodactylus amarali	5.2610	66.1818	0.0065	0.0795	0.0005
Micrablepharus atticolus	3.5765	33.6462	0.0065	0.1063	0.0007
Tropidurus itambere	5.1906	59.1255	0.0065	0.0878	0.0006

Appendix 26: Locality records of *Gymnodactylus amarali* from the Brazilian Cerrado.

Municipality	Locality	State	Latitude	Longitude
Coribe		BA	-13.7587	-44.4187
São Desidério	Estudos da Ferrovia Oeste-Leste/ Roda Velha-BA	BA	-12.7789	-45.8933
Alto Paraíso de Goiás		GO	-14.1622	-47.5233
Alto Paraíso de Goiás	Cerrado rupestre próximo à tapera e Castelinho de pedra, estrada para São Jorge	GO	-14.1631	-47.6193
Cana Brava		GO	-13.5083	-48.3556
Cavalcante	Reserva Bacupari	GO	-13.6424	-47.7217
Cocalzinho de Goiás		GO	-15.6366	-48.5543
Colinas do Sul	LT Serra da Mesa	GO	-13.9903	-48.0922
Minaçu	Cana Brava	GO	-12.7833	-46.8833
Minaçu		GO	-13.8167	-48.3333
Minaçu		GO	-13.4958	-48.3974
Mineiros	Parque Nacional Emas	GO	-17.5591	-52.6749
Monte Alegre de Goiás	Rio Raiz	GO	-13.2500	-46.9000
Monte Alegre de Goiás	Faz. Nossa Senhora do Livramento	GO	-13.2000	-47.1000
Niquelândia		GO	-14.4500	-48.4500
Nova Roma		GO	-13.7500	-46.8833
Pirenópolis		GO	-15.8500	-48.9500
Pirenópolis		GO	-15.8260	-49.0110
Pirenópolis	Cerrado rupestre próximo ao portal vindo de Pirenópolis	GO	-15.8055	-48.8743
Pirenópolis	Cerrado rupestre, estrada para o Morro do Cabeludo	GO	-15.8047	-48.8298
Posse		GO	-14.0833	-46.3333
Rio Verde		GO	-17.8000	-50.9333
São Domingos		GO	-13.4498	-46.4481
Serra da Mesa		GO	-14.2500	-48.5833
Serra Negra		GO	-14.0167	-48.3525
Teresina de Goiás		GO	-13.6938	-47.2399
Alto Parnaíba		MA	-9.1000	-45.9500
Carolina		MA	-7.3339	-47.4147
Carolina		MA	-7.3667	-47.4333
Estreito		MA	-6.5625	-47.4525
Buritis		MG	-15.6178	-46.4233
Januária	PARNA Peruaçu	MG	-15.1233	-44.2401
Unaí		MG	-16.3853	-46.8318
[divisa Canarana/Cocalinho]	São Domingos, Rio das Mortes	MT	-13.5000	-51.4000
Barra do Garças		MT	-15.8833	-52.2500
Barra do Garças	Ponta da Serra do Roncador	MT	-15.3282	-52.2284
Cocalinho		MT	-14.3744	-51.0022
Nova Xavantina	Parque Estadual do Bacaba - Campus Unemat	MT	-14.6858	-52.3358
Nova Xavantina	Rio Noidore	MT	-14.2724	-52.4487
Santa Filomena	Engenheiro Dodt	PI	-8.8000	-45.9333
Almas	RPPN Minnehaha	TO	-11.4737	-47.1211
Almas	EESGT, linha 4	TO	-11.2208	-46.8856
Almas	EESGT, linha mamíferos (rochas)	TO	-11.1792	-46.8396

Municipality	Locality	State	Latitude	Longitude
Almas	EESGT, Ribeirão Cascavel	TO	-11.2442	-46.8126
Barra do Rio São Domingos		TO	-13.4000	-47.2000
Bom Jesus do Tocantins		TO	-9.0022	-47.8624
Caseara	Parque Estadual do Cantão	TO	-9.3723	-49.8430
Combinado	Ferrovia Oeste-Leste	TO	-12.8130	-46.4747
Conceição do Tocantins	Ferrovia Oeste-Leste	TO	-12.4071	-47.1860
Dianópolis		TO	-11.7976	-46.9803
Figueirópolis	Ferrovia Oeste-Leste	TO	-12.1839	-48.9602
Guaraí		TO	-8.8333	-48.5167
Gurupí		TO	-11.7167	-49.0667
Ipueiras		TO	-11.2333	-48.4667
Lajeado	UHE Luís Eduardo Magalhães	TO	-9.8500	-48.3167
Mateiros	Jalapão	TO	-10.7022	-46.4128
Natividade	Torre de Telefone	TO	-11.6938	-47.7016
Palmas		TO	-10.0333	-48.3333
Palmas		TO	-10.1891	-48.1085
Palmas	Taquaruçu	TO	-10.2680	-48.1513
Paranã		TO	-12.6167	-47.8833
Paranã	Ferrovia Oeste-Leste	TO	-12.4847	-47.8137
Pedro Afonso		TO	-9.2044	-48.0168
Peixe		TO	-12.0333	-48.3500
Peixe	Faz. São Francisco	TO	-11.9064	-48.6992
Peixe	UHE Peixe Angical	TO	-12.0350	-48.6006
Pium	Parque Estadual do Cantão	TO	-9.9789	-50.0372
Pium	Piau ou Piaus	TO	-10.4422	-49.1856
Porto Alegre do Tocantins		TO	-11.6767	-46.9858
Porto Nacional		TO	-10.7000	-48.4167
Porto Nacional	Criação Projeto Irrigação	TO	-10.5485	-48.4840
São Salvador do Tocantins		TO	-12.7333	-48.2333

Appendix 27: Locality records of *Micrablepharus atticolus* from the Brazilian Cerrado.

Municipality	Locality	State	Latitude	Longitude
Brasília	Faz. Água Limpa	DF	-15.9726	-47.9099
Brasília	Airport	DF	-15.8820	-47.9241
Brasília	Área Alfa (CIAB) - CAE a	DF	-16.0120	-47.9442
Brasília	Área Alfa (CIAB) - Campo *Vochysia*	DF	-15.9831	-47.9062
Brasília	Área Alfa (CIAB) - CAW b	DF	-16.0160	-47.9558
Brasília	Área Alfa (CIAB) - CEW d	DF	-16.0060	-47.9617
Brasília	Floresta Nacional de Brasília	DF	-15.7579	-48.0599
Brasília	IBGE - Mata de Galeria	DF	-15.9312	-47.8829
Brasília	IBGE - Projeto Fogo	DF	-15.9505	-47.8678
Brasília	Jardim Botânico de Brasília - Mata de Galeria	DF	-15.8818	-47.8415
Brasília	Parque Nacional de Brasília	DF	-15.6968	-47.9250
Brasília		DF	-16.0089	-47.9497
Brasília		DF	-15.7811	-47.7972
Alexânia	Faz. Cafundó	GO	-16.1481	-48.5793
Aporé	PCH Planalto	GO	-18.7883	-52.3753
Aporé		GO	-18.7680	-52.0454
Arenópolis	PCH Mosquitão - Rio Caiapó	GO	-16.3419	-51.4375
Aruanã		GO	-14.8056	-50.9372
Caldas Novas	UHE Corumbá	GO	-17.7900	-48.6000
Caldas Novas		GO	-17.7333	-48.6167
Campo Alegre de Goiás	AHE Serra do Facão	GO	-17.7486	-47.7030
Campo Alegre de Goiás	AHE Serra do Facão	GO	-17.7484	-47.7056
Catalão	AHE Serra do Facão	GO	-17.9185	-47.7085
Catalão	LT Serra da Mesa	GO	-18.3633	-47.9089
Catalão		GO	-17.9305	-47.6657
Chapadão do Céu	Parque Nacional das Emas - Água Ruim	GO	-18.1833	-52.7435
Minaçu	Serra da Mesa	GO	-13.8306	-48.2932
Mineiros	Assentamento Nascentes do Araguaia	GO	-17.6644	-53.2172
Mineiros	Faz. Babilônia	GO	-17.6561	-52.9097
Mineiros	Parque Nacional das Emas	GO	-18.2543	-52.8865
Pirenópolis		GO	-15.8000	-48.8600
Rio Verde		GO	-17.3386	-50.7152
Santa Rita do Araguaia		GO	-17.2272	-53.1580
Santa Rita do Araguaia		GO	-17.2445	-53.0693
Santa Rita do Araguaia		GO	-17.3167	-53.2000
Santa Rita do Araguaia		GO	-17.3000	-53.2000
Serranópolis	Casarão de Pedra	GO	-18.3308	-51.9664
Arinos	RPPN Arara Vermelha and Vereda do Pacari	MG	-15.4461	-45.8251
Curvelo	Curvelo/Pompéu	MG	-19.0260	-44.7090
Formoso	PARNA Grande Sertão Veredas	MG	-15.3070	-45.9414
Paracatu		MG	-17.4000	-47.3000
Uberlândia		MG	-19.0283	-48.3317
Alcinópolis	Faz. Vista Bonita - Line 1	MS	-17.9892	-53.6276
Alcinópolis	Faz. Vista Bonita - Line 2	MS	-17.9952	-53.6334
Alcinópolis	Faz. Vista Bonita - Line 3	MS	-18.0262	-53.6415
Alcinópolis	Faz. Vista Bonita - Line 4	MS	-18.0173	-53.6541

Municipality	Locality	State	Latitude	Longitude
Bataguassu		MS	-21.8082	-52.5707
Campo Grande	Estância Santa Maria	MS	-20.5101	-54.5258
Costa Rica	PE Nascentes do Taquari - Faz. Mutum	MS	-18.2150	-53.3121
Coxim	Diamante Farm - Jauru district	MS	-18.6916	-54.4110
Três Lagoas	Faz. Barra da Moeda	MS	-20.9829	-51.7892
Três Lagoas	Faz. Canaã	MS	-20.4700	-52.0000
Alto Araguaia	Faz. Bacuri, Bálsamo, and Córrego Fundo	MT	-17.2600	-53.3000
Alto Araguaia	Córrego do Sapo	MT	-17.5470	-53.3209
Alto Araguaia	Faz. Saramandaia - Line 10	MT	-17.9042	-53.4663
Barra do Garças	PCH Toricoejo	MT	-15.2536	-53.1239
Barra do Garças		MT	-15.3597	-52.4971
Brasnorte		MT	-12.4254	-57.9955
Canabrava do Norte	BR 158	MT	-11.2359	-51.6867
Canarana	Faz. Peixe Boi	MT	-13.3313	-52.3987
Chapada dos Guimarães	APM Manso	MT	-14.9100	-55.7000
Chapada dos Guimarães		MT	-15.1071	-55.5396
Chapada dos Guimarães		MT	-14.9934	-55.8865
Chapada dos Guimarães		MT	-15.2700	-55.8400
Cocalinho		MT	-13.8687	-51.1474
Cuiabá	Ribeirão do Forte	MT	-15.3340	-55.9470
Gaúcha do Norte		MT	-13.2333	-53.0667
Nova Lacerda		MT	-14.3005	-59.7713
Nova Nazaré	Pindaíba	MT	-14.3667	-51.7167
Nova Xavantina	Rancho Ponte de Pedra Farm	MT	-14.7951	-52.6417
Nova Xavantina	UNEMAT	MT	-14.6988	-52.3509
Novo Santo Antônio	Parque Estadual do Araguaia	MT	-12.3849	-50.8934
Novo São Joaquim	AHE Agua Limpa	MT	-15.2930	-53.7610
Paranatinga		MT	-13.6768	-54.1087
Ribeirão Cascalheira	BR 158	MT	-12.8488	-51.7459
Ribeirão Cascalheira	Serra do Roncador - Aldeia de Caça	MT	-12.8500	-51.7500
Santa Terezinha	Tapirapé River	MT	-10.6833	-50.6333
São Félix do Araguaia	São Domingos - Rio das Mortes	MT	-11.7500	-50.7333
São José do Rio Claro		MT	-13.5962	-56.7945
Sapezal	UHE Cachoeirão	MT	-13.5333	-58.8000
Sapezal		MT	-13.1522	-58.6535
Sorriso	Boa Esperança	MT	-13.5091	-55.1425
Conceição do Araguaia	São José da Fortaleza Farm - Mata do Buraco	PA	-8.1425	-49.3391
Conceição do Araguaia		PA	-8.1825	-49.5126
Palestina do Pará	UHE Santa Isabel - Pedral do Araguaia	PA	-6.1236	-48.4125
Santana do Araguaia		PA	-9.7271	-50.1813
Santana do Araguaia		PA	-9.6812	-50.1581
Pimenta Bueno		RO	-11.8088	-60.7230
Vilhena	BR-399 km 21-23/km 53-55	RO	-12.4700	-60.5200
Vilhena		RO	-12.4733	-60.2896
Águas de Santa Bárbara	Estação Ecológica Águas de Santa Bárbara	SP	-22.7900	-49.2432
Águas de Santa Bárbara	Estação Ecológica Águas de Santa Bárbara	SP	-22.7870	-49.2346
Angatuba	Faz. Três Lagoas	SP	-23.3831	-48.4536

Municipality	Locality	State	Latitude	Longitude
Cajuru		SP	-21.2757	-47.3101
Descalvado		SP	-21.8810	-47.6541
Santa Rita do Passa Quatro	ARIE Pé do Gigante	SP	-21.6392	-47.6436
Santa Rita do Passa Quatro		SP	-21.6833	-47.4833
São José do Rio Preto		SP	-20.8200	-49.3789
Teodoro Sampaio	Parque Estadual do Morro do Diabo	SP	-22.5833	-52.3000
Caseara		TO	-9.4004	-49.8480
Lagoa da Confusão	Ilha do Bananal	TO	-11.3000	-50.2800
Peixe	UHE Peixe Angical	TO	-12.0300	-48.5500
Pium	Ilha do Bananal	TO	-10.4541	-50.4723
Pium	Parque Estadual do Cantão	TO	-9.9789	-50.0372
Pium	Parque Estadual do Cantão	TO	-9.3951	-50.0015
Pium	Parque Nacional do Araguaia	TO	-10.3860	-50.1330
São Salvador do Tocantins		TO	-12.4858	-48.2650

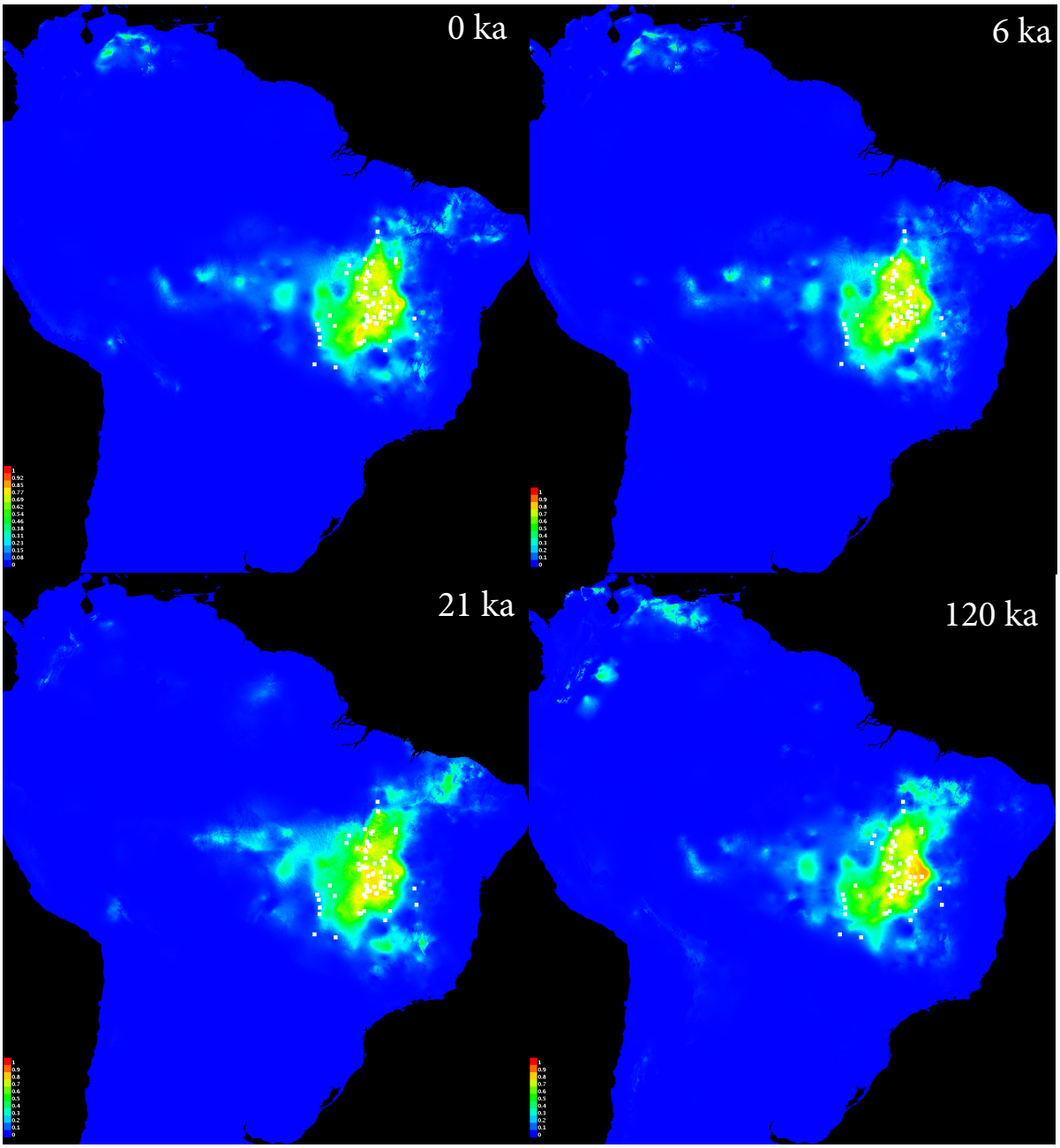
Appendix 28: Locality records of *Tropidurus itambere* from the Brazilian Cerrado.

Municipality	Locality	State	Latitude	Longitude
Brasília	Faz. Água Limpa	DF	-15.7700	-47.9300
Alto Paraíso de Goiás	Cerrado rupestre em frente ao Portal da Chapada	GO	-14.1524	-47.5963
Cristalina	Cerrado Rupestre próximo à BR 040	GO	-16.7892	-47.5765
Alto Paraíso de Goiás	Cerrado rupestre próximo à tapera e Castelinho de pedra, estrada para São Jorge	GO	-14.1631	-47.6193
Pirenópolis	Parque Estadual dos Pirineus - Murinho de Pedra	GO	-15.8068	-48.8488
Cristalina	Pedras próximas à BR 040	GO	-16.7342	-47.6208
Pirenópolis	Serra dos Pireneus	GO	-15.8500	-48.9500
Água Limpa		GO	-18.0700	-48.7600
Alto Paraíso de Goiás		GO	-14.1622	-47.5233
Aporé		GO	-18.9600	-51.9200
Baliza		GO	-16.1900	-52.5400
Caldas Novas		GO	-17.7400	-48.6200
Catalão		GO	-18.1700	-47.9400
Cocalzinho		GO	-15.6366	-48.5543
Colinas do Sul		GO	-13.9903	-48.0922
Cristalina		GO	-16.7600	-47.6100
Flores de Goiás		GO	-14.4400	-47.0400
Jataí		GO	-17.8800	-51.7100
Minaçu		GO	-13.4958	-48.3974
Novo Gama		GO	-16.0500	-48.0300
Pirenópolis		GO	-15.8260	-49.0110
Planaltina de goias		GO	-15.4500	-47.6100
Santo Antonio do Descoberto		GO	-15.9300	-48.2500
São Domingos		GO	-13.4498	-46.4481
São João d'Aliança		GO	-14.7000	-47.5200
Lima Duarte	Parque Estadual do Ibitipoca	MG	-21.8400	-43.7900
Parque Nacional Grande Sertão Veredas	Parque Nacional Grande Sertão Veredas	MG	-15.2233	-45.8122
Três Marias	Ponte de pedra - Condominio Morada dos Peixes	MG	-18.1901	-45.3249
Ingaí	Reserva Biológica Unilavras Boqueirão	MG	-21.3464	-44.9908
Arcos	São Julião	MG	-20.2800	-45.5300
Ouro Branco	Serra do Ouro Branco	MG	-20.5200	-43.6900
São Thomé das Letras	Sobradinho	MG	-21.6613	-44.8870

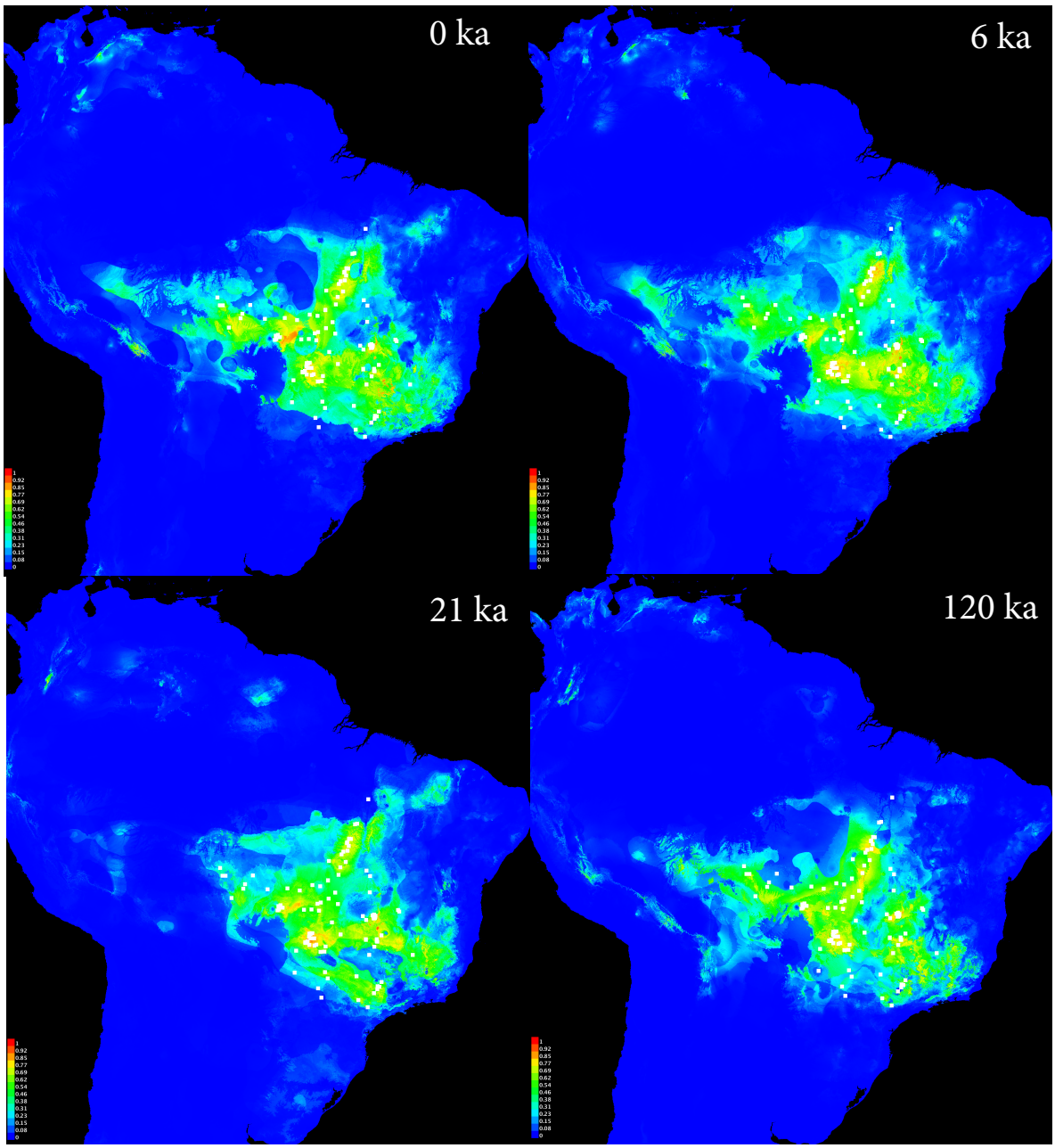
Municipality	Locality	State	Latitude	Longitude
Lima Duarte	Vila do Mogol	MG	-21.7403	-43.8520
Campos Altos		MG	-19.6900	-46.1700
Chapada Gaúcha		MG	-15.3000	-45.6100
Itapeva		MG	-22.7600	-46.2200
Paracatu		MG	-17.2200	-46.8700
Três Marias		MG	-18.2000	-45.2400
Unai		MG	-16.3853	-46.8318
Alcinópolis	Parque Natural Municipal Templo dos Pilares	MS	-18.1494	-53.6778
Miranda	Salobra	MS	-20.2400	-56.3700
Alcinópolis		MS	-18.3200	-53.7000
Aquidauana		MS	-20.4700	-55.7800
Bodoquena		MS	-20.5300	-56.7100
Bonito		MS	-21.1200	-56.4800
Cassilândia		MS	-19.1100	-51.7300
Três Lagoas		MS	-20.7500	-51.6700
Barra do Garças	Parque Estadual da Serra Azul	MT	-15.8578	-52.2545
Alto Araguaia		MT	-17.3100	-53.2100
Barra do Bugres		MT	-15.0700	-57.1800
Barra do Garças		MT	-15.2000	-52.5000
Chapada dos Guimarães		MT	-15.4600	-55.7500
Cuiabá		MT	-15.5900	-56.0900
Itiquira		MT	-17.2000	-54.1500
Nova Bandeirantes		MT	-10.3466	-57.6891
Nova Monte Verde		MT	-10.3322	-57.6505
Nova Xavantina		MT	-14.6858	-52.3358
Ribeirão Cascalheira		MT	**-13.0031**	**-51.7533**
Ribeirão Cascalheira		MT	-12.9400	-51.8200
Ribeirão Cascalheira		MT	-12.4700	-52.3700
São Geraldo do Araguaia		PA	-6.1691	-48.7872
Tibaji		PR	-24.5000	-50.4100
Campinas	Faz. Manga	SP	-23.0000	-47.0000
Botucatu	Rubião Júnior	SP	-22.8800	-48.4400
São Luís do Paraitinga	Serra de Itambé	SP	-23.2200	-45.3100
Adolfo		SP	-21.2300	-49.6400

Municipality	Locality	State	Latitude	Longitude
Araguaquara		SP	-21.7900	-48.1700
Atibaia		SP	-23.1582	-46.5342
Barretos		SP	-20.5500	-48.5600
Cabrália Paulista		SP	-22.4500	-49.3300
Franca		SP	-20.5300	-47.4000
Ituverava		SP	-20.3300	-47.7800
Jaboticabal		SP	-21.2500	-48.3200
Maracaí		SP	-21.6100	-50.6600
Nova Europa		SP	-21.7700	-48.5600
Penápolis		SP	-21.4200	-50.0700
Piedade		SP	-23.7100	-47.4200
Piraju		SP	-23.1900	-49.3800
Ribeirão Preto		SP	-21.1700	-47.8100
Rio Claro		SP	-22.4100	-47.5600
São Carlos		SP	-22.0100	-47.8900
São João da Boa Vista		SP	-21.9600	-46.7900
São Roque		SP	-23.5200	-47.1300
Socorro		SP	-22.5900	-46.5200
Sorocaba		SP	-23.5000	-47.4500
Tabatinga		SP	-21.7100	-48.6800
Valinhos		SP	-22.9333	-46.9167
Valinhos		SP	-22.9700	-46.9900
Valinhos		SP	-22.9333	-46.9166
Vinhedo		SP	-23.0300	-46.9700
Vista Alegre do Alto		SP	-21.1700	-48.6200
Votorantim		SP	-23.5400	-47.4300
Peixe		TO	-12.0333	-48.3500
São Salvador do Tocantins		TO	-12.7333	-48.2333

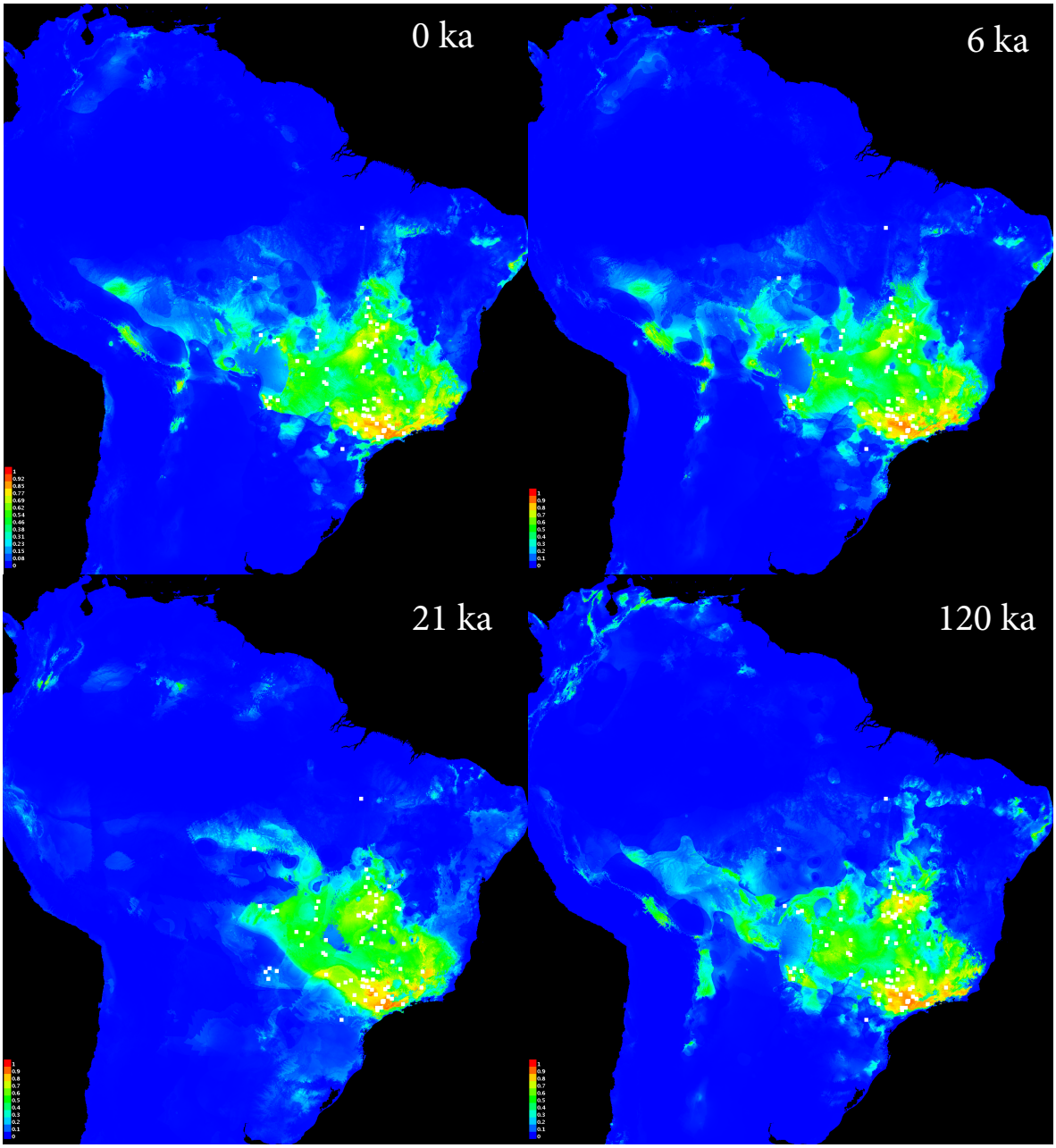
Appendix 29: Species distribution models of *Gymnodactylus amarali* under past and current environmental conditions. SDMs shown are from current (0 ka) climate, mid-Holocene (6 ka), Last Glacial Maximum (LGM, 21 ka), and Last Interglacial (LIG, 120 ka).



Appendix 30: Species distribution models of *Micrablepharus atticolus* under past and current environmental conditions. SDMs shown are from current (0 ka) climate, mid-Holocene (6 ka), Last Glacial Maximum (LGM, 21 ka), and Last Interglacial (LIG, 120 ka).



Appendix 31: Species distribution models of *Tropidurus itambere* under past and current environmental conditions. SDMs shown are from current (0 ka) climate, mid-Holocene (6 ka), Last Glacial Maximum (LGM, 21 ka), and Last Interglacial (LIG, 120 ka).



References

Ab' Sáber, A.N. (1954) O Planalto dos Parecis, na região de Diamantino (Mato Grosso). *Boletim Paulista de Geografia*, **17**, 63–69.

Ab'Sáber, A.N. (1974) O domínio morfoclimático semiárido das caatingas Brasileiras. *Geomorfologia*, **43**, 1–19.

Ab'Sáber, A.N. (1998) Participação das depressões periféricas e superfícies aplainadas na compartimentação do planalto brasileiro - considerações finais e conclusões. *Revista do Instituto Geológico*, **19**, 51–69.

Aberer, A.J., Kobert, K. & Stamatakis, A. (2014) ExaBayes: Massively parallel Bayesian tree inference for the whole-genome era. *Molecular Biology and Evolution*, **31**, 2553–2556.

Aldrich, S., Walker, R., Simmons, C., Caldas, M. & Perz, S. (2012) Contentious Land Change in the Amazon's Arc of Deforestation. *Annals of the Association of American Geographers*, **102**, 103-128.

Almeida, F.C., Bonvicino, C.R. & Cordeiro-Estrela, P. (2007) Phylogeny and temporal diversification of Calomys (Rodentia, Sigmodontinae): implications for the biogeography of an endemic genus of the open/dry biomes of South America. *Molecular Phylogenetics and Evolution*, **42**, 449–66.

Andrew, R.L., Bernatchez, L., Bonin, A., Buerkle, C.A., Carstens, B.C., Emerson, B.C., Garant, D., Giraud, T., Kane, N.C., Rogers, S.M., Slate, J., Smith, H., Sork, V.L., Stone, G.N., Vines, T.H., Waits, L., Widmer, A. & Rieseberg, L.H. (2013) A road map for molecular ecology. *Molecular Ecology*, **22**, 2605–26.

Antonelli, A. & Sanmartín, I. (2011) Why are there so many plant species in the Neotropics? *Taxon*, **60**, 403–414.

Arbogast, B.S. & Kenagy, G.J. (2001) Comparative Phylogeography as an Integrative Approach to Historical Biogeography. *Journal of Biogeography*, 819–825.

Arias, F., de Carvalho, C.M., Zaher, H. & Rodrigues, M.T. (2014a) A New Species of *Ameivula* (Squamata, Teiidae) from Southern Espinhaço Mountain Range, Brazil. *Copeia*, **2014**, 95-105.

Arias, Federico J., Teixeira, M., de Carvalho, Celso M., Recoder, R., Zaher, H. & Rodrigues, Miguel T. (2014b) Whiptail lizards in South America: a new *Ameivula* (Squamata, Teiidae) from Planalto dos Gerais, Eastern Brazilian Cerrado. *Amphibia-Reptilia*, **35**, 227-242.

Arnegard, M.E., McGee, M.D., Matthews, B., Marchinko, K.B., Conte, G.L., Kabir, S., Bedford, N., Bergek, S., Chan, Y.F., Jones, F.C., Kingsley, D.M., Peichel, C.L. & Schluter, D. (2014) Genetics of ecological divergence during speciation. *Nature*, **511**, 307-11.

Austin, M. (2007) Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecological Modelling*, **200**, 1–19.

Avise, J.C. (1992) Molecular Population Structure and the Biogeographic History of a Regional Fauna: A Case History with Lessons for Conservation Biology. *Oikos*, **63**, 62–76.

Avise, J.C. (1998) The history and purview of phylogeography: a personal reflection. *Molecular Ecology*, **7**, 371–379.

Avise, J.C. (2001) *Phylogeography: the History and Formation of Species*. Harvard University Press, Cambridge.

Avise, J.C., Walker, D.E. & Johns, G.C. (1998) Speciation durations and Pleistocene effects on vertebrate phylogeography. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **265**, 1707–1712.

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

Baby, P., Rochat, P., Mascle, G. & Hérail, G. (1997) Neogene shortening contribution to crustal thickening in the back arc of the Central Andes. *Geology*, **25**, 883.

Báez, A.M. & de Gasparini, Z.B. (1979) The South American herpetofauna: an evaluation of the fossil record. *The South American herpetofauna: its origin, evolution and dispersal* (ed. by W.E. Duellman), pp. 29–55. Museum of Natural History, The University of Kansas.

Bagley, J., Sandel, M., Travis, J., Lozano-Vilano, M. & Johnson, J. (2013) Paleoclimatic modeling and phylogeography of least killifish, *Heterandria formosa*: insights into Pleistocene expansion-contraction dynamics and evolutionary history of North American Coastal Plain freshwater biota. *BMC Evolutionary Biology*, **13**, 223.

Bagley, J.C. & Johnson, J.B. (2014a) Phylogeography and biogeography of the lower Central American Neotropics: diversification between two continents and between two seas. *Biological Reviews*, **89**, 767–790.

Bagley, J.C. & Johnson, J.B. (2014b) Testing for shared biogeographic history in the lower Central American freshwater fish assemblage using comparative phylogeography: concerted, independent, or multiple evolutionary responses? *Ecology and evolution*, **4**, 1686–1705.

Balian, E.V., Segers, H., Lévèque, C. & Martens, K. (2007) The Freshwater Animal Diversity Assessment: an overview of the results. *Hydrobiologia*, **595**, 627–637.

Barber, B. & Klicka, J. (2010) Two pulses of diversification across the Isthmus of Tehuantepec in a montane Mexican bird fauna. *Proceedings of the Royal Society B: Biological Sciences*, **277**, 2675–2681.

Barbosa, A.R., Fiorini, C.F., Silva-Pereira, V., Mello-Silva, R. & Borba, E.L. (2012) Geographical Genetic Structuring and Phenotypic Variation in the Vellozia Hirsuta (Velloziaceae) Ochlospecies Complex. *American Journal of Botany*, **99**, 1477-1488.

Bauer, A.M., Parham, J.F., Brown, R.M., Stuart, B.L., Grismer, L., Papenfuss, T.J., Böhme, W., Savage, J.M., Carranza, S. & Grismer, J.L. (2011) Availability of new Bayesian-delimited gecko names and the importance of character-based species descriptions. *Proceedings of the Royal Society B: Biological Sciences*, **278**, 490–492.

Beaumont, M.A., Zhang, W. & Balding, D.J. (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–2035.

Beerli, P. & Felsenstein, J. (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 4563–8.

Beheregaray, L.B. (2008) Twenty years of phylogeography: the state of the field and the challenges for the Southern Hemisphere. *Molecular Ecology*, **17**, 3754–3774.

Beheregaray, L.B. & Caccone, A. (2007) Cryptic biodiversity in a changing world. *Journal of Biology*, **6**, 9.

Beheregaray, L.B., Cooke, G., Chao, N. & Landguth, E.L. (2015) Ecological speciation in the tropics: Insights from comparative genetic studies in Amazonia. *Frontiers in Genetics*, **5:477**

Bell, R.C., MacKenzie, J.B., Hickerson, M.J., Chavarrîa, K.L., Cunningham, M., Williams, S. & Moritz, C. (2012) Comparative multi-locus phylogeography confirms multiple vicariance events in co-distributed rainforest frogs. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 991–999.

Ben-Hur, A., Ong, C.S., Sonnenburg, S., Schölkopf, B. & Rätsch, G. (2008) Support Vector Machines and Kernels for Computational Biology. *PLoS Computational Biolology: Education*, **4**, e1000173.

Bermingham, E. & Moritz, C. (1998) Comparative phylogeography: concepts and applications. *Molecular Ecology*, **7**, 367–369.

Bernatchez, L. & Wilson, C.C. (1998) Comparative phylogeography of Nearctic and Palearctic fishes. *Molecular Ecology*, **7**, 431-452.

Bichteler, J. (1991) Geologists and Gray Literature. *Science & Technology Libraries*, **11**, 39–50.

Bickford, D., Lohman, D.J., Sodhi, N.S., Ng, P.K., Meier, R., Winker, K., Ingram, K.K. & Das, I. (2007) Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, **22**, 148–55.

Bini, L.M., Diniz-Filho, J.A.F., Rangel, T.F.L.V.B., Bastos, R.P. & Pinto, M.P. (2006) Challenging Wallacean and Linnean shortfalls: knowledge gradients and conservation planning in a biodiversity hotspot. *Diversity and Distributions*, **12**, 475–482.

Blach-Overgaard, A., Kissling, W.D., Dransfield, J., Balslev, H. & Svenning, J.C. (2013) Multimillion-year climatic effects on palm species diversity in Africa. *Ecology*, **94**, 2426–2435.

Blair, C., Mendez de la Cruz, F.R., Law, C. & Murphy, R.W. (2015) Molecular phylogenetics and species delimitation of leaf-toed geckos (Phyllodactylidae: *Phyllodactylus*) throughout the Mexican tropical dry forest. *Molecular Phylogenetics and Evolution*, **84**, 254–265.

Bokma, F. (2009) Problems detecting density-dependent diversification on phylogenies. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 993–994.

Bonatelli, I.A., Perez, M.F., Peterson, A.T., Taylor, N.P., Zappi, D.C., Machado, M.C., Koch, I., Pires, A.H. & Moraes, E.M. (2014) Interglacial microrefugia and diversification of

a cactus species complex: phylogeography and palaeodistributional reconstructions for *Pilosocereus aurisetus* and allies. *Molecular Ecology*, **23**, 3044–3063.

Borcard, D. & Legendre, P. (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, **153**, 51–68.

Borcard, D., Legendre, P. & Drapeau, P. (1992) Partialling out the spatial component of ecological variation. *Ecology*, **73**, 1045–1055.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A. & Drummond, A.J. (2014) BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Computional Biology*, **10**, e1003537.

Brandley, M.C., Bragg, J.G., Singhal, S., Chapple, D.G., Jennings, C.K., Lemmon, A.R., Lemmon, E.M., Thompson, M.B. & Moritz, C. (2015) Evaluating the performance of anchored hybrid enrichment at the tips of the tree of life: a phylogenetic analysis of Australian *Eugongylus* group scincid lizards. *BMC Evolutionary Biology*, **15**, 62–.

Bridgewater, S., Ratter, J.A. & Ribeiro, J.F. (2004) Biogeographic patterns, beta-diversity and dominance in the cerrado biome of Brazil. *Biodiversity and Conservation*, **13**, 2295–2318.

Brito, D. (2010) Overcoming the Linnean shortfall: Data deficiency and biological survey priorities. *Basic and Applied Ecology*, **11**, 709-713.

Brown, J.H. & Lomolino, M.V. (1998) *Biogeography*, 2nd edn. Sinauer Associates, Sunderland, MA.

Browning, S.R. & Browning, B.L. (2011) Haplotype phasing: existing methods and new developments. *Nature reviews. Genetics*, **12**, 703-14.

Buffon, G.L.L., Comte de (1761) *Histoire naturelle générale et particuliére*. Imprimerie Royale, Paris.

Burbrink, F.T., Yao, H., Ingrasci, M., Bryson, R.W., Jr., Guiher, T.J. & Ruane, S. (2011) Speciation at the Mogollon Rim in the Arizona Mountain Kingsnake (Lampropeltis pyromelana). *Molecular Phylogenetics and Evolution*, **60**, 445–54.

Bush, M.B. (1994) Amazonian speciation: a necessarily complex model. *Journal of Biogeography*, **21**, 5–17.

Buuren, S.V. & Groothuis-Oudshoorn, K. (2011) MICE: Multivariate Imputation by Chained Equations in R. *Journal of statistical software*, **45**, 1–67.

Byron, K.W. (1983) Some Observations of the Use of Discriminant Analysis in Ecology. *Ecology*, **64**, 1283-1291.

Camargo, A., Sinervo, B. & Sites Jr., J.W. (2010) Lizards as model organisms for linking phylogeographic and speciation studies. *Molecular Ecology*, **19**, 3250–3270.

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

Camolez, T. & Zaher, H. (2010) Levantamento, identificação e descrição da fauna de Squamata do Quaternário brasileiro (Lepidosauria). *Arquivos de Zoologia*, **41**, 1–96.

Candeiro, C.R.A. (2007) Paleogeographic distribution of the terrestrial squamate reptiles from the cretaceous of Brazil. *Bioscience Journal*, **23**, 65–74.

Candeiro, C.R.A., Nava, W., Martinelli, A.G., Forasiepi, A.M., Scanferla, C.A. & Muzzopappa, P. (2009) New lizard Record (Diapsida, Lepidosauria) from the Upper Cretaceous Adamantina Formation, Brazil. *Bulletin of Geosciences*, **84**, 573–576.

Candolle, A.P.d. (1855) *Géographie Botanique Raisonée*. V. Masson, Paris.

Carnaval, A.C. & Moritz, C. (2008) Historical climate modelling predicts patterns of current biodiversity in the Brazilian Atlantic forest. *Journal of Biogeography*, **35**, 1187–1201.

Carnaval, A.C., Hickerson, M.J., Haddad, C.F.B., Rodrigues, M.T. & Moritz, C. (2009) Stability predicts genetic diversity in the Brazilian Atlantic Forest hotspot. *Science*, **323**, 785–789.

Carnaval, A.C., Waltari, E., Rodrigues, M.T., Rosauer, D., VanDerWal, J., Damasceno, R., Prates, I., Strangas, M., Spanos, Z., Rivera, D., Pie, M.R., Firkowski, C.R., Bornschein, M.R., Ribeiro, L.F. & Moritz, C. (2014) Prediction of phylogeographic endemism in an environmentally complex biome. *Proceedings of the Royal Society B: Biological Sciences*, **281**, 20141461.

Carstens, B., Lemmon, A.R. & Lemmon, E.M. (2012) The promises and pitfalls of next-generation sequencing data in phylogeography. *Systematic Biology*, **61**, 713-5.

Carstens, B.C. & Richards, C.L. (2007) Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution*, **61**, 1439–1454.

Carstens, B.C., Pelletier, T.A., Reid, N.M. & Satler, J.D. (2013) How to fail at species delimitation. *Molecular Ecology*, **22**, 4369–83.

Carvalho, A.L.G. (2013) On the distribution and conservation of the South American lizard genus *Tropidurus* Wied-Neuwied, 1825 (Squamata: Tropiduridae). *Zootaxa*, **3640**, 42–56.

Carvalho, A.L.G., de Britto, M.R. & Fernandes, D.S. (2013) Biogeography of the lizard genus *Tropidurus* Wied-Neuwied, 1825 (Squamata: Tropiduridae): distribution, endemism, and area relationships in South america. *PloS one*, **8**, e59736.

Cassimiro, J. & Rodrigues, M.T. (2009) A new species of lizard genus *Gymnodactylus* Spix, 1825 (Squamata: Gekkota: Phyllodactylidae) from Serra do Sincorá, northeastern Brazil, and the status of *G. carvalhoi* Vanzolini, 2005. *Zootaxa*, **2008**, 38–52.

Castoe, T., Doan, T. & Parkinson, C. (2004) Data partitions and complex models in Bayesian analysis: The phylogeny of Gymnophthalmid lizards. *Systematic Biology*, **53**, 448–469.

Castro, A.A.J.F., Martins, F.R., Tamashiro, J.Y. & Shepherd, G.J. (1999) How rich is the flora of Brazilian cerrados? *Annals of the Missouri Botanical Garden*, **86**, 192–224.

Cavalcanti, R.B. & Joly, C.A. (2002) Biodiversity and conservation priorities in the Cerrado region. *The Cerrados of Brazil: Ecology and Natural History of a Neotropical Savanna* (ed. by P.S. Oliveira and R.J. Marquis), pp. 351–367. Columbia University Press, New York.

Ceccarelli, F.S., Sharkey, M.J. & Zaldivar-Riveron, A. (2012) Species identification in the taxonomically neglected, highly diverse, neotropical parasitoid wasp genus *Notiospathius* (Braconidae: Doryctinae) based on an integrative molecular and morphological approach. *Molecular Phylogenetics and Evolution*, **62**, 485–95.

Chang, C.C. & Lin, C.J. (2011) LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, **2**, 1–27.

Collevatti, R.G., Grattapaglia, D. & Hay, J.D. (2003) Evidences for multiple maternal lineages of *Caryocar brasiliense* populations in the Brazilian Cerrado based on the analysis of chloroplast DNA sequences and microsatellite haplotype variation. *Molecular Ecology*, **12**, 105–115.

Collevatti, R.G., Rabelo, S.G. & Vieira, R.F. (2009) Phylogeography and disjunct distribution in *Lychnophora ericoides* (Asteraceae), an endangered cerrado shrub species. *Annals of Botany*, **104**, 655-664.

Collevatti, R.G., de Castro, T.G., Lima, J.D. & Telles, M.P.D. (2012a) Phylogeography of *Tibouchina papyrus* (Pohl) Toledo (Melastomataceae), an endangered tree species from rocky savannas, suggests bidirectional expansion due to climate cooling in the Pleistocene. *Ecology and evolution*, **2**, 1024-1035.

Collevatti, R.G., Terribile, L.C., Diniz-Filho, J.A. & Lima-Ribeiro, M.S. (2015) Multi-model inference in comparative phylogeography: an integrative approach based on multiple lines of evidence. *Frontiers in genetics*, **6**, 31.

Collevatti, R.G., Lima-Ribeiro, M.S., Souza-Neto, A.C., Franco, A.A., de Oliveira, G. & Terribile, L.C. (2012b) Recovering the Demographical History of a Brazilian Cerrado Tree Species *Caryocar brasiliense*: Coupling Ecological Niche Modeling and Coalescent Analyses. *Natureza & Conservacao*, **10**, 169-176.

Collevatti, R.G., Terribile, L.C., Lima-Ribeiro, M.S., Nabout, J.C., Oliveira, G., Rangel, T.F., Rabelo, S.G. & Diniz, J.A.F. (2012c) A coupled phylogeographical and species distribution modelling approach recovers the demographical history of a Neotropical seasonally dry forest tree species. *Molecular Ecology*, **21**, 5845-5863.

Collevatti, R.G., Terribile, L.C., de Oliveira, G., Lima-Ribeiro, M.S., Nabout, J.C., Rangel, T.F., Diniz-Filho, J.A.F. & Pearson, R. (2013) Drawbacks to palaeodistribution modelling: the case of South American seasonally dry forests. *Journal of Biogeography*, **40**, 345–358.

Colli, G.R. (2005) As origens e a diversificação da herpetofauna do Cerrado. *Cerrado: Ecologia, Biodiversidade e Conservação.* (ed. by A. Scariot, J.C. Souza-Silva and J.M. Felfili), pp. 247–264. Ministério do Meio Ambiente, Brasília.

Colli, G.R., Bastos, R.P. & Araujo, A.F.B. (2002) The character and dynamics of the Cerrado herpetofauna. *The Cerrados of Brazil: Ecology and Natural History of a Neotropical Savanna* (ed. by P.S. Oliveira and R.J. Marquis), pp. 223–241. Columbia University Press, New York.

Colli, G.R., Mesquita, D.O., Rodrigues, P.V.V. & Kitayama, K. (2003a) Ecology of the gecko *Gymnodactylus geckoides amarali* in a Neotropical Savanna. *Journal of Herpetology*, **37**, 694–706.

Colli, G.R., Giugliano, L.G., Mesquita, D.O. & França, F.G.R. (2009) A new species of *Cnemidophorus* from the Jalapão region, in the central Brazilian Cerrado. *Herpetologica*, **65**, 311–327.

Colli, G.R., Costa, G.C., Garda, A.A., Kopp, K.A., Mesquita, D.O., Péres Jr, A.K., Valdujo, P.H., Vieira, G.H.C. & Wiederhecker, H.C. (2003b) A critically endangered new species of *Cnemidophorus* (Squamata, Teiidae) from a Cerrado enclave in southwestern Amazonia, Brazil. *Herpetologica*, **59**, 76–88.

Colli, G.R., Caldwell, J.P., Costa, G.C., Gainsbury, A.M., Garda, A.A., Mesquita, D.O., Filho, C.M.M., Soares, A.H.B., Silva, V.N., Valdujo, P.H., Vieira, G.H.C., Vitt, L.J., Werneck, F.P., Wiederhecker, H.C. & Zatz, M.G. (2003c) A new species of *Cnemidophorus* (Squamata, Teiidae) from the Cerrado biome in central Brazil. *Occasional Papers Of The Oklahoma Museum Of Natural History*, **14**, 1–14.

Condit, R., Ashton, P.S., Baker, P., Bunyavejchewin, S., Gunatilleke, S., Gunatilleke, N., Hubbell, S.P., Foster, R.B., Itoh, A. & LaFrankie, J.V. (2000) Spatial patterns in the distribution of tropical tree species. *Science*, **288**, 1414.

Condon, M.A., Scheffer, S.J., Lewis, M.L. & Swensen, S.M. (2008) Hidden neotropical diversity: greater than the sum of its parts. *Science*, **320**, 928-31.

Cooke, G.M., Landguth, E.L. & Beheregaray, L.B. (2014) Riverscape genetics identifies replicated ecological divergence across an Amazonian ecotone. *Evolution*, **68**, 1947–1960.

Cortes, C. & Vapnik, V. (1995) Support-vector networks. *Machine learning*, **20**, 273–297.

Costa, G.C., Nogueira, C., Machado, R.B. & Colli, G.R. (2007) Squamate richness in the Brazilian Cerrado and its environmental climatic associations. *Diversity and Distributions*, **13**, 714–724.

Costa, G.C., Nogueira, C., Machado, R.B. & Colli, G.R. (2010) Sampling bias and the use of ecological niche modeling in conservation planning: a field evaluation in a biodiversity hotspot. *Biodiversity and Conservation*, **19**, 883–899.

Costa, G.C., Wolfe, C., Shepard, D.B., Caldwell, J.P. & Vitt, L.J. (2008) Detecting the influence of climatic variables on species distributions: a test using GIS niche-based

models along a steep longitudinal environmental gradient. *Journal of Biogeography*, **35**, 637–646.

Costa, H.C. & Bérnils, R.S. (2014) Répteis brasileiros: lista de espécies. *Herpetologia Brasileira*, **3**, 74–84.

Costello, M.J., May, R.M. & Stork, N.E. (2013) Can we name Earth's species before they go extinct? *Science*, **339**, 413–6.

Cracraft, J. (2002) The seven great questions of systematic biology: an essential foundation for conservation and the sustainable use of biodiversity. *Annals of the Missouri Botanical Garden*, 127–144.

Crawford, A.J., Lips, K.R. & Bermingham, E. (2010) Epidemic disease decimates amphibian abundance, species diversity, and evolutionary history in the highlands of central Panama. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 13777-82.

Crisp, M.D., Arroyo, M.T., Cook, L.G., Gandolfo, M.A., Jordan, G.J., McGlone, M.S., Weston, P.H., Westoby, M., Wilf, P. & Linder, H.P. (2009) Phylogenetic biome conservatism on a global scale. *Nature*, **458**, 754–756.

Cushman, S.A. & Landguth, E.L. (2010) Spurious correlations and inference in landscape genetics. *Molecular Ecology*, **19**, 3592-602.

de Freitas, J.L., Strüssmann, C., de Carvalho, M.A., Kawashita-Reibeiro, R.A. & Mott, T. (2011) A new species of *Bachia* Gray, 1845 (Squamata: Gymnophthalmidae) from the Cerrado of midwestern Brazil. *Zootaxa*, **2737**, 61–68.

de Lima, N.E., Lima-Ribeiro, M.S., Tinoco, C.F., Terribile, L.C. & Collevatti, R.G. (2014a) Phylogeography and ecological niche modelling, coupled with the fossil pollen record, unravel the demographic history of a Neotropical swamp palm through the Quaternary. *Journal of Biogeography*, **41**, 673–686.

de Lima, N.E., Lima-Ribeiro, M.S., Tinoco, C.F., Terribile, L.C., Collevatti, R.G. & Svenning, J.C. (2014b) Phylogeography and ecological niche modelling, coupled with the fossil pollen record, unravel the demographic history of a Neotropical swamp palm through the Quaternary. *Journal of Biogeography*, n/a–n/a.

de Queiroz, K. (1998) The general lineage concept of species, species criteria, and the process of speciation: A conceptual unification and terminological recommendations. *Endless forms: Species and speciation* (ed. by D.J. Howard and S.H. Berlocher), pp. 57–75. Oxford University Press, New York.

de Queiroz, K. (2007) Species concepts and species delimitation. *Systematic Biology*, **56**, 879–86.

de Ré, F.C., Gustani, E.C., Oliveira, A.P.F., Machado, L.P.B., Mateus, R.P., Loreto, E.L.S. & Robe, L.J. (2014) Brazilian populations of *Drosophila maculifrons* (Diptera: Drosophilidae): low diversity levels and signals of a population expansion after the Last Glacial Maximum. *Biological Journal of the Linnean Society*, **112**, 55–66.

Degnan, J.H. & Rosenberg, N.A. (2006) Discordance of species trees with their most likely gene trees. *PLoS genetics*, **2**, e68.

Degnan, J.H. & Rosenberg, N.A. (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution*, **24**, 332–40.

Delgado-Salinas, A., Bibler, R. & Lavin, M. (2006) Phylogeny of the genus *Phaseolus* (Leguminosae): a recent diversification in an ancient landscape. *Systematic Botany*, **31**, 779–791.

Diniz-Filho, J.A., Loyola, R.D., Raia, P., Mooers, A.O. & Bini, L.M. (2013) Darwinian shortfalls in biodiversity conservation. *Trends in Ecology & Evolution*, **28**, 689–695.

Diniz-Filho, J.A.F., Bastos, R.P., Rangel, T.F.L.V.B., Bini, L.M., Carvalho, P. & Silva, R.J. (2005) Macroecological correlates and spatial patterns of anuran description dates in the Brazilian Cerrado. *Global Ecology and Biogeography*, **14**, 469–477.

Diniz-Filho, J.A.F., Bini, L.M., Pinto, M.P., Rangel, T., Carvalho, P. & Bastos, R.P. (2006) Anuran species richness, complementarity and conservation conflicts in Brazilian Cerrado. *Acta Oecologica*, **29**, 9–15.

Diniz-Filho, J.A.F., Bini, L.M., Vieira, C.M., Blamires, D., Terribile, L.C., Bastos, R.P., de Oliveira, G. & Barreto, B.D.S. (2008) Spatial patterns of terrestrial vertebrate species richness in the Brazilian Cerrado. *Zoological Studies*, **47**, 146–157.

DKRZ (1992) The ECHAM3 atmospheric general circulation model. DKRZ Technical Report No. 6. Deutsches Klimarechenzentrum (DKRZ) Modellbetreuungsgruppe. In, Hamburg, Germany.

Domingos, F.M.C.B., Bosque, R.J., Cassimiro, J., Colli, G.R., Rodrigues, M.T., Santos, M.G. & Beheregaray, L.B. (2014) Out of the deep: Cryptic speciation in a Neotropical gecko (Squamata, Phyllodactylidae) revealed by species delimitation methods. *Molecular Phylogenetics and Evolution*, **80**, 113–124.

Donoghue, M.J. (2008) A phylogenetic perspective on the distribution of plant diversity. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 11549–11555.

Drummond, A.J., Suchard, M.A., Xie, D. & Rambaut, A. (2012) Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–73.

Duellman, W.E. (1979) *The South American herpetofauna: its origin, evolution and dispersal*. Museum of Natural History, The University of Kansas, Lawrence, Kansas, USA.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792-1797.

Eiten, G. (1972) The cerrado vegetation of Brazil. *The Botanical Review*, **38**, 201–341.

Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J.,

Townsend Peterson, A., J. Phillips, S., Richardson, K., Scachetti-Pereira, R., E. Schapire, R., Soberón, J., Williams, S., S. Wisz, M. & E. Zimmermann, N. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

Emerson, B.C. & Gillespie, R.G. (2008) Phylogenetic analysis of community assembly and structure over space and time. *Trends in Ecology & Evolution*, **23**, 619-30.

Ence, D.D. & Carstens, B.C. (2011) SpedeSTEM: a rapid and accurate method for species delimitation. *Molecular Ecology Resources*, **11**, 473–80.

Esselstyn, J.A., Evans, B.J., Sedlock, J.L., Anwarali Khan, F.A. & Heaney, L.R. (2012) Single-locus species delimitation: a test of the mixed Yule-coalescent model, with an empirical application to Philippine round-leaf bats. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 3678–86.

Estes, R. & Báez, A. (1985) Herpetofaunas of North and South America during the late Cretaceous and Cenozoic: Evidence for interchange? *The Great American Biotic Interchange.* (ed. by F.G. Stehli and S.D. Webb), pp. 139–195. Plenum Press, New York.

Excoffier, L. & Lischer, H.E. (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564-7.

Fagundes, N.J., Ray, N., Beaumont, M., Neuenschwander, S., Salzano, F.M., Bonatto, S.L. & Excoffier, L. (2007) Statistical evaluation of alternative models of human evolution. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 17614–17619.

Faria, M.B., Nascimento, F.F., Oliveira, J.A. & Bonvicino, C.R. (2013) Biogeographic determinants of genetic diversification in the mouse opossum *Gracilinanus agilis* (Didelphimorphia: Didelphidae). *The Journal of Heredity*, **104**, 613–26.

Faria, R., Renaut, S., Galindo, J., Pinho, C., Melo-Ferreira, J., Melo, M., Jones, F., Salzburger, W., Schluter, D. & Butlin, R. (2014) Advances in ecological speciation: an integrative approach. *Molecular Ecology*, **23**, 513–521.

Faria, R.G. & Araujo, A.F.B. (2004) Sintopy of two *Tropidurus* lizard species (Squamata: Tropiduridae) in a rocky Cerrado habitat in Central Brazil. *Brazilian Journal of Biology*, **64**, 775–786.

Fearnside, P.M. (2005) Deforestation in Brazilian Amazonia: History, Rates, and Consequences. *Conservation Biology*, **19**, 680–688.

Feldman, C.R. & Spicer, G.S. (2006) Comparative phylogeography of woodland reptiles in California: repeated patterns of cladogenesis and population expansion. *Molecular Ecology*, **15**, 2201–2222.

Fernandes, A.M., Wink, M. & Aleixo, A. (2012) Phylogeography of the chestnut-tailed antbird (*Myrmeciza hemimelaena*) clarifies the role of rivers in Amazonian biogeography. *Journal of Biogeography*, **39**, 1524–1535.

Ferreira, A., Silva, D., Van Sluys, M. & Dolder, H. (2009) Seasonal changes in testicular and epididymal histology of the tropical lizard, *Tropidurus itambere* (Rodrigues, 1987), during its reproductive cycle. *Brazilian Journal of Biology*, **69**, 429–435.

Fjeldsaå, J., Ehrlich, D., Lambin, E. & Prins, E. (1997) Are biodiversity 'hotspots' correlated with current ecoclimatic stability? A pilot study using the NOAA-AVHRR remote sensing data. *Biodiversity and Conservation*, **6**, 401–422.

Forest, F., Crandall, K.A., Chase, M.W. & Faith, D.P. (2015) Phylogeny, extinction and conservation: embracing uncertainties in a time of urgency. *Philosophical Transactions of the Royal Society of London B. Biological Sciences*, **370**, 20140002.

Fouquet, A., Cassini, C.S., Haddad, C.F.B., Pech, N. & Rodrigues, M.T. (2013) Species delimitation, patterns of diversification and historical biogeography of the Neotropical frog genus *Adenomera* (Anura, Leptodactylidae). *Journal of Biogeography*,

Fouquet, A., Cassini, C.S., Haddad, C.F.B., Pech, N. & Rodrigues, M.T. (2014) Species delimitation, patterns of diversification and historical biogeography of the Neotropical frog genus *Adenomera* (Anura, Leptodactylidae). *Journal of Biogeography*, **41**, 855–870.

Fouquet, A., Gilles, A., Vences, M., Marty, C., Blanc, M. & Gemmell, N.J. (2007) Underestimation of species richness in neotropical frogs revealed by mtDNA analyses. *PloS one*, **2**, e1109.

Fouquet, A., Noonan, B.P., Rodrigues, M.T., Pech, N., Gilles, A. & Gemmell, N.J. (2012) Multiple quaternary refugia in the eastern Guiana shield revealed by comparative phylogeography of 12 frog species. *Systematic Biology*, **61**, 461–89.

Franco, F.F. & Manfrin, M.H. (2013) Recent demographic history of cactophilic *Drosophila* species can be related to Quaternary palaeoclimatic changes in South America. *Journal of Biogeography*, **40**, 142–154.

Freeland, J.R., Kirk, H. & Petersen, S. (2011) Molecular Markers in Ecology. *Molecular Ecology* (ed. by J.R. Freeland, H. Kirk and S. Petersen), pp. 35–75. John Wiley & Sons, Ltd, West Sussex, UK.

Freire, E.M.X. (1998) Diferenciação geográfica em *Gymnodactylus Darwini* (Gray, 1845) (Sauria, Gekkonidae). *Papéis Avulsos de Zoologia*, **40**, 311–322.

Frost, D.R., Rodrigues, M.T., Grant, T. & Titus, T.A. (2001) Phylogenetics of the lizard genus *Tropidurus* (Squamata : Tropiduridae : Tropidurinae): Direct optimization, descriptive efficiency, and sensitivity analysis of congruence between molecular data and morphology. *Molecular Phylogenetics and Evolution*, **21**, 352–371.

Fu, Y.-X. (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915–925.

Fujisawa, T. & Barraclough, T.G. (2013) Delimiting Species Using Single-Locus Data and the Generalized Mixed Yule Coalescent Approach: A Revised Method and Evaluation on Simulated Data Sets. *Systematic Biology*, **62**, 707–724.

Fujita, M.K., McGuire, J.A., Donnellan, S.C. & Moritz, C. (2010) Diversification and persistence at the arid-monsoonal interface: Australia-wide biogeography of the Bynoe's gecko (*Heteronotia binoei*; Gekkonidae). *Evolution*, **64**, 2293–2314.

Fujita, M.K., Leache, A.D., Burbrink, F.T., McGuire, J.A. & Moritz, C. (2012) Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology & Evolution*, **27**, 480–8.

Funk, W.C., Caminer, M. & Ron, S.R. (2012a) High levels of cryptic species diversity uncovered in Amazonian frogs. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 1806–1814.

Funk, W.C., McKay, J.K., Hohenlohe, P.A. & Allendorf, F.W. (2012b) Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution*, **27**, 489–96.

Furley, P.A. (1999) The nature and diversity of neotropical savanna vegetation with particular reference to the Brazilian cerrados. *Global Ecology and Biogeography*, **8**, 223–241.

Gainsbury, A.M. & Colli, G.R. (2003) Lizard assemblages from natural Cerrado enclaves in southwestern Amazonia: the role of stochastic extinctions and isolation. *Biotropica*, **35**, 503–519.

Gamble, T., Daza, J.D., Colli, G.R., Vitt, L.J. & Bauer, A.M. (2011) A new genus of miniaturized and pug-nosed gecko from South America (Sphaerodactylidae: Gekkota). *Zoological Journal of the Linnean Society*, **163**, 1244–1266.

Gamble, T., Colli, G.R., Rodrigues, M.T., Werneck, F.P. & Simons, A.M. (2012) Phylogeny and cryptic diversity in geckos (*Phyllopezus*; Phyllodactylidae; Gekkota) from South America's open biomes. *Molecular Phylogenetics and Evolution*, **62**, 943–53.

Garrick, R.C., Bonatelli, I.A.S., Hyseni, C., Morales, A., Pelletier, T.A., Perez, M.F., Rice, E., Satler, J.D., Symula, R.E., Thomé, M.T.C. & Carstens, B.C. (2015) The evolution of phylogeographic data sets. *Molecular Ecology*, **24**, 1164-1171.

Gavrilets, S. (2003) Perspective: models of speciation: what have we learned in 40 years? *Evolution*, **57**, 2197–2215.

Gehara, M., Crawford, A.J., Orrico, V.G., Rodriguez, A., Lotters, S., Fouquet, A., Barrientos, L.S., Brusquetti, F., De la Riva, I., Ernst, R., Urrutia, G.G., Glaw, F., Guayasamin, J.M., Holting, M., Jansen, M., Kok, P.J., Kwet, A., Lingnau, R., Lyra, M., Moravec, J., Pombal, J.P., Jr., Rojas-Runjaic, F.J., Schulze, A., Senaris, J.C., Sole, M., Rodrigues, M.T., Twomey, E., Haddad, C.F., Vences, M. & Kohler, J. (2014) High levels of diversity uncovered in a widespread nominal taxon: continental phylogeography of the neotropical tree frog *Dendropsophus minutus*. *PloS one*, **9**, e103958.

Geraldes, M.C., Van Schmus, W.R., Condie, K.C., Bell, S., Teixeira, W. & Babinski, M. (2001) Proterozoic geologic evolution of the SW part of the Amazonian Craton in Mato Grosso state, Brazil. *Precambrian Research*, **111**, 91–128.

Geurgas, S.R., Rodrigues, M.T. & Moritz, C. (2008) The genus *Coleodactylus* (Sphaerodactylinae, Gekkota) revisited: a molecular phylogenetic perspective. *Molecular Phylogenetics and Evolution*, **49**, 92–101.

Giovanelli, J.G.R., de Siqueira, M.F., Haddad, C.F.B. & Alexandrino, J. (2010) Modeling a spatially restricted distribution in the Neotropics: How the size of calibration area affects the performance of five presence-only methods. *Ecological Modelling*, **221**, 215–224.

Giugliano, L.G., Nogueira, C.C., Valdujo, P.H., Collevatti, R.G. & Colli, G.R. (2013) Cryptic diversity in South American Teiinae (Squamata, Teiidae) lizards. *Zoologica Scripta*, **42**, 473–487.

Glor, R.E., Kolbe, J.J., Powell, R., Larson, A. & Losos, J.B. (2003) Phylogenetic analysis of ecological and morphological diversification in Hispaniolan trunk-ground anoles (*Anolis cybotes* group). *Evolution*, **57**, 2383–2397.

Gomes, J.B.V., Curi, N., Motta, P.E.F., Ker, J.C., Marques, J. & Schulze, D.G. (2004) Análise de componentes principais de atributos físicos, químicos e mineralógicos de solos do bioma Cerrado. *Revista Brasileira de Ciência do Solo*, **28**, 137–153.

Goodland, R. (1971) A physiognomic analysis of the 'Cerrado' vegetation of Central Brasil. *The Journal of Ecology*, **59**, 411–419.

Graham, A. (2011) The age and diversification of terrestrial New World ecosystems through Cretaceous and Cenozoic time. *American Journal of Botany*, **98**, 336–351.

Gronau, I., Hubisz, M.J., Gulko, B., Danko, C.G. & Siepel, A. (2011) Bayesian inference of ancient human demography from individual genome sequences. *Nature Genetics*, **43**, 1031-4.

Guillot, G., Renaud, S., Ledevin, R., Michaux, J. & Claude, J. (2012) A unifying model for the analysis of phenotypic, genetic, and geographic data. *Systematic Biology*, **61**, 897–911.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010) New algorithms and methods to estimate Maximum-Likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, **59**, 307–321.

Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological modelling*, **135**, 147–186.

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993–1009.

Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H. & Bustamante, C.D. (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS genetics*, **5**, e1000695.

Haffer, J. (1969) Speciation in Amazonian forest birds. *Science*, **165**, 131–137.

Haffer, J. (1997) Alternative models of vertebrate speciation in Amazonia: an overview. *Biodiversity and Conservation*, **6**, 451–476.

Harvey, M.B. & Gutberlet, R.L. (2000) A phylogenetic analysis of the tropidurine lizards (Squamata : Tropiduridae), including new characters of squamation and epidermal micro structure. *Zoological Journal of the Linnean Society*, **128**, 189–233.

He, Q., Edwards, D.L. & Knowles, L.L. (2013) Integrative testing of how environments from the past to the present shape genetic structure across landscapes. *Evolution*, **67**, 3386–3402.

Heads, M. (2015) The relationship between biogeography and ecology: envelopes, models, predictions. *Biological Journal of the Linnean Society*, **115**, 456–468.

Hebert, P.D., Penton, E.H., Burns, J.M., Janzen, D.H. & Hallwachs, W. (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 14812–7.

Heled, J. & Drummond, A.J. (2010) Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, **27**, 570–80.

Hey, J. (2010) Isolation with migration models for more than two populations. *Molecular Biology and Evolution*, **27**, 905–20.

Hey, J., Waples, R.S., Arnold, M.L., Butlin, R.K. & Harrison, R.G. (2003) Understanding and confronting species uncertainty in biology and conservation. *Trends in Ecology & Evolution*, **18**, 597–603.

Hickerson, M.J. & Cunningham, C.W. (2005) Contrasting Quaternary histories in an ecologically divergent sister pair of low-dispersing intertidal fish (*Xiphister*) revealed by multilocus DNA analysis. *Evolution*, **59**, 344–360.

Hickerson, M.J., Stahl, E.A. & Lessios, H.A. (2006) Test for Simultaneous Divergence Using Approximate Bayesian Computation. *Evolution*, **60**, 2435–2453.

Hickerson, M.J., Stahl, E. & Takebayashi, N. (2007) msBayes: pipeline for testing comparative phylogeographic histories using hierarchical approximate Bayesian computation. *BMC Bioinformatics*, **8**, 268.

Hickerson, M.J., Stone, G.N., Lohse, K., Demos, T.C., Xie, X., Landerer, C. & Takebayashi, N. (2014) Recommendations for using msBayes to incorporate uncertainty in selecting an ABC model prior: a response to Oaks et al. *Evolution*, **68**, 284–294.

Hijmans, R.J. (2014) *raster: Geographic data analysis and modeling. R package version 2.3-0*. CRAN.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**

Holt, B.G., Lessard, J.P., Borregaard, M.K., Fritz, S.A., Araujo, M.B., Dimitrov, D., Fabre, P.H., Graham, C.H., Graves, G.R., Jonsson, K.A., Nogues-Bravo, D., Wang, Z., Whittaker, R.J., Fjeldsa, J. & Rahbek, C. (2013) An update of Wallace's zoogeographic regions of the world. *Science*, **339**, 74–78.

Hoorn, C., Wesselingh, F.P., ter Steege, H., Bermudez, M.A., Mora, A., Sevink, J., Sanmartin, I., Sanchez-Meseguer, A., Anderson, C.L., Figueiredo, J.P., Jaramillo, C., Riff, D., Negri, F.R., Hooghiemstra, H., Lundberg, J., Stadler, T., Sarkinen, T. & Antonelli, A. (2010) Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science*, **330**, 927–31.

Houle, D. (1992) Comparing evolvability and variability of quantitative traits. *Genetics*, **130**, 195–204.

Huang, W., Takebayashi, N., Qi, Y. & Hickerson, M.J. (2011) MTML-msBayes: approximate Bayesian comparative phylogeographic inference from multiple taxa and multiple loci with rate heterogeneity. *BMC Bioinformatics*, **12**, 1.

Huelsenbeck, J. & Rannala, B. (2004) Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Systematic Biology*, **53**, 904-13.

Hugall, A., Moritz, C., Moussalli, A. & Stanisic, J. (2002) Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail *Gnarosophia bellendenkerensis* (Brazier 1875). *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 6112–6117.

Hugall, A.F. & Lee, M.S. (2004) Molecular claims of Gondwanan age for Australian agamid lizards are untenable. *Molecular Biology and Evolution*, **21**, 2102-10.

Humboldt, A.V. (1849) *Aspects of nature, in different lands and different climates with scientific elucidations*. J. Murray, London.

Hutchinson, M.N., Sistrom, M.J., Donnellan, S.C. & Hutchinson, R.G. (2014) Taxonomic revision of the Australian arid zone lizards *Gehyra variegata* and *G. montium* (Squamata, Gekkonidae) with description of three new species. *Zootaxa*, 221-41.

Isaac, N.J., Mallet, J. & Mace, G.M. (2004) Taxonomic inflation: its influence on macroecology and conservation. *Trends in Ecology & Evolution*, **19**, 464-9.

IUCN (2014) *The IUCN Red List of Threatened Species. Version 2014.3.* Available at: http://www.iucnredlist.org (accessed 20 March 2015

Jablonski, D., Roy, K. & Valentine, J.W. (2006) Out of the tropics: evolutionary dynamics of the latitudinal diversity gradient. *Science*, **314**, 102–106.

Jarvis, E.D., Mirarab, S., Aberer, A.J., Li, B., Houde, P., Li, C., Ho, S.Y.W., Faircloth, B.C., Nabholz, B., Howard, J.T., Suh, A., Weber, C.C., da Fonseca, R.R., Li, J., Zhang, F., Li, H., Zhou, L., Narula, N., Liu, L., Ganapathy, G., Boussau, B., Bayzid, M.S.,

Zavidovych, V., Subramanian, S., Gabald√≥n, T., Capella-Gutiérrez, S., Huerta-Cepas, J., Rekepalli, B., Munch, K., Schierup, M., Lindow, B., Warren, W.C., Ray, D., Green, R.E., Bruford, M.W., Zhan, X., Dixon, A., Li, S., Li, N., Huang, Y., Derryberry, E.P., Bertelsen, M.F., Sheldon, F.H., Brumfield, R.T., Mello, C.V., Lovell, P.V., Wirthlin, M., Schneider, M.P.C., Prosdocimi, F., Samaniego, J.A., Velazquez, A.M.V., Alfaro-N√∫√±ez, A., Campos, P.F., Petersen, B., Sicheritz-Ponten, T., Pas, A., Bailey, T., Scofield, P., Bunce, M., Lambert, D.M., Zhou, Q., Perelman, P., Driskell, A.C., Shapiro, B., Xiong, Z., Zeng, Y., Liu, S., Li, Z., Liu, B., Wu, K., Xiao, J., Yinqi, X., Zheng, Q., Zhang, Y., Yang, H., Wang, J., Smeds, L., Rheindt, F.E., Braun, M., Fjeldsa, J., Orlando, L., Barker, F.K., J√∏nsson, K.A., Johnson, W., Koepfli, K.-P., O,ÄôBrien, S., Haussler, D., Ryder, O.A., Rahbek, C., Willerslev, E., Graves, G.R., Glenn, T.C., McCormack, J., Burt, D., Ellegren, H., Alstr√∂m, P., Edwards, S.V., Stamatakis, A., Mindell, D.P., Cracraft, J., Braun, E.L., Warnow, T., Jun, W., Gilbert, M.T.P. & Zhang, G. (2014) Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, **346**, 1320-1331.

Jezkova, T., Leal, M. & Rodríguez-Robles, J.A. (2009) Living together but remaining apart: comparative phylogeography of Anolis poncensis and A. cooki, two lizards endemic to the aridlands of Puerto Rico. *Biological Journal of the Linnean Society*, **96**, 617–634.

Johns, G.C. & Avise, J.C. (1998) A comparative summary of genetic distances in the vertebrates from the mitochondrial cytochrome b gene. *Molecular Biology and Evolution*, **15**, 1481–1490.

Johnson, M.A., Saraiva, P.M. & Coelho, D. (1999) The role of gallery forests in the distribution of Cerrado mammals. *Revista Brasileira de Biologia*, **59**, 421–427.

Jones, G., Aydin, Z. & Oxelman, B. (2015) DISSECT: an assignment-free Bayesian discovery method for species delimitation under the multispecies coalescent. *Bioinformatics*, **31**, 991-8.

Jones, G.R. (2015) Species delimitation and phylogeny estimation under the multispecies coalescent. *bioRxiv doi:* http://dx.doi.org/10.1101/010199,

Kelly, M., Guo, Q., Liu, D. & Shaari, D. (2007) Modeling the risk for a new invasive forest disease in the United States: An evaluation of five environmental niche models. *Computers, Environment and Urban Systems*, **31**, 689–710.

Kidd, D.M. & Ritchie, M.G. (2006) Phylogeographic information systems: putting the geography into phylogeography. *Journal of Biogeography*, **33**, 1851–1865.

Kiefer, M.C., Van Sluys, M. & Rocha, C.F. (2005) Body temperatures of *Tropidurus torquatus* (Squamata, Tropiduridae) from coastal populations: Do body temperatures vary along their geographic range? *Journal of Thermal Biology*, **30**, 449-456.

Kier, G., Mutke, J., Dinerstein, E., Ricketts, T.H., Küper, W., Kreft, H. & Barthlott, W. (2005) Global patterns of plant diversity and floristic knowledge. *Journal of Biogeography*, **32**, 1107-1116.

King, L.C. (1956) A geomorfologia do Brasil Oriental. *Revista Brasileira de Geografia*, **18**, 147–265.

Klink, C.A. & Machado, R.B. (2005) Conservation of the Brazilian Cerrado. *Conservation Biology*, **19**, 707–713.

Knowles, L.L. & Maddison, W.P. (2002) Statistical phylogeography. *Molecular Ecology*, **11**, 2623–2635.

Knowles, L.L. & Carstens, B.C. (2007) Delimiting species without monophyletic gene trees. *Systematic Biology*, **56**, 887–95.

Knowles, L.L., Carstens, B.C. & Keat, M.L. (2007) Coupling genetic and ecological-niche models to examine how past population distributions contribute to divergence. *Current Biology*, **17**, 940–946.

Kohlsdorf, T. & Wagner, G.P. (2006) Evidence for the reversibility of digit loss: a phylogenetic study of limb evolution in *Bachia* (Gymnophthalmidae: Squamata). *Evolution*, **60**, 1896–1912.

Kohlsdorf, T. & Navas, C.A. (2007) Evolution of jumping capacity in Tropidurinae lizards: does habitat complexity influence obstacle-crossing ability? *Biological Journal of the Linnean Society*, **91**, 393-402.

Kohlsdorf, T., Garland Jr, T. & Navas, C.A. (2001) Limb and tail lengths in relation to substrate usage in *Tropidurus* lizards. *Journal of Morphology*, **248**, 151-164.

Kohlsdorf, T., James, R.S., Carvalho, J.E., Wilson, R.S., Pai-Silva, M.D. & Navas, C.A. (2004) Locomotor performance of closely related *Tropidurus* species: relationships with physiological parameters and ecological divergence. *Journal of Experimental Biology*, **207**, 1183-1192.

Kubatko, L.S., Carstens, B.C. & Knowles, L.L. (2009) STEM: species tree estimation using maximum likelihood for gene trees under coalescence. *Bioinformatics*, **25**, 971-3.

Kuhn, M. (2008) Building predictive models in R using the caret package. *Journal of statistical software*, **28**, 1-26.

Lanfear, R., Calcott, B., Ho, S.Y. & Guindon, S. (2012) Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, **29**, 1695–701.

Lanfear, R., Calcott, B., Kainer, D., Mayer, C. & Stamatakis, A. (2014) Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evolutionary Biology*, **14**, 82.

Lapointe, F.J. & Rissler, L.J. (2005) Congruence, Consensus, and the Comparative Phylogeography of Codistributed Species in California. *American Naturalist*, **166**, 290–299.

Lara, L.E., Naranjo, J.A. & Moreno, H. (2004) Lanín volcano (39.5S), Southern Andes: geology and morphostructural evolution. *Revista geolígica de Chile*, **31**, 241–257.

Lavin, M., Herendeen, P.S. & Wojciechowski, M.F. (2005) Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. *Systematic Biology*, **54**, 575–594.

Leache, A.D. (2009) Species tree discordance traces to phylogeographic clade boundaries in North American fence lizards (Sceloporus). *Systematic Biology*, **58**, 547–559.

Leache, A.D. & Rannala, B. (2011) The accuracy of species tree estimation under simulation: a comparison of methods. *Systematic Biology*, **60**, 126-37.

Leache, A.D., Crews, S.C. & Hickerson, M.J. (2007) Two waves of diversification in mammals and reptiles of Baja California revealed by hierarchical Bayesian analysis. *Biology letters*, **3**, 646-50.

Leaché, A.D. & Fujita, M.K. (2010) Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). *Proceedings of the Royal Society B: Biological Sciences*, **277**, 3071–3077.

Leaché, A.D., Harris, R.B., Maliska, M.E. & Linkem, C.W. (2013a) Comparative Species Divergence across Eight Triplets of Spiny Lizards (*Sceloporus*) Using Genomic Sequence Data. *Genome Biology and Evolution*, **5**, 2410–2419.

Leaché, A.D., Palacios, J.A., Minin, V.N. & Bryson, R.W. (2013b) Phylogeography of the Trans-Volcanic bunchgrass lizard (*Sceloporus bicanthalis*) across the highlands of south-eastern Mexico. *Biological Journal of the Linnean Society*, **110**, 852–865.

Leaché, A.D., Harris, R.B., Rannala, B. & Yang, Z. (2014) The Influence of Gene Flow on Species Tree Estimation: A Simulation Study. *Systematic Biology*, **63**, 17–30.

Leaché, A.D., Chavez, A.S., Jones, L.N., Grummer, J.A., Gottscho, A.D. & Linkem, C.W. (2015) Phylogenomics of Phrynosomatid lizards: conflicting signals from sequence capture versus restriction site associated DNA sequencing. *Genome Biology and Evolution*, **7**, 706–719.

Ledru, M.P. (1993) Late Quaternary environmental and climatic changes in Central Brazil. *Quaternary Research*, **39**, 90–98.

Ledru, M.P., Ceccantini, G., Gouveia, S.E.M., López-Sáez, J.A., Pessenda, L.C.R. & Ribeiro, A.S. (2006) Millenial-scale climatic and vegetation changes in a northern Cerrado (Northeast, Brazil) since the Last Glacial Maximum. *Quaternary Science Reviews*, **25**, 1110–1126.

Leite, R.N., Kolokotronis, S.O., Almeida, F.C., Werneck, F.P., Rogers, D.S. & Weksler, M. (2014) In the wake of invasion: tracing the historical biogeography of the South american cricetid radiation (Rodentia, Sigmodontinae). *PloS one*, **9**, e100687.

Lemmon, A.R. & Moriarty, E.C. (2004) The importance of proper model assumption in Bayesian phylogenetics. *Systematic Biology*, **53**, 265–277.

Lemmon, A.R. & Lemmon, E.M. (2008) A likelihood framework for estimating phylogeographic history on a continuous landscape. *Systematic Biology*, **57**, 544–61.

Lemmon, A.R., Emme, S.A. & Lemmon, E.M. (2012) Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology*, **61**, 727–44.

Lemmon, A.R., Brown, J.M., Stanger-Hall, K. & Lemmon, E.M. (2009) The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference. *Systematic Biology*, **58**, 130-45.

Lemmon, E.M. & Lemmon, A.R. (2013) High-throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, **44**, 99–121.

Lexer, C., Mangili, S., Bossolini, E., Forest, F., Stölting, K.N., Pearman, P.B., Zimmermann, N.E., Salamin, N. & Carine, M. (2013) 'Next generation' biogeography: towards

understanding the drivers of species diversification and persistence. *Journal of Biogeography*, **40**, 1013-1022.

Librado, P. & Rozas, J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Lima, J.R., Mansano, V.F. & Araujo, F.S. (2012) Richness and diversity of Leguminosae in an altitudinal gradient in the tropical semi-arid zone of Brazil. *Journal of Systematics and Evolution*, **50**, 433-442.

Lischer, H.E., Excoffier, L. & Heckel, G. (2014) Ignoring heterozygous sites biases phylogenomic estimates of divergence times: implications for the evolutionary history of *Microtus* voles. *Molecular Biology and Evolution*, **31**, 817–831.

Liu, L. & Yu, L. (2011) Estimating species trees from unrooted gene trees. *Systematic Biology*, **60**, 661-7.

Liu, L., Yu, L., Pearl, D.K. & Edwards, S.V. (2009) Estimating species phylogenies using coalescence times among sequences. *Systematic Biology*, **58**, 468-77.

Lohse, K., Barton, N.H., Melika, G. & Stone, G.N. (2012) A likelihood-based comparison of population histories in a parasitoid guild. *Molecular Ecology*, **21**, 4605-17.

Lomolino, M.V. (2000) Ecology's most general, yet protean pattern: the species-area relationship. *Journal of Biogeography*, **27**, 17–26.

Losos, J.B. (2008) Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecology Letters*, **11**, 995–1003.

Losos, J.B. (2009) *Lizards in an evolutionary tree: ecology and adaptive radiation of anoles*. University of California Press.

Losos, J.B. & Glor, R.E. (2003) Phylogenetic comparative methods and the geography of speciation. *Trends in Ecology & Evolution*, **18**, 220-227.

Lougheed, S.C., Campagna, L., Dávila, J.A., Tubaro, P.L., Lijtmaer, D.A. & Handford, P. (2013) Continental phylogeography of an ecologically and morphologically diverse Neotropical songbird, *Zonotrichia capensis*. *BMC Evolutionary Biology*, **13**, 1–16.

MacArthur, R.H. & MacArthur, J.W. (1961) On bird species diversity. *Ecology*, 594–598.

MacArthur, R.H. & Wilson, E.O. (1963) An Equilibrium Theory of Insular Zoogeography. *Evolution*, **17**, 373–387.

Mace, G.M. (2004) The role of taxonomy in species conservation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **359**, 711–719.

Mace, G.M. & Purvis, A. (2008) Evolutionary biology and practical conservation: bridging a widening gap. *Molecular Ecology*, **17**, 9-19.

Macey, J.R., Schulte II, J.A., Ananjeva, N.B., Larson, A., Rastegar-Pouyani, N., Shammakov, S.M. & Papenfuss, T.J. (1998) Phylogenetic relationships among agamid lizards of the *Laudakia caucasia* species group: Testing hypotheses of biogeographic fragmentation and an area cladogram for the Iranian Plateau. *Molecular Phylogenetics and Evolution*, **10**, 118–131.

Maciel, N.M., Collevatti, R.G., Colli, G.R. & Schwartz, E.F. (2010) Late Miocene diversification and phylogenetic relationships of the huge toads in the *Rhinella marina* (Linnaeus, 1758) species group (Anura: Bufonidae). *Molecular Phylogenetics and Evolution*, **57**, 787–797.

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

Mares, M.A. & Ernest, K.A. (1995) Population and community ecology of small mammals in a gallery forest of central Brazil. *Journal of Mammalogy*, 750–768.

Margules, C.R. & Pressey, R.L. (2000) Systematic conservation planning. *Nature*, **405**, 243–253.

Marinho-Filho, J.S., Rodrigues, F.H.G. & Juarez, K.M. (2002) The Cerrado mammals: diversity, ecology, and natural history. 266-284. *The Cerrados of Brazil: Ecology and Natural History of a Neotropical Savanna. PS Oliveira e RJ Marquis (eds). Columbia University Press, New York, EUA* (ed. by P.S. Oliveira and R.J. Marquis). Columbia University Press.

Marske, K.A., Rahbek, C. & Nogués-Bravo, D. (2013) Phylogeography: spanning the ecology-evolution continuum. *Ecography*, **36**, 1169-1181.

Martins, R.L., Barcelos, A. & Del-Claro, K. (1999) Activity pattern of *Tropidurus torquatus* (Sauria: Tropidnridae) in an urban area of Uberlândia-MG. *Revista de Etologia*, **1**, 19–23.

May, R.M. (2011) Why worry about how many species and their loss? *PLoS Biology*, **9**, e1001130.

McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C. & Brumfield, R.T. (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, **66**, 526–38.

McVay, J.D. & Carstens, B. (2013) Testing monophyly without well-supported gene trees: evidence from multi-locus nuclear data conflicts with existing taxonomy in the snake tribe Thamnophiini. *Molecular Phylogenetics and Evolution*, **68**, 425–31.

Mendonça, R.C., Felfili, J.M., Walter, B.M.T., Silva Júnior, M.C., Rezende, A.V., Filgueiras, T.S. & Nogueira, P.E. (1998) Flora vascular do cerrado. *Cerrado: ambiente e flora* (ed. by M.S. Almeida and S.P. Almeida), pp. 287– 556 Embrapa-CPAC, Planaltina, DF.

Mesquita, D.O., Colli, G.R., França, F.G.R. & Vitt, L.J. (2006) Ecology of a Cerrado lizard assemblage in the Jalapão region of Brazil. *Copeia*, **2006**, 460–471.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A. & Leisch, F. (2014) *e1071: Misc Functions of the Department of Statistics (e1071), TU Wien. R package version 1.6-2.*

Michaux, J.R., Libois, R. & Filippucci, M.G. (2005) So close and so different: comparative phylogeography of two small mammal species, the yellow-necked fieldmouse (*Apodemus flavicollis*) and the woodmouse (*Apodemus sylvaticus*) in the Western Palearctic region. *Heredity*, **94**, 52–63.

Minh, B.Q., Nguyen, M.A. & von Haeseler, A. (2013) Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution*, **30**, 1188-95.

Mirarab, S., Bayzid, M.S. & Warnow, T. (2014) Evaluating summary methods for multilocus species tree estimation in the presence of incomplete lineage sorting. *Systematic Biology*, 1–15.

Mittermeier, R.A., Gil, P.R., Hoffman, M., Pilgrim, J., Brooks, T., Mittermeier, C.G., Lamoreux, J. & Da Fonseca, G.A.B. (2005) *Hotspots revisited: Earth's biologically richest and most endangered terrestrial ecoregions*. The University of Chicago Press.

Monnet, A.C., Jiguet, F., Meynard, C.N., Mouillot, D., Mouquet, N., Thuiller, W. & Devictor, V. (2014) Asynchrony of taxonomic, functional and phylogenetic diversity in birds. *Global Ecology and Biogeography*, **23**, 780-788.

Mora, C., Tittensor, D.P., Adl, S., Simpson, A.G. & Worm, B. (2011) How many species are there on Earth and in the ocean? *PLoS Biology*, **9**, e1001127.

Moraes, E.M., Yotoko, K.S.C., Manfrin, M.H., Solferini, V.N. & Sene, F.M. (2009) Phylogeography of the cactophilic species *Drosophila gouveai*: demographic events and divergence timing in dry vegetation enclaves in eastern Brazil. *Journal of Biogeography*, **36**, 2136–2147.

Moraes-Barros, N., Silva, J.A.B., Miyaki, C.Y. & Morgante, J.S. (2006) Comparative Phylogeography of the Atlantic Forest Endemic Sloth (*Bradypus torquatus*) and the Widespread Three-toed Sloth (*Bradypus variegatus*) (Bradypodidae, Xenarthra). *Genetica*, **126**, 189–198.

Morando, M., Avila, L.J. & Sites Jr, J.W. (2003) Sampling Strategies for Delimiting Species; Genes, Individuals, and Populations in the *Liolaemus elongatus*-kriegi Complex (Squamata; Liolaemidae) in Andean-Patagonian South America. *Systematic Biology*, **52**, 159–185.

Morando, M., Medina, C.D., Avila, L.J., Perez, C.H.F., Buxton, A. & Sites, J.W. (2014) Molecular phylogeny of the New World gecko genus *Homonota* (Squamata: Phyllodactylidae). *Zoologica Scripta*, **43**, 249–260.

Morgan, K., O'Loughlin, S.M., Chen, B., Linton, Y.M., Thongwat, D., Somboon, P., Fong, M.Y., Butlin, R., Verity, R., Prakash, A., Htun, P.T., Hlaing, T., Nambanya, S., Socheat, D., Dinh, T.H. & Walton, C. (2011) Comparative phylogeography reveals a shared impact of pleistocene environmental change in shaping genetic diversity within nine *Anopheles* mosquito species across the Indo-Burma biodiversity hotspot. *Molecular Ecology*, **20**, 4533–49.

Moritz, C. (1994) Defining 'Evolutionarily Significant Units' for conservation. *Trends in Ecology & Evolution*, **9**, 373–375.

Moritz, C. & Faith, D.P. (1998) Comparative phylogeography and the identification of genetically divergent areas for conservation. *Molecular Ecology*, **7**, 419–429.

Moritz, C., Patton, J., Schneider, C. & Smith, T. (2000) Diversification of rainforest faunas: an integrated molecular approach. *Annual Review of Ecology and Systematics*, **31**, 533–563.

Moritz, C., Hoskin, C.J., MacKenzie, J.B., Phillips, B.L., Tonione, M., Silva, N., VanDerWal, J., Williams, S.E. & Graham, C.H. (2009) Identification and dynamics of a cryptic suture zone in tropical rainforest. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 1235–1244.

Motta, P.E.F., Curi, N. & Franzmeier, D.P. (2002) Relation of soils and geomorphic surfaces in the Brazilian cerrado. *The cerrados of Brazil - Ecology and natural history of a*

neotropical savanna. (ed. by P.S. Oliveira and R.J. Marquis), pp. 13–32. Columbia University Press, New York.

Moussalli, A., Moritz, C., Williams, S.E. & Carnaval, A.C. (2009) Variable responses of skinks to a common history of rainforest fluctuation: concordance between phylogeography and palaeodistribution models. *Molecular Ecology*, **18**, 483–499.

Mulcahy, D.G., Noonan, B.P., Moss, T., Townsend, T.M., Reeder, T.W., Sites, J.W., Jr. & Wiens, J.J. (2012) Estimating divergence dates and evaluating dating methods using phylogenomic and mitochondrial data in squamate reptiles. *Molecular Phylogenetics and Evolution*, **65**, 974–991.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B. & Kent, J. (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.

Nair, A., Gopalan, S.V., George, S., Kumar, K.S., Teacher, A.G.F., Merilä, J., Garner, T. & Austin, J. (2012) High cryptic diversity of endemic *Indirana* frogs in the Western Ghats biodiversity hotspot. *Animal Conservation*, **15**, 489–498.

Nakagawa, S. & Freckleton, R.P. (2008) Missing inaction: the dangers of ignoring missing data. *Trends in Ecology & Evolution*, **23**, 592–596.

Navas, C.A. (2002) Herpetological diversity along Andean elevational gradients: links with physiological ecology and evolutionary physiology. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, **133**, 469–485.

Nelson, G. (1978) From Candolle to croizat: Comments on the history of biogeography. *Journal of the History of Biology*, **11**, 269–305.

Niemiller, M.L., Graening, G.O., Fenolio, D.B., Godwin, J.C., Cooley, J.R., Pearson, W.D., Fitzpatrick, B.M. & Near, T.J. (2013) Doomed before they are described? The need for conservation assessments of cryptic species complexes using an amblyopsid cavefish (Amblyopsidae: *Typhlichthys*) as a case study. *Biodiversity and Conservation*, **22**, 1799–1820.

Noble, G.K. (1927) The value of life history data in the study of the evolution of the Amphibia. *Annals of the New York Academy of Sciences*, **30**, 31–128.

Nogueira, C. (2006) *Diversidade e padrões de distribuição da fauna de lagartos do Cerrado*. Universidade de São Paulo, São Paulo.

Nogueira, C. & Rodrigues, M.T. (2006) The genus *Stenocercus* (Squamata: Tropiduridae) in Extra-Amazonian Brazil, with the description of two new species. *South American Journal of Herpetology*, **1**, 149–165.

Nogueira, C., Valdujo, P.H. & França, F.G.R. (2005) Habitat variation and lizard diversity in a Cerrado area of Central Brazil. *Studies on Neotropical Fauna and Environment*, **40**, 105–112.

Nogueira, C., Colli, G.R. & Martins, M. (2009) Local richness and distribution of the lizard fauna in natural habitat mosaics of the Brazilian Cerrado. *Austral Ecology*, **34**, 83–96.

Nogueira, C., Ribeiro, S., Costa, G.C. & Colli, G.R. (2011) Vicariance and endemism in a Neotropical savanna hotspot: distribution patterns of Cerrado squamate reptiles. *Journal of Biogeography*, **38**, 1907-1922.

Nosil, P., Vines, T.H. & Funk, D.J. (2005) Reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution*, **59**, 705–719.

Novaes, R.M.L., Ribeiro, R.A., Lemos-Filho, J.P. & Lovato, M.B. (2013) Concordance between phylogeographical and biogeographical patterns in the Brazilian Cerrado: Diversification of the endemic tree *Dalbergia miscolobium* (Fabaceae). *PloS one*, **8**, e82198.

Nylander, J.A.A. (2004) *MrModeltest v2. Program distributed by the author*. Evolutionary Biology Centre, Uppsala University.

O'Connell, J., Gurdasani, D., Delaneau, O., Pirastu, N., Ulivi, S., Cocca, M., Traglia, M., Huang, J., Huffman, J.E., Rudan, I., McQuillan, R., Fraser, R.M., Campbell, H., Polasek, O., Asiki, G., Ekoru, K., Hayward, C., Wright, A.F., Vitart, V., Navarro, P.,

Zagury, J.F., Wilson, J.F., Toniolo, D., Gasparini, P., Soranzo, N., Sandhu, M.S. & Marchini, J. (2014) A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS genetics*, **10**, e1004234.

O'Meara, B.C. (2010) New heuristic methods for joint species delimitation and species tree inference. *Systematic Biology*, **59**, 59-73.

O'Neill, E.M., Schwartz, R., Bullock, C.T., Williams, J.S., Shaffer, H.B., Aguilar-Miguel, X., Parra-Olea, G. & Weisrock, D.W. (2013) Parallel tagged amplicon sequencing reveals major lineages and phylogenetic structure in the North American tiger salamander (*Ambystoma tigrinum*) species complex. *Molecular Ecology*, **22**, 111-29.

Ogden, R. & Thorpe, R.S. (2002) Molecular evidence for ecological speciation in tropical habitats. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 13612–13615.

Olave, M., Solà, E. & Knowles , L.L. (2014) Upstream Analyses Create Problems with DNA-Based Species Delimitation. *Systematic Biology*, **63**, 263-271.

Oliveira Filho, A.T. & Ratter, J.A. (2000) Padrões florísticos das matas ciliares da região do Cerrado e a evolução das paisagens do Brasil central durante o quaternário tardio. *Matas ciliares: conservação e recuperação* (ed. by R.R. Rodrigues and H.F. Leitão Filho), pp. 73–89. Edusp / Fapesp, São Paulo.

Oliveira, P.S. & Marquis, R.J. (2002) *The cerrados of Brazil: ecology and natural history of a neotropical savanna*. Columbia University Press.

Oliveira-Filho, A.T. & Ratter, J.A. (2002) Vegetation physiognomies and woody flora of the cerrado biome. *The cerrados of Brazil: ecology and natural history of a neotropical savanna.(PS Oliveira & JR Marquis, eds.). Columbia University Press, New York*, 91–119.

Oliver, P.M., Richards, S.J. & Sistrom, M. (2012) Phylogeny and systematics of Melanesia's most diverse gecko lineage (*Cyrtodactylus*, Gekkonidae, Squamata). *Zoologica Scripta*, **41**, 437-454.

Oliver, P.M., Adams, M., Lee, M.S., Hutchinson, M.N. & Doughty, P. (2009) Cryptic diversity in vertebrates: molecular data double estimates of species diversity in a radiation of Australian lizards (*Diplodactylus*, Gekkota). *Proceedings of the Royal Society B: Biological Sciences*, **276**, 2001–2007.

Otto-Bliesner, B.L., Marshall, S.J., Overpeck, J.T., Miller, G.H. & Hu, A. (2006) Simulating Arctic climate warmth and icefield retreat in the last interglaciation. *Science*, **311**, 1751-3.

Padial, J.M., Miralles, A., De la Riva, I. & Vences, M. (2010) The integrative future of taxonomy. *Frontiers in zoology*, **7**, 16.

Pamilo, P. & Nei, M. (1988) Relationships between Gene Trees and Species Trees. *Molecular Biology and Evolution*, **5**, 568–583.

Pante, E., Abdelkrim, J., Viricel, A., Gey, D., France, S.C., Boisselier, M.C. & Samadi, S. (2015) Use of RAD sequencing for delimiting species. *Heredity*, **114**, 450-9.

Papadopoulou, A., Anastasiou, I., Keskin, B. & Vogler, A.P. (2009) Comparative phylogeography of tenebrionid beetles in the Aegean archipelago: the effect of dispersal ability and habitat preference. *Molecular Ecology*, **18**, 2503-17.

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

Park, K.J. & Kanehisa, M. (2003) Prediction of protein subcellular locations by support vector machines using compositions of amino acids and amino acid pairs. *Bioinformatics*, **19**, 1656–1663.

Parra, J.L., McGuire, J.A. & Graham, C.H. (2010) Incorporating Clade Identity in Analyses of Phylogenetic Community Structure: An Example with Hummingbirds. *American Naturalist*, **176**, 573–587.

Pearse, D.E. & Crandall, K.A. (2004) Beyond FST: analysis of population genetic data for conservation. *Conservation Genetics*, **5**, 585–602.

Pearson, R.G., Thuiller, W., Araújo, M.B., Martinez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, T.P. & Lees, D.C. (2006) Model-based uncertainty in species range prediction. *Journal of Biogeography*, **33**, 1704-1711.

Pellegrino, K.C.M., Rodrigues, M.T., Waite, A.N., Morando, M., Yassuda, Y.Y. & Sites, J.W. (2005) Phylogeography and species limits in the *Gymnodactylus darwinii* complex (Gekkonidae, Squamata): genetic structure coincides with river systems in the Brazilian Atlantic Forest. *Biological Journal of the Linnean Society*, **85**, 13–26.

Pellegrino, K.C.M., dos Santos, R.M.L., Rodrigues, M.T., Laguna, M.M., Amaro, R.C. & Yonenaga-Yassuda, Y. (2009) Chromosomal evolution in the Brazilian geckos of the genus *Gymnodactylus* (Squamata, Phyllodactylidae) from the biomes of Cerrado, Caatinga and Atlantic Rain Forest: evidence of Robertsonian fusion events and supernumerary chromosomes. *Cytogenetic and Genome Research*, **127**, 191–203.

Pelletier, T.A. & Carstens, B.C. (2014) Model choice for phylogeographic inference using a large set of models. *Molecular Ecology*, **23**, 3028–3043.

Pennington, R.T., Lavin, M., Prado, D.E., Pendry, C.A., Pell, S.K. & Butterworth, C.A. (2004) Historical climate change and speciation: neotropical seasonally dry forest plants show patterns of both Tertiary and Quaternary diversification. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **359**, 515-537.

Pepper, M., Doughty, P. & Keogh, J.S. (2006) Molecular phylogeny and phylogeography of the Australian *Diplodactylus stenodactylus* (Gekkota; Reptilia) species-group based on

mitochondrial and nuclear genes reveals an ancient split between Pilbara and non-Pilbara *D. stenodactylus*. *Molecular Phylogenetics and Evolution*, **41**, 539–555.

Pepper, M., Doughty, P., Fujita, M.K., Moritz, C. & Keogh, J.S. (2013) Speciation on the rocks: integrated systematics of the *Heteronotia spelea* species complex (Gekkota; Reptilia) from Western and Central Australia. *PloS one*, **8**, e78110.

Pereira, F., Carneiro, J. & Amorim, A. (2008) Identification of species with DNA-based technology: current progress and challenges. *Recent Patents on DNA & Gene Sequences*, **2**, 187–200.

Pfenninger, M. & Schwenk, K. (2007) Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. *BMC Evolutionary Biology*, **7**, 121.

Phillips, S.J. & Dudík, M. (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, **31**, 161–175.

Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

Pianka, E.R. (1966) Convexity, desert lizards, and spatial heterogeneity. *Ecology*, 1055–1059.

Pianka, E.R. (1989) Desert lizard diversity: additional comments and some data. *American Naturalist*, **134**, 344–364.

Pinheiro, F., Cozzolino, S., Draper, D., de Barros, F., Félix, L.P., Fay, M.F. & Palma-Silva, C. (2014) Rock outcrop orchids reveal the genetic connectivity and diversity of inselbergs of northeastern Brazil. *BMC Evolutionary Biology*, **14**, 49.

Pinna, P.H., Mendonça, A.F., Bocchiglieri, A. & Fernandes, D.S. (2010) A new two-pored *Amphisbaena* Linnaeus from the endangered Brazilian Cerrado biome (Squamata: Amphisbaenidae). *Zootaxa*, 44–54.

Pinzon-Navarro, S., Barrios, H., Murria, C., Lyal, C.H. & Vogler, A.P. (2010) DNA-based taxonomy of larval stages reveals huge unknown species diversity in neotropical seed

weevils (genus *Conotrachelus*): relevance to evolutionary ecology. *Molecular Phylogenetics and Evolution*, **56**, 281-93.

Pons, J., Barraclough, T., Gomez-Zurita, J., Cardoso, A., Duran, D., Hazell, S., Kamoun, S., Sumlin, W. & Vogler, A. (2006) Sequence-Based Species Delimitation for the DNA Taxonomy of Undescribed Insects. *Systematic Biology*, **55**, 595–609.

Posada, D. (2008) jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, **25**, 1253–6.

Posadas, P., Crisci, J.V. & Katinas, L. (2006) Historical biogeography: a review of its basic concepts and critical issues. *Journal of Arid Environments*, **66**, 389–403.

Possingham, H.P., Grantham, H. & Rondinini, C. (2007) How can you conserve species that haven't been found? *Journal of Biogeography*, **34**, 758-759.

Prado, C.P., Haddad, C.F. & Zamudio, K.R. (2012) Cryptic lineages and Pleistocene population expansion in a Brazilian Cerrado frog. *Molecular Ecology*, **21**, 921–41.

Prentice, I.C., Cramer, W., Harrison, S.P., Leemans, R., Monserud, R.A. & Solomon, A.M. (1992) A global biome model based on plant physiology and dominance, soil properties and climate. *Journal of Biogeography*, **19**, 117–134.

Provan, J. & Maggs, C.A. (2012) Unique genetic variation at a species' rear edge is under threat from global climate change. *Proceedings of the Royal Society B*, **279**, 39–47.

Pyron, R.A., Hendry, C.R., Chou, V.M., Lemmon, E.M., Lemmon, A.R. & Burbrink, F.T. (2014) Effectiveness of phylogenomic data and coalescent species-tree methods for resolving difficult nodes in the phylogeny of advanced snakes (Serpentes: Caenophidia). *Molecular Phylogenetics and Evolution*, **81**, 221–231.

Quinn, G.P. & Keough, M.J. (2002) *Experimental design and data analysis for biologists*. Cambridge University Press, Cambridge, UK.

R Core Team (2013) *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.*

Rabosky, D.L. & Lovette, I.J. (2008) Explosive evolutionary radiations: decreasing speciation or increasing extinction through time? *Evolution*, **62**, 1866–1875.

Rabosky, D.L. & Lovette, I.J. (2009) Problems detecting density-dependent diversification on phylogenies: reply to Bokma. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 995–997.

Radambrasil (1982-83) *Geologia, Geomorfologia, Pedologia, Vegetação e Uso Potencial da Terra*. Ministério das Minas e Energia, Rio de Janeiro, Brasil.

Rambaut, A. & Drummond, A.J. (2009) *Tracer v1.5, Available from* http://tree.bio.ed.ac.uk/software/tracer/.

Ramos, A.C.S., Lemos-Filho, J.P., Ribeiro, R.A., Santos, F.R. & Lovato, M.B. (2007) Phylogeography of the tree *Hymenaea stigonocarpa* (Fabaceae : Caesalpinioideae) and the influence of quaternary climate changes in the Brazilian Cerrado. *Annals of Botany*, **100**, 1219–1228.

Ramos-Onsins, S.E. & Rozas, J. (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, **19**, 2092–2100.

Rangel, T.F., Colwell, R.K., Graves, G.R., Fučíková, K., Rahbek, C. & Diniz-Filho, J.A.F. (2015) Phylogenetic uncertainty revisited: Implications for ecological analyses. *Evolution*, ***in press***

Rannala, B. & Yang, Z. (2003) Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics*, **164**, 1645–1656.

Rannala, B. & Yang, Z. (2013) Improved reversible jump algorithms for Bayesian species delimitation. *Genetics*, **194**, 245-53.

Ratter, J.A., Ribeiro, J.F. & Bridgewater, S. (1997) The Brazilian cerrado vegetation and threats to its biodiversity. *Annals of Botany*, **80**, 223.

Recoder, R., Teixeira Júnior, M., Camacho, A. & Rodrigues, M.T. (2012) Natural history of the tropical gecko *Phyllopezus pollicaris* (Squamata, Phyllodactylidae) from a sandstone outcrop in Central Brazil. *Herpetology Notes*, **5**, 49–58.

Recoder, R.S., Werneck, F.P., Teixeira, M., Colli, G.R., Sites, J.W. & Rodrigues, M.T. (2014) Geographic variation and systematic review of the lizard genus *Vanzosaura* (Squamata, Gymnophthalmidae), with the description of a new species. *Zoological Journal of the Linnean Society*, **171**, 206–225.

Redding, D.W., Mazel, F. & Mooers, A.Ø. (2014) Measuring evolutionary isolation for conservation. *PloS one*, **9**, e113490.

Redford, K.H. & da Fonseca, G.A.B. (1986) The Role of Gallery Forests in th Zoogeography of the Cerrado's Non-volant Mammalian Fauna. *Biotropica*, 126–135.

Reeder, T.W., Townsend, T.M., Mulcahy, D.G., Noonan, B.P., Wood, P.L., Jr., Sites, J.W., Jr. & Wiens, J.J. (2015) Integrated Analyses Resolve Conflicts over Squamate Reptile Phylogeny and Reveal Unexpected Placements for Fossil Taxa. *PloS one*, **10**, e0118199.

Reid, N.M. & Carstens, B.C. (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.

Ribeiro, S., Castro-Mello, C. & Nogueira, C. (2009) New species of *Anops* Bell, 1833 (Squamata, Amphisbaenia) from Jalapão Region in the Brazilian Cerrado. *Journal of Herpetology*, **43**, 21–28.

Richards, C.L., Carstens, B.C. & Knowles, L.L. (2007) Distribution modelling and statistical phylogeography: an integrative framework for generating and testing alternative biogeographical hypotheses. *Journal of Biogeography*, **34**, 1833–1845.

Riddle, B.R. (1996) The molecular phylogeographic bridge between deep and shallow history in continental biotas. *Trends in Ecology & Evolution*, **11**, 207–211.

Rissler, L.J. & Apodaca, J.J. (2007) Adding more ecology into species delimitation: ecological niche models and phylogeography help define cryptic species in the black salamander (Aneides flavipunctatus). *Systematic Biology*, **56**, 924-42.

Rittmeyer, E.N. & Austin, C.C. (2012) The effects of sampling on delimiting species from multi-locus sequence data. *Molecular Phylogenetics and Evolution*, **65**, 451-63.

Rittmeyer, E.N. & Austin, C.C. (2015) Combined next-generation sequencing and morphology reveal fine-scale speciation in Crocodile Skinks (Squamata: Scincidae: Tribolonotus). *Molecular Ecology*, **24**, 466-83.

Robertson, J.M. & Vega, A. (2011) Genetic and phenotypic variation in a colourful treefrog across five geographic barriers. *Journal of Biogeography*, **38**, 2122–2135.

Robinson, J.D., Bunnefeld, L., Hearn, J., Stone, G.N. & Hickerson, M.J. (2014) ABC inference of multi-population divergence with admixture from unphased population genomic data. *Molecular Ecology*, **23**, 4458–4471.

Rodrigues, M.T. (1987) Sistemática, ecologia e zoogeografia dos *Tropidurus* do grupo torquatus ao sul do Rio Amazonas (Sauria, Iguanidae). *Arquivos de Zoologia*, **31**, 105–230.

Rodrigues, M.T. (1996) A new species of lizard, genus *Micrablepharus* (Squamata: Gymnophthalmidae), from Brazil. *Herpetologica*, **52**, 535–541.

Rodrigues, M.T., Kasahara, S. & Yonenaga-Yassuda, Y. (1988) *Tropidurus psammonastes*: uma nova espécie do grupo torquatus com notas sobre seu cariótipo e distribuição (Sauria, Iguanidae). *Papéis Avulsos de Zoologia*, **36**, 307–313.

Rodrigues, M.T., Pavan, D. & Curcio, F.F. (2007) Two new species of lizards of the genus *Bachia* (Squamata, Gymnophthalmidae) from Central Brazil. *Journal of Herpetology*, **41**, 545–553.

Rodrigues, M.T., Camacho, A., Sales Nunes, P.M., Recoder, R.S., Teixeira, M., Valdujo, P.H., Ghellere, J.M.B., Mott, T. & Nogueira, C. (2008) A new species of the lizard

genus *Bachia* (Squamata: Gymnophthalmidae) from the Cerrados of Central Brazil. *Zootaxa*, **1875**, 39–50.

Rodríguez, J.P., Brotons, L., Bustamante, J. & Seoane, J. (2007) The application of predictive modelling of species distribution to biodiversity conservation. *Diversity and Distributions*, **13**, 243–251.

Rodriguez-Sanchez, F., Hampe, A., Jordano, P. & Arroyo, J. (2010) Past tree range dynamics in the Iberian Peninsula inferred through phylogeography and palaeodistribution modelling: A review. *Review of Palaeobotany and Palynology*, **162**, 507–521.

Romiguier, J., Gayral, P., Ballenghien, M., Bernard, A., Cahais, V., Chenuil, A., Chiari, Y., Dernat, R., Duret, L., Faivre, N., Loire, E., Lourenco, J.M., Nabholz, B., Roux, C., Tsagkogeorga, G., Weber, A.A., Weinert, L.A., Belkhir, K., Bierne, N., Glemin, S. & Galtier, N. (2014) Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature*, **515**, 261–263.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S., Larget, B., Liu, L., Suchard, M.A. & Huelsenbeck, J.P. (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539–42.

Rubin, D.B. (2003) Discussion on multiple imputation. *International Statistical Review*, **71**, 619–625.

Rull, V. (2008) Speciation timing and neotropical biodiversity: the Tertiary-Quaternary debate in the light of molecular phylogenetic evidence. *Molecular Ecology*, **17**, 2722-9.

Rull, V. (2011) Neotropical biodiversity: timing and potential drivers. *Trends in Ecology & Evolution*, **26**, 508–13.

Rundle, H.D. & Nosil, P. (2005) Ecological speciation. *Ecology Letters*, **8**, 336-352.

Salgado-Laboriau, M.L. (2005) Alguns aspectos sobre a Paleoecologia dos Cerrados. *Cerrado: Ecologia, Biodiversidade e Conservação.* (ed. by A. Scariot, J.C. Sousa-Silva and J.M. Felfili), pp. 107–118. Ministério do Meio Ambiente, Brasília, Distrito Federal

Sanmartín, I., Van Der Mark, P. & Ronquist, F. (2008) Inferring dispersal: a Bayesian approach to phylogeny-based island biogeography, with special reference to the Canary Islands. *Journal of Biogeography*, **35**, 428–449.

Santos, M.G., Nogueira, C., Giugliano, L.G. & Colli, G.R. (2014) Landscape evolution and phylogeography of *Micrablepharus atticolus* (Squamata, Gymnophthalmidae), an endemic lizard of the Brazilian Cerrado. *Journal of Biogeography*, **41**, 1506–1519.

Satler, J.D., Carstens, B.C. & Hedin, M. (2013) Multilocus species delimitation in a complex of morphologically conserved trapdoor spiders (Mygalomorphae, Antrodiaetidae, *Aliatypus*). *Systematic Biology*, **62**, 805–23.

Schall, J.J. & Pianka, E.R. (1978) Geographical trends in numbers of species. *Science*, **201**, 679–686.

Scheet, P. & Stephens, M. (2006) A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *The American Journal of Human Genetics*, **78**, 629-644.

Scheffers, B.R., Joppa, L.N., Pimm, S.L. & Laurance, W.F. (2012) What we know and don't know about Earth's missing biodiversity. *Trends in Ecology & Evolution*, **27**, 501–10.

Scheffers, B.R., Edwards, D.P., Diesmos, A., Williams, S.E. & Evans, T.A. (2014) Microhabitats reduce animal's exposure to climate extremes. *Global Change Biology*, **20**, 495–503.

Schipper, J., Chanson, J.S., Chiozza, F., Cox, N.A., Hoffmann, M., Katariya, V., Lamoreux, J., Rodrigues, A.S.L., Stuart, S.N., Temple, H.J., Baillie, J., Boitani, L., Lacher, T.E., Mittermeier, R.A., Smith, A.T., Absolon, D., Aguiar, J.M., Amori, G., Bakkour, N.,

Baldi, R., Berridge, R.J., Bielby, J., Black, P.A., Blanc, J.J., Brooks, T.M., Burton, J.A., Butynski, T.M., Catullo, G., Chapman, R., Cokeliss, Z., Collen, B., Conroy, J., Cooke, J.G., da Fonseca, G.A.B., Derocher, A.E., Dublin, H.T., Duckworth, J.W., Emmons, L., Emslie, R.H., Festa-Bianchet, M., Foster, M., Foster, S., Garshelis, D.L., Gates, C., Gimenez-Dixon, M., Gonzalez, S., Gonzalez-Maya, J.F., Good, T.C., Hammerson, G., Hammond, P.S., Happold, D., Happold, M., Hare, J., Harris, R.B., Hawkins, C.E., Haywood, M., Heaney, L.R., Hedges, S., Helgen, K.M., Hilton-Taylor, C., Hussain, S.A., Ishii, N., Jefferson, T.A., Jenkins, R.K.B., Johnston, C.H., Keith, M., Kingdon, J., Knox, D.H., Kovacs, K.M., Langhammer, P., Leus, K., Lewison, R., Lichtenstein, G., Lowry, L.F., Macavoy, Z., Mace, G.M., Mallon, D.P., Masi, M., McKnight, M.W., Medell√≠n, R.A., Medici, P., Mills, G., Moehlman, P.D., Molur, S., Mora, A., Nowell, K., Oates, J.F., Olech, W., Oliver, W.R.L., Oprea, M., Patterson, B.D., Perrin, W.F., Polidoro, B.A., Pollock, C., Powel, A., Protas, Y., Racey, P., Ragle, J., Ramani, P., Rathbun, G., Reeves, R.R., Reilly, S.B., Reynolds, J.E., Rondinini, C., Rosell-Ambal, R.G., Rulli, M., Rylands, A.B., Savini, S., Schank, C.J., Sechrest, W., Self-Sullivan, C., Shoemaker, A., Sillero-Zubiri, C., De Silva, N., Smith, D.E., Srinivasulu, C., Stephenson, P.J., van Strien, N., Talukdar, B.K., Taylor, B.L., Timmins, R., Tirira, D.G., Tognelli, M.F., Tsytsulina, K., Veiga, L.M., Vi√©, J.-C., Williamson, E.A., Wyatt, S.A., Xie, Y. & Young, B.E. (2008) The status of the world's land and marine mammals: diversity, threat, and knowledge. *Science*, **322**, 225–230.

Schlick-Steiner, B.C., Seifert, B., Stauffer, C., Christian, E., Crozier, R.H. & Steiner, F.M. (2007) Without morphology, cryptic species stay in taxonomic crypsis following discovery. *Trends in Ecology & Evolution*, **22**, 391–392.

Schluter, D. (1984) Body size, prey size and herbivory in the Galapagos lava lizard, *Tropidurus*. *Oikos*, **43**, 291–300.

Schluter, D. (2001) Ecology and the origin of species. *Trends in Ecology & Evolution*, **16**, 372–380.

Schluter, D. (2009) Evidence for ecological speciation and its alternative. *Science*, **323**, 737–741.

Schölkopf, B., Smola, A.J., Williamson, R.C. & Bartlett, P.L. (2000) New support vector algorithms. *Neural computation*, **12**, 1207–1245.

Schulte, U., Hochkirch, A., Lötters, S., Rödder, D., Schweiger, S., Weimann, T. & Veith, M. (2012) Cryptic niche conservatism among evolutionary lineages of an invasive lizard. *Global Ecology and Biogeography*, **21**, 198–211.

Schwartz, R.S. & Mueller, R.L. (2010) Branch length estimation and divergence dating: estimates of error in Bayesian and maximum likelihood frameworks. *BMC Evolutionary Biology*, **10**, 5.

Sclater, P.L. (1858) On the general geographical distribution of the members of the Class Aves. *Journal of the Proceedings of the Linnean Society of London. Zoology*, **2**, 130–136.

Shaw, T.I., Ruan, Z., Glenn, T.C. & Liu, L. (2013) STRAW: Species TRee Analysis Web server. *Nucleic Acids Research*, **41**, W238-41.

Sifeddine, A., Spadano Albuquerque, A.L., Ledru, M.-P., Turcq, B., Knoppers, B., Martin, L., Zamboni de Mello, W., Passenau, H., Landim Dominguez, J.M., Campello Cordeiro, R., Abrão, J.J. & da Silva Pinto Bittencourt, A.C. (2003) A 21 000 cal years paleoclimatic record from Caçó Lake, northern Brazil: evidence from sedimentary and pollen analyses. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **189**, 25–34.

Silva, J.F., Farinas, M.R., Felfili, J.M. & Klink, C.A. (2006) Spatial heterogeneity, land use and conservation in the cerrado region of Brazil. *Journal of Biogeography*, **33**, 536–548.

Silva, J.M.C. (1995) Biogeographic analysis of the South American Cerrado avifauna. *Steenstrupia*, **21**, 49–67.

Silva, J.M.C. (1996) Distribution of Amazonian and Atlantic birds in gallery forests of the Cerrado region, South America. *Ornitologia Neotropical*, **7**, 1–18.

Silva, J.M.C. (1997) Endemic bird species and conservation in the Cerrado region, South America. *Biodiversity and Conservation*, **6**, 435–450.

Silva, J.M.C. & Bates, J.M. (2002) Biogeographic patterns and conservation in the South American cerrado: a tropical savanna hotspot. *BioScience*, **52** 225– 233.

Silva, V.N., Pressey, R.L., Machado, R.B., VanDerWal, J., Wiederhecker, H.C., Werneck, F.P. & Colli, G.R. (2014) Formulating conservation targets for a gap analysis of endemic lizards in a biodiversity hotspot. *Biological Conservation*, **180**, 1–10.

Simon, M.F., Grether, R., de Queiroz, L.P., Skema, C., Pennington, R.T. & Hughes, C.E. (2009) Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by in situ evolution of adaptations to fire. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 20359–64.

Sinervo, B., Mendez-de-la-Cruz, F., Miles, D.B., Heulin, B., Bastiaans, E., Villagran-Santa Cruz, M., Lara-Resendiz, R., Martinez-Mendez, N., Calderon-Espinosa, M.L., Meza-Lazaro, R.N., Gadsden, H., Avila, L.J., Morando, M., De la Riva, I.J., Victoriano Sepulveda, P., Rocha, C.F., Ibarguengoytia, N., Aguilar Puntriano, C., Massot, M., Lepetz, V., Oksanen, T.A., Chapple, D.G., Bauer, A.M., Branch, W.R., Clobert, J. & Sites, J.W., Jr. (2010) Erosion of lizard diversity by climate change and altered thermal niches. *Science*, **328**, 894–899.

Sistrom, M., Donnellan, S.C. & Hutchinson, M.N. (2013) Delimiting species in recent radiations with low levels of morphological divergence: a case study in Australian *Gehyra* geckos. *Molecular Phylogenetics and Evolution*, **68**, 135–43.

Sistrom, M., Edwards, D.L., Donnellan, S. & Hutchinson, M. (2012) Morphological differentiation correlates with ecological but not with genetic divergence in a *Gehyra* gecko. *Journal of Evolutionary Biology*, **25**, 647–60.

Sistrom, M., Hutchinson, M., Bertozzi, T. & Donnellan, S. (2014) Evaluating evolutionary history in the face of high gene tree discordance in Australian Gehyra (Reptilia: Gekkonidae). *Heredity*, **113**, 52–63.

Skinner, A. & Lee, M.S. (2010) Plausibility of inferred ancestral phenotypes and the evaluation of alternative models of limb evolution in scincid lizards. *Biology letters*, **6**, 354-8.

Smith, B.T., Harvey, M.G., Faircloth, B.C., Glenn, T.C. & Brumfield, R.T. (2014a) Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. *Systematic Biology*, **63**, 83-95.

Smith, B.T., McCormack, J.E., Cuervo, A.M., Hickerson, M.J., Aleixo, A., Cadena, C.D., Perez-Eman, J., Burney, C.W., Xie, X., Harvey, M.G., Faircloth, B.C., Glenn, T.C., Derryberry, E.P., Prejean, J., Fields, S. & Brumfield, R.T. (2014b) The drivers of tropical speciation. *Nature*, **515**, 406–409.

Smith, C.I., Tank, S., Godsoe, W., Levenick, J., Strand, E., Esque, T. & Pellmyr, O. (2011) Comparative Phylogeography of a Coevolved Community: Concerted Population Expansions in Joshua Trees and Four Yucca Moths. *PLoS ONE*, **6**, e25628.

Smith, S.A., Stephens, P.R. & Wiens, J.J. (2005) Replicate patterns of species richness, historical biogeography, and phylogeny in Holarctic treefrogs. *Evolution*, **59**, 2433–2450.

Smouse, P.E., Whitehead, M.R. & Peakall, R. (2015) An informational diversity framework, illustrated with sexually deceptive orchids in early stages of speciation. *Molecular Ecology Resources*, ***in press***

Solis-Lemus, C., Knowles, L.L. & Ane, C. (2015) Bayesian species delimitation combining multiple genes and traits in a unified framework. *Evolution*, **69**, 492–507.

Soltis, D.E., Morris, A.B., McLachlan, J.S., Manos, P.S. & Soltis, P.S. (2006) Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*, **15**, 4261-93.

Song, S., Liu, L., Edwards, S.V. & Wu, S. (2012) Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. *Proceedings of the National Academy of Sciences*, **109**, 14942–14947.

Sota, T. & Vogler, A.P. (2003) Reconstructing species phylogeny of the carabid beetles *Ohomopterus* using multiple nuclear DNA sequences: heterogeneous information content and the performance of simultaneous analyses. *Molecular Phylogenetics and Evolution*, **26**, 139-154.

Sousa, V. & Hey, J. (2013) Understanding the origin of species with genome-scale data: modelling gene flow. *Nature reviews. Genetics*, **14**, 404-14.

Spix, J.B.v. & Martius, C.F.P.v. (1824) *Travels in Brazil, in the years 1817-1820*. Longman, Hurst, Rees, Orme, Brown and Green, London.

Stace, C.A. (1989) Guest editorial: Dispersal versus vicariance - no contest! *Journal of Biogeography*, **16**, 201–202.

Stadler, T. & Bokma, F. (2013) Estimating speciation and extinction rates for phylogenies of higher taxa. *Systematic Biology*, **62**, 220–230.

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

Strüssmann, C. & Mott, T. (2009) Sympatric amphisbaenids from Manso Dam region, Mato Grosso State, Western Brazil, with the description of a new two-pored species of *Amphisbaena* (Squamata, Amphisbaenidae). *Studies on Neotropical Fauna and Environment*, **44**, 37–46.

Sunnucks, P. & Hales, D.F. (1996) Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). *Molecular Biology and Evolution*, **13**, 510–524.

Swenson, N. (2008) The past and future influence of geographic information systems on hybrid zone, phylogeographic and speciation research. *Journal of Evolutionary Biology*, **21**, 421–434.

Tabachnick, B.G. & Fidell, L.S. (1996) *Using Multivariate Statistics*, 3rd edn. HarperCollins, New York, USA.

Taberlet, P., Fumagalli, L., Wust-Saucy, A.-G. & Cosson, J.-F. (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453-464.

Tajima, F. (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics*, **105**, 437-460.

Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–9.

Tan, D.S.H., Ang, Y., Lim, G.S., Ismail, M.R.B. & Meier, R. (2010) From 'cryptic species' to integrative taxonomy: an iterative process involving DNA sequences, morphology, and behaviour leads to the resurrection of *Sepsis pyrrhosoma* (Sepsidae: Diptera). *Zoologica Scripta*, **39**, 51–61.

Tancoigne, E. & Dubois, A. (2013) Taxonomy: no decline, but inertia. *Cladistics*, **29**, 567–570.

Teixeira, M., Vechio, F.D., Neto, A.M. & Rodrigues, M.T. (2014) A new two-pored *Amphisbaena* Linnaeus, 1758, from Western Amazonia, Brazil (Amphisbaenia: Reptilia). *South American Journal of Herpetology*, **9**, 62–74.

Teixeira, M.J., Recoder, R.S., Camacho, A., De Sena, M.A., Navas, C.A. & Rodrigues, M.T. (2013) A new species of *Bachia* Gray, 1845 (Squamata: Gymnophthalmidae) from the Eastern Brazilian Cerrado, and data on its ecology, physiology and behavior. *Zootaxa*, **3616**, 173–189.

Thorpe, R.S., Surget-Groba, Y. & Johansson, H. (2008) The relative importance of ecology and geographic isolation for speciation in anoles. *Philosophical Transactions of the Royal Society of London B*, **363**, 3071–3081.

Thorpe, R.S., Brown, R.P., Malhotra, A. & Wuster, W. (1991) Geographic variation and population systematics: distinguishing between ecogenetics and phylogenetics. *Italian Journal of Zoology*, **58**, 329–335.

Tinkle, D.W., Wilbur, H.M. & Tilley, S.G. (1970) Evolutionary strategies in lizard reproduction. *Evolution*, **24**, 55–74.

Toews, D.P. & Brelsford, A. (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, **21**, 3907–30.

Tonini, J., Moore, A., Stern, D., Shcheglovitova, M. & Orti, G. (2015) Concatenation and Species Tree Methods Exhibit Statistically Indistinguishable Accuracy under a Range of Simulated Conditions. *PLoS currents*, **7**

Topp, C.M., Pruett, C.L., McCracken, K.G. & Winker, K. (2013) How migratory thrushes conquered northern North America: a comparative phylogeography approach. *PeerJ*, **1**, e206.

Torres-Carvajal, O. & de Queiroz, K. (2009) Phylogeny of hoplocercine lizards (Squamata: Iguania) with estimates of relative divergence times. *Molecular Phylogenetics and Evolution*, **50**, 31–43.

Townsend, T.M., Mulcahy, D.G., Noonan, B.P., Sites, J.W., Jr., Kuczynski, C.A., Wiens, J.J. & Reeder, T.W. (2011) Phylogeny of iguanian lizards inferred from 29 nuclear loci, and a comparison of concatenated and species-tree approaches for an ancient, rapid radiation. *Molecular Phylogenetics and Evolution*, **61**, 363–380.

Turchetto-Zolet, A.C., Pinheiro, F., Salgueiro, F. & Palma-Silva, C. (2013) Phylogeographical patterns shed light on evolutionary process in South America. *Molecular Ecology*, **22**, 1193-213.

Upham, N.S. & Patterson, B.D. (2012) Diversification and biogeography of the Neotropical caviomorph lineage Octodontoidea (Rodentia: Hystricognathi). *Molecular Phylogenetics and Evolution*, **63**, 417-429.

van Doorn, G.S., Edelaar, P. & Weissing, F.J. (2009) On the origin of species by natural and sexual selection. *Science*, **326**, 1704–1707.

Van-Sluys, M. (1993) The reproductive cycle of *Tropidurus itambere* (Sauria: Tropiduridae) in southeastern Brazil. *Journal of Herpetology*, **27**, 28–32.

Van-Sluys, M. (1997) Home range of the saxicolous lizard *Tropidurus itambere* (Tropiduridae) in southeastern Brazil. *Copeia*, **1997**, 623–628.

Van-Sluys, M., Rocha, C.F.D., Vrcibradic, D., Galdino, C.A.B. & Fontes, A.l.F. (2004) Diet, activity, and microhabitat use of two syntopic *Tropidurus* species (Lacertilia: Tropiduridae) in Minas Gerais, Brazil. *Journal of Herpetology*, **38**, 606-611.

Vanzolini, P.E. (1948) Notas sobre os ofídios e lagartos da Cachoeira de Emas no município de Pirassununga, Estado de São Paulo. *Revista Brasileira de Biologia*, **8**, 377–400.

Vanzolini, P.E. (1953a) Sobre a diferenciação geográfica de *Gymnodactylus geckoides* (Sauria, Gekkonidae). *Papéis Avulsos do Departamento de Zoologia* **11**, 225–262.

Vanzolini, P.E. (1953b) Notas sobre alguns lagartos Sul Americanos (Sauria, Gekkonidae). *Revista Brasileira de Biologia*, **13**

Vanzolini, P.E. (1968a) Geography of the South american Gekkonidae (Sauria). *Arquivos de Zoologia*, **17**, 85–112.

Vanzolini, P.E. (1968b) Lagartos Brasileiros da família Gekkonidae (Sauria). *Arquivos de Zoologia*, **17**, 1–84.

Vanzolini, P.E. (1976) On the lizards of a Cerrado-Caatinga contact: evolutionary and zoogeographical implications (Sauria). *Papéis Avulsos de Zoologia*, **29**, 111–119.

Vanzolini, P.E. (1982) A new *Gymnodactylus* from Minas Gerais, Brazil, with remarks on the genus and on montane endemisms in Brazil (Sauria, Gekkonidae). *Papéis Avulsos de Zoologia*, **34**, 403–413.

Vanzolini, P.E. (1997) The *silvestrii* species-group of *Amphisbaena*, with the description of two new Brasilian species (Reptilia: Amphisbaenia). *Papéis Avulsos de Zoologia. São Paulo*, **40**, 65–85.

Vanzolini, P.E. (2004) On the geographical differentiation of *Gymnodactylus geckoides* Spix, 1825 (Sauria, Gekkonidae): Speciation in the Brazilian caatingas. *Anais da Academia Brasileira de Ciências*, **76**, 663–698.

Vanzolini, P.E. (2005) On *Gymnodactylus amarali* Barbour, 1925, with the description of a new species (Sauria, Gekkonidae). *Anais da Academia Brasileira de Ciências*, **77**, 595–611.

Vanzolini, P.E. & Williams, E.E. (1981) The vanishing refuge: a mechanism for ecogeographic speciation. *Papéis Avulsos de Zoologia*, **34**, 251-255.

Vieira, G.H.C., Mesquita, D.O., Péres Jr, A.K., Kitayama, K. & Colli, G.R. (2000) Natural history: Micrablepharus atticolus. *Herpetological review*, **31**, 241–242.

Vincent, B., Dionne, M., Kent, M.P., Lien, S. & Bernatchez, L. (2013) Landscape genomics in Atlantic Salmon (*Salmo salar*): searching for gene–environment interactions driving local adptation. *Evolution*, **67**, 3469–3487.

Vitt, L.J. (1991) An introduction to the ecology of Cerrado lizards. *Journal of Herpetology*, **25**, 79–90.

Vitt, L.J. & Zani, P.A. (1998) Ecological relationships among sympatric lizards in a transitional forest in the northern Amazon of Brazil. *Journal of Tropical Ecology*, **14**, 63-86.

Vitt, L.J. & Pianka, E.R. (2005) Deep history impacts present-day ecology and biodiversity. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 7877–7881.

Vitt, L.J., Caldwell, J.P., Zani, P.A. & Titus, T.A. (1997) The Role of habitat shift in the evolution of lizard morphology: evidence from tropical *Tropidurus*. *Proceedings of the National Academy of Sciences*, **94**, 3828–3832.

Vitt, L.J., Shepard, D.B., Caldwell, J.P., Vieira, G.H.C., Franca, F.G.R. & Colli, G.R. (2007) Living with your food: geckos (*Gymnodactylus carvalhoi*) in termitaria of Cantão. *Journal of Zoology*, **272**, 321–328.

Wagele, H., Klussmann-Kolb, A., Kuhlmann, M., Haszprunar, G., Lindberg, D., Koch, A. & Wagele, J.W. (2011) The taxonomist - an endangered race. A practical proposal for its survival. *Frontiers in zoology*, **8**, 25.

Wallace, A.R. (1869) *The Malay Archipelago: The land of the Orangutan, and the Bird of Paradise*. Harper & Brothers, New York.

Wang, I.J. & Summers, K. (2010) Genetic structure is correlated with phenotypic divergence rather than geographic isolation in the highly polymorphic strawberry poison-dart frog. *Molecular Ecology*, **19**, 447-58.

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

Watterson, G.A. (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, **7**, 256–276.

Webb, S.D. (1978) A History of Savanna Vertebrates in the New World. Part II: South America and the Great Interchange. *Annual Review of Ecology and Systematics*, **9**, 393–426.

Weir, J.T. & Schluter, D. (2008) Calibrating the avian molecular clock. *Molecular Ecology*, **17**, 2321-8.

Werneck, F.P. (2011) The diversification of eastern South American open vegetation biomes: Historical biogeography and perspectives. *Quaternary Science Reviews*, **30**, 1630–1648.

Werneck, F.P. & Colli, G.R. (2006) The lizard assemblage from Seasonally Dry Tropical Forest enclaves in the Cerrado biome, Brazil, and its association with the Pleistocenic Arc. *Journal of Biogeography*, **33**, 1983–1992.

Werneck, F.P., Colli, G.R. & Vitt, L.J. (2009) Determinants of assemblage structure in Neotropical dry forest lizards. *Austral Ecology*, **34**, 97–115.

Werneck, F.P., Costa, G.C., Colli, G.R., Prado, D.E. & Sites Jr, J.W. (2011) Revisiting the historical distribution of Seasonally Dry Tropical Forests: new insights based on palaeodistribution modelling and palynological evidencegeb. *Global Ecology and Biogeography*, **20**, 272–288.

Werneck, F.P., Gamble, T., Colli, G.R., Rodrigues, M.T. & Sites Jr, J.W. (2012a) Deep diversification and long-term persistence in the South American 'Dry Diagonal': Integrating continent-wide phylogeography and distribution modeling of geckos. *Evolution*, **66**, 3014–3034.

Werneck, F.P., Nogueira, C., Colli, G.R., Sites, J.W. & Costa, G.C. (2012b) Climatic stability in the Brazilian Cerrado: implications for biogeographical connections of South American savannas, species richness and conservation in a biodiversity hotspot. *Journal of Biogeography*, **39**, 1695–1706.

Wheeler, Q.D., Knapp, S., Stevenson, D.W., Stevenson, J., Blum, S.D., Boom, B.M., Borisy, G.G., Buizer, J.L., De Carvalho, M.R., Cibrian, A., Donoghue, M.J., Doyle, V., Gerson, E.M., Graham, C.H., Graves, P., Graves, S.J., Guralnick, R.P., Hamilton, A.L., Hanken, J., Law, W., Lipscomb, D.L., Lovejoy, T.E., Miller, H., Miller, J.S., Naeem, S., Novacek, M.J., Page, L.M., Platnick, N.I., Porter-Morgan, H., Raven, P.H., Solis, M.A., Valdecasas, A.G., Van Der Leeuw, S., Vasco, A., Vermeulen, N., Vogel, J., Walls, R.L., Wilson, E.O. & Woolley, J.B. (2012) Mapping the biosphere: exploring species to understand the origin, organization and sustainability of biodiversity. *Systematics and Biodiversity*, **10**, 1–20.

Whittaker, R.J., Araújo, M.B., Jepson, P., Ladle, R.J., Watson, J.E.M. & Willis, K.J. (2005) Conservation Biogeography: assessment and prospect. *Diversity and Distributions*, **11**, 3–23.

Wiederhecker, H.C., Pinto, A.C.S., Paiva, M.S. & Colli, G.R. (2003) The demography of the lizard *Tropidurus torquatus* (Squamata, Tropiduridae) in a highly seasonal Neotropical savanna. *Phyllomedusa*, **2**, 9–19.

Wiens, J.J. (2004) Speciation and ecology revisited: phylogenetic niche conservatism and the origin of species. *Evolution*, **58**, 193–197.

Wiens, J.J. (2007) Species delimitation: new approaches for discovering diversity. *Systematic Biology*, **56**, 875–8.

Wiens, J.J. & Donoghue, M.J. (2004) Historical biogeography, ecology and species richness. *Trends in Ecology & Evolution*, **19**, 639–644.

Wiens, J.J. & Graham, C.H. (2005) Niche conservatism: Integrating evolution, ecology, and conservation biology. *Annual Review of Ecology, Evolution, and Systematics*, **36**, 519–539.

Wiens, J.J. & Morrill, M.C. (2011) Missing data in phylogenetic analysis: reconciling results from simulations and empirical data. *Systematic Biology*, **60**, 719-31.

Wiens, J.J., Kuczynski, C.A., Smith, S.A., Mulcahy, D.G., Sites, J.W., Jr., Townsend, T.M. & Reeder, T.W. (2008) Branch lengths, support, and congruence: testing the phylogenomic approach with 20 nuclear loci in snakes. *Systematic Biology*, **57**, 420–431.

Williams, E.E. & Vanzolini, P.E. (1966) Studies on Soth American Anoles. *Anolis transversalis* A. Duméril. *Papéis Avulsos do Departamento de Zoologia*, **19**

Winston, J.E. (1999) *Describing species: practical taxonomic procedure for biologists*. Columbia University Press, New York.

Wuster, W., Ferguson, J.E., Quijada-Mascarenas, J.A., Pook, C.E., Da Graca Salomao, M. & Thorpe, R.S. (2005) Tracing an invasion: landbridges, refugia, and the phylogeography of the Neotropical rattlesnake (Serpentes: Viperidae: Crotalus durissus). *Molecular Ecology*, **14**, 1095–1108.

Xi, Z., Liu, L., Rest, J.S. & Davis, C.C. (2014) Coalescent versus concatenation methods and the placement of *Amborella* as sister to Water Lilies. *Systematic Biology*, **63**, 919–932.

Xia, X. (2013) DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. *Molecular Biology and Evolution*, **30**, 1720-8.

Xia, X., Xie, Z., Salemi, M., Chen, L. & Wang, Y. (2003) An index of substitution saturation and its application. *Molecular Phylogenetics and Evolution*, **26**, 1-7.

Xue, C., Li, F., He, T., Liu, G.P., Li, Y. & Zhang, X. (2005) Classification of real and pseudo microRNA precursors using local structure-sequence features and support vector machine. *BMC Bioinformatics*, **6**, 310.

Yang, Z. (2002) Likelihood and Bayes estimation of ancestral population sizes in hominoids using data from multiple loci. *Genetics*, **162**, 1811–1823.

Yang, Z. (2015) BPP 3.1: A Bayesian program for analysis of genomic sequence data under the multispecies coalescent model. *Current Zoology*, ***in press***

Yang, Z. & Rannala, B. (2010) Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences*, **107**, 9264–9269.

Yang, Z. & Rannala, B. (2014) Unguided species delimitation using DNA sequence data from multiple Loci. *Molecular Biology and Evolution*, **31**, 3125-35.

Yonenaga-Yassuda, Y. & Rodrigues, M.T. (1999) Supernumerary chromosome variation, heteromorphic sex chromosomes and banding patterns in microteiid lizards of the genus *Micrablepharus* (Squamata, Gymnophthalmidae). *Chromosome Research*, **7**, 21–29.

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

Zhang, C., Zhang, D.X., Zhu, T. & Yang, Z. (2011) Evaluation of a Bayesian coalescent method of species delimitation. *Systematic Biology*, **60**, 747–61.

Zhu, T. & Yang, Z. (2012) Maximum likelihood implementation of an isolation-with-migration model with three species for testing speciation with gene flow. *Molecular Biology and Evolution*, **29**, 3131-42.