UNIVERSIDADE DE BRASÍLIA

INSTITUTO DE CIÊNCIAS BIOLÓGICAS

PROGRAMA DE PÓS-GRADUAÇÃO EM ZOOLOGIA

FILOGENIA MOLECULAR, BIOGEOGRAFIA E MODELAGEM DE DISTRIBUIÇÃO DAS ABELHAS *LANTHANOMELISSA* – UM GÊNERO ENDÊMICO DOS CAMPOS DO SUL DA AMÉRICA DO SUL

TAÍS MATTOSO DE ANDRADE RIBEIRO

Brasília/DF

2018

UNIVERSIDADE DE BRASÍLIA

INSTITUTO DE CIÊNCIAS BIOLÓGICAS

PROGRAMA DE PÓS-GRADUAÇÃO EM ZOOLOGIA

FILOGENIA MOLECULAR, BIOGEOGRAFIA E MODELAGEM DE DISTRIBUIÇÃO DAS ABELHAS *LANTHANOMELISSA* – UM GÊNERO ENDÊMICO DOS CAMPOS DO SUL DA AMÉRICA DO SUL

TAÍS MATTOSO DE ANDRADE RIBEIRO

Orientador: Prof. Antonio J. C.Aguiar

Dissertação apresentada ao Instituto de Ciências Biológicas da Universidade de Brasília como parte dos requisitos para a obtenção do título de Mestra em Zoologia.

Brasília/DF

Junho de 2018

Agradecimentos

Agradeço especialmente à minha família: meus pais Adriana e Reginaldo, meu irmão Lucas, minha tia Léa e à Mandy por todo o apoio incondicional tanto no âmbito emocional quanto no financeiro além de sempre me incentivarem a crescer. Não teria conseguido nem seria nada sem vocês!

Ao Prof. Antonio Aguiar e à Dra. Aline Martins por serem orientadores absolutamente presentes, sempre se preocupando com o andamento do trabalho e contribuindo de todas as formas possíveis com novas ideias e propostas, nunca me deixando desamparada. Aprendi muito com vocês.

Ao Prof. Daniel Silva, colaborador com experiência em modelagem de distribuição de espécies que é super animado e sempre disposto a ajudar.

À Fundação de Apoio à Pesquisa do Distrito Federal (FAPDF) pela bolsa recebida durante os dois anos de mestrado e recursos de projetos que possibilitaram a realização da pesquisa.

À banca, composta pela Profª Lilian Giugliano, Dra Kelli Ramos e Prof. Renato Caparroz, por terem aceitado o convite e por todas as críticas construtivas que podem trazer para o trabalho. Ao professor Renato também agradeço por ceder o acesso ao espaço e termociclador do Laboratório de Genética e Biodiversidade da UnB para PCRs adicionais, e aos membros deste laboratório pela ajuda e gentileza. Ao Prof. Reginaldo Constantino e novamente ao Prof. Antonio Aguiar pela estruturação do Laboratório de Biologia Molecular do Departamento de Zoologia, onde todas as extrações e quase todas as amplificações de DNA foram feitas, sendo o primeiro trabalho realizado neste laboratório.

Agradeço àqueles que contribuiram permitindo o estudo do material ou através da doação de espécimes para o estudo molecular: Andrea Cocucci, Alicia Sersic, Birgit Harter Marques, Clemens Schlindwein, Reisla Oliveira, Eduardo Almeida, Gabriel Melo, Jerome Rozen, Juan Pablo Torreta, Mabel Lizaraso, Kelli Ramos, Leopoldo Alvarez, Rafael Ferrari, Rodrigo Gonçalves e Vicent Lee. À Mardiore Pinheiro pelas belíssimas fotos de *Lanthanomelissa* e ao Eduardo Almeida pelas diversas dicas.

Aos colegas do Laboratório de Hymenoptera da UnB, ou "Labelhas": Hanna Heid, Matheus Viana, Wagner Silva, Hannah Cascelli, Matheus Carvalho, Bruno Gonçalves e Rogério Andrade, e aos agregados Joice Paz, Jorge Laerson e Victor Silva pela companhia, momentos de descontração, discussões construtivas ou simplesmente por compartilhar fotos fofas de abelhas. Ao Victor também pela ajuda com modelagem e ao Flávio Duque pela revisão ortográfica.

Ao Phillipe Cedro pelo companheirismo, apoio, incentivo, ajuda inestimável com todas as figuras para o trabalho e por querer ajudar sempre com tudo que estivesse ao seu alcance.

Aos Bioamiguinhos Flávio, Fred, Fabíola, Fabiano, Sarah, Nathália, Dariane, Flávia, Julia e Maísa por todos os momentos de diversão e conversas filosóficas sobre o destino da nossa profissão desde a graduação.

A todos vocês: MUITO OBRIGADA!

Resumo

Os Campos Sulinos incluem o bioma Pampa e parte do bioma da Mata Atlântica, representado pelos Campos de Cima da Serra. Pouco se sabe sobre os eventos que levaram à origem destes campos, porém eventos tectônicos no Neogeno e flutuações climáticas no Pleistoceno podem ter influenciado na origem e distribuição de espécies presentes nesta região. Estimativas da origem e divergência de grupos endêmicos representam a principal ferramenta para o entendimento dos processos biogeográficos que podem ter ocorrido sobre estas áreas. Neste sentido, este trabalho busca contribuir para a biogeografia dos Campos Sulinos tentando responder quando e onde ocorreu a origem e diversificação das abelhas coletoras de óleo do gênero *Lanthanomelissa* Holmberg 1903 e como foi a dinâmica da distribuição destas espécies sobre a região desde o último interglacial. Com este objetivo, foi construída a filogenia molecular datada a partir de 37 terminais incluindo mais de uma amostra para cada uma das cinco espécies do gênero e grupos externos, além da reconstrução das possíveis áreas ancestrais. Também foi modelada a distribuição destas espécies para as condições climáticas e de precipitação do presente e em dois tempos pretéritos – o último interglacial (~121 kya) e o último máximo glacial (~21 kya). A origem de *Lanthanomelissa* foi estimada entre o Oligoceno e o Mioceno (ca. 22 Ma) em uma possível área ancestral compartilhada pelo Chaco e pelo Pampa, indicando a origem destes biomas nesta época. Esta área foi, possivelmente, fragmentada pelo complexo de introgressões marinhas no sul da América do Sul, como o Mar Patagônico e Mar Paranense, que fragmentaram a região desde o fim do Oligoceno ao início do Mioceno. Eventos de diversificação teriam ocorrido pela expansão das gramíneas no Mioceno tardio, de maneira semelhante ao observado em espécies do Cerrado. As variações climáticas do Pleistoceno, por serem mais recentes do que a especiação de *Lanthanomelissa* há ca de 17 Ma, influenciaram a distribuição dessas abelhas pela complexa topografia do sudeste da América do Sul. Como também já observado em outros grupos endêmicos do bioma Pampa, para algumas espécies há a distribuição disjunta entre as populações, com uma na região costeira e outra na região interior oeste, provavelmente remanescentes da distribuição destas espécies no último máximo glacial. A filogenia molecular datada e associação com o estudo minucioso da distribuição potencial permitiu delinear a evolução deste grupo de abelhas tão quanto gerar hipóteses sobre o bioma relacionado, permitindo estimar a origem do bioma Pampa para pelo menos o Mioceno inferior, uma idade muito mais antiga que a do Cerrado.

Palavras chave: Hymenoptera, Abelhas coletoras de óleo, Tapinotaspidini, Campos Sulinos, Pampa , Savanas

Abstract

Southern Grasslands include the Pampa and part of the Atlantic Forest biome, represented by the Subtropical Highland Grasslands. Little is known about the events that led to the origin of these grasslands. However, tectonic events in the Neogene and Pleistocene climatic fluctuations could have driven the origin and distribution of species inhabiting this region. Estimates on the origin and divergence of endemic groups are the main tool for the understanding of biogeographic processes that could have happened on these areas. In this sense, this work aims to contribute for the biogeography of the Southern Grasslands by answering when and where the oil-collecting bees from the genus *Lanthanomelissa* Holmberg 1903 originate and diversified and how was the distribution dynamic of these species on this region since the last interglacial. For that we constructed the dated molecular phylogeny from 37 terminals, including more than one specimen for each of the five species from the genus and outgroups. We have also studied the historic biogeography by reconstructing the ancestral areas. Besides that, we modelled the species distribution for the climatic conditions on current and on two past scenarios: the last interglacial (ca. 121 kya) and the last glacial maximum (ca. 21kya). The origin of *Lanthanomelissa* was estimated between the Oligocene and Miocene (ca. 22 Mya), in a possible ancestral area shared by the Chaco and Pampa, indicating the origin of these biomes at this time. This area could have been fragmented by the complex of marine introgressions in the Southern South America, such as the Patagonian and Paranean Seas, which fragmented this region from the late Oligocene to early Miocene. Diversification events could be related to the grassland expansion in late Miocene, similar to what happened with species from Cerrado. Pleistocene climatic fluctuations, being more recent than the speciation of *Lanthanomelissa* ca. 17 Mya, influenced these bees distribution through south eastern South America's complex topography. Similar to other groups endemic to the Pampa biome, for some species there is a separate distribution between populations – one in the coast and one in the interior west region – probably reminiscent from these species distribution on the last glacial maximum. The dated molecular phylogeny in association with the analysis of the potential distribution allowed the understanding of the evolution of this bee genus as well as to estimate the origin of the related Pampa biome to at least the early Miocene, an age much older than the one estimated for the Cerrado.

Keywords: Hymenoptera, Oil-collecting bees, Tapinotaspidini, South Brazilian Grasslands, Pampa, Savannas

Lista de Figuras

Apêndice

Lista de Tabelas

Apêndice

Sumário

I ntrodução Geral

Os campos do sudeste da América do Sul

Campos Sulinos é o termo utilizado para designar as áreas nos três estados da Região Sul do Brasil com vegetação campestre, caracterizada pela predominância de gramíneas (Poaceae) e herbáceas da família Asteraceae (Boldrini, 2009; Overbeck et al., 2015). Segundo a regionalização do IBGE (2004), os Campos Sulinos estão incluídos em dois biomas distintos: nos estados do Paraná, Santa Catarina e norte do Rio Grande do Sul pertencem à Mata Atlântica; e no sul do Rio Grande do Sul pertencem ao Pampa.

Dentro da Mata Atlântica, os Campos Sulinos são representados pelos Campos de Cima da Serra, ou Campos do Planalto Meridional, ou ainda "*Subtropical Highland Grasslands"* (SHG)*,* formando mosaicos com as Florestas de Araucária. Os Campos de Cima da Serra ocupam uma área de mais de 13 mil km², em altitudes que variam de 700 m a 1300 m, chegando até aos 1800 m em algumas regiões (Overbeck et al., 2015). Segundo levantamento florístico realizado por Iganci et al. (2011), estes Campos possuem alto nível de endemismo, sendo que cerca de 25% das suas espécies de plantas são endêmicas. O clima nessa região é úmido, subtropical e subtemperado e o solo é relacionado com afloramentos de rocha basáltica da Serra Geral (Boldrini, 2009). Esta região faz parte da Província Floresta de Araucária na regionalização geográfica proposta por Morrone (2014) e da província do Paraná na proposta de Cabrera & Willink (1973).

O bioma do Pampa tem distribuição no Uruguai, Argentina e no sul do estado brasileiro do Rio Grande do Sul, incluindo a Serra do Sudeste (IBGE 2004), com área de 760 mil km² (Overbeck et al., 2015). São áreas de baixa altitude, com clima temperado e diversas fitofisionomias com composição florística e condições geomorfológicas distintas que ocorrem em mosaico. A regionalização do Pampa proposta pelo IBGE é, em grande parte, congruente com as propostas por Morrone (2014) e Cabrera & Willink (1973), que denominam essa região como "Província Pampeana". Entretanto, a Província Pampeana desses autores difere levemente da proposta do IBGE em termos de delimitação, principalmente por esta última ser restrita ao Brasil.

A origem do Pampa e Campos de Cima da Serra é pouco conhecida. O Pampa é incluídos na subregião Chacoana (Morrone, 2000), juntamente com os biomas da Diagonal de Áreas Abertas, Chaco, Cerrado e Caatinga, sugerindo uma relação entre a origem destes

biomas, que também é corroborada pela inclusão dos Pampas em agrupamento do Cerrado e Chaco por Antonelli et al. (2018; modificado de Olson et al., 2001). Além disto, O Pampa apresenta mosaico com manchas de vegetação de outros biomas como manchas de vegetação xerofítica típica de Chaco além de vegetação florestal típica de Mata Atlântica, assim como há manchas de vegetação campestre nestes outros biomas, sugerindo também relação entre eles (Zanella, 2002; Ramos & Melo, 2010; Ferretti et al., 2014).

Os Campos Sulinos passaram por intensas mudanças em sua vegetação devido a variações climáticas durante o Quaternário. Durante o Último Interglacial (ca 121 mil anos antes do presente (ka ap)) o clima era mais quente e úmido (Otto-Bliesner et al., 2006), favorecendo a expansão de floresta de nos campos. Durante o Último Máximo Glacial (ca. 21 ka ap) a vegetação era majoritariamente campestre e o clima mais seco e frio do que o atual. Essa situação prevaleceu durante o Pleistoceno tardio, o Holoceno inferior e médio. Entretanto, durante o Holoceno Superior, a partir de 3 ka ap, a precipitação aumentou em frequência e abundância e, com isso, a Floresta de Araucária começou a se expandir a partir de refúgios florestais para áreas mais altas ao longo de rios. Após 1 ka ap as condições climáticas se tornaram semelhantes às encontradas atualmente: úmidas e sem estação seca definida, de modo que a expansão florestal propriamente dita começou e a floresta substituiu os campos até mesmo em áreas mais elevadas (Behling, 1997, 2002, Safford, 1999, 2007; Behling et al., 2004; Behling & Pillar, 2007).

As subsequentes expansões e retrações de florestas de Araucária nos Campos de Cima da Serra (SHG) teriam levado à especiação alopátrica e à diversificação das espécies de plantas nesta região (Lorenz-Lemke et al., 2010). Entretanto, em estudo com plantas dos gêneros *Petunia* e *Calibrachoa* (Convolvulaceae), que ocorrem na região dos Campos Sulinos, Fregonezi et al. (2013) sugerem que o Pampa não devem ter sido afetados por essas expansões já que os primeiros registros polínicos das florestas de Araucária na região são de apenas 5 mil anos atrás (Behling & Pillar, 2007). Neste caso, as pressões evolutivas sofridas por estas plantas estariam relacionadas a fatores ecológicos como a interação com polinizadores e a fatores abióticos como diferentes tipos de solo, e não a fatores climáticos (Fregonezi et al., 2013). Essas diferentes pressões evolutivas separam mais ainda as áreas campestres do bioma da Mata Atlântica e do Pampa. Neste contexto, Fregonezi et al. (2013) ressaltam a importância de mais estudos com outras espécies relacionadas a estas áreas, especialmente no contexto das interações entre plantas e polinizadores.

É curioso notar que as condições climáticas atuais na região do Pampa favoreceriam a expansão de florestas e formação de mosaicos expressivos como nos campos da Mata

Atlântica. Porém, apesar de haver presença de floresta, este estrato não chega a dominar a paisagem, que é predominantemente campestre em todas as fisionomias do bioma. Este é o chamado "Problema do Pampa" (Walter 1967, Eriksen 1978, Box 1986 *apud* Overbeck et al., 2007). A resposta para isto provavelmente está ligada aos tipos de solo presentes no Pampa, que não favorecem a presença de florestas (Boldrini, 2009). Além disto, o fogo e pastejo seriam fatores importantes para a manutenção da vegetação nativa campestre.

A conservação dos Campos Sulinos é negligenciada, com menos de 40% da vegetação campestre nativa ainda remanescente (Overbeck et al., 2007). Esta foi, em grande parte, substituída pela agricultura e silvicultura (especialmente *Pinus* e *Eucaliptus*). Outra ameaça grande é o sobrepastejo, ou seja, muito gado mantido por muito tempo em área pequena, que acarreta no esgotamento da vegetação campestre e compactação do solo (Vélez-Martin et al., 2015).

As abelhas do gênero Lanthanomelissa *como modelo de estudo*

As abelhas são as principais agentes polinizadoras, já que dependem inteiramente dos recursos florais, principalmente néctar e pólen, para a sua sobrevivência (Michener, 2007). Além de visitarem as flores para a coleta destes para a sua alimentação, algumas abelhas também visitam plantas específicas em busca de outros recursos, tais como óleos florais (Vogel, 1974). As abelhas coletoras de óleo pertencem às subfamílias Melittinae (*Rediviva* e *Macropis*), e Apinae, e nesta inclui o gênero Paleotropical *Ctenoplectra* e as linhagens neotropicais *Centris, Epicharis,* Tetrapediini e Tapinotaspidini (Rasmussen et al., 2000; Renner & Schaefer, 2010; Martins et al., 2014).

Dentro da tribo Tapinotaspidini, o gênero *Lanthanomelissa* tem sua distribuição fortemente associada aos campos sulinos, com distribuição restrita no sudeste da América do Sul, na Argentina, Uruguai, Paraguai e Brasil, incluindo os estados de São Paulo, Paraná, Santa Catarina e Rio Grande do Sul (Urban, 1995; Aguiar, 2012; Souza, 2017). É composto por cinco espécies: *L. betinae* Urban*,* 1995*, L. clementis* Urban, 1995*, L. discrepans* Holmerg, 1903*, L. magaliae* Urban, 1995 e *L. pampicola* Urban, 1995. São abelhas especializadas na coleta de óleo em flores do gênero *Sisyrinchium* (Iridaceae) e, tanto machos quanto fêmeas, possuem cerdas foliáceas no basitarso anterior especializadas para a coleta deste recurso (Vogel, 1974).

As abelhas do gênero *Lanthanomelissa* – endêmicas dos campos do sudeste da América do Sul– são o foco do presente estudo que tenta responder as seguintes perguntas: (1)

Quando e em que região biogeográfica *Lanthanomelissa* originou-se? (2) A origem do gênero estaria associada a períodos de expansão campestre? (3) Flutuações climáticas do Quaternário e trocas entre vegetação campestre e florestal alteraram a distribuição destas abelhas?

Para isso, foram aplicadas ferramentas da filogenética molecular, biogeografia histórica e modelagem de distribuição de espécies. O trabalho será apresentado a seguir, como um capítulo único, em inglês e formatado para a publicação na revista *Journal of Biogeography*.

Time of diversification and niche modelling of the oil-collecting bee genus *Lanthanomelissa* (Apidae) shed light on the evolution of grasslands in south-eastern South America

(este capítulo foi escrito seguindo as normas de submissão da revista Journal of Biogeography)

Taís M. de A. Ribeiro[1],

Colaboradores:

Aline C. Martins[2], Daniel de Paiva Silva[3], Antonio J. C. Aguiar[1]

[1]Departamento de Zoologia, Universidade de Brasília, Brasília, Distrito Federal, Brazil;[2]Departamento de Botânica, Universidade Federal do Paraná, 81531-980 Curitiba, Paraná, Brazil.; [3]Instituto Federal Goiano, Campus Urutaí, Urutaí, Goiás, Brazil.

Biosketch

All authors except for Daniel P. Silva are members of the Hymenoptera Laboratory at University of Brasília. The main focus of the group is studying systematic and biogeography of Neotropical bees. Daniel Silva leads the Conservation Biogeography and Macroecology lab at Instituto Federal Goiano. TMAR designed research, collected and analysed data and wrote the manuscript. All authors reviewed and helped writing the manuscript. DPS analysed data. AJCA collected data and designed research. ACM collected and analysed data.

ABSTRACT

Aim: *Lanthanomelissa* bees are restricted to the south-eastern grasslands of South America: an endangered and still poorly known environment. We aimed to understand the origin of this group in time and space and also the influence of Quaternary climatic fluctuations on its evolution, and possible link to the history of the Southern Grasslands.

Location: The Pampa and Subtropical Highland Grasslands in south-eastern South America.

Taxon: Bees of the *Lanthanomelissa* genus (Hymenoptera, Apidae, Tapinotaspidini)

Methods: We inferred phylogenetic relationships and divergence times of *Lanthanomelissa* species using 37 terminals and 3430 nucleotides of three mitochondrial and two nuclear markers. We applied a secondary calibration and uncorrelated log normal clock to estimate divergence times. The dated phylogeny was used to estimate ancestral ranges in BioGeoBEARS under six different models. To analyse species distribution during current and two past climatic scenarios (LIG, ~120 kya and LGM, ~21 kya) we performed an ensemble with the models GLM, Maxent and Random Forest in a dataset of 196 georeferenced occurrence points using all 19 WorldClim bioclimatic variables.

Results: *Lanthanomelissa* originated at the Oligocene-Miocene border in the Chacoan-Pampean region. During Quaternary, large areas remained stable since the last 120 ky for most species, including great part of southern Brazil. During LIG the suitable area was narrower than in LGM.

Main Conclusions: Miocene tectonic events such as the Andean orogeny and the marine incursions that led to the formation of the Patagonian and Paranean seas could have influenced the expansions of flood plains and grasslands leading to the origin and diversification of *Lanthanomelissa*. Quaternary climatic fluctuations influenced mostly in distribution and intraspecific relations in *Lanthanomelissa*.

Keywords: Grasslands, Molecular phylogenetics, historical biogeography, Hymenoptera, oil-collecting bees, Pampa, Tapinotaspidini

1. INTRODUCTION

South American Grasslands, or Campos Sulinos, comprehend a large open vegetation area included in two Brazilian biomes: Atlantic Forest and Pampa. On Atlantic Forest they are represented by the Subtropical Highland Grasslands (SHG), which are patches of grasslands that occur in mosaic with *Araucaria* Forest at elevations from 700 to 1300 m (Overbeck et al., 2015). The Pampa, include the southern Rio Grande do Sul (IBGE, 2004) and parts of Uruguay and Argentina (Cabrera & Willink, 1973; Overbeck et al., 2007; Morrone, 2014). Its floristic composition is variable according to altitudinal changes.

The origin of the Pampa and SHG are still unresolved. The Pampa is part of the Chacoan biogeographic subregion, together with the biomes included in the South American Diagonal of Open Formations: Chaco, Cerrado and Caatinga; suggesting that their origins could be connected (Morrone, 2000). Biogeographic studies with solitary bees (Zanella, 2002; Ramos & Melo, 2010) and lizards (Ferretti et al., 2014) suggest that the Pampa is related to Chaco, Atlantic forest and Montes (Porzecanski & Cracraft, 2005), as it contains patches of typical vegetation from these biomes in mosaic, such as xerophytic cactus *Opuntia* from Chaco, and patches of Atlantic forest.

Neogene events such as tectonics and marine introgressions (Hoorn et al., 2010; Hughes et al., 2013; Antonelli et al., 2018) and Pleistocene climatic fluctuations (Hewitt, 2004; Carnaval & Moritz, 2008; Carnaval et al., 2009; Ramos-Fregonezi et al., 2015; Silva et al., 2018) were suggested to drive speciation and diversification in south-eastern South America biota. During the quaternary there were several events of global climate cooling and warming (Hewitt, 2004). In the Last Interglacial (LIG ca. 120 kya) climate was warmer and humid (Otto-Bliesner et al., 2006). According to palynologic studies carried out in the last decades (Behling, 1997, 2002, Safford, 1999, 2007; Behling et al., 2004; Behling & Pillar, 2007), these climate fluctuations led to deep changes in south-eastern South America vegetation. In the Last Glacial Maximum (LGM, ca 21 kya (thousand years ago)), grasslands were dominant and climate was drier and colder than in the present. Only starting at 3 kya, precipitation became more frequent and abundant allowing the expansion of *Araucaria* Forest even in more elevated regions.

In SHG, these fluctuations were suggested to be important drivers of speciation in this region, while for Pampa these are not strongly observed, since there are only recent registers

for *Araucaria* Forest. In Pampa, abiotic factors (e.g. soil diversity) and ecological relations (e.g. intraspecific competition and pollination) are suggested as selective pressures. However, how those factors interact (Fregonezi et al., 2013) and the relationship between these two areas using groups that inhabit both environments (Peres et al., 2015) are unclear.

Bees are the most important angiosperm pollinators, relying entirely on flower resources, such as nectar and pollen for feeding themselves and their larvae (Michener, 2007). Some bees also collect floral oils for feeding their larvae and water-proofing their brood cells (Vogel, 1974).

Bees from the tribe Tapinotaspidini are all floral oil-collectors. Within this tribe, the genus *Lanthanomelissa* (Figure 1) is endemic to the Campos of southern Brazil, Argentina and Paraguay, including Pampa and SHG. The genus comprises five species: *L. betinae* Urban, 1995, *L. clementis* Urban, 1995, *L. discrepans* Holmberg, 1903, *L. magaliae* Urban, 1995 and *L. pampicola* Urban, 1995 (Urban, 1995). However, recent morphological phylogeny and taxonomic revision of the genus (Souza, 2017) suggest the inclusion of the species *Chalepogenus parvus* Roig-Alsina, 1997 as a *Lanthanomelissa* species. Moreover, Michener & Moure (1957) state that the *Chalepogenus* species *C. goeldianus* (Friese, 1899), *C. neffi* Roig-Alsina, 1999 and *C. luciane* (Urban, 1995) should constitute the subgenus *Lanthanomelissa* (*Lanthanella*), relation which is also corroborated by the morphological phylogeny (Souza, 2017). The taxonomic treatment in this work was conservative for *Lanthanomelissa*, considering only the five species according to Urban (1995), despite these possible new arrangements for the species of outgroup taxa.

These bees are oil-collecting specialists in flowers of the genus *Sisyrinchium* (Iridaceae) (Cocucci & Vogel, 2001). This is the most diverse genus from the family Iridaceae in the New World (ca. 140 species) and has a wide distribution, covering a great part of the American continent. The highest diversity of *Sisyrinchium* overlaps with the distribution of *Lanthanomelissa* bees – their main pollinators (Cocucci & Vogel, 2001; Chauveau et al., 2011, 2012).

The study of bee speciation through molecular phylogenetics and distribution models is interesting for understanding biogeographic patterns (e.g: Hedtke et al., 2013; Hines, 2008; Praz and Packer, 2014; Ramírez et al., 2010), as they – specially the ones that are oligolectic to some plants – are intrinsicaly connected to the vegetation that characterizes a biome. In this context, we studied the origin and distribution of five species from the oil-collecting bee genus *Lanthanomelissa*. Due to its restricted distribution and stable taxonomy,

Lanthanomelissa is an interesting model for the biogeography of the grassland areas in south-eastern South America.

To investigate the diversification of *Lanthanomelissa* in time and space, as well as the biogeographic history of the south-eastern South America grasslands, we used molecular phylogenetics, ancestral area reconstructions and species distribution modelling. Specifically, we wanted to address the following questions: (1) When and in which biogeographic region *Lanthanomelissa* arose? (2) The origin of this genus is associated to periods of climatic conditions favourable to the expansion of grasslands in South America? (3) Did Quaternary climatic fluctuations and related shifts in vegetation alter the distribution of *Lanthanomelissa*?

Figure 1. *Lanthanomelissa discrepans* in *Sisyrinchium* flower. Photo: Mardiore Pinheiro

2. MATERIALS AND METHODS

2.1 *Taxon sampling*

We analysed all five species recognized in *Lanthanomelissa* (sensu Urban, 1995) as ingroup in the molecular phylogeny. We have chosen specimens from different localities in order to obtain potential variation for each species. We made an effort to sample all the known distribution from the species, including high and lowland grasslands in the Brazilian states of São Paulo, Santa Catarina, Paraná, and Rio Grande do Sul, including the Serra do Sudeste. In total, we have generated 93 sequences, all deposited in GenBank (Table 1). As outgroup, we used 16 sequences from the Tapinotaspidini genera *Chalepogenus, Arhysoceble* and *Trigonopedia*, obtained from A. Aguiar. A previous analysis containing all genera of the

tribe (Aguiar et al. in prep) already observed these genera as sister groups of *Lanthanomelissa*. Specimen data and GenBank numbers are listed in Table 1.

Most specimens analysed were preserved in 95% EtOH, but we have also included pinned specimens collected less than 10 years ago (Table A 2). Vouchers are deposited in University of Brasilia's Entomological Collection or in the collections listed on Table 1.

2.2 *Molecular data sampling*

For most specimens, we extracted DNA from the whole body of the bees, by soaking them in a mixture of 180 µl digestion buffer and 20 µl Proteinase K. We preferred this non-destructive procedure since it allows eventual morphological analysis. For some specimens, specially the fresher ones, we extracted DNA from wing muscle tissue and others from mid legs. For DNA extractions we used the following kits: Purelink Miniprep (Invitrogen), Nucleospin XS (Machery Nagel), Nucleospin Insect (Machery Nagel), and DNeasy Blood and Tissue (Qiagen) according to the respective instructions for animal tissue in the handbooks. Tissues extracted and extraction kits used for each specimen are provided on Table A 2. For pinned specimens, we adopted the following modifications proposed by (Evangelista et al., 2017): Incubation in 70% EtOH for 24h for decontamination and rehydration; and incubation in Elution Buffer for 24h prior to digestion. We immersed the samples on a mixture of 180 µl of Digestion buffer and 20 µl of Proteinase K, provided in extraction kits for 2 – 36 h at room temperature and then let them stay overnight (ca. 12h) on a dry bath set at 55ºC. We followed steps for binding and washing DNA according to the kits manuals and eluted using 50 µl of elution buffer. We performed a second elution with the same volume and stored separately.

We amplified and sequenced mitochondrial markers cytochrome oxidase subunit I (COI, ~700 bp), cytochrome B (CytB, ~600 bp) and 16S (~500 bp). We used mitochondrial markers because they mutate at high speeds and then provide useful information as well as differentiation between species (Danforth et al., 2005, 2013). We also sequenced the nuclear markers longwave rhodopsin (LW- rhodopsin, ~800 bp) and elongation factor 1-α F2 copy (EF-1α, ~1100 pb). We amplified the genes through Polymerase Chain Reaction (PCR) under the following conditions: an initial denaturation at 95°C for 5 min, followed by 36 cycles of denaturation at 95°C for 60 s, annealing at 50–56°C, 60 s, and extension at 68°C, 60–90 s. Primers and specific conditions for each locus amplified are listed on Table A 3. For all

reactions there was a positive control (DNA extracted from fresh *Apis mellifera* muscle tissue) and a negative control (with ddH_2O replacing the DNA volume).

To check if the desired fragments were amplified in PCR, we combined 2 µl of each product with 2 µl of a mixture of Gel Red and Blue Juice. We loaded this mixture into a 1% agarose gel for electrophoresis at constant voltage of 95 V and electrical current set at 200 A for 30 – 45 min. We conducted all extraction and amplification procedures in the Molecular Biology Laboratory at the Zoology Department of University of Brasilia. We sent samples that showed a bright band after electrophoresis to the company Macrogen (South Korea) for purification and sequencing. We assembled assembled, trimmed, edited and BLAST searched the yielded sequences using the GenBank database on Geneious 8.1.9 (Kearse et al., 2012)

Table 1. Taxon sampling used in this study from A. *Lanthanomelissa* and B. outgroup, with collecting data and voucher information, for five markers: the nuclear Elongation factor 1-α (EF-1-α) and Long wave Rhodopsin (LW-Rhodopsin) and the mitochondrial Cytochrome-oxidase subunit 1(CO1), Cytochrome B (CytB), and the ribosomal 16S. Voucher: entomological collection number from DZUB (Department of Zoology, University of Brasilia)

Species	Collecting Data	Voucher	Extraction Code	CO1	CytB	16S	EF-1α	LW-Rhodopsin
A. *Lanthanomelissa*								
Lanthanomelissa betinae	"Brasil, Paraná, Quatro Barras, 9.nov.2011"	-	AA179	-	-	-	KX064547	-
Lanthanomelissa betinae	"Brasil, RS, Canela, P.E. Caracol, -29.307670 -50.841542, 04/11/16"	DZUB3488	TR035	MH213656	MH213629	MG894390	-	-
Lanthanomelissa betinae	"Brasil, São Paulo, Cotia, 20.xi.2011"	-	TR012	MH213653	MH213626	MG894379	-	-
Lanthanomelissa betinae	"Canela, RS, pt 4, -29.3462313, -50.8425992,04.xi.2016"	DZUB3471	TR033	MH213655	MH213628	-	-	-
Lanthanomelissa betinae	"Caracol, RS, pt5, 4.xi.2016, -29.307670 -50.841542"	DZUB3456	TR020	MH213654	MH213627	MG894381	-	-
Lanthanomelissa betinae	"Gramado, RS, pt2, 4/11/16, -29.360437 -50.839575"	DZUB3468	TR032	-	-	MG894388	-	-
Lanthanomelissa betinae	"Maracajá, SC; 31/10/2017; -28.841985\ -49.436212"	DZUB3518	TR055	MH213657	-	-	-	-
Lanthanomelissa betinae	"Maracajá, SC; 31/10/2017; -28.841985\ -49.436212"	DZUB3525	TR056	MH213658	MH213630	-	MG888653	MG889473
Lanthanomelissa betinae	"Palmeira, PR, Recanto dos Papagaios, 9.nov.2011"	-	AA253	MH213652	MH213625	-	-	-
Lanthanomelissa clementis	"Brasil, RS, ~19km NW Pinheiro Machado\31º28'13"S 53º27'25"W 300m\05.xi.2016"	DZUB3459	TR023	-	-	MG894384	-	-
Lanthanomelissa clementis	"Brasil, RS, 38km NW São Lourenço do Sul\ 31°14'57"S 52°18'30"W, 250m\05.xi.2016	DZUB3460	TR024	MH213664	-	MG894385	-	-
Lanthanomelissa clementis	"Brasil, RS, ca. 10km E Santana do Livramento\30º52'46"S 55º26'53"W\ 07.xi.2016"	DZUB3457	TR021	MH213662	MH213634	MG894382	-	-

Lanthanomelissa clementis	“Piraquara/PR, 6.xi.2016, -25.4604 -49.0843”	DZUB3491	TR036	MH213666	MH213636	-	-	-
Lanthanomelissa clementis	“Porto Alegre, RS, Fundação Zoobotânica, nov.2015”	DZUB3492	TR037	-	MH213637	-	-	-
Lanthanomelissa clementis	Brasil, RS, 10.5km SW Pinheiro Machado, linha de trem (RS-265), 31°38'12"S 53°27'13"W, 300 m, 2.xi.2012”	-	TR013	MH213661	MH213633	MG894380	-	-
Lanthanomelissa clementis	Brazil, Rio Grande do Sul, Canguçu	-	AA143	MH213660	MH213632	-	KX064535	-
Lanthanomelissa clementis	“Gramado, RS, pt 1, -29.4007, -50.8812, 4.xi.2016, “	DZUB3458	TR022	MH213663	-	MG894383	-	-
Lanthanomelissa clementis	“Gramado, RS, pt2, 4/11/16, , -29.360437 -50.839575”	DZUB3478	TR034	MH213665	MH213635	MG894389	-	-
Lanthanomelissa discrepans	“Brasil, RS, ~40km S Caçapava do Sul (Guaritas)\30°45′33″S 53°31′31″W 320m pt2\01.xi.2017”	DZUB3550	TR065	MH213670	MH213642	-	-	MH213613
Lanthanomelissa discrepans	“Brasil, RS, Caçapava do Sul (Minas de Camaquã)\30°53′6″S 53°28′18″W 220m, pt.7\01.xi.2017”	DZUB3555	TR064	MH213669	MH213641	MH213619	-	-
Lanthanomelissa discrepans	“Caçapava do Sul, RS\02.11.2017\ -30.761; -53.524”	DZUB3529	TR059	MH213668	MH213640	MH213618	MG888656	MG889476
Lanthanomelissa discrepans	“Brasil, RS, Caçapava do Sul (Minas de Camaquã) \ 30°53′7″S 53°28′18″W 250m\ 02.xi.2017”	DZUB3545	TR057	MH213667	MH213639	-	MG888654	MG889474
Lanthanomelissa magaliae	“Brasil, RS, ~40km S Caçapava do Sul (Guaritas)\30°45′33″S 53°31′31″W 320m pt2\01.xi.2017”	DZUB3548	TR067	MH213673	-	-	-	MH213615
Lanthanomelissa magaliae	“Brasil, RS, 3km NE Pedras Altas\31°43'14”S 53°33'40”W 400m\06.xi.2016.”	DZUB3463	TR027	-	MH213643	MG894386	-	-
Lanthanomelissa magaliae	“Brasil, RS, ca. 10km E Santana do Livramento\30º52'46”S 55º26'53”W\ 07.xi.2016	DZUB3464	TR028	-	MH213644	-	-	-

Lanthanomelissa magaliae	“Maracajá, SC; 31/10/2017; -28.841985\ -49.436212\ “	DZUB3509	TR058	MH213671	MH213645	MH213620	MG888655	MG889475
Lanthanomelissa magaliae	“Maracajá, SC; 31/10/2017; -28.841985\ -49.436212\ “	DZUB3513	TR066	MH213672	-	-	-	MH213614
Lanthanomelissa pampicola	“Brasil, RS, ca.30km S Rosário do Sul\30°26'08”S 55°03'01”W 120m\ 07-08.xi.2016”	DZUB3465	TR029	MH213674	MH213646	-	-	-
Lanthanomelissa pampicola	“Brasil, RS, Caçapava do Sul (Minas de Camaquã)\30°53′6″S 53°28′18″W 220m, pt.7\01.xi.2017”	DZUB3554	TR069	MH213676	MH213649	-	-	-
Lanthanomelissa pampicola	“Brasil, RS, Santana do Livramento\30º51'16”S 55º30'41”W 190m\07.xi.2016”	DZUB3466	TR030	MH213675	MH213647	MG894387	-	-
B. Outgroup								
Arhysoceble huberi	Brazil, Ceará, Crateus	DZUB089489	AA169	KX064648	MH213621	-	-	-
Arhysoceble huberi	Brazil, Goias, Flores de Goias	DZUB089479	AA159	-	-	-	KX064542	-
Arhysoceble melampoda	Argentina, Jujuy	DZUB089487	AA167	KX064649	MH213622	-	KX064544	-
Chalepogenus goeldianus	“Caçapava do Sul, RS\03.11.2017\ -30.761\-53.524”	DZUB3559	TR060	MH213650	-	-	MG888657	MG889477
Chalepogenus goeldianus	Brazil, Rio Grande do Sul, Canguçu	DZUB000924	AA002	KX064671	-	-	KX064518	KX064561
Chalepogenus parvus	“ARG, Cordoba”	DZUB089562	TR061	MH213651	MH213624	-	-	MH213610
Chalepogenus parvus	“ARG, Cordoba”	DZUB089519	AA202	-	-	-	-	KX064580
Chalepogenus parvus	“ARG, Cordoba”		AA106	KX064661	-	-	KX064531	
Trigonopedia sp.	Brazil, Paraná, Piraquara	DZUB089530	AA217	KX064636	-	-	KX064557	KX064583

2.3 *Alignment and Phylogenetic analysis*

We aligned sequences for each gene using MAFFT extension (Katoh & Standley, 2013) on Geneious 8.1.9, and made minor adjustments by eye. We used the following parameters: 200PAM/k=2 for the nucleotide scoring matrix; 1.53 for gap opening penalty and 0.123 offset value. For CO1 and CytB we used the algorithm G-INS-i, recommended for sequences with global homology; for the protein-coding nuclear Elongation-Factor 1-α and LW-Rhodopsin we used E-INS-i, recommended for sequences with multiple conserved domains and long gaps; and for 16S we used Q-INS-i, recommended for sequences with secondary structure. We also concatenated the alignments using a Geneious 8.1.9 tool. Despite the presence of missing data for some genes we did not considered as a problem since missing data is not a limitation for phylogenetic trees and can even increase the accuracy in some cases (Wiens, 2003; Jiang et al., 2014).

We constructed trees for each gene and, in the absence of strongly supported incongruences, we concatenated the matrix. We performed Maximum likelihood tree searches and bootstrapping in RaxML vs. RAxML 7.4.2 (Stamatakis, 2006) using the graphical user interface raxmlGUI 1.3 (Silvestro & Michalak, 2012). For the support we used the rapid bootstrap with 1,000 non-parametric bootstrap replicates (Felsenstein, 1985).

For choosing the best evolutionary model and dataset partitions we used PartitionFinder 2.1.1 (Lanfear et al., 2017) with all available models selected by Corrected Akaike Information Criterion (AICc) or Bayesian Information Criterion (BIC) using the greedy algorithm (Lanfear et al., 2012). For each model selection criterion we tested two partition strategies as input data blocks for PartitionFinder: a restricted and a comprehensive one. In the restricted scheme we divided the data matrix by gene and by codon for the mitochondrial markers and for nuclear markers only by intron and exon. In the comprehensive scheme we have also divided the exons by codon. We have also tested partitions for a matrix without introns, with data blocks divided by codon. All schemes and models are available in Table A 4.

We conducted Bayesian tree searches in MrBayes (Ronquist et al., 2012) with 10 million generations and four chains sampled every 1000 generations. We discarded the first 25% of trees as burnin. We used the partition scheme yielded by the comprehensive strategy with models selected by AICc (see above). We analysed the Convergence of Markov Monte Carlo chains (MCMC) in Tracer (Rambaut et al., 2018). We edited all trees in FigTree 1.4.3 (Rambaut, 2016).

2.4 *Divergence times estimates*

To estimate the divergence times of *Lanthanomelissa* species and phylogenetic related genera, we relied on the same matrix composed by five markers and 37 taxa. We performed Bayesian estimates in BEAST version 1.8.4 (Drummond et al., 2012) via the CIPRES server using both the strict clock model and the uncorrelated lognormal relaxed model. We observed the coefficient of variation for the relaxed clock model in Tracer to infer the rate heterogeneity among branches and the clock-likeness of our data (CV approaching zero indicates the data are clock-like). We used the Yule tree speciation model, which is more appropriate when considering sequences from different species (Heled & Drummond, 2012; Drummond et al., 2015) and GTR+I+G substation model with empirical base frequencies. We ran MCMC chains for 20 million generations sampled every 10000 steps and assessed convergence of chains in Tracer (when Effective sample size – ESS - for all parameters was >200). We produced the maximum clade credibility tree in TreeAnotator vs. 1.8.4 (part of Beast package), with a burnin of the first 25% trees. We visualized and edited the trees in FigTree 1.4.3.

To calibrate the tree, we used a secondary calibration approach, since there is no known fossil for Tapinotaspidini and related groups. We calibrated the tree on the crown node of the clade formed by *Lanthanomelissa*, *Arhysoceble* and *Chalepogenus* using the age estimated by Aguiar et al. (in prep.): 33.22 (95% HPD 26,82-40,7). We applied a normal distribution prior (mean: 33, stdev: 4) on the node representing the three genera in our sampling (Figure A 9). Beast analysis could not recover two clades, which we then enforced to be monophyletic since they were well supported in the Bayesian phylogenetic analysis: 1. *Chalepogenus parvus* + *Lanthanomelissa* and 2. *Arhysoceble* + *C. parvus* + *C. goeldianus* and *Lanthanomelissa* (see Figure 2).

2.5 *Ancestral Area Reconstruction*

We derived biogeographic areas from the provinces defined in Morrone (2014) according to the occurrence points for each species. We coded the geographical range of *Lanthanomelissa* and outgroup as A: Atlantic province, B: Cerrado province, C: Chacoan

province, D: Caatinga province, E: Pampean province, F: *Araucaria* Forest province. All these provinces are in the Chacoan subregion. Cerrado, Chacoan, Pampean and Caatinga provinces are in Chacoan dominion, while *Araucaria* Forest and Atlantic provinces are in Parana dominion. For outgroup lineages we coded the areas representative for all species of the lineage. We coded the lineage *Chalepogenus goeldianus* to represent the distribution of all its three species, *C. goeldianus*, *C. neffi*, and *C. luciane*. The same was applied to *Arhysoceble* and *Trigonopedia*. Despite the maps from Morrone (2014) suggest the Atlantic province covering all the costal Santa Catarina and highlands of São Paulo, we considered the records on Cotia (São Paulo) and Maracajá (Santa Catarina) as patches of *Araucaria* forest and Pampa on the mosaic in the Atlantic Forest. We did the same for records in Parana province. For discussion purposes we also consider the regionalization introduced by Olson et al. (2001) as modified by Antonelli et al. (2018).

For ancestral area estimation we relied on the R package 'BioGeoBEARS' (Matzke, 2013), which evaluates several biogeographic models: models DEC (Ree & Smith, 2008), a modified version of DIVA (Ronquist, 1997), named DIVA-like, and a modified version of BayArea (Landis et al., 2013) or BayArea-like. The contribution of evolutionary processes (i.e., range expansion, range extinctions, vicariance, founder-event speciation, within-area speciation, founder-speciation event) is evaluated. To assess the fitness of the models we conducted likelihood ratio tests based on AICc scores.

2.6 *Occurrence data sampling*

We obtained 196 georreferenced points (Table A 1) for all the five *Lanthanomelissa* species from literature and voucher labels deposited in Departamento de Zoologia da Universidade de Brasília (DZUB), Departamento de Zoologia da Universidade do Paraná (DZUP), Museu de Ciências e Tecnologia da PUCRS (MCTP), Fundação Zoobotânica do Rio Grande do Sul (FZB/RS), American Museum of Natural History (AMNH), Coleção Sersic-Cocucci, Departamento de Botânica da Universidade Nacional de Cordoba (UNC), Museu Argentino de La Plata – Universidade Nacional de La Plata (UNLP), Universidade do Extremo Sul Catarinense (UNESC), Faculdade de Ciências e Letras de Ribeirão Preto USP - Coleção Camargo (RPSP), Museu de Zoologia da Universidade de São Paulo (MZSP). When coordinates were not available, we extracted them from the localities or cities provided on labels by searching them on Google Maps. We have also obtained data from the online

databases SpeciesLink (http://www.splink.org.br) and GBIF (Global Biodiversity Information Facility, https://www.gbif.org/).

2.7 *Environmental data*

To estimate potential geographic range of *Lanthanomelissa* species across a glacial-interglacial cycle we performed species distribution models based on current (1970 – 2000), Last Glacial Maximum (LGM, 22 ky) and Last Interglacial (LIG, 120 ky, Otto-Bliesner et al., 2006) bioclimatic variables from WorldClim (http://www.worldclim.org/) in 2.5 arcminutes resolution. For the LGM and current climatic scenarios we performed the simulations based on the Community Climate System Models (CCSM) general circulation model. We resampled map resolution for LIG variables from 30 arcseconds to 2.5 arcminutes using the 'raster' package (Hijmans, 2017) in R 3.4.2 (R Core Team, 2017), implemented in RStudio 1.0.153 (RStudio Team, 2016). In this resolution we have compiled 73 unique points, being 52 for *Lanthanomelissa betinae*, 48 for *L. clementis,* 38 for *L. discrepans,* 21 for *L. magaliae* and 14 for *L. pampicola*

2.8 *Species distribution modelling*

To avoid collinearity and model overfitting (Jiménez-Valverde et al., 2011) we standardized the 19 bioclimatic variables by subtracting the mean value for each cell and then divided this result by the standard deviation, so that all variables vary from -1 to +1 and have average equal to zero and variance equal to one. Then we ran a Principal Component Analysis (PCA) creating new orthogonal spatialized principal components (PCs) to remove the collinearity of the variables. We performed this analysis first for current climate and then projected its linear coefficients into the past climates (LGM and LIG) so that the PCs generated for the past were dependent to the current scenario.

Maximum and minimum latitudes and longitudes from the occurrence points were parameters for the extent area modelled (longitude: minimum -70 and maximum -45; latitude: minimum -40 and maximum -20). Since the models consider only climate data, they do not acknowledge biotic factors or species dispersal ability (Soberon & Peterson, 2005). In this context, an extent that is too wide and not congruent with the occurrence data would insert too much noise.

We performed all modelling analysis in R, using a script for Ecological Niche Modelling: the ENMTheMetaLand (Andrade & Velazco, https://github.com/andrefaa/ENM_TheMetaLand/tree/master), which uses the following packages: 'raster', 'sp' (Pebesma & Bivand, 2005), 'dismo' (Hijmans et al., 2017), 'kernlab' (Karatzoglou et al., 2004), 'xlsx' (Dragulescu, 2014), 'randomForest' (Liaw & Wiener, 2002), 'mda' (Hastie et al., 2017), 'rgdal' (Bivand et al., 2017), 'dummies' (Brown, 2012), 'MASS' (Venables & Ripley, 2002), 'ade4' (Dray & Dufour, 2007), 'gam' (Hastie, 2018), 'mvtnorm' (Genz & Bretz, 2009), 'progress' (Csardi & FitzJohn, 2016), 'maxnet' (Phillips, 2017), 'maptools' (Bivand & Lewin-Koh, 2017), 'XML' (Lang & the CRAN Team, 2018), 'maxlike' (Royle et al., 2012), 'mgcv' (Wood et al., 2016), 'plyr' (Wickham, 2011), 'GRaF' (Golding, 2014), 'RStoolbox' (Leutner et al., 2018), 'stringi' (Gagolewski, 2018), 'flexclust' (Scharl & Leisch, 2006), 'ape' (Paradis et al., 2004), 'modEvA' (Barbosa et al., 2016) and 'SDMTools' (VanDerWal et al., 2014).

We partitioned the dataset of occurrences of *Lanthanomelissa* species with checkerboard partition, where data is divided in two subsets (50% each). The first subset produces the potential distribution and the second evaluates it and vice-versa. Then, we performed an ensemble through a consensus from algorithms that had TSS values above mean with three algorithms: the machine learning algorithms Maxent (Maximum Entropy, Phillips et al., 2004) and Random Forest (Breiman, 2001) and the statistical algorithm GLM (Generalized Linear Models, McCullagh and Nelder, 1989). We evaluated model performance based on True Skill Statistics values (TSS, Allouche et al., 2006), a threshold-dependent statistics varying between -1 to +1 in which values above 0.5 are acceptable and values above 0.7 are considered good. We have also evaluated the AUC (Area Under the Curve, Allouche et al., 2006) values, considered good when above 0.8.

We also performed a stable area analysis in R using the 'raster' package. For that, for each species, we multiplied the binary rasters yielded by the ensemble with all models from the three scenarios (LIG, LGM and current) to find the intersection between them and plot them together.

3 RESULTS

3.1 *Molecular Phylogeny*

The aligned data matrix comprised 109 sequences and 3430 nucleotides, in which 88 sequences belong to *Lanthanomelissa* specimens plus 21 from the outgroup. We have obtained 32 sequences from cytochrome oxidase subunit I (COI), 28 from cytochrome B (CytB) and 16 from the ribosomal gene 16S. For nuclear markers, 10 were from elongation factor 1-α (EF-1α) and 14 from Long Wave Rhodopsin (LW-Rhodopsin). We have generated individual gene trees and a concatenated tree with all five genes (Figure A 1 to Figure A 5).

Bayesian and Maximum Likelihood inference trees recovered the same highly supported clades (Figure 2, Figure 3). The trees show *Arhysoceble* as sister to a clade containing *Chalepogenus goeldianus, Chalepogenus parvus* and all *Lanthanomelissa* species (1 Bayesian Posterior Probability (BPP)). All relations within the outgroup are well supported (100% Bootstrap Support (BS), 1 BPP). In Bayesian tree *C. goeldianus* is sister to *C. parvus* and *Lanthanomelissa* (1 BPP) and *C. parvus* is sister to *Lanthanomelissa* (0.53 BPP). However, in Maximum Likelihood tree these relations are inverted as *C. parvus* is sister to *C. goeldianus* and *Lanthanomelissa* (97% BS) while *C. goeldianus* is sister to *Lanthanomelissa* (55% BS). *Lanthanomelissa* constitute a monophyletic group (81% BS, 0.99 BPP) and all the species are well supported as monophyletic (BS > 88%, BPP > 0.95). In this clade, *L. betinae* is sister to all the other species (81% BS, 0.99 BPP). In both trees *L. discrepans* and *L. magaliae* are monophyletic (73% BS, 0.97 BPP). These two species constitute a clade sister to another containing *L. pampicola* and *L. clementis* (87% BS, 1 BPP), however in Maximum Likelihood tree these two species are weakly supported as sister groups (54% BS) while in Bayesian tree this relation is highly supported (0.97 BPP).

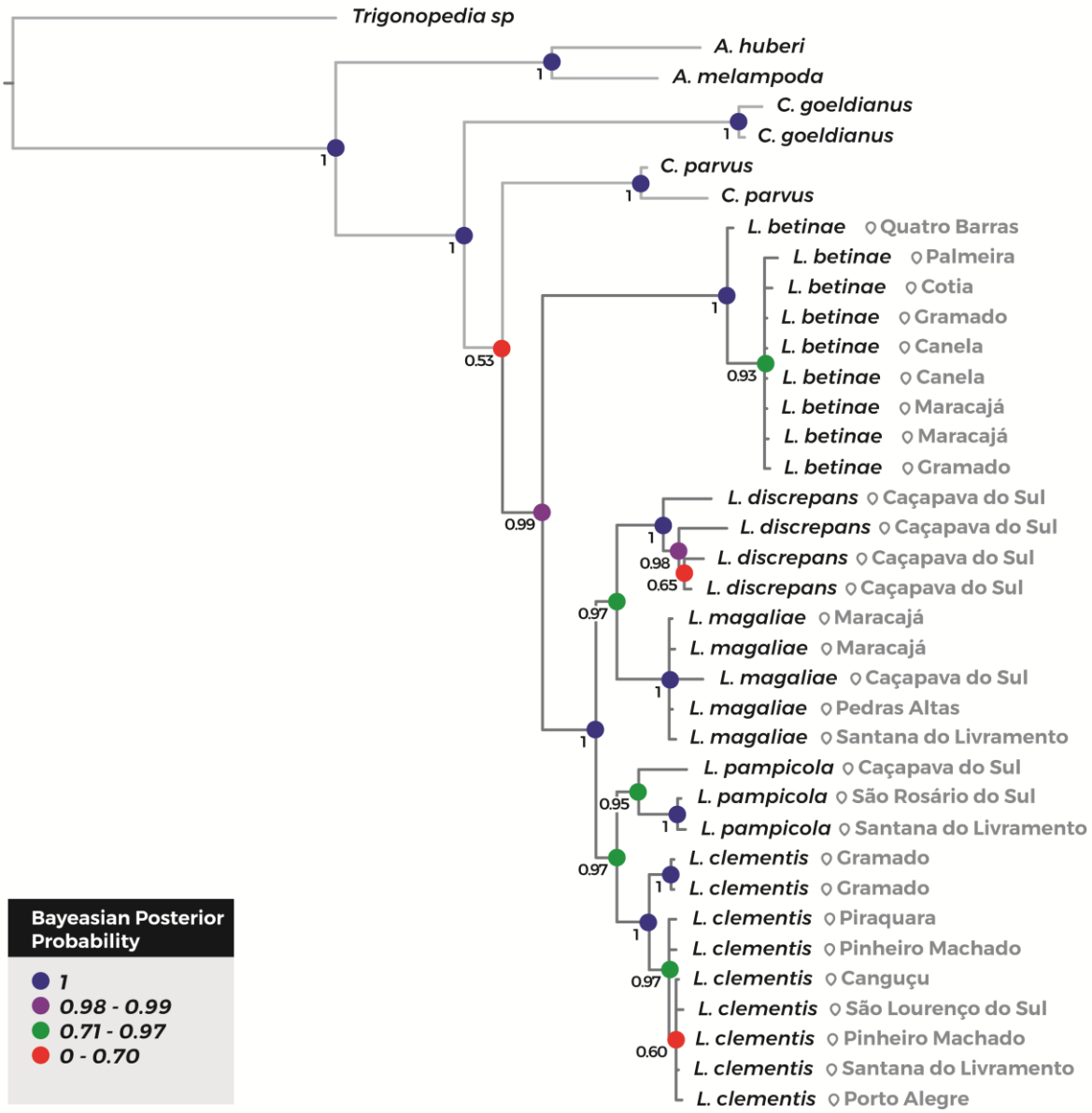

Figure 2. Bayesian consensus tree of *Lanthanomelissa* with representatives of *Arhysoceble*, *Chalepogenus* and *Trigonopedia* as outgroup based on a concatenated matrix comprising 37 terminals and 3430 nucleotides obtained from mitochondrial CO1 and CytB, ribosomal 16S and nuclear EF1-α and LW-Rhodopsin using 14 schemes selected by PartitionFinder using the corrected Akaike Information Criterion. Node circles indicate Bayesian posterior probabilities (BPP), in which blue circles indicate highly supported nodes (1 BPP), purple indicate good support (0.98 – 0.99 BPP), green indicate accepted support (0.71 – 0.97) and red indicate low supported nodes (BPP ≤ 0.70).

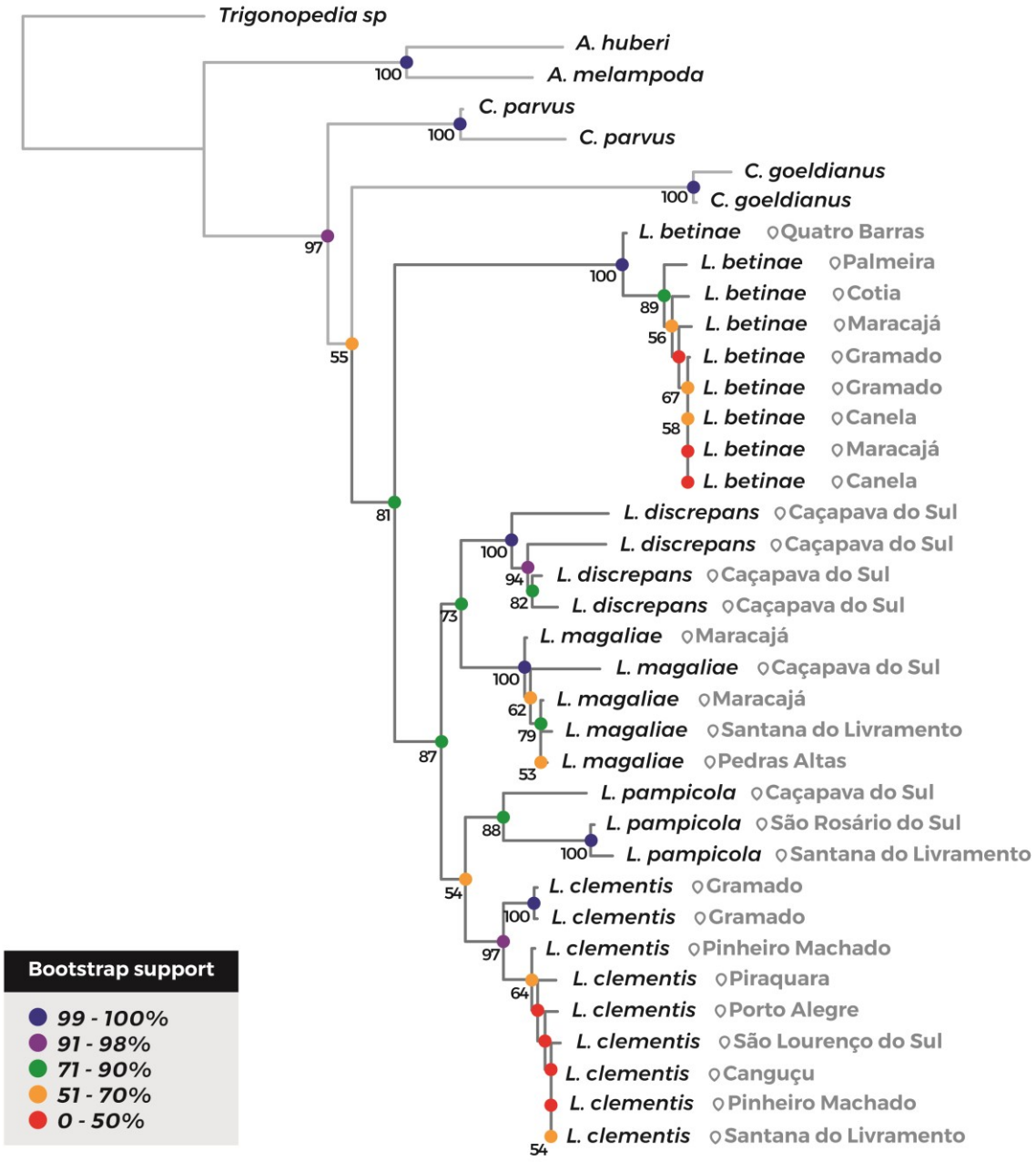

Figure 3. Best scoring Maximum Likelihood tree of *Lanthanomelissa* with representatives of *Arhysoceble*, *Chalepogenus* and *Trigonopedia* as outgroup based on a concatenated matrix comprising 37 terminals and 3430 nucleotides from five genes: mitochondrial CO1 and CytB, ribosomal 16S and protein coding nuclear EF1-α and LW-Rhodopsin. Node circles indicate bootstrap support, in which blue indicate highly supported nodes (99 – 100%), purple indicate high support (91 – 98%), green indicate good support (71 – 90%), orange indicate acceptable support (51 – 70%) and red diamonds indicate low supported nodes (≤ 50%).

3.2 *Divergence times estimates*

The coefficient of variation for the relaxed clock model was closer to 1 indicating that the model is non-clock like; therefore the relaxed uncorrelated molecular clock was preferred. According to the Bayesian time tree (Figure 4), the stem group *Arhysoceble* + *Chalepogenus* + *Lanthanomelissa* had its origin in the Eocene, at 36.87 (25.28 - 52.93, 95% HPD) Mya while the crown group age is estimated as 32.03 (24.64 - 39.73) Mya, in Oligocene. The crown group of *Chalepogenus goeldianus* + *Chalepogenus parvus* + *Lanthanomelissa* has its age estimated at 24.40 (16.33 - 32.34) Mya. Crown age for *Chalepogenus parvus* + *Lanthanomelissa* is estimated at 21.39 (14.29 – 29.9) Mya in Miocene. Crown age for *Lanthanomelissa* is estimated at 17.49 (11.64 – 25.17) Mya. *Lanthanomelissa betinae* has its age estimated at 3.62 (1.32 – 7.93) Mya. Crown age for all the other *Lanthanomelissa* is estimated at 13.27 (8.57 – 19.9) Mya. The crown group *L. pampicola* + *L. clementis* has its age estimated at 10.34 (6.14 – 16) Mya while crown age for *L. pampicola* is 6.84 (3.42 – 11.49) Mya and for *L. clementis* is 5.71 (2.46 – 10.28) Mya. Crown age for the clade *L. discrepans* + *L. magaliae* is 10.39 (6.07 – 16.01) Mya while *L. discrepans* has its age estimated at 6.12 (3.03 – 10.07) Mya and *L. magaliae* has origin estimated at 3.65 (1.1 – 7.62) Mya, in Pliocene.

3.3 *Ancestral area reconstruction*

Multi-model analysis in BioGeoBEARS yielded BayArealike as the best fitting model for our data, as it showed the higher log likelihood and lowest AICc (LnL = -29.95, AICc = 64.24, Table 2). According to this analysis (Figure 4), the ancestral area for the most recent common ancestor of the lineage *Arhysoceble* + *C. goeldianus* + *C. parvus* comprised Chacoan, Pampean and *Araucaria* Forest provinces. The latter province is not inferred for the most recent common ancestor of the two *Chalepogenus* + *Lanthanomelissa*, which are Chacoan + Pampean provinces. For *Lanthanomelissa* the ancestral area inferred is Pampean and this is the current area for most species. However, *L. betinae* and *L. clementis* are also distributed in the *Araucaria* Forest province.

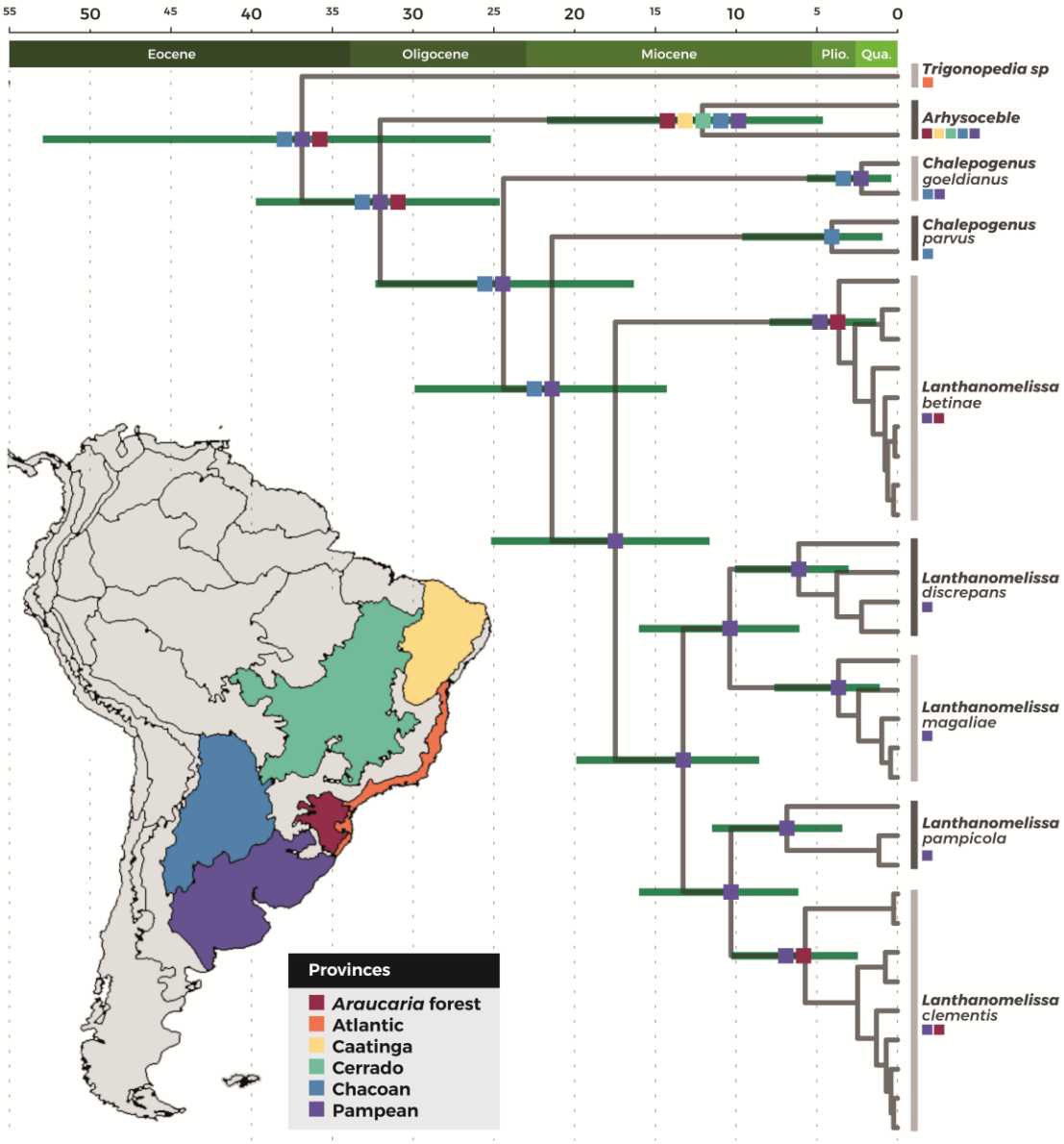

Figure 4. Time calibrated phylogeny and historical biogeography of *Lanthanomelissa* with representatives of *Arhysoceble, Chalepogenus* and *Trigonopedia* as outgroup based on a concatenated matrix comprising 37 terminals and 3430 nucleotides with 20 million generations, age of the root node sampled from a normal distribution (mean =33 and s.d. = 4). Map shows provinces from Morrone (2014) used for biogeographic reconstruction. Coloured squares after species names represent current and at nodes historical distributions inferred in BioGeoBEARS under the Bayarealike model (lnL = -29.95). Horizontal green bars at nodes indicate 95% HPD of estimated divergence times and top bar indicates epochs and Quaternary period. Abbreviations: Plio: Pliocene; Qua: Quaternary.

Table 2 Statistics obtained by BioGeoBEARS for each model tested, including log likelihood (LnL), parameters considered (d: dispersion, e: extinction, j: founder-event speciation), corrected Akaike Information Criterion (AICc), Akaike Information Criterion (AIC) and weighted AICc and AIC.

		Parameters				weighted		weighted
Model	**LnL**	**D**	**E**	**J**	**AICc**	**AICc**	**AIC**	**AIC**
DEC	-111.3	0.02	0.0079	0	227	4.60E-36	226.6	4.60E-36
DEC+J	-111.3	0.02	0.0079	1.00E-05	229.3	1.40E-36	228.6	1.70E-36
DIVALIKE	-121.2	0.021	0.0027	0	246.8	2.20E-40	246.5	2.20E-40
DIVALIKE+J	-121.3	0.021	0.0028	0.0001	246.9	2.20E-40	246.5	2.20E-40
BAYAREALIKE	-29.95	0.003	0.015	0	64.24	1	63.89	1
BAYAREALIKE+J	-55.97	0.0032	0.003	0.0039	118.7	1.50E-12	117.9	1.80E-12

3.4 *Distribution of* Lanthanomelissa

From the three models tested, GLM had the higher values of TSS (True Skill Statistics) and AUC (Area Under the Curve). All values are listed in Table 3 and suitability maps are shown in Figure 5. For all species in LGM suitability extrapolates current continental borders because the sea level was retracted at that time (Salgado-Labouriau et al., 1998).

For *Lanthanomelissa betinae* the suitable area in LIG (Figure 5a) was narrow, including most of what is now the Brazilian states of Rio Grande do Sul and Santa Catarina. In LGM (Figure 5b) this distribution was expanded northwards to Paraguay and the northeast of Argentina and the Brazilian states of São Paulo and Mato Grosso do Sul. However, the higher suitability was in the coast. The current suitable area (Figure 5c) is expanded southwards, covering the entire Brazilian South region and Uruguay. For this species the stable area (Figure 6a) predicted covers most of Rio Grande do Sul, excluding the Pampa region, west Santa Catarina and a narrow area in the coast of São Paulo.

For *Lanthanomelissa clementis* the suitable area in LIG (Figure 5d) covered what is now the Brazilian southern region, including part of São Paulo and Uruguay. In LGM (Figure 5e) the suitable area was wider to the north and west; still it does not cover Uruguay and the south of Rio Grande do Sul. Current suitable area is narrower, mostly in Brazilian southern region but excluding the north of Paraná and the west of Rio Grande do Sul and a gap in south-eastern Rio Grande do Sul, in a lower altitude area (Figure 5f). Suitability is also shown in São Paulo near the two occurrence points in the city of Cotia. Stable area (Figure 6b)

covers most of Brazilian Southern Region, excluding north Paraná, and southern Rio Grande do Sul and the gap in the lower area. The area in São Paulo is also stable.

For *Lanthanomelissa discrepans,* the suitability area was narrow in LIG (Figure 5g), covering mostly eastern Argentina, southern Uruguay, and eastern Brazilian south region. For LGM (Figure 5h), the suitability was higher in the coast and the suitable area included part of eastern Argentina and Paraguay, the greatest part of Uruguay and of the states of Rio Grande do Sul and Santa Catarina, also covering eastern Paraná and São Paulo. From all *Lanthanomelissa* species, *L. discrepans* shows the widest suitable areas for the present (Figure 5i), covering an area similar to the one predicted for LGM but including the whole Uruguay and a greater portion of Rio Grande do Sul. A wide stable area is shown in Argentinian and Uruguayan pampa (Figure 6c).

The suitable area for *Lanthanomelissa magaliae* in LIG (Figure 5j) covered a small part of Argentina and Paraguay and almost the whole Brazilian Southern region and Uruguay, except for a gap in the north eastern part of Rio Grande do Sul that is maintained in all three scenarios. In LGM (Figure 5k), the gap is wider, and the distribution is expanded both north and southwards. In current distribution (Figure 5l), the gap is mostly the same, but the area is much narrower, covering mostly the western part of the states. Stable area for this species (Figure 6d) is large, covering part of Paraguay, Argentina and most of Uruguay and almost the whole Brazilian Southern region, excluding the previously mentioned gap in Rio Grande do Sul.

This gap also appears for *L. pampicola,* but it is less expressive. In LIG (Figure 5m) the suitable area reaches east Argentina, although the suitability is higher in south Rio Grande do Sul. In LGM (Figure 5n), suitability area is shifted northwards, covering Paraguay, west Paraná and Santa Catarina and the whole Rio Grande do Sul, no longer reaching Argentina. For current distribution (Figure 5o), suitable area shrinks and concentrates with higher suitability in Rio Grande do Sul and covers part of Santa Catarina, Paraná and Uruguay. Stable area for *L. pampicola* (Figure 6e) is very small, covering only patches in Rio Grande do Sul.

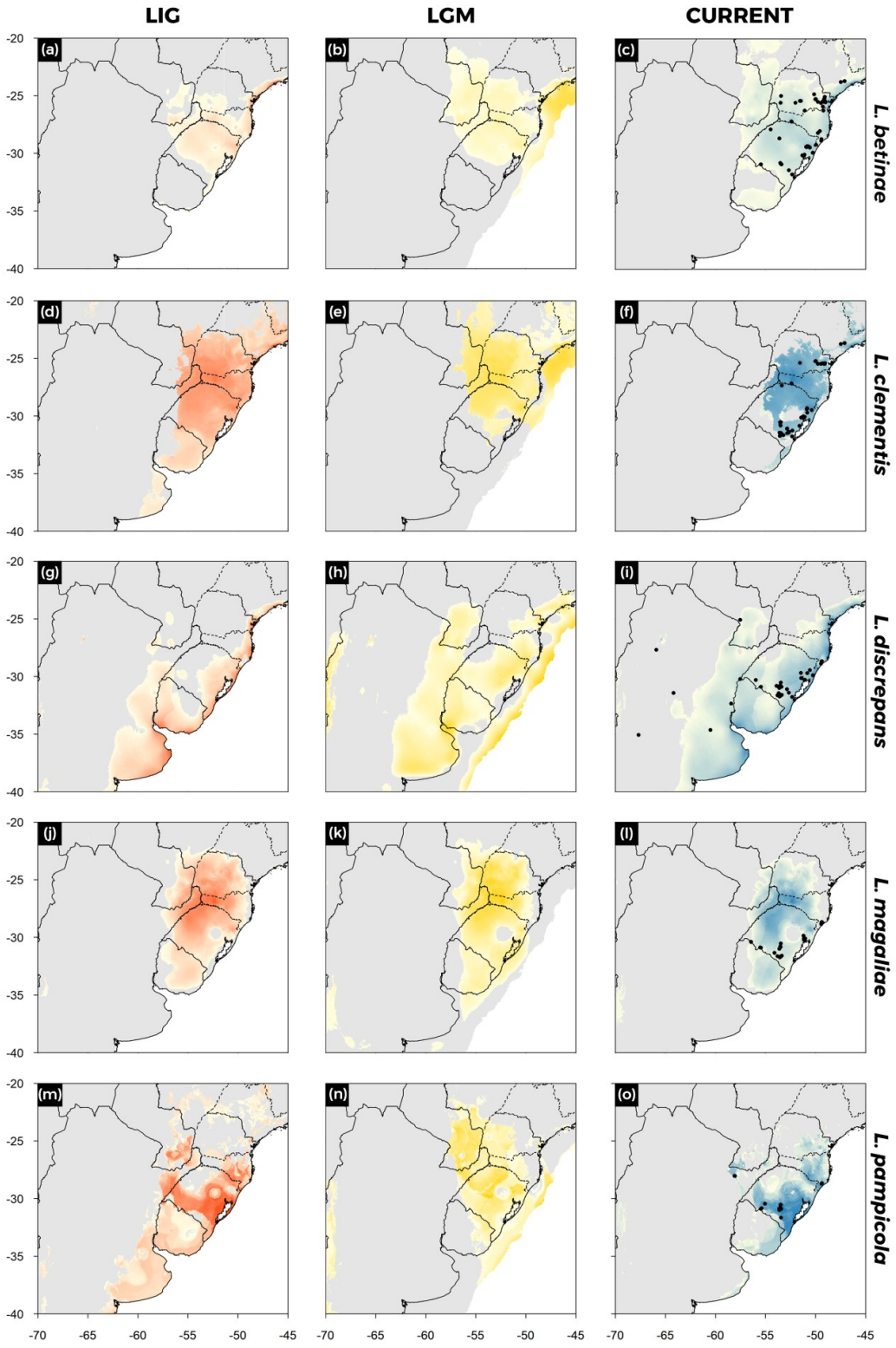

Figure 5. Suitability maps for *Lanthanomelissa betinae* (a, b, c), *L. clementis* (d, e, f), *L. discrepans* (g, h, i), *L. magaliae* (j, k, l) and *L. pampicola* (m, n, o) for Last Interglacial (LIG, 120 ky) (a, d, g, j, m), Last Glacial Maximum (LGM, 22 ky) (b, e, h, k, n) and current climate (c, f, i, l, o) using ensemble of the algorithms Maxent, Random Forest and GLM. Deeper colours indicate areas of higher suitability.

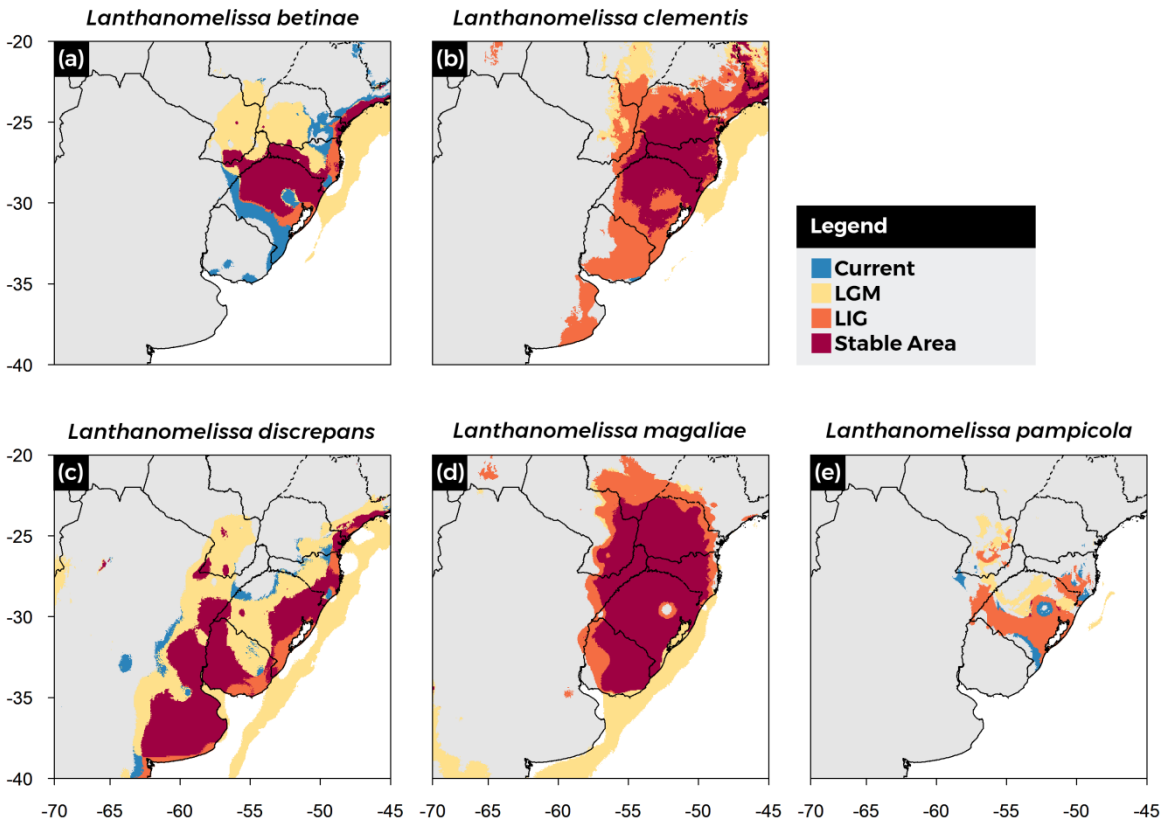

Figure 6. Potential range shifts for *Lanthanomelissa betinae* (a), *L. clementis* (b), *L. discrepans* (c), *L. magaliae* (d) and *L. pampicola* (e). Stable areas are shown in purple, areas predicted as suitable in the Last Interglacial (LIG) are shown in orange, areas predicted in the Last Glacial Maximum (LGM) are in yellow and areas in current distribution are shown in blue.

Table 3. Statistic values of TSS (True Skilled Statistics) and AUC (Area Under the Curve) for all models tested (GLM, Maxent, Random Forest) and ensemble

Species	Algorithm	TSS	AUC
Lanthanomelissa betinae	GLM	0.96	0.98
Lanthanomelissa betinae	Maxent	0.80	0.92
Lanthanomelissa betinae	Random Forests	0.87	0.97
Lanthanomelissa betinae	Ensemble	0.96	0.97
Lanthanomelissa clementis	GLM	0.92	0.96
Lanthanomelissa clementis	Maxent	0.79	0.91
Lanthanomelissa clementis	Random Forests	0.94	0.98
Lanthanomelissa clementis	Ensemble	0.96	0.93
Lanthanomelissa discrepans	GLM	0.81	0.94
Lanthanomelissa discrepans	Maxent	0.62	0.80
Lanthanomelissa discrepans	Random Forests	0.63	0.88
Lanthanomelissa discrepans	Ensemble	0.81	0.94
Lanthanomelissa magaliae	GLM	0.96	0.99
Lanthanomelissa magaliae	Maxent	0.75	0.94
Lanthanomelissa magaliae	Random Forests	0.86	0.95
Lanthanomelissa magaliae	Ensemble	0.96	1.00
Lanthanomelissa pampicola	GLM	0.68	0.78
Lanthanomelissa pampicola	Maxent	0.65	0.83
Lanthanomelissa pampicola	Random Forests	0.84	0.95
Lanthanomelissa pampicola	Ensemble	0.84	0.92

4 DISCUSSION

4.1 *Origin and Diversification of* Lanthanomelissa

The origin of *Lanthanomelissa* was estimated at the transition from Oligocene to Miocene (21.39, 14.29 – 29.9 Mya (Figure 4)). The Andean uplift could have driven this origin, since it was a major driver of speciation in South America, having several discrete tectonic events that broadly affected the region at different time periods (Antonelli et al., 2009).

In Oligocene, there was a relative quiescence in tectonic events (Ortiz-Jaureguizar & Cladera, 2006). However, convergence between Nazca and South American tectonic plates reactivated the main magmatic belt, especially in central and occidental Argentina, Bolivia and Peru (Ortiz-Jaureguizar & Cladera, 2006). During late Oligocene to Miocene (20 – 26

Mya), Andean morphological structure in Chile started to develop and the ocean levels increased, allowing the formation of the Patagonian Sea, an Atlantic transgression that extended from southern Patagonia to Bolivia. This sea could have been the first event of divergence affecting the common ancestral of the lineage *C. parvus* + *Lanthanomelissa* in western and eastern lineages on the ancestral area of southern South America. The ancestral area of this lineage represents a shared area of what are now Pampean and Chacoan provinces that could have been fragmented by the Patagonian Sea. Vegetation at this time was very hybrid and patches of grasslands and open vegetation areas in which the ancestral could inhabit were present (Ortiz-Jaureguizar & Cladera, 2006). The origin of the lineage *C. goeldianus* + *C. parvus* + *Lanthanomelissa* in a shared area between the Chaco and Pampa suggests that these two areas could have arisen from the same ancestral area.

The diversification of *Lanthanomelissa* (at 17.48, 11.64 – 25.17 Mya) could be related to the Middle Miocene Climatic Optimum at 15 – 17 Mya, which was the peak of a global warming phase that started in late Oligocene (26 – 27 Mya) (Zachos et al., 2001). By mid Miocene, tectonic activity that led to the ending of Central Andean orogeny became more intense impacting the climate and biodiversity (Ortiz-Jaureguizar & Cladera, 2006). During this time, temperature and precipitation were higher and marine incursions led to the formation of the Paranean Sea at ca. 14 Mya, which was bordered by the Andes at west and by the Brazilian shield at east, which comprises the region of what are now Uruguay and the South, Southern, Central-West and North-Eastern Brazilian regions (Räsänen et al., 1995).

The ancestral area of the lineage, including all *Lanthanomelissa* species, except *L. betinae*, could have been fragmented by Paranean Sea and established in Middle to Upper Miocene during the Quechua phase of Andean orogeny, when this sea retracted and in its place widespread plains were formed, in the so called “The Age of the Southern Plains” (Donato et al., 2003; Ortiz-Jaureguizar & Cladera, 2006).

In the Quechua phase, Patagonian Andean Cordillera was elevated creating a rain-shadow effect to humid winds from east and from west increasing the aridity on the ancestral area (Ortiz-Jaureguizar & Cladera, 2006), which could have induced diversification of the clades *L. pampicola* + *L. clementis* and *L. discrepans* + *L. magaliae,* both at ca 10 (6 – 16) Mya. The higher aridity caused by the rain-shadow effect led to open vegetation habitats predominance in southern South America and the expansion of the Diagonal of Open Formations, with strong presence of C3 grasses and increased presence of C4 grasses (Strömberg, 2011).

In late Miocene (ca. 7 Mya), there was an expansion in C4 grasslands that could be related to low CO_2 pressure, and to increased seasonality and aridity caused by tectonic events (Pagani et al., 1999). This time is mostly congruent with the origin and establishment of the species of *Lanthanomelissa* and to the occupation by *L. betinae* and *L. clementis* of the *Araucaria* Forest province. The origin and diversification of plants endemic to the Cerrado (at ca. 4 – 10 Mya, (Simon et al., 2009)) are related to the worldwide expansion of grasslands. The Cerrado savannah and the Pampa could have gone through similar diversification processes associated to the expansion of grasslands. However, the origin of the Pampa at the transition from Oligocene to Miocene is older than the one suggested for the Cerrado.

Lanthanomelissa is specialist for oil-collecting in flowers from the genus *Sisyrinchium,* Iridaceae (Cocucci & Vogel, 2001), which have its origin in Late Miocene at 8 (6 – 11 95% HPD) Mya (Martins et al., 2015), coincident with the origin of the clades *L. pampicola* + *L. clementis* (10.34, 6.14– 16 Mya) and *L. discrepans* + *L. magaliae* (10.38, 6.07 – 16). This and the higher diversity of *Sisyrinchium* in the same regions that *Lanthanomelissa* inhabits, supports the suggestion that *Lanthanomelissa* and *Sisyrinchium* could have played important roles in each other diversifications (Cocucci & Vogel, 2001; Chauveau et al., 2011). However, further studies covering the biogeography of these plants and its implications for *Lanthanomelissa* distribution are needed to better understand how this could have happened.

4.2 *Quaternary distribution*

As the main events of origin and speciation in *Lanthanomelissa* happened in Miocene, climatic fluctuations in Pleistocene had no influence in these events. However, these fluctuations affected the distribution and possibly the intraspecific diversification of these bees.

The Last Interglacial (LIG, ~120 kya) had the highest global temperatures of the last 250 ky and is characterized by wet conditions, higher sea levels and expansion of forest biomes over arid open formations (Otto-Bliesner et al., 2006). This could have caused the shown restriction on the distribution of most species from *Lanthanomelissa* at this time (Figure 5), suggesting a restriction in the grasslands, which were probably more expressive towards the south.

During the Last Glacial Maximum (LGM, ~21 kya) the sea level was retracted by at least 100 m (Salgado-Labouriau et al., 1998). According to our species' distribution modelling, the suitability for most species shifted towards the north and this east coast at this time. This is

corroborated by the presence of a mosaic of grasslands and *Araucaria* forest in the coast of southern Brazil, from São Paulo to Rio Grande do Sul (Behling, 2002; Behling et al., 2004; Overbeck et al., 2007; Pessenda et al., 2009; Leite et al., 2016; Silva et al., 2018). It is possible that this type of vegetation extended to the coast of Uruguay, as this area is also predicted as suitable for *Lanthanomelissa* in LGM. However, for *L. magaliae* and *L. pampicola* the suitability in LGM is shifted to the west at lower latitudes, covering part of Paraguay. After LGM, at ca 4 kya, as temperatures got higher the *Araucaria* Forest could expand, limiting once again the occurrence of the grasslands (Behling et al., 2004) as well as the distribution of *Lanthanomelissa,* as evident in models for current distribution of *L. clementis.*

The distribution of *L. discrepans* (and to a lesser extent *L. pampicola*) is divided by a central gap suggesting the division of inland western and eastern clusters as well as a costal cluster in LGM. This structure was also observed in this area for plants endemic to the Southern Grasslands (Pinheiro et al., 2011; Longo et al., 2014; Silva et al., 2018). The species could have dispersed through the "Portal de Torres" (Rambo, 1950), a migratory route from the coastal Atlantic forest to the inland grasslands (Pinheiro et al., 2011; Barros et al., 2018; Silva et al., 2018).

Our modelled current distribution predicts as suitable for *Lanthanomelissa* some areas where occurrence was not yet recorded. For instance in *L. magaliae* occurrence data are present mostly in lowland Pampa, while stable area covers the *Araucaria* Forest. This shows the need for better sampling effort in the areas predicted as suitable for a better understanding of the distribution of these bees. Integrative modelling approaches using occurrence data of the oil-host plant *Sisyrinchium* as a variable limiting the distribution of *Lanthanomelissa* and the phylogeography of one of these bee species would also bring valuable insights on the biogeographic history of the south-eastern South America grasslands.

5 CONCLUSIONS

The origin of the lineage *C. parvus* + *Lanthanomelissa* in a shared area between Chaco and Pampa by early Miocene was probably related to the action of marine transgressions that fragmented southern South America, followed by aridification events triggered by the rain-shadow effect caused by Andean uplift. These areas of endemism could have arisen from a common ancestral derived from this shared area that was probably older than Cerrado. The

speciation of *Lanthanomelissa* was probably influenced by the expansion of C4 grasslands over southern South America, including the Pampa and Atlantic forest.

Although Pleistocene climatic fluctuations are suggested as a main driver of speciation in Southern South America, *Lanthanomelissa* as many other groups (Hoorn et al., 2010) had its divergence time in Miocene, supporting the thought that Neogene events had a deeper effect in the cladogenesis and extinction of biodiversity and Quaternary events were better observed at dispersion shifts in community and population levels (Ortiz-Jaureguizar & Cladera, 2006; Werneck et al., 2011).

Future prospects for a deeper understanding of biogeography of *Lanthanomelissa* and origin of the Campos are integrated studies between these bees and its oil-host plants, (*Sisyrinchium*) both in phylogenetic and species distribution modelling approaches. Besides that, more inclusive field work involving a larger sampling area would provide a more comprehensive idea of the current distribution of *Lanthanomelissa.* Another interesting approach would be choosing one species of *Lanthanomelissa* to run phylogeographical analysis in order to understand the process underlying the intraspecific diversification patterns.

REFERENCES

Aguiar, A.J.C. (2012) Available at: http://www.moure.cria.org.br/catalogue.

Allouche, O., Tsoar, A., & Kadmon, R. (2006) Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, **43**, 1223–1232.

Antonelli, A., Nylander, J.A.A., Persson, C., & Sanmartin, I. (2009) Tracing the impact of the Andean uplift on Neotropical plant evolution. *Proceedings of the National Academy of Sciences*, **106**, 9749–9754.

Antonelli, A., Zizka, A., Carvalho, F.A., Scharn, R., Bacon, C.D., Silvestro, D., & Condamine, F.L. (2018) Amazonia is the primary source of Neotropical biodiversity. *Proceedings of the National Academy of Sciences USA*, **115**, 6034–6039.

Arias, M.C., Silvestre, D., Oliveira Francisco, F. De, Weinlich, R., & Sheppard, W.S. (2008) An oligonucleotide primer set for PCR amplification of the complete honey bee mitochondrial genome. *Apidologie*, **39**, 475–480.

Barbosa, A.M., Brown, J.A., Jimenez-Valverde, A., & Real, R. (2016) modEvA: Model Evaluation and Analysis. .

Barros, M.J.F., Diniz-Filho, J.A.F., Freitas, L.B., Barros, M.J.F., Diniz-Filho, J.A.F., & Freitas, L.B. (2018) Ecological drivers of plant genetic diversity at the southern edge of geographical distributions: Forestal vines in a temperate region. *Genetics and Molecular Biology*, **41**, 318–326.

Behling, H. (1997) Late quaternary vegetation, climate and fire history of the *Araucaria* forest and campos region from Serra Campos Gerais, Parana State (South Brazil). *Review of Palaeobotany and Palynology*, **97**, 109–121.

Behling, H. (2002) South and southeast Brazilian grassland during Late Quaternary times: a synthesis. *Palaeogeography Palaeoclimatology Palaeoecology*, **177**, 19–27.

Behling, H. & Pillar, V.D. (2007) Late Quaternary vegetation, biodiversity and fire dynamics on the southern Brazilian highland and their implication for conservation and management of modern Araucaria forest and grassland ecosystems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **362**, 243–251.

Behling, H., Pillar, V.D.P., Orlóci, L., & Bauermann, S.G. (2004) Late Quaternary *Araucaria* forest, grassland (Campos), fire and climate dynamics, studied by high-resolution pollen, charcoal and multivariate analysis of the Cambará do Sul core in southern Brazil. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **203**, 277–297.

Belshaw, R. & Quicke, D.L. (1997) A molecular phylogeny of the Aphidiinae (Hymenoptera: Braconidae). *Molecular Phylogenetics and Evolution*, **7**, 281–293.

Bivand, R., Keitt, T., & Rowlingson, B. (2017) rgdal: Bindings for the "Geospatial" Data Abstraction Library. .

Bivand, R. & Lewin-Koh, N. (2017) maptools: Tools for Reading and Handling Spatial Objects. .

Boldrini, I. (2009) A flora dos Campos do Rio Grande do Sul. *Campos Sulinos - conservação e uso da biodiversidade* (ed. by V.D.P. Pillar, S.C. Müller, Z.M. de S. Castilhos, and A.V.Á. Jacques), pp. 63–87. Ministério do Meio Ambiente, Brasília.

Breiman, L. (2001) Random Forests. *Mach Learn*, **45**, 5–32.

Brown, C. (2012) dummies: Create dummy/indicator variables flexibly and efficiently. .

Cabrera, A.L. & Willink, A. (1973) *Biogeografia da America Latina.* OEA, Washington.

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

Chauveau, O., Eggers, L., Raquin, C., Silvério, A., Brown, S., Couloux, A., Cruaud, C., Kaltchuk-Santos, E., Yockteng, R., Souza-Chies, T.T., & Nadot, S. (2011) Evolution of oil-producing trichomes in *Sisyrinchium* (Iridaceae): Insights from the first comprehensive phylogenetic analysis of the genus. *Annals of Botany*, **107**, 1287–1312.

Chauveau, O., Eggers, L., Souza-Chies, T.T., & Nadot, S. (2012) Oil-producing flowers within the Iridoideae (Iridaceae): evolutionary trends in the flowers of the New World genera. *Annals of botany*, **110**, 713–729.

Cocucci, A.A. & Vogel, S. (2001) Oil-producing flowers of *Sisyrinchium* species (Iridaceae) and their pollinators in southern South America. *Flora*, **196**, 26–46.

Csardi, G. & FitzJohn, R. (2016) progress: Terminal Progress Bars. .

Danforth, B.N. (1999) Phylogeny of the Bee Genus *Lasioglossum* (Hymenoptera : Halictidae) Based on Mitochondrial COI Sequence Data. *Systematic Entomology*, **24**, 377–393.

Danforth, B.N., Cardinal, S., Praz, C., Almeida, E. a B., & Michez, D. (2013) The impact of molecular data on our understanding of bee phylogeny and evolution. *Annual review of entomology*, **58**, 57–78.

Danforth, B.N., Lin, C.-P., & Fang, J. (2005) How do insect nuclear ribosomal genes compare to protein-coding genes in phylogenetic utility and nucleotide substitution patterns? *Systematic Entomology*, **30**, 549–562.

Donato, M., Posadas, P., Miranda-Esquivel, D.R., Ortiz-Jaureguizar, E., & Cladera, G. (2003) Historical biogeography of the Andean region: evidence from *Listroderina* (Coleoptera: Curculionidae: Rhytirrhinini) in the context of the South American geobiotic scenario. *Biological Journal of the Linnean Society*, **80**, 339–352.

Dragulescu, A.A. (2014) xlsx: Read, write, format Excel 2007 and Excel 97/2000/XP/2003 files. .

Dray, S. & Dufour, A.B. (2007) The ade4 package: implementing the duality diagram for

ecologists. *Journal of Statistical Software*, **22**, 1–20.

Drummond, A.J., Rambaut, A., & Bouckaert, R. (2015) Divergence Dating Tutorial with BEAST 2.0. .

Drummond, A.J., Suchard, M.A., Xie, D., & Rambaut, A. (2012) Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular biology and evolution*, **29**, 1969–1973.

Evangelista, O., Sakakibara, A.M., Cryan, J.R., & Urban, J.M. (2017) A phylogeny of the treehopper subfamily Heteronotinae reveals convergent pronotal traits (Hemiptera: Auchenorrhyncha: Membracidae). *Systematic Entomology*, **42**, 410–428.

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

Ferretti, N., Pérez-Miles, F., & González, A. (2014) Historical relationships among Argentinean biogeographic provinces based on mygalomorph spider distribution data (Araneae: Mygalomorphae). *Studies on Neotropical Fauna and Environment*, **49**, 1–10.

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular marine biology and biotechnology*, **3**, 294–299.

Fregonezi, J.N., Turchetto, C., Bonatto, S.L., & Freitas, L.B. (2013) Biogeographical history and diversification of *Petunia* and *Calibrachoa* (Solanaceae) in the Neotropical Pampas grassland. *Botanical Journal of the Linnean Society*, **171**, 140–153.

Gagolewski, M. (2018) R package stringi: Character string processing facilities. .

Genz, A. & Bretz, F. (2009) *Computation of Multivariate Normal and t Probabilities.* Springer-Verlag, Heidelberg.

Golding, N. (2014) GRaF: Species distribution modelling using latent Gaussian random fields. .

Hastie, T. (2018) gam: Generalized Additive Models. .

Hastie, T., Tibshirani, R., Leisch, F., Hornik, K., & Ripley, B.D. (2017) mda: Mixture and Flexible Discriminant Analysis. .

Hedtke, S.M., Patiny, S., & Danforth, B.N. (2013) The bee tree of life: a supermatrix approach to apoid phylogeny and biogeography. *BMC evolutionary biology*, **13**, 138.

Heled, J. & Drummond, A.J. (2012) Calibrated tree priors for relaxed phylogenetics and divergence time estimation. *Systematic Biology*, **61**, 138–149.

Hewitt, G.M. (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **359**, 183–195.

Hijmans, R.J. (2017) raster: Geographic Data Analysis and Modeling. .

Hijmans, R.J., Phillips, S., Leathwick, J., & Elith, J. (2017) dismo: Species Distribution Modeling. .

Hines, H.M. (2008) Historical biogeography, divergence times, and diversification patterns of

bumble bees (Hymenoptera: Apidae: *Bombus*). *Systematic biology*, **57**, 58–75.

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

Hughes, C.E., Pennington, R.T., & Antonelli, A. (2013) Neotropical Plant Evolution: Assembling the Big Picture. *Botanical Journal of the Linnean Society*, **171**, 1–18.

Iganci, J.R. V, Heiden, G., Miotto, S.T.S., & Pennington, R.T. (2011) Campos de Cima da Serra: The Brazilian Subtropical Highland Grasslands show an unexpected level of plant endemism. *Botanical Journal of the Linnean Society*, **167**, 378–393.

Instituto Brasileiro de Geografia e Estatística (2004) Available at: https://ww2.ibge.gov.br/home/presidencia/noticias/21052004biomashtml.shtm.

Jiang, W., Chen, S.-Y., Wang, H., Li, D.-Z., & Wiens, J.J. (2014) Should genes with missing data be excluded from phylogenetic analyses? *Molecular Phylogenetics and Evolution*, **80**, 308–318.

Jiménez-Valverde, A., Peterson, A.T., Soberón, J., Overton, J.M., Aragón, P., & Lobo, J.M. (2011) Use of niche models in invasive species risk assessments. *Biological Invasions*, **13**, 2785–2797.

Karatzoglou, A., Smola, A., Hornik, K., & Zeileis, A. (2004) kernlab -- An {S4} Package for Kernel Methods in {R}. *Journal of Statistical Software*, **11**, 1–20.

Katoh, K. & Standley, D.M. (2013) MAFFT Multiple Sequence Alignment Software Version 7 : Improvements in Performance and Usability Article Fast Track. *Molecular biology and evolution*, **30**, 772–780.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., & Drummond, A. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647–1649.

Landis, M.J., Matzke, N.J., Moore, B.R., & Huelsenbeck, J.P. (2013) Bayesian analysis of biogeography when the number of areas is large. *Systematic Biology*, **62**, 789–804.

Lanfear, R., Calcott, B., Ho, S.Y., & Guindon, S. (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, **29**, 1695–1701.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., & Calcott, B. (2017) Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, **34**, 772–773.

Lang, D.T. & the CRAN Team (2018) XML: Tools for Parsing and Generating XML Within R and S-Plus. .

Leite, Y.L.R., Costa, L.P., Loss, A.C., Rocha, R.G., Batalha-Filho, H., Bastos, A.C.,

Quaresma, V.S., Fagundes, V., Paresque, R., Passamani, M., & Pardini, R. (2016) Neotropical forest expansion during the last glacial period challenges refuge hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, **113**, 1008–1013.

Leutner, B., Horning, N., & Schwalb-Willmann, J. (2018) RStoolbox: Tools for Remote Sensing Data Analysis. .

Liaw, A. & Wiener, M. (2002) Classification and Regression by randomForest. *R News*, **2**, 18–22.

Longo, D., Lorenz-Lemke, A.P., Mäder, G., Bonatto, S.L., & Freitas, L.B. (2014) Phylogeography of the *Petunia integrifolia* complex in southern Brazil. *Botanical Journal of the Linnean Society*, **174**, 199–213.

Lorenz-Lemke, A.P., Togni, P.D., Mäder, G., Kriedt, R.A., Stehmann, J.R., Salzano, F.M., Bonatto, S.L., & Freitas, L.B. (2010) Diversification of plant species in a subtropical region of eastern South American highlands: A phylogeographic perspective on native *Petunia* (Solanaceae). *Molecular Ecology*, **19**, 5240–5251.

Lunt, D.H., Zang, D.X., Szymura, J.M., & Hewitt, G.M. (1996) The insect cytochrome oxidase I gene: evolutionary patterns and conserved primers for phylogenetic studies. *Insect Mol. Biol.*, **5**, 153–165.

Mardulyn, P. & Whitfield, J.B. (1999) Phylogenetic signal in the COI, 16S, and 28S genes for inferring relationships among genera of Microgastrinae (Hymenoptera; Braconidae): evidence of a high diversification rate in this group of parasitoids. *Molecular phylogenetics and evolution*, **12**, 282–294.

Martins, A.C., Melo, G.A.R., & Renner, S.S. (2014) The corbiculate bees arose from New World oil-collecting bees: Implications for the origin of pollen baskets. *Molecular phylogenetics and evolution*, **80**, 88–94.

Martins, A.C., Melo, G.A.R., & Renner, S.S. (2015) Gain and loss of specialization in two oil-bee lineages, Centris and Epicharis (Apidae). *Evolution*, **69**, 1835–1844.

Matzke, N.J. (2013) BioGeoBEARS: BioGeography with Bayesian (and Likelihood) Evolutionary Analysis in R Scripts. R package, version 0.2.1. .

McCullagh, P. & Nelder, J. (1989) *Generalized linear models.* Chapman and Hall/CRC,

Michener, C.D. (2007) *The Bees of the World.* The Johns Hopkins University Press, Baltimore, Maryland, USA.

Michener, C.D. & Moure, J.S. (1957) A Study of the Classification of the More Primitive non-parasitic Anthophorine Bees (Hymenoptera, Apoidea). *Bulletin of the American Museum of Natural History*, **112**, 395–452.

Morrone, J.J. (2000) What is the Chacoan subregion? *Neotrópica*, **46**, 51–68.

Morrone, J.J. (2014) Biogeographical regionalisation of the Neotropical region. *Zootaxa*, **3782**, 1–110.

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N.,

Underwood, E.C., D'amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., Loucks, C.J., Allnutt, T.F., Ricketts, T.H., Kura, Y., Lamoreux, J.F., Wettengel, W.W., Hedao, P., & Kassem, K.R. (2001) Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience*, **51**, 933–938.

Ortiz-Jaureguizar, E. & Cladera, G.A. (2006) Paleoenvironmental evolution of southern South America during the Cenozoic. *Journal of Arid Environments*, **66**, 498–532.

Otto-Bliesner, B.L., Marshall, S.J., Overpeck, J.T., Gifford, H., Miller, A.H., & CAPE Last Interglacial Project Members (2006) Simulating Arctic Climate Warmth and Icefield Retreat in the Last Interglaciation. *Science*, **311**, 1751–1753.

Overbeck, G.E., Boldrini, I.I., Carmo, M.R.B. do, Garcia, É.N., Moro, R.S., Pinto, C.E., Trevisan, R., & Zannin, A. (2015) Fisionomia dos Campos. *Os Campos do Sul* (ed. by V.D.P. Pillar and O. Lange), Rede Campos Sulinos - UFRGS, Porto Alegre.

Overbeck, G.E., Müller, S.C., Fidelis, A., Pfadenhauer, J., Pillar, V.D., Blanco, C.C., Boldrini, I.I., Both, R., & Forneck, E.D. (2007) Brazil's neglected biome: The South Brazilian Campos. *Perspectives in Plant Ecology, Evolution and Systematics*, **9**, 101–116.

Pagani, M., Freeman, K.H., & Arthur, M.A. (1999) Late Miocene atmospheric CO2 concentrations and the expansion of C4 grasses. *Science*, **285**, 876–879.

Paradis, E., Claude, J., & Strimmer, K. (2004) A{PE}: analyses of phylogenetics and evolution in {R} language. *Bioinformatics*, **20**, 289–290.

Pebesma, E.J. & Bivand, R.S. (2005) Classes and methods for spatial data in {R}. *R News*, **5**, 9–13.

Peres, E.A., Sobral-Souza, T., Perez, M.F., Bonatelli, I.A.S., Silva, D.P., Silva, M.J., & Solferini, V.N. (2015) Pleistocene Niche Stability and Lineage Diversification in the Subtropical Spider Araneus omnicolor (Araneidae). *PLoS ONE*, **10**, 1–20.

Pessenda, L.C.R., De Oliveira, P.E., Mofatto, M., de Medeiros, V.B., Francischetti Garcia, R.J., Aravena, R., Bendassoli, J.A., Zuniga Leite, A., Saad, A.R., & Lincoln Etchebehere, M. (2009) The evolution of a tropical rainforest/grassland mosaic in southeastern Brazil since 28,000 14C yr BP based on carbon isotopes and pollen records. *Quaternary Research*, **71**, 437–452.

Phillips, S. (2017) maxnet: Fitting "Maxent" Species Distribution Models with "glmnet." .

Phillips, S.J., Dudík, M., & Schapire, R.E. (2004) A Maximum Entropy Approach to Species Distribution Modeling. *Proceeding of the Twenty-first International Conference on Machine Learning*, **69**, 655–662.

Pinheiro, F., De Barros, F., Palma-Silva, C., Fay, M.F., Lexer, C., & Cozzolino, S. (2011) Phylogeography and genetic differentiation along the distributional range of the orchid *Epidendrum fulgens*: a Neotropical coastal species not restricted to glacial refugia. *Journal of Biogeography*, **38**, 1923–1935.

Porzecanski, A.L. & Cracraft, J. (2005) Cladistic analysis of distributions and endemism (CADE): Using raw distributions of birds to unravel the biogeography of the South American aridlands. *Journal of Biogeography*, **32**, 261–275.

Praz, C.J. & Packer, L. (2014) Molecular Phylogenetics and Evolution Phylogenetic position of the bee genera *Ancyla* and *Tarsalia* (Hymenoptera : Apidae): A remarkable base compositional bias and an early Paleogene geodispersal from North America to the Old World. *Molecular Phylogenetics and Evolution*, **81**, 258–270.

R Core Team (2017) R: A language and environment for statistical computing. .

Rambaut, A. (2016) FigTree, a graphical viewer of phylogenetic trees. .

Rambaut, A., Drummond, A.J., Xie, D., Baele, G., & Suchard, M.A. (2018) Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Systematic biology*, **032**, 1–3.

Rambo, B. (1950) A porta de Torres. *Anais Botânicos do Herbário Barbosa Rodrigues*, **2**, 125–136.

Ramírez, S.R., Roubik, D.W., Skov, C.E., & Pierce, N.E. (2010) Phylogeny, diversification patterns and historical biogeography of euglossine orchid bees (Hymenoptera: Apidae). *Biological Journal of the Linnean Society*, **100**, 552–572.

Ramos-Fregonezi, A.M., Fregonezi, J.N., Cybis, G.B., Fagundes, N.J., Bonatto, S.L., & Freitas, L.B. (2015) Were sea level changes during the Pleistocene in the South Atlantic Coastal Plain a driver of speciation in *Petunia* (Solanaceae)? *BMC Evolutionary Biology*, **15:92**, 11.

Ramos, K.S. & Melo, G.A.R. (2010) Taxonomic revision and phylogenetic relationships of the bee genus *Parapsaenythia* Friese (Hymenoptera, Apidae, Protandrenini), with biogeographic inferences for the South American Chacoan Subregion. *Systematic Entomology*, **35**, 449–474.

Räsänen, M.E., Linna, a M., Santos, J.C., & Negri, F.R. (1995) Late miocene tidal deposits in the amazonian foreland basin. *Science (New York, N.Y.)*, **269**, 386–390.

Rasmussen, C., Rasmussen, C., Olesen, J.M., & Olesen, J.M. (2000) Oil flowers and oil-collecting bees. *Scandinavian Association for Pollination Ecology honours Knut Faegri*, **39**, 23–31.

Ree, R.H. & Smith, S.A. (2008) Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology*, **57**, 4–14.

Renner, S.S. & Schaefer, H. (2010) The evolution and loss of oil-offering flowers: new insights from dated phylogenies for angiosperms and bees. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **365**, 423–435.

Ronquist, F. (1997) Dispersal-vicariance analysis: A new approach to the quantification of historical biogeography. *Systematic Biology*, **46**, 195–203.

Ronquist, F., Teslenko, M., Mark, P. van der, Ayres, D.L., Darling, A., Höhna, 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**, 539–542.

Royle, J.A., Chandler, R.B., Yackulic, C., & Nichols, J.D. (2012) Likelihood analysis of species occurrence probability from presence-only data for modelling species distributions. *Methods in Ecology and Evolution*, **3**, 545–554.

RStudio Team (2016) RStudio: Integrated Development for R. .

Safford, H.D. (1999) Brazilian Páramos I. An introduction to the physical environment and vegetation of the campos de altitude. *Journal of Biogeography*, **26**, 693–712.

Safford, H.D. (2007) Brazilian Páramos IV. Phytogeography of the campos de altitude. *Journal of Biogeography*, **34**, 1701–1722.

Salgado-Labouriau, M.L., Barberi, M., Ferraz-Vicentini, K.R., & Parizzi, M.G. (1998) A dry climatic event during the late Quaternary of tropical Brazil. *Review of Palaeobotany and Palynology*, **99**, 115–129.

Scharl, T. & Leisch, F. (2006) The Stochastic QT--clust Algorithm: Evaluation of Stability and Variance on Time--course Microarray Data. 1015–1022.

Schwarz, M.P., Tierney, S.M., Cooper, S.J.B., & Bull, N.J. (2004) Molecular phylogenetics of the allodapine bee genus Braunsapis: A-T bias and heterogeneous substitution parameters. *Molecular Phylogenetics and Evolution*, **32**, 110–122.

Silva, G.A.R., Antonelli, A., Lendel, A., Moraes, E. de M., & Manfrin, M.H. (2018) The impact of early Quaternary climate change on the diversification and population dynamics of a South American cactus species. *Journal of Biogeography*, **45**, 76–88.

Silvestro, D. & Michalak, I. (2012) RaxmlGUI: A graphical front-end for RAxML. *Organisms Diversity and Evolution*, **12**, 335–337.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., & Flook, P. (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann Entomol Soc Am.*, **87**, 651–701.

Simon, M.F., Grether, R., Queiroz, L.P. De, 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. *PNAS*, **106**, 20359–20364.

Soberon, J. & Peterson, T.A. (2005) Interpretation of Models of Fundamental Ecological Niches and Species Distributional Areas. *Biodiversity Informatics*, **2**, 1–10.

Souza, H.H.S. de (2017) *Sistemática do gênero Lanthanomelissa Holmberg (Hymenoptera, Apidae: Tapinotaspidini).* Universidade de Brasília,

Stamatakis, A. (2006) RAxML-VI-HPC : maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics Applications*, **22**, 2688–2690.

Strömberg, C.A.E. (2011) Evolution of Grasses and Grassland Ecosystems. *Annual Review of Earth and Planetary Sciences*, **39**, 517–544.

Urban, D. (1995) Espécies novas de *Lanthanomelissa* Holmberg e *Lanthanella* Michener & Moure (Hymenoptera, Anthophoridae, Exomalopsinae). *Revista Brasileira de Zoologia*, **12**, 767–777.

VanDerWal, J., Falconi, L., Januchowski, S., Shoo, L., & Storlie, C. (2014) SDMTools: Species Distribution Modelling Tools: Tools for processing data associated with species distribution modelling exercises. .

Vélez-Martin, E., Rocha, C.H., Blanco, C., Azambuja, B.O., Hasenack, H., & Pillar, V.D.P. (2015) Conversão e Fragmentação. *Os Campos do Sul* (ed. by V.D.P. Pillar and O. Lange), Rede Campos Sulinos - UFRGS, Porto Alegre.

Venables, W.N. & Ripley, B.D. (2002) *Modern Applied Statistics with S.* Springer, New York.

Vogel, S. (1974) *Ölblumen und ölsammelnde Bienen.* Akademie der Wissenschaften und der Literatur, Tropische und Subtropische Pflanzenwelt, Wiesbaden, Germany.

Werneck, F.P., Costa, G.C., Colli, G.R., Prado, D.E., & Sites, 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.

Wickham, H. (2011) The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical Software*, **40**, 1–29.

Wiens, J.J. (2003) Missing Data, Incomplete Taxa, and Phylogenetic Accuracy. *Syst. Biol*, **52**, 528–538.

Wood, S.N., Pya, N., & Safken, B. (2016) Smoothing parameter and model selection for general smooth models (with discussion). *Journal of the American Statistical Association*, **111**, 1548–1575.

Zachos, J., Pagani, H., Sloan, L., Thomas, E., & Billups, K. (2001) Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science*, **292**, 686–693.

Zanella, F.C. V. (2002) Systematics and Biogeography of the Bee Genus *Caenonomada* Ashmead, 1899 (Hymenoptera: Apidae: Tapinotaspidini). *Studies on Neotropical Fauna and Environment*, **37**, 249–261.

Apêndice

Figuras

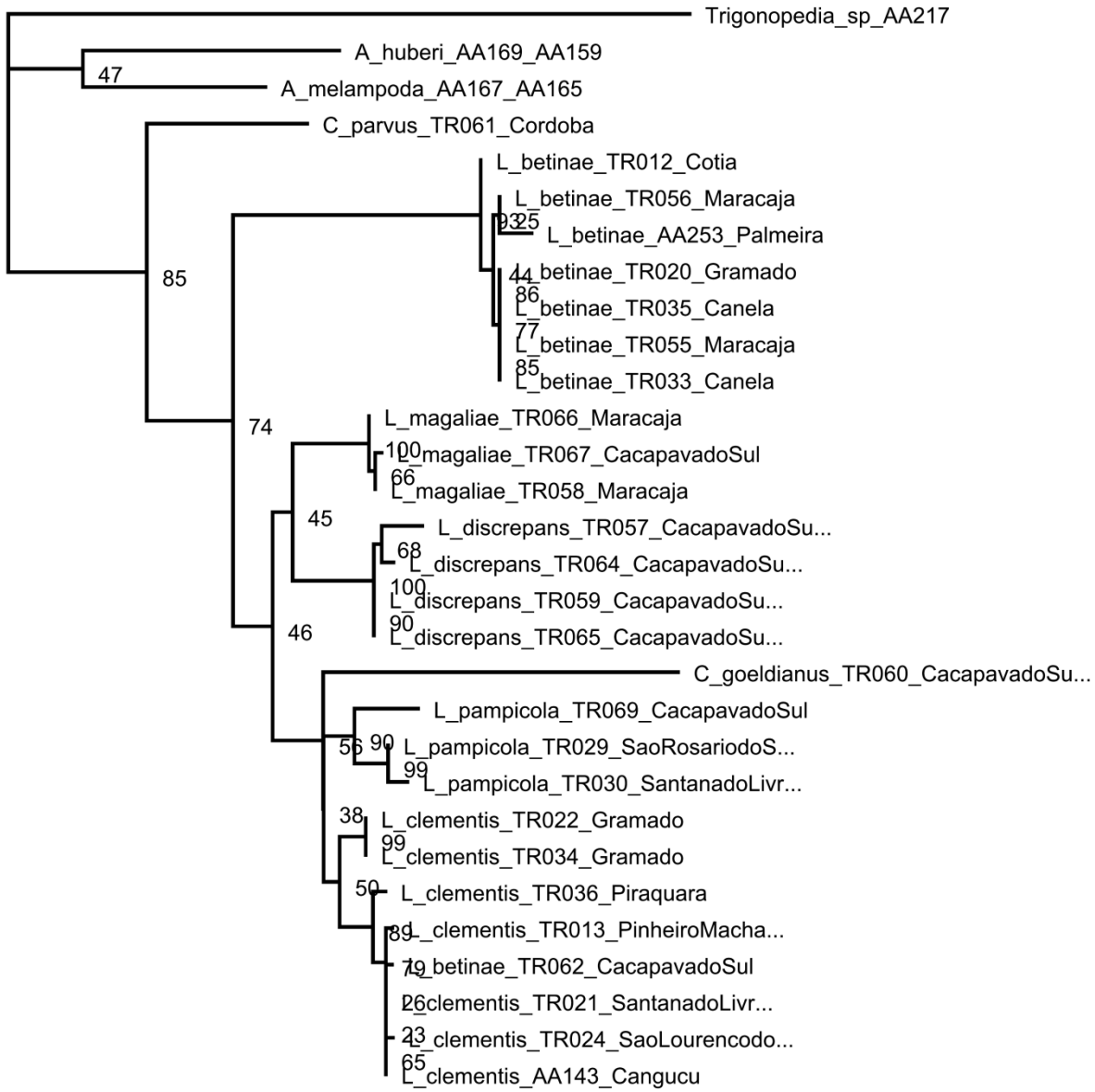

Figure A 1. Best scoring Maximum Likelihood tree of *Lanthanomelissa* with representatives of *Arhysoceble, Chalepogenus* and *Trigonopedia* as outgroups based on a matrix comprising 31 terminals and 684 nucleotides obtained from mitochondrial gene COI. Node numbers indicate bootstrap support (Accepted support > 0.7).

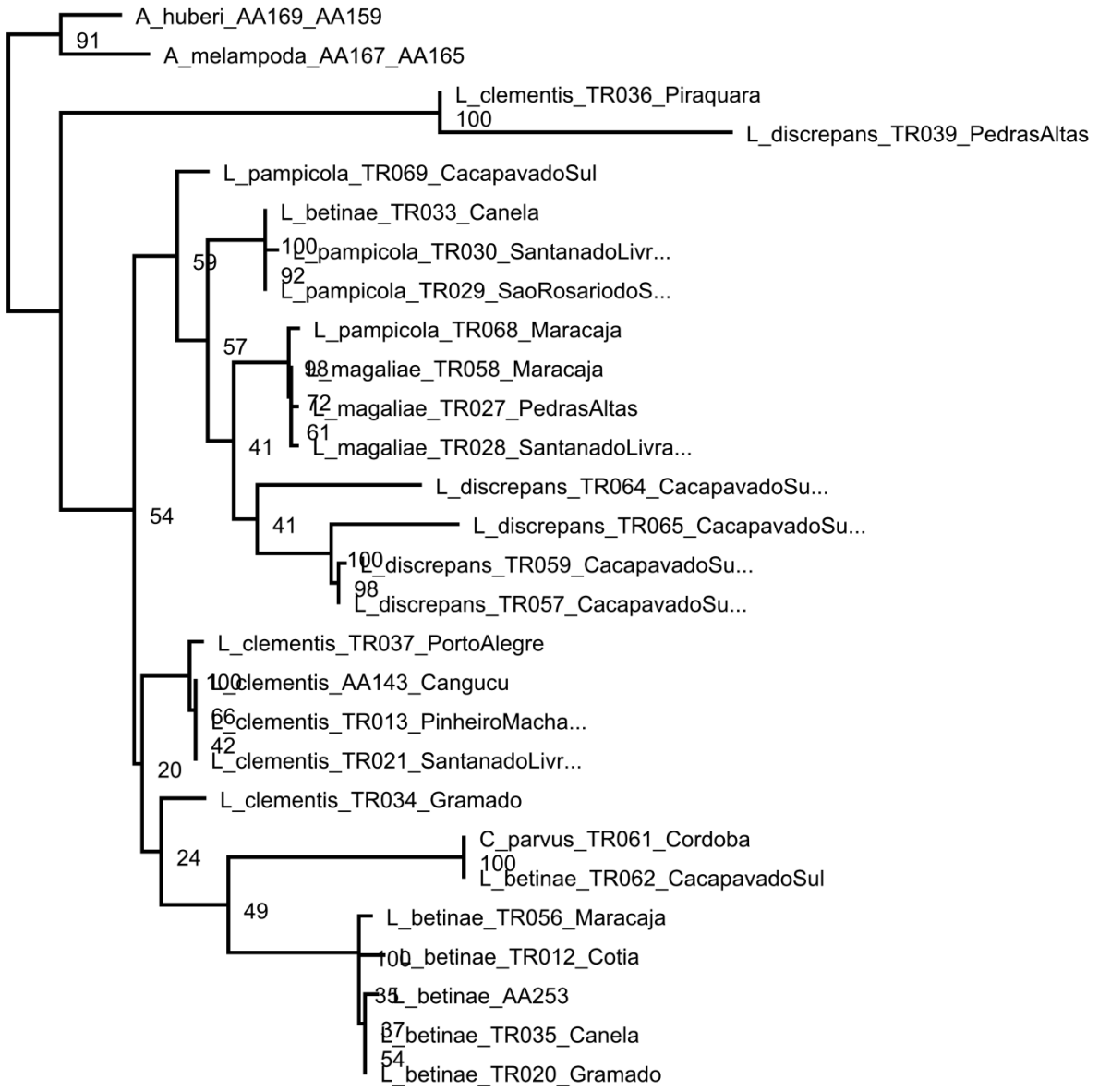

Figure A 2. Best scoring Maximum Likelihood tree of *Lanthanomelissa* with representatives of *Arhysoceble, Chalepogenus* and *Trigonopedia* as outgroups based on a matrix comprising 23 terminals and 483 nucleotides obtained from mitochondrial gene CytB. Node numbers indicate bootstrap support (Accepted support > 0.7).

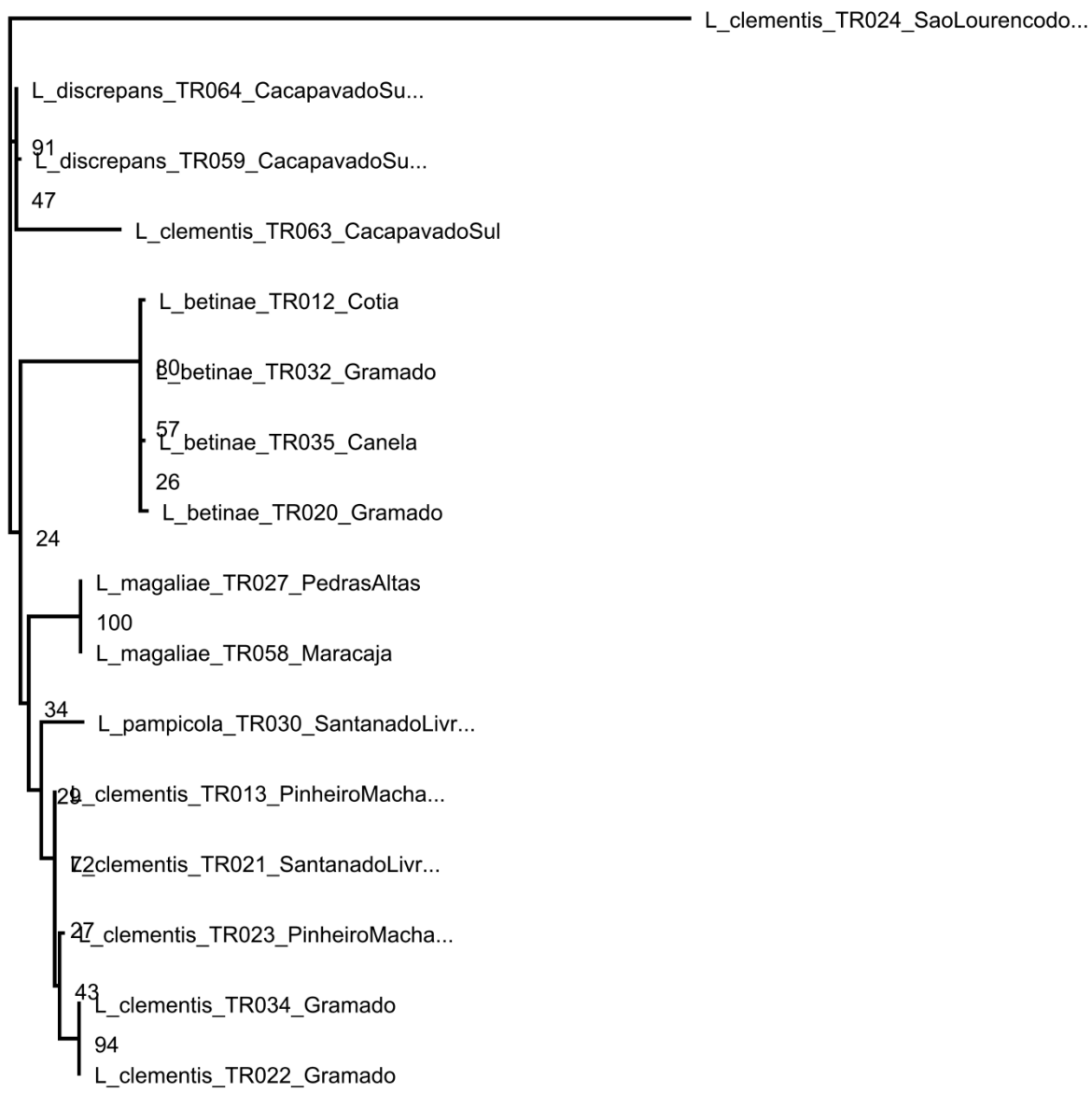

Figure A 3. Best scoring Maximum Likelihood tree of *Lanthanomelissa* with representatives of *Arhysoceble, Chalepogenus* and *Trigonopedia* as outgroups based on a matrix comprising 16 terminals and 575 nucleotides obtained from ribosomal gene 16S. Node numbers indicate bootstrap support (Accepted support > 0.7).

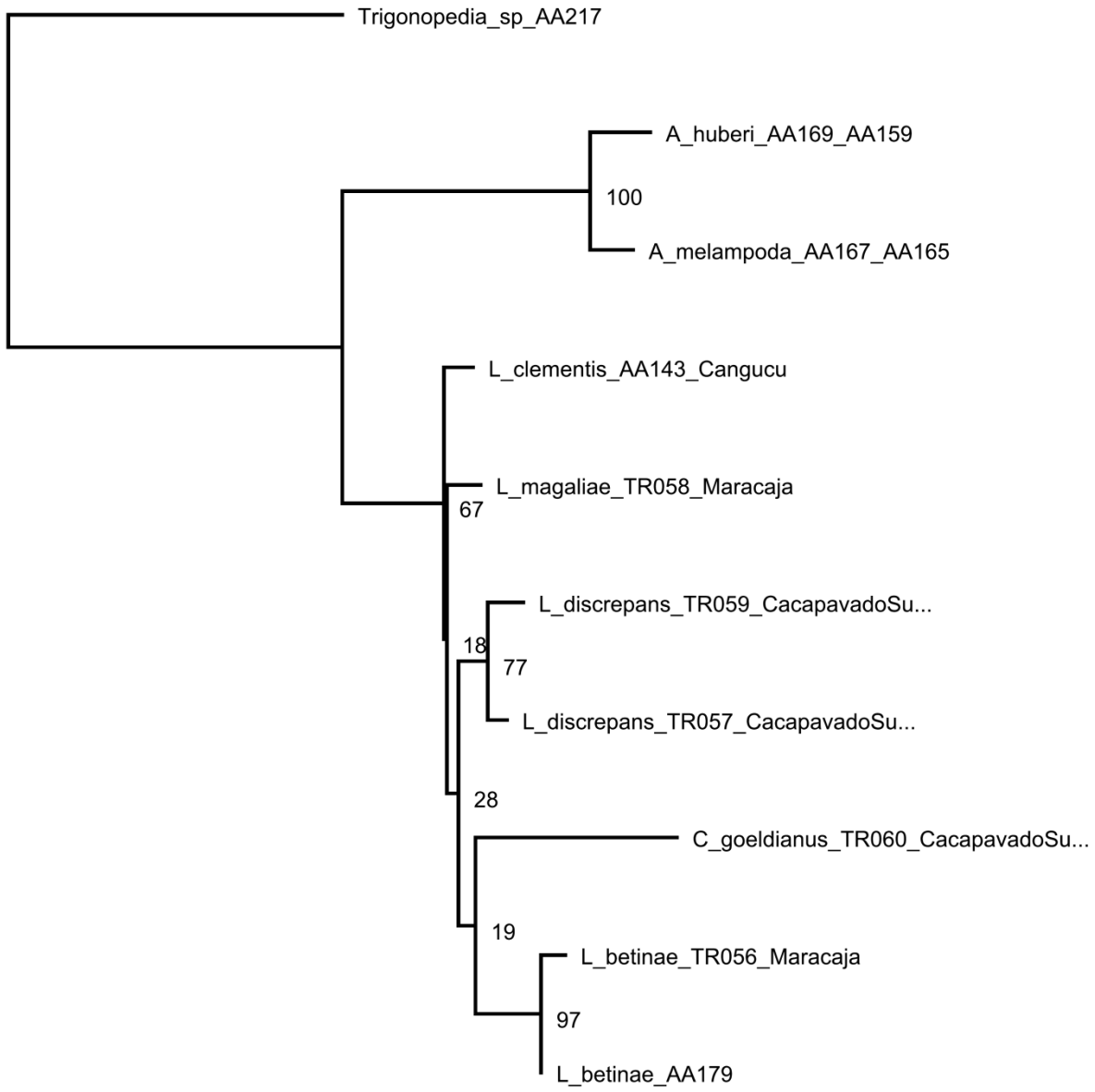

Figure A 4. Best scoring ML tree of *Lanthanomelissa* with representatives of *Arhysoceble, Chalepogenus* and *Trigonopedia* as outgroups based on a matrix comprising 12 terminals and 1048 nucleotides obtained from nuclear gene EF1-α. Node numbers indicate bootstrap support (MLBS > 0.7).

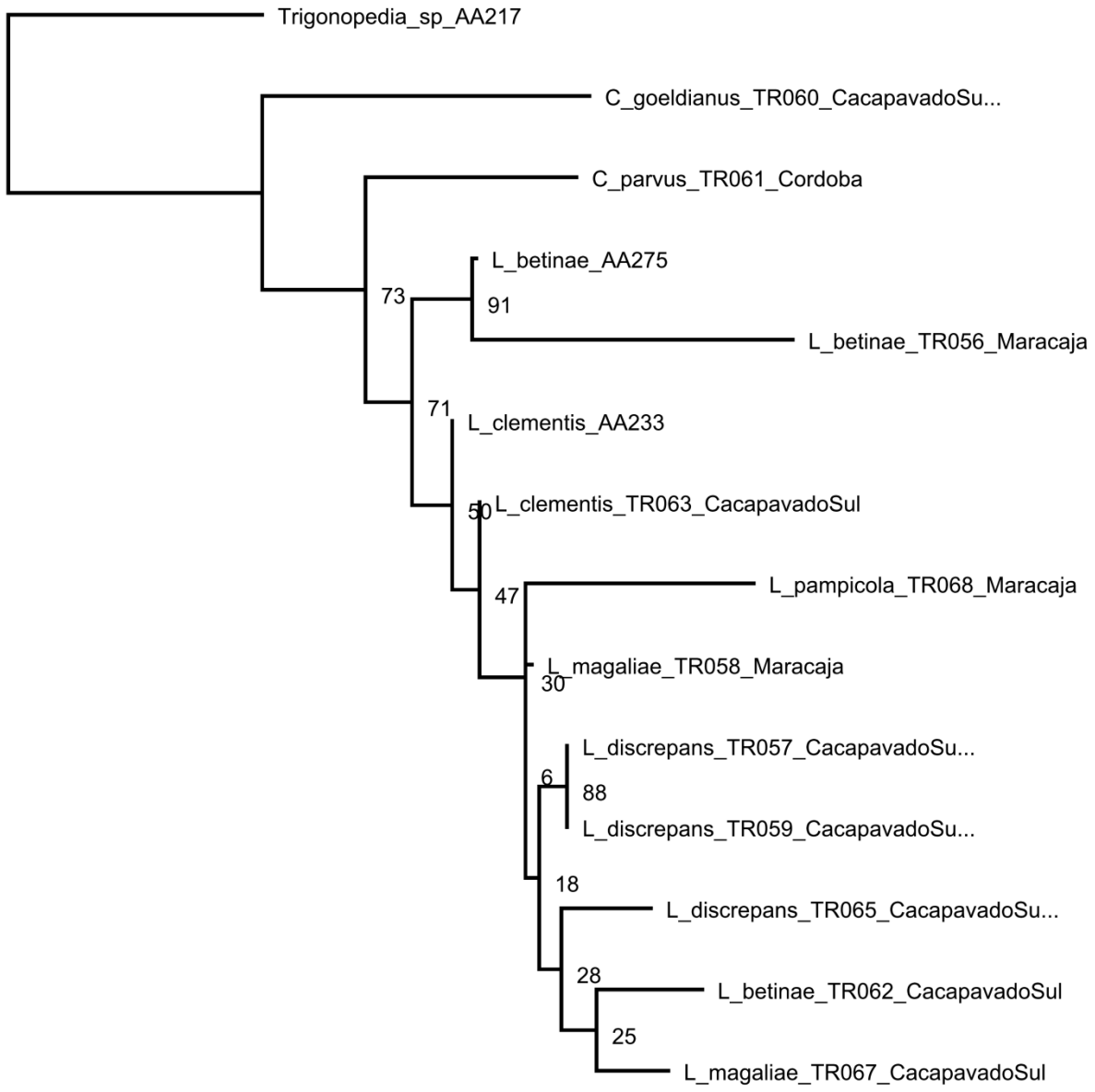

Figure A 5. Best scoring Maximum Likelihood tree of *Lanthanomelissa* with representatives of *Arhysoceble, Chalepogenus* and *Trigonopedia* as outgroups based on a matrix comprising 15 terminals and 852 nucleotides obtained from nuclear gene LW-Rhodopsin. Node numbers indicate bootstrap support (Accepted support > 0.7).

Figure A 6. Bayesian consensus tree of *Lanthanomelissa* with representatives of *Arhysoceble, Chalepogenus* and *Trigonopedia* as outgroups based on a concatenated matrix comprising 39 terminals and 3417 nucleotides obtained from concatenated mitochondrial CO1 and CytB, ribosomal 16S and nuclear EF1-α and LW-Rhodopsin using 10 partition schemes selected by PartitionFinder using the corrected Akaike Information Criterion. Node numbers indicate Bayesian posterior probabilities.

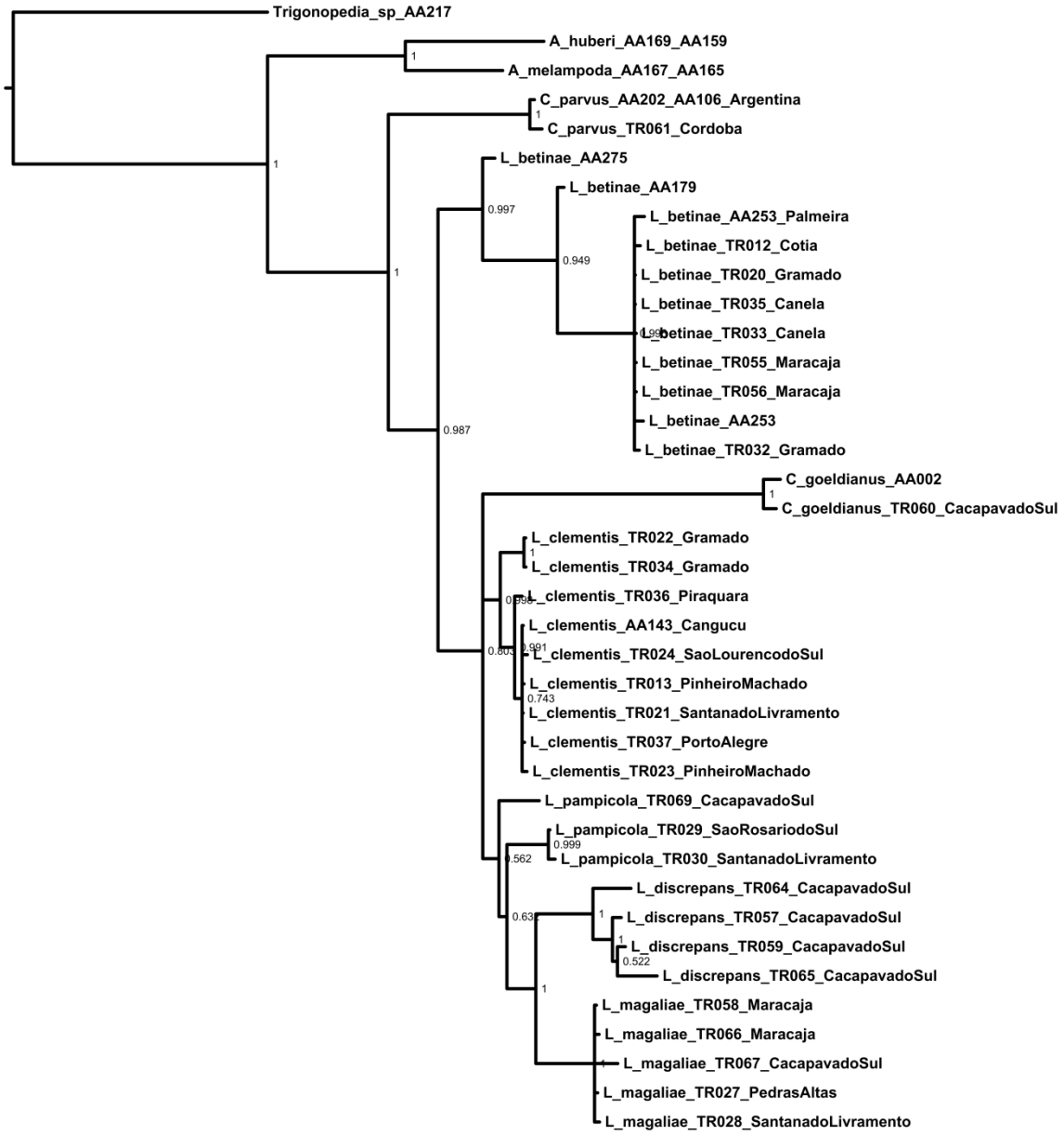

Figure A 7. Bayesian consensus tree of *Lanthanomelissa* with representatives of *Arhysoceble, Chalepogenus* and *Trigonopedia* as outgroups based on a concatenated matrix comprising 39 terminals and 2989 nucleotides obtained from concatenated mitochondrial CO1 and CytB, ribosomal 16S and nuclear EF1-α and LW-Rhodopsin excluding introns and using partition for all codons and models selected by AICc. Node numbers indicate Bayesian posterior probabilities.

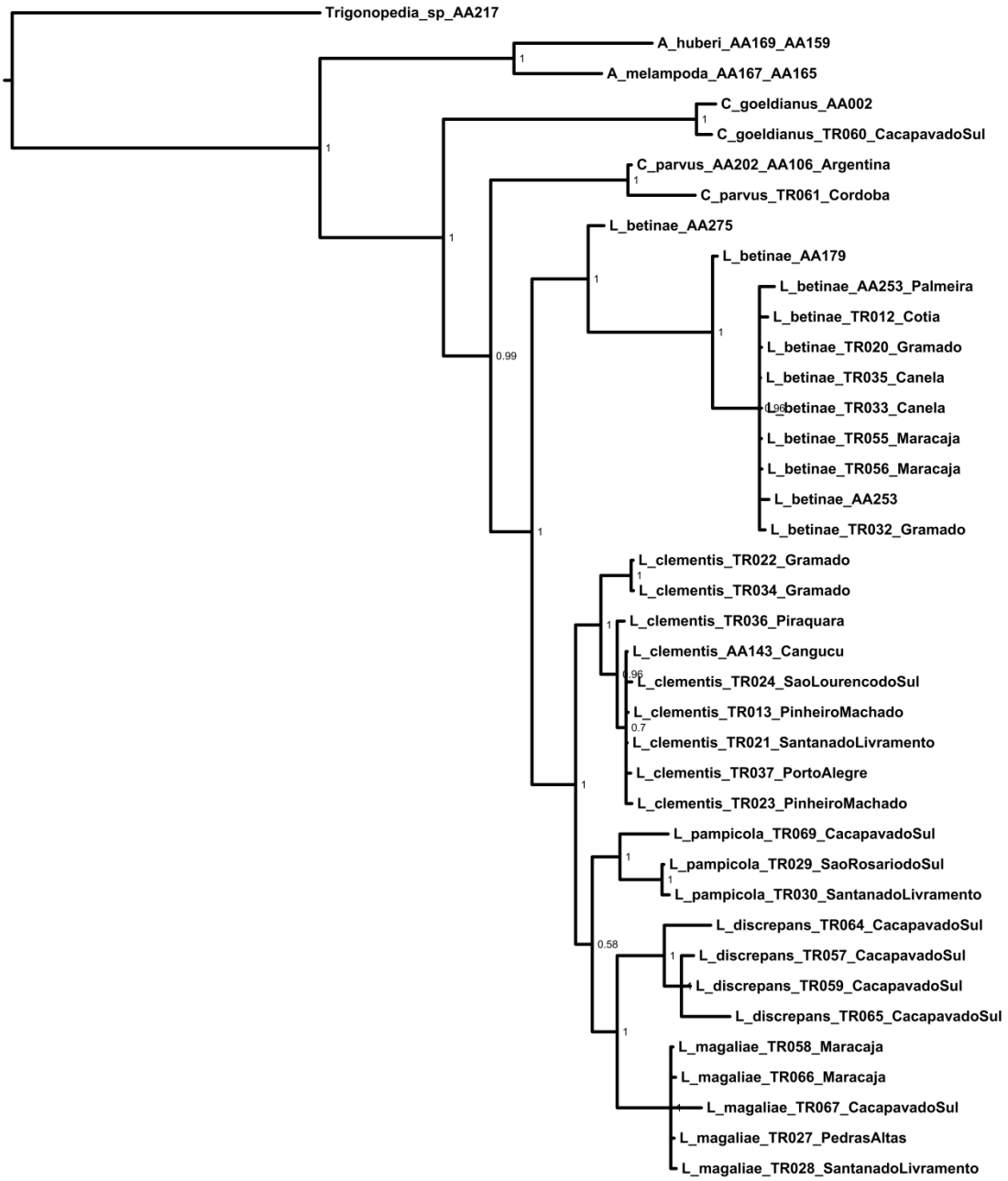

Figure A 8. Bayesian consensus tree of *Lanthanomelissa* with representatives of *Arhysoceble, Chalepogenus* and *Trigonopedia* as outgroups based on a concatenated matrix comprising 39 terminals and 3417 nucleotides obtained from concatenated mitochondrial CO1 and CytB, ribosomal 16S and nuclear EF1-α and LW-Rhodopsin using 14 partition schemes selected by PartitionFinder using the Bayesian Information Criterion. Node numbers indicate Bayesian posterior probabilities.

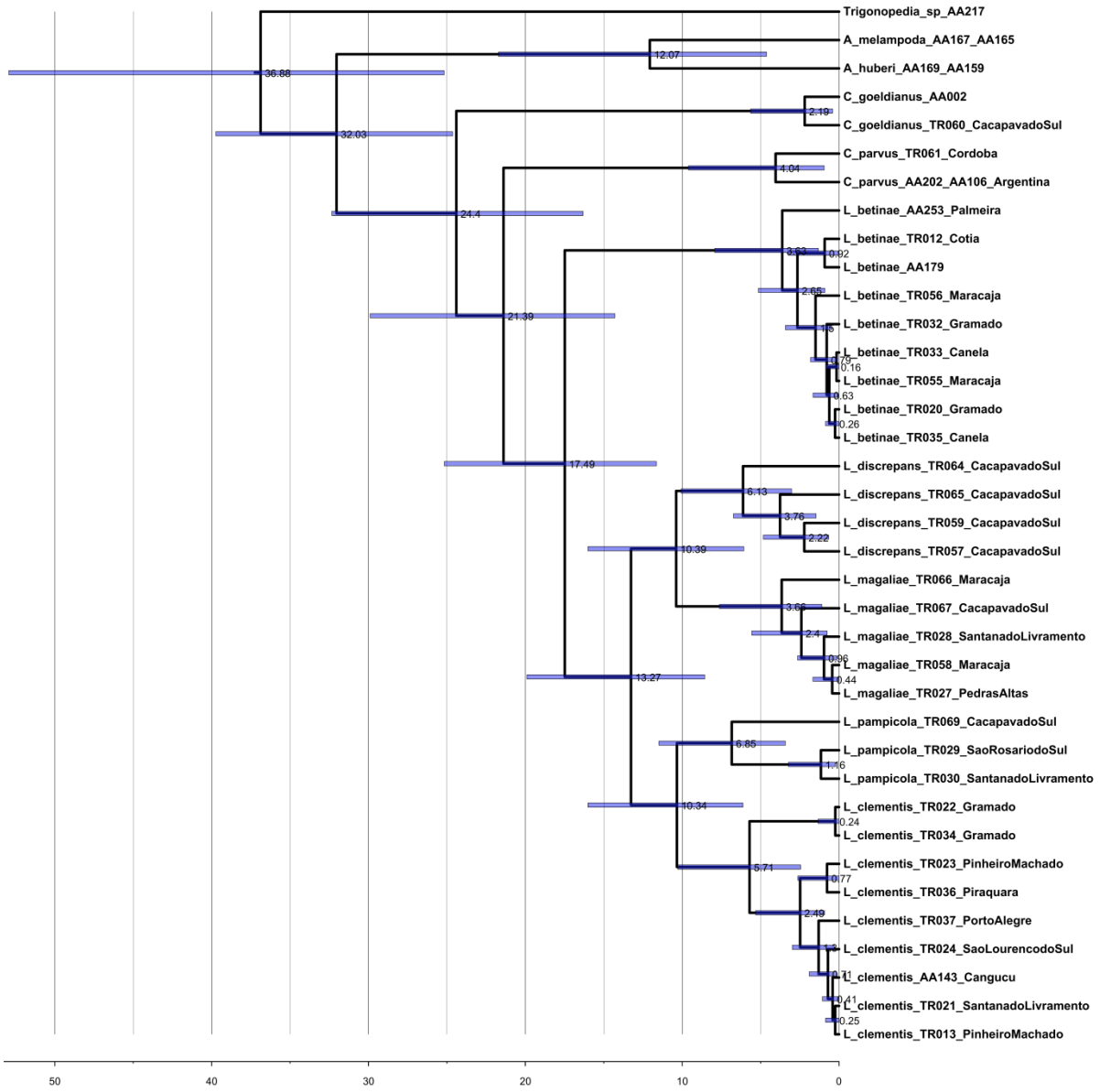

Figure A 9. Time calibrated phylogeny of *Lanthanomelissa* with representatives of *Arhysoceble*, *Chalepogenus* and *Trigonopedia* as outgroups based on a concatenated matrix comprising 37 terminals and 3430 nucleotides with 20 million generations, age of the root node sampled from a normal distribution with a mean of 33 and a s.d. of four. Horizontal bars indicate 95% HPD of estimated divergence times. Star represents node used for calibration.

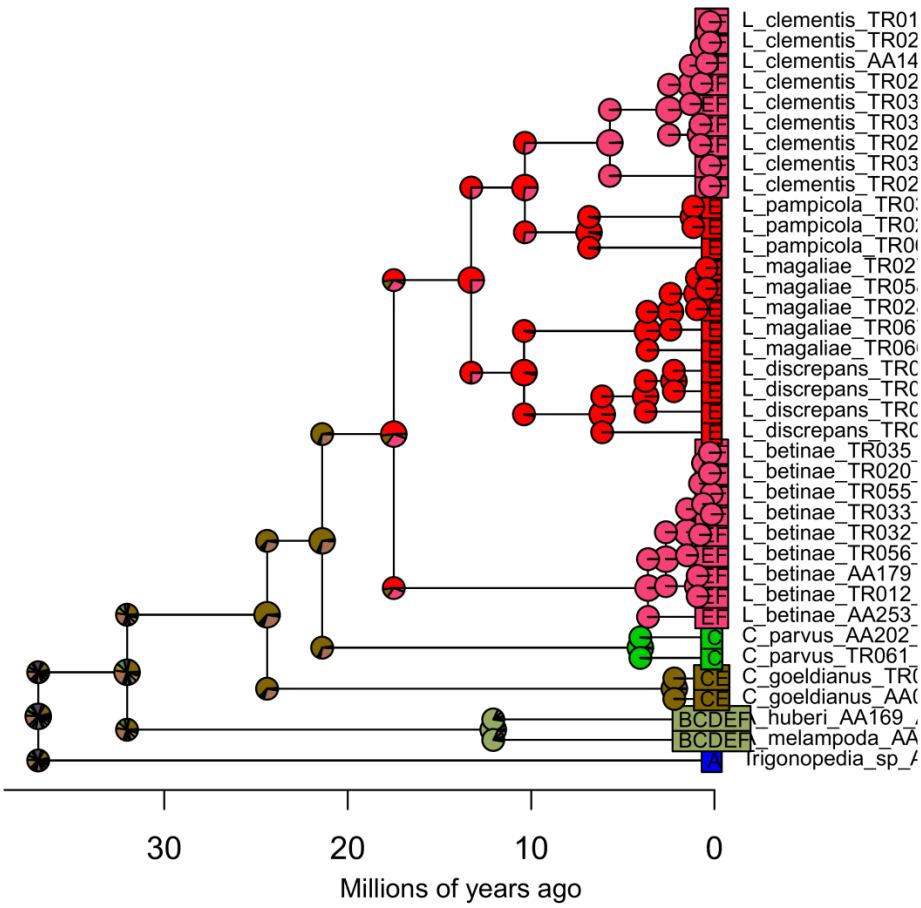

Figure A 10. Ancestral area reconstruction from BAYAREALIKE model implemented in BioGeoBEARS with logarithm likelihood (LnL = -29.95). Pie charts at nodes indicate probability for each group of areas. Letters in squares at the tips indicate the areas as follows: A: Atlantic, B: Cerrado, C: Chacoan, D: Caatinga, E: Pampean, F: Araucaria.

Tables

Table A 1. Geographic coordinates (in decimals) obtained for the five *Lanthanomelissa* species from literature, voucher labels and databases GBIF and Specieslink

Species	Longitude	Latitude
Lanthanomelissa betinae	-55.448056	-30.879444
Lanthanomelissa betinae	-54.482781	-27.870999
Lanthanomelissa betinae	-53.605986	-28.639474
Lanthanomelissa betinae	-53.524000	-30.761000
Lanthanomelissa betinae	-53.500000	-30.833333
Lanthanomelissa betinae	-53.498856	-25.543328
Lanthanomelissa betinae	-53.455278	-24.955833
Lanthanomelissa betinae	-53.410000	-30.885278
Lanthanomelissa betinae	-52.679347	-31.396459
Lanthanomelissa betinae	-52.416679	-27.160938
Lanthanomelissa betinae	-52.383333	-27.183333
Lanthanomelissa betinae	-52.337722	-31.764788
Lanthanomelissa betinae	-52.017300	-25.544500
Lanthanomelissa betinae	-51.555900	-25.392800
Lanthanomelissa betinae	-51.468855	-25.383333
Lanthanomelissa betinae	-51.316991	-30.108499
Lanthanomelissa betinae	-51.236911	-30.122218
Lanthanomelissa betinae	-51.233980	-30.122674
Lanthanomelissa betinae	-51.122798	-30.058093
Lanthanomelissa betinae	-51.116667	-30.066667
Lanthanomelissa betinae	-51.080818	-26.235592
Lanthanomelissa betinae	-50.936509	-29.379558
Lanthanomelissa betinae	-50.881238	-29.400750
Lanthanomelissa betinae	-50.873900	-29.378600
Lanthanomelissa betinae	-50.841542	-29.307670
Lanthanomelissa betinae	-50.839575	-29.360437
Lanthanomelissa betinae	-50.832125	-29.351834
Lanthanomelissa betinae	-50.683600	-29.333800
Lanthanomelissa betinae	-50.580609	-29.444088
Lanthanomelissa betinae	-50.281648	-29.885392
Lanthanomelissa betinae	-50.269700	-29.886700
Lanthanomelissa betinae	-50.152744	-24.805856
Lanthanomelissa betinae	-50.021272	-25.246354
Lanthanomelissa betinae	-49.999000	-25.238000
Lanthanomelissa betinae	-49.983333	-25.233333
Lanthanomelissa betinae	-49.948540	-29.195263
Lanthanomelissa betinae	-49.766850	-25.466700
Lanthanomelissa betinae	-49.762100	-28.163900
Lanthanomelissa betinae	-49.588660	-28.007451
Lanthanomelissa betinae	-49.455654	-25.575726
Lanthanomelissa betinae	-49.436212	-28.841985
Lanthanomelissa betinae	-49.408306	-28.701298
Lanthanomelissa betinae	-49.269951	-25.885609
Lanthanomelissa betinae	-49.266128	-25.429371

Lanthanomelissa betinae	-49.250000	-25.416670
Lanthanomelissa betinae	-49.236472	-25.451320
Lanthanomelissa betinae	-49.234802	-25.447252
Lanthanomelissa betinae	-49.231349	-25.426888
Lanthanomelissa betinae	-49.230700	-26.230500
Lanthanomelissa betinae	-49.142918	-25.467425
Lanthanomelissa betinae	-49.119141	-25.211618
Lanthanomelissa betinae	-49.091308	-25.037229
Lanthanomelissa betinae	-49.084300	-25.460400
Lanthanomelissa betinae	-49.083333	-25.033333
Lanthanomelissa betinae	-49.056300	-25.408900
Lanthanomelissa betinae	-49.010278	-25.516667
Lanthanomelissa betinae	-48.983333	-25.500000
Lanthanomelissa betinae	-47.474900	-23.759200
Lanthanomelissa betinae	-47.087019	-23.670306
Lanthanomelissa clementis	-53.594567	-31.441442
Lanthanomelissa clementis	-53.561111	-31.720556
Lanthanomelissa clementis	-53.550000	-30.533333
Lanthanomelissa clementis	-53.537778	-31.656667
Lanthanomelissa clementis	-53.523889	-30.760000
Lanthanomelissa clementis	-53.500000	-30.833333
Lanthanomelissa clementis	-53.491400	-30.512200
Lanthanomelissa clementis	-53.486900	-30.541400
Lanthanomelissa clementis	-53.483451	-30.512781
Lanthanomelissa clementis	-53.477778	-31.641389
Lanthanomelissa clementis	-53.456944	-31.470278
Lanthanomelissa clementis	-53.453611	-31.636667
Lanthanomelissa clementis	-53.424722	-31.613889
Lanthanomelissa clementis	-53.386000	-27.346500
Lanthanomelissa clementis	-53.380573	-31.575113
Lanthanomelissa clementis	-53.214700	-31.636500
Lanthanomelissa clementis	-52.866667	-31.100000
Lanthanomelissa clementis	-52.843333	-31.476111
Lanthanomelissa clementis	-52.833333	-31.083333
Lanthanomelissa clementis	-52.703611	-31.418611
Lanthanomelissa clementis	-52.687470	-31.441442
Lanthanomelissa clementis	-52.679347	-31.396459
Lanthanomelissa clementis	-52.675600	-31.395000
Lanthanomelissa clementis	-52.674778	-31.364778
Lanthanomelissa clementis	-52.416679	-27.160938
Lanthanomelissa clementis	-52.383333	-27.183333
Lanthanomelissa clementis	-52.342500	-31.771900
Lanthanomelissa clementis	-52.337722	-31.764788
Lanthanomelissa clementis	-52.308333	-31.249167
Lanthanomelissa clementis	-51.619800	-30.848100
Lanthanomelissa clementis	-51.595300	-30.873900
Lanthanomelissa clementis	-51.555900	-25.392800
Lanthanomelissa clementis	-51.433667	-30.125920
Lanthanomelissa clementis	-51.325000	-30.113900

Lanthanomelissa clementis	-51.236911	-30.122218
Lanthanomelissa clementis	-51.233980	-30.122674
Lanthanomelissa clementis	-51.230000	-30.033100
Lanthanomelissa clementis	-51.179334	-30.051661
Lanthanomelissa clementis	-51.116667	-30.066667
Lanthanomelissa clementis	-51.085976	-30.068943
Lanthanomelissa clementis	-50.881238	-29.400750
Lanthanomelissa clementis	-50.842599	-29.346231
Lanthanomelissa clementis	-50.841542	-29.307670
Lanthanomelissa clementis	-50.839575	-29.360437
Lanthanomelissa clementis	-50.836778	-29.712361
Lanthanomelissa clementis	-50.832125	-29.351834
Lanthanomelissa clementis	-50.376724	-29.472666
Lanthanomelissa clementis	-50.021272	-25.246354
Lanthanomelissa clementis	-49.983333	-25.233333
Lanthanomelissa clementis	-49.766667	-25.466667
Lanthanomelissa clementis	-49.465108	-25.465108
Lanthanomelissa clementis	-49.266128	-25.429371
Lanthanomelissa clementis	-49.142918	-25.467425
Lanthanomelissa clementis	-49.084300	-25.460400
Lanthanomelissa clementis	-49.056300	-25.408900
Lanthanomelissa clementis	-47.474900	-23.759200
Lanthanomelissa clementis	-47.087019	-23.670306
Lanthanomelissa discrepans	-67.697395	-35.066837
Lanthanomelissa discrepans	-65.923889	-27.686944
Lanthanomelissa discrepans	-64.186802	-31.419855
Lanthanomelissa discrepans	-60.500991	-34.638486
Lanthanomelissa discrepans	-58.440937	-32.349702
Lanthanomelissa discrepans	-57.543749	-30.207029
Lanthanomelissa discrepans	-57.528440	-25.090984
Lanthanomelissa discrepans	-55.974990	-30.294211
Lanthanomelissa discrepans	-55.448056	-30.879444
Lanthanomelissa discrepans	-53.895000	-30.813100
Lanthanomelissa discrepans	-53.806010	-31.621729
Lanthanomelissa discrepans	-53.708925	-30.868046
Lanthanomelissa discrepans	-53.688333	-31.444444
Lanthanomelissa discrepans	-53.616667	-30.983333
Lanthanomelissa discrepans	-53.616389	-30.870833
Lanthanomelissa discrepans	-53.594567	-31.729488
Lanthanomelissa discrepans	-53.537778	-31.656667
Lanthanomelissa discrepans	-53.525278	-30.759167
Lanthanomelissa discrepans	-53.483451	-30.512781
Lanthanomelissa discrepans	-53.471667	-30.885000
Lanthanomelissa discrepans	-53.453611	-31.636667
Lanthanomelissa discrepans	-53.410000	-30.885278
Lanthanomelissa discrepans	-53.380573	-31.575113
Lanthanomelissa discrepans	-52.866667	-31.100000
Lanthanomelissa discrepans	-52.679347	-31.396459
Lanthanomelissa discrepans	-52.675600	-31.395000

Lanthanomelissa discrepans	-52.337722	-31.764788
Lanthanomelissa discrepans	-51.555900	-25.392800
Lanthanomelissa discrepans	-51.466900	-29.682500
Lanthanomelissa discrepans	-51.433667	-30.125920
Lanthanomelissa discrepans	-51.063294	-30.269872
Lanthanomelissa discrepans	-50.836778	-29.712361
Lanthanomelissa discrepans	-50.580609	-29.444088
Lanthanomelissa discrepans	-50.269700	-29.886700
Lanthanomelissa discrepans	-50.021272	-25.246354
Lanthanomelissa discrepans	-49.466983	-28.879637
Lanthanomelissa discrepans	-49.408306	-28.701298
Lanthanomelissa discrepans	-49.370252	-28.670384
Lanthanomelissa discrepans	-49.266128	-25.429371
Lanthanomelissa discrepans	-49.250000	-25.416670
Lanthanomelissa magaliae	-56.450997	-30.384509
Lanthanomelissa magaliae	-56.441389	-30.400833
Lanthanomelissa magaliae	-55.528333	-30.873056
Lanthanomelissa magaliae	-55.448056	-30.879444
Lanthanomelissa magaliae	-54.106900	-31.331400
Lanthanomelissa magaliae	-53.806010	-31.621729
Lanthanomelissa magaliae	-53.616667	-30.983333
Lanthanomelissa magaliae	-53.561111	-31.720556
Lanthanomelissa magaliae	-53.525278	-30.759167
Lanthanomelissa magaliae	-53.523889	-30.760000
Lanthanomelissa magaliae	-53.500000	-30.833333
Lanthanomelissa magaliae	-53.483451	-30.512781
Lanthanomelissa magaliae	-53.453611	-31.636667
Lanthanomelissa magaliae	-53.424722	-31.613889
Lanthanomelissa magaliae	-53.380573	-31.575113
Lanthanomelissa magaliae	-51.201449	-29.850881
Lanthanomelissa magaliae	-51.085976	-30.068943
Lanthanomelissa magaliae	-51.063294	-30.269872
Lanthanomelissa magaliae	-51.033300	-30.083300
Lanthanomelissa magaliae	-49.436212	-28.841985
Lanthanomelissa magaliae	-49.408306	-28.701298
Lanthanomelissa magaliae	-49.370300	-28.670400
Lanthanomelissa magaliae	-49.309405	-28.713883
Lanthanomelissa pampicola	-58.098536	-28.035259
Lanthanomelissa pampicola	-58.066667	-28.016667
Lanthanomelissa pampicola	-55.448056	-30.879444
Lanthanomelissa pampicola	-55.387332	-30.831966
Lanthanomelissa pampicola	-55.050278	-30.435556
Lanthanomelissa pampicola	-53.616667	-30.983333
Lanthanomelissa pampicola	-53.616389	-30.870833
Lanthanomelissa pampicola	-53.523889	-30.760000
Lanthanomelissa pampicola	-53.500000	-30.833333
Lanthanomelissa pampicola	-53.491400	-30.512200
Lanthanomelissa pampicola	-53.486900	-30.541400
Lanthanomelissa pampicola	-53.483451	-30.512781

Lanthanomelissa pampicola	-53.453611	-31.636667
Lanthanomelissa pampicola	-53.431079	-30.899777
Lanthanomelissa pampicola	-49.408306	-28.701298
Lanthanomelissa pampicola	-49.370300	-28.670400

Table A 2. DNA Extraction information for each specimen with voucher information, part of the animal body used, date and DNA extraction kit used for each specimens

Species	Voucher	Nº DNA	Extraction date	Conservation Vial	Tissue	Extraction Kit (Manufactor)
Lanthanomelissa discrepans	-	TR011	24/07/2015	Ethanol	Whole body	Purelinkminiprep (Invitrogen)
Lanthanomelissa betinae	-	TR012	24/07/2015	Ethanol	Whole body	Purelinkminiprep (Invitrogen)
Lanthanomelissa clementis	-	TR013	24/07/2015	Ethanol	Whole body	Purelinkminiprep (Invitrogen)
Lanthanomelissa betinae	UNB 3456	TR020	03/03/2017	Ethanol	Wing muscle	Nucleospin Tissue XS (Machery Nagel)
Lanthanomelissa clementis	UNB 3457	TR021	03/03/2017	Ethanol	Wing muscle	Nucleospin Tissue XS (Machery Nagel)
Lanthanomelissa clementis	UNB 3458	TR022	03/03/2017	Ethanol	Midleg	Nucleospin Tissue XS (Machery Nagel)
Lanthanomelissa clementis	UNB 3459	TR023	03/03/2017	Ethanol	Midleg	Nucleospin Tissue XS (Machery Nagel)
Lanthanomelissa clementis	UNB 3460	TR024	03/03/2017	Ethanol	Midleg	Nucleospin Tissue XS (Machery Nagel)
Lanthanomelissa discrepans	UNB 3461	TR025	25/05/2017	Ethanol	Wing muscle	Nucleo Spin Insect (Machery Nagel)
Lanthanomelissa discrepans	UNB 3462	TR026	03/03/2017	Ethanol	Midleg	Nucleospin Tissue XS (Machery Nagel)
Lanthanomelissa magaliae	UNB 3463	TR027	03/03/2017	Ethanol	Midleg	Nucleospin Tissue XS (Machery Nagel)
Lanthanomelissa magaliae	UNB 3464	TR028	03/03/2017	Ethanol	Midleg	Nucleospin Tissue XS (Machery Nagel)
Lanthanomelissa pampicola	UNB 3465	TR029	03/03/2017	Ethanol	Midleg	Nucleospin Tissue XS (Machery Nagel)
Lanthanomelissa pampicola	UNB 3466	TR030	03/03/2017	Ethanol	Midleg	Nucleospin Tissue XS (Machery Nagel)
Lanthanomelissa betinae	UNB 3468	TR032	14/03/2017	Ethanol	Whole body	Nucleospin Tissue XS (Machery Nagel)
Lanthanomelissa betinae	UNB 3471	TR033	14/03/2017	Ethanol	Whole body	Nucleospin Tissue XS (Machery Nagel)
Lanthanomelissa clementis	UNB 3478	TR034	14/03/2017	Ethanol	Whole body	Nucleospin Tissue XS (Machery Nagel)
Lanthanomelissa betinae	UNB 3488	TR035	14/03/2017	Ethanol	Whole body	Nucleospin Tissue XS (Machery Nagel)
Lanthanomelissa clementis	UNB 3491	TR036	14/03/2017	Ethanol	Whole body	Nucleospin Tissue XS (Machery Nagel)
Lanthanomelissa clementis	UNB 3492	TR037	14/03/2017	Ethanol	Whole body	Nucleospin Tissue XS (Machery Nagel)
Lanthanomelissa discrepans	UNB 64154	TR039	15/05/2017	Ethanol	Whole body + Midleg	Nucleo Spin Insect (Machery Nagel)
Lanthanomelissa betinae	UNB 64170	TR040	15/05/2017	Ethanol	Whole body + Midleg	Nucleo Spin Insect (Machery Nagel)
Lanthanomelissa pampicola	UNB 3504	TR042	25/05/2017	Ethanol	Wing muscle	Nucleo Spin Insect (Machery Nagel)
Lanthanomelissa betinae	UNB 3502	TR043	25/05/2017	Ethanol	Wing muscle	Nucleo Spin Insect (Machery Nagel)
Lanthanomelissa betinae	UNB 64171	TR044	13/07/2017	Dried pinned	Whole body	Dneasy Blood & Tissue (Qiagen)
Lanthanomelissa magaliae	UNESC 6523	TR045	13/07/2017	Dried pinned	Whole body	Dneasy Blood & Tissue (Qiagen)
Lanthanomelissa betinae	UNB 3505	TR046	13/07/2017	Ethanol	Wing muscle	Dneasy Blood & Tissue (Qiagen)
Lanthanomelissa magaliae	UNB 3507	TR047	13/07/2017	Ethanol	Wing muscle	Dneasy Blood & Tissue (Qiagen)

Lanthanomelissa discrepans	UNB 64162	TR049	25/09/2017	Dried pinned	Whole body	Dneasy Blood & Tissue (Qiagen)
Lanthanomelissa discrepans	UNB 64111	TR050	25/09/2017	Dried pinned	Whole body	Dneasy Blood & Tissue (Qiagen)
Lanthanomelissa magaliae	UNB 64119	TR051	25/09/2017	Dried pinned	Whole body	Dneasy Blood & Tissue (Qiagen)
Lanthanomelissa magaliae	UNESC 3146	TR052	25/09/2017	Dried pinned	Whole body	Dneasy Blood & Tissue (Qiagen)
Lanthanomelissa pampicola	UNB 64157	TR053	25/09/2017	Dried pinned	Whole body	Dneasy Blood & Tissue (Qiagen)
Lanthanomelissa pampicola	RPSP 124218	TR054	25/09/2017	Dried pinned	Whole body	Dneasy Blood & Tissue (Qiagen)
Lanthanomelissa betinae	UNB 3518	TR055	10/11/2017	Ethanol	Wing muscle	Dneasy Blood & Tissue (Qiagen)
Lanthanomelissa betinae	UNB 3525	TR056	10/11/2017	Ethanol	Wing muscle	Dneasy Blood & Tissue (Qiagen)
Lanthanomelissa discrepans	UNB 3545	TR057	10/11/2017	Ethanol	Wing muscle	Dneasy Blood & Tissue (Qiagen)
Lanthanomelissa magaliae	UNB 3509	TR058	10/11/2017	Ethanol	Wing muscle	Dneasy Blood & Tissue (Qiagen)
Lanthanomelissa discrepans	UNB 3529	TR059	10/11/2017	Ethanol	Wing muscle	Dneasy Blood & Tissue (Qiagen)
Chalepogenus goeldianus	UNB 3559	TR060	10/11/2017	Ethanol	Wing muscle	Dneasy Blood & Tissue (Qiagen)
Chalepogenus parvus	UNB 89562	TR061	10/11/2017	Ethanol	Wing muscle	Dneasy Blood & Tissue (Qiagen)
Lanthanomelissa betinae	UNB 3528	TR062	21/02/2018	Ethanol	Wing muscle	Dneasy Blood & Tissue (Qiagen)
Lanthanomelissa clementis	UNB 3547	TR063	21/02/2018	Ethanol	Wing muscle	Dneasy Blood & Tissue (Qiagen)
Lanthanomelissa discrepans	UNB 3555	TR064	21/02/2018	Ethanol	Wing muscle	Dneasy Blood & Tissue (Qiagen)
Lanthanomelissa discrepans	UNB 3550	TR065	21/02/2018	Ethanol	Wing muscle	Dneasy Blood & Tissue (Qiagen)
Lanthanomelissa magaliae	UNB 3513	TR066	21/02/2018	Ethanol	Wing muscle	Dneasy Blood & Tissue (Qiagen)
Lanthanomelissa magaliae	UNB 3548	TR067	21/02/2018	Ethanol	Wing muscle	Dneasy Blood & Tissue (Qiagen)
Lanthanomelissa pampicola	UNB 3508	TR068	21/02/2018	Ethanol	Wing muscle	Dneasy Blood & Tissue (Qiagen)
Lanthanomelissa pampicola	UNB 3554	TR069	21/02/2018	Ethanol	Wing muscle	Dneasy Blood & Tissue (Qiagen)

Table A 3. DNA regions sequenced, base pairs, primers used and bibliographic references.

DNA Region	Base pairs	Primer	Primer Sequence (5' - 3')	Reference
Cytochrome oxidase subunit 1[1]	700	HCO	GGT CAA CAA ATC ATA AAG ATA TTG G	Folmer et al., 1994
		LCO	TAA ACT TCA GGG TGA CCA AAA AAT CA	Folmer et al., 1994
	1279	UEA7	TAC AGT TGG AAT AGA CGT TGA TAC	Lunt et al., 1996
		UEA10	TCC AAT GCA CTA ATC TGC CAT ATT A	Lunt et al., 1996
	1279	M414	CCT TTT ATA ATT GGA GGA TTT GG	Schwarz et al., 2004
		M399	TCA TCT AAA AAC TTT AAT TCC TG	Schwarz et al., 2004
Cytochrome B[2]	600	MTD26	TAT GTA CTA CCA TGA GGA CAA ATA TC	Simon et al., 1994
		AMB16	ATT AC CCT CCT AAT TTA TTA GGA AT	Arias et al., 2008
16S[3]	528	16SF	TGA TAA AAA GAA ATA TTT TGA	Simon et al., 1994
		16SR	CCC GGT AAA ATT AAA ATA TAA A	Simon et al., 1994
	500	874-16S1R	TAT AGA TAG AAA CCA AYC TG	Belshaw and Quicke, 1997
		16SWb	CAC CTG TTT ATC AAA AAC AT	Mardulyn and Whitfield, 1999
Elongation Factor 1α4	1100	HAF2For 1	GGY AAA GGW TCC TTC AAR TAT GC	Danforth, 1999
		F2Rev1	AAT CAG CAG CAC CTT TAG GTG G	Danforth, 1999
		576R	CAC CMA CCA GAC CTA CCG AC	This work
		429Fb	TCA AGG GAT GGA CTG TTG AG	This work
LW-Rhodopsin[5]	800	OpsinRe	ATA TGG AGT CCA NGC CAT RAA CCA	Mardulyn and Whitfield, 1999
		OpsinFor	AAT TGC TAT TAY GAR ACN TGG GT	Mardulyn and Whitfield, 1999

PCR Conditions

1. 95ºC for 60 s, 52ºC for 60 s, 68ºC for 60 s (36 cycles)
2. 95ºC for 60 s, 50-52ºC for 60 s, 68ºC for 60 s (36 cycles)
3. 95ºC for 60 s, 54ºC for 60 s, 68ºC for 90 s (36 cycles)
4. 95ºC for 60 s, 54-56ºC for 60 s, 68ºC for 60 s - 90 s (36 cycles)
5. 95ºC for 60 s, 52ºC for 60 s, 68ºC for 90 s (36 cycles)

Table A 4. Partition strategies tested in PartitionFinder with algorithm for model selection, number of subsets yielded, gene regions partitioned, best model selected and number of sites for partition

Partition strategy	Model selection	Number of subsets	Partition name	Best model	Sites
Restricted	AICc	10	COI 1st codon	K81UF+G	226
			COI 2nd codon, CytB 3rd codon	GTR+I	378
			COI 3rd codon	TRN	225
			CytB 1st codon	F81+I	154
			CytB 2nd codon	GTR+G	153
			16S	GTR+I	548
			EF-1α exon 2, Opsin exon 2, EF-1α exon 1	TRNEF+G	1038
			Opsin intron 2, EF intron, Opsin exon 3	TRN+G	359
			Opsin exon 1	TIMEF	232
			Opsin intron 1	TVM+I	104
	BIC	7	CytB 2nd codon, COI 1st codon	HKY+G	379
			16S, COI 2nd codon, CytB 3rd codon	K81UF+I	926
			COI 3rd codon, CytB 1st codon	F81+I	379
			EF-1α exon 2, Opsin exon 2, EF-1α exon 1	TRNEF+G	1038
			Opsin exon 3, EF intron, Opsin intron 2	HKY+G	359
			Opsin exon 1	JC	232
			Opsin intron 1	JC	104
Comprehensive	AICc	14	COI 1st codon	K81UF+G	226
			Opsin exon 3/3rd codon, COI 2nd codon, CytB 3rd codon	GTR+I	389
			COI 3rd codon	TRN	225
			CytB 1st codon	F81+I	154
			CytB 2nd codon	GTR+G	153
			16S	GTR+I	548
			Opsin exon 3/1st codon, EF-1α exon 1/1st codon, Opsin exon 2/2nd codon	GTR+I	286
			EF-1α exon 1/2nd codon, Opsin exon 2/3rd codon, EF-1α exon 2/2nd codon	TVM+I	346
			EF-1α exon 2/3rd codon, EF-1α exon 1/3rd codon, Opsin exon 2/1st codon	HKY+G	346
			EF-1α intron, Opsin intron 2	HKY+G	324

			EF-1α exon 2/1st codon	TRN	72
			Opsin exon 1/1st codon, Opsin exon 3/2nd codon	F81	90
			Opsin exon 1/2nd codon, Opsin exon 1/3rd codon	TIMEF	154
			Opsin intron 1	TVM+I	104
	BIC	8	CytB 2nd codon, COI 1st codon	HKY+G	379
			16S, COI 2nd codon, CytB 3rd codon	K81UF+I	926
			Opsin exon 1/1st codon, CytB 1st codon, COI 3rd codon	F81+I	457
			Opsin exon 2/2nd codon, EF-1α exon 1/1st codon, EF-1α ex2 1st codon	F81+I	346
			EF-1α exon 1/2nd codon, Opsin exon 2/3rd codon, Opsin exon 3/1st codon, EF-1α exon 2/2nd codon	JC+I	358
			Opsin intron 2, EF-1α intron, EF-1α ex2/3rd codon, EF-1α exon 1/3rd codon, Opsin exon 2/1st codon, Opsin exon 3/3rd codon	HKY+G	681
			Opsin exon 3/2nd codon, Opsin exon 1/3rd codon, Opsin exon 1/2nd codon	JC	166
			Opsin intron 1	JC	104
No intron	AICc	11	CytB 2nd codon, COI 1st codon	GTR+I+G	379
			Opsin exon 3/3rd codon, CytB 3rd codon, COI 2nd codon	GTR+I	389
			COI 3rd codon	TRN	225
			CytB 1st codon	F81+I	154
			16S	GTR+I	548
			Opsin exon 3/1st codon, Opsin exon 2/2nd codon, EF-1α exon 1/1st codon	GTR+I	286
			EF-1α exon 2/2nd codon, EF-1α exon 1/2nd codon, Opsin exon 2/3rd codon	TVM+I	346
			EF-1α exon 2/3rd codon, Opsin exon 2/1st codon, EF-1α exon 1/3rd codon	HKY+G	346
			EF-1α exon 2/1st codon	TRN	72
			Opsin exon 1/1st codon, Opsin exon 3/2nd codon	F81	90

			Opsin exon 1/3rd codon, Opsin exon 1/2nd codon	TIMEF	154
Comprehensive, Smaller matrix	AICc	14	COI 1st codon	GTR+I	228
			COI 2nd codon	TRN	228
			CytB 1st codon, COI 3rd codon	GTR+G	382
			CytB 2nd codon, 16S	GTR+I+G	701
			CytB 3rd codon, Opsin exon 1/3rd codon	F81+I	233
			EF-1α exon 1/1st codon, EF-1α exon 2/1st codon, Opsin exon 2/3rd codon	TVM+I	343
			EF-1α exon 2/2nd codon, Opsin exon 2/1st codon, EF-1α exon 1/2nd codon	HKY+G	344
			Opsin exon 2/2nd codon, EF-1α exon 1/3rd codon	GTR+I	275
			Opsin exon 3/3rd codon, Opsin intron 2, EF-1α intron	TRN+G	338
			EF-1α exon 2/3rd codon	TRN	68
			Opsin exon 3/1st codon, Opsin exon 1/1st codon	TRNEF	94
			Opsin exon 1/2nd codon	TVMEF	79
			Opsin intron 1	TVM+I	103
			Opsin exon 3/2nd codon	K81+I	14