- 1 Systematics of the oil bee genus *Lanthanomelissa* and its implications to the biogeography of southern
- 2 grasslands

5

7

13

- 4 Running head: Lanthanomelissa bees systematics and biogeography
- 6 Taís M. A. Ribeiro<sup>1,4</sup>, Aline C. Martins<sup>2</sup>, Daniel P. Silva<sup>3</sup>, Antonio J. C. Aguiar<sup>1</sup>
- 8 <sup>1</sup>Departamento de Zoologia, Universidade de Brasília, Brasília, Distrito Federal, Brazil
- 9 <sup>2</sup>Departamento de Ecologia, Universidade de Brasília, Brasília, Distrito Federal, Brazil
- 10 <sup>3</sup>Instituto Federal Goiano, Campus Urutaí, Urutaí, Goiás, Brazil
- <sup>4</sup>Department of Entomology, University of Maryland, College Park, Maryland, USA.
- 12 Corresponding author: taismaribeiro@gmail.com

#### **Abstract**

14

38 39

40

15 Lanthanomelissa is a controversial taxonomic group of bees that was treated as subgenus of 16 Chalepogenus or an independent genus. All of its species are endemic to the south-eastern 17 grasslands of South America, an endangered and still poorly known environment. We aimed 18 19 to 20 understand the origin of this group of bees in time and space as well as the influence of Quaternary climatic fluctuations on its current distribution and possible link to the history of 21 22 the Southern Grasslands. We inferred phylogenetic relationships of *Lanthanomelissa* species 23 using 37 terminals and 3430 nucleotides of three mitochondrial and two nuclear markers and estimated divergence times and ancestral geographic range. We performed an ensemble with 24 the algorithms SVM, Maxent, and Random Forest in a dataset of 192 georeferenced 25 occurrence points using 19 WorldClim bioclimatic variables to analyse species distribution 26 27 during the current and two past climatic scenarios (LIG, ~120 kya and LGM, ~21 kya). The results support the monophyly of the genus and taxonomic changes including the species 28 29 Lanthanomelissa parva n. comb. on Lanthanomelissa, and the treatment of the goeldianus group of Chalepogenus as the genus Lanthanella. The genus originated at the Oligocene-30 Miocene border in the Chacoan-Pampean region, and the glacial-interglacial models indicated 31 expansion in Last Glacial Maximum and retraction in Last Interglacial. Expansion and 32 33 retraction of Lanthanomelissa distribution in the last glacial-interglacial indicated grasslands distributional shifts during periods of climate cooling and warming. The diversification of 34 Lanthanomelissa supported the estimated expansion of southern grasslands in South America 35 36 concurrently with the origin of Cerrado during the late Miocene. Their origin was 37 approximately synchronized with their exclusive floral host, Sisyrinchium (Iridaceae).

Keywords: Neotropical, Grasslands, oil-collecting bees, Pampa, Iridaceae, Hymenoptera

#### 1. Introduction

 Bees are the most important angiosperm pollinators, relying entirely on flower resources (Vogel 1974; Michener 2007; Ollerton *et al.* 2011), with differing degrees of specialization (Fenster *et al.* 2004). Specialized bee-plant relationships tend to restrict their distribution according to their mutualist partner (Vamosi and Vamosi 2010). *Lanthanomelissa* bees are endemic to the Campos of southern Brazil, Argentina, and Paraguay, including the Pampa and Subtropical Highland Grasslands. The tribe where it belongs – the Tapinotaspidini – originated in early Palaeocene savannas and diversified in open vegetation environments until conquering forests in Eocene (Aguiar *et al.* 2020). Tapinotaspidini bees nest in the soil, as many other groups of bees who are intimately associated to open arid regions (Michener 1979), but they overcome the risks offered by humidity by the use of floral oils in nest lining (Aguiar *et al.* 2004).

The genus *Lanthanomelissa* is a controversial taxon of bees that has been treated either as a distinct genus with two subgenera, as a subgenus of *Chalepogenus*, or as an unique genus (Michener and Moure 1957; Michener 2007; Aguiar 2012). The classification of the two genera, *Chalepogenus* and *Lanthanomelissa*, in generic or subgeneric level was considered artificial by Michener (1963), and must be subject to further studies using phylogenetic methods.

The species of the oil-collecting bee genus *Lanthanomelissa* are solitary bees that nest on the ground close to populations of *Sisyrinchium* (Iridaceae) (Rozen *et al.* 2006), flowers in which they are specialists and resource-dependent (Cocucci and Vogel 2001). *Sisyrinchium* offer oil and pollen to these bees and about 35% of its species produce oil in trichomatic glands, while the remaining are pollen offering flowers and do not offer oil (Chauveau *et al.* 2011). The species in this genus are widely distributed in Neotropical Region, always in open vegetation environments, but its highest diversity is found around Parana Basin in Southern South America, overlapping with the distribution of *Lanthanomelissa* bees, their main pollinators (Cocucci and Vogel 2001; Chauveau *et al.* 2011; Chauveau *et al.* 2012). The closest relatives of *Lanthanomelissa* are part of the genus *Chalepogenus* (Aguiar *et al.* 2020) that explore plant species from the *Nierembergia* (Solanaceae) genus for oil and pollen (Cocucci 1991). The origin of *Sisyrinchium* and *Nierembergia* stem groups are close and estimated at about 13-20 Ma (Martins *et al.* 2015).

Lanthanomelissa and most of the oil-producing Sisyrinchium are endemic to Southern South American Grasslands. These grasslands comprehend a large open vegetation area occupying parts of southern Brazil, Argentina, and Uruguay included in two biomes: Atlantic Forest (Subtropical Highland Grasslands) and Pampa (Cabrera & Willink, 1973; IBGE, 2004; Overbeck et al., 2007; Morrone, 2014). Subtropical Highland Grasslands in Atlantic Forest occurs in patches forming a mosaic with Araucaria Forest at elevations from 700 to 1300 m

a.sl. (Overbeck *et al.* 2015). 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 (Morrone 2000). The origin and relations of the Pampa and Subtropical Highland Grasslands are still barely known. The Miocene climatic cooling, Andean Uplift, and the evolution of C4 grasses have been argued as the main factors influencing open vegetation worldwide and the formation of the Dry Diagonal in South America (Simon *et al.* 2009; Pennington and Hughes 2014).

The present biome configuration of forested and open vegetation remained relatively stable since the Pliocene (5.3 - 2.5 Ma). During the last 2.5 Ma in the Quaternary, however, fluctuations between global climate cooling and warming events (Hewitt 2004) could have led to periods of expansion and retraction of grasslands in south-eastern South America vegetation (Behling 1997; Safford 1999; Behling 2002; Behling *et al.* 2004; Safford 2007; Behling and Pillar 2007). Biogeographic studies with solitary bees (Zanella 2002; Ramos and Melo 2010), spiders (Ferretti *et al.* 2014), and birds (Porzecanski and Cracraft 2005) suggest that the Pampa is related to Chaco, Atlantic Forest, and Monte vegetation. This is reinforced by the occurrence of xerophytic elements from Chaco, like the cactus *Opuntia*, and patches of Atlantic forest (Boldrini 2009). Since bees are intrinsically connected to the vegetation, their origin and diversification could help understand biome evolution. In this context, we investigated the early origin and geographic distribution through time of the endemic bee genus *Lanthanomelissa*, through molecular phylogenetics associated to biogeography and distribution models.

We were interested in understanding the origins, distributions and persistence of these bees in south-eastern South America grasslands. Moreover, we were interested in finding out if *Lanthanomelissa* originate synchronously with its unique oil host *Sisyrinchium*. To answer those questions, we reconstructed the diversification of *Lanthanomelissa* in time using molecular phylogenetic tools coupled with ancestral range reconstructions and species distribution modelling, to find out when and under which conditions this bee group arose and how was the influence of recent climatic fluctuations on its distribution. We hypothesized that *Lanthanomelissa*'s origin, diversification and Quaternary fluctuations are directly associated to the origin of South America grasslands and that of its oil hosts, the plants of the genus *Sisyrinchium*.

#### 2. Material and methods

2.1 Taxon sampling, DNA sequencing, and phylogenetic analysis

We analysed all five species recognized in *Lanthanomelissa* (sensu Urban 1995), with more than one sample per species in order to maximize the variation within the species and

118

119

120

121122

123124

125

126

127128

129

130

131

132133

134

135136

137138

139

140

141

142

143

144145

146

147148

149

150

151

152

153154

155

investigate each species dispersal history. We tried to sample all the known distribution of *Lanthanomelissa*, including high and lowland grasslands. The genera *Chalepogenus*, *Arhysoceble*, and *Trigonopedia* were used as outgroups, since they are the closest relatives to *Lanthanomelissa*, according to the comprehensive analysis for the Tapinotaspidini tribe by (Aguiar *et al.* 2020). We used specimens preserved in 95% EtOH, but also included pinned specimens collected less than 10 years ago for DNA extraction (Table S2). We tested several commercial DNA extraction kits (Table S2) following their recommended protocols. We adopted the following modifications for old museum specimens: Incubation in 70% EtOH for 24h for decontamination and rehydration; and incubation in Elution Buffer for 24h prior to digestion (Evangelista *et al.* 2017). We used different body parts aiming to improve the quality and quantity of DNA extracted (Table S2), maintaining the body integrity to assure further morphological studies.

We have newly generated 93 DNA sequences, plus 16 sequences from (Aguiar et al. 2020), all deposited in GenBank (Table S1). Vouchers are deposited in University of Brasilia's Entomological Collection or in the collections listed in Table S1. We sequenced five genes: the mitochondrial genes cytochrome oxidase subunit I (COI, ~700 bp), cytochrome B (CytB, ~600 bp), and 16S (~500 bp); and the nuclear long wavelength rhodopsin (LW- rhodopsin,  $\sim 800$  bp), and elongation factor  $1-\alpha$  F2 copy (EF- $1\alpha$ ,  $\sim 1100$  bp). We chose both fast evolving mitochondrial genes and slow evolving protein coding genes to provide useful information as well as differentiation between species (Danforth et al. 2005; Danforth et al. 2013). Table S3 shows primers and PCR conditions. We conducted all extraction and amplification procedures in the Molecular Biology Laboratory at the Zoology Department of University of Brasilia and sent samples for purification and sequencing in the company Macrogen (South Korea). We assembled, trimmed, edited, and BLAST-searched the yielded sequences using the GenBank database on Geneious 8.1.9 (Kearse et al. 2012). We aligned sequences in MAFFT extension (Katoh and Standley 2013) on Geneious 8.1.9, using 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 made minor adjustments by eye.

We divided the data matrix by gene, codon and, for the nuclear markers, also by intron and exon and searched for best models and partition scheme in PartitionFinder (Lanfear *et al.* 2012), selecting with corrected Akaike Information Criterion. All schemes and models are available in Table S4. We performed Maximum likelihood tree searches in RAxML 7.4.2 (Stamatakis 2006) using the graphical user interface raxmlGUI 1.3 (Silvestro and Michalak 2012) and 1,000 non-parametric bootstrap replicates (Felsenstein 1985). We conducted Bayesian tree searches in MrBayes (Ronquist *et al.* 2012) using a matrix composed by 37 taxa

and 3430 nucleotides with 10 million generations and four chains sampled every 1000 generations. We used 14 dataset partitions, as yielded by PartitionFinder 2.1.1 (Lanfear *et al.* 2017) searches, with models selected by Corrected Akaike Information Criterion (AICc) using the greedy algorithm (Lanfear *et al.* 2012) (Methods S1, Table S4). We discarded the first 25% of trees as burnin. 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).

## 2.2 Divergence times estimates

Since there is no known fossil for Tapinotaspidini we used a secondary calibration point from the widely sampled phylogeny of Tapinotaspidini tribe (Aguiar *et al.* 2020). The age of most recent common ancestor of *Lanthanomelissa* + *Arhysoceble* + *Chalepogenus* is 28 Ma (95% HPD 22.98-34.89). We therefore applied a normal distribution prior (with values of mean: 28, stdev: 4.0).

We estimated the divergence times for *Lanthanomelissa* in BEAST version 1.8.4 (Drummond *et al.* 2012) via the CIPRES server (Miller *et al.* 2010), relying on the same 37 taxa matrix. We estimated the divergence times for *Lanthanomelissa* with the uncorrelated lognormal relaxed model. We used the Yule tree speciation model, which is more appropriate when considering sequences from different species (Heled and Drummond 2012; Drummond *et al.* 2015) and HKY substitution model with empirical base frequencies. We ran a Markov chain Monte Carlo (MCMC) for 50 million generations sampled every 10000 generations. We assessed convergence of chains in Tracer vs. 1.6 (Rambaut *et al.* 2018), considering an Effective Sample Size (ESS) for all parameters > 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 vs. 1.4.3 (Rambaut 2016).

#### 2.3 Ancestral Area Reconstruction

We derived six biogeographic areas from the classification of Neotropical Region defined in Morrone (2014), representing the geographic range of *Lanthanomelissa* and outgroups: A) Atlantic, B) Cerrado, C) Chaco, D) Caatinga, E) Pampa, F) *Araucaria* Forest. For outgroups, we coded the geographic areas representative for all species of the lineages. For ancestral area estimation we relied on the R package 'BioGeoBEARS' (Matzke 2013), which evaluates the contribution of evolutionary processes (i.e., range expansion, range extinctions, vicariance, founder-event speciation) through the biogeographic models: DEC (Ree and Smith 2008), and modified versions of DIVA (Ronquist 1997), and BayArea

- 193 (Landis et al. 2013) named DIVA-like and BayArea-like. We do not implement DEC+J
- model since it has been demonstrated to be a poor model of geographic range evolution (Ree
- and Sanmartín 2018). We conducted likelihood ratio tests based on AICc scores to assess the
- fitness of the models.

206

207 208

209

210

211212

213

214215

216

217

218219

220

221

222

223224

225

226

227

228

229

230

- 197 We coded the geographic area according to the vegetation type of the collection locality, even
- when representing a minor patch of vegetation located inside a biome. For instance, although
- the classification from Morrone (2014) suggested the Atlantic province covering all the costal
- Santa Catarina and highlands of São Paulo states, we considered the records on Cotia (São
- Paulo) and Maracajá (Santa Catarina) as occurring in patches of grasslands in the Atlantic
- Forest, and have coded them as Araucaria Forest Province. We did the same for records in
- 203 Parana province. For discussion purposes we also consider the regionalization introduced by
- Olson et al. (2001) as modified by Antonelli et al. (2018).

## 2.4 Occurrence and environmental data sampling

We obtained 192 georeferenced points (Table S6) for all Lanthanomelissa species from literature and voucher labels deposited in entomological collections (Methods S1). When geographic coordinates were not available on the labels, we used Google Maps (Google Inc. 2018) to acquire approximate geographical coordinates from city downtown. We have also obtained data from the online databases SpeciesLink (www.splink.org.br) and GBIF (Global Biodiversity Information Facility, www.gbif.org/). The following entomological collections provided geographic location points for Lanthanomelissa: 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). We have compiled 170 unique points, being 52 for L. betinae, 48 for L. clementis, 35 for L. discrepans, 21 for L. magaliae, and 14 for L. pampicola.

To estimate potential geographic range of *Lanthanomelissa* species across a glacial-interglacial cycle, we performed species distribution models based on current (from the years 1970 to 2000), Last Glacial Maximum (LGM, 22 ky ago) and Last Interglacial (LIG, 120 ky ago) bioclimatic variables from WorldClim (Hijmans *et al.* 2005; Otto-Bliesner *et al.* 2006; Fick and Hijmans 2017) in 2.5 arcminutes resolution.

For the Last Glacial Maximum and current climatic scenarios, we performed the simulations based on the Community Climate System Models (CCSM) general circulation

model in 2.5 arcminutes resolution. To correspond to the other environmental layers, we resampled map resolution for Last Interglacial 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). 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. We then ran a Principal Component Analysis (PCA) creating new orthogonal spatialized principal components (PCs) to remove variable collinearity. 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.

Since the models consider only climate data, they do not acknowledge biotic factors or species dispersal ability (Soberón 2007), so we restricted the modelled area using maximum and minimum latitudes and longitudes from the occurrence points as parameters for the extent area modelled (longitude: minimum -70 and maximum -45; latitude: minimum -40 and maximum -20).

## 2.5 Species distribution modelling

231

232

233234

235

236

237238

239240

241

242

243

244245

246

247248

249 250

251

252

253

254

255

256257

258

259

260

261262

263

264

265

266

We performed all ecologic niche modelling analysis with the R package ENMTML (Andrade et al. 2020). Methods S1 provide information about occurrence data and environmental data sampling. We partitioned Lanthanomelissa species occurrence dataset with checkerboard partition, in which data was divided in two subsets (50% each). The first subset produced the potential distribution and the second evaluated 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), Random Forest (Breiman 2001) and SVM (Support Vector Machines, Schölkopf et al., 2001; Tax & Duin, 2004). We evaluated model performance based on the AUC (Area Under the Curve, Allouche et al., 2006) values, which is a threshold independent matrix varying from 0 to 1 and considered good when above 0.8. We have also evaluated True Skill Statistics values (TSS, Allouche et al., 2006), a threshold-dependent metric varying between -1 to +1 in which values above 0.5 are acceptable and values above 0.7 are considered good. To estimate stable areas 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 3,430 nucleotides, in which 88 sequences belong to Lanthanomelissa specimens and 21 to the outgroup. Bayesian and Maximum Likelihood inference trees both recovered similar highly supported clades (Figure S2, Figure S3) The trees show Arhysoceble as sister to a clade containing Chalepogenus goeldianus, Chalepogenus parvus and Lanthanomelissa (Bayesian Posterior Probability (BPP) = 1 and bootstrap support (BS) = 97). Lanthanomelissa species are also all well supported. In Bayesian tree, C. goeldianus is sister to C. parvus + 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).

#### 3.2 Divergence times and ancestral range reconstruction

According to the Bayesian time tree (Figure 1, Figure S3), the crown group Arhysoceble + Chalepogenus + Lanthanomelissa had its origin estimated in the Oligocene, at 26.91 (18.56 – 34.68, 95% HPD) Mya. The crown group of Chalepogenus parvus+ Lanthanomelissa had its age estimated at 20.91 (13.24–28.38) Mya, in Miocene. Crown age for C. parvus + Lanthanomelissa is estimated at 17.47 (11.16–24.79) Mya. Crown age for Lanthanomelissa is estimated at 14.31 (8.58–20.56) Mya. Individual Lanthanomelissa species had their crown ages estimated between 6.14 (2.59 – 10.16, 95% HPD for L. pampicola) and 3.12 (1.07 – 6.01 for L. magaliae).

Multi-model analysis in BioGeoBEARS yielded BayArealike including founder-event speciation event j 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 S5). According to this analysis (Figure 1), the ancestral area for the most recent common ancestor of the lineage Arhysoceble + C. goeldianus + C. parvus comprised Chacoan, Pampean and Araucaria Forest provinces. However, for the most recent common ancestor of the two Chalepogenus + Lanthanomelissa, the area comprised only Chacoan + Pampean provinces. For Lanthanomelissa the ancestral

area inferred is *Araucaria* Forest + Pampean, and this is the current area for *L. betinae* and *L. clementis*, while the other three species are limited to Pampean province.

#### 3.3 Distribution modelling of Lanthanomelissa

 Figure 2 depicts the distribution modelling for all *Lanthanomelissa* species for current climate, Last Interglacial, Last Glacial Maximum, and stable areas. Figure S5 shows separately each modelling results and all statistic values are listed in Table S7. Potential distribution of *Lanthanomelissa* species based on current climate indicates that all species could occur in larger areas of grassland vegetation, but either collecting effort was not yet exhaustive or other factors, such as mutualistic partner, limits their distribution (see blue areas in Figure 2). For all *Lanthanomelissa* species the distribution modelled for the Last Interglacial (LIG) extended to the south, including mostly southern Brazil and Uruguay, almost reaching the latitude of -40 (Figure 2). On the Last Glacial Maximum (LGM) this suitability was shifted northwards. For all species in LGM, the suitability extrapolated current continental borders because the sea level was retracted at that time.

Stable areas predicted for *L. betinae* and for *L. clementis* (Figure 2 a, b) included all three South Brazilian states (Paraná, Santa Catarina, and Rio Grande do Sul) and even a small part of São Paulo. For these two species, however, the stable areas included only a small portion of the Pampa. For *L. discrepans* (Figure 2c) the stable area is larger than for the other species, including parts of eastern Argentina, and probably expanding southwards pass latitude -40S. Northwards, it reaches part of Paraguay and a small portion of the Brazilian state of Santa Catarina. Stable areas predicted for *L. magaliae* and for *L. pampicola* (Figure 2 d, e) are very restricted in relation to the other species, limited to only a small portion of Rio Grande do Sul for *L. magaliae*, while for *L. pampicola* it includes patches in Argentina and Paraguay.

#### **4 DISCUSSION**

#### 4.1 Phylogenetic relationships and taxonomic considerations

Our phylogenetic results indicated *Lanthanomelissa* as a monophyletic genus sister to species of the paraphyletic genus *Chalepogenus* (Aguiar *et al.* 2020). The species currently described in *Chalepogenus* form two paraphyletic groups: the most speciose are composed mostly by species occurring in the Monte vegetation in Argentina (*Chalepogenus* s.s.); the remaining *Chalepogenus* are related to *Lanthanomelissa*, related to Pampas and Southern grasslands, as shown in the present paper. This result reinforces the need of a taxonomic reclassification of the species of *Chalepogenus* related to *Lanthanomelissa*, as we presented

below. Additionally, the species grouped in the *Chalepogenus* s.s. should be subject of further taxonomic review.

As indicated in our phylogeny (Fig 1, S1 and S2) Chalepogenus goeldianus is sister to all Lanthanomelissa plus Chalepogenus parvus. Therefore, we propose that C. goeldianus and related species (not sampled in our phylogeny) must be reclassified in Lanthanella Michener & Moure (1957). Lanthanella was proposed as subgenus of Lanthanomelissa by Michener and Moure (1957) as monotypic for L. (L.) goeldiana, and posteriorly Moure (1992) surged the group to genus level. Urban (1995) described Lanthanella luciane Urban 1995, and Roig-Alsina (1999) transferred L. goeldiana and L. luciane to Chalepogenus, also describing a new species to this group, which he named Chalepogenus neffi. These three species were proposed by Roig-Alsina (1999) as goeldianus species group and the present results support the significance of this clade as the distinct genus Lanthanella therefore composed by Lanthanella goeldiana (Friese, 1899), Lanthanella luciane Urban, 1995, and Lanthanella neffi n. comb. Roig-Alsina (1999) and reinforced by morphological characters such as identical genital capsules, lower apices of hind tibia acute (Roig-Alsina 1999),

Due to their sister position to *Lanthanomelissa*, we propose *Lanthanomelissa parva* n. comb., instead of designating a new genus. As a result, *Lanthanomelissa* is now composed by six species, supported also by morphological characters such as the two submarginal cells in the fore wing, reduced body size, and similar terminalia of males.

# 4.2. Lanthanomelissa origin indicates the presence of grasslands in South America in early Miocene

The origin of the stem group *Lanthanomelissa* was estimated at the transition from Oligocene to Miocene (20.91, 13.24 – 28.38 Mya) in an ancestral area comprising Pampean and Chacoan biogeographical provinces. Further, these two areas could have been fragmented by the Patagonian Sea, an ocean transgression extended from southern Patagonia to Bolivia and facilitated by Andean orogeny process (Ortiz-Jaureguizar and Cladera 2006). The crown group *Lanthanomelissa* originated at 14.31 (8.58 – 20.56 Mya) in the Middle Miocene Climatic Optimum, which was the peak of a global warming phase that started in late Oligocene (Zachos *et al.* 2001) and favoured the raise of many animal and plant groups specially in South America (Hoorn *et al.* 2010).

The origin of the crown group *Lanthanomelissa* is associated to the beginning of a phase of climatic cooling and expansion of open vegetation environments in many parts of the world, including South America. From 15 My Andean Cordillera elevations created a rainshadow effect to humid winds from west increasing the aridity in eastern South America (Hartley 2003). The higher aridity associated to the C4 grasses expansion led to the predominance of open vegetation habitats in southern South America (Lehmann *et al.* 2011; Strömberg 2011; Werneck 2011; Hoetzel *et al.* 2013; Pennington and Hughes 2014).

Moreover, in late Miocene, the low CO<sub>2</sub> pressure coupled led to a worldwide expansion of C4 grasslands, more flammable than common grasses, and the origin of savannas (Pennington and Hughes 2014), which rise to dominance in the Pliocene (Simon *et al.* 2009). Similarly, the origin of Pampa inferred here at ca. 10 Mya could have been a result from the expansion of the grasslands associated with *Araucaria* forest. This suggests that the Cerrado and the Pampa could have gone through similar diversification processes associated to the expansion of grasslands. The expansion of grasslands could have also influenced the origin of the clade *L. pampicola* + *L. discrepans* + *L. magaliae* already in a Pampa area, differently from *L. betinae* and *L. clementis*, which maintained their distribution in the *Araucaria* forest province.

## 4.3 Synchronous origin of oil source plants and Lanthanomelissa bees

380

381

382

383 384

385

386

387 388

389

390 391 392

393

394 395

396

397

398

399 400

401

402

403 404

405

406 407

408 409

410

411

412 413

414

415

416 417

Lanthanomelissa crown age is coincident with the stem origin of its preferred oil host, plants species from the Sisyrinchium genus, as well as is the origin of the stem node, thus including Chalepogenus parvus and C. goeldianus, coincident with Nierembergia (see Martins et al., 2015 for Nierembergia and Sisyrinchium ages). The coincident age, spatial distribution, and close mutualistic association (Cocucci and Vogel 2001; Chauveau et al. 2011) supports our suggestion that Lanthanomelissa and Sisyrinchium could have played important roles in each other's diversification and biogeographic history in South America. The gains of glandular trichomes in species of Sisyrinchium, especially the oil secreting ones, have been hypothesized as one of their main drivers of diversification in South America (Chauveau et al. 2011). Although the presence of oil glandular trichomes is not ancestral in Sisyrinchium, it occurred in early diverging clades (Chauveau et al. 2011), suggesting an ancient relationship with oil bee pollinators. Further analysis on diversification of these plant clades is necessary to understand the role of oil bee pollinators in speciation or extinction. Lanthanomelissa's sister lineages, i.e. lineages of Chalepogenus and Arhysoceble, would have exploited other plants for oils, as demonstrated by time of origin we estimated, current preference for oil hosts (Cocucci 1991; Renner and Schaefer 2010; Martins and Alves-dos-Santos 2013), and estimated origin of oil producing Solanaceae and Plantaginaceae (Martins et al. 2014; Martins et al. 2015). The relationship of the lineage Arhysoceble-Lanthanomelissa with floral oil families is intimately associated to the expansion of open vegetation areas in South America (Aguiar et al. in prep). The origin of floral oil glands in these younger plant clades evolved as an exaptation of the oil-bee-syndrome evolved with the Malpighiaceae since Palaeocene (Aguiar et al. in prep; Renner & Schaefer, 2010).

4.4 Climatic fluctuations drove Lanthanomelissa distribution towards grasslands areas in the Quaternary

According to our species distribution modelling, the suitability for most species shifted north- and eastwards in the LGM. 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). The distribution of *L. discrepans* and *L. pampicola* are 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).

At LGM, grasslands were dominant in south eastern South America and climate was drier and colder than in the present (Behling 1997), and possibly extended to the coast of Uruguay, as this area is also predicted as suitable for *Lanthanomelissa*. About 4 Kya, with increasing temperature the *Araucaria* Forest expanded, 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 fluctuation among grasslands and *Araucaria* forests have been suggested as important drivers of speciation in Subtropical Highland Grasslands, but not in Pampa where registers for *Araucaria* Forest are too recent (Fregonezi *et al.* 2013). Instead, abiotic factors (e.g. soil diversity) and ecological relations (e.g. intraspecific competition and pollination) could have been stronger selective pressures in this region. However, few studies use species that cohabit both environments to address the interaction between these abiotic factors and how the two grassland areas relate (Fregonezi *et al.* 2013; Peres *et al.* 2015).

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 2), suggesting a restriction in the grasslands, which were probably more expressive towards the south.

The stable areas indicated by *Lanthanomelissa* species in the last glacial-interglacial cycle varies among species, but a clear pattern of stability in highland grasslands can be observed in *L. betinae* and *L. clementis*. In colder periods, the highlands act as refugia to those species, since it favoured expansion of grasslands (Behling 1997). Other species respond differently to stable areas analysis, but we can see an overlap between stable areas that approximately correspond to the Southern Atlantic Forest refugia described in Costa et al. (2018). This region has been dynamic over climatic fluctuations of Quaternary, presenting forest refugia in the grassland phase, which rapidly expanded over grasslands when the climate became suitable (Costa *et al.* 2018).

The potential distribution based on current climate indicated large areas as suitable for most *Lanthanomelissa* species, including areas of grasslands and Chaco. Collecting effort has been not yet exhaustive, but possibly other factors, such as *Sisyrinchium* plants distribution limits their distribution.

Lanthanomelissa are dependent on their flowers, although the contrary is not always valid (Cocucci and Vogel 2001; Chauveau et al. 2012). However, the congruence of the diversification of those bees with the origin of its oil host indicates that those groups could have been drivers of diversification to each other. 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. Even so, understanding the systematics and the evolutionary history of bees can help understand biome evolution, as the origin of the genus Lanthanomelissa in an ancestral area shared by Pampa and Chaco indicates that it is related to the origin of the South America Grasslands.

## **Figures**

 **Figure 1**. 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 50 million generations, age of the root node sampled from a normal distribution (mean =28 and s.d. = 4). Map shows provinces from Morrone (2014) used for biogeographic reconstruction. Coloured squares after species names and at nodes represent current historical distributions inferred in BioGeoBEARS under the Bayarealike+j model (lnL = -28.65). Horizontal grey bars at nodes indicate 95% HPD of estimated divergence times and bottom bar indicates epochs and Quaternary period. Abbreviations: Eo: Eocene, Plio: Pliocene; Qua: Quaternary. Picture shows a pinned specimen of *Lanthanomelissa discrepans*. Bottom: Temperature variation for the last 35 My adapted from (Zachos et al., 2001), indicating Mid-Miocene Climatic Optimum and estimated origin of *Sisyrinchium* and *Nierembergia* (95% hpd, Martins et al., 2015).

**Figure 2**. 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.

# Data availability statement

493

499

500

511

512

513

- 494 Specimen voucher information and geographic localities used in this study are provided as
- Supporting Information (Table S1 and S6 respectively). Genetic sequences were deposited in
- 496 GenBank and accession numbers are provided in Supporting Information (Table S1).
- Sequence matrix, phylogenetics trees and time-calibrated tree are available in Dryad Digital
- Repository (#will be submitted as soon as the paper became accepted).

#### Acknowledgments

- 501 We thank Andrea Cocucci, Alicia Sérsic, Birgit Harter Marques, Clemens Schlindwein,
- 502 Reisla Oliveira, Eduardo Almeida, Gabriel Melo, Jerome Rozen, Juan Pablo Torreta, Mabel
- Lizaraso, Kelli Ramos, Leopoldo Alvarez, Rafael Ferrari, Rodrigo Gonçalves, and Vicent Lee
- 504 for providing specimens for this study. We also thank Lilian Giugliano, Kelli Ramos and
- Anahí Espíndola for suggestions on previous versions of this manuscript. We thank P. Cedro
- for designing figures, M. Cavalcante, and V. Silva for helping with R programming and F.
- 507 Duque for grammar revision. TMAR received scholarships from Fundação de Apoio à
- Pesquisa do Distrito Federal (FAPDF) and ACM received scholarships from CAPES and
- 509 CNPq during the development of this research. AJCA thanks CAPES and FAPDF for grant
- support (process n. 193.000.893/2015).

#### References

- 514 Aguiar, A. J. C. (2012). Tapinotaspidini Roig-Alsina & Michener, 1993. Moure, J. S., Urban,
- 515 D. & Melo, G. A. R. (Orgs). Catalogue of Bees (Hymenoptera, Apoidea) in the Neotropical
- Region online version. Available at: http://www.moure.cria.org.br/catalogue [accessed 10]
- 517 July 2018]
- Aguiar, A. J. C., Melo, G. A. R., Vasconcelos, T. N. C., Gonçalves, R. B., Giugliano, L., and
- Martins, A. C. (2020). Biogeography and early diversification of Tapinotaspidini oil-bees
- support presence of Paleocene savannas in South America. Molecular Phylogenetics and
- 521 Evolution **143**, 106692. doi:10.1016/j.ympev.2019.106692
- Allouche, O., Tsoar, A., and Kadmon, R. (2006). Assessing the accuracy of species
- 523 distribution models: Prevalence, kappa and the true skill statistic (TSS). Journal of Applied
- 524 *Ecology* **43**, 1223–1232. doi:10.1111/j.1365-2664.2006.01214.x
- Andrade, A. F. A. de, Velazco, S. J. E., and De Marco Júnior, P. (2020). ENMTML: An R
- package for a straightforward construction of complex ecological niche models.
- 527 Environmental Modelling and Software 125. doi:10.1016/j.envsoft.2019.104615
- 528 Antonelli, A., Zizka, A., Carvalho, F. A., Scharn, R., Bacon, C. D., Silvestro, D., and
- 529 Condamine, F. L. (2018). Amazonia is the primary source of Neotropical biodiversity.
- 530 Proceedings of the National Academy of Sciences USA 115, 6034–6039.
- 531 doi:10.1073/pnas.1713819115
- Barros, M. J. F., Diniz-Filho, J. A. F., Freitas, L. B., Barros, M. J. F., Diniz-Filho, J. A. F.,

- and 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*
- 535 *Biology* **41**, 318–326. doi:10.1590/1678-4685-gmb-2017-0031
- 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
- 538 Palaeobotany and Palynology **97**, 109–121. doi:10.1016/S0034-6667(96)00065-6
- Behling, H. (2002). South and southeast Brazilian grassland during Late Quaternary times: a
- synthesis. *Palaeogeography Palaeoclimatology Palaeoecology* **177**, 19–27.
- 541 doi:10.1016/S0031-0182(01)00349-2
- Behling, H., and Pillar, V. D. (2007). Late Quaternary vegetation, biodiversity and fire
- 543 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.
- 546 doi:10.1098/rstb.2006.1984
- Behling, H., Pillar, V. D. P., Orlóci, L., and Bauermann, S. G. (2004). Late Quaternary
- 548 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. doi:10.1016/S0031-
- 551 0182(03)00687-4
- 552 Boldrini, I. (2009). A flora dos Campos do Rio Grande do Sul. In 'Campos Sulinos -
- conservação e uso da biodiversidade'. (Eds 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. doi:10.1023/A:1010933404324
- Cabrera, A. L., and Willink, A. (1973). 'Biogeografia da America Latina'. (OEA:
- 557 Washington.)
- Chauveau, O., Eggers, L., Raquin, C., Silvério, A., Brown, S., Couloux, A., Cruaud, C.,
- Kaltchuk-Santos, E., Yockteng, R., Souza-Chies, T. T., and 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.
- 562 doi:10.1093/aob/mcr080
- Chauveau, O., Eggers, L., Souza-Chies, T. T., and Nadot, S. (2012). Oil-producing flowers
- within the Iridoideae (Iridaceae): evolutionary trends in the flowers of the New World genera.
- 565 Annals of Botany **110**, 713–729. doi:10.1093/aob/mcs134
- 566 Cocucci, A. A. (1991). Pollination Biology of *Nierembergia* (Solanaceae). *Plant Systematics*
- 567 and Evolution **174**, 17–35.
- Cocucci, A. A., and Vogel, S. (2001). Oil-producing flowers of Sisyrinchium species
- 569 (Iridaceae) and their pollinators in southern South America. Flora 196, 26–46.
- 570 Costa, G. C., Hampe, A., Ledru, M.-P., Martinez, P. A., Mazzochini, G. G., Shepard, D. B.,
- Werneck, F. P., Moritz, C., and Carnaval, A. C. (2018). Biome stability in South America
- over the last 30 kyr: Inferences from long-term vegetation dynamics and habitat modelling.
- 573 *Global Ecology and Biogeography* **27**, 285–297. doi:10.1111/geb.12694

- Danforth, B. N., Cardinal, S., Praz, C., Almeida, E. A. B., and Michez, D. (2013). The impact
- of molecular data on our understanding of bee phylogeny and evolution. *Annual review of*
- 576 *entomology* **58**, 57–78. doi:10.1146/annurev-ento-120811-153633
- 577 Danforth, B. N., Lin, C.-P., and Fang, J. (2005). How do insect nuclear ribosomal genes
- compare to protein-coding genes in phylogenetic utility and nucleotide substitution patterns?
- 579 Systematic Entomology **30**, 549–562. doi:10.1111/j.1365-3113.2005.00305.x
- Drummond, A. J., Rambaut, A., and Bouckaert, R. (2015). Divergence Dating Tutorial with
- 581 BEAST 2.0.
- Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian Phylogenetics
- with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29**, 1969–1973.
- 584 doi:10.1093/molbev/mss075
- Evangelista, O., Sakakibara, A. M., Cryan, J. R., and Urban, J. M. (2017). A phylogeny of the
- treehopper subfamily Heteronotinae reveals convergent pronotal traits (Hemiptera:
- Auchenorrhyncha: Membracidae). *Systematic Entomology* **42**, 410–428.
- 588 doi:10.1111/syen.12221
- Felsenstein, J. (1985). Confidence Limits on Phylogenies: an Approach Using the Bootstrap.
- 590 Evolution **39**, 783–791. doi:10.1111/j.1558-5646.1985.tb00420.x
- Fenster, C. B., Armbruster, W. S., Wilson, P., Dudash, M. R., and Thomson, J. D. (2004).
- 592 Pollination Syndromes and Floral Specialization. Annual Review of Ecology, Evolution, and
- 593 *Systematics* **35**, 375–403. doi:10.1146/annurev.ecolsys.34.011802.132347
- Ferretti, N., Pérez-Miles, F., and González, A. (2014). Historical relationships among
- Argentinean biogeographic provinces based on mygalomorph spider distribution data
- 596 (Araneae: Mygalomorphae). Studies on Neotropical Fauna and Environment 49, 1–10.
- 597 doi:10.1080/01650521.2014.903616
- 598 Fick, S. E., and Hijmans, R. J. (2017). Worldclim 2: New 1-km spatial resolution climate
- surfaces for global land areas. *International Journal of Climatology*.
- Fregonezi, J. N., Turchetto, C., Bonatto, S. L., and Freitas, L. B. (2013). Biogeographical
- 601 history and diversification of *Petunia* and *Calibrachoa* (Solanaceae) in the Neotropical
- Pampas grassland. Botanical Journal of the Linnean Society 171, 140–153.
- 603 doi:10.1111/j.1095-8339.2012.01292.x
- 604 Hartley, A. J. (2003). Andean uplift and climate change. *Journal of the Geological Society*
- 605 **160**, 7–10. doi:10.1144/0016-764902-083
- 606 Heled, J., and Drummond, A. J. (2012). Calibrated tree priors for relaxed phylogenetics and
- divergence time estimation. Systematic Biology **61**, 138–149. doi:10.1093/sysbio/syr087
- Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary.
- 609 Philosophical Transactions of the Royal Society B: Biological Sciences **359**, 183–195.
- 610 doi:10.1098/rstb.2003.1388
- Hijmans, R. J. (2017). raster: Geographic Data Analysis and Modeling.
- 612 Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A. (2005). Very high
- 613 resolution interpolated climate surfaces for global land areas. *International Journal of*

- 614 *Climatology* **25**, 1965–1978. doi:10.1002/joc.1276
- Hoetzel, S., Dupont, L., Schefuß, E., Rommerskirchen, F., and Wefer, G. (2013). The role of
- fire in Miocene to Pliocene C4 grassland and ecosystem evolution. *Nature Geoscience* 6,
- 617 1027.
- 618 Hoorn, C., Wesselingh, F. P., Ter Steege, H., Bermudez, M. A., Mora, A., Sevink, J.,
- 619 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., and Antonelli, A.
- 621 (2010). Amazonia through time: Andean uplift, climate change, landscape evolution, and
- 622 biodiversity. *Science* **330**, 927–931. doi:10.1126/science.1194585
- 623 Instituto Brasileiro de Geografia e Estatística (2004). Mapa da vegetação do Brasil e Mapa de
- 624 Biomas do Brasil. Available at:
- https://ww2.ibge.gov.br/home/presidencia/noticias/21052004biomashtml.shtm
- Jiménez-Valverde, A., Peterson, A. T., Soberón, J., Overton, J. M., Aragón, P., and Lobo, J.
- 627 M. (2011). Use of niche models in invasive species risk assessments. *Biological Invasions* 13,
- 628 2785–2797. doi:10.1007/s10530-011-9963-4
- Katoh, K., and Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software
- 630 Version 7□: Improvements in Performance and Usability Article Fast Track. *Molecular*
- 631 *Biology and Evolution* **30**, 772–780. doi:10.1093/molbev/mst010
- 632 Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S.,
- 633 Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., and Drummond,
- A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the
- organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649.
- 636 Landis, M. J., Matzke, N. J., Moore, B. R., and Huelsenbeck, J. P. (2013). Bayesian analysis
- of biogeography when the number of areas is large. Systematic Biology **62**, 789–804.
- 638 doi:10.1093/sysbio/syt040
- 639 Lanfear, R., Calcott, B., Ho, S. Y., and Guindon, S. (2012). PartitionFinder: combined
- selection of partitioning schemes and substitution models for phylogenetic analyses.
- 641 *Molecular Biology and Evolution* **29**, 1695–1701.
- Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., and Calcott, B. (2017).
- Partition of evolution for molecular Partition of evolution for molecular
- and morphological phylogenetic analyses. *Molecular Biology and Evolution* **34**, 772–773.
- 645 doi:10.1093/molbev/msw260
- 646 Lehmann, C. E. R., Archibald, S. A., Hoffmann, W. A., and Bond, W. J. (2011). Deciphering
- the distribution of the savanna biome. *New Phytologist*, 197–209.
- 648 Leite, Y. L. R., Costa, L. P., Loss, A. C., Rocha, R. G., Batalha-Filho, H., Bastos, A. C.,
- 649 Quaresma, V. S., Fagundes, V., Paresque, R., Passamani, M., and Pardini, R. (2016).
- Neotropical forest expansion during the last glacial period challenges refuge hypothesis.
- 651 Proceedings of the National Academy of Sciences of the United States of America 113, 1008–
- 652 1013. doi:10.1073/pnas.1513062113
- 653 Longo, D., Lorenz-Lemke, A. P., Mäder, G., Bonatto, S. L., and Freitas, L. B. (2014).
- Phylogeography of the *Petunia integrifolia* complex in southern Brazil. *Botanical Journal of*

- 655 *the Linnean Society* **174**, 199–213. doi:10.1111/boj.12115
- Martins, A. C., and Alves-dos-Santos, I. (2013). Floral-oil-producing Plantaginaceae species:
- 657 geographical distribution, pollinator rewards and interactions with oil-collecting bees. *Biota*
- 658 *Neotropica* **13**, 77–89.
- 659 Martins, A. C., Melo, G. A. R., and Renner, S. S. (2015). Gain and loss of specialization in
- two oil-bee lineages, Centris and Epicharis (Apidae). Evolution 69, 1835–1844.
- 661 doi:10.1111/evo.12689
- Martins, A. C., Scherz, M. D., and Renner, S. S. (2014). Several origins of floral oil in the
- Angelonieae, a southern hemisphere disjunct clade of Plantaginaceae. *American Journal of*
- 664 Botany **101**, 2113–2120. doi:10.3732/ajb.1400470
- Matzke, N. J. (2013). BioGeoBEARS: BioGeography with Bayesian (and Likelihood)
- 666 Evolutionary Analysis in R Scripts. R package, version 0.2.1.
- Michener, C. D. (1963). Some Future Developments in Taxonomy. Systematic Zoology 12,
- 668 151–172.
- Michener, C. D. (2007). 'The Bees of the World' 2nd ed. (The Johns Hopkins University
- 670 Press: Baltimore, Maryland, USA.) doi:10.1017/CBO9781107415324.004
- 671 Michener, C. D., and Moure, J. S. (1957). A Study of the Classification of the More Primitive
- 672 non-parasitic Anthophorine Bees (Hymenoptera, Apoidea). Bulletin of the American Museum
- 673 *of Natural History* **112**, 395–452.
- 674 Miller, M. A., Pfeiffer, W., and Schwartz, T. (2010). Creating the CIPRES Science Gateway
- 675 for inference of large phylogenetic trees. In 'Gateway Computing Environments'. pp. 1–8.
- 676 (New Orleans, USA.)
- 677 Morrone, J. J. (2014). Biogeographical regionalisation of the Neotropical region. Zootaxa
- 678 **3782**, 1–110.
- Morrone, J. J. (2000). What is the Chacoan subregion? *Neotropica* **46**, 51–68.
- Moure, J. S. (1992). *Lissopedia*, gen.n. de Paratetrapediini para a Região Neotropical, com as
- descrições de três espécies no. Revista Brasileira de Zoologia 9, 305–317.
- 682 Ollerton, J., Winfree, R., and Tarrant, S. (2011). How many flowering plants are pollinated by
- 683 animals? *Oikos* **120**, 321–326. doi:10.1111/j.1600-0706.2010.18644.x
- 684 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., and
- Kassem, K. R. (2001). Terrestrial Ecoregions of the World: A New Map of Life on Earth.
- 688 BioScience **51**, 933–938. doi:10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2
- Ortiz-Jaureguizar, E., and Cladera, G. A. (2006). Paleoenvironmental evolution of southern
- 690 South America during the Cenozoic. *Journal of Arid Environments* **66**, 498–532.
- 691 doi:10.1016/j.jaridenv.2006.01.007
- 692 Otto-Bliesner, B. L., Marshall, S. J., Overpeck, J. T., Gifford, H., Miller, A. H., and CAPE
- 693 Last Interglacial Project Members (2006). Simulating Arctic Climate Warmth and Icefield

- Retreat in the Last Interglaciation. *Science* **311**, 1751–1753.
- 695 Overbeck, G. E., Boldrini, I. I., Carmo, M. R. B. do, Garcia, É. N., Moro, R. S., Pinto, C. E.,
- 696 Trevisan, R., and Zannin, A. (2015). Fisionomia dos Campos. In 'Os Campos do Sul'. (Eds V.
- 697 D. P. Pillar and O. Lange.) (Rede Campos Sulinos UFRGS: Porto Alegre.)
- 698 Overbeck, G. E., Müller, S. C., Fidelis, A., Pfadenhauer, J., Pillar, V. D., Blanco, C. C.,
- Boldrini, I. I., Both, R., and Forneck, E. D. (2007). Brazil's neglected biome: The South
- 700 Brazilian Campos. *Perspectives in Plant Ecology, Evolution and Systematics* **9**, 101–116.
- 701 doi:10.1016/j.ppees.2007.07.005
- Pennington, R. T., and Hughes, C. E. (2014). The remarkable congruence of New and Old
- 703 World savanna origins. *New Phytologist* **204**, 4–6. doi:10.1111/nph.12996
- Peres, E. A., Sobral-Souza, T., Perez, M. F., Bonatelli, I. A. S., Silva, D. P., Silva, M. J., and
- Solferini, V. N. (2015). Pleistocene Niche Stability and Lineage Diversification in the
- 706 Subtropical Spider *Araneus omnicolor* (Araneidae). *PLoS ONE* **10**, 1–20.
- 707 doi:10.1371/journal.pone.0121543
- 708 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., and Lincoln Etchebehere,
- M. (2009). The evolution of a tropical rainforest/grassland mosaic in southeastern Brazil since
- 711 28,000 14C yr BP based on carbon isotopes and pollen records. Quaternary Research 71,
- 712 437–452. doi:10.1016/j.yqres.2009.01.008
- Phillips, S. J., Dudík, M., and Schapire, R. E. (2004). A Maximum Entropy Approach to
- 714 Species Distribution Modeling. Proceeding of the Twenty-first International Conference on
- 715 *Machine Learning* **69**, 655–662. doi:10.1145/1015330.1015412
- 716 Pinheiro, F., De Barros, F., Palma-Silva, C., Fay, M. F., Lexer, C., and Cozzolino, S. (2011).
- 717 Phylogeography and genetic differentiation along the distributional range of the orchid
- 718 Epidendrum fulgens: a Neotropical coastal species not restricted to glacial refugia. Journal of
- 719 *Biogeography* **38**, 1923–1935. doi:10.1111/j.1365-2699.2011.02539.x
- Porzecanski, A. L., and Cracraft, J. (2005). Cladistic analysis of distributions and endemism
- 721 (CADE): Using raw distributions of birds to unravel the biogeography of the South American
- 722 aridlands. Journal of Biogeography **32**, 261–275. doi:10.1111/j.1365-2699.2004.01138.x
- 723 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., and Suchard, M. A. (2018). Posterior
- summarisation in Bayesian phylogenetics using Tracer 1.7. Systematic Biology 32, 1–3.
- 727 doi:10.1093/sysbio/syy032/4989127
- 728 Rambo, B. (1950). A porta de Torres. Anais Botânicos do Herbário Barbosa Rodrigues 2,
- 729 125–136.
- Ramos, K. S., and Melo, G. A. R. (2010). Taxonomic revision and phylogenetic relationships
- of the bee genus *Parapsaenythia* Friese (Hymenoptera, Apidae, Protandrenini), with
- 532 biogeographic inferences for the South American Chacoan Subregion. Systematic Entomology
- 733 **35**, 449–474. doi:10.1111/j.1365-3113.2009.00518.x

- Ree, R. H., and Sanmartín, I. (2018). Conceptual and statistical problems with the DEC+J
- model of founder-event speciation and its comparison with DEC via model selection. *Journal*
- 736 *of Biogeography* **45**, 741–749. doi:10.1111/jbi.13173
- 737 Ree, R. H., and Smith, S. A. (2008). Maximum likelihood inference of geographic range
- evolution by dispersal, local extinction, and cladogenesis. Systematic Biology 57, 4–14.
- 739 doi:10.1080/10635150701883881
- 740 Renner, S. S., and Schaefer, H. (2010). The evolution and loss of oil-offering flowers: new
- insights from dated phylogenies for angiosperms and bees. *Philosophical transactions of the*
- 742 Royal Society of London. Series B, Biological sciences **365**, 423–435.
- 743 doi:10.1098/rstb.2009.0229
- Roig-Alsina, A. (1999). Revisión de las abejas colectoras de aceites del género *Chalepogenus*
- Holmberg. Rev. Mus. Argentino Cienc. Nat. 1, 67–101.
- Ronquist, F. (1997). Dispersal-vicariance analysis: A new approach to the quantification of
- historical biogeography. Systematic Biology 46, 195–203. doi:10.1093/sysbio/46.1.195
- 748 Ronquist, F., Teslenko, M., Mark, P. van der, Ayres, D. L., Darling, A., Höhna, S., Larget, B.,
- Liu, L., Suchard, M. A., and Huelsenbeck, J. P. (2012). MrBayes 3.2□: Efficient Bayesian
- 750 Phylogenetic Inference and Model Choice Across a Large Model Space. Systematic Biology
- 751 **61**, 539–542. doi:10.1093/sysbio/sys029
- 752 Rozen, J. G., Melo, G. A. R., Aguiar, A. J. C., and Alves-Dos-Santos, I. (2006). Nesting
- 753 Biologies and Immature Stages of the Tapinotaspidine Bee Genera *Monoeca* and
- 754 Lanthanomelissa and of Their Osirine Cleptoparasites Protosiris and Parepeolus
- 755 (Hymenoptera: Apidae: Apinae). American Museum Novitates 3501, 1. doi:10.1206/0003-
- 756 0082(2006)501[0001:NBAISO]2.0.CO;2
- 757 RStudio Team (2016). RStudio: Integrated Development for R.
- 758 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.
- 760 doi:doi:10.1046/j.1365-2699.1999.00313.x
- Safford, H. D. (2007). Brazilian Páramos IV. Phytogeography of the Campos de Altitude.
- 762 Journal of Biogeography **34**, 1701–1722. doi:10.1111/j.1365-2699.2007.01732.x
- 763 Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., and Williamson, R. C. (2001).
- 764 Estimating the Support of a High-Dimensional Distribution. *Neural Computation* **13**, 1443–
- 765 1471. doi:10.1162/089976601750264965
- Silva, G. A. R., Antonelli, A., Lendel, A., Moraes, E. de M., and Manfrin, M. H. (2018). The
- 767 impact of early Quaternary climate change on the diversification and population dynamics of
- a South American cactus species. *Journal of Biogeography* **45**, 76–88. doi:10.1111/jbi.13107
- 769 Silvestro, D., and Michalak, I. (2012). RaxmlGUI: A graphical front-end for RAxML.
- 770 *Organisms Diversity and Evolution* **12**, 335–337. doi:10.1007/s13127-011-0056-0
- 771 Simon, M. F., Grether, R., de Queiroz, L. P., Skema, C., Pennington, R. T., and Hughes, C. E.
- 772 (2009). Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by in situ
- evolution of adaptations to fire. *Proceedings of the National Academy of Sciences* **106**,
- 774 20359–20364. doi:10.1073/pnas.0903410106

- Soberón, J. (2007). Grinnellian and Eltonian niches and geographic distributions of species.
- 776 Ecology Letters **10**, 1115–1123. doi:10.1111/j.1461-0248.2007.01107.x
- 777 Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses
- with thousands of taxa and mixed models. *Bioinformatics Applications* **22**, 2688–2690.
- 779 doi:10.1093/bioinformatics/btl446
- 780 Strömberg, C. A. E. (2011). Evolution of Grasses and Grassland Ecosystems. *Annual Review*
- 781 *of Earth and Planetary Sciences* **39**, 517–544. doi:10.1146/annurev-earth-040809-152402
- Tax, D. M. J., and Duin, R. P. W. (2004). Support Vector Data Description. *Machine*
- 783 *Learning* **54**, 45–66. doi:10.1023/B:MACH.0000008084.60811.49
- 784 Urban, D. (1995). Espécies novas de Lanthanomelissa Holmberg e Lanthanella Michener &
- Moure (Hymenoptera, Anthophoridae, Exomalopsinae). Revista Brasileira de Zoologia 12,
- 786 767–777.

- Vamosi, J. C., and Vamosi, S. M. (2010). Key innovations within a geographical context in
- 788 flowering plants: towards resolving Darwin's abominable mystery. *Ecology Letters* 13, 1270–
- 789 1279. doi:10.1111/j.1461-0248.2010.01521.x
- 790 Vogel, S. (1974). 'Ölblumen und ölsammelnde Bienen'. (Akademie der Wissenschaften und
- 791 der Literatur, Tropische und Subtropische Pflanzenwelt: Wiesbaden, Germany.)
- 792 Werneck, F. P. (2011). The diversification of eastern South American open vegetation
- 593 biomes: Historical biogeography and perspectives. Quaternary Science Reviews 30, 1630–
- 794 1648. doi:10.1016/j.quascirev.2011.03.009
- 795 Zachos, J., Pagani, M., Sloan, L., Thomas, E., and Billups, K. (2001). Trends, Global
- Rhythms, Aberrations in Global Climate 65Ma to Present. *Science* **292**, 686–693.
- 797 doi:10.1126/science.1059412
- 798 Zanella, F. C. V. (2002). Systematics and Biogeography of the bee genus *Caenonomada*
- 799 Ashmead, 1899 (Hymenoptera: Apidae: Tapinotaspidini). Studies on Neotropical Fauna and
- 800 Environment 37, 249–261. doi:10.1076/snfe.37.3.249.8571











