

1 AGE, ORIGIN, AND BIOGEOGRAPHY: UNVEILING THE FACTORS 2 BEHIND THE DIVERSIFICATION OF DUNG BEETLES

3 Orlando Schwery^{1,2*} and Brian C. O'Meara²

4 ¹ New Mexico Consortium, Los Alamos, NM, USA

5 ² Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN,
6 USA

7 * Corresponding author: oschwery@vols.utk.edu

8 **Keywords:** dung beetles, Scarabaeinae, diversification, speciation, extinction, macroevolution,
9 phylogenetics, biogeography

10

11 **Abstract**

12 The remarkable diversity and global distribution of dung beetles has long attracted the interest of
13 researchers. However, there is still an ongoing debate on their origin, the reasons behind their
14 diversity, and their path to global distribution. The two most prominent hypotheses regarding
15 their origin and biogeographic history involve either vicariance events after the breakup of
16 Gondwana, or an African origin and subsequent dispersal. One of the key reasons why the
17 question is still disputed is a dependence on knowing the age of the dung beetles – a Mesozoic
18 origin would favor the scenario of Gondwanan vicariance, a Cenozoic origin would suggest the
19 out-of-Africa scenario. To help settle this longstanding question, we provide a taxonomically
20 expanded phylogeny, with divergence times estimated under two calibration schemes suggesting
21 an older or younger origin respectively. Using model-based inference, we estimate the ancestral
22 area of the group and test for the influence of ranges on diversification rates. Our results support
23 the hypothesis of an old age for Scarabaeinae and Gondwanan origin but remain ambiguous
24 about the exact relation of range on lineage diversification.

25

26 **Introduction**

27 Dung beetles are a remarkable group of insects. Their unusual lifestyle requiring the dung of
28 other animals to feed and reproduce gave rise to a host of morphological and behavioral
29 specializations, as adaptations to the various ecological peculiarities they face in their worldwide
30 distribution (Hanski & Cambefort 1991). While their diversity of around 5,300 species is
31 comparatively modest among beetles, their dung-processing activity makes them one of the most
32 important groups of insects, both ecologically (Nichols *et al.* 2008) and economically (Losey &
33 Vaughan 2006). Their unusual life history has attracted considerable interest of researchers, in
34 ecology and evolution (Hanski & Cambefort 1991; Scholtz, Davis & Kryger 2009), conservation
35 (Spector 2006), and even developmental biology (Moczek 2011). Despite that, key aspects
36 concerning their origin and the factors behind their diversity are still unknown and subject of
37 debate in the field. The ecology of extant species has been well studied (Hanski & Cambefort
38 1991), but the main hindrance to understanding historical, evolutionary aspects of their biology
39 is the lack of a comprehensive and reliable phylogeny (Tarasov & Génier 2015), and with that,
40 validation for their taxonomy (Tarasov & Dimitrov 2016).

41 Probably the two most debated questions of dung beetle evolution revolve around their
42 geographic origin and what led to their current diversity and distribution. A relation to their
43 associated dung producers is suspected (Scholtz, Davis & Kryger 2009; Gunter *et al.* 2016), but
44 disagreement exists over whether the success of dung beetles has always been an association
45 with mammals, or whether dinosaurs were involved early in their history. Regarding their
46 geographic origin and subsequent spread, the main competing hypotheses are whether it was
47 Gondwanan-vicariance (Davis, Scholtz & Philips 2002) or dispersal out of Africa (Sole &
48 Scholtz 2010). The answers to all these hypotheses partially hinge on the question of how old the
49 group is. It has long been debated whether Scarabaeinae are of Mesozoic or Cenozoic origin –
50 the former would make Gondwanan vicariance and feeding on dinosaur dung plausible, the latter
51 would rule it out. Earlier attempts to determine the age of dung beetles using various approaches
52 led to widely different estimates, ranging from mid Mesozoic to early Cenozoic (Hanski &
53 Cambefort 1991; Scholtz & Chown 1995; Krell 2000; Davis, Scholtz & Philips 2002; Krell
54 2006). More recent attempts using phylogenies also ranged from Cretaceous to Eocene-

55 Oligocene (Wirta, Orsini & Hanski 2008; Ahrens, Schwarzer & Vogler 2014; Mlambo, Sole &
56 Scholtz 2015; Gunter *et al.* 2016), shifting the support for early and late origins of Scarabaeinae
57 over time (Scholtz, Davis & Kryger 2009). At this time, the field is still divided: an old Mesozoic
58 origin of dung beetles, and with that a biogeographical scenario of Gondwanan vicariance and
59 subsequent dispersal has been supported by some studies (Davis, Scholtz & Philips 2002; Gunter
60 *et al.* 2016; Gunter *et al.* 2018), whereas a young Cenozoic origin and therefore an out-of-Africa
61 dispersal scenario has been supported by others (Monaghan *et al.* 2007; Sole & Scholtz 2010;
62 Davis, Scholtz & Sole 2016).

63 To address these questions, we provide an extended phylogeny of Scarabaeinae, with divergence
64 times estimates based on calibrations reflecting two different assumptions of their maximal age.
65 We use this phylogeny and the inferred age of the group to address the question of their
66 geographical origin using model-based methods for ancestral range estimation; and the question
67 of how it relates to their lineage diversification using range-dependent diversification rate
68 estimation. In particular, focusing on their Gondwanan origin and subsequent dispersals to areas
69 of (what was formerly) Laurasia and Madagascar, we address the question of whether the
70 dispersal into new areas was promoting dispersal, or whether their place of origin is the main
71 source of diversification. Finally, we test a hypothesis that links dung beetle diversity to the dung
72 producers present in different regions (Davis & Scholtz 2001; Scholtz, Davis & Kryger 2009).
73 Specifically, linking beetle diversity to the size of mammal dung producers, and the diversity of
74 different dung types they produced, suggesting that areas with larger mammals and more diverse
75 droppings would allow for a more diverse dung beetle fauna.

76 **Materials and Methods**

77 *Phylogenetic Analysis*

78 We wrote a script in R (R Development Core Team 2014) using the packages *reutils* (Schöfl
79 2016) and *seqinr* (Charif *et al.* 2005) to query GenBank (Sayers *et al.* 2019) for any nucleotide
80 sequences matching the organism label “Scarabaeinae” and that carried either “COI”, “16S”,
81 “18S”, “28S”, or “CAD” in the title. We downloaded the resulting accessions using the packages
82 *ape* (Paradis, Claude & Strimmer 2004) and *insect* (Wilkinson *et al.* 2018), removed any

83 duplicates and saved them as FASTA files. The 28S accessions included both 28SD2 and 28SD3
84 sequences, which were sorted and aligned separately. The retrieved COI sequences largely
85 covered two adjacent regions of the gene, and since only few accessions covered both regions,
86 we decided to split them into two separate alignments (further called COI-1 and COI-2
87 respectively). If a marker had taxa with several accessions, the longest sequence was chosen
88 unless it proved to be clearly different from the other accessions for that taxon. Taxa which were
89 only determined to genus level were only included if they were the only representative of that
90 genus.

91 The accessions of the seven markers were aligned separately using AliView v.1.26 (Larsson
92 2014): They were aligned using MAFFT globalpair, and then visually inspected and adjusted
93 manually where necessary. During this process, any sequences that showed considerable
94 mismatch with the rest of the alignment were submitted to BLAST (Altschul *et al.* 1997) to
95 detect any mislabeled sequences from other organismal groups. Similarly, quick RAxML
96 (Stamatakis 2014) runs were performed for each aligned marker separately, and were tested for
97 generic monophyly and long branches. Generic monophyly was tested using the package
98 *MonoPhy* (Schwery & O’Meara 2016). Long branches were determined using the package *ips*;
99 terminal branches were considered exceptionally long if their length was more than 0.25 times
100 the maximum tip height in the tree, or if the product of their length and height was more than two
101 times the interquartile range away from the third quartile of all tip length and tip height products
102 in the tree. Any taxa that stood out by branch length or placement were checked using BLAST as
103 well. The accession numbers of the sequences used in the final alignment can be found in Table
104 S1.

105 Poorly aligned or divergent regions in each alignment were removed using Gblocks (Castresana
106 2000) (settings: minimum bases for conserved regions >0.5x alignment length, minimum for
107 flanking regions >0.7x alignment length, maximum contiguous nonconserved sites: 8, minimum
108 block length: 4 (noncoding) or 5 (coding), gap positions allowed with > half of sequences having
109 gap). Finally, all alignments were concatenated using the package *evobiR* (Blackmon & Adams
110 2015). Partitioning the alignment by each marker (the two marker parts in case of COI), we used
111 PartitionFinder v.2.1.1 (Lanfear *et al.* 2017) to determine the best substitution models for each

112 partition, using the greedy algorithm (Lanfear *et al.* 2012), and PhyML (Guindon *et al.* 2010).
113 Additionally, the package *ClockstaR* (Duchene, Molak & Ho 2014) was used to determine
114 whether the partitions should have different clock models.

115 Using the concatenated alignment and the determined substitution models, a maximum
116 likelihood phylogeny was constructed with RAxML v.8.2.4 (Stamatakis 2014), using the rapid
117 hill climbing algorithm with 100 rapid bootstrap samples. The ingroup and four clades that
118 would later be used for fossil calibrations (see below) were constrained to be monophyletic.
119 Because of branch support issues, we used the online implementation of RogueNaRok (Aberer,
120 Krompass & Stamatakis 2013) to find potential rogue taxa. Taxa whose exclusion would lead to
121 a raw improvement of more than 1 were inspected and 136 sequences were eventually dropped
122 and a final RAxML tree was built from this reduced alignment. To account for weakly supported
123 clade relationships in subsequent analyses, ten sets of bootstrapped alignments were created, for
124 each of which a separate tree was built in the same manner as from the full alignment.
125 Monophyly on genus and tribe level was assessed for all trees using the R package *MonoPhy*
126 (Schwery & O'Meara 2016).

127 ***Divergence Time Estimation***

128 Reliable Scarabaeinae fossils that have recently been re-examined by Tarasov *et al.* (2016) were
129 chosen to calibrate tree nodes during divergence time estimation. Of the 35 fossils previously
130 assigned to Scarabaeinae, they considered only 21 to be assigned reliably on the basis of their
131 morphological characters. From among these, the earliest fossils with confident generic
132 placement were chosen for each genus that we could assign a node to. The minimal age of the
133 fossils was used as minimum constraint for the stem age of the corresponding clade. The oldest
134 reliable fossil dung beetle, *Lobateuchus parisii*, was used to constrain the subfamily, and the
135 minimal age estimate for *Juraclopus rohendorfi*, the oldest fossil of the family Scarabaeidae, was
136 used as the minimal age constraint for the crown of the whole tree.

137 In order to get times to constrain the maximal age of these clades, we consulted previous studies
138 that had estimated the age of Scarabaeinae. On the younger end of the spectrum, some studies
139 refer to Wirta, Orsini and Hanski (2008), or Mlambo, Sole and Scholtz (2015) for young

140 estimates of the age of dung beetles (33.9ma or 56ma respectively). However, neither of those
141 estimates are particularly useful to calibrate the age of the whole group, as Wirta, Orsini and
142 Hanski (2008) focused on Helictopleurini of Madagascar, and Mlambo, Sole and Scholtz (2015)
143 exclusively on African dung beetles. While some more dated phylogenies of clades within the
144 subfamily exist, the only examples that include an age for the whole group are of higher taxa in
145 which the dung beetles are nested. Ahrens, Schwarzer and Vogler (2014) constructed a tree of
146 146 taxa of Scarabaeoidea, in which the crown Scarabaeinae was estimated at 89.6ma (ranging
147 from 83.5-98.1ma), while Gunter *et al.* (2016) in a phylogeny of 445 taxa of Scarabaeoidea
148 estimated it to range from 118.8-131.6ma. While various studies estimated the ages within
149 Scarabaeoidea (Ahrens, Schwarzer & Vogler 2014; McKenna *et al.* 2015; Toussaint *et al.* 2017),
150 the ages of Scarabaeinae and Scarabaeidae could not always be obtained from them –
151 consequently, we relied solely on Gunter *et al.* (2016) for an age estimate of Scarabaeidae
152 (116.85-199.64ma).

153 Because no clade can be older than the clade in which it is nested, we constrained the maximal
154 stem ages for all constrained genera (as well as the maximal crown age of all Scarabaeinae) to be
155 the maximal estimated age of the Scarabaeinae, and the maximal crown age of the whole tree
156 (being the stem age of Scarabaeinae) to be the maximal estimated age of Scarabaeidae. Given the
157 disagreement of estimated ages in the literature, we set up two different sets of constraints: an
158 ‘old’ one, with the Scarabaeinae and Scarabaeidae ages as estimated by Gunter *et al.* (2016), and
159 a ‘young’ one, with the Scarabaeinae and Scarabaeidae ages as estimated by Ahrens, Schwarzer
160 and Vogler (2014) – with the latter actually being the age of Scarabaeoidea, due to the lack of
161 available ages for the actual Scarabaeidae (apart from that by Gunter *et al.* (2016)). While
162 McKenna *et al.* (2015) actually estimated a younger age for Scarabaeoidea, their estimate is
163 younger than the age of the oldest fossil of that group, which is why we chose to use the next
164 older estimate by Ahrens, Schwarzer and Vogler (2014) ranging from 167.2-181.8ma. The
165 different fossil constraints used can be seen in Table 2.

166 Those two sets of age constraints were then used to estimate divergence times for the ML tree
167 obtained through RAXML and all ten bootstrap trees, using penalized likelihood in treePL (Smith

168 & O'Meara 2012). We performed a thorough search with random cross validation, which was
169 preceded by a preliminary run to estimate the best tuning parameters and smoothing factor.

170 ***Occurrence Data and Ancestral Range Estimation***

171 We used the package *rgbif* (Chamberlain *et al.* 2016; Chamberlain & Boettiger 2017) to
172 download occurrence data from GBIF (<http://data.gbif.org>) for all taxa determined to species
173 level. Using an R script, the retrieved records were cleaned from empty or invalid entries, with
174 regards to the basis of record, identification date, country and coordinates, and sets with unique
175 countries or coordinates per taxon respectively were made. Name validity of unavailable taxa
176 was checked using the package *taxize* (Chamberlain & Szocs 2013), and eventually searched on
177 genus level. For any taxa that were still missing, occurrence information was searched for in the
178 literature.

179 We defined the areas taxa occur in as continents (Africa, Asia, Europe, North America, South
180 America, Oceania), and defined Madagascar as a separate area, given the high number of taxa
181 endemic to it (Miraldo, Wirta & Hanski 2011). However, we did not define India as a separate
182 area, and considered it part of Asia. India has a peculiar tectonic history of initially staying
183 connected to Madagascar until breaking off around 87ma and drifting northwards to collide with
184 Asia around 55-43ma (Seton *et al.* 2012). Not explicitly including it as a separate area means a
185 major simplification for the model, but also ignoring potentially relevant possible scenarios. For
186 example, lineages that have entered India from Madagascar before their breakup, could have
187 been evolving in relative isolation there until coming into contact with Eurasia. However, we
188 suspect that testing that kind of scenario would require more detailed geographical resolution,
189 and thus may warrant a dedicated separate study.

190 Each taxon's range was defined as the one or several areas it occurred in, based on the collected
191 information from GBIF and the literature. Where information on continent was missing, it was
192 inferred from coordinates and countries of occurrence, and inconsistencies between these were
193 checked and corrected. Taxa occurring in many continents were inspected for the extent of
194 overlap and reliability of the records. In doing so, we also paid attention to species that were

195 introduced to areas by humans, and removed such occurrences from the species' range, thereby
196 only assigning its presumed natural range.

197 Given the potential age of the group, it is conceivable that tectonic plate movement played a role
198 in their dispersal and distribution. We thus constructed a stratified dispersal matrix representing
199 the changing strength of dispersal barriers between the continents over time. We divided the last
200 200ma before the present into five time slices, following the tectonic events described in Seton *et al.*
201 *al.* (2012), stretching from 200-150ma (Pangaea intact), 150-110ma (breakup of Pangaea into
202 Gondwana and Laurasia, Madagascar breaking off of Africa, though still connected via
203 Antarctica), 110-50ma (breakup of Gondwana into Africa, Australia-Antarctica, and South
204 America, the latter still connected to Antarctica via a land bridge), 50-20ma (Australia separates
205 from Antarctica, Laurasia breakup into Laurentia and Eurasia (Hosner, Braun & Kimball 2015),
206 South America properly disconnected from Antarctica (Reguero *et al.* 2014)), and 20-0ma (land
207 bridges (some temporary) establish at Beringia (Hosner, Braun & Kimball 2015), the Isthmus of
208 Panama, and between Australia and Asia, and Eurasia and Africa).

209 Inspired by Toussaint, Bloom and Short (2017), we recognized dispersal barriers of different
210 strengths: 1) directly adjacent areas (barrier of strength 0), 2) areas connected by land bridge
211 (0.15), 3) areas separated by a small distance of water (0.25), and 4) by a large distance of water
212 (0.75). Having to pass through another area was considered a barrier of strength 0.5, and the case
213 of facing more than 3 barriers was assigned a strength of 0.95; the corresponding dispersal
214 multiplier was 1 - strength of barriers. The route of smallest resistance was picked for each
215 possible dispersal case, adding up the different barriers encountered (matrices see Table S4).

216 While it would make sense that different dispersal barriers at different times would affect the
217 dispersal dynamics of the group, it is also possible that only time differences or only dispersal
218 barriers did. Furthermore, the partially arbitrary choice of time intervals could affect the
219 inference as well. Thus, we also calculated averaged dispersal matrices to be used without time
220 stratification. For this purpose, the dispersal multipliers of each time slice were weighted by the
221 duration of that time slice relative to the total time from the beginning of that slice until the
222 present. These weighted multipliers were added up for each transition and then divided by the
223 sum of weights, so they would add to one. This weighting is intended to represent the fact that

224 the more recent positioning of the continental plates should have had a higher impact on the
225 current distribution of extant taxa. Two matrices (for 3 and 4 time slices respectively, see Table
226 S5) were constructed that way, to be used for either the phylogeny dated with the younger or
227 older calibration times.

228 We used the package *BioGeoBEARS* (Matzke 2013b; Matzke 2013a) to estimate the ancestral
229 ranges of the dung beetles. The package implements some of the most popular models for
230 ancestral range estimation, DEC (Ree & Smith 2008), DIVA (Ronquist 1997), and BayArea
231 (Landis *et al.* 2013), in the same framework, allowing them to easily be compared against each
232 other to test different biogeographical hypotheses. While the DEC model is a maximum
233 likelihood implementation just as originally implemented in Lagrange, the DIVA model was
234 originally implemented using parsimony, and in *BioGeoBEARS* the processes DIVA assumes are
235 modeled under maximum likelihood. Similarly, BayArea was originally implemented as
236 Bayesian, and is represented in *BioGeoBEARS* as a maximum likelihood interpretation of the
237 same. Thus, these two models should be referred to as DIVALIKE and BAYAREALIKE
238 respectively.

239 A popular feature of *BioGeoBEARS* is the addition of “jump-dispersal” to any of these models.
240 By adding the additional jump parameter (“+j”), one allows for founder events in the model,
241 meaning that at a speciation event, one descending lineage stays in the ancestral range, while the
242 other descendant jumps to a new area (Matzke 2014). However, there has recently been some
243 debate regarding the validity of the +j models (Ree & Sanmartín 2018; Klaus & Matzke 2019),
244 and this type of cladogenetic event is believed to be more important in island systems than in
245 non-island clades (Matzke 2013b). For those reasons, we decided not to employ the j parameter.

246 However, since the dispersal multiplier matrices defined above are somewhat arbitrary, we do
247 pair them with a parameter (w) that scales the matrix and that can be optimized as well,
248 constituting the +w models (Dupin *et al.* 2017). The parameter w is used to exponentiate the
249 dispersal multiplier before it is multiplied by the dispersal rate and is 1 by default. Thus, if the
250 multipliers play an important role in the biogeographical history of the group, w is likely to be
251 estimated to be larger than 1, whereas it should be estimated to be lower than 1 if the dispersal
252 matrix does not add much to explain the patterns. While this does not allow to modulate the

253 relative strengths of the different multipliers, it seems reasonable to expect that w would be
254 estimated to downweigh the importance of a grossly unrealistic set of multipliers.

255 We employed each of the three base models (DEC, DIVALIKE, BAYAREALIKE) in four
256 different ways: 1) as a basic model (estimating d and e), 2) with the non-stratified manual
257 dispersal multiplier matrix and estimated w parameter, 3) as a time stratified model, and 4) as a
258 time stratified model with a manual dispersal multiplier matrix for each time slice, and estimated
259 w parameter. In all those analyses, we constrained the maximal range any lineage can occupy to
260 3, as none of the extant species occupy more than 3 areas. To account for the branch support
261 issues, we ran all these models both on the trees with old and young calibrations, as well as the
262 ten bootstrap trees each.

263 Finally, preliminary analyses produced curious results, particularly in the time stratified model,
264 where lineages within clades that were entirely present in one area (*e.g.* Madagascar) would
265 commonly disperse to a neighboring area (*e.g.* Africa) right after a speciation event, only to
266 return to the ancestral area again. It was presumed that clades in which every species inhabited
267 the same two areas (*e.g.* the Americas) would be responsible for this, as models as DEC do not
268 include the required scenario (cladogenesis in widespread taxa where both daughters inherit
269 widespread range). Thus, such cases required the above pattern of one daughter losing and re-
270 gaining the widespread range, thereby inflating the inferred dispersal rate and forcing other
271 lineages to do the same. We therefore created test data sets where we identified the area in which
272 species most commonly co-occurred (the Americas), or the two pairs of most co-inhabited areas
273 (the Americas and Eurasia), and coded those as one, thus getting a 6-area and 5-area dataset to be
274 analyzed separately.

275 *Diversification Analysis*

276 We tested two sets of biogeography-related hypotheses: whether areas with larger mammals and
277 more diverse droppings are associated with higher diversification rates of dung beetles, and
278 whether diversification rates were raised when dung beetles gained access to new habitats,
279 dispersing from Gondwana to Madagascar or Laurasia. The former hypothesis was derived from
280 Davis and Scholtz (2001); (also elaborated in Scholtz, Davis & Kryger 2009). They classify

281 mammalian dung into four types based on size and physico-chemical characteristics: 1) small dry
282 pellets from small to medium herbivores, 2) small odiferous droppings from omni- or carnivores,
283 3) large, dry, coarse-fibered droppings from large non-ruminant herbivores, and 4) large, moist,
284 fine-fibered pads from large ruminants. The number of those types of dung available, as well as
285 the (fairly correlated) body size of mammals was tallied for different biogeographical regions,
286 and shown to relate to aspects of dung beetle diversity (Davis & Scholtz 2001; Scholtz, Davis &
287 Kryger 2009).

288 Explicitly biogeographic diversification models such as GeoSSE (Goldberg, Lancaster & Ree
289 2011) or GeoHiSSE (Beaulieu & O'Meara 2016; Caetano, O'Meara & Beaulieu 2018) only test
290 diversification between two areas (and three ranges: endemic to one or the other area, or
291 widespread), making explicitly testing hypotheses involving more areas (not to speak of all
292 seven) impossible. However, in the case of only three areas, testing different combinations of
293 two against each other can still yield relevant insights. To this end, we re-coded the distribution
294 data from above to merge occurrences 1) of areas with large mammals and diverse droppings
295 (Afro-Eurasia: Africa, Europe, and Asia), of medium sized mammals and less diverse droppings
296 (North and South America), and those of those with small mammals and the least diverse
297 droppings (East Gondwanan Fragments: Oceania and Madagascar).; and 2) those of the areas
298 formerly making up Gondwana (South America, Africa, and Oceania), those of former Laurasia
299 (North America, Europe, and Asia), while keeping Madagascar as an area of its own. We then
300 formatted these two data sets to suit the different GeoSSE tests: a set where we join the areas
301 with medium sized mammals and intermediate droppings-diversity with the areas of large
302 mammals and diverse droppings, and one where we join them with the small mammal and low
303 droppings-diversity instead. Another set of where we join Madagascar with Gondwana, where
304 we join it with Laurasia, and where we leave it as an area separate from the rest.

305 For each of the resulting five data sets (and for both the young and old tree respectively), we
306 inferred maximum likelihood estimates with GeoSSE, under a set of different model constraints.
307 The constraints used were combinations of equal speciation between areas, equal extinction
308 between areas, equal dispersal rates between areas, and variation where speciation was set to
309 zero in one area and forced to be equal in the other and widespread lineages, and vice versa. The

310 full and constrained models were then compared using likelihood ratio tests and their Δ AIC
311 score. The best models for each combination of data set and tree were subsequently used to get
312 posterior distributions of the parameter estimates using MCMC, with an exponential prior related
313 to the Kendall-Moran estimate for net diversification rate (Kendall 1949; Moran 1951). A short
314 preliminary chain of 100 generations was run and the distances between the 5% and 95%
315 quantiles for each parameter were used to set the tuning parameter w for the slice sampler. Then,
316 each dataset was run for 20,000 generations. Convergence was assessed by the convergence
317 parameter of the function, by visual inspection of the log likelihood trace, and calculating
318 effective sample size using *effectiveSize* from the R package *coda* (Plummer *et al.* 2004). We
319 then compared the 95% quantiles of the posterior distribution for all estimated parameters to see
320 whether they overlap.

321 **Results**

322 *Phylogeny and Divergence Times*

323 Out of 122 genera represented in this phylogeny, 33 were monophyletic, 28 were non-
324 monophyletic, and 61 were monotypic (Table S2). When considering the bootstrap trees, 14
325 genera were consistently monophyletic, 22 consistently non-monophyletic, whereas the
326 remaining 25 varied between trees. On the tribal level, we recovered Eucraniini, Ateuchini,
327 Eurysterniini, Gymnopleurini, and Sisyphini as monophyletic (Table S3). Of these, only the last
328 three consistently so, with Onitini being monophyletic in some bootstrap replicates. Upon closer
329 inspection of the reasons for each tribe's non-monophyly (Table 1), we see that at least part of it
330 results from taxonomic issues. For Scarabaeini, Phanaeini, and Onitini, non-monophyly is due to
331 a few *incertae sedis* taxa, most of which monotypic. Oniticellini and Onthophagini have similar
332 issues where one intruder of the former is a member of the latter, but also the whole of
333 Oniticellini is nested within Onthophagini. Dichotomiini, Deltochilini and Canthonini are
334 scattered in clumps across the tree, with the latter two intermingling a lot, while Coprini come
335 out in three separate clades.

336 The estimated ages of the calibrated nodes for both the young and old calibration are given in
337 Table 2. It is apparent that under both calibration schemes, the crown age of the Scarabaeinae has

338 hit its constrained maximum age, while the stem ages of the constrained nodes are not reaching
339 their maximum constraints but are often meeting or exceeding the minimum age constraint of the
340 subfamily's crown. The bootstrap branch support values can be seen in Figure 1.

341 *Ancestral Range Estimation*

342 For all analyses of the 7-area data set on the young and old maximum likelihood trees and their
343 respective sets of 10 bootstrap trees, the analyses recovered the unstratified DEC model with
344 manual dispersal multipliers (DEC+w) as the best-fitting model. In all trees, w – the exponent for
345 the dispersal multipliers – was estimated to be larger than 1 (1.71-2.83), indicating its weight in
346 the dispersal process being higher than the initial dispersal matrix suggested. The estimated
347 dispersal and extirpation rates vary between the trees but are comparable and at the very least in
348 the same order of magnitude. For the 6-area and 5-area datasets, which model was inferred to fit
349 best varied widely among the different trees.

350 The ancestral range of the whole subfamily was estimated to be Africa, Oceania, and South
351 America for both the young and old tree. However, in the alternative trees based on bootstrap
352 replicates, the estimated ancestral range could include Madagascar instead of South America or
353 instead of Oceania, or could just include Africa and Oceania (Table S6). Interestingly,
354 reconstructions with root states other than Africa, Oceania, and South America tended to be more
355 ambiguous. Overall, the estimated ancestral ranges seemed to suggest that numerous clades
356 mostly stay in the same areas (particularly Africa, Oceania, and Madagascar), with some more
357 dispersal within the Americas and Eurasia, as well as within the clade comprising the
358 Onthophagini and Oniticellini (Figure S1).

359 *Diversification Analyses*

360 The best fitting models for each data set and tree, according to likelihood ratio tests and Δ AIC
361 scores, are given in Table 3. The MCMC analyses converged and yielded reasonably high
362 effective sampling sizes with no value below 295. The 95% quantiles of the posterior
363 distributions for each parameter estimate is given in Table 4.

364 Regardless of whether the areas with medium sized mammals and intermediate diversity in
365 droppings were joined with the areas of large or small mammals, there was no significant
366 difference between them in terms of diversification rates (if the best model did not constrain
367 them; Figure 2, Figure 3). When the areas of medium and small mammals are joined, the
368 analysis attributes any difference in diversity between them to higher dispersal out of Afro-
369 Eurasia into the other areas.

370 Despite some differences in which model fitted the Gondwanan origin data best, the results
371 between the old tree (Figure 4) and the young tree (Figure 5) are largely consistent. They show
372 that lineages in Gondwana have a significantly lower speciation rate than lineages outside of it,
373 while lineages in Laurasia disperse into the other areas at a higher rate than vice-versa, while
374 there are no significant differences in either diversification nor dispersal between Madagascar
375 and the other areas.

376 **Discussion**

377 *Phylogeny and Taxonomy*

378 With 541 represented in-group species, this is to date the most inclusive dated species-level
379 molecular phylogeny of Scarabaeinae. Most previously inferred phylogenies of the group were
380 either constrained to specific sub-clades or regions (e.g. Davis & Scholtz 2001; Wirta, Orsini &
381 Hanski 2008; Sole & Scholtz 2010; Wirta *et al.* 2010; Mlambo, Sole & Scholtz 2015;
382 Breeschoten *et al.* 2016; Gunter *et al.* 2018), or part of a larger phylogeny where the main focus
383 accordingly was not or not only on dung beetles (Ahrens, Schwarzer & Vogler 2014; Kim &
384 Farrell 2015; Gunter *et al.* 2016; Toussaint *et al.* 2017). A recent large un-dated phylogeny by
385 Tarasov and Dimitrov (2016) based on 8 gene regions had a similar amount of terminals (547
386 with outgroup), though it constitutes a smaller sample of actual species diversity, as many
387 species were represented by multiple accessions and many were not determined to species level.
388 The levels of generic and tribal monophyly in this new phylogeny (Table S2, Table S3) and the
389 causes of it (Table 1) would suggest a reasonable level of agreement between this tree and
390 current taxonomy. The lack of support for many groupings within the tree (Figure 1), particularly
391 in the backbone, is cause for concern, as it not only suggests shortcomings in the inferred tree,

392 but also casts doubt upon the reliability of analyses results derived from that tree. Other attempts
393 at phylogenies of the overall Scarabaeinae were plagued with similar patterns of low branch
394 support (Tarasov & Dimitrov 2016), which suggests a general issue in the study of this group.
395 Molecular phylogenies have challenged the traditional classification, particularly the tribal
396 monophyly of Canthonini and Dichotomiini (Monaghan *et al.* 2007), as well as Coprini,
397 Onthophagini and Oniticellini, while the monophyly of the remaining tribes still seems supported
398 (Scholtz, Davis & Kryger 2009). Tarasov and Dimitrov (2016) note how their own results are
399 consistent with those of previous phylogenetic studies of Scarabaeinae (Ocampo & Hawks 2006;
400 Monaghan *et al.* 2007; Vaz-de-Mello 2007; Wirta, Orsini & Hanski 2008; Sole & Scholtz 2010;
401 Wirta *et al.* 2010; Mlambo, Sole & Scholtz 2014; Gunter *et al.* 2016), as well as with a large
402 phylogeny based on morphology (Tarasov & Génier 2015). They furthermore observed that the
403 studies to date tend to resolve old nodes and more recent nodes, but not intermediate ones, and
404 that the same set of problematic tribes mentioned above are consistently not monophyletic. All of
405 this seems to be reflected in our results as well, particularly when considering the extent of
406 monophyly problems (Table 1). Using the suggested new classification by Tarasov and Dimitrov
407 (2016) does not yield much improvement. While Eucraniini and Eurysternini are strictly
408 monophyletic in both, monophyly gained through reclassification in Dichotomiini, Canthonini,
409 and Scarabaeini is offset by the loss of it in Ateuchini, Gymnopleurini, and Sisyphini. Both
410 classifications show similar consistency across the set of bootstrap trees.

411 ***Ancestral Range Estimation and the Origin of Scarabaeinae***

412 The two main hypotheses regarding where dung beetles originated are an origin in Gondwana
413 followed by vicariance events after the breakup of the supercontinent (Hanski & Cambefort
414 1991; Davis & Scholtz 2001; Davis, Scholtz & Philips 2002), and an origin in Africa and
415 subsequent dispersal out of it (Sole & Scholtz 2010). One major point of conflict between those
416 two ideas was the age of Scarabaeinae: Gondwanan vicariance would necessitate the group to be
417 of Mesozoic, rather than Cenozoic origin, as they would have to exist and be widespread enough
418 before the continental breakup (110ma according to Sanmartin and Ronquist (2004), 93ma
419 according to Scotese (1993)) in order for vicariance events to be plausible.

420 To answer the question of biogeographic origin in absence of an appropriate phylogeny, some
421 workers relied on classification (Hanski & Cambefort 1991; Davis & Scholtz 2001), considering
422 widespread tribes to be ancient and predating the Gondwana-breakup, in turn giving rise to
423 younger, less widespread tribes. Later attempts combined the relative age of phylogenies
424 (Monaghan *et al.* 2007; Wirta, Orsini & Hanski 2008) with fast and slow rates of molecular
425 sequence divergence in insects to get maximal and minimal divergence time estimates (Scholtz,
426 Davis & Kryger 2009), concluding that even the slowest known divergence rates would not
427 support the idea of a pre-Gondwanan-breakup origin of dung beetles. Sole and Scholtz (2010)
428 subsequently used a time calibrated phylogeny of the African representatives of Canthonini and
429 Dicotomiini to address the question, finding the divergence times between dung beetles and their
430 outgroup, and of the crown of dung beetles to be considerably younger than the breakup of
431 Gondwana (56ma and 40ma respectively), thus further supporting the later out-of-Africa
432 scenario.

433 As for the divergence times inferred in this study (Table 2), the older age of 131.6ma would
434 place the origin of dung beetles well before the complete breakup of Gondwana, whereas the
435 younger estimate of 98.1ma appears ambiguous, depending on when the actual separation of
436 Gondwana was completed, and depending on the accuracy of this age estimate. In either case
437 however, the inference of older origins, similar to other recent estimates that considered more
438 than just a subset of Scarabaeinae (Ahrens, Schwarzer & Vogler 2014; Gunter *et al.* 2016),
439 makes a Gondwanan origin and thus the potential for vicariance after its breakup seem like a
440 plausible option again.

441 The reconstruction of what is essentially Gondwana as the ancestral range at the crown of our
442 phylogenies seems to further support the idea of Gondwanan-vicariance. However, the DEC
443 model is known to have a bias towards widespread ancestors (Clark *et al.* 2008; Ree & Smith
444 2008; Buerki *et al.* 2011; Matzke 2014), even if not as strongly as DIVA (Kodandaramaiah
445 2010). Therefore, since we constrained the number of areas a lineage can maximally inhabit to
446 three, the inference of those three areas as the origin could possibly be an artefact. With regards
447 to the branch support issues, it would appear that the consistency with which the same model
448 was preferred across all trees using the 7-area dataset, and the relative stability of the inferred

449 origin to be Gondwana, or parts thereof (Table S6), could be seen as a sign that the result is
450 robust enough. However, the wildly varying best supported model under the reduced area
451 datasets (6 and 5 areas) is cause for concern. It could be argued, that leaving those pairs or areas
452 separate (North and South America, Europe and Asia respectively), might lead to a distribution
453 of areas across clades that is not broken up when they are rearranged in the bootstrap trees, thus
454 implying the same dispersal mechanisms. On the contrary, lumping them could lead to single-
455 distribution clades being broken up, thus changing the number of implied dispersal and
456 vicariance events. However, this is rather speculative and requires further investigation. Finally,
457 while the high estimates for w suggest that the specified dispersal multipliers are a relevant
458 improvement of the model overall, this does not guarantee that they are an accurate
459 representation of dung beetle dispersal probabilities at the given times. While they might capture
460 some large-scale dispersal constraints, the relative magnitude of the dispersal multipliers
461 between specific continents could still be inaccurate, *e.g.* because of the way the barriers were
462 specified and the values they were assigned. Sensitivity tests, or even adding an approach using
463 actual distance, could help to inform us about this.

464 In any case, given the branch support issues, we would refrain from a too detailed interpretation
465 of the ranges reconstructed at internal nodes. However, it is notable that the reconstruction
466 suggests that many clades can be found which seem to be predominantly confined to one (or
467 few) areas, particularly to Oceania, Madagascar, and Africa, with some more dispersal in
468 American or Eurasian clades. An exception seems to be the clade comprising Onthophagini and
469 Oniticellini, which seems to have much more area changes. This could indicate that the
470 Onthophagini and Oniticellini are different from the remaining dung beetles in their
471 biogeographic history and maybe their dispersal abilities. It could however also just reflect a lack
472 of sampling in that clade.

473 *Diversification Analyses*

474 The GeoSSE results for the hypothesis on whether mammal size and available dung-diversity in
475 different biogeographical areas affected the diversification rates of Scarabaeinae seem to suggest
476 that this is not the case. The increased dispersal from Afro-Eurasia into the other regions is
477 certainly interesting and could reflect different plausible dispersal events. But given that the test

478 was not set up to address these alternatives, we would suggest exercising care not to
479 overinterpret this pattern. When Davis and Scholtz (2001) reported on the patterns among
480 mammals, the diversity of their droppings, and dung beetle diversity in different areas of the
481 world, they noted that it was particularly related to tribal diversity (and generic diversity within
482 tribes) in those areas, but less correlated to genus or species diversity. They interpret this as a
483 sign that the influence of these patterns in mammal body size and droppings on dung beetle
484 diversification happened earlier in their evolutionary history, around the time when the tribes
485 split. While the idea that mammal dung availability could have been a crucial influence on beetle
486 diversification in the past appears very sensible, it would seem that the arguments presented in
487 this particular case are problematic. Firstly, it relates an extant pattern in mammals to past effects
488 in beetles assuming that the distribution of mammals and their droppings across the world was
489 comparable between then and now. However, those extant patterns were almost certainly not
490 constant over the timespan of dung beetle diversification, with the most recent event that
491 changed those patterns already being the Late Quaternary megafaunal extinctions (Stuart 2015).
492 Furthermore, assuming from the ages in this current tree (Table 2), as well as past estimates (e.g.
493 Gunter *et al.* 2016), the ages of many tribes may either be older than the rise of mammal
494 diversity, or would at least coincide with a time when mammal diversity would not be expected
495 to compare well to the extant one. Finally, considering the fact that the pattern does not correlate
496 well with species level diversity, the method we employed might also be suboptimal to test the
497 implied scenario of past influence, and an approach that allows the influence of mammal size and
498 droppings to be constrained to particular time-intervals might be more suitable.

499 The result of the Gondwanan origin GeoSSE analysis might initially seem surprising, as a large
500 portion of Scarabaeinae species diversity is found in areas that were formerly Gondwana
501 (Scholtz, Davis & Kryger 2009) and the former-Gondwanan lineages were not underrepresented
502 in comparison to the others (Gondwana: 320 taxa, Laurasia: 143 taxa, Madagascar: 114 taxa).
503 But considering the possible Gondwanan origin of the whole group, and the perceived inertia of
504 clades (diversifying in an area rather than dispersing more), one might suggest that the larger
505 absolute number of species in former Gondwanaland results from diversifying at a lower rate but
506 over longer timespans than the on average probably younger Laurasian or Malagasy lineages.
507 This would also explain the lack of rate difference between the latter and Laurasia and

508 Gondwana combined. The higher dispersal out of Laurasia was not expected, assuming the
509 general direction of dung beetle dispersal was outward from Gondwana. But in the light of
510 higher dispersal outside of Gondwana, this could reflect the dispersal of few lineages out of
511 Gondwana, where they diversified, and subsequently a few of those Laurasian taxa returned to
512 former Gondwana (*e.g.* in the case of taxa which returned from North America to South America
513 after closing of the isthmus). The lack of difference in dispersal rate between Madagascar and the
514 rest is plausible as well, knowing that most Malagasy dung beetles are part of one of few clades
515 that entered Madagascar and diversified there, with very few dispersals back (Miraldo *et al.*
516 2011; Sole *et al.* 2011). This is also reflected by the fact that the inferred rate between
517 Madagascar and the rest is comparatively low (Table 4). All in all, those results would support
518 the hypothesis that access to new areas was associated with a rise of diversification rates in dung
519 beetles. However, more fine-scale analyses are needed to confirm the implied scenarios behind
520 this. Also, while overall sampling frequency in this phylogeny was considered in this analysis, it
521 cannot be ruled out that some of these results are artefacts of uneven sampling between groups
522 within the subfamily. Until an even more complete tree of Scarabaeinae is available, finding
523 ways to correct for such sampling biases would be advisable.

524 **Acknowledgements**

525 We would like to express our thankfulness for helpful discussions and comments at various
526 stages of this project to Kimberly Sheldon, Nicole Gunter, Sergei Tarasov, Nick Matzke, Jim
527 Fordyce, Colin Sumrall, Dave Bapst, Michelle Lawing, Jeremy Beaulieu, Luna Sanchez-Reyes,
528 Veronica Brown, Marisol Sanchez, Todd Pierson, Hailee Korotkin, Claire Winfrey, Jess Dreyer,
529 and Maggie Mamantov.

530 Finally, we thank the Society of Systematic Biologists and the Systematics Association for their
531 financial support towards this project.

532

533 **References**

- 534 Aberer, A.J., Krompass, D. & Stamatakis, A. (2013) Pruning Rogue Taxa Improves Phylogenetic
535 Accuracy: An Efficient Algorithm and Webservice. *Systematic Biology*, **62**, 162-166.
- 536 Ahrens, D., Schwarzer, J. & Vogler, A.P. (2014) The evolution of scarab beetles tracks the
537 sequential rise of angiosperms and mammals. *Proceedings of the Royal Society B-*
538 *Biological Sciences*, **281**.
- 539 Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W. & Lipman, D.J.
540 (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search
541 programs. *Nucleic Acids Research*, **25**, 3389-3402.
- 542 Beaulieu, J.M. & O'Meara, B.C. (2016) Detecting Hidden Diversification Shifts in Models of
543 Trait-Dependent Speciation and Extinction. *Systematic Biology*, **65**, 583-601.
- 544 Blackmon, H. & Adams, R. (2015) EvobiR: tools for comparative analyses and teaching
545 evolutionary biology.
- 546 Breeschoten, T., Doorenweerd, C., Tarasov, S. & Vogler, A.P. (2016) Phylogenetics and
547 biogeography of the dung beetle genus *Onthophagus* inferred from mitochondrial
548 genomes. *Molecular Phylogenetics and Evolution*, **105**, 86-95.
- 549 Buerki, S., Forest, F., Alvarez, N., Nylander, J.A.A., Arrigo, N. & Sanmartin, I. (2011) An
550 evaluation of new parsimony-based versus parametric inference methods in
551 biogeography: a case study using the globally distributed plant family Sapindaceae.
552 *Journal of Biogeography*, **38**, 531-550.
- 553 Caetano, D.S., O'Meara, B.C. & Beaulieu, J.M. (2018) Hidden state models improve state-
554 dependent diversification approaches, including biogeographical models. *Evolution*, **72**,
555 2308-2324.
- 556 Castresana, J. (2000) Selection of conserved blocks from multiple alignments for their use in
557 phylogenetic analysis. *Molecular Biology and Evolution*, **17**, 540-552.
- 558 Chamberlain, S., Ram, K., Barve, V. & Mcglinn, D. (2016) rgbif: Interface to the Global
559 "Biodiversity" Information Facility "API". R package version 0.9. 8.
- 560 Chamberlain, S.A. & Boettiger, C. (2017) R Python, and Ruby clients for GBIF species
561 occurrence data. PeerJ Preprints.
- 562 Chamberlain, S.A. & Szocs, E. (2013) taxize: taxonomic search and retrieval in R.
563 *F1000Research*, **2**, 191-191.
- 564 Charif, D., Thioulouse, J., Lobry, J.R. & Perriere, G. (2005) Online synonymous codon usage
565 analyses with the ade4 and seqinR packages. *Bioinformatics*, **21**, 545-547.
- 566 Clark, J.R., Ree, R.H., Alfaro, M.E., King, M.G., Wagner, W.L. & Roalson, E.H. (2008) A
567 comparative study in ancestral range reconstruction methods: retracing the uncertain
568 histories of insular lineages. *Systematic Biology*, **57**, 693-707.

- 569 Davis, A.L., Scholtz, C.H. & Sole, C.L. (2016) Biogeographical and co-evolutionary origins of
570 scarabaeine dung beetles: Mesozoic vicariance versus Cenozoic dispersal and dinosaur
571 versus mammal dung. *Biological Journal of the Linnean Society*, **120**, 258-273.
- 572 Davis, A.L.V. & Scholtz, C.H. (2001) Historical vs. ecological factors influencing global
573 patterns of scarabaeine dung beetle diversity. *Diversity and Distributions*, **7**, 161-174.
- 574 Davis, A.L.V., Scholtz, C.H. & Philips, T.K. (2002) Historical biogeography of scarabaeine
575 dung beetles. *Journal of Biogeography*, **29**, 1217-1256.
- 576 Duchene, S., Molak, M. & Ho, S.Y.W. (2014) ClockstaR: choosing the number of relaxed-clock
577 models in molecular phylogenetic analysis. *Bioinformatics*, **30**, 1017-1019.
- 578 Dupin, J., Matzke, N.J., Särkinen, T., Knapp, S., Olmstead, R.G., Bohs, L. & Smith, S.D. (2017)
579 Bayesian estimation of the global biogeographical history of the Solanaceae. *Journal of*
580 *Biogeography*, **44**, 887-899.
- 581 Goldberg, E.E., Lancaster, L.T. & Ree, R.H. (2011) Phylogenetic Inference of Reciprocal Effects
582 between Geographic Range Evolution and Diversification. *Systematic Biology*, **60**, 451-
583 465.
- 584 Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010) New
585 Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the
586 Performance of PhyML 3.0. *Systematic Biology*, **59**, 307-321.
- 587 Gunter, N.L., Monteith, G.B., Cameron, S.L. & Weir, T.A. (2018) Evidence from Australian
588 mesic zone dung beetles supports their Gondwanan origin and Mesozoic diversification
589 of the Scarabaeinae. *Insect Systematics & Evolution*, **1**, 1-27.
- 590 Gunter, N.L., Weir, T.A., Slipinksi, A., Bocak, L. & Cameron, S.L. (2016) If Dung Beetles
591 (Scarabaeidae: Scarabaeinae) Arose in Association with Dinosaurs, Did They Also Suffer
592 a Mass Co-Extinction at the K-Pg Boundary? *Plos One*, **11**.
- 593 Hanski, I. & Cambefort, Y. (1991) Dung beetle ecology. *Dung beetle ecology.*, i-xii, 1-481.
- 594 Hosner, P.A., Braun, E.L. & Kimball, R.T. (2015) Land connectivity changes and global cooling
595 shaped the colonization history and diversification of New World quail (Aves:
596 Galliformes: Odontophoridae). *Journal of Biogeography*, **42**, 1883-1895.
- 597 Kendall, D.G. (1949) Stochastic processes and population growth. *Journal of the Royal*
598 *Statistical Society. Series B (Methodological)*, **11**, 230-282.
- 599 Kim, S.I. & Farrell, B.D. (2015) Phylogeny of world stag beetles (Coleoptera: Lucanidae)
600 reveals a Gondwanan origin of Darwin's stag beetle. *Molecular Phylogenetics and*
601 *Evolution*, **86**, 35-48.
- 602 Klaus, K.V. & Matzke, N.J. (2019) Statistical Comparison of Trait-dependent Biogeographical
603 Models indicates that Podocarpaceae Dispersal is influenced by both Seed Cone Traits
604 and Geographical Distance. *Systematic Biology*.
- 605 Kodandaramaiah, U. (2010) Use of dispersal–vicariance analysis in biogeography—a critique.
606 *Journal of Biogeography*, **37**, 3-11.

- 607 Krell, F.-T. (2000) The fossil record of Mesozoic and Tertiary Scarabaeoidea (Coleoptera:
608 Polyphaga). *Invertebrate Systematics*, **14**, 871-905.
- 609 Krell, F.-T. (2006) Fossil record and evolution of Scarabaeoidea (Coleoptera: Polyphaga). *The
610 Coleopterists Bulletin*, **60**, 120-143.
- 611 Landis, M.J., Matzke, N.J., Moore, B.R. & Huelsenbeck, J.P. (2013) Bayesian Analysis of
612 Biogeography when the Number of Areas is Large. *Systematic Biology*, **62**, 789-804.
- 613 Lanfear, R., Calcott, B., Ho, S.Y.W. & Guindon, S. (2012) PartitionFinder: Combined Selection
614 of Partitioning Schemes and Substitution Models for Phylogenetic Analyses. *Molecular
615 Biology and Evolution*, **29**, 1695-1701.
- 616 Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T. & Calcott, B. (2017) PartitionFinder 2:
617 New Methods for Selecting Partitioned Models of Evolution for Molecular and
618 Morphological Phylogenetic Analyses. *Molecular Biology and Evolution*, **34**, 772-773.
- 619 Larsson, A. (2014) AliView: a fast and lightweight alignment viewer and editor for large
620 datasets. *Bioinformatics*, **30**, 3276-3278.
- 621 Losey, J.E. & Vaughan, M. (2006) The economic value of ecological services provided by
622 insects. *Bioscience*, **56**, 311-323.
- 623 Matzke, N.J. (2013a) BioGeoBEARS: BioGeography with Bayesian (and likelihood)
624 evolutionary analysis in R Scripts. *R package, version 0.2*, **1**, 2013.
- 625 Matzke, N.J. (2013b) Probabilistic Historical Biogeography: New Models for Founder-Event
626 Speciation, Imperfect Detection, and Fossils Allow Improved Accuracy and Model-
627 Testing. *Frontiers of Biogeography*, **5**, 242-248.
- 628 Matzke, N.J. (2014) Model Selection in Historical Biogeography Reveals that Founder-Event
629 Speciation Is a Crucial Process in Island Clades. *Systematic Biology*, **63**, 951-970.
- 630 McKenna, D.D., Farrell, B.D., Caterino, M.S., Farnum, C.W., Hawks, D.C., Maddison, D.R.,
631 Seago, A.E., Short, A.E.Z., Newton, A.F. & Thayer, M.K. (2015) Phylogeny and
632 evolution of Staphyliniformia and Scarabaeiformia: forest litter as a stepping stone for
633 diversification of nonphytophagous beetles. *Systematic Entomology*, **40**, 35-60.
- 634 Miraldo, A., Hewitt, G.M., Paulo, O.S. & Emerson, B.C. (2011) Phylogeography and
635 demographic history of *Lacerta lepida* in the Iberian Peninsula: multiple refugia, range
636 expansions and secondary contact zones. *Bmc Evolutionary Biology*, **11**.
- 637 Miraldo, A., Wirta, H. & Hanski, I. (2011) Origin and diversification of dung beetles in
638 Madagascar. *Insects*, **2**, 112-127.
- 639 Mlambo, S., Sole, C.L. & Scholtz, C.H. (2014) Affinities of the Canthonini dung beetles of the
640 Eastern Arc Mountains. *Organisms Diversity & Evolution*, **14**, 115-120.
- 641 Mlambo, S., Sole, C.L. & Scholtz, C.H. (2015) A molecular phylogeny of the African
642 Scarabaeinae (Coleoptera: Scarabaeidae). *Arthropod Systematics & Phylogeny*, **73**, 303-
643 321.

- 644 Moczek, A. (2011) Evolution and development: onthophagus beetles and the evolutionary
645 developmental genetics of innovation, allometry and plasticity. *Ecology and evolution of*
646 *dung beetles*, 126-151.
- 647 Monaghan, M.T., Inward, D.J.G., Hunt, T. & Vogler, A.P. (2007) A molecular phylogenetic
648 analysis of the Scarabaeinae (dung beetles). *Molecular Phylogenetics and Evolution*, **45**,
649 674-692.
- 650 Moran, P. (1951) Estimation methods for evolutive processes. *Journal of the Royal Statistical*
651 *Society: Series B (Methodological)*, **13**, 141-146.
- 652 Nichols, E., Spector, S., Louzada, J., Larsen, T., Amequita, S., Favila, M.E. & Scarabaeinae Res,
653 N. (2008) Ecological functions and ecosystem services provided by Scarabaeinae dung
654 beetles. *Biological Conservation*, **141**, 1461-1474.
- 655 Ocampo, F.C. & Hawks, D.C. (2006) Molecular phylogenetics and evolution of the food
656 relocation behaviour of the dung beetle tribe Eucraniini (Coleoptera : Scarabaeidae :
657 Scarabaeinae). *Invertebrate Systematics*, **20**, 557-570.
- 658 Paradis, E., Claude, J. & Strimmer, K. (2004) APE: Analyses of Phylogenetics and Evolution in
659 R language. *Bioinformatics*, **20**, 289-290.
- 660 Plummer, M., Best, N., Cowles, K. & Vines, K. (2004) CODA: Output Analysis and Diagnostics
661 for MCMC, R package version 0.13-3, 2008. *CRAN: The Comprehensive R Archive*
662 *Network*.
- 663 R Development Core Team (2014) R: A language and environment for statistical computing. R
664 Foundation for Statistical Computing, Vienna, Austria.
- 665 Ree, R.H. & Sanmartín, I. (2018) Conceptual and statistical problems with the DEC+ J model of
666 founder-event speciation and its comparison with DEC via model selection. *Journal of*
667 *Biogeography*, **45**, 741-749.
- 668 Ree, R.H. & Smith, S.A. (2008) Maximum likelihood inference of geographic range evolution
669 by dispersal, local extinction, and cladogenesis. *Systematic Biology*, **57**, 4-14.
- 670 Reguero, M.A., Gelfo, J.N., López, G.M., Bond, M., Abello, A., Santillana, S.N. & Marensi,
671 S.A. (2014) Final Gondwana breakup: the Paleogene South American native ungulates
672 and the demise of the South America–Antarctica land connection. *Global and Planetary*
673 *Change*, **123**, 400-413.
- 674 Ronquist, F. (1997) Dispersal-vicariance analysis: A new approach to the quantification of
675 historical biogeography. *Systematic Biology*, **46**, 195-203.
- 676 Sanmartin, I. & Ronquist, F. (2004) Southern hemisphere biogeography inferred by event-based
677 models: plant versus animal patterns. *Systematic Biology*, **53**.
- 678 Sayers, E.W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K.D. & Karsch-Mizrachi, I. (2019)
679 GenBank. *Nucleic Acids Research*, **47**, D94-D99.
- 680 Schöfl, G. (2016) reutils: Talk to the NCBI EUtils. *R package, version 0.2.3*.
- 681 Scholtz, C. & Chown, S. (1995) The evolution of habitat use and diet in the Scarabaeoidea: a
682 phylogenetic approach [pp. 355–374]. *Biology, Phylogeny, and Classification of*

- 683 *Coleoptera: Papers Celebrating the 80th Birthday of Roy A. Crowson. J. Pakaluk and SA*
684 *Slipinski (eds.). Muzeum i Instytut Zoologii PAN, Warszawa, Poland.*
- 685 Scholtz, C.H., Davis, A.L.V. & Kryger, U. (2009) *Evolutionary biology and conservation of*
686 *dung beetles*. Pensoft Sofia.
- 687 Schwery, O. & O'Meara, B.C. (2016) MonoPhy: a simple R package to find and visualize
688 monophyly issues. *PeerJ Computer Science*, **2**, e56.
- 689 Scotese, C. (1993) Maps of generalized continental positions and orography produced by the
690 palaeomap project, University of Texas, Arlington. *Terrestrial ecosystems through time*.
- 691 Seton, M., Müller, R., Zahirovic, S., Gaina, C., Torsvik, T., Shephard, G., Talsma, A., Gurnis,
692 M., Turner, M. & Maus, S. (2012) Global continental and ocean basin reconstructions
693 since 200 Ma. *Earth-Science Reviews*, **113**, 212-270.
- 694 Smith, S.A. & O'Meara, B.C. (2012) treePL: divergence time estimation using penalized
695 likelihood for large phylogenies. *Bioinformatics*, **28**, 2689-2690.
- 696 Sole, C.L. & Scholtz, C.H. (2010) Did dung beetles arise in Africa? A phylogenetic hypothesis
697 based on five gene regions. *Molecular Phylogenetics and Evolution*, **56**, 631-641.
- 698 Sole, C.L., Wirta, H., Forgie, S.A. & Scholtz, C.H. (2011) Origin of Madagascan Scarabaeini
699 dung beetles (Coleoptera: Scarabaeidae): dispersal from Africa. *Insect Systematics &*
700 *Evolution*, **42**, 29-40.
- 701 Spector, S. (2006) Scarabaeine dung beetles (Coleoptera: Scarabaeidae: Scarabaeinae): an
702 invertebrate focal taxon for biodiversity research and conservation. *The Coleopterists*
703 *Bulletin*, **60**, 71-83.
- 704 Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of
705 large phylogenies. *Bioinformatics*, **30**, 1312-1313.
- 706 Stuart, A.J. (2015) Late Quaternary megafaunal extinctions on the continents: a short review.
707 *Geological Journal*, **50**, 338-363.
- 708 Tarasov, S. & Dimitrov, D. (2016) Multigene phylogenetic analysis redefines dung beetles
709 relationships and classification (Coleoptera: Scarabaeidae: Scarabaeinae). *Bmc*
710 *Evolutionary Biology*, **16**.
- 711 Tarasov, S. & Génier, F. (2015) Innovative Bayesian and parsimony phylogeny of dung beetles
712 (Coleoptera, Scarabaeidae, Scarabaeinae) enhanced by ontology-based partitioning of
713 morphological characters. *Plos One*, **10**, e0116671.
- 714 Tarasov, S., Vaz-de-Mello, F.Z., Krell, F.T. & Dimitrov, D. (2016) A review and phylogeny of
715 Scarabaeine dung beetle fossils (Coleoptera: Scarabaeidae: Scarabaeinae), with the
716 description of two *Canthochilum* species from Dominican amber. *Peerj*, **4**.
- 717 Toussaint, E.F., Bloom, D. & Short, A.E. (2017) Cretaceous West Gondwana vicariance shaped
718 giant water scavenger beetle biogeography. *Journal of Biogeography*, **44**, 1952-1965.
- 719 Toussaint, E.F.A., Seidel, M., Arriaga-Varela, E., Hajek, J., Kral, D., Sekerka, L., Short, A.E.Z.
720 & Fikacek, M. (2017) The peril of dating beetles. *Systematic Entomology*, **42**, 1-10.

- 721 Vaz-de-Mello, F. (2007) Revision taxonomica e analysis phylogenetico de la tribu Ateuchini.
722 *Xalapa, Veracruz, Mexico: Instituto de Ecologia AC.*
- 723 Wilkinson, S.P., Davy, S.K., Bunce, M. & Stat, M. (2018) Taxonomic identification of
724 environmental DNA with informatic sequence classification trees.
- 725 Wirta, H., Orsini, L. & Hanski, I. (2008) An old adaptive radiation of forest dung beetles in
726 Madagascar. *Molecular Phylogenetics and Evolution*, **47**, 1076-1089.
- 727 Wirta, H., Viljanen, H., Orsini, L., Montreuil, O. & Hanski, I. (2010) Three parallel radiations of
728 Canthonini dung beetles in Madagascar. *Molecular Phylogenetics and Evolution*, **57**,
729 710-727.
- 730

731 Tables and Figures

Table 1 Tribal Monophyly-Issues Dung Beetles

Monophyly status and reasons for non-monophyly for all tribes in the full matrix trees. Mon.=monophyly-status, #Tips=number of taxa assigned to this tribe, ΔTips=number of additional taxa in this clade (descendants of same MRCA), #Intr.=number of tribes among intruder taxa, Intruders=Genera or Tribes intruding (numbers in parentheses are numbers of tips, all taxa *incertae sedis* unless indicated otherwise), #Outl.=number of taxon-members outside the main clade.

Tribe	Mon.	#Tips	ΔTips	#Intr.	Intruders	#Outl.	Outliers
Eucraniini	Yes	8	0	0		NA	
Ateuchini	Yes	2	0	0		NA	
Eurysternini	Yes	6	0	0		NA	
Gymnopleurini	Yes	8	0	0		NA	
Sisyphini	Yes	6	0	0		NA	
Scarabaeini	No	37	1	1	<i>Neateuchus</i> (1)	0	
Phanaeini	No	44	2	1	<i>Diabroctis</i> (1), <i>Dendropaemon</i> (1)	0	
Onitini	No	7	2	1	<i>Cheironitis</i> (2)	0	
Oniticellini	No	36	10	2	<i>Tiniocellus</i> (2), <i>Tragiscus</i> (1), <i>Drepanocerus</i> (4), <i>Cyptochirus</i> (1), <i>Heterosyphus</i> (1), <i>Proagoderus</i> (1), Onthophagini)	0	
Onthophagini	No	119	53	1	<i>Hyalonthophagus</i> (2), <i>Cleptocaccobius</i> (1), <i>Euonthophagus</i> (2), <i>Milichus</i> (1)	16	<i>Onthophagus diabolicus</i> and 15 more.
Deltochilini	No	132	414	1	Canthonini (2)	112	<i>Arachnodes emmae</i> and 111 more.
Canthonini	No	52	494	1	Deltochilini (7)	25	<i>Megathopa villosa</i> and 24 more.
Coprini	No	23	501	1	<i>Paracopris</i> (1)	10	<i>Coptodactyla storeyi</i> and 9 more.
Dichotomiini	No	36	510	0		24	<i>Canthidium haroldi</i> and 23 more.

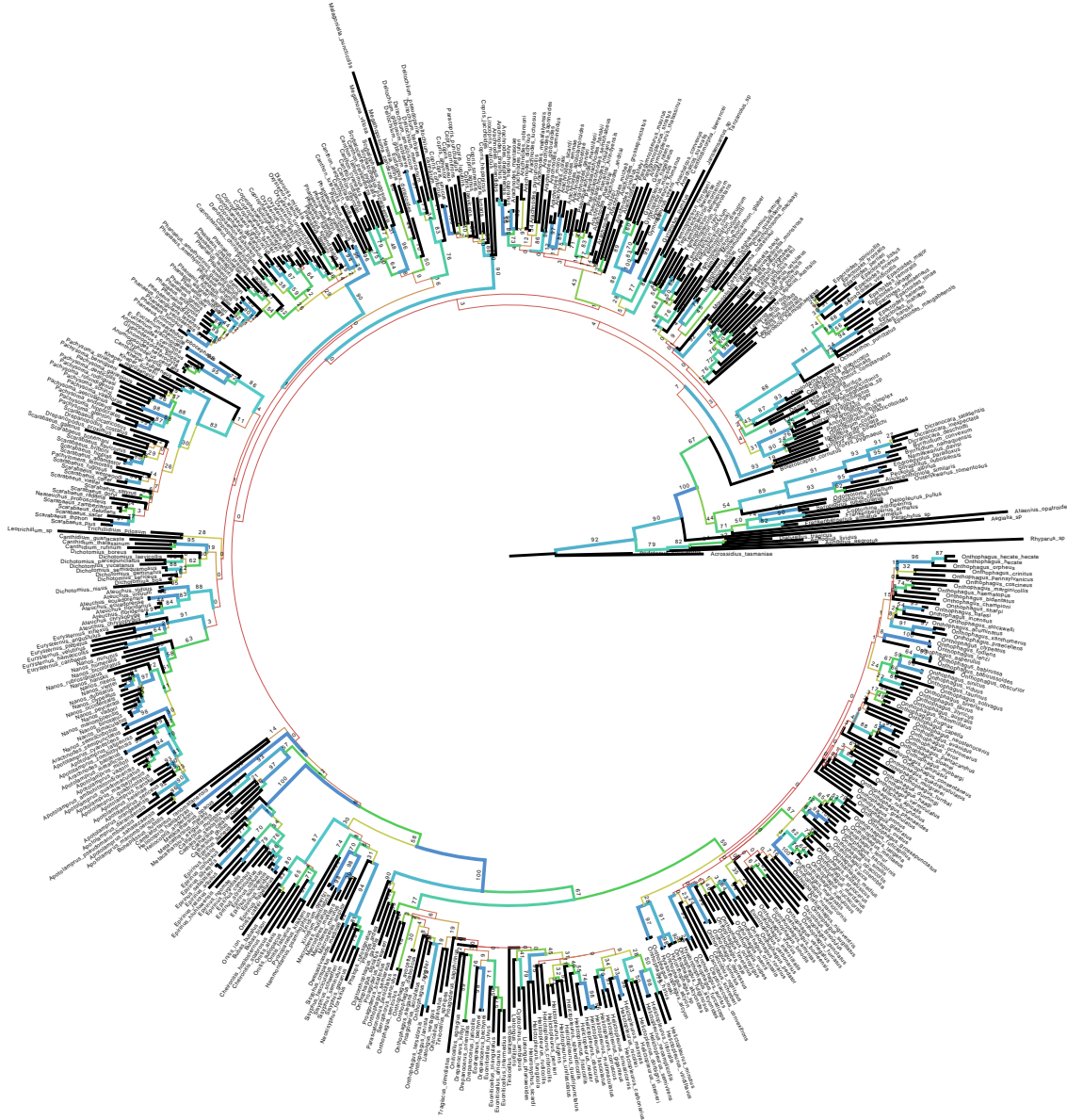


Figure 1: Branch Support Dung Beetle Phylogeny.

Bootstrap branch length support on the un-dated RAxML phylogeny. Colder colors indicating higher branch support, decreasing branch width with decreasing support (terminal branches are black).

Table 2 Node Calibrations and Estimated Ages.

For each node age calibration, the corresponding clade, fossil, location on the branch (crown or stem) are given. For each calibration scheme ('young' vs. 'old'), the age constraints used and the estimated node ages on crown and stem, are indicated for each calibrated node, followed by the difference between the estimates.

Clade	Fossil	Stem/ crown	Young				Old				Age Diff.	
			Min cons.	Max cons.	Crown age	Stem age	Min cons.	Max cons.	Crown age	Stem age	Min	Max
<i>Copris</i>	<i>Copris kartlinus</i>	stem	3.6	98.1	85.80	85.89	3.6	131.6	59.80	59.87	26.00	26.02
<i>Gymnopleurus</i>	<i>Gymnopleurus sisyphus</i>	stem	13.5	98.1	35.53	52.97	13.5	131.6	50.45	74.30	14.91	21.33
<i>Onthophagus</i>	<i>Onthophagus bisoninus</i>	stem	13.5	98.1	57.74	70.81	13.5	131.6	85.03	100.58	27.29	29.77
<i>Heliocopris</i>	<i>Heliocopris antiquus</i>	stem	18.7	98.1	1.46	78.79	18.7	131.6	2.09	110.90	0.63	32.11
Scarabaeinae	<i>Lobateuchus parisii</i>	crown	53	98.1	98.10		53	131.6	131.60		33.50	
Scarabaeidae	<i>Juraclonus rohdendorfi</i>	crown	152	181.8			152	199.64				

Table 3 Best Fitting Models for Biogeographical Diversification Hypotheses.

Best fitting models of ML GeoSSE analyses for each data set and each tree, as determined by likelihood ratio tests and Δ AIC scores. Models where lambda was constrained to be equal in both areas, also constrained lambda to be zero for widespread taxa.

Tested Pairing	Young Tree	Old Tree
Large Mammals vs. rest	equal mu, equal dispersal	equal mu, equal dispersal
Small Mammals vs. rest	equal lambda, equal mu	equal lambda, equal mu
Gondwana vs. rest	equal mu	equal mu, equal dispersal
Laurasia vs . rest	full model	equal lambda, equal mu
Madagascar vs. rest	equal dispersal	equal mu, equal dispersal

Table 4 Posterior Distribution of GeoSSE Rate Estimates.

95% quantiles of posterior distributions for diversification and dispersal rates estimated under the best scoring model for each data set and each tree. s=speciation rate, x=extinction rate, r=net diversification rate (s-x), d= dispersal rate.

732

Pairings	Quant.	sA	sB	sAB	xA	xB	rA	rB	rAB	dA	dB	Likelihood
	Young Tree											
Only Large Mam. Vs. rest	2.50%	0.0968	0.0942	0.0382		0.0192	0.0636	0.0600	0.0382		0.0008	-2436.8114
	97.50%	0.1449	0.1437	3.9306		0.0785	0.0803	0.0799	3.9306		0.0021	-2428.7662
Only Small Mam. vs. rest	2.50%		0.1076			0.0342		0.0613		0.0002	0.0015	-2436.8855
	97.50%		0.1552			0.0922		0.0756		0.0014	0.0042	-2431.4368
Madagascar vs. rest	2.50%	0.1041	0.0930	0.1085	0.0006	0.0191	0.0770	0.0602	0.1085		0.0003	-2381.0849
	97.50%	0.2102	0.1423	8.0402	0.1276	0.0798	0.1158	0.0756	8.0402		0.0010	-2373.0650
Laurasia vs. rest	2.50%	0.0819	0.0863	0.0016	0.0006	0.0160	0.0581	0.0528	0.0016	0.0223	0.0003	-2619.1510
	97.50%	0.1308	0.1409	0.0651	0.0674	0.0857	0.0928	0.0724	0.0651	0.0577	0.0036	-2610.9780
Gondwana vs. rest	2.50%	0.0964	0.0683	0.0002		0.0008	0.0863	0.0593	0.0002	0.0051	0.0028	-2629.9069
	97.50%	0.1228	0.0924	0.0309		0.0305	0.1054	0.0721	0.0309	0.0156	0.0056	-2622.8756
Pairings	Old Tree											
Only Large Mam. vs. rest	2.50%	0.0636	0.0619	0.0286		0.0093	0.0448	0.0426	0.0286		0.0005	-2632.4797
	97.50%	0.0952	0.0942	1.5142		0.0486	0.0560	0.0557	1.5142		0.0013	-2624.6484
Only Small Mam. vs. rest	2.50%		0.0698			0.0183		0.0435		0.0001	0.0010	-2633.0250
	97.50%		0.0994			0.0546		0.0530		0.0009	0.0029	-2627.6812
Madagascar vs. rest	2.50%	0.0726	0.0575	0.0952		0.0043	0.0558	0.0437	0.0952		0.0002	-2578.7214
	97.50%	0.1066	0.0887	6.9666		0.0438	0.0753	0.0544	6.9666		0.0007	-2571.5387
Laurasia vs. rest	2.50%		0.0602			0.0103		0.0394		0.0137	0.0002	-2810.4804
	97.50%		0.0851			0.0443		0.0513		0.0327	0.0027	-2805.3483
Gondwana vs. rest	2.50%	0.0627	0.0477	0.0001		0.0003	0.0580	0.0437	0.0001		0.0028	-2829.6036
	97.50%	0.0780	0.0611	0.0174		0.0152	0.0702	0.0512	0.0174		0.0045	-2822.8190

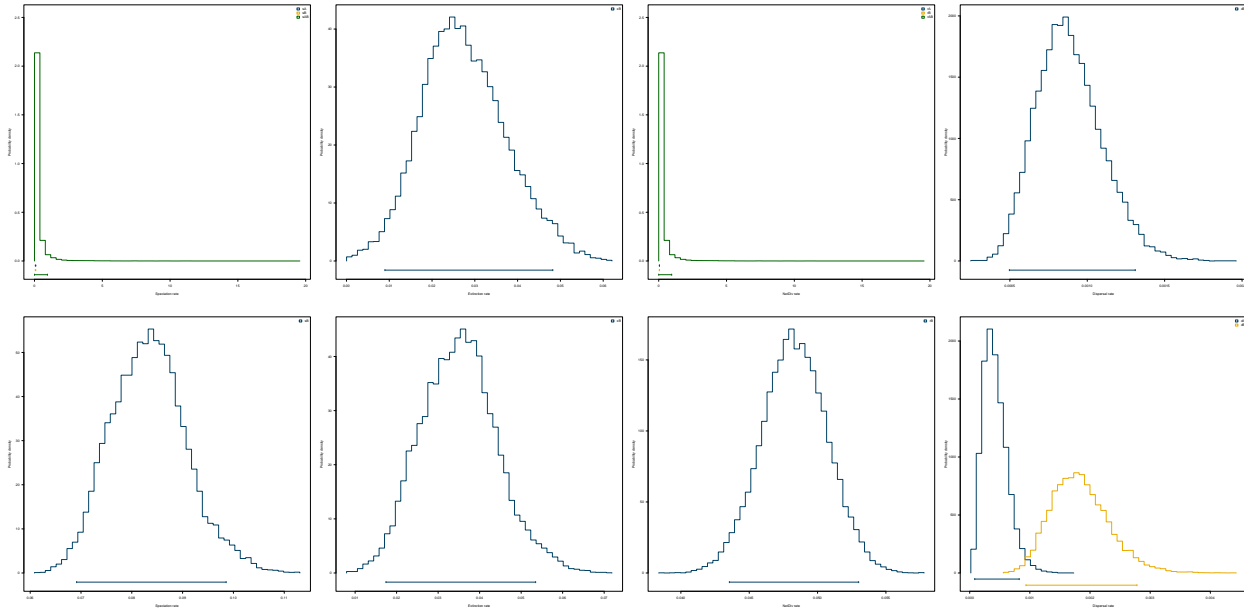


Figure 2: GeoSSE Results Dung Producers Old Tree.

Posterior distributions of area-dependent rates estimated by GeoSSE. Left to right: speciation rates, extinction rates, net diversification rates (speciation – extinction), dispersal rates. **Top:** rates in areas with large and medium sized mammals and high and intermediate droppings-diversity (Afro-Eurasia and Americas, blue), and rates in areas with small sized mammals and low droppings-diversity (East Gondwanan Fragments, yellow). **Bottom:** rates in areas with large sized mammals and high droppings-diversity (Afro-Eurasia, yellow), and rates of areas with low and medium sized mammals and low and intermediate droppings-diversity (Americas and East Gondwanan Fragments, blue). Rates of widespread taxa are colored green; where only one rate was estimated, it is colored blue. Dispersal rates of an area reflect dispersal *out* of said area.

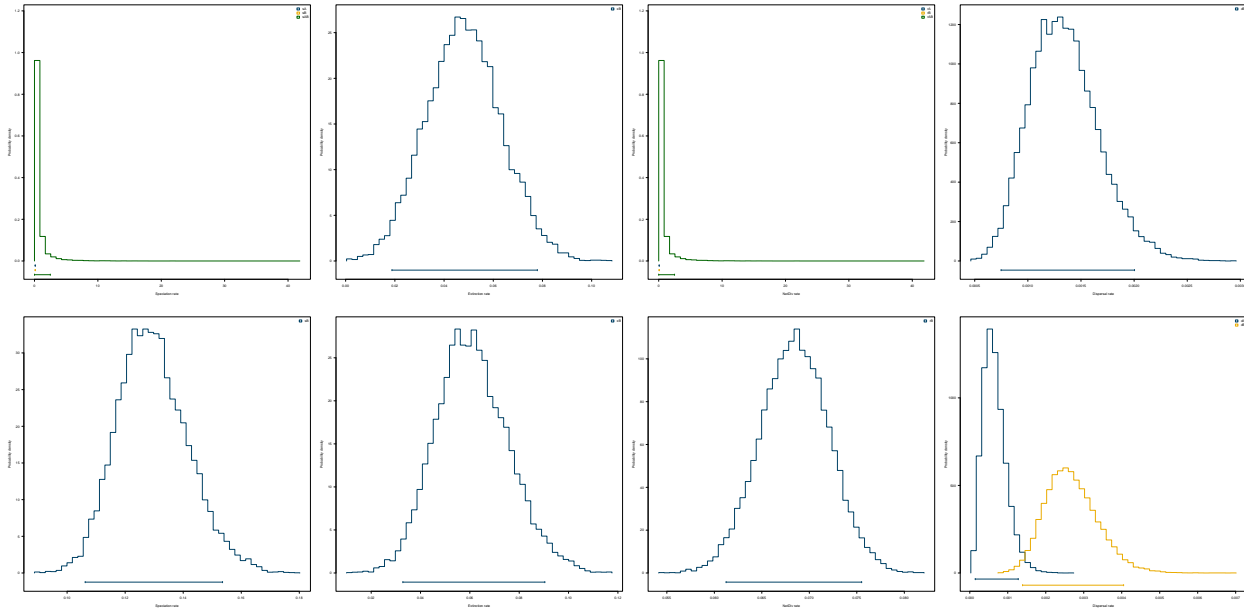


Figure 3: GeoSSE Result Dung Producers Young Tree.

Posterior distributions of area-dependent rates estimated by GeoSSE. Left to right: speciation rates, extinction rates, net diversification rates (speciation – extinction), dispersal rates. **Top:** rates in areas with large and medium sized mammals and high and intermediate droppings-diversity (Afro-Eurasia and Americas, blue), and rates in areas with small sized mammals and low droppings-diversity (East Gondwanan Fragments, yellow). **Bottom:** rates in areas with large sized mammals and high droppings-diversity (Afro-Eurasia, yellow), and rates of areas with low and medium sized mammals and low and intermediate droppings-diversity (Americas and East Gondwanan Fragments, blue). Rates of widespread taxa are colored green; where only one rate was estimated, it is colored blue. Dispersal rates of an area reflect dispersal *out* of said area.

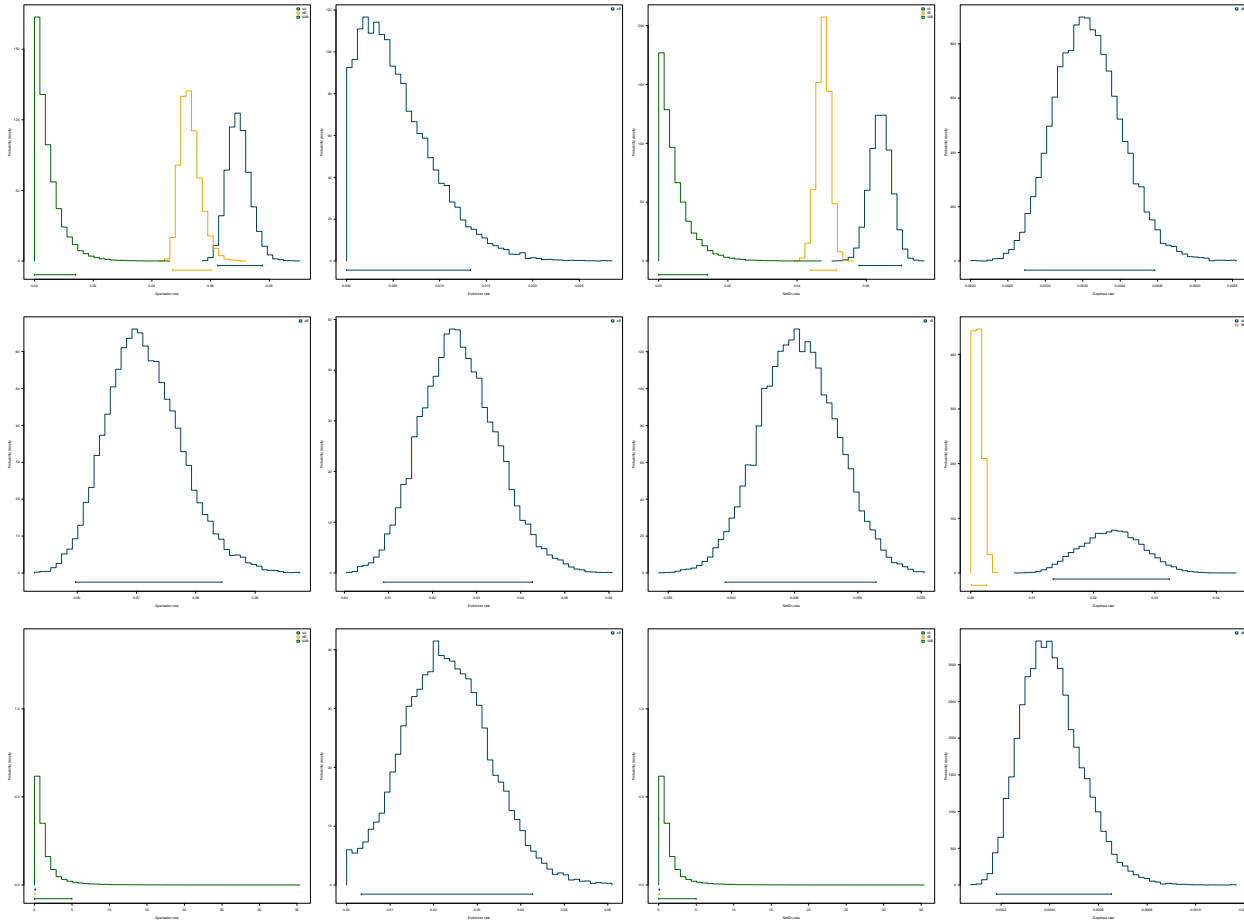


Figure 4: GeoSSE Results Out-Of-Gondwana Old Tree.

Posterior distributions of area-dependent rates estimated by GeoSSE. Left to right: speciation rates, extinction rates, net diversification rates (speciation – extinction), dispersal rates. **Top:** rates in Laurasia and Madagascar (blue), and rates in Gondwana (yellow). **Center:** rates in Laurasia (blue), and rates in Gondwana and Madagascar (yellow). **Bottom:** rates in Madagascar (blue), and rates in Gondwana and Laurasia. Rates of widespread taxa are colored green; where only one rate was estimated, it is colored blue. Dispersal rates of an area reflect dispersal *out* of said area.

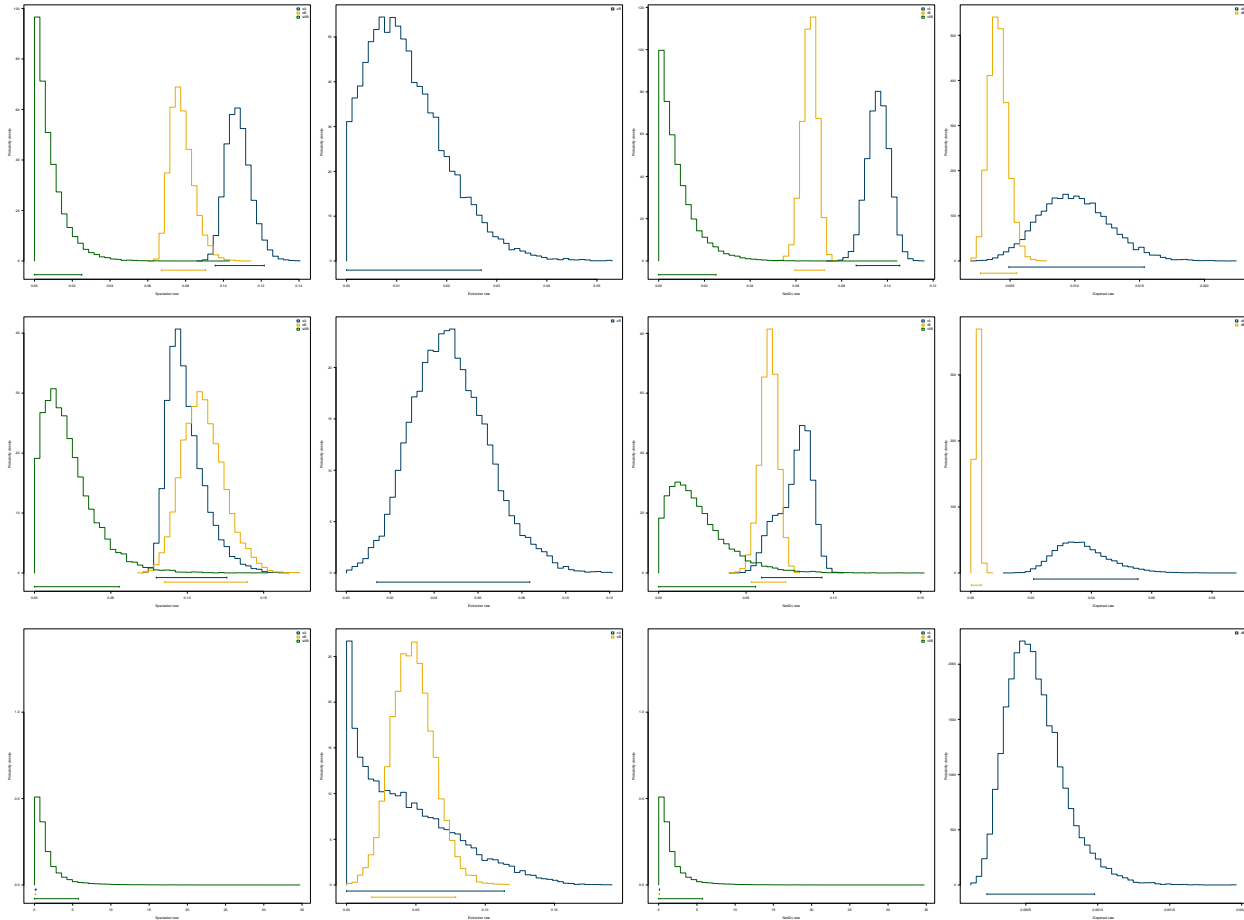


Figure 5: GeoSSE Results Out-Of-Gondwana Young Tree.

Posterior distributions of area-dependent rates estimated by GeoSSE. Left to right: speciation rates, extinction rates, net diversification rates (speciation – extinction), dispersal rates. **Top:** rates in Laurasia and Madagascar (blue), and rates in Gondwana (yellow). **Center:** rates in Laurasia (blue), and rates in Gondwana and Madagascar (yellow). **Bottom:** rates in Madagascar (blue), and rates in Gondwana and Laurasia. Rates of widespread taxa are colored green; where only one rate was estimated, it is colored blue. Dispersal rates of an area reflect dispersal *out* of said area.