Phylogenomic analyses reveal a Gondwanan origin and repeated out of India colonizations into Asia by tarantulas (Araneae: Theraphosidae)

Saoirse Foley [Corresp., 1, 2], Henrik Krehenwinkel [3], Dong-Qiang Cheng [2], William H Piel [1, 2, 4]

1 Department of Biological Science, National University of Singapore, Singapore, Singapore

2 Division of Science, Yale-NUS College, Singapore, Singapore

3 Department of Biogeography, Universität Trier, Trier, Germany

4 Lee Kong Chian Natural History Museum, National University of Singapore, Singapore, Singapore

Corresponding Author: Saoirse Foley
Email address: schioedtei@gmail.com

The study of biogeography seeks taxa that share a key set of characteristics, such as timescale of diversification, dispersibility, and ecological lability. Tarantulas are ideal organisms for studying evolution over continental-scale biogeography given their time period of diversification, their mostly long-lived sedentary lives, low dispersal rate, and their nevertheless wide circumtropical distribution. In tandem with a time-calibrated transcriptome-based phylogeny generated by PhyloBayes, we estimate the ancestral ranges of ancient tarantulas using two methods, DEC+j and BBM, in the context of their evolution. We recover two ecologically distinct tarantula lineages that evolved on the Indian Plate before it collided with Asia, emphasizing the evolutionary significance of the region, and show that both lineages diversified across Asia at different times. Our most ancestral tarantulas emerge on the Americas and Africa 120Ma–105.5Ma. We provide support for a dual colonization of Asia by two different tarantula lineages that occur at least 20 million years apart, as well as a Gondwanan origin for the group. We determine that their current distributions are attributable to a combination of Gondwanan vicariance, continental rafting, and geographic radiation. We also discuss emergent patterns in tarantula habitat preferences through time.

1 **PHYLOGENOMIC ANALYSES REVEAL A GONDWANAN ORIGIN AND**

2 **REPEATED OUT OF INDIA COLONIZATIONS INTO ASIA BY**

3 **TARANTULAS (ARANEAE: THERAPHOSIDAE)**

4 Saoirse Foley[1,2, †,], Henrik Krehenwinkel[3], Dong-Qiang Cheng[2], William H. Piel[1,2,4].

5 [1] Department of Biological Science, National University of Singapore, 117543, Singapore.

6 [2] Division of Science, Yale-NUS College, 10 College Avenue West, 138609, Singapore.

7 [3] Department of Biogeography, Trier University, Trier, Germany

8 [4] Lee Kong Chian Natural History Museum, National University of Singapore, 117377, Singapore.

9 [†] Present Address: Department of Biological Sciences, Carnegie Mellon University, 5000 Forbes Ave,
10 Pittsburgh, PA 15213, USA.

12 Corresponding author:

13 Saoirse Foley[1,2, †,]

14 Department of Biological Sciences, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA
15 15213, USA

16 schioedtei@gmail.com

ABSTRACT

The study of biogeography seeks taxa that share a key set of characteristics, such as timescale of diversification, dispersibility, and ecological lability. Tarantulas are ideal organisms for studying evolution over continental-scale biogeography given their time period of diversification, their mostly long-lived sedentary lives, low dispersal rate, and their nevertheless wide circumtropical distribution. In tandem with a time-calibrated transcriptome-based phylogeny generated by PhyloBayes, we estimate the ancestral ranges of ancient tarantulas using two methods, DEC+j and BBM, in the context of their evolution. We recover two ecologically distinct tarantula lineages that evolved on the Indian Plate before it collided with Asia, emphasizing the evolutionary significance of the region, and show that both lineages diversified across Asia at different times. Our most ancestral tarantulas emerge on the Americas and Africa 120Ma–105.5Ma. We provide support for a dual colonization of Asia by two different tarantula lineages that occur at least 20 million years apart, as well as a Gondwanan origin for the group. We determine that their current distributions are attributable to a combination of Gondwanan vicariance, continental rafting, and geographic radiation. We also discuss emergent patterns in tarantula habitat preferences through time.

Keywords: Araneae; Biogeography; Indian Plate; Tarantulas; Gondwana; Theraphosidae; Time-Calibration.

INTRODUCTION

Due to the historically close geographies of several continental land masses prior to the breakup of the supercontinents Laurasia and Gondwana, many related taxa are presently found on opposite sides of the world. For instance, it has been shown that rodents and primates (De Oliveira, Molina & Marroig, 2009) in the Neotropics likely arrived there from Africa. Several mass faunal exchanges between continents have occurred when two land masses eventually meet, such as the "Great American Interchange" (Marshall *et al.*, 1982) after the formation of the Isthmus of Panama (O'Dea *et al.*, 2016), the dispersal of several taxa "out of India" and into Asia / Indomalaya (Karanth, 2006), and exchanges between Africa and Eurasia (Hedges, 2001; Springer *et al.*, 2011). If we are to understand the evolutionary patterns involved in the diversification of taxa and the generation of biodiversity, it becomes important to consider the roles that continental drift may play, particularly in cases where evolutionary novelties have emerged in the diversification process.

Studies that use model taxa to understand biogeographic patterns are limited by the fact that the ideal model organism should have both a low dispersal ability and a wide (if not global) distribution. Given their low motility and wide array of dispersal strategies, plants have often been the focus of biogeographic studies (e.g. Zhou *et al.*, 2018; McGlone *et al.*, 2018; Fariña *et al.*, 2018; Echeverría-Londoño *et al.*, 2018). Mygalomorph spiders from the family Theraphosidae, also known as tarantulas, are also useful models for understanding biogeographic patterns. They possess several intriguing adaptations that have received significant attention in recent years, such as their stridulatory capacity (Galleti-Lima & Guadanucci, 2018, 2019), urticating bristle evolution (Bertani, 2003; Kaderka *et al.*, 2019; Foley *et al.*, 2019), colouration

59 (Foley, Saranathan & Piel, 2020), and venom compositions (Santana *et al.*, 2017; Lüddecke,
60 Vilcinskas & Lemke, 2019). Robust, subfamily-level phylogenies are late in coming (Lüddecke
61 *et al.*, 2018; Foley *et al.*, 2019), but these phylogenies have verified the monophylies of all
62 studied subfamilies, except Ischnocolinae (with Selenogyrinae being the only subfamily not
63 represented by these studies). It has only recently become possible to study tarantula adaptations
64 in an explicitly phylogenetic context.

65   This diverse suite of intriguing adaptations undoubtedly contributed to the success of
66 these creatures. They are quite widespread and are found throughout the subtropical regions of
67 every continent (Gallon, 2000; World Spider Catalog, 2020). However, while many mature male
68 tarantulas are known to wander in search of females (Prentice, 1992; Pérez-Miles *et al.*, 2005;
69 Pérez-Miles *et al.*, 2007), most juvenile and female tarantulas rarely venture far from their
70 retreats (Yáñez & Floater, 2000; Schultz & Schultz, 2009). Spiderlings seldom disperse far from
71 the burrows of their mothers, leading to dense aggregations of burrows in relatively small areas
72 (Reichling, 2000; Shillington & McEwen, 2006). There are no reports of ballooning tarantulas
73 like in other mygalomorph families (Coyle, 1983; Coyle *et al.*, 1985; Buzatto, Haeusler &
74 Tamang, 2021), but some mygalomorphs are known to be highly divergent across small areas
75 (Starrett & Hedin, 2007). These behaviours do not portend that tarantulas would be successful
76 dispersers, yet they have spread across the globe and have colonized strikingly different
77 ecological niches.

78   Previous studies estimate that tarantulas emerged between 150Ma–71Ma (Bond *et al.*,
79 2014) or ~107Ma (Opatova *et al.*, 2020), which is compatible with a Gondwanan origin. Indeed,
80 some tarantulas (Selenocosmiinae) are suggested to be North Gondwanan taxa (West, Nunn &
81 Hogg, 2012). Given their habitat specificity and low dispersal ability, yet wide geographic
82 distribution, tarantulas are good model organisms for addressing biogeographic questions. In
83 particular, two distinct lineages of Asian tarantulas were recovered in the most recent subfamily-
84 level tarantula phylogeny (Foley *et al.*, 2019). Estimates surrounding the timing of the collision
85 between the Indian Subcontinent and Asia vary, but most studies agree on a late Paleocene /
86 early Eocene timescale around 55Ma (Hu *et al.*, 2016; Zheng & Wu, 2018; Westerweel *et al.*,
87 2019). Migrations from the Indian Subcontinent to Asia are known to have played a significant
88 role in generating the high faunal diversity observed in Asia today (Karanth, 2006; Morley *et al.*,
89 2016; Morley, 2018; Garg & Biju, 2019), even amongst spiders (Li, Shao & Li, 2020) and other
90 invertebrates (Joshi, Karanth & Edgecombe, 2020). It has also been shown that other spiders of a
91 comparable age originated on the Gondwanan supercontinent, i.e. South America and Africa,
92 before eventually colonizing Asia and Oceania (Chamberland *et al.*, 2018). We expect to recover
93 a similar pattern for tarantulas.

94   Here, we test the Gondwanan origin hypothesis for tarantulas using a broad sampling of
95 molecular data and taxa. We also hypothesize that Asian tarantulas may have originated on the
96 Indian Subcontinent and subsequently dispersed throughout Asia. We test this by inferring the
97 historical patterns that contributed to present-day theraphosid distributions. We use a subfamily-
98 level chronogram of Theraphosidae and estimate their ancestral ranges under two different
99 models, allowing us to interpret the node ages in the context of geographical events. We then
100 discuss a timeline of events that may have led to the emergence of theraphosid subfamilies, with
101 emphasis on those found in Asia, and present some hypotheses on how those events impacted
102 their current distributions.

103

METHODS

Data acquisition

106 Although theraphosids are the primary focus of this study, the poor fossil record in this group
107 required that we expand our sampling to include other Mygalomorphae with richer fossil records
108 so that these could assist in molecular clock calibration. Only two tarantula fossils are known,
109 and neither can be firmly placed onto the tarantula phylogeny; one from Burmese amber dating
110 to ~100Mya (Wunderlich & Müller, 2020), and one from Chiapas amber dating to somewhere
111 between the late Oligocene and mid Miocene (possibly as young as 16Mya, Dunlop, Harms &
112 Penney, 2008). Inferring calibration points from prior studies (i.e. secondary calibration) has also
113 been discouraged (Schenk, 2016).

114 Transcriptome data for 29 theraphosids and 18 other mygalomorphs was obtained from
115 the publicly available sequence read archive (SRA) database (http://www.ncbi.nlm.nih.gov/sra).
116 Additionally, to improve the rooting of the family, we included a new transcriptome from a field-
117 collected barychelid, *Rhianodes atratus* (collected under permit number NP/RP18-046, issued by
118 the National Parks Board of Singapore). RNA was extracted using a TRIzol total RNA extraction
119 protocol (Simms, Cizdziel & Chomczynski, 1993). The RNA was subsequently sent to a
120 commercial company (NovogeneAIT) for paired-end sequencing on an Illumina HiSeq 4000.
121 The resulting transcriptome is available on NCBI's Sequence Read Archive (SRA) under
122 bioproject number PRJNA635363 (accession GIXJ00000000). Accession numbers for a total of
123 48 transcriptomes are included in supplementary Table S1. Assembly data for the *Rhianodes*
124 *atratus* (Barychelidae) transcriptome is available in supplementary Table S2.

125

A core ortholog variant to obtain a DNA supermatrix

127 Transcriptomes were assembled using Trinity v2.6.6 (Grabherr *et al.*, 2011), and protein coding
128 regions were predicted using TransDecoder v5.5.0 (Haas *et al.*, 2013). The core-ortholog (OG)
129 pipeline as per Garrison *et al.* (2016) was implemented for spider-specific ortholog selection
130 across our dataset. The initial set of core orthologs was generated from a selection of publicly
131 available transcriptomes, as per Cheng and Piel (2018), and a set of 4,446 profile hidden Markov
132 models (pHMMs) was generated following their methods. HaMStR v13.1 (Ebersberger, Strauss
133 & von Haeseler, 2009) inferred orthology between these pHMMs and our dataset. Groups of
134 orthologous genes were then pooled together and subjected to several filtering steps.

135 Software used during the filtering process include the following: MAFFT (Katoh *et al.*,
136 2005) for alignment during filtering; ALISCORE (Misof & Misof, 2009; Kück *et al.*, 2010) for
137 alignment trimming; ALICUT (Kück, 2009) for excising ambiguous regions; Infoalign (Rice,
138 Longden & Bleasby, 2000) for generating consensus sequences from OGs. Generally, the
139 filtering process and criteria was the same as in Foley *et al.* (2019) with one key exception –
140 instead of removing any OG that was sampled for half (or fewer) of the taxa included, we chose
141 to only retain OGs that were present in all taxa, discarding the rest. Despite this stringency, an
142 adequate set of 743 OG alignments remained, each of which included all 48 taxa. To obtain a
143 more neutral, granular dataset, these protein alignments were converted back to DNA alignments
144 using localized tBLASTn searches for each taxon (Camacho *et al.*, 2009) against their

145 corresponding Trinity outputs, in tandem with a set of custom Perl scripts for parsing. Gene trees
146 for each of the 743 alignments were estimated using RAxML (Stamatakis, 2014). These
147 departures from previous filtering criteria were informed by the downstream needs of our
148 analysis pipeline.

149    To exclude OGs with conflicting signals, we identified divergent gene trees using
150 Robinson-Foulds (RF) metrics (Robinson & Foulds, 1981) in the R package "phytools" (Revell,
151 2012). Given that the theraphosid tree from Foley *et al.* (2019) received 100% bootstrap support
152 at almost all nodes, we used this as the reference tree to compare with each gene tree. In order to
153 calculate RF values, each of our 743 gene trees was pruned to match the set of taxa in the
154 reference tree. Any OG DNA alignment whose corresponding gene tree received an RF score of
155 ≥ 6 was discarded, leaving 125 OGs.

156    Next, TranslatorX (Abascal, Zardoya & Telford, 2010) was used to realign the DNA
157 sequences for the 125 OGs according to our protein alignments. Visualization of OGs in
158 DAMBE v7 (Xia & Xie, 2001) showed that phylogenetically informative content was being
159 distorted due to sequence saturation (Xia *et al.*, 2003; Philippe *et al.*, 2011), so every third base
160 was subsequently removed to mitigate this effect. Lastly, FASconCAT (Kück & Meusemann,
161 2010) was used to concatenate all 125 OGs into a supermatrix, and GBLOCKS (Castresana,
162 2002) was subsequently used to trim and excise ambiguous regions. The result was a 96%
163 complete DNA alignment of 48 taxa and 89,302 characters.

164

165 *Time calibration with PhyloBayes*

166 BEAST v1.10.4 (Drummond & Rambaut, 2007) was used to generate an initial starting
167 phylogeny for time calibration, running for 100 million generations with 20% of the trees
168 discarded as burnin. Divergence times were estimated in PhyloBayes v4.1 (Lartillot, Lepage &
169 Blanquart, 2009) using this tree, the DNA supermatrix, and four fossils to infer calibration
170 points. We followed Opatova *et al.* (2020) in treating the oldest mygalomorph fossil
171 *(Rosamygale grauvogely*, Selden & Gall, 1992) as an Avicularioidea crown group, and use a
172 subset of their fossil calibrations.

173    The first calibrated node corresponds to the Avicularioidea-Atypoidea split. The
174 *Rosamygale grauvogely* fossil dated to 242 Million years ago (Ma), which represents the
175 minimum bound for this node, while 323Ma serves as the maximum bound and corresponds to
176 the oldest age of the Bashkirian stage from which the oldest known spider fossil (*Arthrolycosa*
177 sp.) hails (Selden *et al.*, 2014; Garwood *et al.*, 2016). Secondly, the split between Nemesioidina
178 and Bemmeridae / Theraphosidae was assigned a minimum bound of 125Ma based on the
179 *Cretamygale chasei* fossil (Selden, 2002), and a maximum bound of 242Ma based on
180 *Rosamygale grauvogely*. Lastly, the split between Antrodiaetidae and Atypidae was assigned a
181 minimum bound of 100Ma based on a fossil of *Ambiortiphagus ponomarenkoi* (Eskov &
182 Zonstein, 1990), with *Rosamygale grauvogely* once again providing the maximum bound of
183 242Ma for this node.

184    Four independent PhyloBayes chains were each run for c. 24,000 cycles under an
185 autocorrelated lognormal model with a GTR matrix, and otherwise default parameters. Several
186 advantages of autocorrelated lognormal models over uncorrelated rates models have been

187 demonstrated for a variety of datasets (Lepage *et al.*, 2007; Paradis, 2013), and they have been
188 shown to yield more plausible results in similar studies to ours (Sharma *et al.*, 2018). Burn-in
189 was adjusted to 10% to maximize the number of estimated sample size values that exceeded 200,
190 yielding a total of 86,933 cycles across all chains. Divergence times were estimated using the
191 "readdiv" command in PhyloBayes for each chain, and the results were summarized with
192 TreeAnnotator. The resulting time-calibrated mygalomorph phylogeny is available in
193 supplementary Figure S1, and we proceeded with the tarantula subsection of this phylogeny for
194 further analyses.

195

196 *Biogeographic analysis*

197 The tarantulas included in this study were assigned to one or more of the following five
198 biogeographic ranges as determined by the World Spider Catalog (2020): (A) Indian
199 Subcontinent; (B) Asia (non-Indian Indomalaya); (C) East of the Wallace Line; (D) Americas;
200 and (E) Africa. Taxon range codings are included in supplementary Table S1. We also accounted
201 for proximity between areas by excluding certain combinations of ancestral area (e.g. Asia plus
202 Americas. Full list in supplementary Table S3). These areas were chosen based on the
203 biogeographic realms first proposed by Udvardy (1975), but we treat India separately to the
204 remainder of Indomalaya given that we expect the age of theraphosids to exceed the collision of
205 India and Asia.

206   The tarantula subsection (with outgroups) of our tree from PhyloBayes served as input
207 for RASP (Yu *et al.*, 2015), which estimated ancestral range distributions under two different
208 models that were chosen via model testing (Table 1) in BioGeoBEARS (Matzke, 2013). Despite
209 BAYAREALIKE+j being predicted as the most favorable model, its assumption of a static
210 geological history (Landis *et al.*, 2013) rendered it unsuitable. Hence, we ran the analysis using
211 our second highest scoring model — the dispersal-extinction cladogenesis with jump dispersal
212 (DEC+j) model (Ree & Smith, 2008). However, such "+j" models have been divisive, with some
213 criticizing the "+j" parameter (Ree & Sanmartín, 2018) and others advocating for its inclusion
214 (McDonald-Spicer *et al.*, 2019; Klaus & Matzke, 2019). Therefore, Bayesian Binary MCMC
215 (BBM) was added as a second method, and was ran under the JC69 model for 1 million
216 generations across ten chains with sampling every 100 generations (S. S. Ali *et al.*, 2011). As the
217 model testing found no statistically significant difference between DEC and DEC+j (Table 1),
218 DEC was added as a third method. RASP was also used to calculate an event matrix for node 21.

219

220 **RESULTS**

221 *Ancestral ranges and ages of theraphosid clades*

222 Our DEC+j (Figure 1) and BBM (supplementary Figure S3) results are largely in agreement,
223 disagreeing at only at two nodes (per Figure 1) — node 20 (which DEC+j estimates as present on
224 both Africa and the Americas, whereas BBM estimates the Americas only), and node 28 (which
225 DEC+j estimates as present on Asia only, whereas BBM estimates it as present on Asia and East
226 of the Wallace line). Full range probabilities for DEC+j and BBM are given in supplementary
227 Figures S2 and S3 respectively. The DEC reconstruction is available in supplementary Figure

228 S4, and although there are some differences between this method and the others, the larger
229 concepts discussed are not impacted.

230   Table 2 contains divergence time estimates as 95% HPDs for the theraphosid subsection
231 of the mygalomorph phylogeny, while supplementary Table S4 contains those estimates for all
232 nodes. Our earliest theraphosid node (20) is recovered 120Ma–115.5Ma in the Americas (Figure
233 1), and is found sister to *Rhianodes atratus*, indicating that Barychelidae are sister to
234 Theraphosidae. When considered alongside the divergence times, these estimations are
235 consistent with a Gondwanan origin for Theraphosidae.

236   The Indian Plate is first recovered as an ancestral area at node 25 108Ma–103.5Ma. All
237 ancestral range estimations recover two ecologically distinct Indian Plate lineages descending
238 from node 25. The first such node (26) emerges 99Ma–95Ma, yielding several ground-dwelling
239 taxa. It diverges into the Indian Thrigmopoeinae lineage, and the ancestor to Selenocosmiinae
240 (node 27), which emerges 57.4Ma–55Ma. Node 27 represents the first appearance of Asia as an
241 ancestral area in both DEC+j and BBM, albeit alongside the Indian Subcontinent (Figure 1). This
242 node then diverges into the *Haplocosmia* lineage, and the ancestor to *Selenocosmia* and
243 *Phlogiellus*, which also emerges as present east of the Wallace line around 47.5Ma-45.5Ma
244 (Figure 1).

245   The second descendant of the first Indian Subcontinent node (25) corresponds to the last
246 common ancestor of Harpactirinae and Ornithoctoninae (node 29), emerging 100Ma–96Ma. This
247 node subsequently diverges into the ancestor of Harpactirinae and Stromatopelminae (node 30),
248 and yet another Indian Subcontinent node corresponding to the ancestors of the primarily
249 arboreal Ornithoctoninae, and arboreal Poecilotheriinae (node 33). After the divergence of
250 Ornithoctoninae and Poecilotheriinae, we see Asia represented as an ancestral area for the second
251 time in both DEC+j and BBM as the ancestor to Ornithoctoninae (node 34) emerges 35Ma–
252 33Ma. DEC, however, places the ancestor to Poecilotheriinae and Ornithoctoninae on both the
253 Indian Plate and Asia (supplementary Figure S4).

254   The ancestor to the primarily African Eumenophorinae, and the *Catumiri* lineage (node
255 21) emerges in the Americas 112Ma–107.5Ma in all three analyses, with DEC also placing it in
256 Africa. The event matrices of all analyses for this node are given in Table 3.

257

DISCUSSION

The importance of India and the dual colonizations of Asia

260 The "out of India" hypothesis (McKenna, 1973) has been accepted as a potential distribution
261 pattern for the radiations of a diverse array of taxa (Karanth, 2006; Conti *et al.*, 2002; Joshi,
262 Karanth & Edgecombe, 2020). Our results are consistent with two independent "out of India"
263 dispersals of tarantulas into Asia (Figures 1 and 2). This "dual colonization of Asia via India"
264 pattern has also been demonstrated in centipedes (Joshi, Karanth & Edgecombe, 2020). Both of
265 our Asian tarantula lineages were estimated to arrive there during timescales consistent with Li,
266 Shao, and Li (2020), who determined that Ochyroceratid spiders likely arrived in Asia via India
267 55Ma–38Ma, and also with the India / Asia collision timescales (Hu *et al.*, 2016; Westerweel *et*
268 *al.*, 2019).

269   All three analyses place the ancestor to Selenocosmiinae (node 27) as present on the
270   Indian Subcontinent and Asia 57.5Ma–55Ma (Figure 1) — the first entry of theraphosids into
271   Asia per DEC+j and BBM. Furthermore, both of these analyses recover the ancestor to
272   *Selenocosmia* and *Phlogiellus* (node 28) in both Asia and east of the Wallace line, with DEC
273   recovering the node as Asia-only. Interestingly, despite our analyses suggesting a Gondwanan
274   origin for Theraphosidae, this pattern suggests that tarantulas were not always present in
275   Oceania, and instead is consistent with Selenocosmiinae having diversified across Asia,
276   eventually crossing the Wallace line sometime after the India / Asia collision (possibly as early
277   as 47Ma, supplementary Figure S3), while the terrestrial Thrigmopoeinae remained in India
278   (Figure 2). The second colonization event of Asia by theraphosids occurred with the common
279   ancestors to Ornithoctoninae (node 34, Figure 1), which emerged in Asia at least 20 million
280   years later (35Ma–33Ma), while Poecilotheriinae remained in India (Figure 2). DEC suggests
281   that this lineage colonized Asia even earlier, placing the ancestors to Poecilotheriinae (node 33)
282   on both Asia and the Indian Plate 72Ma–69Ma, but this seems too early to be plausible as i) it
283   predates the collision of the Indian Plate into the rest of Asia, and ii) no extant members of
284   Poecilotheriinae exist in the rest of Asia, and DEC did not predict an extinction event at this
285   node.

286   Our results indicate that both of these Asian lineages diverged while the Indian Plate was
287   still rafting towards Asia, as the collision would likely not have happened before node 25
288   diverges to yield nodes 26 and 29 (Table 2), both of which also emerge on the Indian Plate in our
289   DEC+j and BBM analyses (Figure 1 and supplementary Figure 3), while DEC also places node
290   29 on the African continent in addition to the Indian Plate. Interestingly, the two lineages also
291   appear to be ecologically divergent. Selenocosmiinae is a subfamily that almost exclusively
292   contains terrestrial members, whereas Ornithoctoninae is a largely arboreal subfamily with
293   terrestrial members.

294   A theraphosid fossil was recently uncovered in Burmese amber, dating to ~100Ma
295   (Wunderlich & Müller, 2020). While the identity of the fossil cannot adequately be established
296   (though the authors did speculate that it may represent an ancestor to Selenocosmiinae), Burmese
297   amber is thought to be of Gondwanan origin (Oliveira *et al.*, 2016; Poinar Jr, 2018). The location
298   and age of this fossil could plausibly correspond to one of the ancestral tarantula groups at either
299   node 21 or 24, and confers some support towards the validity of our ancestral area estimations
300   and dating.

301

302   *The Catumiri lineage as a biogeographic relic, and a journey to Africa for Eumenophorinae*

303   Our estimate for the age of our most ancestral theraphosid node (20) was 120Ma–105.5Ma
304   (Table 2), which is consistent with previous estimates based on fewer ingroups that dated the
305   divergence of tarantulas to occur 150Ma–71Ma (Bond *et al.*, 2014) or ~107Ma (Opatova *et al.*,
306   2020). All three analyses recover this node in the Americas, but DEC also places it on Africa.
307   South America finally split from Africa 105Ma–100Ma (Heine, Zoethout & Müller, 2013;
308   Matthews *et al.*, 2016; Olyphant, Johnson & Hughes, 2017). While oceanic crossings were still
309   required, it has been suggested that dispersal between both continents may have been facilitated
310   via the Rio Grande Rise and Walvis Ridge up until ~40Ma, as these might have shortened the
311   terrestrial distance between both continents (De Oliveira, Molina & Marroig, 2009).

312   The ancestor to the primarily African Eumenophorinae and the *Catumiri* lineage (node
313   21, Figure 1) is recovered by all analyses in the Americas 112Ma–107.5Ma (with DEC also
314   placing it in Africa) — a period when South America and Africa were likely still connected. As a
315   member of the taxonomically troubled Ischnocolinae, *Catumiri* and its relatives have been
316   considered to be among the most basal theraphosids (Raven, 1985; Schmidt, 2003; Guadanucci,
317   2014). The findings from Opatova *et al.* (2020) and Foley *et al.* (2019) seem to agree with this
318   with respect to *Catumiri*, which is recovered as an early-branching theraphosid in both works.
319   *Catumiri* was also the only neotropical taxon from Foley *et al.* (2019) that was not recovered in
320   an otherwise monophyletic group of neotropical theraphosids. With a potential maximum age of
321   just under 107.5Ma (Table 2), we suggest that the *Catumiri* lineage represents an ancient
322   biogeographical relic that diversified and remained in South America while the continents drifted
323   apart — potentially even before South America split from Africa. This could also plausibly
324   explain why Ischnocolinae (the subfamily in which *Catumiri* is included) was the only
325   paraphyletic subfamily recovered by Foley *et al.* (2019).

326   Meanwhile, the ancestor to Eumenophorinae emerges solely in Africa in all three
327   analyses 58Ma–55.5Ma. This result leaves two potential biogeographic origins for
328   Eumenophorinae: i) a vicariant event caused by the split between South America and Africa,
329   which resulted in them being separated from the *Catumiri* lineage; or ii) a dispersal event, where
330   some of the ancestors of the *Catumiri* lineage crossed from South America to Africa and
331   subsequently diversified. Per the event matrix in Table 3, the first scenario is supported by DEC
332   (supplementary Figure S4), which recovers the ancestors to both groups on both South America
333   and Africa, whereas DEC+j and BBM favor a dispersal event, which is more consistent with a
334   scenario where Eumenophorinae crossed back to Africa from South America, which may still
335   have been possible as per the results from De Oliveira, Molina, and Marroig (2009).

336   A second emergence of tarantulas on Africa occurs at the ancestor to Harpactirinae and
337   Stromatopelminae 73Ma–69.5Ma (node 30, Figure 1), having apparently crossed back via the
338   Indian Plate. The upward bound of this arrival can be attributed to the minimum age of its parent
339   node (29), which is estimated at 96Ma (Table 2), and a minimum of 69.5Ma (minimum age of
340   the ancestor to Harpactirinae and Stromatopelminae). Crossing between Africa and the Indian
341   Plate via Madagascar could have been possible up until around 90Ma to 85Ma (J. R. Ali &
342   Aitchison, 2008), and our estimates hence afford a short window for this node to cross to Africa.
343   There are a handful of extant Malagasy theraphosids (World Spider Catalog, 2020) that could
344   provide further support for this link, though sample material from these taxa was not available.
345   This "back to Afria" event would represent an intriguing departure from typical patterns
346   associated with Gondwanan vicariance.

347

348   *Tarantula radiations and lifestyle switches*

349   There are three nodes in particular (20, 35, and 37, Figure 1) where the two descendent lineages
350   demonstrate high fidelity to habitat choices. For node 20, one descendant lineage corresponds to
351   the fossorial Selenocosmiinae and Thrigmopoeinae, while the other corresponds to the arboreal
352   Poecilotheriinae and Ornithoctoninae (except *Haplopelma*, which is fossorial). Node 35 splits
353   into Harpactirinae (fossorial) and Stromatopelminae (arboreal). Node 37 splits into
354   Theraphosinae (fossorial), and Aviculariinae, Psalmopoeinae (both arboreal) and

355   Schismatothelinae (fossorial). These habitat choices are maintained throughout radiations across
356   large areas, indicating that, based on our phylogeny and sampling, tarantula lineages may be
357   demonstrating niche conservatism in showing high fidelity to lifestyle and habitat choices.

358   This idea is also relevant to our two independent lineages of tarantulas that colonized
359   Asia. After the terrestrial Selenocosmiinae diversified across Asia, there may have been an
360   opportunity for Ornithoctoninae to capitalise on an open arboreal niche, and subsequently
361   colonize Asia and further diversify while largely retaining this newly acquired arboreal
362   preference. The two lineages that did not colonize Asia / remained on the Indian Plate are also
363   ecologically divergent, with Poecilotheriinae being an arboreal subfamily and Thrigmopoeinae
364   occupying terrestrial niches, suggesting that their lifestyle differences may have facilitated their
365   co-diversification throughout the Indian Subcontinent. The divergence times imply that the
366   arboreal Poecilotheriinae could be as old as 69Ma, and the terrestrial Thrigmopoeinae could be
367   as old as 95Ma (Table 2), which implies that this co-diversification might have occurred before
368   the Indian Plate fully collided with Asia.

369   Perhaps these radiations can be attributed to an ancient switch in lifestyle that each
370   ecologically distinct subfamily to become successful by exploiting different ecological niches.
371   Although the subfamilies in the phylogeny used here are not represented by all members, we
372   expect that the addition of further taxa would verify this idea, given that the niches of most
373   higher theraphosid taxa appear to be conserved over large timescales, and that lifestyle changes
374   are observed at the base of many theraphosid radiations. With the advent of further data from
375   more representative taxa, we encourage future studies to test whether these lifestyle switches
376   directly precede evolutionary rate shifts.

377

378   **CONCLUSIONS**

379   The Indian Plate played a pivotal role in the history of Theraphosidae and facilitated the
380   colonization of Asia by two ecologically distinct tarantula lineages at least 20 million years
381   apart. Ancient tarantulas appear to have undergone several diversifications on India while it was
382   still rafting, affirming the evolutionary significance of the subcontinent. We show that tarantulas
383   demonstrate pronounced niche conservation, which could explain the ecological divergences
384   observed in the two independent Indian subcontinent lineages that colonized Asia. We propose a
385   Gondwanan origin for Theraphosidae, although their present-day distributions suggest that
386   continental rafting and geographic radiations facilitated their colonizations of subsequent
387   landmasses.

388

389   **ACKNOWLEDGEMENTS**

390   Thanks are due to Danwei Huang and Killian O' Connor for their valuable discussions during
391   this project. We are grateful to David Court for his assistance with specimen collecting.
392   Dinosaurpictures.org kindly allowed us to use their "Ancient Earth" maps to make Figure 2.

393   References

394

395

396   Abascal, F., Zardoya, R., & Telford, M. J. (2010). TranslatorX: multiple alignment of nucleotide sequences
397   guided by amino acid translations. *Nucleic acids research, 38*, W7-13.

398   Ali, J. R., & Aitchison, J. C. (2008). Gondwana to Asia: plate tectonics, paleogeography and the biological
399   connectivity of the Indian sub-continent from the Middle Jurassic through latest Eocene (166–35
400   Ma). *Earth-Science Reviews, 88*(3-4), 145-166.

401   Ali, S. S., Yu, Y., Pfosser, M., & Wetschnig, W. (2011). Inferences of biogeographical histories within
402   subfamily Hyacinthoideae using S-DIVA and Bayesian binary MCMC analysis implemented in RASP
403   (Reconstruct Ancestral State in Phylogenies). *Annals of Botany, 109*(1), 95-107.

404   Bertani, R. (2003). Release of urticating hairs by *Avicularia versicolor* (Walckenaer, 1837)(Araneae,
405   Theraphosidae). *Bull. Soc. ent. Fr, 44*, 29-92.

406   Bond, J. E., Garrison, N. L., Hamilton, C. A., Godwin, R. L., Hedin, M., & Agnarsson, I. (2014). Phylogenomics
407   resolves a spider backbone phylogeny and rejects a prevailing paradigm for orb web evolution.
408   *Current Biology, 24*(15), 1765-1771.

409   Buzatto, B. A., Haeusler, L., & Tamang, N. (2021). Trapped indoors? Long-distance dispersal in
410   mygalomorph spiders and its effect on species ranges. *Journal of Comparative Physiology A*.
411   doi:10.1007/s00359-020-01459-x

412   Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009).
413   BLAST+: architecture and applications. *BMC bioinformatics, 10*(1), 421.

414   Castresana, J. (2002). GBLOCKS: selection of conserved blocks from multiple alignments for their use in
415   phylogenetic analysis. *Copyrighted by J. Castresana, EMBL*.

416   Catalog, W. S. 2020. The World Spider Catalog, Natural History Museum Bern, online at
417   http://wsc.nmbe.ch, version 20.0, accessed on 26 November 2019.

418   Chamberland, L., McHugh, A., Kechejian, S., Binford, G. J., Bond, J. E., Coddington, J., Dolman, G., Hamilton,
419   C. A., Harvey, M. S., & Kuntner, M. (2018). From Gondwana to GAAR landia: Evolutionary history
420   and biogeography of ogre-faced spiders (Deinopis). *Journal of Biogeography, 45*(11), 2442-2457.

421   Cheng, D.-Q., & Piel, W. H. (2018). The origins of the Psechridae: Web-building lycosoid spiders. *Molecular
422   Phylogenetics and Evolution, 125*, 213-219.

423   Conti, E., Eriksson, T., Schönenberger, J., Sytsma, K. J., & Baum, D. A. (2002). Early Tertiary out-of-India
424   dispersal of Crypteroniaceae: evidence from phylogeny and molecular dating. *Evolution, 56*(10),
425   1931-1942.

426   Coyle, F. A. (1983). Aerial dispersal by mygalomorph spiderlings (Araneae, Mygalomorphae). *Journal of
427   Arachnology*, 283-286.

428   Coyle, F. A., Greenstone, M. H., Hultsch, A.-L., & Morgan, C. E. (1985). Ballooning mygalomorphs:
429   Estimates of the masses of Sphodros and Ummidia ballooners(Araneae: Atypidae, Ctenizidae).
430   *Journal of Arachnology, 13*(3), 291-296.

431   De Oliveira, F. B., Molina, E. C., & Marroig, G. (2009). Paleogeography of the South Atlantic: a route for
432   primates and rodents into the New World? In *South American primates* (pp. 55-68): Springer.

433   Drummond, A. J., & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC
434   evolutionary biology, 7*(1), 214.

435   Dunlop, J. A., Harms, D., & Penney, D. (2008). A fossil tarantula (Araneae: Theraphosidae) from Miocene
436   Chiapas amber, Mexico. *Revista ibérica de aracnología, 15*, 9-17.

437   Ebersberger, I., Strauss, S., & von Haeseler, A. (2009). HaMStR: profile hidden markov model based search
438   for orthologs in ESTs. *BMC evolutionary biology, 9*(1), 157.

439 Echeverría-Londoño, S., Enquist, B. J., Neves, D. M., Violle, C., Boyle, B., Kraft, N. J., Maitner, B. S., McGill,
440   B., Peet, R. K., & Sandel, B. (2018). Plant functional diversity and the biogeography of biomes in
441   North and South America. *Frontiers in Ecology and Evolution, 6*(DEC).
442 Eskov, K. Y., & Zonstein, S. (1990). First Mesozoic mygalomorph spiders from the Lower Cretaceous of
443   Siberia and Mongolia, with notes on the system and evolution of the infraorder Mygalomorphae
444   (Chelicerata: Araneae). *Neues Jahrbuch für Mineralogie, Geologie und Paläontologie,*
445   *Abhandlungen, 178*, 325-368.
446 Fariña, J. M., He, Q., Silliman, B. R., & Bertness, M. D. (2018). Biogeography of salt marsh plant zonation
447   on the Pacific coast of South America. *Journal of Biogeography, 45*(1), 238-247.
448 Foley, S., Lüddecke, T., Cheng, D.-Q., Krehenwinkel, H., Künzel, S., Longhorn, S. J., Wendt, I., von Wirth, V.,
449   Tänzler, R., Vences, M., & Piel, W. H. (2019). Tarantula phylogenomics: A robust phylogeny of deep
450   theraphosid clades inferred from transcriptome data sheds light on the prickly issue of urticating
451   setae evolution. *Molecular Phylogenetics and Evolution, 140*, 106573.
452   doi:https://doi.org/10.1016/j.ympev.2019.106573
453 Foley, S., Saranathan, V., & Piel, W. H. (2020). The evolution of colouration and opsins in tarantulas.
454   *Proceedings of the Royal Society B: Biological Sciences, 287*(1935), 20201688.
455   doi:10.1098/rspb.2020.1688
456 Galleti-Lima, A., & Guadanucci, J. P. L. (2018). Morphology of setae on the coxae and trochanters of
457   theraphosine spiders (Mygalomorphae: Theraphosidae). *The Journal of Arachnology, 46*(2), 214-
458   226.
459 Galleti-Lima, A., & Guadanucci, J. P. L. (2019). Comparative morphology of stridulating setae of
460   Theraphosinae (Araneae: Theraphosidae). *Zoologischer Anzeiger*.
461 Gallon, R. C. (2000). The Natural History of tarantula spiders. *Natural History, 8*, 18.
462 Garg, S., & Biju, S. (2019). New microhylid frog genus from Peninsular India with Southeast Asian affinity
463   suggests multiple Cenozoic biotic exchanges between India and Eurasia. *Scientific Reports, 9*(1),
464   1906.
465 Garrison, N. L., Rodriguez, J., Agnarsson, I., Coddington, J. A., Griswold, C. E., Hamilton, C. A., Hedin, M.,
466   Kocot, K. M., Ledford, J. M., & Bond, J. E. (2016). Spider phylogenomics: untangling the Spider Tree
467   of Life. *PeerJ, 4*, e1719.
468 Garwood, R. J., Dunlop, J. A., Selden, P. A., Spencer, A. R., Atwood, R. C., Vo, N. T., & Drakopoulos, M.
469   (2016). Almost a spider: a 305-million-year-old fossil arachnid and spider origins. *Proceedings of*
470   *the Royal Society B: Biological Sciences, 283*(1827), 20160125.
471 Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L.,
472   Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F.,
473   Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., & Regev, A. (2011). Full-length
474   transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol, 29*(7),
475   644-652. doi:https://www.doi.org/10.1038/nbt.1883
476 Guadanucci, J. P. L. (2014). Theraphosidae phylogeny: relationships of the 'Ischnocolinae' genera
477   (Araneae, Mygalomorphae). *Zoologica Scripta, 43*(5), 508-518.
478 Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D.,
479   Li, B., & Lieber, M. (2013). De novo transcript sequence reconstruction from RNA-seq using the
480   Trinity platform for reference generation and analysis. *Nature protocols, 8*(8), 1494.
481 Hedges, S. B. (2001). Afrotheria: Plate tectonics meets genomics. *Proceedings of the National Academy of*
482   *Sciences, 98*(1), 1-2.
483 Heine, C., Zoethout, J., & Müller, R. D. (2013). Kinematics of the South Atlantic rift. *arXiv preprint*
484   *arXiv:1301.2096*.

485 Hu, X., Garzanti, E., Wang, J., Huang, W., An, W., & Webb, A. (2016). The timing of India-Asia collision
486   onset – Facts, theories, controversies. *Earth-Science Reviews, 160*, 264-299.
487   doi:https://doi.org/10.1016/j.earscirev.2016.07.014
488 Joshi, J., Karanth, P. K., & Edgecombe, G. D. (2020). The out-of-India hypothesis: evidence from an ancient
489   centipede genus, Rhysida (Chilopoda: Scolopendromorpha) from the Oriental Region, and
490   systematics of Indian species. *Zoological Journal of the Linnean Society, 189*(3), 828-861.
491 Kaderka, R., Bulantová, J., Heneberg, P., & Řezáč, M. (2019). Urticating setae of tarantulas (Araneae:
492   Theraphosidae): Morphology, revision of typology and terminology and implications for
493   taxonomy. *PLOS ONE, 14*(11), e0224384-e0224384. doi:10.1371/journal.pone.0224384
494 Karanth, P. K. (2006). Out-of-India Gondwanan origin of some tropical Asian biota. *Current Science, 90*(6),
495   789-792.
496 Katoh, K., Kuma, K.-i., Toh, H., & Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple
497   sequence alignment. *Nucleic acids research, 33*(2), 511-518.
498 Klaus, K. V., & Matzke, N. J. (2019). Statistical Comparison of Trait-Dependent Biogeographical Models
499   Indicates That Podocarpaceae Dispersal Is Influenced by Both Seed Cone Traits and Geographical
500   Distance. *Systematic Biology, 69*(1), 61-75. doi:10.1093/sysbio/syz034
501 Kück, P. (2009). ALICUT: a Perlscript which cuts ALISCORE identified RSS. Version 2. Bonn (Germany):
502   Department of Bioinformatics, Zoologisches Forschungsmuseum A. Koenig (ZMFK) [cited 22nd
503   June 2018].
504 Kück, P., & Meusemann, K. (2010). FASconCAT: convenient handling of data matrices. *Molecular
505   Phylogenetics and Evolution, 56*(3), 1115-1118.
506 Kück, P., Meusemann, K., Dambach, J., Thormann, B., von Reumont, B. M., Wägele, J. W., & Misof, B.
507   (2010). Parametric and non-parametric masking of randomness in sequence alignments can be
508   improved and leads to better resolved trees. *Frontiers in zoology, 7*(1), 10.
509 Landis, M. J., Matzke, N. J., Moore, B. R., & Huelsenbeck, J. P. (2013). Bayesian Analysis of Biogeography
510   when the Number of Areas is Large. *Systematic Biology, 62*(6), 789-804.
511   doi:10.1093/sysbio/syt040
512 Lartillot, N., Lepage, T., & Blanquart, S. (2009). PhyloBayes 3: a Bayesian software package for phylogenetic
513   reconstruction and molecular dating. *Bioinformatics, 25*(17), 2286-2288.
514 Lepage, T., Bryant, D., Philippe, H., & Lartillot, N. (2007). A general comparison of relaxed molecular clock
515   models. *Molecular biology and evolution, 24*(12), 2669-2680.
516 Li, F., Shao, L., & Li, S. (2020). Tropical niche conservatism explains the Eocene migration from India to
517   Southeast Asia in ochyroceratid spiders. *Systematic Biology*.
518 Lüddecke, T., Krehenwinkel, H., Canning, G., Glaw, F., Longhorn, S. J., Tänzler, R., Wendt, I., & Vences, M.
519   (2018). Discovering the silk road: Nuclear and mitochondrial sequence data resolve the
520   phylogenetic relationships among theraphosid spider subfamilies. *Molecular Phylogenetics and
521   Evolution, 119*, 63-70.
522 Lüddecke, T., Vilcinskas, A., & Lemke, S. (2019). Phylogeny-Guided Selection of Priority Groups for Venom
523   Bioprospecting: Harvesting Toxin Sequences in Tarantulas as a Case Study. *Toxins, 11*(9), 488.
524 Marshall, L. G., Webb, S. D., Sepkoski, J. J., & Raup, D. M. (1982). Mammalian evolution and the great
525   American interchange. *Science, 215*(4538), 1351-1357.
526 Matthews, K. J., Maloney, K. T., Zahirovic, S., Williams, S. E., Seton, M., & Mueller, R. D. (2016). Global
527   plate boundary evolution and kinematics since the late Paleozoic. *Global and Planetary Change,
528   146*, 226-250.
529 Matzke, N. J. (2013). BioGeoBEARS: BioGeography with Bayesian (and likelihood) evolutionary analysis in
530   R Scripts. *R package, version 0.2, 1*, 2013.

531 McDonald-Spicer, C., Knerr, N. J., Encinas-Viso, F., & Schmidt-Lebuhn, A. N. (2019). Big data for a large
532   clade: Bioregionalization and ancestral range estimation in the daisy family (Asteraceae). *Journal*
533   *of Biogeography, 46*(2), 255-267. doi:https://doi.org/10.1111/jbi.13496
534 McGlone, M. S., Heenan, P., Millar, T., & Cieraad, E. (2018). The biogeography, origin and characteristics
535   of the vascular plant flora and vegetation of the New Zealand mountains. *Mountains, Climate and*
536   *Biodiversity*, 375.
537 McKenna, M. C. (1973). Sweepstakes, filters, corridors, Noah's arks, and beached Viking funeral ships in
538   palaeogeography. *Implications of continental drift to the earth sciences*, 293-308.
539 Misof, B., & Misof, K. (2009). A Monte Carlo approach successfully identifies randomness in multiple
540   sequence alignments: a more objective means of data exclusion. *Systematic Biology, 58*(1), 21-34.
541 Morley, R. (2018). Assembly and division of the South and South-East Asian flora in relation to tectonics
542   and climate change. *Journal of Tropical Ecology, 34*(4), 209-234.
543 Morley, R., Plath, M., Zhang, Y., & JT, L. (2016). Biotic interchange between the Indian subcontinent and
544   mainland Asia through time.
545 O'Dea, A., Lessios, H. A., Coates, A. G., Eytan, R. I., Restrepo-Moreno, S. A., Cione, A. L., Collins, L. S., De
546   Queiroz, A., Farris, D. W., & Norris, R. D. (2016). Formation of the Isthmus of Panama. *Science*
547   *advances, 2*(8), e1600883.
548 Oliveira, I. d. S., Bai, M., Jahn, H., Gross, V., Martin, C., Hammel, J. U., Zhang, W., & Mayer, G. (2016).
549   Earliest onychophoran in amber reveals Gondwanan migration patterns. *Current Biology, 26*(19),
550   2594-2601.
551 Olyphant, J. R., Johnson, R. A., & Hughes, A. N. (2017). Evolution of the Southern Guinea Plateau:
552   Implications on Guinea-Demerara Plateau formation using insights from seismic, subsidence, and
553   gravity data. *Tectonophysics, 717*, 358-371.
554 Opatova, V., Hamilton, C. A., Hedin, M., De Oca, L. M., Kral, J., & Bond, J. E. (2020). Phylogenetic
555   systematics and evolution of the spider infraorder Mygalomorphae using genomic scale data.
556   *Systematic Biology, 69*(4), 671-707.
557 Paradis, E. (2013). Molecular dating of phylogenies by likelihood methods: a comparison of models and a
558   new information criterion. *Molecular Phylogenetics and Evolution, 67*(2), 436-444.
559 Pérez-Miles, F., Postiglioni, R., Montes-de-Oca, L., Baruffaldi, L., & Costa, F. G. (2007). Mating system in
560   the tarantula spider *Eupalaestrus weijenberghi* (Thorell, 1894): evidences of monandry and
561   polygyny. *Zoology, 110*(4), 253-260.
562 Pérez-Miles, F., Costa, F. G., Toscano-Gadea, C., & Mignone, A. (2005). Ecology and behaviour of the 'road
563   tarantulas' *Eupalaestrus weijenberghi* and *Acanthoscurria suina* (Araneae, Theraphosidae) from
564   Uruguay. *Journal of Natural History, 39*(6), 483-498.
565 Philippe, H., Brinkmann, H., Lavrov, D. V., Littlewood, D. T. J., Manuel, M., Wörheide, G., & Baurain, D.
566   (2011). Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS*
567   *biology, 9*(3), e1000602.
568 Poinar Jr, G. (2018). Burmese amber: evidence of Gondwanan origin and Cretaceous dispersal. *Historical*
569   *Biology*.
570 Prentice, T. R. (1992). A new species of North American tarantula, *Aphonopelma paloma* (Araneae,
571   Mygalomorphae, Theraphosidae). *Journal of Arachnology*, 189-199.
572 Raven, R. J. (1985). The spider infraorder Mygalomorphae (Araneae): cladistics and systematics. *Bulletin*
573   *of the AMNH, 182* (article 1).
574 Ree, R. H., & Sanmartín, I. (2018). Conceptual and statistical problems with the DEC+J model of founder-
575   event speciation and its comparison with DEC via model selection. *Journal of Biogeography, 45*(4),
576   741-749. doi:https://doi.org/10.1111/jbi.13173
577 Ree, R. H., & Smith, S. A. (2008). Maximum likelihood inference of geographic range evolution by dispersal,
578   local extinction, and cladogenesis. *Systematic Biology, 57*(1), 4-14.

579 Reichling, S. B. (2000). Group dispersal in juvenile *Brachypelma vagans* (Araneae, Theraphosidae). *The*
580   *Journal of Arachnology, 28*(2), 248-251.
581 Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things).
582   *Methods in Ecology and Evolution, 3*(2), 217-223.
583 Rice, P., Longden, I., & Bleasby, A.2000. EMBOSS: the European molecular biology open software suite.
584   In: Elsevier current trends.
585 Robinson, D. F., & Foulds, L. R. (1981). Comparison of phylogenetic trees. *Mathematical biosciences, 53*(1-
586   2), 131-147.
587 Santana, R., Perez, D., Dobson, J., Panagides, N., Raven, R., Nouwens, A., Jones, A., King, G., & Fry, B.
588   (2017). Venom profiling of a population of the theraphosid spider *Phlogius crassipes* reveals
589   continuous ontogenetic changes from juveniles through adulthood. *Toxins, 9*(4), 116.
590 Schenk, J. J. (2016). Consequences of secondary calibrations on divergence time estimates. *PLOS ONE,*
591   *11*(1), e0148228.
592 Schmidt, G. (2003). *Die Vogelspinnen: Eine weltweite Übersicht* (Vol. 641): Westarp-Wissenschaften.
593 Schultz, S. A., & Schultz, M. J. (2009). *The Tarantula Keeper's Guide, 3rd Edition*: Barron's Educational
594   Series New York, NY.
595 Selden, P. A. (2002). First British Mesozoic spider, from Cretaceous amber of the Isle of Wight, southern
596   England. *Palaeontology, 45*(5), 973-983.
597 Selden, P. A., & Gall, J.-C. (1992). A Triassic mygalomorph spider from the northern Vosges, France.
598 Selden, P. A., Shcherbakov, D. E., Dunlop, J. A., & Eskov, K. Y. (2014). Arachnids from the Carboniferous of
599   Russia and Ukraine, and the Permian of Kazakhstan. *Paläontologische Zeitschrift, 88*(3), 297-307.
600 Sharma, P. P., Baker, C. M., Cosgrove, J. G., Johnson, J. E., Oberski, J. T., Raven, R. J., Harvey, M. S., Boyer,
601   S. L., & Giribet, G. (2018). A revised dated phylogeny of scorpions: Phylogenomic support for
602   ancient divergence of the temperate Gondwanan family Bothriuridae. *Molecular Phylogenetics*
603   *and Evolution, 122*, 37-45.
604 Shillington, C., & McEwen, B. (2006). Activity of juvenile tarantulas in and around the maternal burrow.
605   *The Journal of Arachnology, 34*(1), 261-266.
606 Simms, D., Cizdziel, P. E., & Chomczynski, P. (1993). TRIzol: A new reagent for optimal single-step isolation
607   of RNA. *Focus, 15*(4), 532-535.
608 Springer, M. S., Meredith, R. W., Janecka, J. E., & Murphy, W. J. (2011). The historical biogeography of
609   Mammalia. *Philosophical Transactions of the Royal Society B: Biological Sciences, 366*(1577),
610   2478-2502.
611 Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
612   phylogenies. *Bioinformatics, 30*(9), 1312-1313.
613 Starrett, J., & Hedin, M. (2007). Multilocus genealogies reveal multiple cryptic species and biogeographical
614   complexity in the California turret spider Antrodiaetus riversi (Mygalomorphae, Antrodiaetidae).
615   *Molecular Ecology, 16*(3), 583-604.
616 Udvardy, M. D. (1975). *A classification of the biogeographical provinces of the world* (Vol. 8): International
617   Union for Conservation of Nature and Natural Resources Morges.
618 West, R. C., Nunn, S. C., & Hogg, S. (2012). A new tarantula genus, *Psednocnemis*, from West Malaysia
619   (Araneae: Theraphosidae), with cladistic analyses and biogeography of Selenocosmiinae Simon
620   1889. *Zootaxa, 3299*(1), 1-43.
621 Westerweel, J., Roperch, P., Licht, A., Dupont-Nivet, G., Win, Z., Poblete, F., Ruffet, G., Swe, H. H., Thi, M.
622   K., & Aung, D. W. (2019). Burma Terrane part of the Trans-Tethyan arc during collision with India
623   according to palaeomagnetic data. *Nature Geoscience*, 1-6.
624 Wunderlich, J., & Müller, P. (2020). New and already described fossil spiders (Araneae) of 20 families in
625   Mid Cretaceous Burmese amber with notes on spider phylogeny, evolution and classification.
626   *Beitr. Araneol., 13*, 22-164.

627 Xia, X., & Xie, Z. (2001). DAMBE: software package for data analysis in molecular biology and evolution.
628   *Journal of heredity, 92*(4), 371-373.
629 Xia, X., Xie, Z., Salemi, M., Chen, L., & Wang, Y. (2003). An index of substitution saturation and its
630   application. *Molecular Phylogenetics and Evolution, 26*(1), 1-7.
631 Yáñez, M., & Floater, G. (2000). Spatial distribution and habitat preference of the endangered tarantula,
632   *Brachypelma klaasi* (Araneae: Theraphosidae) in Mexico. *Biodiversity & Conservation, 9*(6), 795-
633   810.
634 Yu, Y., Harris, A. J., Blair, C., & He, X. (2015). RASP (Reconstruct Ancestral State in Phylogenies): a tool for
635   historical biogeography. *Molecular Phylogenetics and Evolution, 87*, 46-49.
636 Zheng, Y., & Wu, F. (2018). The timing of continental collision between Indian and Asia. *Science Bulletin*.
637 Zhou, B., Tu, T., Kong, F., Wen, J., & Xu, X. (2018). Revised phylogeny and historical biogeography of the
638   cosmopolitan aquatic plant genus Typha (Typhaceae). *Scientific Reports, 8*(1), 8813.

639

Figure 1

Ancestral ranges as estimated under the DEC+j model

Subfamilies are given for each species, along with a timescale in millions of years. Node numbers are provided, and each node is color coded to correspond with the areas highlighted on the map. Letters on the map correspond to range codings given in the methods section. The stars on the tree represent independent colonizations of Asia by theraphosids, having arrived there from the Indian Subcontinent. Map credit: https://mapchart.net/, 2021. Licensed under CC BY 4.0 SA.
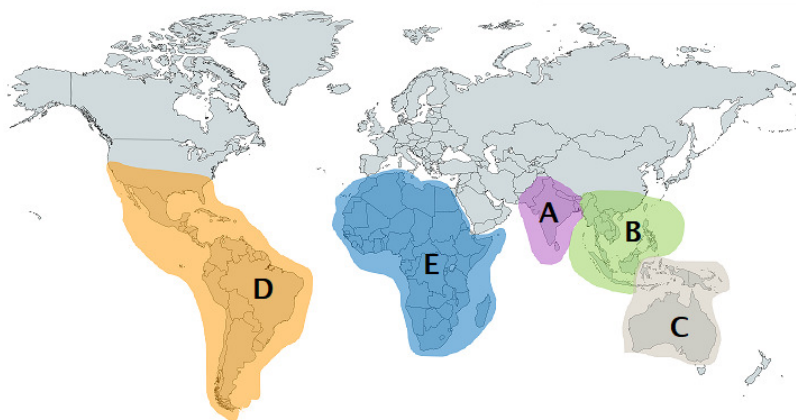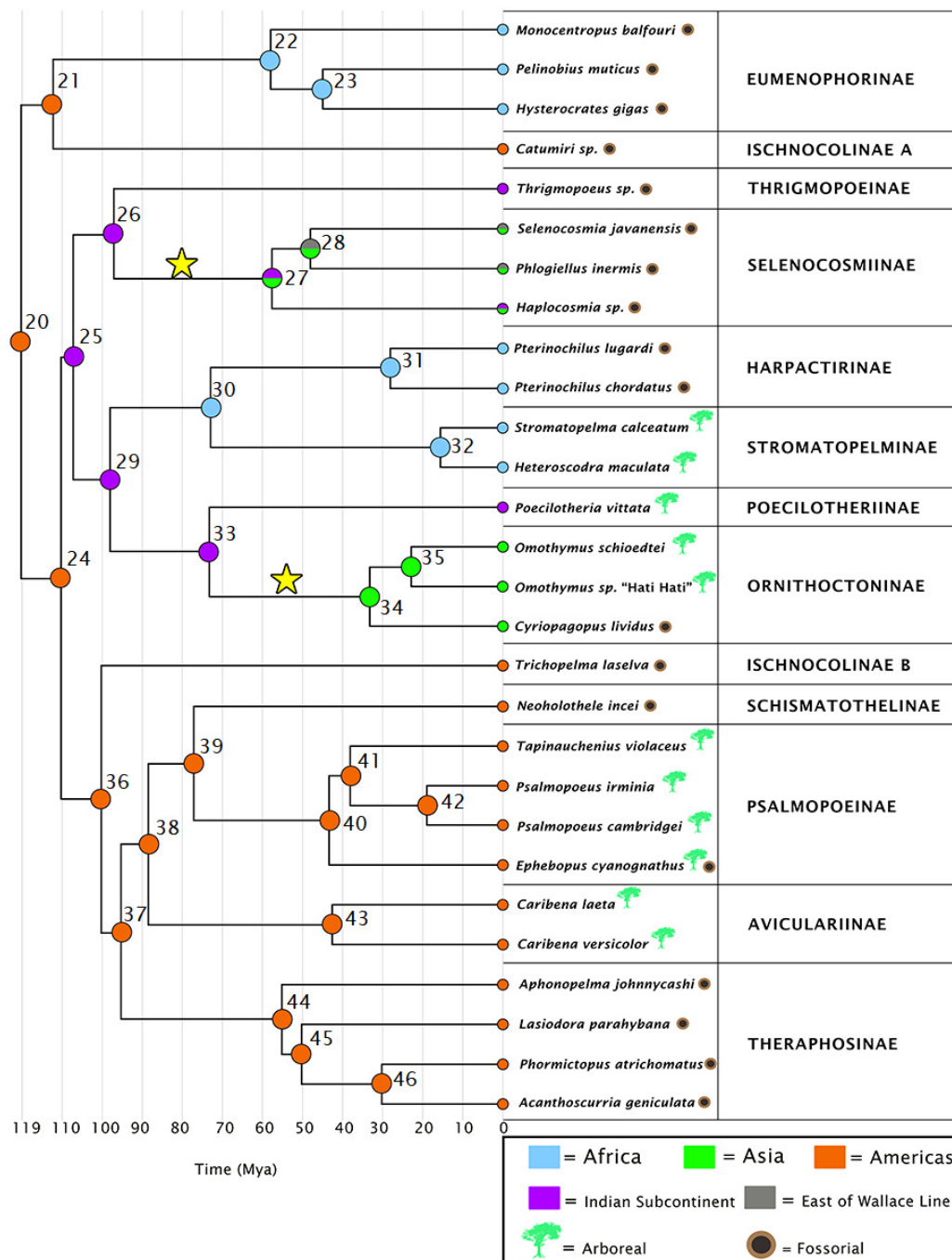
Figure 2

The dual colonization of Asia

Two independent tarantula lineages emerge out of India. a) the first theraphosids on the Indian Plate diverge into the ancestor of Selenocosmiinae and Thrigmopoeinae (~98Ma) and into the ancestor of Ornithoctoninae and Poecilotheriinae (~99Ma); b) present-day distributions of Selenocosmiinae. They entered Asia via India ~57Ma and crossed the Wallace line sometime after ~47Ma. Some Selenocosmiinae (e.g. Haplocosmia nepalensis) may also be found on the Indian Subcontinent today, where Thrigmopoeinae remain; c) present-day distributions of Ornithoctoniinae, which entered Asia via India ~34.5Ma but did not manage to cross the Wallace line. Meanwhile, Poecilotheriinae remain on the Indian Subcontinent. Map credit: https://dinosaurpictures.org/, 2021. Licensed under CC BY 4.0 SA.
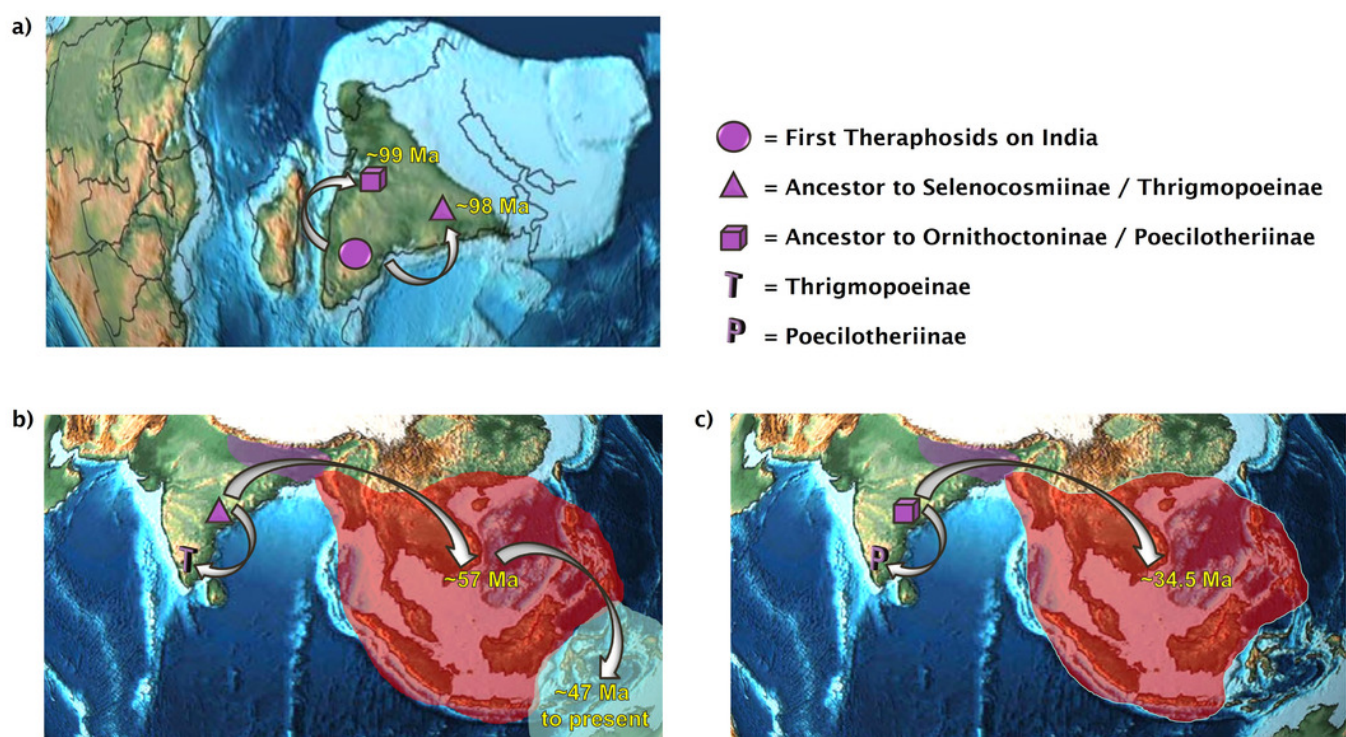
Table 1: Model testing

Model appropriateness as determined by model testing in BioGeoBEARS. While BAYAREALIKE+j emerges as the most appropriate, its assumption of a static geologic history led us to choose DEC+j instead (as indicated by bold font). Asterisks indicate a significant difference between the appropriateness of a model with and without the +j. Hence, we also include DEC in supplementary Figure S3.

Model	LnL	AICc	AICc_wt	Δ −j/+j
DEC	-34.83	74.13	0.22	
DEC+J	**-33.25**	**73.51**	**0.3**	
DIVALIKE	-38.51	81.51	0.0055	*
DIVALIKE+J	-34.18	75.35	0.12	
BAYAREALIKE	-48.13	100.7	3.70E-07	*
BAYAREALIKE+J	-33.1	73.21	0.35	

Table 2: Divergence time estimates

Divergence times with error margins per 95% HPDs as estimated by PhyloBayes. Times are given in millions of years and are rounded to the closest 0.5Ma. Node numbers as per Figure 1.

Node Number	Estimated Age	95% HPD Max	95% HPD Min
20	119	120	115.5
21	111	112	107.5
22	57	58	55.5
23	45	45.5	43.5
24	109	111	106.5
25	107	108	103.5
26	98	99	95
27	57	57.5	55
28	47	47.5	45.5
29	99	100	96
30	72	73	69.5
31	28	28.5	27
32	16	16.5	15
33	71.5	72	69
34	34.5	35	33
35	23.5	24	22
36	100	101	97
37	95	96	92.5
38	89	90	86.5
39	77	78	74.5
40	42	42.5	40.5
41	39	39.5	37.5
42	18.5	18.5	17.5
43	41.5	42	39.5
44	55.5	56	53.5
45	50.5	51	48.5
46	29.5	30	28.5

Table 3: Event matrix at node 21

The ancestor to both Eumenophorinae and the Catumiri lineage, node 21, can provide key insights into tarantula biogeography. DEC+j and BBM favor a dispersal event influencing this node, which is more consistent with members of Eumenophorinae crossing from South America to Africa. However, DEC favors vicariance, which would attribute the present-day distribution to continental drift.

Event	DEC+j	BBM	DEC
Dispersal	2	2	0
Vicariance	1	1	1
Extinction	0	0	0