Unnoticed in the tropics: phylogenomic resolution of the poorly known arachnid order Ricinulei (Arachnida)

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

| Citation       | Fernández, Rosa, and Gonzalo Giribet. 2015. “Unnoticed in the tropics: phylogenomic resolution of the poorly known arachnid order Ricinulei (Arachnida).” Royal Society Open Science 2 (6): 150065. doi:10.1098/rsos.150065. http://dx.doi.org/10.1098/rsos.150065. |
|----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Published Version | doi:10.1098/rsos.150065                                                                                                                                                                                                                                        |
| Citable link    | http://nrs.harvard.edu/urn-3:HUL.InstRepos:23845335                                                                                                                                                                                                                   |
| Terms of Use    | This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA                                                                 |
Unnoticed in the tropics: phylogenomic resolution of the poorly known arachnid order Ricinulei (Arachnida)

Rosa Fernández and Gonzalo Giribet

Museum of Comparative Zoology, Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA

Ricinulei are among the most obscure and cryptic arachnid orders, constituting a micro-diverse group with extreme endemism. The 76 extant species described to date are grouped in three genera: *Ricinoides*, from tropical Western and Central Africa, and the two Neotropical genera *Cryptocellus* and *Pseudocellus*. Until now, a single molecular phylogeny of Ricinulei has been published, recovering the African *Ricinoides* as the sister group of the American *Pseudocellus* and providing evidence for the diversification of the order predating the fragmentation of Gondwana. Here, we present, to our knowledge, the first phylogenomic study of this neglected arachnid order based on data from five transcriptomes obtained from the five major mitochondrial lineages of Ricinulei. Our results, based on up to more than 2000 genes, strongly support a clade containing *Pseudocellus* and *Cryptocellus*, constituting the American group of Ricinulei, with the African *Ricinoides* nesting outside. Our dating of the diversification of the African and American clades using a 76 gene data matrix with 90% gene occupancy indicates that this arachnid lineage was distributed in the South American, North American and African plates of Gondwana and that its diversification is concordant with a biogeographic scenario (both for pattern and tempo) of Gondwanan vicariance.

1. Introduction

Ricinulei (originally known as Cryptostemmatoidae [1]) are among the most obscure and cryptic of the arachnid orders. They are characterized by having in the anterior region of the prosoma a hinged plate, the cucullus, that acts as a hood covering the mouthparts, by a locking mechanism between the prosoma and the opisthosoma (a trait shared with trigonotarbids, an extinct lineage) that can be uncoupled during mating and egg-laying, and by a modified third leg in males for sperm transfer,
among other characters. A total of 76 living Ricinulei species are currently accepted [2,3] in three genera: *Ricinoides* Ewing, 1929 from tropical West Africa (from The Gambia to Gabon), *Cryptocellus* Westwood, 1874 from tropical South America and Central America (Guyana to Peru to Honduras), and *Pseudocellus* Platnick, 1980 from North and Central America (southern USA to Panama) [4] (figure 1).

Despite abundant recent taxonomic work (e.g. [2,3,5–10]), and some phylogenetic and biogeographic studies [11], Ricinulei remains an obscure group, as it was in 1964 when Savory [12] stated that ‘the discovery of each new specimen is still something of a zoological triumph’. Seventy-six years ago, Gertsch *et al.* [13] found the first North American Ricinulei and reported that only at 30 specimens were known for the Americas at the time. Ricinulei have remained a neglected and undersampled group of arthropods until the present, and only a few species are known from more than a handful of specimens. In *Cryptocellus*, three species are still only known by males, six by females only and two only by nymphs [5,6,8,14–21].

With an important fossil record dating back to the Carboniferous [22,23], the phylogenetic position of Ricinulei remains contentious [24]. While virtually all studies have recovered the monophyly of Euchelicerata (=Xiphosura + Arachnida), the monophyly of Arachnida is more controversial and the position of Ricinulei is still unclear, having been recovered as sister group to Acariformes + Parasitiformes [25], Parasitiformes [26], Solifugae [27] or Xiphosura (the later two hypothesis recovered in the same study but with different gene matrices [24]), or recovered as a basal group of Arachnida, excluding Acariformes [28].

As for the phylogenetic relationships within Ricinulei, their internal relationships are virtually restricted to a recent study focusing on the African species belonging to the genus *Ricinoides* [11]. Murienne *et al.* [11] explored the evolutionary relationships between the three currently recognized genera, finding that the African *Ricinoides* was sister group to the American *Pseudocellus*, therefore
suggesting that the entire diversification of this arachnid order predated the fragmentation of Gondwana. This biogeographic hypothesis had been previously proposed based on morphological data [29], and may be supported by the presence of fossil Ricinulei from Myanmar [30].

Here, we revisit the internal phylogeny of Ricinulei and present, to our knowledge, the first phylogenomic study of its three extant genera to test the possible paraphyly of the New World clade and to shed further light on the diversification of this cryptic animal group.

2. Material and methods

Seventy-nine Ricinulei specimens belonging to the three described genera were collected by sifting leaf litter, or with a Winkler apparatus (table 1). Newly sequenced specimens were collected under permit no. 17 from ARAP (Panama, 27 February 2013), no. 032 from Ministry of Scientific Research and Innovation (Republic of Cameroon, 11 March 2009) and no. 369419 from IBAMA (Brazil, 5 June 2012). We sequenced the mitochondrial marker cytochrome c oxidase subunit I (COI) to check the main mitochondrial groups in order to direct transcriptome sequencing efforts, as preliminary results suggested the existence of a high genetic variability within two of the three genera (table 1). Total DNA was extracted from a single leg of each animal using Qiagen’s DNEasy® tissue kit. The COI gene was sequenced as described in Murienne et al. [11]. The sequence-editing software GENEIOUS v. 6.1.3 [31] was used to read the chromatograms obtained from the automatic sequencer, to assemble both strands for each overlapping fragment and to edit the sequence data. Although alignment was trivial, sequences were aligned in MUSCLE through the online server of EMBL-EBI [32].

Uncorrected p-distances between each specimen were calculated and plotted in a heatmap, and maximum-likelihood (ML) and Bayesian inference (BI) phylogenetic hypotheses were generated with RAXML v. 8.0.24 [33] and MrBayes v. 3.2.3 [34] as implemented in the CIPRES Science Gateway [35]. These analyses highlighted five mitochondrial clades: Pseudocellus specimens formed a single clade with less genetic variability than Cryptocellus or Ricinoides, while the other two genera were subdivided into two clades each, exhibiting high genetic variability (see Results and discussion).

Based on these analyses, five Ricinulei specimens representing the three currently recognized genera and the phylogenetic span of the two more diverse genera (Cryptocellus becki, Cryptocellus sp. nov., Pseudocellus pearsei, Ricinoides atewa and Ricinoides karschii) were selected for transcriptomic analysis. The transcriptomes of P. pearsei and R. atewa were recently published by our laboratory [24]. Additional arachnid transcriptomes were used as outgroups [24,36] (see Data accessibility; table 2). Note that Cryptocellus sp. nov. was collected twice and therefore appears with a different MCZ catalogue numbers in the COI tree (IZ-128904) and the phylogenomic tree (IZ-30913), but they correspond to the same species. Further details can be found in MCZbase, the database of the Museum of Comparative Zoology (http://mcz.harvard.edu/collections/searchcollections.html).

Total RNA was extracted with a standard trizol-based method using TRIzol (Life Sciences). After total RNA precipitation, mRNA purification was done with the Dynabeads mRNA Purification Kit (Invitrogen) following manufacturer’s instructions. Quality of mRNA was assessed with a pico RNA assay in an Agilent 2100 Bioanalyzer (Agilent Technologies), and quantity was measured with a Qubit fluorometer (Life Technologies). cDNA libraries were constructed in the Apollo 324 automated system using the PrepX mRNA kit (Wafergen). Concentration of the cDNA libraries was measured through a dsDNA high-sensitivity (HS) assay in a Qubit fluorometer (Invitrogen). cDNA libraries were constructed in the Apollo 324 automated system using the PrepX mRNA kit (Wafergen). Concentration of the cDNA libraries was measured through a dsDNA high-sensitivity (HS) assay in a Qubit fluorometer (Invitrogen). Library quality and size selection were checked in an Agilent 2100 Bioanalyzer (Agilent Technologies) with the HS DNA assay. All samples were sequenced on an Illumina HiSeq 2500 platform with paired-end reads of 150 bp at the FAS Center for Systems Biology, Harvard University.

Demultiplexed Illumina HiSeq 2500 sequencing results, in FASTQ format, were retrieved, each sample being quality-filtered according to a threshold average quality score of 30 based on a Phred scale and adaptor trimmed using TRIMGALORE! 0.3.3 [37]. Ribosomal RNA and mitochondrial DNA were filtered out via BOWTIE v. 1.0.0 [38]. Strand specific de novo assemblies were done individually in TRINITY [39] using paired read files, a strand specificity flag and path reinforcement distance enforced to 75. Raw reads have been deposited in the National Center for Biotechnology Information Sequence Read Archive database (table 2). Redundancy reduction was done with CD-HIT-EST [40] in the raw assemblies (95% global similarity). Resulting assemblies were processed in TRANSDECODER [39] to identify candidate open-reading frames (ORFs) within the transcripts. In order to remove the variation in the coding regions of our assemblies due to alternative splicing, closely related paralogs and allelic diversity, predicted peptides were then processed with a further filter to select only one peptide per putative unigene, by...
Table 1. Specimens sequenced for the COI marker. (DNA number, MCZ voucher number, repository, species, country, locality, coordinates and GenBank accession numbers are indicated.)

| DNA no. | MCZ voucher | repository | species | country | region | latitude | longitude | accession no. COI |
|---------|-------------|------------|---------|---------|--------|----------|-----------|------------------|
| DNA107037 | IZ-130034  | Cryptocelhus becki | Brazil | Amazonas, Manaus, Reserva Florestal Adolfo Ducke | 2.934 | 59.9107 | KR180414 |
| DNA107038 | IZ-130035  | Cryptocelhus becki | Brazil | Amazonas, Manaus, Reserva Florestal Adolfo Ducke | 2.934 | 59.9707 | KR180421 |
| DNA107039 | IZ-130037  | Cryptocelhus cf. becki | Brazil | Amazonas, BR-319, Taboca, Módulo 3 de Pesquisa do PPBio, Trilha N, Parcela 1500 | 1.028 | 62.08722 | KR180410 |
| DNA107040 | IZ-130038  | Cryptocelhus iaci | Brazil | Roraima, Barreira Branca, Comunidade Caripeta, Rio Jufari, Municipalidade Caracaraí, Arquiverão da Mariu e Baixo Rio Branco, Médio Rio Negro | 1.011 | 62.1409 | KR180416 |
| DNA105542 | Cryptocelhus cf. ileepui | Ecuador | Jatun Sacha Foundation, Upper Napo River, Napo Province | 3.7888 | 69.99027 | KR180412 |
| DNA102710 | Cryptocelhus peckorum | Colombia | Track to Calderón, Km 22 N of Leticia, Departamento del Amazonas | 4.4472 | 69.9867 | JX951321 |
| DNA102711 | IZ-130028  | Cryptocelhus peckorum | Colombia | comunidad Moniaya Aman, Km 9.5 N of Leticia, Departamento del Amazonas | 4.120 | 69.92222 | KR180411 |
| DNA102698 | Cryptocelhus sp. | Costa Rica | Limon Province, Cahuita Limon, Reserva Biológica Hito Ce Cecre | 6.7617 | 83.025 | KR180405 |
| DNA102701 | Cryptocelhus sp. | Costa Rica | Limon Province, Cahuita Limon, Reserva Biológica Hito Ce Cecre | 6.7617 | 83.025 | KR180407 |
| DNA102702 | Cryptocelhus sp. | Costa Rica | Puntarenas province, Cajón, Loc. Curri, Close to River Caño Blanco | 6.7617 | 83.025 | KR180407 |
| DNA102703 | Cryptocelhus sp. | Costa Rica | Puntarenas Province, Peninsula de Osa, Agua Buena de Rincón, Fundación Neotrópica | 6.7617 | 83.025 | KR180408 |
| DNA103735 | IZ-80067    | Cryptocelhus sp. | Costa Rica | 13 km SSW Pto. Jimenez, Puntarenas | 8.40667 | 83.32833 | JX951410 |
| DNA105541 | Cryptocelhus sp. | Costa Rica | La Selva | 8.84666 | 83.59624 | KR180401 |
| GHT417  | Cryptocelhus sp. | Costa Rica | Cartago, Parque Nacional Tapantí, Macizo de la Muerte, Sendoer Natural Arboles caidos | 9.751 | 83.77626 | KR180419 |
| GHT418  | Cryptocelhus sp. | Costa Rica | Cartago, Parque Nacional Tapantí, Macizo de la Muerte, Sendoer Natural Arboles caidos | 9.751 | 83.77626 | KR180420 |
| IZ-127849 | TRS05072702L506 | Cryptocelhus sp. | French Guiana | Nouragues Field Station, XII Trail 1°30′ forest; leaf litter; Winkler sample | 4.08875 | 52.67617 | KR180413 |
| IZ-127863 | JSG061000704L507 | Cryptocelhus sp. | Guyana | Upper Takutu–Upper Essequibo: Acarai Mts, nr Rome's Camp, 264 m; 58° 56.7607′ W, 1° 23.334′ N; 7 x 2006; J. Sosa-Calvo; 1°30′ forest; leaf litter; Winkler sample | 1.3889 | 58.94612 | KR180417 |

(Continued.)
| DNA no. | MCZ voucher | repository | species          | country     | region                                         | latitude  | longitude | accession no. COI |
|---------|-------------|------------|------------------|-------------|------------------------------------------------|-----------|-----------|------------------|
| IZ-127864 | SCG06101001 |            | Cryptocellus sp. | Guyana      | Upper Takutu–Upper Essequibo; Acanal Mts, nr Romeo's Camp; 294 m; 58° 56.789′ W, 1° 23.06′ N; 10 x 2006; 1′ forest; rotten wood; Winkler sample | 1.38433   | —         | KRI180418      |
| IZ-83251  |              |            | Cryptocellus sp. | Nicaragua   | RN El Musún, 3 km NW Rio Blanco               | 12.95877  | —         | KRI180404      |
| IZ-124839 |              |            | Cryptocellus sp. | Nicaragua   | RN Cerro Musín                              | 12.95934  | —         | KRI180403      |
| IZ-124836 |              |            | Cryptocellus sp. | Nicaragua   | PN Cerro Saslay                              | 13.76867  | —         | KRI180402      |
| IZ-124835 |              |            | Cryptocellus sp. | Nicaragua   | PN Cerro Saslay                              | 13.77005  | —         | KRI180422      |
| IZ-127866/ IZ-124833 |          |            | Cryptocellus sp. | Nicaragua   | RN Kahika Creek                             | 12.67292  | —         | KRI180423      |
| DNA102709 |              |            | Cryptocellus cf. | Panama      | Prov. Chiriquí: Reserva Forestal Fortuna, Quebrada Honda, hectare PANGCODING inventory | 8.75008   | —         | KRI180409      |
| IZ-127862 |              |            | Cryptocellus cf. | Panama      | Prov. Chiriquí: Reserva Forestal Fortuna, Quebrada Honda | 8.75008   | —         | KRI180424      |
| IZ-128904.1 |             |            | Cryptocellus sp. | Panama      | Smithsonian Research Field Station, Bocas del Toro | 9.35215   | —         | KRI180408      |
| IZ-128904.2 |             |            | Cryptocellus sp. | Panama      | Smithsonian Research Field Station, Bocas del Toro | 9.35215   | —         | KRI180425      |
| IZ-89406  |              |            | Pseudocellus sp. | Guatemala   | 5 km SE Antigua                              | 14.53577  | —         | KRI180445      |
| IZ-83165  |              |            | Pseudocellus sp. | Guatemala   | Cerro Carmona, Finca El Pilar                | 14.53452  | —         | KRI180444      |
| IZ-89422  |              |            | Pseudocellus sp. | Guatemala   | 4 km S Vol. Atitlán                          | 14.54915  | —         | KRI180441      |
| IZ-89536  |              |            | Pseudocellus sp. | Guatemala   | 5 km NW Morales                              | 15.5107   | —         | KRI180439      |
| IZ-89548  |              |            | Pseudocellus sp. | Guatemala   | 5 km NW Morales                              | 15.51405  | —         | KRI180440      |
| IZ-99283  |              |            | Pseudocellus sp. | Guatemala   | Refugio El Quetzal                           | 14.56339  | —         | KRI180442      |
| IZ-98418  |              |            | Pseudocellus sp. | Honduras    | P. N. La Muralia                             | 15.09996  | —         | KRI180438      |
| IZ-98424  |              |            | Pseudocellus sp. | Honduras    | 13 km. E Nuevo Ootepeque                     | 14.45603  | —         | KRI180437      |
| IZ-99190  |              |            | Pseudocellus sp. | Honduras    | 5 km SE Antigua                              | 14.53862  | —         | KRI180449      |
| IZ-99193  |              |            | Pseudocellus sp. | Honduras    | Parque Nacional La Muralia                   | 15.09387  | —         | KRI180443      |
| DNA1033734 |              |            | Pseudocellus sp. | Honduras    |                                           | 15.58333  | —         | JX531409      |
| DNA102697 |              |            | Pseudocellus sp. | Honduras    |                                           | 15.58333  | —         | JX531409      |
| IZ-130036 |              |            | Pseudocellus gertschi | Mexico      | Estación Biológica UNAM, Los Tuxlas, Veracruz | 18.57983  | —         | KRI180436      |
| IZ-136272 |              |            | Pseudocellus monjanzi | Mexico      | Cueva de San Francisco, Municipio La Trinitaria, Chiapas | 16.09971  | —         | KRI180447      |
| IZ-136270 |              |            | Pseudocellus sbordonii | Mexico      | Dentro de la Cueva de las Abejas, Municipio San Fernando, Chiapas | 16.8487   | —         | KRI180448      |
| IZ-79891  |              |            | Pseudocellus sp. | Mexico      | 4 km SE Custepec                             | 15.71018  | —         | KRI180426      |
| IZ-79891.1 |              |            | Pseudocellus sp. | Mexico      |                                           |           |           | KRI180452      |
| IZ-79891.2 |              |            | Pseudocellus sp. | Mexico      |                                           |           |           | KRI180453      |
| IZ-79891.3 |              |            | Pseudocellus sp. | Mexico      |                                           |           |           | KRI180454      |
| IZ-79891.4 |              |            | Pseudocellus sp. | Mexico      |                                           |           |           | KRI180455      |
| IZ-79891.5 |              |            | Pseudocellus sp. | Mexico      |                                           |           |           | KRI180456      |
| IZ-79966  |              |            | Pseudocellus sp. | Mexico      | Mpio. Angel Albino Corzo, Res. Biosfera El Triunfo, Campamento El Quetzal | 15.72025  | —         | KRI180427      |
| IZ-79966.1 |              |            | Pseudocellus sp. | Mexico      |                                           |           |           | KRI180457      |
| IZ-80001  |              |            | Pseudocellus sp. | Mexico      | Mpio. Angel Albino Corzo, Reserva Biosfera El Triunfo, Campamento El Quetzal | 15.71032  | —         | KRI180428      |
| IZ-80001.1 |              |            | Pseudocellus sp. | Mexico      |                                           |           |           | KRI180458      |
| IZ-80001.2 |              |            | Pseudocellus sp. | Mexico      |                                           |           |           | KRI180459      |
| IZ-80010  |              |            | Pseudocellus sp. | Mexico      | Mpio. Angel Albino Corzo, Reserva Biosfera El Triunfo, Campamento El Quetzal | 15.72216  | —         | KRI180429      |
| IZ-80010.1 |              |            | Pseudocellus sp. | Mexico      |                                           |           |           | KRI180460      |
| IZ-80010.2 |              |            | Pseudocellus sp. | Mexico      |                                           |           |           | KRI180461      |

(Continued.)
| DNA no. | MCZ voucher | repository | species | country | region | latitude | longitude | accession no. | no. COI |
|--------|-------------|------------|---------|---------|--------|----------|-----------|--------------|--------|
| IZ-200003.3 | | | | | | | | | |
| IZ-80010.4 | | | | | | | | | |
| IZ-80010.5 | | | | | | | | | |
| IZ-80022 | Pseudocellus sp. | Mexico | Mpio. Angel Albino Corzo, Reserva Biosfera El Triunfo, Campamento El Quetzal | 15.70997 | — | 92.92994 | | KR180462 |
| IZ-80025 | Pseudocellus sp. | Mexico | Mpio. Angel Albino Corzo, Reserva Biosfera El Triunfo, Campamento El Quetzal | 15.70775 | — | 92.93121 | | KR180431 |
| IZ-80025.1 | | | | | | | | | |
| IZ-80041 | Pseudocellus sp. | Mexico | Mpio. Angel Albino Corzo, Reserva Biosfera El Triunfo, Campamento El Quetzal | 15.72178 | — | 92.94544 | | KR180432 |
| IZ-80041.1 | | | | | | | | | |
| IZ-80091 | Pseudocellus sp. | Mexico | Mpio. Angel Albino Corzo, Reserva Biosfera El Triunfo, Campamento El Quetzal | 15.71115 | — | 92.92832 | | KR180433 |
| IZ-80112 | Pseudocellus sp. | Mexico | Mpio. Angel Albino Corzo, Reserva Biosfera El Triunfo, Campamento El Quetzal | 15.72122 | — | 92.93913 | | KR180434 |
| IZ-80112.1 | | | | | | | | | |
| IZ-80243 | Pseudocellus sp. | Mexico | Mpio. Angel Albino Corzo, Reserva Biosfera El Triunfo, Campamento El Quetzal | 15.70819 | — | 92.9307 | | KR180435 |
| IZ-80243.1 | | | | | | | | | |
| DNA103736 | IZ-79799 | Pseudocellus sp. | Mexico | 3 km SE Custepec | 15.77566 | — | 92.93817 | | JX91471 |
| DNA103736.4 | IZ-79799.4 | Pseudocellus sp. | Mexico | 3 km SE Custepec | 15.77566 | — | 92.93817 | | KR180450 |
| DNA105539 | IZ-130046 | Pseudocellus boneti | Mexico | Cueva de Michapa, Town of Michapa, Morelos | 18.70278 | — | 99.49417 | | KR180446 |
| DNA105539.2 | IZ130046.2 | Pseudocellus boneti | Mexico | Cueva de Michapa, Town of Michapa, Morelos | 18.70278 | — | 99.49417 | | KR180451 |
| DNA104741 | IZ-130090 | Ricinaoides cf. olouanoua | Cameroon | Ototomo Forest, near Ngoumou, Central Province | 3.64538 | 11.29033 | | JX91472 |
| DNA104742 | IZ-130091 | Ricinaoides cf. olouanoua | Cameroon | Ototomo Forest, near Ngoumou, Central Province | 3.64447 | 11.29107 | | JX91413 |
| DNA104744 | IZ-130092 | Ricinaoides cf. olouanoua | Cameroon | Ototomo Forest, near Ngoumou, Central Province | 3.66153 | 11.30262 | | JX91415 |
| DNA104745 | IZ-130093 | Ricinaoides cf. olouanoua | Cameroon | Ototomo Forest, near Ngoumou, Central Province | 3.66195 | 11.30825 | | JX91416 |
| DNA105538 | IZ-130094 | Ricinaoides cf. olouanoua | Cameroon | Ototomo Forest, near Ngoumou, Central Province | 3.64513 | 11.29678 | | JX91419 |
| DNA104746 | IZ-130083 | Ricinaoides karschi | Cameroon | Campo Reserve, ca 25 km South of Kribi, Litoral Prov. | 2.74108 | 9.8818 | | JX91417 |
| DNA102686 | IZ-130085 | Ricinaoides cf. karschi | Equatorial Guinea | South of Ebo, P.N. de los Altos de Nso, Acombie District | 1.25278 | 11.05278 | | JX91397 |
| DNA102687 | IZ-130084 | Ricinaoides cf. karschi | Equatorial Guinea | South of Ebo, P.N. de los Altos de Nso, Acombie District | 1.25278 | 11.05278 | | JX91398 |
| DNA102682 | IZ-130082 | Ricinaoides gemmifera | Equatorial Guinea | Region Continental, P.N. de Monte Alén: Itinerario Pedagógico | 1.69806 | 10.31339 | | JX91396 |
| DNA104743 | IZ-130058 | Ricinaoides cf. karschi | Gabon | Reserve du Plateau d’Ipassa, Makokou, Ogoué-Ivindo | 0.50639 | 12.79422 | | JX91414 |
| DNA104747 | IZ-130086 | Ricinaoides cf. karschi | Gabon | Reserve du Plateau d’Ipassa, Makokou, Ogoué-Ivindo | 0.50448 | 12.79525 | | JX91418 |
| DNA102691 | AMNH LP465B | Ricinoides feae | Guinea-Bissau | 12.08156 | — | 14.80103 | | JX91399 |
choosing the longest ORF per TRINITY subcomponent with a Python script. Peptide sequences with all final candidate ORFs were retained as multifasta files. We assigned predicted ORFs into orthologous groups across all samples using OMA stand-alone v0.99y (orthologous matrix [41]). All-by-all local alignments were parallelized across 100 cores of a single compute node, implementing a custom Bash script allowing for execution of independent threads with at least 3 s between each instance of OMA to minimize risk of collisions. Further details and protocols are described elsewhere [36].

Three different amino acid supermatrices were constructed. First, a large matrix was obtained by concatenating the set of orthogroups containing eight or more taxa, yielding a supermatrix with 2177 genes (supermatrix 1: 50% gene occupancy; 568.293 amino acids). To increase gene occupancy and to reduce the percentage of missing data, a second matrix was created by selecting the orthologues contained in 13 or more taxa (supermatrix 2: 476 genes; 75% gene occupancy, 99.933 amino acids), and a third matrix was built choosing the orthologues present in 16 or more taxa (supermatrix 3: 76 genes; 90% gene occupancy; 129.19 amino acids). ML inference was conducted with PhyML-PCMA (supermatrices 2 and 3) [42] and PhyML implementing the integrated branch length option (supermatrix 3) [43]. Bootstrap support values were based on 100 replicates. We selected 20 PCs in the PhyML-optimization function, applying no filter; the supernetworks were visualized in SPLITS TREE v. 4.13.1 [47].

To discern whether compositional heterogeneity among taxa and/or within each individual orthologue alignment was affecting phylogenetic results, we further analysed supermatrices 2 and 3 (76 and 476 genes) in BACOCA v. 1.1 [45]. The relative composition frequency variability (RCFV) values (that measures the absolute deviation from the mean for each amino acid for each taxon) was plotted in a heatmap using the R package gplots with an R script modified from [45].

To investigate potential incongruence between individual gene trees, best-scoring ML trees were inferred for each gene included in each supermatrix under the PROTGAMMALG4 with RAxML v. 8.0.1 [33]. Gene trees were decomposed into quartettes with SUPERQ v. 1.1 [46] and a supernetwork assigning edge lengths based on quartette frequencies was inferred selecting the ‘balanced’ edge-weight optimization function, applying no filter; the supernetworks were visualized in SPLITSTREE v. 4.13.1 [47].
A key aspect of ricinuleid systematics is their tempo of evolution and whether it is consistent with a biogeographic scenario of Gondwanan vicariance, so we used the 76 gene dataset for dating. The fossil record of Ricinulei is impressive considering the current low diversity and restricted distribution, confined to the tropical regions of both sides of the Atlantic. Selden [23] revised the fossil ricinuleids and erected the clade Palaeoricinulei for the extinct species, limiting Neoricinulei to the extant ones. At the time, Palaeoricinulei included several Carboniferous species, the oldest being *Curculioides adompha*, from rocks of the upper Namurian B stage of the Ruhr area, Germany, while the remaining species were Westphalian in age, from the USA and the UK [23]. Subsequently, a species from fossiliferous Cretaceous

| outgroups                          | source            | MCZ acc. no. | BioProject (PRJNA) | run (SRR)  |
|-----------------------------------|-------------------|--------------|--------------------|------------|
| *Peripatopsis overbergiensis*     | Onychophora       | de novo (Illumina HiSeq) | IZ-131372  | 236 598 | 1 145 776 |
| *Scutigera coleoptrata*           | Myriapoda, Chilopoda | de novo (Illumina HiSeq) | IZ-20415  | 237 135 | 1 158 078 |
| *Metasiro americanus*             | Chelicerata, Opiliones | GenBank (Illumina GAII) | —         | 181 108 | 618 563 |
| *Centruroides vittatus*           | Chelicerata, Scorpiones | de novo (Illumina HiSeq) | —         | 236 506 | 1 146 578 |
| *Mastigoproctus giganteus*        | Chelicerata, Thelyphonida | de novo (Illumina GAII) | IZ-29741  | 236 514 | 1 145 698 |
| *Damon variegatus*                | Chelicerata, Amphiyygi | de novo (Illumina GAII) | IZ-29740  | 236 494 | 1 145 694 |
| *Limulus polyphemus*              | Chelicerata, Xiphosura | de novo (Illumina HiSeq) | IZ-29738  | 236 515 | 1 145 732 |
| *Liphistius malayanus*            | Chelicerata, Araneae | de novo (Illumina HiSeq) | IZ-29742  | 236 495 | 1 145 736 |
| *Iodes scapularis*                | Chelicerata, Parasitiformes | GenBank (whole genome) | —         | —     | —     |
| *Tetranychus urticae*             | Chelicerata, Acaniformes | GenBank (whole genome) | —         | —     | —     |
| *Synephyrus apimelus*             | Chelicerata, Pseudoscorpiones | de novo (Illumina HiSeq) | —         | 236 503 | 1 146 578 |
| *Eremobates sp.*                  | Chelicerata, Solifugae | de novo (Illumina GAII) | IZ-49755  | 236 507 | 1 146 672 |
| **Ricinulei**                     |                   |              |                    |            |
| *Pseudocellus pearsei*            |                   | de novo (Illumina HiSeq) | IZ-16426  | 236 504 | 1 146 686 |
| *Ricinoides atewa*                |                   | de novo (Illumina HiSeq) | IZ-130073 (see also IZ-130074) | 236 505 | 1 145 743 |
| *Ricinoides karschii*             |                   | de novo (Illumina HiSeq) | IZ-130083  | 281 072 | 1 972 991 |
| *Cryptocellus becki*              |                   | de novo (Illumina HiSeq) | IZ-136532 (nymph) | 281 078 | 1 979 416 |
| *Cryptocellus sp. nov.*           |                   | de novo (Illumina HiSeq) | IZ-30913 (female) | 281 669 | 1 982 218 |

Table 2. List of transcriptomes analysed in this study. (Each ricinulei specimen is hyperlinked to its entry in the MCZ database (Harvard University).)

rsos.royalsocietypublishing.org
amber of Myanmar was described [30], which has been recently constrained to the earliest Cenomanian age [48]. The age of 98.79 ± 0.62 Ma can be used as a maximum limit for the burmite (either at or after). Although described as a Palaeorcinulei, we consider that the Myanmar fossil belongs to crown-group Neorcinulei, and we use this age as a constraint for the extant taxa.

As for the outgroups, the split between Onychophora and Arthropoda was dated between 528 Ma (the minimum age for Arthropoda used by Lee et al. [49] on the basis of the earliest Rusophycus traces) and 558 Ma, used as the root of Panarthropoda [49]. The Siluro-Devonian scutigeromorph centipede Crussolum [50,51] constitutes the oldest centipede fossil. We thus apply 418 Ma to the split between Scutigera and Chelicera. We used Lunataspis aurora, considered as the oldest xiphosuran (ca 445 Ma), to date the split between Xiphosura and Arachnida [52]. The split between Scorpiones and Tetrapulmonata was dated to 312 Ma, 411 Ma for Opiliones, 308 Ma for Solifugae and 411 Ma for Acari (see a review in [22]).

Divergence dates were estimated using the Bayesian relaxed molecular clock approach as implemented in PhyloBayes v. 3.3f [54] under the autocorrelated lognormal and uncorrelated gamma models and two independent MCMC chains (10,000–12,000 cycles). For dating, we followed a recent review of the oldest occurrences of each arachnid taxon by Dunlop [22] and employed the conservative approach of using the oldest occurrence of a crown-group to constrain the split from its sister group. The calibration constraints were used with soft bounds [55] under a birth–death prior.

### 3. Results and discussion

Analysis of the COI dataset including 103 specimens clearly identifies the presence of five major Ricinulei lineages, although the COI data fail to find monophyly of Cryptocellus (figure 2a). These results, even with a much larger sampling of Neotropical species, are not too different from those presented by Murienne et al. [11]. These five lineages, however, defined the five clades for which species were selected for the subsequent phylogenomic analyses (figure 1), the focus of the remainder of the discussion.

This is, to our knowledge, the first study addressing the phylogenetic reconstruction of the order Ricinulei beyond the resolution provided by Sanger sequencing. All the recovered phylogenomic trees are concordant and clearly show a split between two major clades: one formed by the African genus Ricinoides, and a second one that includes Pseudocellus and the two Cryptocellus (figure 2a), supporting an early split of the Afrotropical and Neotropical species. By contrast, prior work [11] recovered the African Ricinoides as sister to the Neotropical Pseudocellus. From the three genera, Pseudocellus shows more homogeneity than the other two genera in the Sanger-based data analysis, while the African Ricinoides and the Neotropical Cryptocellus appear to have deep structure with two major clades each (figure 2b; [11]). However, the phylogenomic data strongly support monophyly of both Ricinoides and Cryptocellus (figure 2a) and show no conflict at the gene-tree level (figure 3).

Our results are also congruent with early vicariance during the early evolution of extant Ricinulei at the initial breakup of Gondwana. The dating analyses further corroborate the vicariance hypothesis, as we found that the split between Ricinoides and the clade formed by Pseudocellus and Cryptocellus dates back at least to the Early Cretaceous (105–195 Ma), refuting the need of transoceanic dispersal to explain their current distribution (figure 2b,c), even when considering the Myanmar Cretaceous fossils, as these are probably a sister group to the extant clade and therefore may have diverged much earlier in the Mesozoic. In the South Atlantic, ocean floor extension began within continental South America at 150 Ma, inducing a rift zone between South America and Africa. Spreading extended southward along the South Atlantic ridge with a northward propagation leading to seafloor spreading in the ‘Central’ segment by 120 Ma and in the ‘Equatorial’ segment by 110 Ma. From 100 Ma, the Middle and South Atlantic Ridges were well established and rifting in the interior of Africa ceased at about 85 Ma (figure 2b,c; [56,57]). These dates are thus concordant with our phylogenomic dating.

Cladogenesis of the Neotropical genera is slightly more recent (from the Late Cretaceous to the Middle Jurassic; 80–167 Ma), but still occurring potentially before the fragmentation of the South American, African and North American plates, reinforcing vicariance as a main force of diversification in Ricinulei (figure 2c). The development of the Caribbean is tied to the rifting of the central Atlantic during the break up of Pangea, which extended into the Caribbean during the Triassic to the Early Cretaceous. Spreading along the Central Atlantic Ridge continued into the proto-Caribbean Sea until 100 Ma [56], and the initiation of the Panama–Costa Rica Arc occurred around 80–88 Ma [58]. The reciprocal monophyly
of Cryptocellus and Pseudocellus indicates a possible vicariant model of cladogenesis between these two genera, the former predominantly South American, the latter predominantly Caribbean, Meso-American and North American. Future studies should determine the age of the diversification of Pseudocellus and its potential for understanding the palaeogeography of the Caribbean region [59].

Ricinulei constitute a poorly studied arachnid order which once had a broader distribution, including species in southeast Asia [30], but is now restricted to the tropical regions of West Africa and the Americas. Our data however show that this arachnid order has persisted largely unchanged for over 100 Myr, with a conservative phylogenetic pattern able to trace not only old continental movements, but also preserving regional information about the persistence of forests through time [11]. Similar patterns of vicariant diversification are common in other soil-dwelling and saproxylic animal groups originating in Gondwana, including velvet worms [60], centipedes [61] and caecilians [62]. Ricinulei is thus more than just another obscure animal group, and should be studied as a relictual arachnid order with the potential of providing a modern explanation to recalcitrant questions such as ancient Caribbean biogeography.

Data accessibility. All COI sequences were deposited in GenBank. The accession number for the sequence of each species is indicated in table 1. The raw data of the new transcriptomes generated for this study were deposited in the Sequence Read Archive database (SRA) of NCBI. Accession numbers are shown in table 2.

Authors’ contributions. R.F. and G.G. conceived the ideas. G.G. and several collaborators conducted fieldwork. R.F. conducted molecular work and analyses. Both authors wrote the manuscript.

Competing interests. We declare we have no competing interests.

Funding. Collection of live Ricinulei in Cameroon was supported by a Putnam Expedition grant from the MCZ; collecting in the Brazilian Amazon was supported by a CAPES grant to A. L. Tourinho. This work was supported by internal funds from the Museum of Comparative Zoology and by NSF grant no. 1144417 to G.G. and Gustavo...
Figure 3. (a) Heatmap showing the RCFV values (that measures the absolute deviation from the mean for each amino acid for each taxon) in supermatrices 2 (476 genes, right) and 3 (right, 76 genes). (b) Supernetwork visualization of individual gene trees in supermatrices 2 (right) and 3 (left). The lack of reticulation indicates no conflict between individual gene trees.

Hormiga (Collaborative Research: ARTS: Taxonomy and systematics of selected Neotropical clades of arachnids). R.F. was supported by the Fundación Ramón Areces.

Acknowledgements. We thank the many colleagues who accompanied us in different collecting trips, including Ligia Benavides, Jérome Murienne, Julián Bueno, Jesús A. Cruz López, Christopher Laumer, Ana Tourinho, Willians Porto and Pío Colmenares. Special thanks go to Jack Longino’s LLAMA project (NSF DEB-064015), which provided numerous samples. Prashant Sharma provided the de novo assemblies of outgroups. Ligia Benavides, Ana Tourinho and Gustavo Hormiga are further acknowledged for collaborating with us on ricinuleid research. The computations in this paper were run on the Odyssey cluster supported by the FAS Division of Science, Research Computing Group at Harvard University. Two anonymous reviewers provided comments that helped to improve this article.

References

1. Westwood JO. 1874 Thesaurus Entomologicus Oxoniensis; or, Illustrations of new, rare, and interesting insects, for the most part contained in the collections presented to the University of Oxford by the Rev. F.W. Hope... with forty plates from drawings by the author, p. 205. Oxford, UK: Clarendon Press.
2. Botero-Trujillo R. 2014 A new Colombian species of Cryptocellus (Arachnida, Ricinulei), with notes on the taxonomy of the genus. Zootaxa 3814, 121–132. (doi:10.11646/zootaxa.3814.1.7)
3. Tourinho AL, Lo-Man-Hung NF, Salvatierra L. 2014 A new Amazonian species of Cryptocellus (Arachnida, Ricinulei), with descriptions of its integumental structures and all free-living life stages. Zootaxa 3814, 81–95. (doi:10.11646/zootaxa.3814.1.4)
4. Harvey MS. 2003 Catalogue of the smaller arachnid orders of the World, p. 385. Melbourne, Australia: CSIRO Publishing.
5. Tourinho AL, Lo-Man-Hung NF, Bronaldo AB. 2010 A new species of Ricinulei of the genus Cryptocellus from northern Brazil. Zootaxa 2684, 63–68.
6. Tourinho AL, Saturnino R. 2010 On the Cryptocellus peckorum and Cryptocellus adsi groups, and description of a new species of Cryptocellus from Brazil (Arachnida: Ricinulei). J. Arachnol. 38, 425–432. (doi:10.1636/0046-5972-38.4.425)
7. Valdez-Mondragón A, Francke OF. 2011 Four new species of the genus Pseudocellus (Arachnida:
America in the mantle reference frame: an update. Geol. Soc. Lond. Spec. Publ. 328, 1. (doi:10.1144/SP328.1)

59. Iturralde-Vinent MA. 2006 Meso-Cenozoic Caribbean paleogeography: implications for the historical biogeography of the region. Int. Geol. Rev. 48, 791–827. (doi:10.2747/0020-6814.48.9.791).

60. Murienne J, Daniels SR, Buckley TR, Mayer G, Giribet G. 2014 A living fossil tale of Pangean biogeography. Proc. R. Soc. B 281, 20132648. (doi:10.1098/rspb.2013.2648)

61. Giribet G, Edgecombe GD. 2006 The importance of looking at small-scale patterns when inferring Gondwanan biogeography: a case study of the centipede Paralamyctes (Chilopoda, Lithobiomorpha, Henicopidae). Biol. J. Linn. Soc. 89, 65–78. (doi:10.1111/j.1095-8312.2006.00658.x)

62. San Mauro D, Gower DJ, Oommen OV, Wilkinson M, Zardoya R. 2004 Phylogeny of caecilian amphibians (Gymnophiona) based on complete mitochondrial genomes and nuclear RAG1. Mol. Phylogenet. Evol. 33, 413–427. (doi:10.1016/j.ympev.2004.05.014)