Another one bites the dust: taxonomic sampling of a key genus in phylogenomic datasets reveals more non-monophyletic groups in traditional scorpion classification

Carlos E. Santibáñez-López [A,B,D], *Andrés A. Ojanguren-Affilastro* [C] *and*
Prashant P. Sharma [A]

[A]Department of Integrative Biology, University of Wisconsin–Madison, 430 Lincoln Drive, Madison,
WI 53706, USA.
[B]Present address: Department of Biology, Eastern Connecticut State University, 83 Windham Street,
Willimantic, CT 06266, USA.
[C]División Aracnología, Museo Argentino de Ciencias Naturales 'Bernardino Rivadavia', Av. Ángel Gallardo 470,
Buenos Aires, Argentina.
[D]Corresponding author. Email: caecentrus@gmail.com

Abstract. Historically, morphological characters have been used to support the monophyly, composition, and phylogenetic relationships of scorpion families. Although recent phylogenomic analyses have recovered most of these traditional higher-level relationships as non-monophyletic, certain key taxa have yet to be sampled using a phylogenomic approach. Salient among these is the monotypic genus *Caraboctonus* Pocock, 1893, the type species of the family Caraboctonidae Kraepelin, 1905. Here, we examined the putative monophyly and phylogenetic placement of this family, sampling the library of *C. keyserlingi* Pocock, 1893 using high throughput transcriptomic sequencing. Our phylogenomic analyses recovered Caraboctonidae as polyphyletic due to the distant placement of the genera *Caraboctonus* and *Hadrurus* Thorell, 1876. *Caraboctonus* was stably recovered as the sister-group of the monotypic family Superstitioniidae Stahnke, 1940, whereas *Hadrurus* formed an unstable relationship with *Uroctonus* Thorell, 1876 and *Belisarius* Simon, 1879. Four-cluster likelihood mapping revealed that the instability inherent to the placement of *Hadrurus*, *Uroctonus* and *Belisarius* was attributable to significant gene tree conflict in the internodes corresponding to their divergences. To redress the polyphyly of Caraboctonidae, the following systematic actions have been taken: (1) the family Caraboctonidae has been delimited to consist of 23 species in the genera *Caraboctonus* and *Hadruroides* Pocock, 1893; (2) Caraboctonidae, previously included in the superfamily Iuroidea Thorell, 1876 or as *incertae sedis*, is transferred to the superfamily Caraboctonoidea (**new rank**); (3) the superfamily Hadruroidea (**new rank**) is established and the status of Hadrurinae Stahnke, 1973 is elevated to family (Hadruridae **new status**) including 9 species in the genera *Hadrurus* and *Hoffmannihadrurus* Fet & Soleglad, 2004 and (4) we treat *Uroctonus* and *Belisarius* as *insertae sedis* with respect to superfamilial placement. Our systematic actions engender the monophyly of both Iuroidea and Caraboctonidae. Future phylogenomic investigations should target similar taxon-poor and understudied lineages of potential phylogenetic significance, which are anticipated to reveal additional non-monophyletic groups.

Introduction

The advent of phylogenomics has heralded tumultuous changes in the phylogenetic relationships of scorpions. Previous phylogenetic hypotheses that were grounded in morphological cladistic analyses were already characterised by conflicting interpretations of characters and incongruent relationships (Stockwell 1989; Soleglad and Fet 2003, 2005; Prendini and Wheeler 2005). Although some relationships have been validated by phylogenomic datasets, major changes to higher-level systematics of scorpions facilitated by sampling of transcriptomic and genomic data have included: the redefinition of Buthida Soleglad and Fet, 2003 and Iurida Soleglad and Fet, 2003 as a result of placement of the minor families Pseudochactidae Gromov, 1988 and Chaerilidae Pocock, 1893 (Sharma *et al.* 2015); the resurrection of the superfamily Bothriuroidea Simon, 1880 as a distinct lineage from Scorpionoidea Latreille, 1802 (Sharma *et al.* 2015, 2018); the erection of the superfamily Superstitionioidea as distinct from the chactoids (Santibáñez-López *et al.* 2019); and the resurrection of the superfamily Vaejovoidea, with its

recognition as the sister-group of Scorpionoidea (Santibáñez-López *et al.* 2019). The ensuing systematic changes have rendered the monophyly of such higher-level groups as the parvorders Buthida and Iurida, and most of the superfamilies.

One group of scorpions that continues to defy superfamilial placement includes the erstwhile 'vaejovid' genus *Uroctonus* Thorell, 1876, the putative troglotayosicid *Belisarius xambeui* Simon, 1879 and the genus *Hadrurus* Thorell, 1876, nominally in the family Caraboctonidae Kraepelin, 1905 (Santibáñez-López *et al.* 2019). These three lineages have been recovered forming an unstable grade or possibly a clade subtending Vaejovoidea + Scorpionoidea (Santibáñez-López *et al.* 2019). Caraboctonidae constitutes a key target for expanded taxonomic sampling due to its atypical biogeographic pattern. Caraboctonidae comprises four genera and 32 species (Soleglad and Fet 2003; Prendini and Wheeler 2005; Fet and Soleglad 2008). This family has a disjunct distribution, with two genera (*Caraboctonus* Pocock, 1893 and *Hadruroides* Pocock, 1893) containing 23 species occurring in South America and two genera (*Hadrurus* and *Hoffmannihadrurus* Fet & Soleglad, 2004) containing nine species occurring in North America. Caraboctonidae includes several of the largest scorpion species worldwide (species of *Hadrurus* and *Hoffmannihadrurus* are up to 15 cm long), with most of these inhabiting deserts.

Historically, this family has been revised extensively using morphological characters, and these revisionary iterations have changed the composition, status, or the phylogenetic placement of Caraboctonidae. Stockwell (1989) included three subfamilies and six genera within family Iuridae: *Anuroctonus*, *Caraboctonus*, *Hadruroides*, *Hadrurus*, *Iurus* and *Paraiurus* (= *Calchas*). Later, in an extended morphological revision of scorpions, Soleglad and Fet (2003) raised the status of two of those three subfamilies to families (Iuridae and Caraboctonidae) and transferred *Anuroctonus* to the subfamily Uroctoninae in the Neotropical family Chactidae. In an exhaustive critique, Prendini and Wheeler (2005) rejected most of the systematic work of Soleglad and Fet (2003) and transferred *Anuroctonus* back to the subfamily Hadrurinae, which was reinstated along with subfamilies Caraboctoninae and Iurinae. These actions were reversed by Soleglad and Fet (2005) on the basis that Prendini and Wheeler's (2005) critique, while meticulously detailed and enthusiastically intensive, contributed neither data nor an original phylogenetic analysis to substantiate their preferred classification. In a different cladistic analysis, Fet and Soleglad (2008) again elevated the status of Caraboctoninae to family level (Fig. 1).

Despite these findings, phylogenomic analyses of matrices of over 5000 loci have shown that *Hadrurus* is clearly not closely related to *Iurus* (Sharma *et al.* 2015, 2018; Santibáñez-López *et al.* 2018, 2019), nor is it the sister-group of *Anuroctonus* (contra Prendini and Wheeler 2005); *Anuroctonus* has been placed with other true (i.e. non-vaejovoid) Chactoidea. Nevertheless, Caraboctonidae has been represented only by
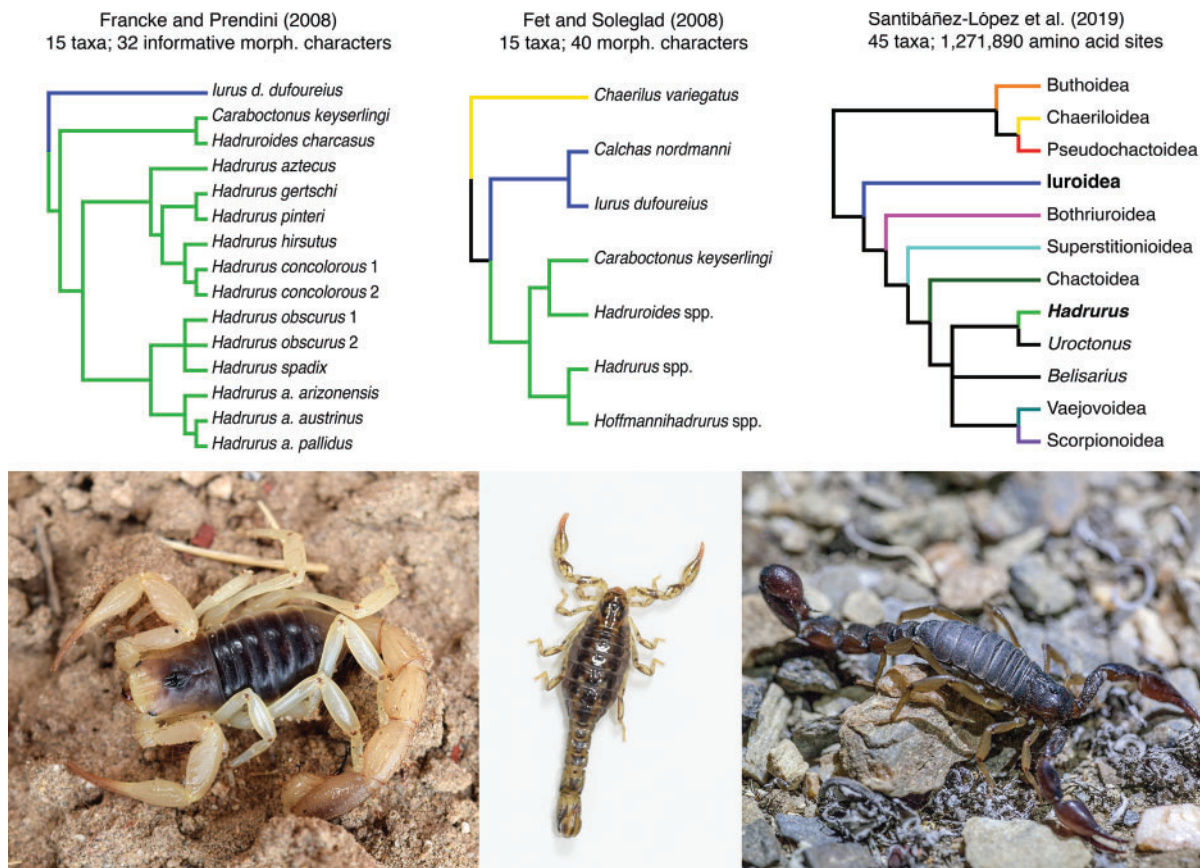

Fig. 1. Selected historical hypotheses of scorpion systematics. Colours in trees correspond to superfamilial boundaries. Photographs of live habitus in bottom row (left to right) are juvenile *Hadrurus arizonensis* (photo: C. Santibáñez), adult female of *Superstitionia donensis* (photo: C. Santibáñez) and adult male of *Caraboctonus keyserlingi* (photo: Alberto Castex).

Hadrurus arizonensis Ewing, 1928 in phylogenomic studies; other key lineages of Caraboctonidae have not been sampled using transcriptomic datasets. Salient among these is *Caraboctonus keyserlingi* Pocock, 1893, the type species of the family and sole member of its genus. This enigmatic species inhabits the lowlands of central and northern Chile. Although several morphological analyses (e.g. Soleglad and Fet 2003; Prendini and Wheeler 2005; Fet and Soleglad 2008) established the relationship of the southern caraboctonids (i.e. *Caraboctonus* and *Hadruroides*) with their northern allies (*Hadrurus* and *Hoffmannihadrurus*), no molecular study has been conducted to support the monophyly of this family as currently established.

To improve the resolution of the *Uroctonus–Belisarius– Hadrurus* soft polytomy along the backbone of the scorpion tree of life, we targeted the sampling of *Caraboctonus*, which we anticipated to be recovered as the sister-group of *Hadrurus* and stabilise this node of the phylogeny. We thus sampled specimens of *C. keyserlingi* from Chile, sequenced its transcriptome, and added it to our recent phylogenomic dataset (Santibáñez-López *et al*. 2019), with the goal of dissecting nodal stability using several criteria for matrix composition and inference of gene-tree conflict. Here, we show that Caraboctonidae is diphyletic; *C. keyserlingi* was unambiguously recovered as the sister-taxon of *Superstitionia donensis* Stahnke, 1940, whereas the position of *Hadrurus* was congruent with previous works. We therefore reassess relevant morphological characters in light of the phylogenomic tree topology and undertake the necessary taxonomic actions to redress the non-monophyly of Caraboctonidae.

Material and methods

Taxonomy and imaging

Examined specimens were deposited in the following institutions.

FMNH   The Field Museum of Natural History, Chicago, IL, USA

MACN   Museo Argentino de Ciencias Naturales 'Bernardino Rivadavia' Buenos Aires, Argentina

DMNS   Denver Museum of Nature and Science, Denver, CO, USA

Specimens were photographed using a motorised Nikon SMZ25 (Tokyo, Japan) driven by Elements software and in-built tools for focus stacking. A 405-nm (DAPI, 4′,6-diamidino-2-phenylindole) filter was used to photograph cuticular fluorescence. Photographs at MACN were taken under UV light using a Leica M205A (Wetzlar, Germany) motorised stereomicroscope and a Leica DFC295 camera; the focal planes were stacked using Helicon Focus (ver. 3.10.3, https:// www.heliconsoft.com/heliconsoft-products/helicon-focus/). The following material was examined for comparison: *Superstitionia donensis* Stahnke, 1940 (FMNH): Arizona, USA, 1 male and 1 female; *Caraboctonus keyserlingi* (MACN): Coquimbo, Chile, 1 male and 1 female; *Hadrurus arizonensis* (DMNS): Arizona, USA, 1 male; and *Uroctonus mordax* Thorell, 1876 (DMNS): California, USA, 1 male.

Taxon sampling

Living specimens of *C. keyserlingi* were collected by hand with the aid of ultraviolet lamps at night from Totoralillo in the Coquimbo Region, Chile (30°03′45″S, 71°22′03″W), on 19 March 2018 by AOA and J. Pizarro-Araya. Scorpions were vivisected into RNAlater solution (Thermofisher, Waltham, MA, USA). To avoid gut content sequencing, only the cephalothorax and the telson were dissected from these specimens into Trizol Trireagent for RNA extraction (Thermofisher). Total RNA was extracted and sequenced following our previous protocols (Sharma *et al*. 2015; Santibáñez-López *et al*. 2018). *De novo* assembly of the *C. keyserlingi* library was conducted using Trinity ver. 2.8 (Grabherr *et al*. 2011), with transcript coding sequence files processed with TransDecoder (ver. 3.0.1, see https://github.com/ TransDecoder/TransDecoder/releases/tag/v3.0.1; Haas *et al*. 2013). Other transcriptomes previously published by us (Santibáñez-López *et al*. 2019) or other original sources (e.g. Sharma *et al*. 2014, 2015; Cid-Uribe *et al*. 2018) were included in our analyses.

Orthology assessment and phylogenomic analyses

The taxon sampling comprised 47 terminals, with Iurida represented by 40 species. Outgroup taxa consisted of five buthids, one spider, and one horseshoe crab. Collecting and accession data are listed in Table S1. Transcriptomes were combined and a *de novo* homology search was conducted using the phylogenetically informed orthology criterion implemented in UPhO (ver. 1.0, see https://github.com/ ballesterus/UPhO; Ballesteros and Hormiga 2016). Initial searches used a database (consisting of the orthologous sequences from Santibáñez-López *et al*. 2019) against the library of *C. keyserlingi* using *blastp*, with clusters of homologous sequences processed using mcl (inflation parameter (i) = 6) (Dongen 2000). This strategy (query *v.* database) has proven fast and accurate in recovering orthologs for scorpion phylogenetics (Santibáñez-López *et al*. 2018). In total, 3716 clusters were produced with at least 30 species (64% of the dataset). Gene family trees (GFT) were estimated for each cluster with IQ-TREE (ver. 1.6.6, see http://www.cibiv.at/ software/iqtree; Nguyen *et al*. 2015) using the LG (general matrix) + R4 model (this model was favoured based on its efficiency and suitability in other studies; i.e. Ballesteros and Sharma 2019). From this collection of GFT, groups of orthologous sequences were identified using UPhO, enforcing the presence of 24 (m24), 35 (m35), 37 (m37), 40 (m40) or 45 (m45) species per orthogroup and a minimum branch support above 75. In-paralogs, alleles, duplicates or splice variants retained in the orthogroups were resolved in favour of the longest sequence. The individual orthogroups were aligned with MAFFT (ver. 7.0, see https://mafft.cbrc.jp/alignment/ software/; Katoh and Standley 2013), gap-masked with trimAl (ver. 1.2, see http://trimal.cgenomics.org/; Capella-Gutiérrez *et al*. 2009), and cleaned by removing sequences with fewer than 50 amino acids and less than 25% unambiguous sites (with the script *Al2Phylo.py* from UPhO). Phylogenetic inference of the orthologous gene tree (OGT) was computed with IQ-TREE, coupled with model selection of substitution and heterogeneity based on the Bayesian Information Criterion with ModelFinder (see http://www.iqtree.org/; Kalyaanamoorthy *et al*. 2017) and 1000 ultrafast bootstraps to assess branch support (Hoang *et al*. 2018; using -mset LG, JTT, WAG, -st AA -bb 1000). Subsequently, cleaned sequences recovered from the collection of OGTs were concatenated in supermatrices (five

in total) using the script *geneStitcher.py* (see https://github.com/ballesterus/Utensils/blob/master/geneStitcher.py; Ballesteros and Hormiga 2016). All matrices were partitioned by locus. Maximum likelihood (ML) analyses of the five concatenated matrices listed above were also conducted with IQ-TREE using the same parameters as above except for the precomputed best substitution models from the collection of OGTs (-spp partition_m*.nex). Species trees were estimated with ASTRAL-II (see https://github.com/smirarab/ASTRAL; Mirarab and Warnow 2015) using the collection of ML OGTs to account for potentially deleterious effects of concatenating loci. We also visualised conflicting bipartitions among the ML gene-tree topologies by constructing supernetworks using the SuperQ method (Grüenewald *et al*. 2013).

Efforts to obtain fresh tissue for sequencing from any member of genus *Hadruroides* and *Hoffmannihadrurus* were unsuccessful. Fortuitously, one mitochondrial maker (16S rRNA) was available in GenBank for *Hadruroides charcasus* (Karsch, 1879), *Hoffmannihadrurus aztecus* (Pocock, 1902) and 20 of the species in our transcriptomic dataset (including *C. keyserlingi* and *Uroctonus mordax*, see Table S2). To assess the placement of *Hadruroides* and *Hoffmannihadrurus*, the 16S rRNA partition was aligned and concatenated to 607 amino acid partitions (m40) resulting in a new matrix (m40 + 16S = 608 partitions). A GTR (generalised time reversible) + F + R4 substitution model (as selected by ModelFinder) was applied to the 16S rRNA partition; all amino acid substitution models were unchanged for the remaining 607 loci. Phylogenetic inference of this dataset was performed using IQ-TREE only.

Matrix composition and informativeness

Summary statistics from each partition, including compositional heterogeneity (relative composition frequency variability), were computed using BaCoCa (ver.1.105, https://github.com/PatrickKueck/BaCoCa; Kück and Struck 2014), with the locus length reported for each partition using the script *Alistats.py* (see https://github.com/ballesterus/UPhO/blob/master/Alistats.py; Ballesteros and Hormiga 2016), and gap content and informative characters reported by IQ-TREE. In addition, mean pair-wise sequence identity was calculated for each alignment as described by Sharma *et al*. (2014) and Ballesteros and Sharma (2019). Using four-cluster likelihood mapping (Strimmer and von Haeseler 1997), the information content of the individual gene alignments (1871 loci) and its correspondent matrix (m24) were evaluated as quartets in IQ-TREE (-lmap). Two hypotheses were tested, and for these, clusters for the quartet mapping were defined as follows: h1, testing the position of *Caraboctonus* (*Caraboctonus*, *Hadrurus*, *Superstitionia* Stahnke, 1940, *Uroctonus*); and h2, testing the position of *Hadrurus* (*Hadrurus*, *Uroctonus*, *Belisarius* and the clade Scorpionoidea + Vaejovoidea).

Results

Phylogenetic analyses

Phylogenetic orthology assessment using UPhO recovered between 214 (m45) and 1871 (m24) orthologs (Table 1). Results from the ML analyses and species-tree reconciliation of all matrices agreed with the previous higher-

level relationships recovered in an array of our previous phylogenomic investigations of scorpion relationships (Sharma *et al*. 2015, 2018; Santibáñez-López *et al*. 2018, 2019; Fig. 2). Within Iurida, the relationship of all superfamilies was consistent with previous results.

Caraboctonus keyserlingi was consistently recovered as sister group to *S. donensis* in the ML analyses of five matrices (m24 to m45; Fig. 2*A*). *Hadrurus* was recovered in one of two positions: (1) sister group to *U. mordax* with high nodal support [bootstrap support (BS) = 100%] in m24 and m35; or (2) sister group to Scorpionoidea + Vaejovoidea (m37, m40 and m45, albeit with low or no nodal support).

ML analysis of the matrix m40 + 16S recovered *C. keyserlingi* as sister species to *H. charcasus* (albeit without support; Fig. 2*B*); the remaining topology was unaffected, although bootstrap values on neighbouring nodes were depressed by the inclusion of the data-poor 16S rRNA partition. Nevertheless, support for the monophyly of a clade including *Caraboctonus*, *Superstitionia* and *Hadruroides* was supported in this analysis (BS = 89%). *Hoffmannihadrurus* was recovered by this same analysis as the sister group of *Hadrurus* with 100% nodal support, in accordance with high sequence similarity between *Hadrurus* and *Hoffmannihadrurus* 16S rRNA.

Species tree analyses of OGTs from m24 to m45 recovered *C. keyserlingi* as sister group to *S. donensis* (Fig. S1–S5). OGTs from m24 and m35 recovered the same ASTRAL-II topology (*Hadrurus* + *Uroctonus*) as their ML counterparts (Fig. S1–S2). Both ML and species tree analyses showed either consistently high or increasing nodal support (BS) with the size of the data matrix, both for the position of *Caraboctonus*, and the monophyly of the clade *Hadrurus* + *Uroctonus* (Fig. 3*A–C*). Other highly stable nodes (e.g. Scorpionoidea + Vaejovoidea) had maximal nodal support regardless of the matrix size (Fig. 3*A–C*, Fig. S3–S5). Lastly, gene tree conflict was inferred using supernetworks for two datasets: 1871 orthologs (m24) and m45 (the densest matrix, with 214 orthologs). Both supernetworks showed some conflict in the basal relationships of Iurida (Fig. S6), but with clear visualisation of support subtending all stable nodes (e.g., superfamilies).

Quartet likelihood mapping and matrix composition

The quartet likelihood mapping (QLM) of h1 (*Caraboctonus*, *Hadrurus*, *Superstitionia*, *Uroctonus*) using the m24 supermatrix recovered the quartet (*Caraboctonus* + *Superstitionia*), which was consistent with IQ-TREE and ASTRAL-II results, with 100% frequency (Q1 in Fig. 3*D, F*). Sampling of quartets across 1871 individual loci recovered a similar result, with the majority of partitions (61.3%) supporting

Table 1. Statistic summary of the Maximum likelihood analyses of seven matrices with IQ-TREE in this study

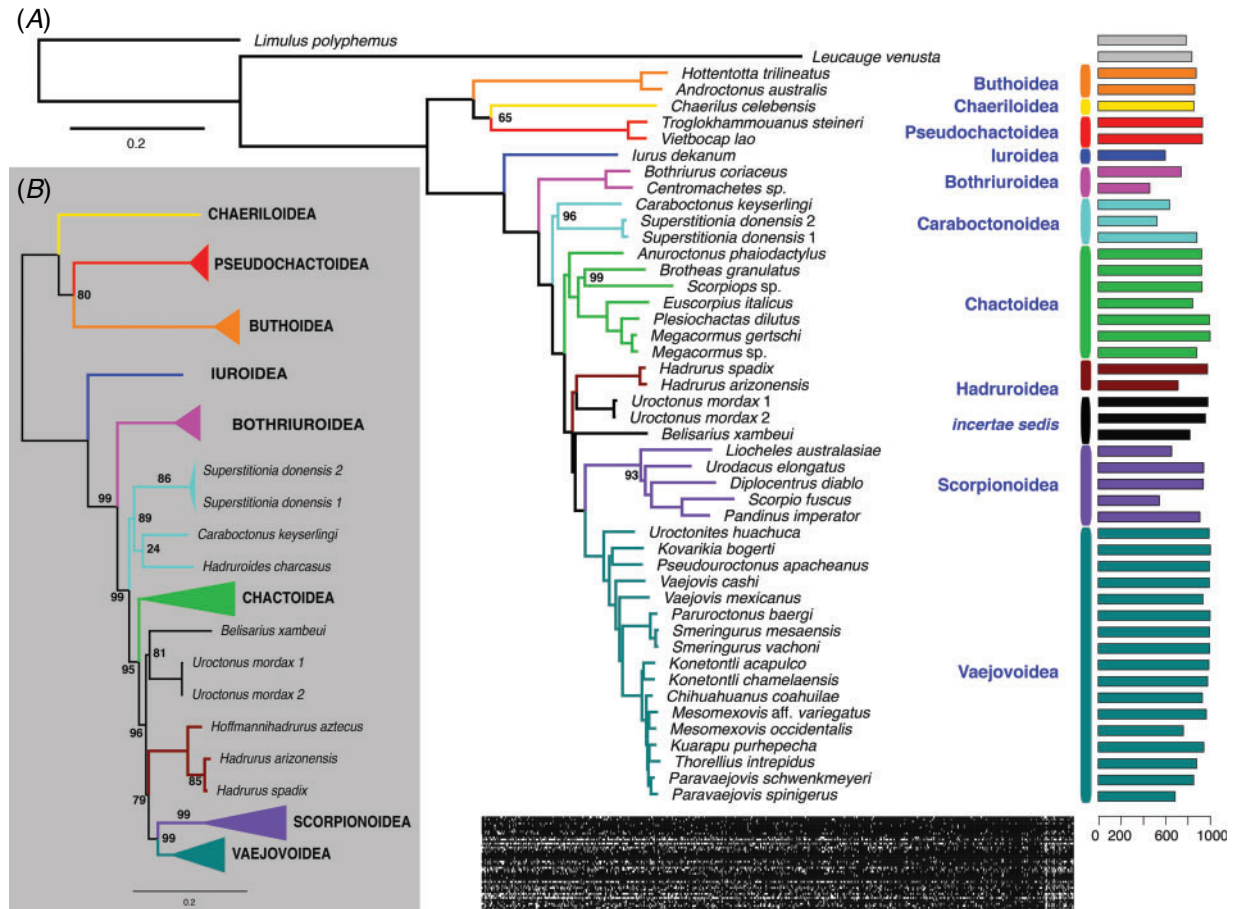| Matrix	No. partitions	Total sites	Missing data
m24	1871	1 165 325	28.07%
m35	1012	519 005	16.11%
m37	849	415 820	12.44%
m40	607	25 505	8.47%
m45	214	89 372	3.06%

Fig. 2. (*A*) Maximum likelihood (ML) tree topology recovered from the analysis of 1012 loci (m35; lnL = −8707220.6924). Numbers on nodes indicate bootstrap support; nodes without numbers were maximally supported. Bars to the right of each terminal name indicate the gene occupancy of each taxon, with an overview of matrix density shown at the bottom of the tree. (*B*) ML tree topology recovered from the analysis of 607 loci (m40) plus mitochondrial 16S rRNA. Numbers on nodes indicate bootstrap support values below 100%.

the expected topology (Fig. 3*D*, *F*). In contrast, QLM of h2 (*Hadrurus*, *Uroctonus*, *Belisarius*, Scorpionoidea + Vaejovoidea) using the m24 supermatrix recovered the clade (*Hadrurus* + *Uroctonus*) for only 78.4% of all quartets (Fig. 3*E*, *G*). Paralleling this outcome, quartet mapping of individual loci recovered only 22.9% of all quartets supporting this topology, with 21.3% supporting an alternative one (*Uroctonus* + *Belisarius*; Fig. 3*E*, *G*). These results suggest that the recovery of (*Caraboctonus* + *Superstitionia*) is strongly supported with little conflict between partitions, whereas partitions exhibit considerable conflict with respect to the relationship (*Hadrurus* + *Uroctonus*).

Discussion

Phylogenetic position and composition of Caraboctonidae

The first phylogenomic analysis of Scorpiones (Sharma *et al.* 2015) revealed discordance between the traditional morphology-based higher-level systematics and the result of phylogenomic analyses, but suffered from limited taxonomic sampling. Among the surprising incongruences revealed by that analysis was the position of Iuridae [represented by *Iurus dekanum* (Roewer,

1943)] as distantly related to Caraboctonidae (represented by *Hadurus arizonensis*); these families were traditionally grouped as the superfamily Iuroidea, based on trichobothrial patterns or cheliceral dentition (Stockwell 1989; Soleglad and Fet 2003; Prendini and Wheeler 2005; Fet and Soleglad 2008). Moreover, in a recent study, *Hadrurus*, *Uroctonus* and *Belisarius* were recovered as closely related, but were considered *incertae sedis* with respect to superfamilial placement; while clearly distinct from all superfamilies, their relationships were nevertheless unstable with respect to each other, likely owing to the short branch lengths subtending their relationships (Santibáñez-López *et al.* 2019). The sampling of additional Caraboctonidae was therefore anticipated to mitigate this putative soft polytomy.

Contrary to our expectations, our results recapitulated a recurring phenomenon in scorpion systematics – that the traditional classification of scorpions is at odds with their phylogeny. Our results clearly refute the monophyly of Caraboctonidae as currently defined, as well as its traditional placement as part of Iuroidea (Fig. 2). Rather, *C. keyserlingi* formed a clade with *S. donensis*, an outcome highly supported throughout our analyses. Although this grouping has never been
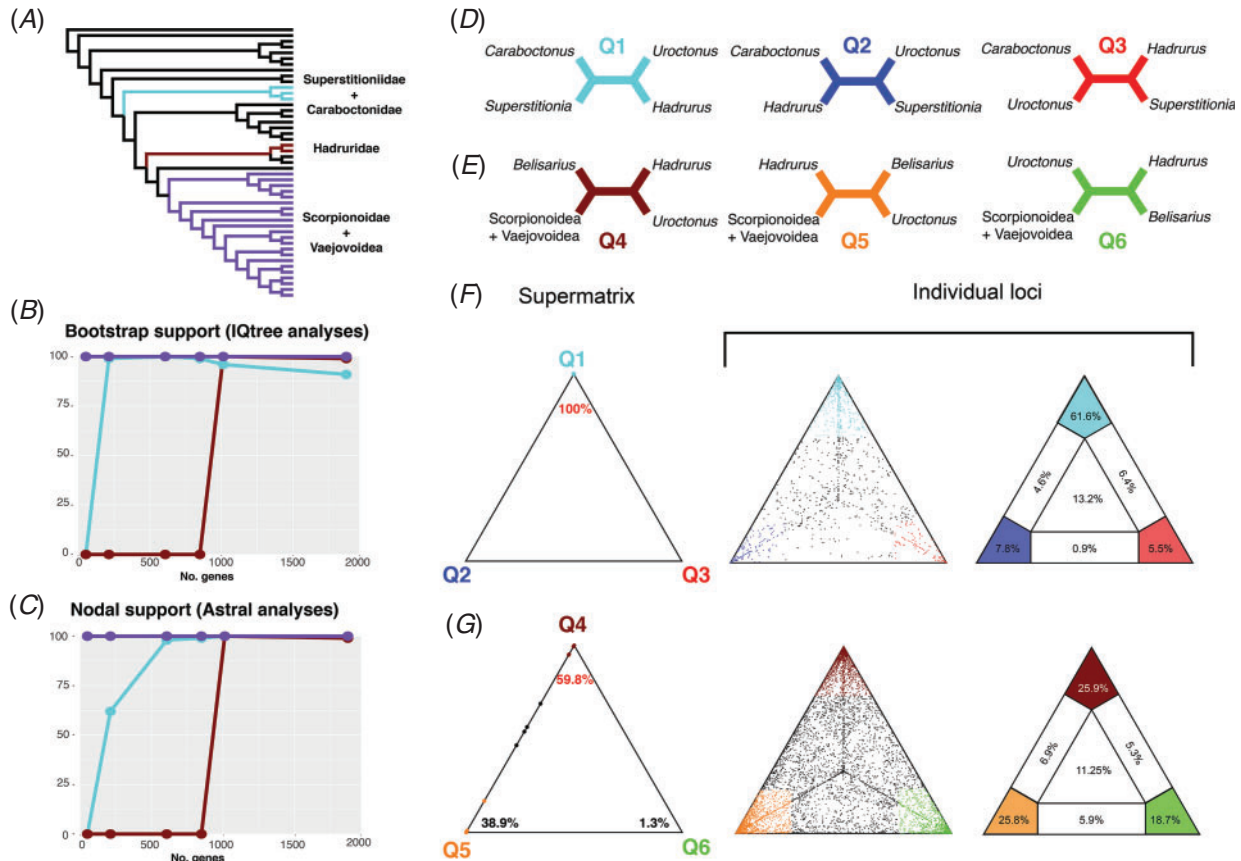
Fig. 3. (*A*) Maximum likelihood (ML) tree topology recovered from the analysis of m35 with taxon names removed and selected branches for the quartet likelihood mapping (QLM) highlighted to indicate superfamily (as shown in Fig. 2). (*B*, *C*) Distribution of branch support (expressed as bootstrap values) as a function of increasing matrix size in the ML analysis with (*B*) IQ-TREE and (*C*) local posterior probability in the species-tree reconciliation with ASTRAL-II. (*D–G*) QLM of the six alternative quartet topologies to test the position of (*D*, *F*) *Caraboctonus* and (*E*, *G*) *Hadrurus*. The right column in *F* and *G* shows the results from the concatenated matrix (1871 loci, m24) with mapping of the quartets and their respective percentage of the informative regions of the map. The centre column in *F* and *G* aggregates the mapping of 10% of the random quartets from the quartets analysed in 1871 loci, with the summary distribution of the proportion of the informative areas shown in the left columns in *F* and *G*.

suggested by morphological data, here we propose the inclusion of family Caraboctonidae (*Caraboctonus* + *Hadruroides*) as members of the superfamily Caraboctonoidea (**new superfamily assignment**, see section below) and diagnose it based on shared tarsal armature, sternum shape, trichobothrial pattern, ocular pattern and hemispermatophore bauplan. Soleglad and Fet (2003) illustrated the presence of spinule clusters in the telotarsus of *Iurus dufourensis* (Brulle, 1832) (with notes on *Calchas nordmanni* Birula, 1899), *Caraboctonus keyserlingi* (with notes on *Hadruroides charcasus*), and *Superstitionia donensis*, and contrasted them with the tarsal armature of *Hadrurus*. Based on their interpretation and our observations, iurids (sister-group to the rest of scorpions) share the median row of spinule clusters and superstitionids (as defined here), with these spinules truncated in *Iurus* (absent in adults of *Calchas nordmanni*), but not in *Caraboctonus* and *Hadruroides*, and with the clusters not as concentrated in *Superstitionia*. The sternum in all species is type II, subpentagonal but clearly wider than long (Soleglad and Fet 2003). Trichobothrial pattern is type C (Vachon 1974), with

chelal trichobothrium V1 and V2 not aligned in *Superstitionia* and *Caraboctonus*. The lateral ocelli pattern is type 4C (*sensu* Loria and Prendini 2014) and the spermatophore corresponds to the 'one fold' bauplan according to Monod *et al*. (2017).

Hadruroidea, a scorpion superfamily supported by phylotranscriptomic analyses

Despite widespread use of molecular sequence data for phylogenetic reconstruction in some scorpion genera or families (e.g. González-Santillán and Prendini 2015; Ojanguren-Affilastro *et al*. 2016; Santibáñez López *et al*. 2017; Esposito *et al*. 2018; Esposito and Prendini 2019), no molecular phylogeny has been published for the family Caraboctonidae. Morphological analyses of *Hadrurus* have consistently supported the relationship between iurids and caraboctonids, grounded in analyses of cheliceral dentition and trichobothrial patterns (Fet *et al*. 2004; Francke and Prendini 2008; Soleglad and Fet 2010; Fet and Soleglad 2008). Sharma *et al*. (2015) showed this relationship was not
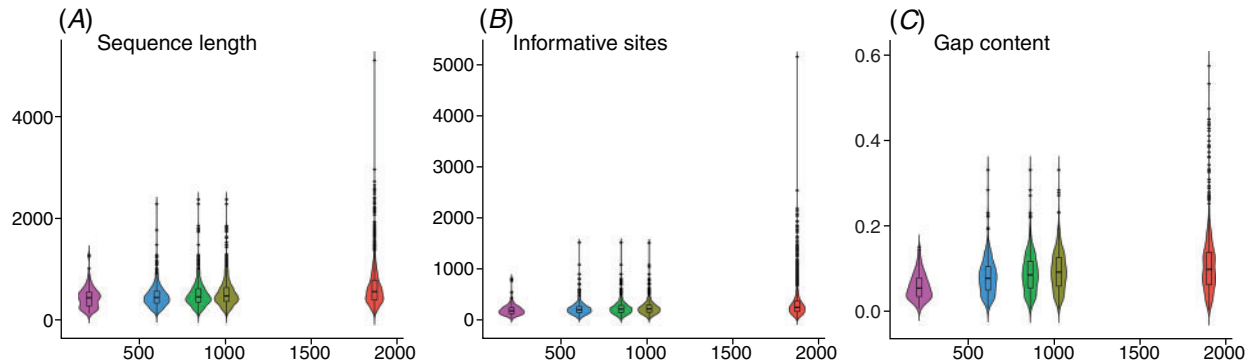
Fig. 4. Violin plots of (*A*) sequence length, (*B*) informative sites and (*C*) gap content per partition for each of five matrices.

congruent with phylotranscriptomic data and suggested that *Hadrurus* (and by extrapolation, Caraboctonidae) was part of the paraphyletic superfamily Chactoidea. Consistent with previous phylogenomic analyses, *Hadrurus* formed a clade with *Uroctonus* and was distantly related to both *Iurus* or *Caraboctonus* throughout our analyses with significant support, suggesting morphological convergence in the cheliceral dentition and trichobothrial patterns of these genera. Using denser matrices (between 214- and 849-locus datasets), *Hadrurus* was recovered as sister-group to the clade (Scorpionoidea + Vaejovoidea). Larger matrices (>1000 genes) recovered instead the clade (*Hadrurus* + *Uroctonus*) with maximal nodal support under both supermatrix and coalescent methods. Upon exploring sources of bias with BaCoCa, no partition was found to be aberrant with respect to compositional heterogeneity (data available upon request).

To assess whether the accruing support for the clade (*Hadrurus* + *Uroctonus*) was artificially driven by dataset patchiness, we examined such sources of systematic error as missing data, informativeness of sites and sequence length. Consistent with previous investigations of phylotranscriptomic datasets (e.g. Braun *et al*. 2015; Aharon *et al*. 2019), our results showed that large matrices may bear the highest proportions of missing data, but they also include the longest partitions with the most informative sites (Fig. 4*A*–*C*). In all analyses, short internodes subtending the splits between *Hadrurus*, *Belisarius* and *Uroctonus* may be indicative of a series of rapid divergences, with the ensuing possibility of incomplete lineage sorting at these internodes (Degnan and Rosenberg 2009). However, no tendency for weaker support on these branches or disagreement with the species tree analyses was found in our study, in agreement with results in other investigations (e.g. Lambert *et al*. 2015; Ballesteros and Sharma 2019).

Traditional scorpion morphology has been suggested to be prone to convergence at higher taxonomic levels, making diagnoses for these new scorpion groupings challenging (Sharma *et al*. 2015). Although the clade (*Hadrurus* + *Uroctonus*) is clearly supported, these genera have never been classified in the same superfamily. Upon scouring the literature on scorpion morphology (Stockwell 1989; Soleglad and Fet 2003, 2004, 2010; Prendini and Wheeler 2005; Fet and Soleglad 2008; Francke and Prendini 2008; Volschenk *et al*.

2008; Loria and Prendini 2014; Monod *et al*. 2017), we encountered a character pertaining to setae on legs as the most stable to unite these two genera. However, our results do not strongly support this clade; therefore, we treat *Uroctonus* and *Belisarius* as *insertae sedis* with respect to superfamilial placement. To redress the non-monophyly of Caraboctonidae, we emend the classification of scorpions by elevating the rank to a superfamily, as follows.

Taxonomy

Superfamily **CARABOCTONOIDEA** Kraepelin, 1905 (= Superstitionioidea Stanhke, 1940)

Composition

Following Santibáñez-López *et al*. (2019), we include in this superfamily two families: Caraboctonidae and Superstitioniidae.

Distribution

North America (USA, Mexico), South America (Ecuador, Peru, Chile).

Diagnosis

Carapace with median eyes raised (Fig. 5*A*, *B*), lateral ocelli pattern type 4C (*sensu* (Loria and Prendini 2014). Sternum type II, wider than long (Soleglad and Fet 2003). Trichobothrial pattern type C (Vachon 1974) (Fig. 6, 7). Spinule clusters on ventral surface of the four walking legs (Fig. 8, see also fig. 9–12 of Fet *et al*. 2004). One fold hemispermatophore bauplan (Monod *et al*. 2017).

Family **CARABOCTONIDAE** Kraepelin, 1905

Type genus

Caraboctonus Pocock, 1893, by subsequent designation Kraepelin, 1905.

Type species

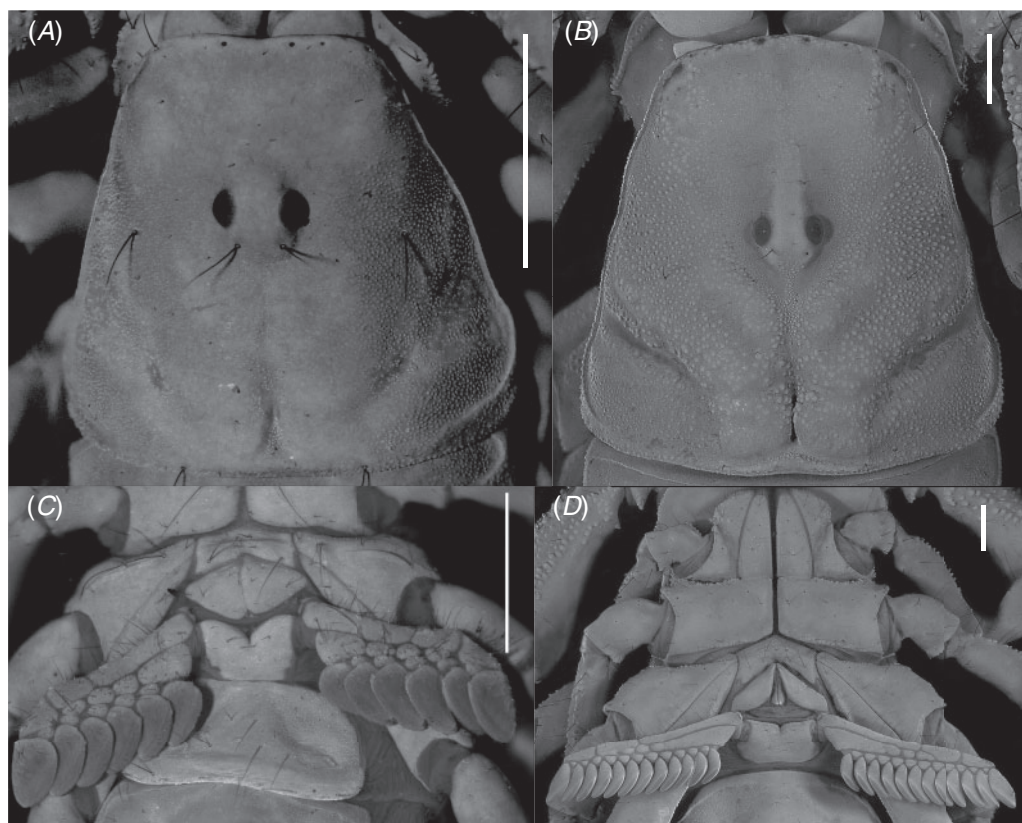*Caraboctonus keyserlingi* Pocock, 1893, by monotypy.

Fig. 5. Superfamily Caraboctonoidea. (*A, C*) Family Superstitioniidae, *Superstitionia donensis*, adult male, (*A*) carapace dorsal aspect, (*C*) sternum, genital operculum and pectines. (*B, D*) Family Caraboctonidae, *Caraboctonus keyserlingi*, adult male, (*B*) carapace dorsal aspect, (*D*) sternum, genital operculum and pectines. Scale bars: 1 mm.

Fig. 6. Superfamily Caraboctonoidea. Dextral pedipalp chela, retrolateral aspect. (*A*) Family Superstitioniidae, *Superstitionia donensis*, adult male. (*B*) Family Caraboctonidae, *Caraboctonus keyserlingi*, adult male. Trichobothria terminology follows Vachon (1974). Scale bars: 1 mm.
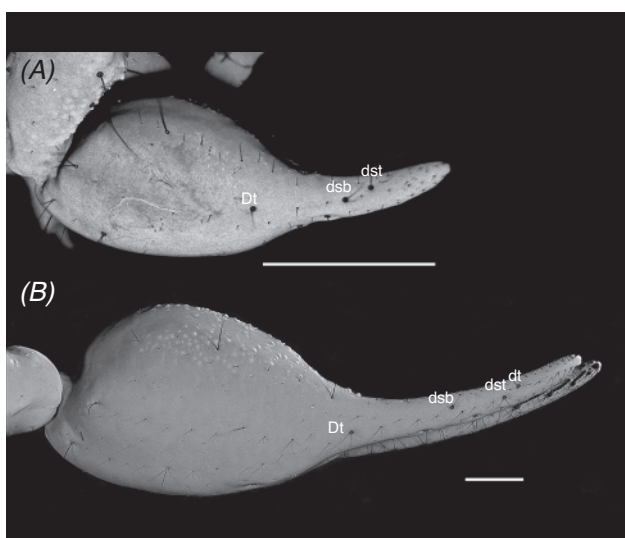
Fig. 7. Superfamily Caraboctonoidea. Dextral pedipalp chela, dorsal aspect. (*A*) Family Superstitioniidae, *Superstitionia donensis*, adult male. (*B*) Family Caraboctonidae, *Caraboctonus keyserlingi*. Trichobothria terminology follows Vachon (1974). Scale bars: 1 mm.
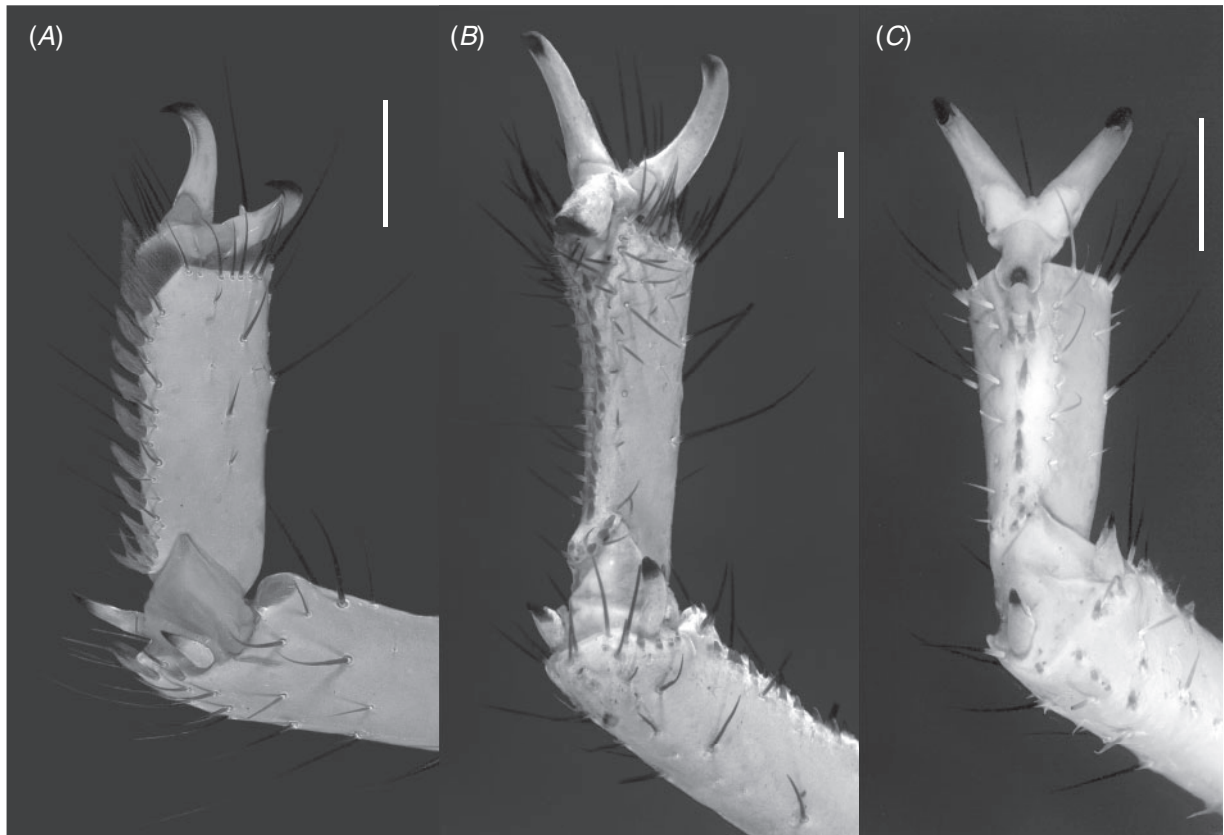
Fig. 8. Leg I, telotarsi, retroventral aspect illustrating the (*A*) spinule cluster patterning in *Caraboctonus keyserlingi* (adult female), and the single ventral row of spinules, flanked by macrosetae in (*B*) *Hadrurus hirsutus* and (*C*) *Uroctonus mordax*. Scale bars: 0.5 mm.

Composition

Following Soleglad and Fet (2003), this family includes two genera, the monotypic *Caraboctonus* and *Hadruroides* with 22 species.

Distribution

South America (Ecuador, Peru and Chile; Galapagos Islands and mainland Ecuador).

Family **SUPERSTITIONIIDAE** Stanhke, 1940

Type genus

Superstitionia Stanhke, 1940.

Type species

Superstitionia donensis Stanhke, 1940, by monotypy.

Composition

This family includes one genus and one species.

Distribution

North America (southern USA and northern Mexico).

Superfamily **HADRUROIDEA** Stahnke, 1974

Composition

Following this work, we include one family in this new taxon: Hadruridae (**new status**).

Distribution

North America (USA: Arizona, California, Nevada, Utah, Washington; Mexico: Baja California, Guerrero, Morelos, Oaxaca, Puebla).

Diagnosis

Lateral ocelli pattern type 4C (*sensu* Loria and Prendini 2014). Trichobothria pattern type 'C' (Vachon 1974). Increased neobothriotaxy on pedipalpal chela manus in genus *Hadrurus*, and on ventral surface of patella in *Hadrurus* and *Hoffmannihadrurus*. A single ventral row of spinules, flanked by macrosetae on the four walking legs (Fig. S7, see also fig. 17–20 in Fet *et al*. 2004 and fig. 12 in Soleglad and Fet 2003).

Remarks

Other characters, such as the relative position of trichobothria to each other, should be viewed sceptically until the homology of this character system is revised.

Family **HADRURIDAE** Stahnke, 1974

Type genus

Hadrurus Thorell, 1876, by subsequent designation:

Type species

Buthus hirsutus Wood, 1863 (= *Hadrurus hirsutus* (Wood, 1863)).

Composition

This family includes one family (Hadruridae) with two genera (*Hadrurus* and *Hoffmannihadrurus*).

Distribution

As for the superfamily.

Diagnosis

As for the superfamily.

Conclusion

This study resolves the status, composition and phylogenetic position of the family Caraboctonidae. New or revised taxonomic ranks proposed here redress previous non-monophyletic groups. In agreement with previous works, we established a new superfamily and raised a subfamily to family status to accommodate our results. This newly phylogenetic framework will provide a basis for future inquiries of scorpion evolution.

Conflict of interest

The authors declare that they have no conflicts of interest.

References

Aharon, S., Ballesteros, J. A., Crawford, A. R., Friske, K., Gainett, G., Langford, B., Santibáñez-López, C. E., Ya'aran, S., Gavish-Regev, E., and Sharma, P. P. (2019). The anatomy of an unstable node: a Levantine relict precipitates phylogenomic dissolution of higher-level relationships of the armored harvestmen (Arachnida: Opiliones: Laniatores). *Invertebrate Systematics* **33**, 697–717. doi:10.1071/IS19002

Ballesteros, J. A., and Hormiga, G. (2016). A new orthology assessment method for phylogenomic data: unrooted phylogenetic orthology. *Molecular Biology and Evolution* **33**, 2117–2134. doi:10.1093/molbev/msw069

Ballesteros, J. A., and Sharma, P. P. (2019). A critical appraisal of the placement of Xiphosura (Chelicerata) with account of known sources of phylogenetic error. *Systematic Biology* **68**, 1–14. doi:10.1093/sysbio/syz011

Braun, E. L., Kimball, R. T., Hosner, P. A., Faircloth, B. C., and Glenn, T. C. (2015). Avoiding missing data biases in phylogenomic inference: an empirical study in the landfowl (Aves: Galliformes). *Molecular Biology and Evolution* **33**, 1110–1125. doi:10.1093/molbev/msv347

Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973. doi:10.1093/bioinformatics/btp348

Cid-Uribe, J. I., Santibáñez-López, C. E., Meneses, E. P., Batista, C. V. F., Jiménez-Vargas, J. M., Ortiz, E., and Possani, L. D. (2018). The diversity of venom components of the scorpion species *Paravaejovis schwenkmeyeri* (Scorpiones: Vaejovidae) revealed by transcriptome and proteome analyses. *Toxicon* **151**, 47–62. doi:10.1016/j.toxicon.2018.06.085

Degnan, J. H., and Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution* **24**, 332–340. doi:10.1016/j.tree.2009.01.009

Dongen, S. (2000). 'Graph Clustering by Low Simulation.' (University of Utrecht: Utrecht, Sweden.)

Esposito, L. A., and Prendini, L. (2019). Island ancestors and New World biogeography: a case study from the scorpions (Buthidae: Centruroidinae). *Scientific Reports* **9**, 3500. doi:10.1038/s41598-018-33754-8

Esposito, L. A., Yamaguti, H. Y., Pinto-da-Rocha, R., and Prendini, L. (2018). Plucking with the plectrum: phylogeny of the New World buthid scorpion subfamily Centruroidinae Kraus, 1955 (Scorpiones: Buthidae) reveals evolution of three pecten–sternite stridulation organs. *Arthropod Systematics & Phylogeny* **76**, 87–122.

Fet, V., and Soleglad, M. E. (2008). Cladistic analysis of superfamily Iuroidea, with emphasis on subfamily Hadrurinae (Scorpiones: Iurida). *Boletin de la SEA* **43**, 255–281.

Fet, V., Soleglad, M. E., Neff, D. P. A., and Stathi, I. (2004). Tarsal armature in the superfamily Iuoridea (Scorpiones: Iurida). *Revista Iberica de Aracnologia* **10**, 17–40.

Francke, O. F., and Prendini, L. (2008). Phylogeny and classification of the giant hairy scorpions, *Hadrurus* Thorell (Iuridae Thorell): a reappraisal. *Systematics and Biodiversity* **6**, 205–223. doi:10.1017/S1477200008002697

González-Santillán, E., and Prendini, L. (2015). Phylogeny of the North American vaejovid scorpion subfamily Syntropinae Kraepelin, 1905, based on morphology, mitochondrial and nuclear DNA. *Cladistics* **31**, 341–405. doi:10.1111/cla.12091

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., and Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**, 644–652. doi:10.1038/nbt.1883

Grüenewald, S., Spillner, A., Bastkowski, S., Boegershausen, A., and Moulton, V. (2013). SuperQ: computing supernetworks from quartets. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **10**, 151–160. doi:10.1109/TCBB.2013.8

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman,

R., William, T., Dewey, C. N., Henschel, R., LeDuc, R. D., Friedman, N., and Regev, A. (2013). *De novo* transcript sequence reconstruction from RNA-Seq using the Trinity platform for reference generation and analysis. *Nature Protocols* **8**, 1494–1512. doi:10.1038/nprot.2013.084

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* **35**, 518–522. doi:10.1093/molbev/msx281

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**, 587–589. doi:10.1038/nmeth.4285

Katoh, K., and Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**, 772–780. doi:10.1093/molbev/mst010

Kück, P., and Struck, T. H. (2014). BaCoCa – a heuristic software tool for the parallel assessment of sequence biases in hundreds of gene and taxon partitions. *Molecular Phylogenetics and Evolution* **70**, 94–98. doi:10.1016/j.ympev.2013.09.011

Lambert, S. M., Reeder, T. W., and Wiens, J. J. (2015). When do species-tree and concatenated estimates disagree? An empirical analysis with higher-level scincid lizard phylogeny. *Molecular Phylogenetics and Evolution* **82**, 146–155. doi:10.1016/j.ympev.2014.10.004

Loria, S. F., and Prendini, L. (2014). Homology of the lateral eyes of scorpiones: a six-ocellus model. *PLoS One* **9**, e112913. doi:10.1371/journal.pone.0112913

Mirarab, S., and Warnow, T. (2015). ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* **31**, 44–52. doi:10.1093/bioinformatics/btv234

Monod, L., Cauwet, L., González-Santillán, E., and Huber, S. (2017). The male sexual apparatus in the order Scorpiones (Arachnida): a comparative study of functional morphology as a tool to define hypotheses of homology. *Frontiers in Zoology* **14**, 51. doi:10.1186/s12983-017-0231-z

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32**, 268–274. doi:10.1093/molbev/msu300

Ojanguren-Affilastro, A. A., Mattoni, C. I., Ochoa, J. A., Ramírez, M. J., Ceccarelli, F. S., and Prendini, L. (2016). Phylogeny, species delimitation and convergence in the South American bothriurid scorpion genus *Brachistosternus* Pocock 1893: integrating morphology, nuclear and mitochondrial DNA. *Molecular Phylogenetics and Evolution* **94**, 159–170. doi:10.1016/j.ympev.2015.08.007

Prendini, L., and Wheeler, W. C. (2005). Scorpion higher phylogeny and classification, taxonomic anarchy, and standards for peer review in online publishing. *Cladistics* **21**, 446–494. doi:10.1111/j.1096-0031.2005.00073.x

Santibáñez-López, C. E., Kriebel, R., and Sharma, P. P. (2017). *eadem figura manet*: measuring morphological convergence in diplocentrid scorpions (Arachnida: Scorpiones: Diplocentridae) under a multilocus phylogenetic framework. *Invertebrate Systematics* **31**, 233–248. doi:10.1071/IS16078

Santibáñez-López, C. E., Kriebel, R., Ballesteros, J. A., Rush, N., Witter, Z., Williams, J., Janies, D. A., and Sharma, P. P. (2018). Integration of phylogenomics and molecular modeling reveals lineage-specific diversification of toxins in scorpions. *PeerJ* **6**, e5902–e5923. doi:10.7717/peerj.5902

Santibáñez-López, C. E., González-Santillán, E., Monod, L., and Sharma, P. P. (2019). Phylogenomics facilitates stable scorpion systematics: reassessing the relationships of Vaejovidae and a new higher-level classification of Scorpiones (Arachnida). *Molecular Phylogenetics and Evolution* **135**, 22–30. doi:10.1016/j.ympev.2019.02.021

Sharma, P. P., Kaluziak, S. T., Pérez-Porro, A. R., González, V. L., Hormiga, G., Wheeler, W. C., and Giribet, G. (2014). Phylogenomic interrogation of Arachnida reveals systemic conflicts in phylogenetic signal. *Molecular Biology and Evolution* **31**, 2963–2984. doi:10.1093/molbev/msu235

Sharma, P. P., Fernández, R., Esposito, L. A., González-Santillán, E., and Monod, L. (2015). Phylogenomic resolution of scorpions reveals multilevel discordance with morphological phylogenetic signal. *Proceedings of the Royal Society B: Biological Sciences* **282**, 20142953. doi:10.1098/rspb.2014.2953

Sharma, P. P., Baker, C. M., Cosgrove, J. G., Johnson, J. E., Oberski, J. T., Raven, R. J., Harvey, M. S., Boyer, S. L., and Giribet, G. (2018). A revised dated phylogeny of scorpions: phylogenomic support for ancient divergence of the temperate Gondwanan family Bothriuridae. *Molecular Phylogenetics and Evolution* **122**, 37–45. doi:10.1016/j.ympev.2018.01.003

Soleglad, M. E., and Fet, V. (2003). High-level systematics and phylogeny of the extant scorpions (Scorpiones: Orthosterni). *Euscorpius* **2003**, 1–56. doi:10.18590/euscorpius.2003.vol2003.iss11.1

Soleglad, M. E., and Fet, V. (2004). The systematics of the scorpion subfamily Uroctoninae (Scorpiones: Chactidae). *Revista Iberica de Aracnologia* **10**, 81–128.

Soleglad, M. E., and Fet, V. (2005). Contributions to scorpion systematics I. On recent changes in high-level taxonomy. *Euscorpius* **31**, 1–13.

Soleglad, M., and Fet, V. (2010). Further observations on scorpion genera *Hadrurus* and *Hoffmannihadrurus* (Scorpiones, Caraboctonidae). *ZooKeys* **59**, 1–13. doi:10.3897/zookeys.59.546

Stockwell, S. A. (1989). Revision of the phylogeny and higher classification of scorpions (Chelicerata). Ph.D. Thesis, University of Berkeley, Berkeley, CA.

Strimmer, K., and von Haeseler, A. (1997). Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 6815–6819. doi:10.1073/pnas.94.13.6815

Vachon, M. (1974). Etude des caracteres utilises pour classer les familles et les genres de Scorpions (Arachnides). 1. La trichobothriotaxie en arachnologie. Sigles trichobothriaux et types de trichobothriotaxie chez les Scorpions. *Bulletin du Muséum National d'Histoire Naturelle* **140**, 857–958.

Volschenk, E. S., Mattoni, C. I., and Prendini, L. (2008). Comparative anatomy of the mesosomal organs of scorpions (Chelicerata, Scorpiones), with implications for the phylogeny of the order. *Zoological Journal of the Linnean Society* **154**, 651–675. doi:10.1111/j.1096-3642.2008.00426.x