Spotlight Article

Syst. Biol. 0(0):1–9, 2022
© The Author(s) 2022. Published by Oxford University Press, on behalf of the Society of Systematic Biologists. All rights reserved.
For permissions, please email: journals.permissions@oup.com
https://doi.org/10.1093/sysbio/syac021

Phylogenomics of Scorpions Reveal Contemporaneous Diversification of Scorpion Mammalian Predators and Mammal-Active Sodium Channel Toxins

CARLOS E. SANTIBÁÑEZ-LÓPEZ[1,2,*], SHLOMI AHARON[3], JESÚS A. BALLESTEROS[1], GUILHERME GAINETT[1], CAITLIN M. BAKER[1,4], EDMUNDO GONZÁLEZ-SANTILLÁN[5], MARK S. HARVEY[6], MOHAMED K. HASSAN[7], ALI HUSSEIN ABU ALMAATY[7], SHOROUK MOHAMED ALDEYARBI[7], LIONEL MONOD[8], ANDRÉS OJANGUREN-AFFILASTRO[9], RICARDO PINTO-DA-ROCHA[10], YORAM ZVIK[11], EFRAT GAVISH-REGEV[3], PRASHANT P. SHARMA[1]

[1]*Department of Integrative Biology, University of Wisconsin–Madison, Madison, WI 53706, USA;* [2]*Department of Biology, Western Connecticut State University, 181 White Street, Danbury, CT 06810, USA;* [3]*National Natural History Collections, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel;* [4]*Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA;* [5]*Colección Nacional de Arácnidos, Instituto de Biología, Universidad Nacional Autónoma de México, 3er Circuito s/n Ciudad Universitaria, Alcaldía Coyoacán, C.P. 04510 Mexico city, Mexico;* [6]*Department of Terrestrial Zoology, Western Australian Museum, Locked Bag 49, Welshpool DC 6986, Western Australia, Australia;* [7]*Zoology Department, Faculty of Science, Port Said University, Port Fouad, Egypt;* [8]*Deípartement des arthropodes et d'entomologie I, Museúm d'histoire naturelle, Route de Malagnou 1, 1208 Geneva, Switzerland;* [9]*División Arachnología, Museo Argentino de Ciencias Naturales, Av. Aíngel Gallardo 470, C1405DJR Buenos Aires, Argentina;* [10]*Departamento de Zoologia, Instituto de Biociências, Universidade de São Paulo, Rua do Matão, travessa 14, 321, São Paulo, Cep: 05508-900, Brazil; and* [11]*Hoopoe Ornithology & Ecology Center, Yeroham, Israel & Ben-Gurion University of the Negev, 8410501 Be'er-Sheva, Israel*
[*]*Correspondence to be sent to: Department of Biology, Western Connecticut State University, 181 White Street, Danbury, CT 06810, USA; E-mail: santibanezlopezc@wcsu.edu.*
Efrat Gavish-Regev and Prashant P. Sharma are the co-senior authors.

Received 23 November 2020; reviews returned 1 March 2022; accepted 5 March 2022
Associate Editor: Ken Halanych

Abstract.—Scorpions constitute a charismatic lineage of arthropods and comprise more than 2500 described species. Found throughout various tropical and temperate habitats, these predatory arachnids have a long evolutionary history, with a fossil record that began in the Silurian. While all scorpions are venomous, the asymmetrically diverse family Buthidae harbors nearly half the diversity of extant scorpions, and all but one of the 58 species that are medically significant to humans. However, the lack of a densely sampled scorpion phylogeny has hindered broader inferences of the diversification dynamics of scorpion toxins. To redress this gap, we assembled a phylogenomic data set of 100 scorpion venom gland transcriptomes and genomes, emphasizing the sampling of highly toxic buthid genera. To infer divergence times of venom gene families, we applied a phylogenomic node dating approach for the species tree in tandem with phylostratigraphic bracketing to estimate the minimum ages of mammal-specific toxins. Our analyses establish a robustly supported phylogeny of scorpions, particularly with regard to relationships between medically significant taxa. Analysis of venom gene families shows that mammal-active sodium channel toxins (NaTx) have independently evolved in five lineages within Buthidae. Temporal windows of mammal-targeting toxin origins are correlated with the basal diversification of major scorpion mammal predators such as shrews, bats, and rodents. These results suggest an evolutionary model of relatively recent diversification of buthid NaTx homologs in response to the diversification of scorpion predators. [Adaptation; arachnids; phylogenomic dating; phylostratigraphy; venom.]

Venom is a complex mixture of bioactive compounds secreted in specialized animal organs and used to disrupt biochemical and physiological processes in target organisms (Casewell et al. 2013; King and Hardy 2013). Across Metazoa, venoms have evolved over 100 times, often in tandem with specialized structures for production and delivery (Schendel et al. 2019). Among arthropods, one of the most notorious venomous groups is scorpions. Scorpion envenomation causes nearly an order of magnitude greater fatalities worldwide than snakebites, and particularly so in developing and rural subtropical regions. Intensive functional study of specific peptides has uncovered significant biomedical applications in scorpion venom, such as the identification of antimicrobial and antitumor agents, fluorescent "tumor paint," and transport molecules for molecular cargo (e.g., Veiseh et al. 2007; Rapôso 2017).

A major dimension of scorpion diversity is found in their venoms, which are rich in molecules with a broad array of biological targets, such as ion channel toxins (Na^+, K^+, Cl^-, and Ca^{2+}), mucopolysaccharides, and enzymes (Froy et al. 1999; Possani et al. 1999; Cao et al. 2013). All scorpions are venomous and thought to possess insect-targeting ion channel toxins, which facilitate prey capture (e.g., Gordon et al. 2007). These toxins are stabilized by various types of folds, such as the cysteine-stabilized α-helix and β-sheet (CSαβ), the disulfide-directed beta-hairpin (DDH), and the inhibitor cystine knot (ICK). Intriguingly, both the extant diversity of scorpions, as well as the toxicity of their venom to mammals, is asymmetrically distributed. Of the nearly 2500 described scorpion species, ca. 1200 are members of Buthidae ("thick-tailed scorpions"; Fig. 1), one of 22 extant scorpion families. Buthids include nearly all significantly venomous scorpion species, as well most of the known molecular diversity of scorpion venom (Santibáñez-López et al. 2019a). Salient components of the buthid venom cocktail are ion channel toxins that

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syac021/6554541 by University of Wisconsin-Madison Libraries user on 28 March 2022

FIGURE 1. Exemplars of scorpion diversity. Top row (Buthidae): *Buthus israelensis* and *Androctonus crassicauda* (R. Livne); *Buthacus leptochelys* (J. Ove Rein); *Centruroides meisei* and *Tityus serrulatus* (B. Myers). Bottom row (Iurida): *Belisarius xambeui* (G. Giribet); *Palaeocheloctonus pauliani* and *Opistophthalmus carinatus* (J. Ove Rein); *Hadrurus obscurus* and *Anuroctonus bajae* (C. Santibáñez-López). All photographs published with permission.

specifically target mammalian ion channels; these are inferred to function as antipredator deterrents (Niermann et al. 2020). Such neurotoxins operate by blocking action potentials at nerve synapses, precipitating symptoms of neurotoxicosis such as intense pain, hypersalivation, muscle spasms, asphyxia, and paralysis (Wang et al. 2011). Groups like *Androctonus, Buthus, Centruroides, Hottentotta, Leiurus, Parabuthus*, and *Tityus* all contain multiple highly toxic and medically significant species known for the potency of their neurotoxins (Santos et al. 2016; Niermann et al. 2020).

Surprisingly, evolutionary relationships of medically relevant scorpions remain poorly understood. The higher-level molecular phylogeny of buthids was first inferred using a 296-bp fragment of 16S rRNA sampling 17 species, with limited resolution of basal relationships (Fet et al. 2003). Most subsequent molecular phylogenetic studies of buthids used Sanger-sequenced data sets to address relationships of derived groups, often within specific geographic terranes (e.g., Ojanguren-Affilastro et al. 2017; Suranse et al. 2017). Mitochondrial phylogenies of medically significant buthids have not yielded support for basal relationships (Borges and Graham 2014), whereas phylogenomic studies of scorpions have mostly focused on the systematically complex Iurida (Sharma et al. 2015, 2018; Santibáñez-López et al. 2018, 2019a, 2019b, 2020).

The lack of a robust and densely sampled molecular phylogeny of scorpions has hindered understanding of the origins of mammal-active toxins. It was previously thought that ancient divergences of buthid toxins reflected biogeographic divisions between Old World and New World Buthidae, implying divergences exceeding 180 Myr in age (Froy and Gurevitz 2003). These ancient age estimates are at odds with the distribution of mammal-active toxins in both the Paleotropics and the Neotropics—how would a toxin diversify to target mammal ion channels millions of years before the appearance of the targets?

To close these knowledge gaps, we assembled a phylogenomic data set of 100 scorpion terminals, with transcriptomic data derived from resting venom glands, together with fossil-calibrated estimation of divergence times. Within Buthidae, we emphasized sampling of highly toxic species from the southwestern United States, the Neotropics, and the Middle East. Leveraging a large body of biochemical and functional literature, we mapped the known bioactivity of toxins onto multifamily gene trees of scorpion venom peptides. Here, we show that Buthidae is a surprisingly young lineage that diversified in the Cretaceous; mammal-active toxins of buthids have originated independently in five major clades, contemporaneously with diversification windows of major mammalian predators of scorpions.

MATERIALS AND METHODS

Extended methods are provided in the Supplementary Text available on Dryad at http://dx.doi.org/10.5061/dryad.m0cfxpp25.

Species Sampling and Phylogenomic Analyses

Scorpions were hand collected in Brazil, Egypt, Israel, and the United States, commonly with the aid of ultraviolet (405 nm) lighting. A subset of well-studied species was obtained through captive breeding programs maintained by us and colleagues in the scorpion venomics community. This was required for the manipulation of live animals to drain venom glands and profile transcription levels during venom synthesis. Milking and dissection of venom glands, RNA extraction, library construction, and paired-end transcriptome sequencing on the Illumina HiSeq 2500 platform were performed for 43 species. New data sets were combined with 56 RNA-Seq data sets we generated previously, as well as one genome (*Centruroides sculpturatus*; Schwager et al. 2017; Santibáñez-López et al. 2018). Twenty outgroup

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syac021/6554541 by University of Wisconsin-Madison Libraries user on 28 March 2022

species were including in the analysis, spanning tetrapulmonates, pseudoscorpions, harvestmen, and three horseshoe crabs, following recent updates to chelicerate phylogeny (Ontano et al. 2021; Ballesteros et al. 2022). Collecting, vouchering, and accession data are provided in Tables S1–S4 of the Supplementary material available on Dryad.

Orthologous loci were drawn from Markov Cluster Algorithm clustering of 3534 orthogroups computed from a larger analysis of Chelicerata and outgroup taxa (Ballesteros and Sharma 2019). Untrimmed alignments were used to produce a hidden Markov profile using *hmmerbuild* from *hmmer* v.3.2.1 (Mistry et al. 2013). Each proteome/transcriptome of the species of interest was then searched (*hmmersearch*) for matches against the collection of profiles, with an expectation threshold of e $<10^{-20}$; for cases with more than one hit per locus, the sequence with the best score was selected. Clustering of putative orthologs was tested by comparing each orthogroup's constituent sequences with the proteome of *Drosophila melanogaster* and removing from orthogroups any sequences with mismatching functional annotations (Ballesteros et al. 2022). Gene trees were constructed using IQ-TREE v.1.6 (Nguyen et al. 2014) using ModelFinder Plus (Kalyaanamoorthy et al. 2017) for automated model fitting. Alignments and/or gene trees were visually inspected for chimeric transcripts or paralogous sequences, as evidenced by large numbers of mismatches in alignments, in tandem with anomalously long root-to-tip distances ($>5\times$ average patristic root-to-tip distance). Such anomalous sequences typically corresponded to data-poor legacy data sets (e.g., the genome of *Mesobuthus martensii*) and were discarded. Inclusion of the pyrosequenced genome of *M. martensii* and Sanger-sequenced EST libraries from other buthid species was trialed separately from main analyses, due to the amount of missing data incurred.

Three matrices were assembled with minimum taxon occupancy thresholds: Matrix 1 (at least 115 species), Matrix 2 (at least 109 species), and Matrix 3 (at least 103 species). Phylogenetic inference of concatenated matrices was computed with IQ-TREE, implementing the best-fitting amino acid substitution model per partition. Nodal support values (ultrafast bootstrap replicates [BS]; gene and site concordance factors [gCF, sCF]) were calculated using IQ-TREE. Given the large number of short, shallow internodes we encountered in Buthidae, species tree estimation was also performed using the coalescent summary method implemented in ASTRAL III (Mirarab and Warnow 2015), which can infer shallow-level relationships effectively in the presence of incomplete lineage sorting. Maximum likelihood (ML) gene trees were used as inputs to ASTRAL. Finally, we explored tree inference using site heterogeneous models (CAT+GTR+ Γ) using the software PhyloBayes-mpi v.1.8 (Lartillot et al. 2013) for Matrix 1; these runs failed to converge after 35 weeks of continuous computation and are not reported.

Divergence Time Estimation

Divergence time estimation was computed on Matrix 1 (selected for its number of genes and high completeness) using codeml and MCMCtree (both part of the PAML v. 4.8 software package; Yang 2007; dos Reis and Yang 2019), implementing a likelihood approximation of branch lengths using a multivariate normal distribution. The tree topology inferred from Matrix 2 (selected for a balance of completeness, matrix size, and high nodal support) was used as the input tree and calibrated using fifteen fossil placements (Supplementary Text available on Dryad and Table S5 of the Supplementary material available on Dryad). Four Bayesian inference chains were run for 2.5 M postburnin generations (burnin of 25,000 generations). Fossils used to inform the dating consisted of six ingroup and nine outgroup node calibrations. All calibrations were implemented either as a soft minimum only, or as a soft minimum and soft maximum ages. Fossil calibrations and implementation are provided in Table S5 of the Supplementary material available on Dryad.

Evolution of Scorpion Toxins

Homologs of CSαβ, ICK, and DDH from scorpion venom were retrieved from the complete data set used in our previous scorpion venomics analyses (Santibáñez-López et al. 2018; Santibáñez-López et al. 2019a), as well as UniProt. These peptides are thought to be homologous (Norton and Pallaghy 1998; Santibáñez-López and Possani 2015) (Table S6 of the Supplementary material available on Dryad).

To test relationships of toxin homologs, gene trees were inferred using IQ-TREE for the entire data set (1353 CSαβ-ICK scorpion toxins, with 41 DDH scorpion toxins as outgroups), and for each of the four main clades recovered: (i) sodium channel toxins (NaTx); (ii) potassium channel toxins (KTx); (iii) chlorine channel toxins (ClTx); and (iv) calcins. To accommodate possible nonhomology of CSαβ and ICK/DDH peptides (an alternative interpretation of protein folds; Zhang et al. 2017), we ran a separate family of analyses, wherein relationships between CSαβ were inferred independently of ICK and DDH homologs. Tree inference heuristics were identical to the analysis above.

To test for the signature of common versus independent evolution of bioactivity in venom genes, comparative analyses between subclades recovered within NaTx included searches for repetitive motifs in their mature peptide using Multiple Em for Motif Elicitation (MEME v. 5.1; Bailey et al. 2015), using the gene family tree inferred above as the input. As gene trees can often lack sufficient signal in short peptides like toxins, we separately analyzed the mature peptide of the two main clusters within the NaTx (Aah2-like and Cn2-like) using CLANS clustering (Frickey and Lupas 2004). Poorly connected networks of sequences that shared a specific bioactivity were inferred to constitute independent evolutionary gains of a given molecular function.

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syac021/6554541 by University of Wisconsin-Madison Libraries user on 28 March 2022

To infer the ages of mammal-specific toxins, we employed a phylostratigraphic bracketing approach, which consists of estimating ages of genes from the chronogram of the species tree; ages of the most inclusive clades of taxa are used as minimum age estimates for gene age. Divergence times for scorpion mammal predators such as Herpestidae (Carnivora), Chiroptera, Eulipotyphla, and Rodentia were retrieved from two analyses of mammal diversification times (Meredith et al. 2011; Upham et al. 2019) for comparison with the inferred ages of mammal-active toxins.

Results

Phylogenomic Analysis

Three phylogenomic matrices were constructed, spanning 192–660 loci (53,333–185,631 aligned amino acid sites), using taxon minimum occupancy thresholds of 103, 109, and 115 terminals per locus (88.0%, 91.4%, and 94.1% complete, respectively). All matrices exhibited desirable properties with respect to phylogenomic biases, such as low saturation levels and long-branch scores (Figs. S1–S3 of the Supplementary material available on Dryad). ML analyses of concatenated data sets consistently recovered, with maximal nodal support, the monophyly of scorpions, and the established basal split between Buthida (consisting of Buthidae, Chaerilidae, and Pseudochactidae) and Iurida (the remaining scorpion families) (Sharma et al. 2015). Matrices 2 and 3 recovered *Lychas variatus* as the sister group of the remaining buthids (BS: 87–95%; gCF: 75.2–78.9; sCF: 74.0–74.1 sCF), whereas Matrix 1 recovered a nested placement of *L. variatus* with weak nodal support (BS: 72%; gCF: 13.0; sCF: 32.1; Figs. S4–S6 of the Supplementary material available on Dryad). All three data sets recovered with maximal support four major clades within the buthids, consisting of (i) the largely Palearctic "*Buthus* group," (ii) *Ananteris* + *Babycurus*, (iii) the "*Uroplectes* group," and (iv) the "*Tityus* group" (Fig. S7 of the Supplementary material available on Dryad; Supplementary Data available on Dryad). Relationships between these groups across all supermatrix analyses supported the sister group relationship of the *Tityus* and *Uroplectes* groups, with this lineage in turn sister group to (*Ananteris* + *Babycurus*). Relationships within Iurida closely reflected tree topologies reported in our previous works and are not discussed further herein (Sharma et al. 2015, 2018; Santibáñez-López et al. 2019a, 2019b; Santibáñez-López et al. 2020). Relationships reconstructed under the multispecies coalescent using ASTRAL were largely congruent, excepting the placement of *Lychas*, which was also unsupported (posterior probability [PP] = 0.47–0.80 across analyses; Fig. S8 of the Supplementary material available on Dryad).

Molecular Dating

Divergence times were inferred under two models of rate evolution, a correlated rates model and an independent rates model. Both clock models recovered comparable inferences of basal scorpion diversification, with a split between Buthida and Iurida dating to the Carboniferous-Permian (291–320 Ma; highest posterior density [HPD] interval: 247–352 Ma) (Fig. 2). The diversification of Buthidae was estimated to span the end-Jurassic to the Early Cretaceous (95% HPD: 105–161 Ma). Divergences within major buthid clades fell within the Late Cretaceous to Paleogene, excepting the *Buthus* group, wherein most divergences were estimated to occur in the Neogene.

Evolution of Venom Peptides

Putative CSαβ and ICK venom components were identified in 100 terminals ($N = 605$) and were complemented with additional sequences from the UniProt databases ($N = 789$). Using gene tree reconstructions, we documented 1353 CSαβ-ICK peptide homologs spanning 151 scorpion species in 19 families (from which 48% were buthids; Table S6 of the Supplementary material available on Dryad). ML analysis of the 1353 CSαβ-ICK matrix and 41 DDH sequences as outgroups (399 amino acid sites) recovered a gene family tree subdivided into three major clades: (i) NaTx; (ii) calcins; and (iii) KTx, including the nested ClTx (Figs. S9 and S10 of the Supplementary material available on Dryad). Our results revealed that calcins are phylogenetically restricted to iurids (Figs. S9 and S11 of the Supplementary material available on Dryad), whereas ClTx is restricted to a subset of Old World buthids (Figs. S9 and S12 of the Supplementary material available on Dryad). ML analysis of CSαβ only recovered the independent origin of NaTx and KTx, in agreement with our topology recovered using the combined data set, as well as previous studies (Zhang et al. 2017). Similarly, ML analysis of ICK-DDH reflected dynamics reconstructed elsewhere (Smith et al. 2011; Santibáñez-López et al. 2018).

To reconstruct the evolutionary dynamics of these homologs, we leveraged a large body of literature documenting biochemical and functional properties of these peptides, where known (e.g., Possani et al. 1999; Gordon et al. 2007; Smith et al. 2011; Gao et al. 2013), mapping these onto the gene family tree (Fig. 3). Peptide functions were classified following the characterization in Uniprot. A complete list of functional properties mapped onto the tree is provided in Table S6 of the Supplementary material available on Dryad.

Within NaTx, mammal-active toxins were recovered as eight derived clusters within three main groups: (i) Acra3-like; (ii) Aah2-like; and (iii) Cn2-like (Figs. 3a,b). The Aah2-like gene family was found exclusively in Buthida (mostly from Old World Buthids) and encodes for peptides that are arthropod-active, insect- and mammal-active, and mammal-active (Fig. 3b). In contrast, within the Cn2-like gene family, we found one cluster restricted to Iurida and six clusters exclusively

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syac021/6554541 by University of Wisconsin-Madison Libraries user on 28 March 2022

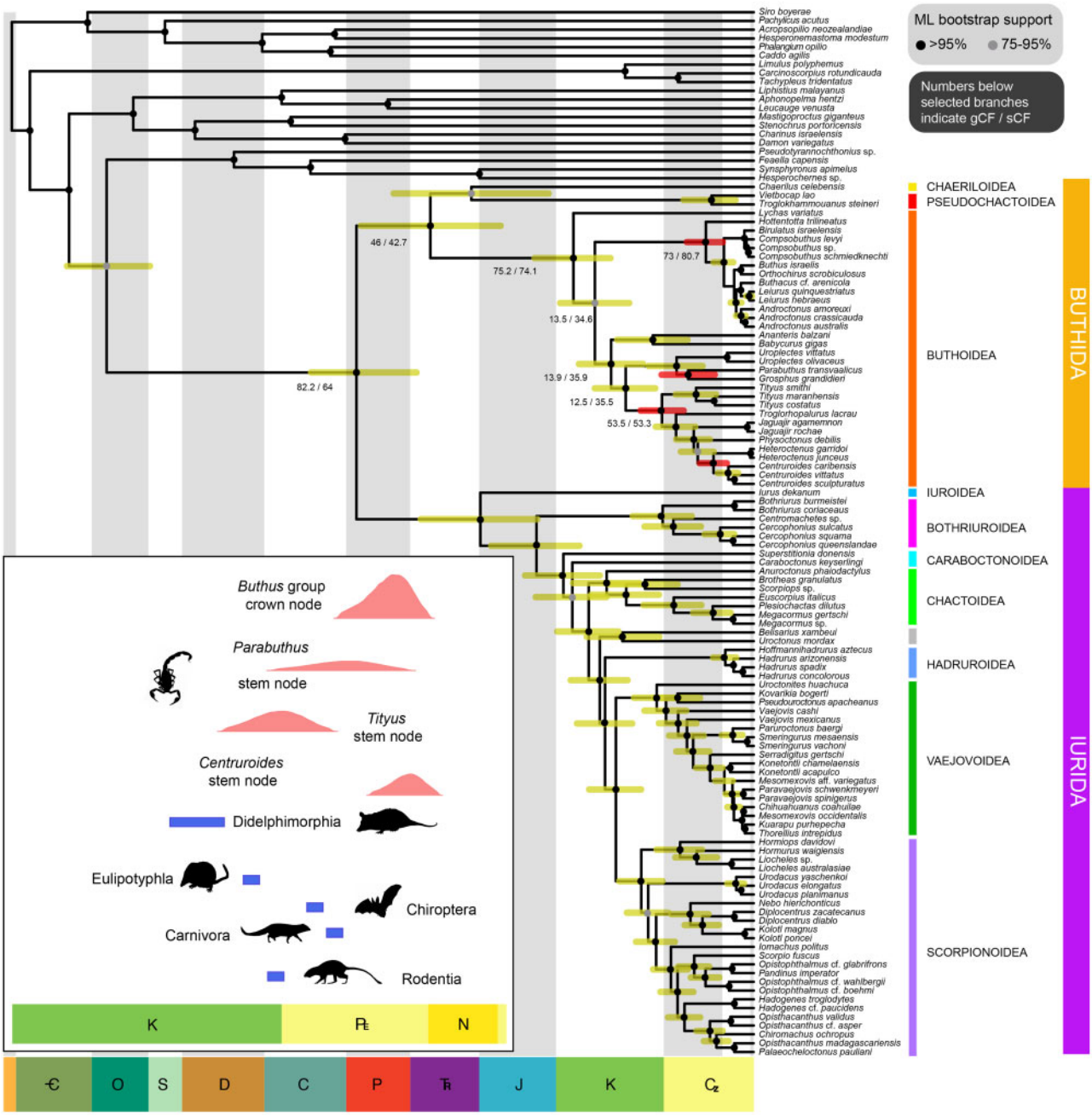

FIGURE 2. Chronogram of scorpion relationships derived from ML analysis of 192 loci (53,333 amino acid sites). Nodal support metrics are provided in Supplementary material available on Dryad. Node ages are computed in a time-calibrated analysis using 15 fossil calibrations. Yellow bars depict 95% credibility intervals of node ages, whereas red bars depict 95% credibility intervals for the most toxic buthids. Inset: Posterior distribution of node ages corresponding to mammal-active toxin origins (red), compared to the posterior distribution of the four major mammal orders (blue) that include major scorpion predators. Mammal order ages from Upham et al. (2019).

in Buthida. Among these, two clusters with mammal-active targets were found uniquely in *Centruroides* and *Tityus*. Our search for motifs with MEME (Bailey et al. 2015) showed a short nonunique motif of 10 amino acids (GXXCWCXXLPD) for members of the Aah2-like gene family, and no specific motif for members of the Cn2-like gene family (Figs. S13–S19 of the Supplementary material available on Dryad). More specifically, no conserved or repetitive motifs were found in the mammal-active toxins of either the Aah2 or Cn2 gene families (Fig. S19 of the Supplementary material available on Dryad). Given the short sequence length of NaTx, we also assessed patterns of relatedness between the 661 mature peptide sequences from our data set using CLANS clustering (Frickey and Lupas 2004). Consistent with the gene tree analyses, this approach recovered eight separate groups

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syac021/6554541 by University of Wisconsin-Madison Libraries user on 28 March 2022

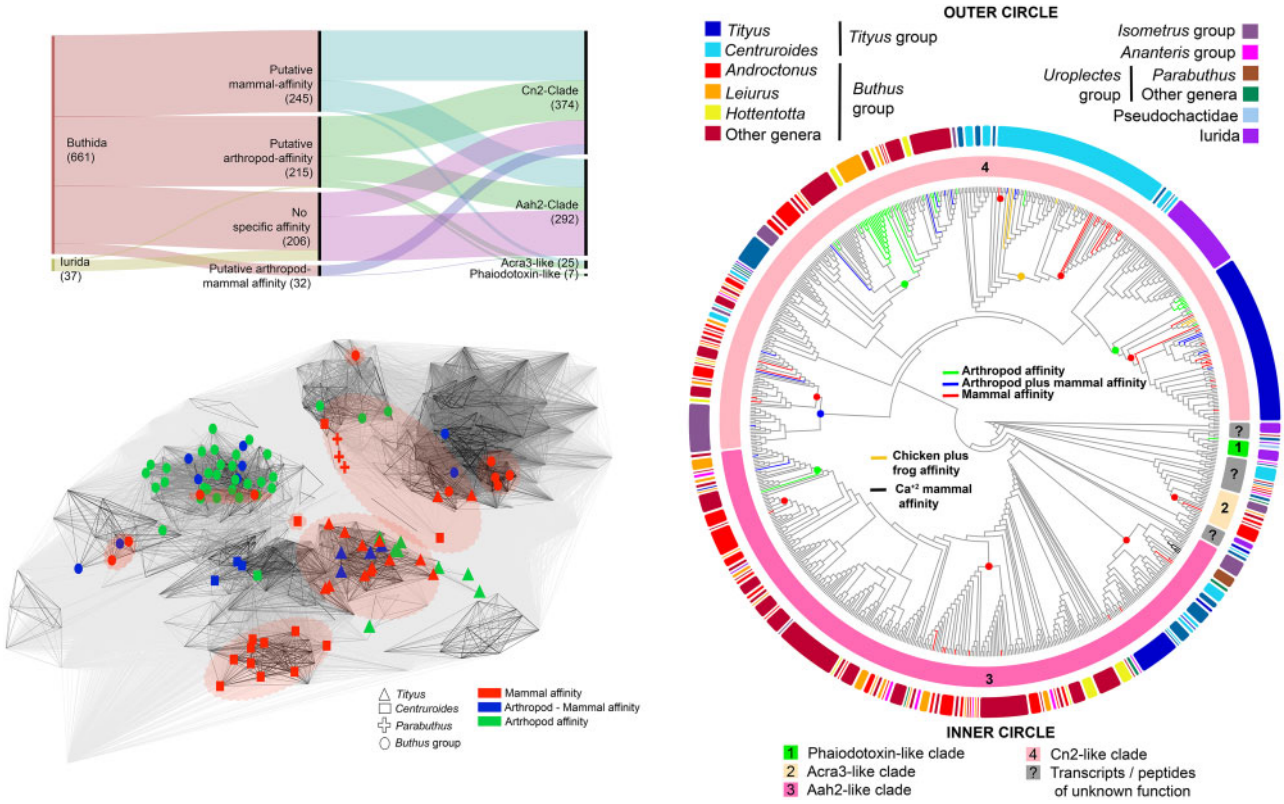

FIGURE 3. Evolutionary analyses of the sodium channel toxin family (NaTx). a) Alluvial plot summarizing the activity of the NaTx found in scorpion venom (center) of each parvorder (left), and their subtype (right). Numbers represent the total transcripts and/or peptides found in our transcriptomic analyses and UniProt. b) NaTx activity (inner bars) plotted onto the NaTx gene tree subdivided into seven subclades (four well-known plotted onto the alluvial plot in (a)). In gray: transcript clades with unknown function. Circles on nodes represent inferred ancestral bioactivity. c) 3D CLANS clustering of the mature peptide amino acid sequence. Colors correspond to NaTx activity. Red dotted lines show eight clusters that include at least one peptide with known mammal ion channel activity.

of peptides with affinity for mammalian ion channels (Fig. 3c).

To infer the age of mammal-active genes, we employed a phylostratigraphic bracketing approach, using the stem age of the most inclusive buthid clade containing a given mammal-active toxin as a proxy for the minimum estimated gene age. Stem ages of *Centruroides*, *Tityus*, *Parabuthus*, and the node uniting *Hottentotta* with the remaining *Buthus* group scorpions thus implied the earliest diversification of mammal-specific toxins in a temporal window spanning. 20-83 Ma (Fig. 2, inset; Fig. S20 of the Supplementary material available on Dryad).

DISCUSSION

Contemporaneous Diversification of Buthid Mammal-Active Toxins and Mammalian Predators of Scorpions

Previous genomic resources sampling scorpion venom diversity, as well as the ensuing inferences of scorpion toxin evolution, have focused on data from mostly medically significant species, such as *Leiurus quinquestriatus*, *Androctonus australis*, and *Centruroides sculpturatus* (e.g., Froy et al. 1999; Schwager et al. 2017). We focused herein on comparative analyses sampling venom gene expression broadly across scorpion phylogeny, toward characterizing the evolutionary dynamics that precipitated scorpion toxicity to mammals. We also endeavored to increase high-quality transcriptomic resources for medically significant species in *Androctonus*, *Centruroides*, and *Tityus*. Through these data sets, we discovered that certain classes of toxins are phylogenetically restricted; as examples, chlorotoxins occur only in Old World Buthidae; scorpines occur in all families except Buthidae; and calcins are restricted to Iurida (Fig. S9 of the Supplementary material available on Dryad).

Fossil-based phylogenomic dating recovered a surprisingly young age for the basal diversification of Buthidae in the Late Mesozoic, contrary to previous conjectures of a pre-Jurassic age of buthids that were grounded in biogeographic patterns (Froy and Gurevitz 2003). We inferred eight independent origins of NaTx with known specificity for mammalian targets spanning five separate buthid lineages, with three of these occurring in genera sampled with multiple terminals. Phylostratigraphic bracketing of these toxins' origins recovered age estimates broadly overlapping the basal diversification dates of several mammal orders that

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syac021/6554541 by University of Wisconsin-Madison Libraries user on 28 March 2022

include scorpion predators (Fig. 2). These mammal-active buthid NaTx were nested within a cluster of NaTx that target arthropod tissues across all scorpions (Fig. 3).

The relatively young ages of highly toxic buthid scorpions, in addition to the derived position of toxin genes with an affinity for mammalian ion channels in the broader tree of insect-targeting toxins (e.g., the position of the phaiodotoxin-like clade, Fig. 3), point to the cooption of scorpion mammal-targeting toxins from insect-targeting ancestral peptides, reflecting the derivation of an antipredator defensive adaptation from peptides previously used to target prey, as evidenced by intermediate peptides with dual affinities (e.g., Gordon et al. 2007). Similar dynamics have recently been revealed in the highly venomous Australian funnel-web spiders, wherein δ-hexatoxins exhibit high evolutionary conservation, reflecting a defensive role for deterring vertebrate predators (Herzig et al. 2020). By contrast to these spiders, molecular signatures of selection revealed no consistent pattern of amino acid sequence evolution across groups of scorpion mammal-active toxins, consistent with the inference of independent evolutionary origins of antivertebrate defensive peptides (Fig. S19 of the Supplementary material available on Dryad).

Notably, counter-adaptations to scorpion venom are known to occur in some scorpion predators. For example, the grasshopper mouse *Onychomys torridus* exhibits reduced sensitivity to pain caused by the sting of the Arizona bark scorpion *C. sculpturatus* (Rowe et al. 2013). The mechanism of this counter-adaptation was shown to be amino acid variants of a voltage-gated Na^+ channel in *O. torridus* that have evolved to selectively bind *C. sculpturatus* toxins, blocking action potential propagation. Parallel evolution of resistance to the venom of *C. sculpturatus* via modification of Na^+ ion channels has also been suggested in the bat *Antrozous pallidus* (Hopp et al. 2017). Comparable molecular dynamics underlying the evolution of resistance to snake venom have evolved several times in snake predators, such as mongooses, honey badgers, and hedgehogs (Barchan et al. 1992; Drabeck et al. 2015; Holding et al. 2016; Khan et al. 2020).

Given the diversity of Buthidae and the molecular complexity of their venoms, broader phylogenetically informed surveys of venom gland transcriptomics may uncover additional origins of mammalian-targeting toxin functions in scorpions. Venom gland transcriptomic databases sampling poorly studied species may also offer additional targets for beneficial biomedical applications that are not represented among established scorpion biomedical research programs.

Implications for Buthid Systematics

The systematic history of scorpions was previously dominated by morphological analyses, with marked contention between competing interpretations of homologies and cladistic practices. Intensive surveys of trichobothrial position by Fet et al. (2005) were previously used to delimit six major groups of scorpions (the *Buthus*, *Ananteris*, *Tityus*, *Charmus*, *Isometrus*, and *Uroplectes* groups), with the Paleartic *Buthus* groups comprising the sister lineage of the remaining Buthidae. The present study recovered many of the groupings within Buthidae that were established on the basis of trichobothrial characters (sensory setae on the pedipalps). Our results accord precisely with the morphological conception of the *Buthus* group, the *Uroplectes* group, and the *Tityus* group. Some analyses additionally recovered the basally branching placement of the *Buthus* group, albeit weakly (Figs. S4 and S7 of the Supplementary material available on Dryad).

Our results differed with morphological delimitations of buthid relationships only with respect to the composition of the two groups that were also poorly resolved by the morphological data in that study (the *Isometrus* and *Ananteris* groups; Fet et al. 2005). One lineage not sampled herein using transcriptomic data is the enigmatic *Charmus* group, which is geographically restricted to the Indian subcontinent, southeast Asia, and parts of East Africa. A separate analysis we performed—in which we combined Matrix 2 with available Sanger-sequenced data for some buthid lineages as well as a pyrosequenced genome for *M. martensii*—also revealed that *Charmus* is part of the group that includes the Malagasy endemic genus *Grosphus* and the southern African *Parabuthus* and *Uroplectes* (Fig. S7 of the Supplementary material available on Dryad). These outcomes vindicate the interpretation and utility of trichobothrial arrangements established by Fet et al. (2005) for inferring shallow-level relationships within scorpion families.

Supplementary Material

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.m0cfxpp25.

Funding

Fieldwork in Australia was supported by intramural UW-Madison funds to P.P.S. This work was supported by National Science Foundation [IOS-1552610 to P.P.S.]; Fundação de Amparo à Pesquisa do Estado de São Paulo BIOTA [2013/50297-0] and National Science Foundation [DOB-1343578 to R.P.R.]; and Binational Science Foundation [2019216 to P.P.S. and E.G.R.]. Fieldwork in Israel was supported by National Geographic Society Expeditions Council [NGS-271R-18 to J.A.B.].

Acknowledgments

We are indebted to Brandon Myers for providing some of the captive-bred specimens analyzed in this study. Steve Goodman provided tissue of *Grosphus grandidieri* from Madagascar. An earlier draft of this work was improved by comments from Lourival D. Possani, Bryan G. Fry, Yehu Moran, the Associate Editor, and three anonymous reviewers. Sequencing was performed

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syac021/6554541 by University of Wisconsin-Madison Libraries user on 28 March 2022

at the UW-Madison Biotechnology Center. Access to computing nodes for intensive tasks was provided by the Center for High Throughput Computing (CHTC) and facilitated by Christina Koch; and by the Bioinformatics Resource Center (BRC) of UW-Madison. Specimens in Israel were collected under permits 2017/41718 and 2018/42037 (Israel National Parks Authority).

References

Ballesteros J.A., Sharma P.P. 2019. A critical appraisal of the placement of Xiphosura (Chelicerata) with account of known sources of phylogenetic error. Syst. Biol. 68:896–917.

Ballesteros J.A., Santibáñez-López C.E., Baker C.M., Benavides L.R., Cunha T.J., Gainett G., Ontano A.Z., Setton E.V.W., Arango C.P., Gavish-Regev E., Harvey M.S., Wheeler W.C., Hormiga G., Giribet G., Sharma P.P. 2022. Comprehensive species sampling and sophisticated algorithmic approaches refute the monophyly of Arachnida. Mol. Biol. Evol. 39:msac021.

Bailey T.L., Johnson J., Grant C.E., Noble W.S. 2015. The MEME suite. Nucleic Acids Res. 43:W39–W49.

Barchan D., Kachalsky S., Neumann D., Vogel Z., Ovadia M., Kochva E., Fuchs S. 1992. How the mongoose can fight the snake: the binding site of the mongoose acetylcholine receptor. Proc. Natl. Acad. Sci. USA 89:7717–7721.

Borges A., Graham M.R. 2014. Phylogenetics of scorpions of medical importance. In: Gopalakrishnakone P., Calvete J.J., editors. Venom genomics and proteomics. Dordrecht, Germany: Springer. p. 81–104.

Cao Z., Yu Y., Wu Y., Hao P., Di Z., He Y., Chen Z., Yang W., Shen Z., He X., Sheng J., Xu X., Pan B., Feng J., Yang X., Hong W., Zhao W., Li Z., Huang K., Li T., Kong Y., Liu H., Jiang D., Zhang B., Hu J., Hu Y., Wang B., Dai J., Yuan B., Feng Y., Huang W., Xing X., Zhao G., Li X., Li Y., Li W. 2013. The genome of *Mesobuthus martensii* reveals a unique adaptation model of arthropods. Nat. Commun. 4:2602.

Casewell N.R., Wüster W., Vonk F.J., Harrison R.A., Fry B.G. 2013. Complex cocktails: the evolutionary novelty of venoms. Trends Ecol. Evol. 28:219–229.

dos Reis M., Yang Z. 2019. Bayesian molecular clock dating using genome-scale datasets. In: Anisimova M., editor. Evolutionary genomics. Methods in molecular biology, vol. 1910. New York, NY: Humana. p. 309-330.

Drabeck D.H., Dean A.M., Jansa S.A. 2015. Why the honey badger don't care: convergent evolution of venom-targeted nicotinic acetylcholine receptors in mammals that survive venomous snake bites. Toxicon 99:68–72.

Fet V., Gantenbein B., Gromov A.V., Lowe G., Lourenço W.R. 2003. The first molecular phylogeny of Buthidae (Scorpiones). Euscorpius 4:1–10.

Fet V., Soleglad M., Lowe G. 2005. A new trichobothrial character for the high-level systematics of Buthoidea (Scorpiones: Buthida). Euscorpius 23:1–40.

Frickey T., Lupas A. 2004. CLANS: a Java application for visualizing protein families based on pairwise similarity. Bioinformatics 20:3702–3704.

Froy O., Gurevitz M. 2003. New insight on scorpion divergence inferred from comparative analysis of toxin structure, pharmacology and distribution. Toxicon 42:549–555.

Froy O., Sagiv T., Porech M., Urbach D., Zillberberg N., Gurevitz M. 1999. Dynamic diversification from a putative common ancestor of scorpion toxins affecting sodium, potassium, and chloride channels. J. Mol. Evol. 48:187–196.

Gao B., Harvey P.J., Craik D.J., Ronjat M., De Waard M., Zhu S. 2013. Functional evolution of scorpion venom peptides with an inhibitor cystine knot fold. Biosci. Rep. 33(3):e00047.

Gordon D., Karbat I., Illan N., Cohen L., Kahn R., Gilles N., Dong K., Stühmer W., Tytgat J., Gurevitz M. 2007. The differential preference of scorpion α-toxins for insect or mammalian sodium channels: implications for improved insect control. Toxicon 49:452–472.

Herzig V., Sunagar K., Wilson D.T.R., Pineda S.S., Israel M.R., Dutertre S., McFarland B.S., Undheim E.A.B., Hodgson W.C., Alewood P.F., Lewis R.J., Bosmans F., Vetter I., King G.F., Fry B.G. 2020. Australian funnel-web spiders evolved human-lethal δ-hexatoxins for defense against vertebrate predators. Proc. Natl. Acad. Sci. USA 117(40):24920–24928.

Holding M.L., Drabeck D.H., Jansa S.A., Gibbs H.L. 2016. Venom resistance as a model for understanding the molecular basis of complex coevolutionary adaptations. Integr. Comp. Biol. 56:1032–1043.

Hopp B.H., Arvinson R.S., Adams M.E., Razak K.A. 2017. Arizona bark scorpion venom resistance in the pallid bat, *Antrozous pallidus*. PLoS One 12:e0183215.

Kalyaanamoorthy S., Minh B., Wong T.K.F., von Haeseler A., Jermiin L.S. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14:587–589.

Khan M.A., Dashevsky D., Kerkkamp H., Kordis D., de Bakker M.A.G., Wouters R., van Thiel J., den Brouw B., Vonk F., Kini R.M., Nazir J., Fry B.G., Richardson M.K. 2020. Widespread evolution of molecular resistance to snake venom alpha neurotoxins in vertebrates. Toxins 12:638

King G.F., Hardy M.C. 2013. Spider-venom peptides: structure, pharmacology, and potential for control of insect pests. Ann. Rev. Entomol. 58:475–96.

Lartillot N, Rodrigue N, Stubbs D, Richer J. 2013. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. Syst Biol. 62(4):611–615.

Meredith R.W., Janecka J.E., Gatesy J., Ryder O.A., Fisher C.A., Teeling E.C., Goodbla A., Eizirik E., Simão T.L., Stadler T., Rabosky D.L. 2011. Impacts of the cretaceous terrestrial revolution and KPg extinction on mammal diversification. Science 334(6055):521–524.

Mirarab S., Warnow T. 2015. ASTRAL-II: coalescent- based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 31:i44–i52.

Mistry J., Finn R.D., Eddy S.R., Bateman A., Punta M. 2013. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Res. 41:e121.

Niermann C., Tate T.G., Suto A.L., Barajas R., While H.A., Guswiler O.D., Secor S.M., Rowe A.H., Rowe M.P. 2020 Defensive venoms: is pain sufficient for predator deterrence? Toxins 12:260.

Nguyen L.T., Schmidt H.A., Haeseler A., Minh B.Q. 2014. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32:268–274.

Norton R.S., Pallaghy P.K. 1998. The cystine knot structure of ion channel toxins and related polypeptides. Toxicon 36:1573–1583.

Ojanguren-Affilastro A.A., Adilardi R.S., Mattoni C.I., Ramírez M.J., Ceccarelli F.S. 2017. Dated phylogenetic studies of the southernmost American buthids (Scorpiones; Buthidae). Mol. Phylogenet. Evol. 110:39–49.

Ontano A.Z., Gainett G., Aharon S., Ballesteros J.A., Benavides L.R., Corbett K.F., Gavish-Regev E., Harvey M.S., Monsma S., Santibáñez-López C.E., Setton E.V.W., †Zehms J.T., Zeh J.A., Zeh D.W., Sharma P.P. 2021. Taxonomic sampling and rare genomic changes overcome long-branch attraction in the phylogenetic placement of pseudoscorpions. Mol. Biol. Evol. 38:2446–2467.

Oukkache N., ElJaoudi R., Chgoury F., Rocha M.T., Sabatier J.M. 2015. Characterization of Am IT, an anti-insect β-toxin isolated from the venom of scorpion *Androctonus mauretanicus*. Sheng Li Xue Bao 67:295–304.

Possani L.D., Becerril B., Delepierre M., Tytgat J. 1999. Scorpion toxins specific for Na^+- channels. Eur. J. Biochem. 264(2):287–300.

Rapôso C. 2017. Scorpion and spider venoms in cancer treatment: state of the art, challenges, and perspectives. J. Clin. Transl. Res. 3:233–249.

Rowe A.H., Xiao Y., Rowe M.P., Cummins T.R., Zakon H.H. 2013. Voltage-gated sodium channel in grasshopper mice defends against bark scorpion toxin. Science 342:441–446.

Santibáñez-López C.E., Possani L.D. 2015. Overview of the Knottin scorpion toxin-like peptides in scorpion venoms: insights on their classification and evolution. Toxicon 107:317–326.

Santibáñez-López C.E., Kriebel R., Ballesteros J.A., Rush N., Witter Z., Williams J., Janies D.A., Sharma P.P. 2018. Integration of

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syac021/6554541 by University of Wisconsin-Madison Libraries user on 28 March 2022

phylogenomics and molecular modeling reveals lineage-specific diversification of toxins in scorpions. PeerJ 6:e5902–5923.
Santibáñez-López C.E., Graham M.R., Sharma P.P., Ortiz E., Possani L.D. 2019a. Hadrurid scorpion toxins: evolutionary conservation and selective pressures. Toxins 11:637.
Santibáñez-López C.E., González-Santillán E., Monod L., Sharma P.P. 2019b. Phylogenomics facilitates stable scorpion systematics: reassessing the relationships of Vaejovidae and a new higher-level classification of Scorpiones (Arachnida). Mol. Phylogenet. Evol. 135:22–30.
Santibáñez-López C.E., Ojanguren-Affilastro A.A., Sharma P.P. 2020. Another one bites the dust: taxonomic sampling of a key genus in phylogenomic datasets reveals more non-monophyletic groups in traditional scorpion classification. Invertebr. Syst. 34:133–143.
Santos M.S., Silva C.G.L., Neto B.S., Júnior C.R.P., Lopes V.H.G., Júnior A.G.T., Bezerra D.A., Luna J.V.C.P., Cordeiro J.B., Júnior J.G., Lima M.A.P. 2016. Clinical and epidemiological aspects of scorpionism in the world: a systematic review. Wild Environ. Med. 27: 504–518.
Schendel V., Rash L.D., Jenner R.A., Undheim E.A.B. 2019. The diversity of venom: the importance of behavior and venom system morphology in understanding its ecology and evolution. Toxins 11:666.
Schwager E.E., Sharma P.P., Clarke T., Leite D.J., Wierschin T., Pechmann M., Akiyama-Oda Y., Esposito L., Bechsgaard J., Bilde T., Buffry A.D., Chao H., Dinh H., Doddapaneni H., Dungan S., Eibner C., Extavour C.G., Funch P., Garb J., Gonzalez L.B., Gonzalez V.L., Griffiths-Jones S., Han Y., Hayashi C., Hilbrant M., Hughes D.S.T., Janssen R., Lee S.L., Maeso I., Murali S.C., Muzny D.M., Nunes da Fonseca R., Paese C.L.B., Qu J., Roshaugen M., Schomburg C., Schönauer A., Stollewerk A., Torres-Oliva M., Turetzek N., Vanthournout B., Werren J.H., Wolf C., Worley K.C., Bucher G., Gibbs R.A., Coddington J., Oda H., Stanke M., Ayoub N.A., Prpic N.M., Flot J.F., Posnien N., Richards S., McGregor A.P. 2017. The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. BMC Biol. 15:62.
Sharma P.P., Fernández R., Esposito L.A., González-Santillán E., Monod L. 2015. Phylogenomic resolution of scorpions reveals multilevel discordance with morphological phylogenetic signal. Proc. R. Soc. B 282:20142953.
Sharma P.P., Baker C.M., Cosgrove J.G., Johnson J.E., Oberski J.T., Raven R.J., Harvey M.S., Boyer S.L., Giribet G. 2018. A revised dated phylogeny of scorpions: phylogenomic support for ancient divergence of the temperate Gondwanan family Bothriuridae. Mol. Phylogenet. Evol. 122:37–45.
Smith J.J., Jill J.M., Little M.J., Nicholson G.M., King G.F., Alewood P.F. 2011. Unique scorpion toxin with a putative ancestral fold provides insight into evolution of the inhibitor cystine knot motif. Proc. Natl. Acad. Sci. USA 108(26):10478–10483.
Suranse V., Sawant N.S., Paripatydar S.V., Krutha K., Paingankar M.S., Padhye A.D., Bastawade D.B., Dahanukar N. 2017. First molecular phylogeny of scorpions of the family Buthidae from India. Mitochondrial DNA A 28:606–611.
Upham N.S., Esselstyn J.A., Jetz W. 2019. Inferring the mammal tree: species-level sets of phylogenies for questions in ecology, evolution, and conservation. PLoS Biol. 17: e3000494.
Veiseh M., Gabikian P., Bahrami S.B., Veiseh O., Zhang M., Hackman R.C., Ravanpay A.C., Stroud M.R., Kusama Y., Hansen S.J., Kwok D., Munoz N.M., Sze R.W., Grady W.M., Greenberg N.M., Ellebogen R.G., Olson J.M. 2007. Tumor paint: a chlorotoxin:Cy5.5 bioconjugate for intraoperative visualization of cancer foci. Cancer Res. 67:6882–6888.
Wang J., Yarov-Yarovoy V., Kahn R., Gordon D., Gurevitz M., Scheuer T., Catterall W.A. 2011. Mapping the receptor site for α-scorpion toxins on a Na+ channel voltage sensor. Proc. Natl. Acad. Sci. USA 108(37):15426–15431.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24:1586–1591.
Zhang S., Gao B., Zhu S. (2017) Independent origins of scorpion toxins affecting potassium and sodium channels. In: Gopalakrishnakone P., Malhotra A., editors. Evolution of venomous animals and their toxins. Toxinology. Dordrecht: Springer. p. 167–186.

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syac021/6554541 by University of Wisconsin-Madison Libraries user on 28 March 2022