Development Genes and Evolution
https://doi.org/10.1007/s00427-019-00644-6

ORIGINAL ARTICLE

Developmental gene expression as a phylogenetic data class: support for the monophyly of Arachnopulmonata

Erik D. Nolan [1,2] · Carlos E. Santibáñez-López [1,3] · Prashant P. Sharma [1]

Received: 23 September 2019 / Accepted: 10 December 2019

Abstract

Despite application of genome-scale datasets, the phylogenetic placement of scorpions within arachnids remains contentious between two different phylogenetic data classes. Paleontologists continue to recover scorpions in a basally branching position, partly owing to their morphological similarity to extinct marine orders like Eurypterida (sea scorpions). Phylogenomic datasets consistently recover scorpions in a derived position, as the sister group of Tetrapulmonata (a clade of arachnids that includes spiders). To adjudicate between these hypotheses using a rare genomic change (RGC), we leveraged the recent discovery of ancient paralogy in spiders and scorpions to assess phylogenetic placement. We identified homologs of four transcription factors required for appendage patterning (*dachshund*, *homothorax*, *extradenticle*, and *optomotor blind*) in arthropods that are known to be duplicated in spiders. Using genomic resources for a spider, a scorpion, and a harvestman, we conducted gene tree analyses and assayed expression patterns of scorpion gene duplicates. Here we show that scorpions, like spiders, retain two copies of all four transcription factors, whereas arachnid orders like mites and harvestmen bear a single copy. A survey of embryonic expression patterns of the scorpion paralogs closely matches those of their spider counterparts, with one paralog consistently retaining the putatively ancestral pattern found in the harvestman, as well as the mite, and/or other outgroups. These data comprise a rare genomic change in chelicerate phylogeny supporting the inference of a distal placement of scorpions. Beyond demonstrating the diagnostic power of developmental genetic data as a phylogenetic data class, a derived placement of scorpions within the arachnids, together with an array of stem-group Paleozoic scorpions that occupied marine habitats, effectively rules out a scenario of a single colonization of terrestrial habitat within Chelicerata, even in tree topologies contrived to recover the monophyly of Arachnida.

Keywords Arachnida · Rare genomic change · Arthropoda · Phylogenomics

Introduction

The advent of current generation sequencing technologies has greatly benefitted the practice of molecular systematics. However, certain recalcitrant nodes in the Tree of Life remain staunchly unresolved despite quantity of sequence data deployed to address phylogenetic relationships. Among the most intractable empirical problems in phylogenetics are nodes characterized by the combination of (1) ancient and rapid diversification ("bushes" in the Tree of Life, sensu Rokas and Carroll 2006) and (2) accelerated evolution of one or more ingroup lineages, exacerbating long branch attraction artifacts (Bergsten 2005). A notable empirical case is this phylogeny of Chelicerata (the subphylum of arthropods that includes sea spiders, horseshoe crabs, and arachnids) (Regier et al. 2010; Sharma et al. 2014a; Lozano-Fernandez et al. 2019; Ballesteros and Sharma 2019).

This article is part of the Special Issue "Crossroads in Spider Research—evolutionary, ecological, and economic significance"

Communicated by Matthias Pechmann

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00427-019-00644-6) contains supplementary material, which is available to authorized users.

✉ Prashant P. Sharma
prashant.sharma@wisc.edu

[1] Department of Integrative Biology, University of Wisconsin-Madison, Madison, WI 53706, USA

[2] Department of Developmental Biology, Washington University of St. Louis, St. Louis, MO 63110, USA

[3] Department of Biology, Eastern Connecticut State University, 83 Windham Street, Willimantic, CT 06266, USA

It is generally held by morphologists and paleontologists that the two extant marine orders of Chelicerata (Pycnogonida and Xiphosura) form a grade at the base of a monophyletic Arachnida (Shultz 1990, 2007; Legg et al. 2013; Garwood and Dunlop 2014; Aria and Caron 2019) (Fig. 1; see "Materials and methods"). The attendant evolutionary scenario postulates that Arachnida is the consequence of a single colonization of land. Key to this scenario is the position of scorpions, which superficially resemble the extinct marine order Eurypterida (sea scorpions). Scorpions bear book lungs, respiratory structures that are interpreted to be internalized book gills, and are thus held to represent a stepping stone to the evolution of the remaining arachnid orders (Dunlop 1998; Scholtz and Kamenz 2006; reviewed by Sharma (2017). This idea has become so deeply engrained in the literature that some workers do not test the monophyly of arachnids, using Xiphosura (horseshoe crabs) as the default outgroup to root the tree (Shultz 1990, 2007; Garwood et al. 2016).

The entrenchment of arachnid monophyly and a basally branching placement of scorpions (either alone or with Opiliones; Legg et al. 2013; Aria and Caron 2019) has long been refuted by molecular sequence data, which generally recover both Scorpiones and Xiphosura as nested within Arachnida across an array of datasets, analytical approaches, and taxonomic sampling strategies (Meusemann et al. 2010; Sharma et al. 2014a; Ballesteros and Sharma 2019;

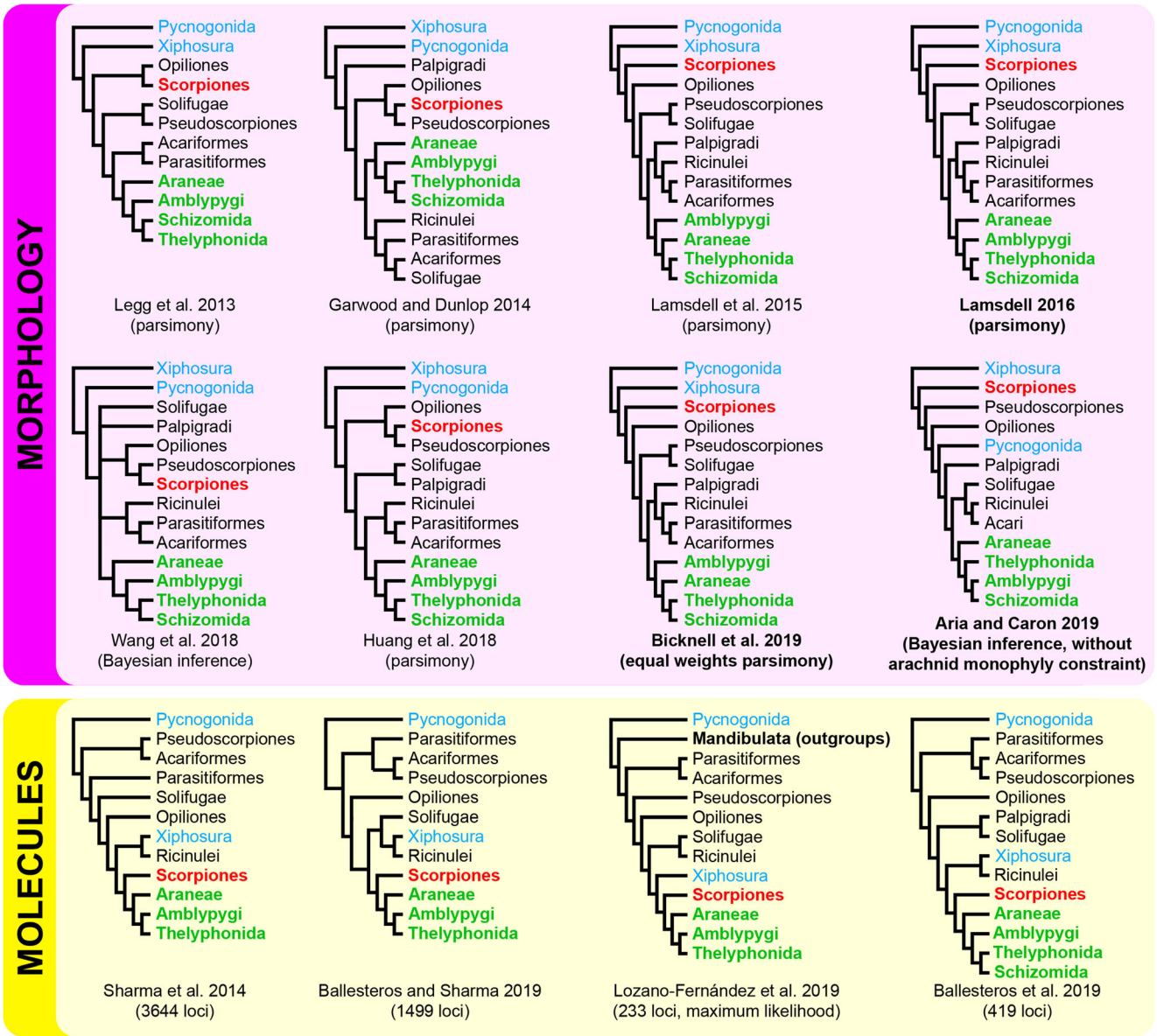

Fig. 1 Selection of phylogenetic hypotheses from the literature pertaining to the phylogeny of Chelicerata. Colors in branches correspond to labels in the center. Boldface text indicates data matrices reanalyzed by us in this study. For the Aria and Caron (2019) Bayesian inference analysis, the MrBayes tree block was run as specified by the authors, removing only the constraint for arachnid monophyly ($\ln L = -3904.47$)

Ballesteros et al. 2019). Interestingly, we and others have previously shown that arachnid monophyly can indeed be recovered with maximal support (Sharma et al. 2014a; Lozano-Fernandez et al. 2019; Ballesteros and Sharma 2019), but this result is highly unstable and unreproducible across datasets. Sharma et al. (2014a) recovered the monophyly of Arachnida as a function of retaining only the 500 slowest-evolving genes that are less prone to the effects of saturation, but they also recovered multiple clades that were recovered by slowly-evolving genes and were mutually exclusive; the recovery of a lineage in a slowly-evolving dataset alone is therefore not sufficient evidence of its topological accuracy. Among other approaches, Ballesteros and Sharma (2019) applied the gene-wise log likelihood score framework of Shen et al. (2017) to show that the monophyly of Arachnida could be recovered with maximal support in a trivial and contrived way if the investigator predicated the selection of genes in the matrix using the requirement of support for Arachnida. Problematically, the same approach could be used to obtain nonsensical phylogenies with maximal nodal support, such as the sister group relationship of Euchelicerata to Pancrustacea (fulfilling a nonsensical notion of "making horseshoe crabs, crabs again" as an example of an absurd relationship). After systematic dissection of phylogenetic signal Ballesteros and Sharma (2019) concluded that arachnids were not monophyletic and horseshoe crabs represented a relictual aquatic arachnid lineage.

Lozano-Fernandez et al. (2019) recovered the monophyly of arachnids as well in a 233-locus matrix with greater taxonomic sampling, but arachnid monophyly was only recovered in a Bayesian inference run that did not reach convergence; summary statistics of their Bayesian inference analysis in PhyloBayes-mpi (Lartillot and Philippe 2004; Lartillot et al. 2007) after excluding six "unstable" taxa (whose placement was stable and highly supported in their maximum likelihood analyses), were *maxdiff* = 0.28 and minimum sample size = 63. For reference, convergence of chains in PhyloBayes-mpi's documentation is defined as *maxdiff* < 0.1, and the effective sample size in any Bayesian inference analysis should exceed 200 to be considered consistent with convergence. In that same study, every other analysis of every other matrix recovered the non-monophyly of arachnids, with maximum likelihood analysis of their preferred matrix also failing to recover the monophyly of Chelicerata (Fig. 1). Curiously, Lozano-Fernandez et al. (2019) expressed unbridled confidence solely in the tree yielding arachnid monophyly, as well as the downstream interpretation of "a single, unreversed, colonization of land underpinning the evolutionary success of this group" (Lozano-Fernandez et al. 2019). While intuitively appealing and conceptually simple, their conclusion is at completely at odds with the evolutionary history and extant diversity of modern aquatic mites (both freshwater and marine groups)—ruling out entirely the notion of unreversed terrestrial colonization with complete epistemological certainty. While a full account of analytical and interpretive anomalies of Lozano-Fernandez et al. (2019) is beyond the scope of this work, we refer interested readers to Supplementary Text 2 of Ballesteros et al. (2019). In that study, which provides the first phylogenomic analysis sampling all extant orders of Chelicerata, our attempts to work with the datasets of Lozano-Fernandez et al. (2019) resulted in the discovery of a series of errors in phylogenomic matrix construction, inconsistency in analyses of saturation, and underlying lack of signal or support for arachnid monophyly in all of the matrices of Lozano-Fernandez et al. (2019). Specifically, we were unable to reproduce the preferred result of Lozano-Fernandez et al. (2019) using PhyloBayes-mpi with identical heuristics, even upon using their preferred tree as a starting point in our analysis (Ballesteros et al. 2019).

Regardless of the monophyly of arachnids, every phylogenomic analysis of Chelicerata to date (including the preferred tree of Lozano-Fernandez et al. 2019) has supported the stable sister group relationship of Scorpiones and Tetrapulmonata (=Arachnopulmonata sensu Sharma et al. 2014a), the group of arachnid orders that bears book lungs (Regier et al. 2010; Meusemann et al. 2010; Ballesteros and Sharma 2019; Ballesteros et al. 2019). Problematically for Lozano-Fernandez et al. (2019), the Paleozoic scorpion fossil record includes several aquatic groups, including specimens with controversial external gills (Selden and Jeram 1989; Poschmann et al. 2008; Dunlop 2010; Waddington et al. 2015). Their derived placement along the stem subtending crown-group scorpions is consistent with two possibilities: scorpions as a whole have either colonized land independently of other arachnid orders (implying convergent evolution of the book lung in Scorpiones and Tetrapulmonata), or the Paleozoic scorpions had secondarily recolonized aquatic habitats in the Paleozoic (Poschmann et al. 2008). The same research team of Lozano-Fernandez et al. (2019) had previously reviewed the scorpion fossil record under a node dating approach (Howard et al. 2019) using a small and incomplete subset of available scorpion phylotranscriptomic datasets produced by us (Sharma et al. 2015a, b, c, 2018; Santibáñez-López et al. 2019), and predicated their analysis on the preconception that all Paleozoic fossil scorpions must be terrestrial, as an extension of their preferred scenario of a single terrestrial colonization. Their interpretation disregarded the reconstruction of marine or brackish habitats in which many of these fossils are found, and requires the assumption of independent, post-mortem transportation events from terrestrial to aquatic habitats for all Paleozoic scorpion taxa. In addition, Howard et al. (2019) disregarded evidence of gill-like structures in Paleozoic taxa like *Waeringoscorpio* (Poschmann et al. 2008; reviewed by Di et al. 2018) due to the incorrect extrapolation that a non-epipodal origin of chelicerate opisthosomal book lungs (Sharma 2017) falsifies

the notion that these fossils bore gills. The contention of Sharma (2017) and Di et al. (2018) was that all chelicerate opisthosomal organs are non-epipodal, including the book lungs of extant arachnopulmonates and the book gills of horseshoe crabs (which likely constitute a fusion of telopodal and exopodal elements; see also the limb architecture of the biramous synziphosurine *Dibasterium durgae* in Briggs et al. 2012). This homology statement does not negate the observation of gills in groups like *Waeringoscorpio*, nor did Howard et al. (2019) offer any competing interpretation of these paired lamellate organs in multiple Paleozoic scorpion specimens.

As these controversies reveal, the phylogenetic placement of scorpions carries enormous weight in the evolutionary history of Chelicerata (Fig. 1). As already previously pointed out by Sharma (2017), arachnid monophyly has proven questionable both in morphological and phylogenomic datasets, with the most recent morphological trees of Chelicerata having to enforce arachnid monophyly to recover this desired outcome (Aria and Caron 2019; compare with Fig. 1), and failing to recover such stable groups as Euchelicerata (Arachnida + Xiphosura), and Pedipalpi (Amblypygi + Uropygi) (Fig. 1). As previously discussed by us (Sharma et al. 2014a; Sharma 2017; Ballesteros and Sharma 2019), the demise of Arachnida in modern datasets is no different from the decay of support for Atelocerata/Tracheata (i.e., Hexapoda + Myriapoda) 20 years ago, with the exception that there have always been fewer morphological data supporting arachnid monophyly than there were supporting Atelocerata. By contrast, the nested placement of scorpions as part of Arachnopulmonata is both robust and reproducible in molecular phylogenomic datasets (Sharma et al. 2014a; Lozano-Fernandez et al. 2019; Ballesteros and Sharma 2019; Ballesteros et al. 2019), analyses of genome architecture (Schwager et al. 2017; Leite et al. 2018), and neontological investigations of functional morphology (Klußmann-Fricke et al. 2014; Klußmann-Fricke and Wirkner 2016; Lehmann and Melzer 2019). Nevertheless, recent paleontological works continue to hold the traditional view of scorpion phylogenetic placement as branching at the base of a monophyletic Arachnida (Legg et al. 2013; Garwood and Dunlop 2014; Lamsdell et al. 2015; Lamsdell 2016; Wang et al. 2018; Huang et al. 2018; Bicknell et al. 2019) or a non-monophyletic Arachnida in a recent analysis (Aria and Caron 2019; Fig. 1).

One phylogenetic data class that may independently adjudicate the nested placement of scorpions stems from gene expression data from the paralogs retained from the shared genome duplication in the common ancestor of spiders and scorpions. Genome duplications have provided valuable benchmarks for assessing the accuracy of phylogenomic methods because their signature is readily recognizable using distributions of genetic distances between paralogs and analyses of gene order, particularly for conserved blocks of genes like Hox clusters (Dehal and Boore 2005; Salichos and Rokas 2014). When applied to a subset of chelicerate taxa, these analyses have supported the inference that the paralogs exhibited by spiders and scorpions resulted from a shared genome duplication, which excluded mites, ticks, and harvestmen (Opiliones) (Schwager et al. 2007; Sharma et al. 2014b, 2015a; Schwager et al. 2017), as well as horseshoe crabs (which have undergone an independent genome duplication/s on their stem; Kenny et al. 2016; Liao et al. 2019). As a result, paralogy has been shown to be systemic in genomes of spiders and scorpions (Leite et al. 2016). Comparative surveys of duplicated embryonic patterning genes in a spider and their single copy ortholog in the harvestman have pinpointed several candidate pairs exemplifying subfunctionalization and neofunctionalization in some paralog pairs of homeobox genes (Leite et al. 2018). However, comparable expression surveys do not exist for scorpions' duplicated paralogs, outside of the four posterior-most Hox gene pairs (Sharma et al. 2014b). Moreover, the expression patterns of duplications beyond homeobox genes has not been assessed in scorpions.

We therefore aimed to test the derived placement of scorpions, using as benchmarks four appendage-patterning genes whose expression domains have been previously well-characterized in spiders and/or harvestmen. These appendage-patterning genes were chosen for (a) clearly divergent expression domains between each pair of paralogs in spiders (e.g., Janssen et al. 2008; Pechmann and Prpic 2009), (b) the representation of both homeobox and non-homeobox transcription factors among the four homologs, and (c) the abundance of comparative datasets for other arthropods and non-arthropod taxa that serve as outgroups. We tested the hypothesis that if the spider and scorpion paralogs were the result of a shared duplication event in the arachnopulmonate ancestor, then the expression patterns of each scorpion gene copy should reflect similar expression domains as its spider ortholog, with the harvestman, mite, and non-chelicerate outgroup single-copy orthologs reflecting a putatively ancestral condition. We thus generated expression data from a spider, a harvestman, and a scorpion for all homologs of the four candidate genes to assess the retention of phylogenetic signal in developmental gene expression.

Materials and methods

A subset of morphological chelicerate phylogenies summarized in Fig. 1 have reported the monophyly of Arachnida, but did not show the internal topology of arachnids. We therefore reconstructed the morphological trees from three studies after downloading the respective data matrices from MorphoBank (Lamsdell et al. 2015) or requesting the data matrix directly from the authors (Bicknell et al. 2019). In the case of Aria and Caron (2019), we reran analyses after

removing the constraint forcing the monophyly of arachnids. Parsimony analyses were conducted in TNT v. 1.5 (Goloboff and Catalano 2016) under equal weights (xinact; hold 1000; xmult:replications 100,000; xmult = level 3; bb; best; unique). Bayesian inference searches were performed in MrBayes v.3.2.6 (Ronquist et al. 2012) under a 1-parameter Markov model. The consensus trees of these analyses are depicted in Fig. 1, focusing only on extant chelicerate relationships. Annotated tree files from these analyses, including tree costs and log likelihood values, are provided in supplementary material.

Adult females of the scorpion *Centruroides sculpturatus* were collected in Arizona (US) using UV illumination at night. Dissection of gravid females, tissue fixation, RNA extraction, and cDNA synthesis all follow our previously published protocols (Sharma et al. 2014b, c). Embryos of the spider *Parasteatoda tepidariorum* and *Phalangium opilio* were obtained from our colonies and fixed as previously described (Setton and Sharma 2018).

Homologs of the genes *dachshund* (*dac*), *homothorax* (*hth*), *extradenticle* (*exd*), and *optomotor blind* (*omb*) were identified by tBLASTn from the genomes of the *Parasteatoda tepidariorum* and *Centruroides sculpturatus* (Schwager et al. 2017), from the transcriptome of the vinegaroon *Mastigoproctus giganteus* (Sharma et al. 2014a), and from developmental transcriptomes and an unpublished genome of *Phalangium opilio* (Sharma et al. 2014a; Leite et al. 2018). Homologs were confirmed using best reciprocal blast hit. Amino acid sequences of panarthropod homologs with known expression patterns, as well as non-eczysozoan outgroups, were aligned using MUSCLE v. 3.6 (Edgar 2004); overhanging ends and ambiguous regions outside of the conserved domain were culled using GBLOCKS v. 0.91b (Castresana 2000). Maximum-likelihood analysis was performed using RAxML v. 8 (Stamatakis 2014) under an LG + Γ model with 500 independent searches and 1000 bootstrap resampling replicates. All alignments are provided as supplementary material.

Procedures for riboprobe synthesis and cloning of amplicons followed our recently detailed protocols (Setton and Sharma 2018). Amplicon identities were verified by Sanger sequencing. Primer sequences for templates are provided as supplementary material. Whole mount in situ hybridization for each species followed previously published protocols as well (Sharma et al. 2012a, 2014c; Setton and Sharma 2018). Approximately 15–20 embryos were assayed for every scorpion homolog, at each of two broadly defined stages: a limb bud stage ("LB") and a post-appendicular segmentation stage (PA). For each pair of paralogs, nearly synchronous clutches from single females were divided and expression examined in equivalent stages. Over 50 spider and harvestmen embryos were assayed for each of the remaining homologs. Sense probes were run as negative controls using half the number of samples (for each species and paralog) as for the anti-sense probes. Embryos were mounted in glycerol and images were captured using a Nikon SMZ 25 fluorescence stereomicroscope mounted with a DS-Fi2 digital color camera driven by Nikon Elements software. Appendage mounts of arachnid embryos were prepared following our previous approaches (Sharma et al. 2012b, 2015b) and imaged using an Olympus BX60 mounted with an Olympus DP72 camera and driven by Olympus cellSens software.

Results

Comparative gene expression of *dac* homologs

dac is expressed in a medial domain of the leg proximo-distal (PD) axis throughout Panarthropoda and in mandibulates (e.g., insects, malacostracan crustaceans) has been shown to be necessary for the medial leg territory. Previous work has shown that the single-copy ortholog of the harvestman *P. opilio* and the mite *Archegozetes longisetosus* is expressed in a conserved manner with respect to outgroup taxa (Sharma et al. 2012b; Barnett and Thomas 2013). Functional interrogation of *dac* in *P. opilio* previously showed that knockdown of *dac* leads to deletions of the entire femur and patella (Sharma et al. 2013). It was subsequently shown that spiders retain two copies of *dac*; *dac-1* is expressed in a manner comparable to the ancestral copy, whereas *dac-2* is expressed only at the patella-tibia boundary (Turetzek et al. 2015).

We recovered two copies of *dac* from the genome of *C. sculpturatus.* Phylogenetic analysis of a 54-taxon, 695-amino acid site alignment under maximum likelihood recovered a topology reflecting independent duplications in Xiphosura; mutually monophyletic clusters of *dac* paralogs corresponding to Arachnopulmonata; and the exclusion of Acari and harvestmen from these duplicated clusters (Fig. 2; $\ln L = -22{,}047.14$). Notably, the three putative copies of scorpion *dac* previously reported by Turetzek et al. (2015) from the genome of *Mesobuthus martensii* constitute only two copies—two of the tree *Mesobuthus* sequences ("Mma07092" and "MMa07100") have no overlapping sequence and correspond to different exons; addition of complete *C. sculpturatus dac* paralogs resulted in their clustering with the *Cscu-dac-2* sequence with near-zero branch lengths, indicating strong sequence conservation in amino acid sequences of the two buthid scorpions.

Cscu-dac-1, which clustered in the gene tree with spider *dac-1*, is expressed at the LB stage in the central nervous system, in several groups of cells in the head lobes, and in a median territory of all developing prosomal appendages. In the ventral ectoderm of the prosoma, the segmentally iterated expression domains of *Cscu-dac-1* are continuous and sinusoidal (Fig. 3A). In the developing appendages of this stage,

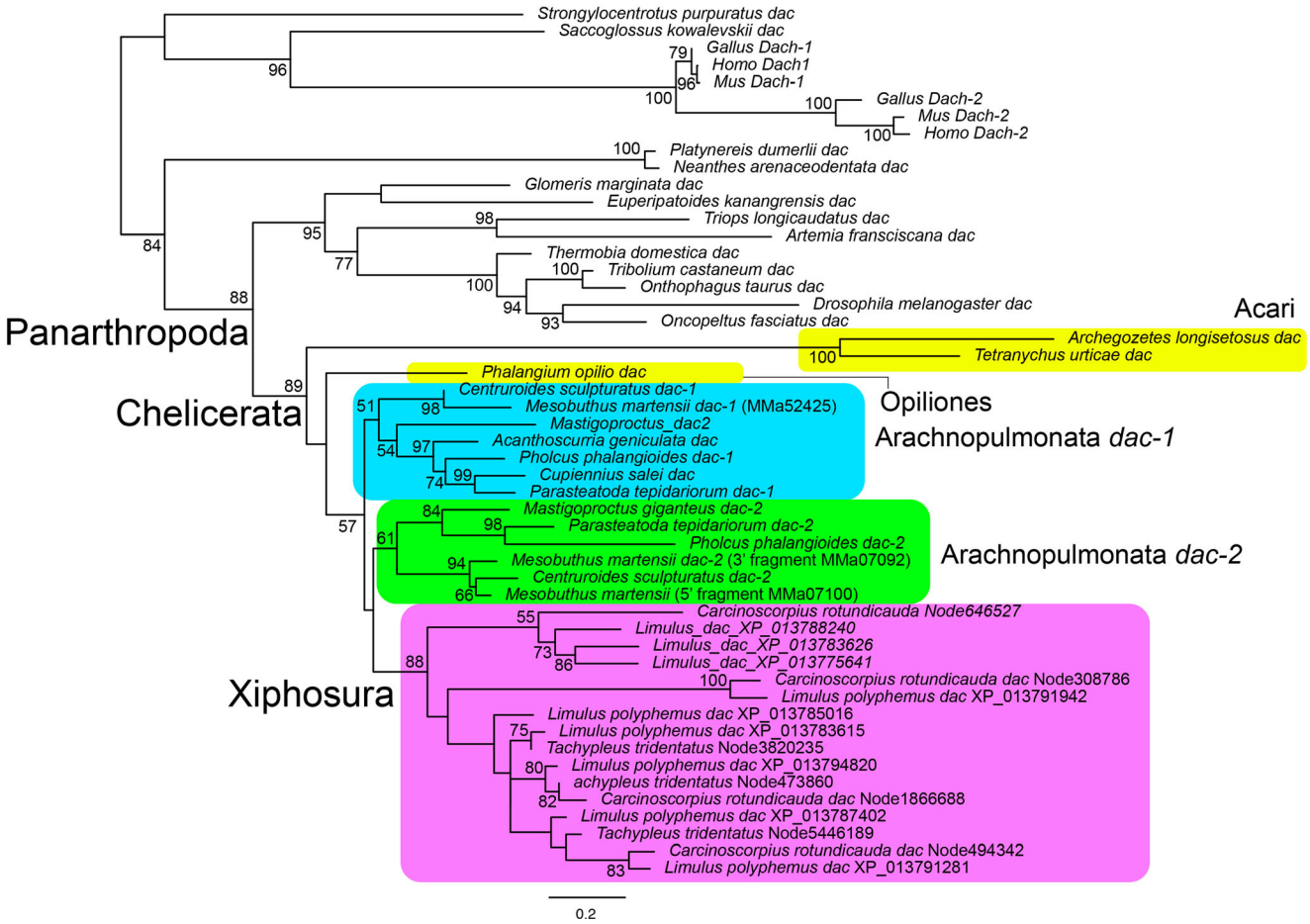

Fig. 2 Maximum likelihood gene tree topology for *dachshund* homologs (ln*L* = − 22,047.14). Numbers on nodes represent bootstrap resampling frequencies. Purple: Xiphosura. Blue: *dac-1* of Arachnopulmonata. Green: *dac-2* of Arachnopulmonata. Yellow: Non-arachnopulmonate arachnid orders

Cscu-dac-1 is expressed in the proximal segment of the three-segmented chelicera (Fig. 3E), like in the corresponding proximal segment of the harvestman chelicera (Sharma et al. 2012b). In addition, a dot of distal expression is detected in the chelicera in the second article of the chela. In the pedipalp and leg (Fig. 3F, G), *Cscu-dac-1* is strongly expressed in a medial territory, with addition discrete points of expression in the bifurcating pedipalpal chela. By the PA stage, expression in the ventral ectoderm and chelicera is diminished (Fig. 3B, H). Strong and complex expression is detected in the eye fields (Fig. 3B). In addition, strong expression is detected in the distal trochanter through the proximal patella of both the pedipalp and the walking leg (Fig. 3I, J).

Cscu-dac-2, which clustered in the gene tree with spider *dac-2*, was also expressed in the ventral ectoderm in early development, but expression patterns were distinct from *Cscu-dac-1* at corresponding stages. At the LB stage, expression in the ventral ectoderm consisted of segmentally iterating, discrete stripes (Fig. 3C), which were diminished but still visible at the PA stage, particularly in the metasomal segments (Fig. 3D). In the appendages, *Cscu-dac-2* was expressed similarly to *Cscu-dac-1*, with additional small expression domains in the distal regions of each appendage (Figs. 3K–2M). But by the PA stage, *Cscu-dac-2* was expressed more heterogeneously along the PD axis of the pedipalp and walking leg, with concentration of expression in the patella-tibia boundary (Fig. 3O–P). These expression patterns of the two paralogs along the PD axis correspond closely to those previously described in spiders (Turetzek et al. 2015).

On a corroborative note, the expression of both *dac* copies in the proximal segment of the scorpion three-segmented chelicera is consistent with our previous interpretation that the loss of this proximal expression domain precipitated the transition from the three- to the two-segmented cheliceral type (Sharma et al. 2012b, 2013).

Comparative gene expression of *hth* homologs

Expression of *Cscu-hth-1* at an early stage of development was previously described by us, as part of a comparative expression survey of *hth* homologs in a harvestman, a scorpion, and a horseshoe crab (Sharma et al. 2015b). In that work, we

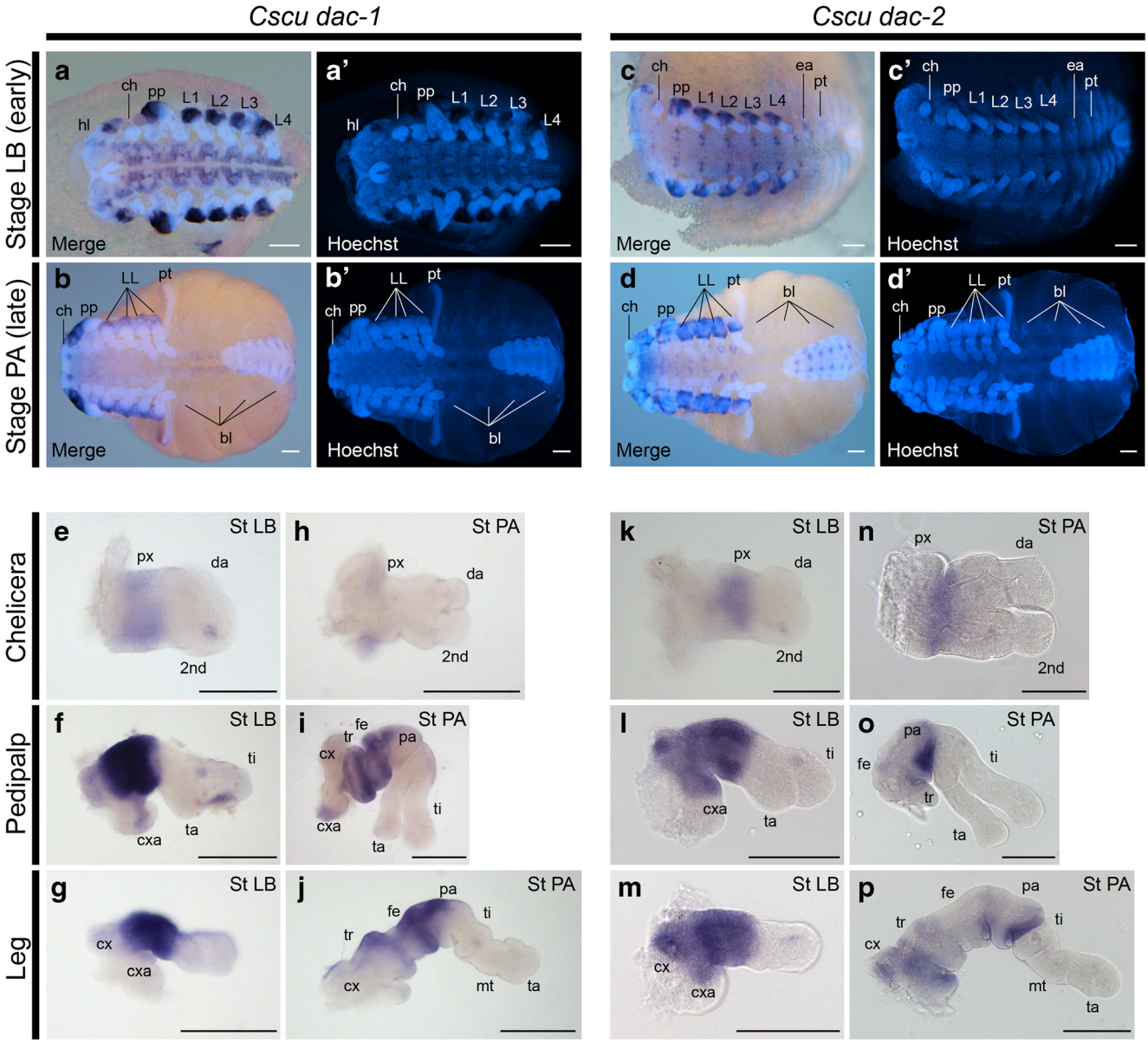

Fig. 3 Expression of *dac-1* and *dac-2* paralogs of the scorpion *C. sculpturatus*. *Cscu-dac-1* (A, B) in the prosoma of an LB stage scorpion embryo (A) and in a whole mount PA stage scorpion embryo (B). *Cscu-dac-2* (C, D) in the prosoma of an LB stage scorpion embryo (C) and in a whole mount PA stage scorpion embryo (D). Appendage mounts stained for *Cscu-dac-1* (E-J) show dynamic expression patterns between LB (E–G) and PA (H–J) stages. Appendage mounts for *Cscu-dac-2* (K–P) show comparable expression to *Cscu-dac-1* at the LB (K–M) stage, but distinguishable expression by the PA stage (N–P). Abbreviations: bl, book lung; ch, chelicera; cx, coxa; cxa, coxapophysis; da, distal article of chelicera; fe, femur; hl, head lobe; L1, first walking leg; mt, metatarsus; pa, patella; pp., pedipalp; pt., pectine; px, proximal segment of chelicera; ta, tarsus; ti, tibia; tr, trochanter; 2nd, second article of chelicera. (A'–D') Counterstaining of embryos in (A-D) with Hoechst 33342. Scale bars: 200 μm

previously showed that knockdown of *Popi-hth* in the harvestman *P. opilio* incurred both proximal leg defects, as well as homeotic chelicera-to-leg and pedipalp-to-leg transformations (Sharma et al. 2015b). This phenotypic spectrum, which constitutes the only functional data for *hth* in Chelicerata, accords closely with counterpart functional data in pancrustacean exemplars like crickets (Ronco et al. 2008). In addition, expression surveys by (Pechmann and Prpic 2009) and (Turetzek et al. 2017) have separately shown that the two copies of *hth* retained in multiple spider species exhibit divergent expression patterns: *hth-1* generally has a continuous expression domain through much of the PD axis, whereas *hth-2* is heterogeneously expressed in stripes corresponding to appendage segmental boundaries. In the mite *A. longisetosus*, the single *hth* copy is expressed comparably to harvestmen and to spider *hth-1*, with broad continuous expression through the proximal and medial territory of the walking leg (Barnett and Thomas 2013).

In the present study, we recovered two copies of *hth* from the genome of *C. sculpturatus*, consistent with the homeobox gene survey of Leite et al. (2018). Phylogenetic analysis of a 46-taxon, 630-amino acid site alignment under maximum likelihood recovered a topology again reflecting independent duplications in Xiphosura; two clusters of *hth* paralogs corresponding to Arachnopulmonata; and the exclusion of Acari and harvestmen from these duplicated clusters (Fig. 4; ln*L* = − 10,871.69).

We thus established that the scorpion *hth* copy whose expression at an early stage of development (LB) was previously described by us was *hth-1* (Fig. 4). Consistent with surveyed spider *hth-1* expression patterns, *Cscu-hth-1* was expressed in a continuous domain in all podomeres of the walking leg except the metatarsus and the tarsus (Fig. 3D–F of Sharma et al. (2015b). Here, we focused our investigation on the older stage for both paralogs.

Expression of the *Cscu-hth-1* at the PA stage consists of a strong, continuous expression domain in the chelicerae, with expression diminished only in the distal fingers of the chela (Fig. 5A), comparable with the expression pattern of this paralog at an earlier stage (Fig. 3D of Sharma et al. (2015b). In the pedipalp, expression of *Cscu-hth-1* changes during development. In early development, expression consists of a continuous domain extending up to the bifurcation of the pedipalpal chela (Fig. 3E of Sharma et al. 2015b). By the PA stage *Cscu-hth-1* is more heterogeneously expressed, with strong expression in the coxa and body wall, the dorsal component of the distal femur, and the patella-tibia boundary (Fig. 5B). Weaker expression domains persist throughout the PD axis in this stage. In the walking leg, *Cscu-hth-1* expression is diminished by the PA stage (Fig. 5C; compare with Fig. 3F of Sharma et al. 2015b), with strong expression restricted to the coxa and body wall, and weak, heterogeneous expression domains throughout the remaining podomeres.

Cscu-hth-2 expression is not distinguishable from *Cscu-hth-1* at the LB stage (not shown), comparably with the *dac* paralogs (ref. Fig. 3). The expression domains of the *hth* paralogs are distinct by the PA stage in each of the prosomal segments. In the chelicera, *Cscu-hth-2* is expressed from the proximal segment to the bifurcation of the chela, as well as the distal tips of both cheliceral fingers; the mobile digit more

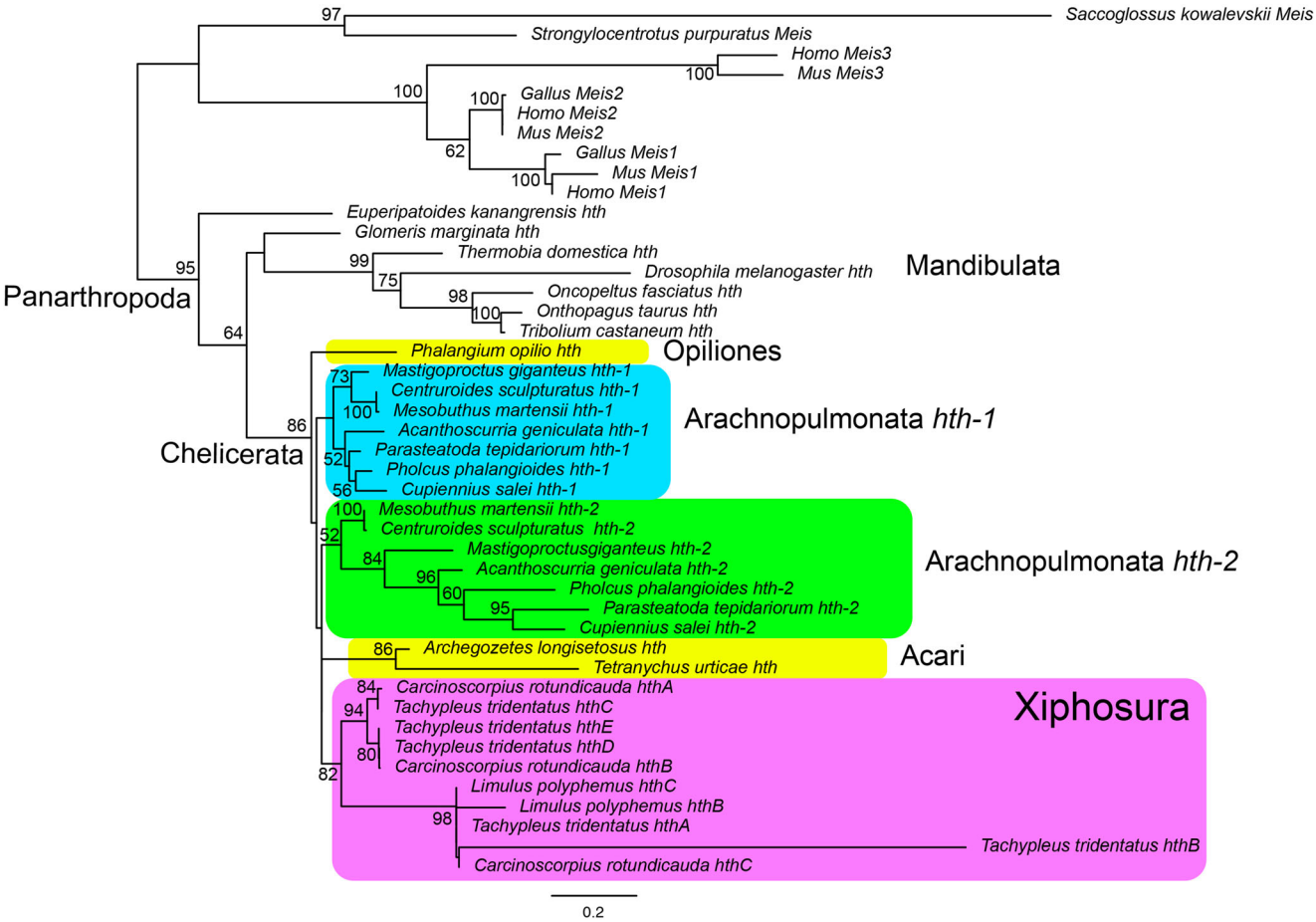

Fig. 4 Maximum likelihood gene tree topology for *homothorax* homologs (ln*L* = − 10,871.69). Numbers on nodes represent bootstrap resampling frequencies. Purple: Xiphosura. Blue: *hth-1* of Arachnopulmonata. Green: *hth-2* of Arachnopulmonata. Yellow: Non-arachnopulmonate arachnid orders

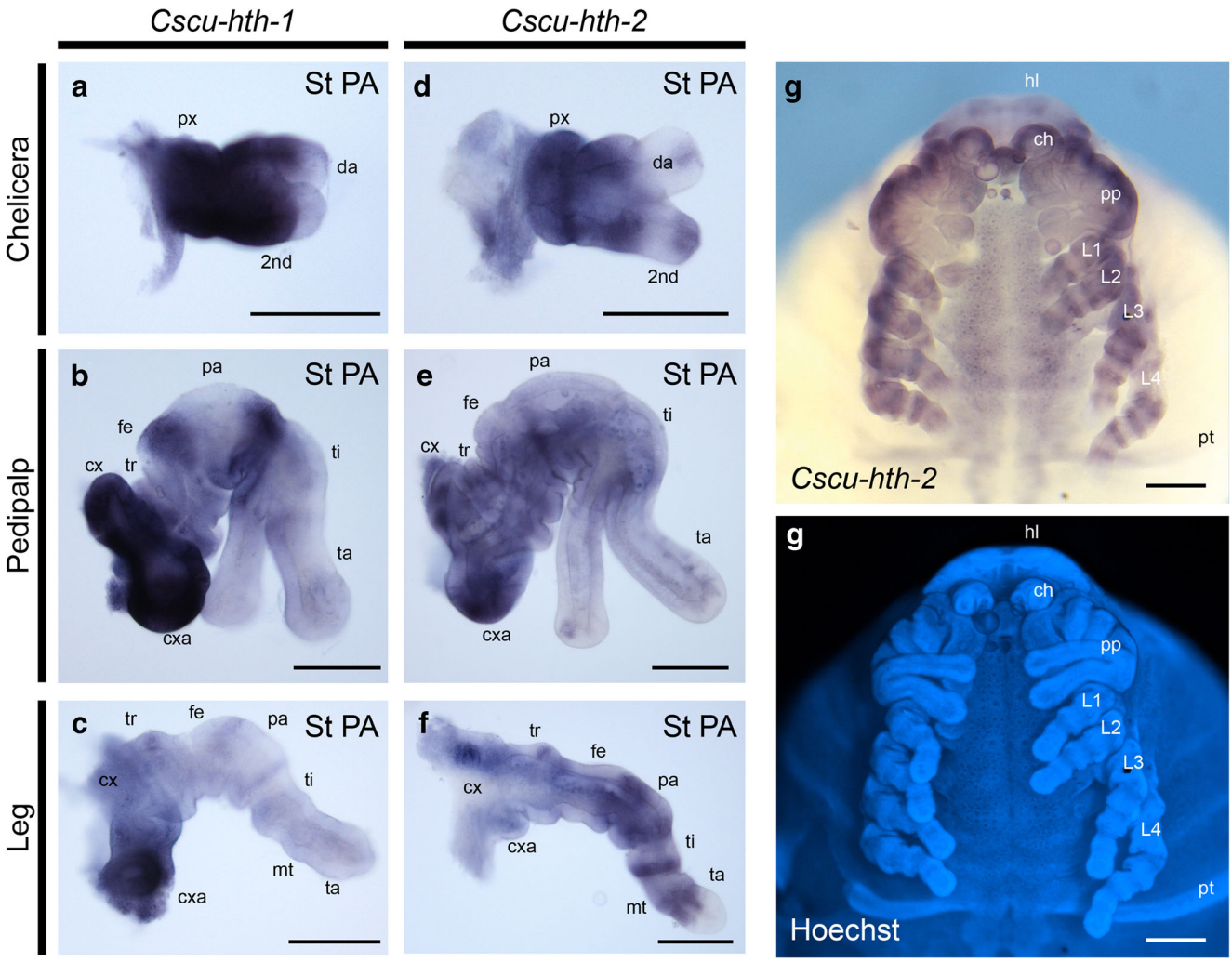

Fig. 5 Expression of *hth-1* and *hth-2* paralogs of the scorpion *C. sculpturatus* at the PA stage. Appendage mounts stained for *Cscu-hth-1* (A–C) and *Cscu-hth-2* (D–F) show distinct expression patterns in each prosomal appendage type. In a whole mount embryo stained for *Cscu-hth-2* (G), stripes of expression are clearly seen in the walking legs, in addition to weaker expression in the ventral ectoderm, head lobes, and pectines. Abbreviations as in Fig. 4. (G') Counterstaining of embryo in (G) with Hoechst 33342. Scale bars: 200 μm

strongly expressed *Cscu-hth-2* than the finger of the second article (Fig. 5D). In the pedipalp, *Cscu-hth-2* expression is nearly continuous from the coxa to the bifurcation of the pedipalpal chela. Additional discrete expression domains are observed in the distal tips of the pedipalpal chela (Fig. 5E). In the walking leg, *Cscu-hth-2* is strongly expressed in heterogeneous domains, with stronger bands of expression at the distal segmental boundaries of the femur, the patella, the tibia, and the metatarsus. The intervening tissue between these boundaries weakly expresses *Cscu-hth-2*; the tarsus does not visibly express this paralog (Fig. 5F). This expression pattern in the leg somewhat resembles the expression of *Cscu-hth-1* at an earlier developmental stage (Fig. 3 of Sharma et al. 2015c), with the distinction of the additional stripe of expression at the metatarsal-tarsal boundary. Heterogeneous and weak expression domains of *Cscu-hth-2* also occur in the head lobes, the ventral ectoderm, and part of the pectine (Fig. 5G).

Comparative gene expression of *exd* homologs

In the leg of the harvestman *P. opilio*, the single-copy *exd* homolog is expressed in two discrete territories: in the proximal podomeres and in the patella-tibia boundary (Sharma et al. 2012b). In the mite *A. longisetosus*, the single-copy *exd* homolog is expressed uniformly in only the proximal part of the PD axis (Barnett and Thomas 2013). No functional data are available for any chelicerate *exd* homologs. The incidence of two copies of *exd* in a spider was first described in the mygalomorph *Acanthoscurria geniculata* (Pechmann and Prpic 2009). In contrast to *dac* and *hth*, the distinction between these paralogs' expression domains is subtle and limited to the chelicerae. In this tarantula, *exd-1* is expressed throughout the cheliceral PD axis, whereas *exd-2* expression extends from the body wall to a point just proximal of the fang-basal segment boundary (Fig. 7E, F of Pechmann and Prpic 2009).

Expression of the two *exd* paralogs in the pedipalps and legs is nearly indistinguishable in this spider, with an additional weak ring of expression at the trochanter-femur boundary distinguishing *exd-2*.

We recovered two copies of *exd* from the genome of *C. sculpturatus*, consistent with the homeobox gene survey of Leite et al. (2018). Phylogenetic analysis of a 40-taxon, 271-amino acid site alignment under maximum likelihood recovered a topology that exhibits little phylogenetic signal within Chelicerata (note polytomies within Chelicerata; Fig. 6A, ln$L$ = − 2436.11). Due to the exclusion of one of the four *exd* homologs of the horseshoe crab *Limulus polyphemus* from a cluster of Xiphosura paralogs, a scenario of duplications specific to Xiphosura was not supported, although both *Limulus polyphemus* and *Tachypleus tridentatus* exhibited the four paralogs expected of the two-fold whole genome duplication previously described at the base of the horseshoe crabs (Kenny et al. 2016; Liao et al. 2019). We also did not observe the clustering of arachnopulmonate *exd* paralogs due to the exclusion of a *Cupiennius salei* spider sequence from the *exd-2* cluster and a *C. sculpturatus* sequence from the *exd-1* cluster. The lack of signal in this alignment is attributable to the brevity of this alignment (< 300 sites, compared with > 600 sites for *dac* and *hth*) and the high conservation of the remaining sites after treatment with GBlocks v. 0.91b. We therefore denoted *Cscu-exd-2* as the copy that clustered with the *exd-2* copy of spiders, and *Cscu-exd-1* as the more distantly branching copy (Fig. 6A).

Expression of *Cscu-exd-1* showed strong expression throughout the chelicera, with slightly weaker expression in the distal tips of the chela (Fig. 6B), consistent with the *exd-1* copy of *A. geniculata*. However, expression of this paralog in the pedipalp and leg diverged from the conserved expression domains of spider counterparts. In the pedipalp, *Cscu-exd-1* was strongly expressed in the coxa and body wall, as well as in stripes at the distal segmental boundaries of the femur and patella, and the joint between the tibia and the mobile digit of the pedipalpal chela. Concentration of transcript abundance was also observed in a subdistal territory of the mobile digit. Weaker expression was observed in the intervening territories of the femur, patella, and proximal tibia (Fig. 6C). In the walking leg, heterogeneous expression in most of the PD axis was similarly observed, with concentrated expression in the coxa and coxapophysis, as well as stripes of expression at the distal segmental borders of the patella and tibia (Fig. 6D). *Cscu-exd-1* expression thus differs from its spider orthologs.

Cscu-exd-2 is expressed differently from *Cscu-exd-1* in all three prosomal appendage types. In the chelicerae, *Cscu-exd-2* is expressed from the body wall to a boundary just proximal of the bifurcation of the chela. In contrast to *Cscu-exd-1*, *Cscu-exd-2* is not visibly expressed in the distal tips of the chelicera (Fig. 6E), like its ortholog in *A. geniculata* (Fig. 7E, F of Pechmann and Prpic 2009). In the pedipalp, *Cscu-exd-2* is strongly expressed in the coxa and coxapophysis, the region of the tibia subtending the fingers of the pedipalpal chela, and a subdistal territory in the fixed finger of the pedipalpal chela. Weak expression of *Cscu-exd-2* is observed between the coxa and the tibia (Fig. 6F). In the walking leg, *Cscu-exd-2* is not strongly expressed in the coxa or coxapophysis. Intriguingly, two clear stripes of expression are detected at the trochanter-femur boundary and a strongly stripe in the patella-tibia joint. The expression of *Cscu-exd-2* is thus strongly comparable to its *A. geniculata* ortholog (Fig. 6F; compare with Fig. 7R of Pechmann and Prpic 2009).

Comparative gene expression of *omb* homologs

A survey of three spider species has shown that two copies of *omb* occur in *Cupiennius salei* and *Tegenaria atrica* with only a single copy previously known for *A. geniculata* and *P. tepidariorum* (Janssen et al. 2008). In that work, Janssen et al. (2008) demonstrated that the *omb-2* paralog is consistently expressed in the dorsal territory of the prosomal appendages, whereas *omb-1* is expressed in that territory as well as at least three mesodermal regions of the pedipalps and walking legs. No chelicerate *omb* expression data exist apart from these spider datasets.

We recovered two copies of *omb* from the genome of *C. sculpturatus*, as well as a single-copy homolog of *omb* in *P. opilio*. In addition, we discovered in the genome of *P. tepidariorum* the missing paralog that was not previously discovered by the approach of Janssen et al. (2008) and added it to the alignment. Phylogenetic analysis of the ensuing 17-taxon, 146-amino acid site alignment under maximum likelihood recovered an *omb* gene tree topology exhibiting almost no phylogenetic signal (Fig. 7A, ln$L$ = − 762.26). Due to the paucity of panarthropod taxa for which *omb* expression domains have been surveyed, as well as the brevity of the conserved region in the alignment, this multiple sequence alignment was the most data-poor of the four we generated and bore almost no informative sites for splits within Arthropoda.

Of the two *omb* copies of *C. sculpturatus* that we discovered, we established that one of these was clearly expressed only in the dorsal territory of the PD axis (*Cscu-omb-2*; Fig. 7E–G), whereas the other was expressed in this dorsal territory as well as three mesodermal domains in the pedipalp and the walking leg (*Cscu-omb-1*; Fig. 7B–D). In the chelicerae, the two copies were additionally distinguished by expression in the proximal segment, with *Cscu-omb-1* expressed only in the dorsal territory of the proximal segment, and *Cscu-omb-2* expressed throughout this podomere.

To enable polarization of sub- versus neofunctionalization between these copies, we additionally generated expression data for the single-copy ortholog of the harvestman. Intriguingly, *Popi-omb* is expressed only in the dorsal territory of the prosomal appendages (Fig. 7H–J) in all embryonic stages surveys (stages 12–15), which suggests that the *omb-*

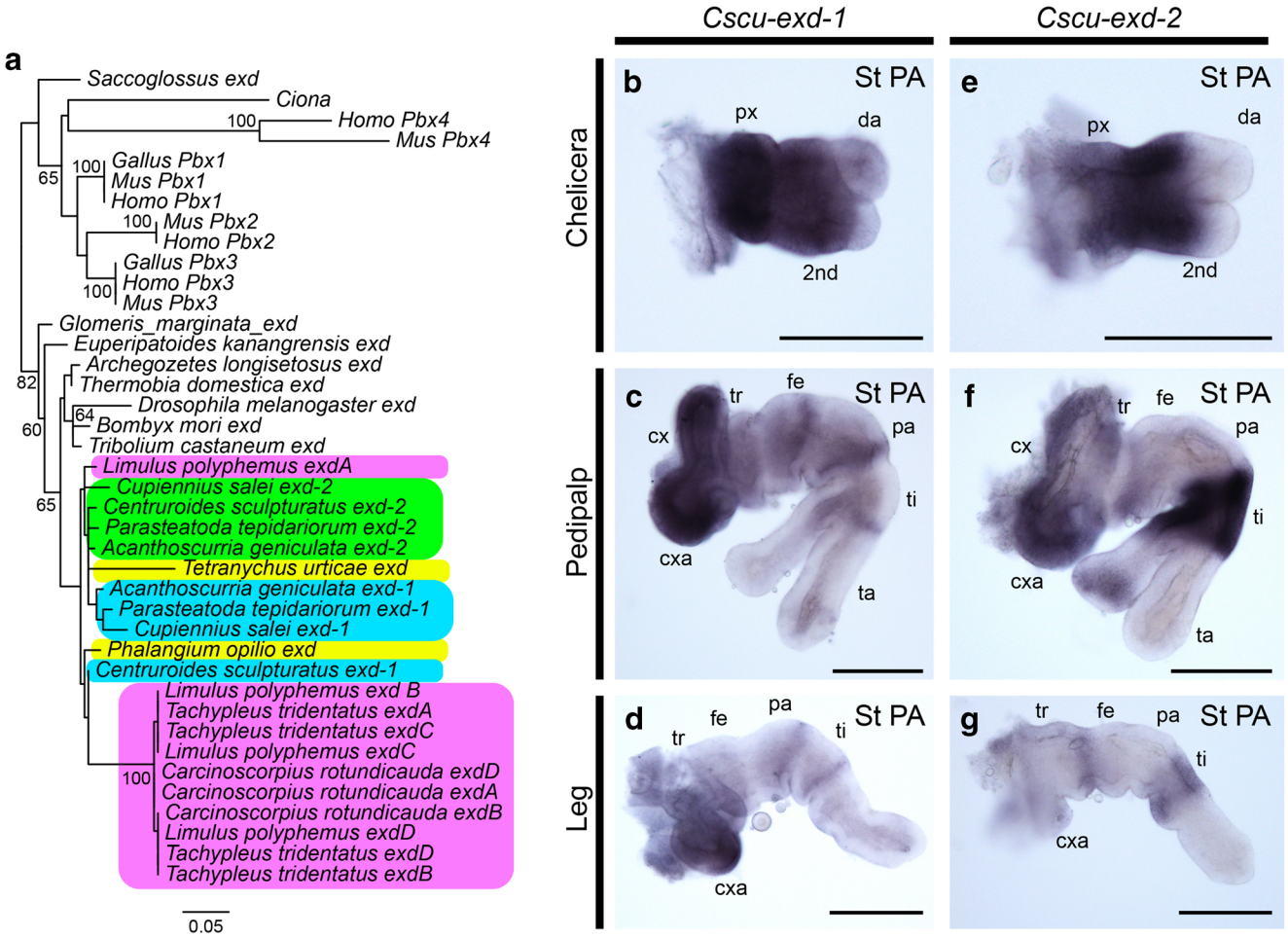

Fig. 6 (A) Maximum likelihood gene tree topology for *extradenticle* homologs (ln*L* = − 2436.11). Numbers on nodes represent bootstrap resampling frequencies. Purple: Xiphosura. Blue: *exd-1* of Arachnopulmonata. Green: *exd-2* of Arachnopulmonata. Yellow: Non-arachnopulmonate arachnid orders. Appendage mounts stained for *exd-1* (B-D) and *exd-2* (E-G) paralogs of the scorpion *C. sculpturatus* reveal distinct expression patterns in each prosomal appendage type by the PA stage. Abbreviations as in Fig. 4. Scale bars: 200 μm

2 copy of arachnopulmonates has retained the ancestral expression domain.

As an internal test of the conservation of expression domains, we assayed the expression pattern of the *P. tepidariorum omb* paralog we discovered, predicting that it should resemble the expression of arachnopulmonate *omb-1* (the copy reported by Janssen et al. 2008 exhibits) exhibits expression domains that resemble other spiders' *omb-2* expression). Consistent with this prediction, *Ptep-omb-1* is expressed in the dorsal territory of all prosomal appendages, with three additional mesodermal domains in the pedipalps and walking legs (Fig. 7K–N).

Discussion

Gene expression patterns of ancient paralogs as a phylogenetic data class

The discovery of rare genomic changes (RGCs), generally defined as any class of infrequently occurring genomic mutations that are inferred to be homoplasy-free (or nearly so), was proposed nearly twenty years ago as a solution to particularly recalcitrant nodes in the Tree of Life (Rokas and Holland 2000). Candidates for RGCs in the two decades hence have include insertions of transposable elements, gene order, gene or genome duplications, gains or losses of genetic modules such as introns and protein functional domains, chromosomal inversions or duplications, and microRNAs. RGCs were historically regarded as ideal characters by virtue of being slow-evolving, clearly interpretable as homologous, and simple to analyze in a parsimony framework due to their rarity. RGCs have been applied with some success in resolving some of the most challenging parts of the Tree of Life and thus provided valuable benchmarks for assessing the accuracy of phylogenomic methods (Hazkani-Covo 2009; Salichos and Rokas 2014; Fröbius and Funch 2017). Propitiously for Chelicerata, the incidence of a distal genome duplication uniting a subset of arachnid orders provides precisely the kind of RGC that can be surveyed efficiently by leveraging high-quality genomic and transcriptomic datasets, in tandem with

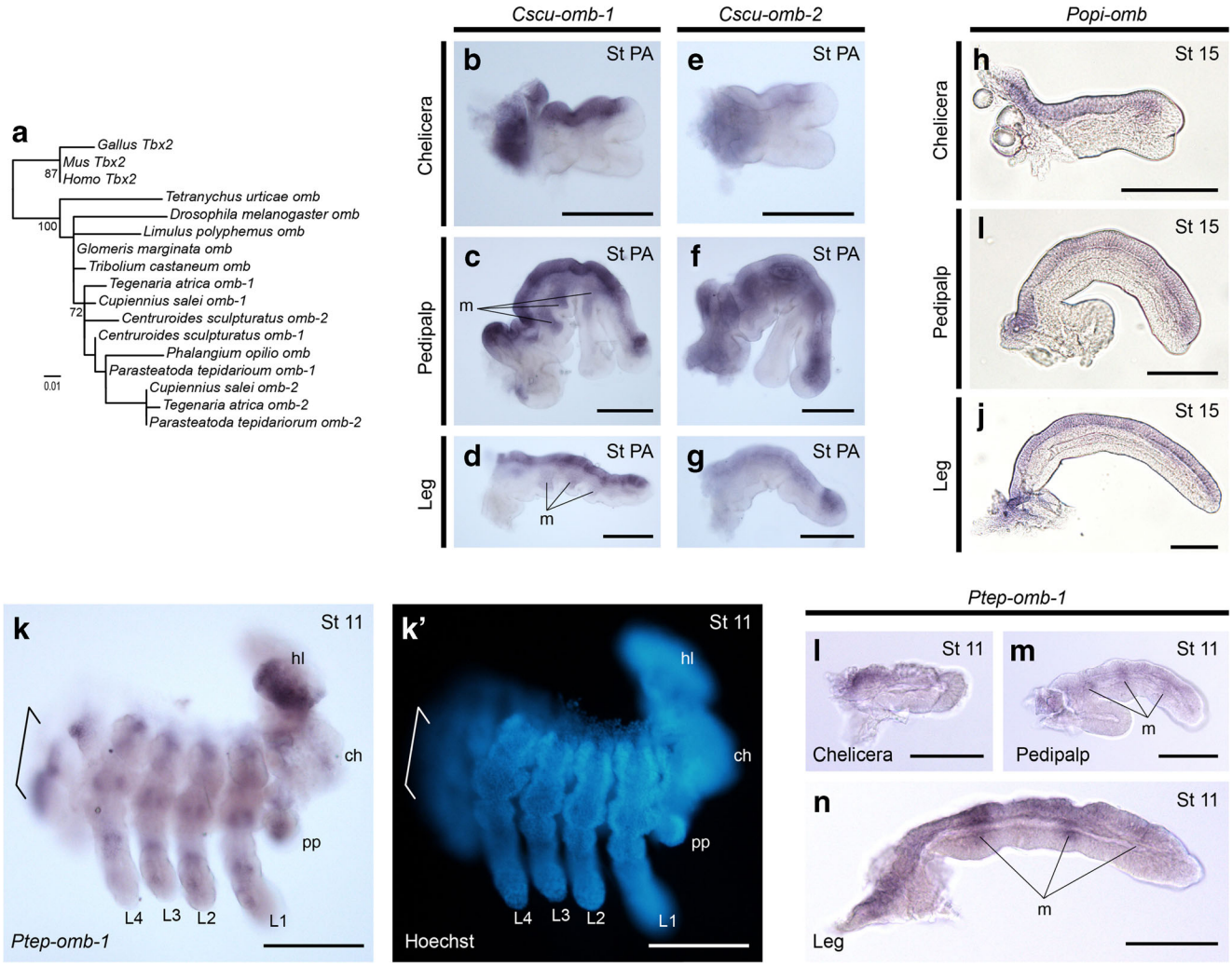

Fig. 7 (A) Maximum likelihood gene tree topology for *optomotor blind* homologs (ln*L* = − 762.26). Numbers on nodes represent bootstrap resampling frequencies. In the scorpions, appendage mounts stained for *Cscu-omb-1* (B–D) and *Cscu-omb-2* (E–G) show distinct expression patterns in each prosomal appendage type. Note the mesodermal ("m") expression unique to *Cscu-omb-1* (m). In the harvestman, appendage mounts stained for *Popi-omb* (H–J) show expression domains restricted to the dorsal territory of all prosomal appendages, comparably to arachnopulmonate *omb-2*. The missing *omb* homolog of the spider, predicted to be orthologous to arachnopulmonate *omb-1*, is shown in (K–N). (K) Lateral view of a bisected stage 11 embryo after removal of yolk, showing strong expression in the head lobe, parts of the opisthosomal body wall (bracket), and in discrete parts of the mesoderm of prosomal appendages. Appendage mounts of *Ptep-omb-1* show the expected pattern of dorsal territory and mesodermal expression (m) in the pedipalp (M) and leg (N), but only the dorsal expression domain in the chelicera (L). Abbreviations as in Fig. 4. Scale bars: 200 μm for B–G, 100 μm for H–N)

developmental study of emerging model organisms (Leite et al. 2018).

In our survey of scorpion leg-patterning genes with known duplicates in spiders, we established that in two cases (*dac* and *hth*), the gene trees produced by multiple sequence alignments recover topologies where paralogs with shared expression patterns were clustered, to the exclusion of the single-copy orthologs of Acari and Opiliones (Figs. 2 and 4). In these same topologies, Xiphosura homologs (or fragments of homologs) formed an independent cluster of paralogs, corroborating the previous inference of independent genome duplications at the base of Xiphosura (twofold duplication) and Arachnopulmonata (Kenny et al. 2016; Schwager et al. 2017; Liao et al. 2019). In addition, the exclusion of Acari and Opiliones from these clusters in the gene trees cannot be attributed to gene loss after a more ancient shared duplication (e.g., in the common ancestor of Chelicerata). Published genomes of several Acari, as well as our unpublished *P. opilio* genome, have revealed a single Hox cluster in these taxa, by comparison with the two Hox clusters of Arachnopulmonata, or the two to four of Xiphosura (Kenny et al. 2016; Leite et al. 2018; Liao et al. 2019). While losses of duplicated genes are common after a genome duplication, the independent losses of entire clusters of genes in multiple arachnid orders, without any signature of such an event in gene trees or in the genomes, is an implausible scenario.

Gene trees are only as informative as their underlying alignments, however, and in the case of *exd* and *omb*, we were unable to retain amino acid alignments of a size sufficient for informing evolutionary relationships between paralogs after trimming with GBlocks v. 0.91b (< 300 sites for both *exd* and *omb*, compared with > 600 sites for *dac* and *hth*). For this reason, we turned to expression patterns to assess the conservation of expression patterns in each of four pairs of paralogous genes. Consistent with a scenario of an ancient shared arachnopulmonate genome duplication, the expression patterns of scorpion gene copies exhibited clearly identifiable similarities to their spider orthologs. Some divergence of expression patterns relative to spiders was observed in the scorpion, particularly for the *hth* and *exd* paralogs. In addition, expression patterns in the scorpion pedipalp in some cases do not resemble known patterns of other chelicerates; this divergence may reflect the disparate morphology of the scorpion pedipalp, which does not resemble the non-chelate counterpart found in spiders and harvestmen.

Nevertheless, in all four surveyed cases, expression patterns in the walking legs facilitated clear identification of orthology, which was consistent with that predicted by the gene trees in the case of *dac* and *hth*. Taken together with expression surveys in outgroup taxa, and principally in the closely related Opiliones, we are therefore able to reconstruct the evolutionary history of the paralogs ensuing from the genome duplication, and for the first time, polarize these characters as states on a phylogenetic tree sampling a minimum of three chelicerate taxa (Fig. 8).

The origin of *dac-2* postdates the evolution of the arachnid patella

One of the interesting implications of this work concerns a previous claim of neofunctionalization of *dac-2* in *P. tepidariorum*. Turetzek et al. (2015) had previously shown that RNAi against the *dac-2* copy of the house spider incurs the loss of the patella-tibia joint. Given that the patella is unique to arachnids and horseshoe crabs, Turetzek et al. (2015) claimed that their functional data substantiated the role of *dac-2* as the developmental causal gene whose neofunctionalization drove the evolution of the patella.

Several aspects of their work call their conclusion into question. First, Turetzek et al. (2015) generated *dac-2* loss-of-function phenotypes, but never did so for *dac-1*. Moreover, no *dac-1* functional data exist in any spider species. The interpretation of neofunctionalization was therefore not substantiated, because it was never shown that the other *dac* copy had retained the ancestral function of patterning the medial appendage territory. Second, Turetzek et al. (2015) did not discuss the only other chelicerate *dac* functional data published at that time. In *P. opilio*, we were able to show that knockdown of *dac* resulted in the loss of the entire femur and patella (Sharma et al. 2013). The putatively ancestral function of the single-copy ortholog in Chelicerata would therefore be interpreted to pattern the entire territory spanning the patella; invoking neofunctionalization within an ancestral domain that is deleted upon knockdown is a problematic inference at best. Third, Turetzek et al. (2015) published a gene tree in which several spiders and the "three" *Mesobuthus martensii* copies (see Fig. 2 for the two scorpion *dac* copies) were used to represent Chelicerata. This taxonomic sampling was also problematic; transcriptomic and genomic datasets were available at the time for Xiphosura, Opiliones, and multiple species of Acari. Given that a patella occurs in all arachnids and horseshoe crabs, if *dac-2* were responsible for the origin of the patella, it would stand to reason that all arachnid and horseshoe crab taxa should exhibit two copies of *dac*, and these should be resolved into each of the two clusters (*dac-1* and *dac-2*).

The gene tree we recovered for *dac* homologs within Chelicerata (Fig. 2) falsifies the interpretation of Turetzek et al. (2015). The origin of *dac-2* does not predate the origin of Arachnopulmonata and therefore cannot be causally linked to the origin of the patella. It is further implausible and ad hoc to propose a scenario in which non-arachnopulmonate arachnid orders initially had a *dac-2* copy but then subsequently lost it, with rescue of the patella-patterning function by another gene; under this scenario, the retained copy in non-arachnopulmonate arachnids would still be nested inside one of the two *dac* clusters retained by Arachnopulmonata. Given that (1) *dac* occurs as single-copy in Acari and Opiliones genomes, (2) Acari and Opiliones exhibit patellae, (3) the four copies of *dac* in Xiphosura are independently derived, (4) Xiphosura also exhibit patellae, and (5) severe *dac* loss-of-function phenotypes of *P. opilio* undergo deletions of the entire femur and patella, a scenario consistent with the gene tree and genomic architectures of chelicerates is that the *dac-2* function reflects a subfunctionalization of the ancestral chelicerate *dac* copy. Thus, the identity of the developmental causal gene responsible for the origin of the patellar segment has not yet been established and remains at large.

On a similar note, Turetzek et al. (2017) surveyed *hth* paralogs in multiple spider species, concluding that these may have exhibited rapid diversification after duplication in spiders, which may be linked to the diversification of spiders. As shown in this work (Fig. 4), the duplication of *hth* and the divergence of its expression patterns is not restricted to spiders, but is shared across all arachnopulmonates (represented in this study by scorpions and the thelyphonid species *Mastigoproctus giganteus*). Given that five arachnid orders putatively share *hth* copies, we speculate that the differences in expression patterns between species reflects drift as a result of functional redundancy; a scenario of rapid neofunctionalization that subtends only extant spider diversity is inconsistent with the timing and incidence of *hth* paralogs in Arachnopulmonata.

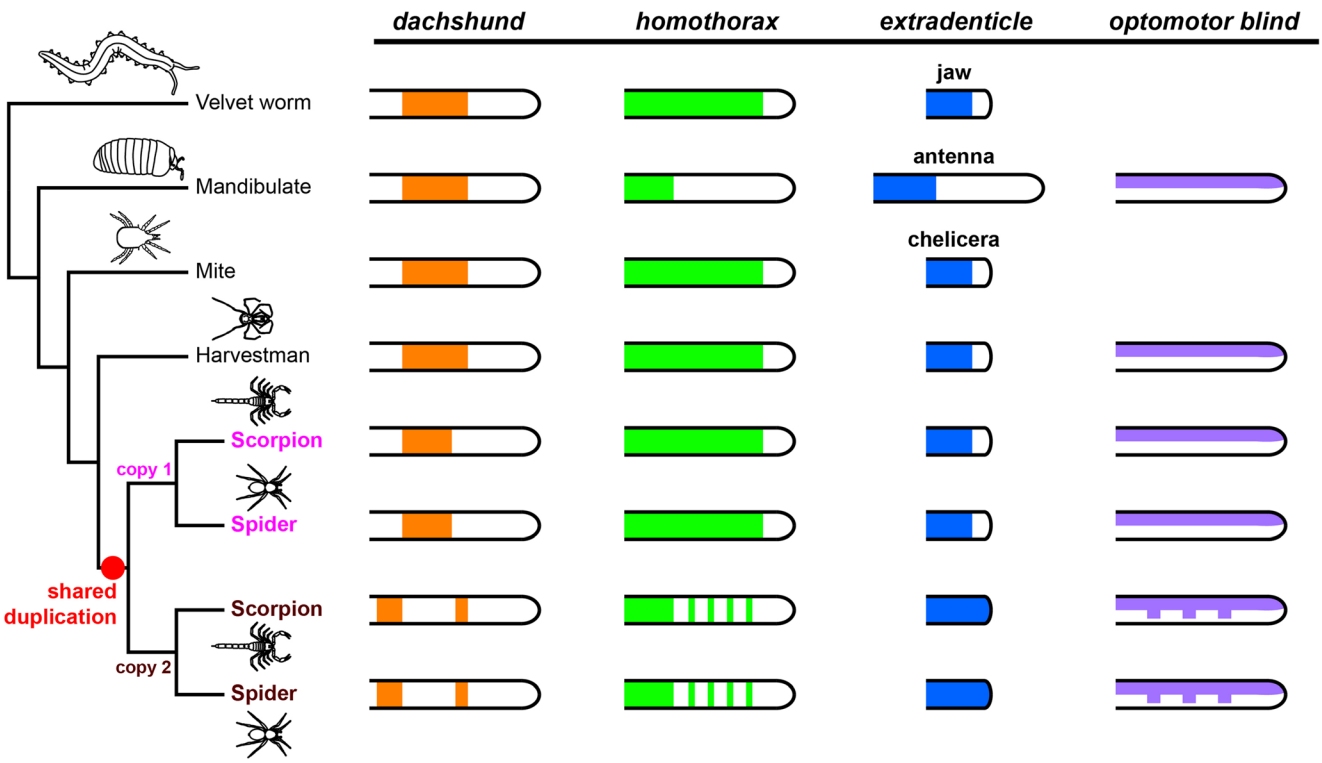

Fig. 8 Summary of gene expression data surveyed for Chelicerata and non-chelicerate outgroups. Schematics represent walking legs for *dac*, *hth*, and *omb*; and deutocerebral appendages (chelicera, antenna, or jaw) for *exd*. For each appendage schematic, proximal is to the left, distal to the right

A new view of chelicerate evolution

The observation of systemic gene duplications in scorpions and other arachnopulmonates (Schwager et al. 2007; Sharma et al. 2014b, 2015a; Schwager et al. 2017; Leite et al. 2018), together with the validation of predicted orthology using gene expression patterns in this work, is strongly consistent with phylogenomic datasets that robustly support recovery of Scorpiones as the sister group of Tetrapulmonata (Sharma et al. 2014a; Ballesteros and Sharma 2019). Taken together with the Paleozoic fossil record of aquatic scorpions, the derived placement of modern Scorpiones in analyses of molecular sequence data, genome architecture, or gene expression patterns precipitates a shift in previous understanding of the evolutionary history of Chelicerata. If modern scorpions are indeed derived, then they must represent either an independent terrestrialization event deeply nested within Chelicerata or the Paleozoic scorpions represent a secondary recolonization of aquatic habitats. The same may hold for Xiphosura, whose placement as the sister group of a monophyletic Arachnida is now highly suspect (Ballesteros and Sharma 2019; Ballesteros et al. 2019).

Phylogenies produced in recent paleontological works continue to reflect the notion of scorpions as a stepping stone between Xiphosura and the remaining "derived" arachnids (Legg et al. 2013; Garwood and Dunlop 2014; Lamsdell et al. 2015; Lamsdell 2016; Huang et al. 2018; Aria and Caron 2019; Bicknell et al. 2019), despite demonstrable evidence that morphological characters are, at best, poor arbiters of deep phylogenetic relationships, especially in groups that exemplify external morphological stasis (e.g., Bieler et al. 2014; Sharma et al. 2015c; Burbrink et al. 2019). In the specific case of scorpions, we previously showed that morphological characters traditionally employed in scorpion morphological phylogenies exhibit high levels of homoplasy or are simply uninformative at higher taxonomic ranks (Fig. S25 of Sharma et al. 2015c). Even at shallow taxonomic scales, we showed that parametric analyses of shape data for structures highly valued in scorpion systematics (e.g., the shape of the carapace and specific pedipalpal podomeres) exhibit measurable morphological convergence and/or broadly uninformative variation in the family Diplocentridae (Santibáñez-López et al. 2017). The only character systems in scorpions shown to exhibit phylogenetic signal at higher taxonomic levels are (1) an internal character pertaining to developmental mode, (2) an internal character pertaining to the distribution of the midgut glands (Sharma et al. 2015c), and (3) the architecture of the hemispermatophore (Monod et al. 2017). This trend should be uniquely troubling to systematic paleontologists, for whom such internal character systems are simply not observable in fossil scorpion taxa assigned to higher level groups in node dating approaches (e.g.,Howard et al. 2019).

To date, no efforts have been undertaken to incorporate the gene duplications of Arachnopulmonata as characters in total evidence matrices combining paleontological and neotological datasets (i.e., "molecular morphology"; but see Hox expression characters in data matrix of Legg et al. 2013), nor have any morphological data matrices ever recovered the monophyly of Arachnopulmonata. Future efforts to integrate such disparate datasets may elucidate a more comprehensive view of chelicerate evolutionary history, and herald a reconciliation between paleontological and neontological investigations. As first steps toward this goal, we advocate unbiased interpretations of morphological character systems in chelicerate systematics, free of preconceptions about terrestrialization, in tandem with sophisticated, modern approaches to collection of morphological data in underexplored character systems, as exemplified by the works of Klußmann-Fricke et al. (2014), Klußmann-Fricke and Wirkner (2016), and Lehmann and Melzer (2019). We further postulate that the incorporation of gene expression patterns and other rare genomic changes as characters in data matrices may guide future reconciliation of disputing datasets.

Acknowledgments We are indebted to Angelika Stollewerk and Matthias Pechmann for inviting this contribution. Russell Bicknell kindly provided a morphological data matrix for reanalysis. Melody Albright assisted with field collection of *C. sculpturatus*. Comments from Emily V.W. Setton, Guilherme Gainett, Jesús A. Ballesteros, and two anonymous reviewers improved a previous draft of the manuscript. Discussions with three colleagues in the paleontological community, Russell Garwood, Jason Dunlop, and Paul Selden, greatly refined some of the discussion on Paleozoic scorpions. James Lamsdell could not be reached for comment.

Funding information This work is based on material supported by National Science Foundation grant IOS-1552610 to PPS.

References

Aria C, Caron J-B (2019) A middle Cambrian arthropod with chelicerae and proto-book gills. Nature 465:215–214. https://doi.org/10.1038/s41586-019-1525-4

Ballesteros JA, Sharma PP (2019) A critical appraisal of the placement of Xiphosura (Chelicerata) with account of known sources of phylogenetic error

Ballesteros JA, Santibáñez-López CE, Kováč L, Gavish-Regev E, Sharma PP (2019) Ordered phylogenomic subsampling enables diagnosis of systematic errors in the placement of the enigmatic arachnid order Palpigradi. Proc R Soc Lond B. https://doi.org/10.1098/rspb.2019.2426

Barnett AA, Thomas RH (2013) The expression of limb gap genes in the mite Archegozetes longisetosusreveals differential patterning mechanisms in chelicerates. Evol Dev 15:280–292. https://doi.org/10.1111/ede.12038

Bergsten J (2005) A review of long-branch attraction. Cladistics 21:163–193. https://doi.org/10.1111/j.1096-0031.2005.00059.x

Bicknell RDC, Lustri L, Brougham T (2019) Revision of "*Bellinurus*" *carteri* (Chelicerata: Xiphosura) from the Late Devonian of Pennsylvania, USA. Comptes Rendus Palevol. https://doi.org/10.1016/j.crpv.2019.08.002

Bieler R, Mikkelsen PM, Collins TM, Glover EA, González VL, Graf DL, Harper EM, Healy J, Kawauchi GY, Sharma PP, Staubach S, Strong EE, Taylor JD, Temkin I, Zardus JD, Clark S, Guzmán A, McIntyre E, Sharp P, Giribet G (2014) Investigating the bivalve Tree of Life—an exemplar-based approach combining molecular and novel morphological characters. Invertebr Syst 28:32–115. https://doi.org/10.1071/IS13010

Briggs DEG, Siveter DJ, Siveter DJ, Sutton MD, Garwood RJ, Legg D (2012) Silurian horseshoe crab illuminates the evolution of arthropod limbs. Proc Natl Acad Sci U S A 109:15702–15705. https://doi.org/10.1073/pnas.1205875109

Burbrink FT, Grazziotin FG, Pyron RA, Cundall D, Donnellan S, Irish F, Keogh JS, Kraus F, Murphy RW, Noonan B, Raxworthy CJ, Ruane S, Lemmon AR, Lemmon EM, Zaher H (2019) Interrogating genomic-scale data for Squamata (lizards, snakes, and Amphisbaenians) shows no support for key traditional morphological relationships. Syst Biol. https://doi.org/10.1093/sysbio/syz062

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17(4):540–552

Dehal P, Boore JL (2005) Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol 3:e314. https://doi.org/10.1371/journal.pbio.0030314

Di Z, Edgecombe GD, Sharma PP (2018) Homeosis in a scorpion supports a telopodal origin of pectines and components of the book lungs. BMC Evol Biol 18:73. https://doi.org/10.1186/s12862-018-1188-z

Dunlop J (1998) The origins of tetrapulmonate book lungs and their significance for chelicerate phylogeny. Proceedings of the 17th European Colloquium of Arachnology 9–16

Dunlop JA (2010) Geological history and phylogeny of Chelicerata. Arthropod Structure and Development 39:124–142. https://doi.org/10.1016/j.asd.2010.01.003

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleis Acids Res 32(5):1792–1797. https://doi.org/10.1093/nar/gkh340

Fröbius AC, Funch P (2017) Rotiferan Hox genes give new insights into the evolution of metazoan bodyplans. Nat Commun 8:9. https://doi.org/10.1038/s41467-017-00020-w

Garwood RJ, Dunlop J (2014) Three-dimensional reconstruction and the phylogeny of extinct chelicerate orders. PeerJ 2:e641. https://doi.org/10.7717/peerj.641/supp-4

Garwood RJ, Dunlop JA, Selden PA, Spencer AR, Atwood RC, Vo NT, Drakopoulos M (2016) Almost a spider: a 305-million-year-old fossil arachnid and spider origins. Proc R Soc B Biol Sci 283:20160125–20160128. https://doi.org/10.1098/rspb.2016.0125

Goloboff PA, Catalano SA (2016) TNT version 1.5, including a full implementation of phylogenetic morphometrics. Cladistics 32:221–238. https://doi.org/10.1111/cla.12160

Hazkani-Covo E (2009) Mitochondrial insertions into primate nuclear genomes suggest the use of numts as a tool for phylogeny. Mol Biol Evol 26:2175–2179. https://doi.org/10.1093/molbev/msp131

Howard RJ, Edgecombe GD, Legg DA et al (2019) Exploring the evolution and terrestrialization of scorpions (Arachnida: Scorpiones) with rocks and clocks:1–16. https://doi.org/10.1007/s13127-019-00390-7

Huang D, Hormiga G, Cai C, Su Y, Yin Z, Xia F, Giribet G (2018) Origin of spiders and their spinning organs illuminated by mid-Cretaceous amber fossils. Nat Ecol Evol 2:623–627. https://doi.org/10.1038/s41559-018-0475-9

Janssen R, Feitosa NM, Damen WGM, Prpic N-M (2008) The T-box genes H15 and optomotor-blind in the spiders Cupiennius salei,

Tegenaria atrica and Achaearanea tepidariorum and the dorsoventral axis of arthropod appendages. Evol Dev 10:143–154. https://doi.org/10.1111/j.1525-142X.2008.00222.x

Kenny NJ, Chan KW, Nong W, Qu Z, Maeso I, Yip HY, Chan TF, Kwan HS, Holland PW, Chu KH, Hui JH (2016) Ancestral whole-genome duplication in the marine chelicerate horseshoe crabs. Heredity 116: 190–199. https://doi.org/10.1038/hdy.2015.89

Klußmann-Fricke BJ, Wirkner CS (2016) Comparative morphology of the hemolymph vascular system in Uropygi and Amblypygi (Arachnida): complex correspondences support Arachnopulmonata. J Morphol:1–20. https://doi.org/10.1002/jmor.20559

Klußmann-Fricke BJ, Pomrehn SW, Wirkner CS (2014) A wonderful network unraveled-detailed description of capillaries in the prosomal ganglion of scorpions. Front Zool 11:28. https://doi.org/10.1186/1742-9994-11-28

Lamsdell JC (2016) Horseshoe crab phylogeny and independent colonizations of fresh water: ecological invasion as a driver for morphological innovation. Palaeontology 59:181–194. https://doi.org/10.1111/pala.12220

Lamsdell JC, Briggs DEG, Liu HP, Witzke BJ, McKay RM (2015) A new Ordovician arthropod from the Winneshiek Lagerstätte of Iowa (USA) reveals the ground plan of eurypterids and chasmataspidids. Sci Nat- Heidelberg 102:63. https://doi.org/10.1007/s00114-015-1312-5

Lartillot N, Philippe H (2004) A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol Biol Evol 21(6):1095–1109. https://doi.org/10.1093/molbev/msh112

Lartillot N, Brinkmann H, Philippe H (2007) Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. BMC Evol Biol 7:S4. https://doi.org/10.1186/1471-2148-7-S1-S4

Legg DA, Sutton MD, Edgecombe GD (2013) Arthropod fossil data increase congruence of morphological and molecular phylogenies. Nat Commun 4:2485. https://doi.org/10.1038/ncomms3485

Lehmann T, Melzer RR (2019) The visual system of Thelyphonida (whip scorpions): support for Arachnopulmonata. Arthropod Struct Dev 51:23–31. https://doi.org/10.1016/j.asd.2019.06.002

Leite DJ, Ninova M, Hilbrant M, Arif S, Griffiths-Jones S, Ronshaugen M, McGregor A (2016) Pervasive microRNA duplication in Chelicerates: insights from the embryonic microRNA repertoire of the spider Parasteatoda tepidariorum. Genome Biol Evol 8:2133–2144. https://doi.org/10.1093/gbe/evw143

Leite DJ, Baudouin-Gonzalez L, Iwasaki-Yokozawa S et al (2018) Homeobox gene duplication and divergence in arachnids. Mol Biol Evol 35:2240–2253. https://doi.org/10.1093/molbev/msy125

Liao YY, Xu PW, Kwan KY, Ma ZY, Fang HY, Xu JY, Wang PL, Yang SY, Xie SB, Xu SQ, Qian D, Li WF, Bai LR, Zhou DJ, Zhang YQ, Lei J, Liu K, Li F, Li J, Zhu P, Wang YJ, Wu HP, Xu YH, Huang H, Zhang C, Liu JX, Han JF (2019) Draft genomic and transcriptome resources for marine chelicerate Tachypleus tridentatus. Nat Publ Group 6:1–10. https://doi.org/10.1038/sdata.2019.29

Lozano-Fernandez J, Tanner AR, Giacomelli M, Carton R, Vinther J, Edgecombe GD, Pisani D (2019) Increasing species sampling in chelicerate genomic-scale datasets provides support for monophyly of Acari and Arachnida. Nat Commun 10:1–8. https://doi.org/10.1038/s41467-019-10244-7

Meusemann K, Reumont Von BM, Simon S et al (2010) A Phylogenomic approach to resolve the arthropod Tree of Life. Mol Biol Evol 27: 2451–2464. https://doi.org/10.1093/molbev/msq130

Monod L, Cauwet L, González-Santillan E, 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. Front Zool 14:51. https://doi.org/10.1186/s12983-017-0231-z

Pechmann M, Prpic N-M (2009) Appendage patterning in the South American bird spider Acanthoscurria geniculata (Araneae: Mygalomorphae). Dev Genes Evol 219:189–198. https://doi.org/10.1007/s00427-009-0279-7

Poschmann M, Dunlop JA, Kamenz C, Scholtz G (2008) The Lower Devonian scorpion *Waeringoscorpio* and the respiratory nature of its filamentous structures, with the description of a new species from the Westerwald area, Germany. Paläontol Z 82(4):418–436. https://doi.org/10.1007/BF03184431

Regier JC, Shultz JW, Zwick A, Hussey A, Ball B, Wetzer R, Martin JW, Cunningham CW (2010) Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. Nature 463:1079–1083. https://doi.org/10.1038/nature08742

Rokas A, Carroll SB (2006) Bushes in the Tree of Life. PLoS Biol 4: 1899–1904. https://doi.org/10.1371/journal.pbio.0040352

Rokas A, Holland P (2000) Rare genomic changes as a tool for phylogenetics. Trends Ecol Evol 15:454–459

Ronco M, Uda T, Mito T, Minelli A, Noji S, Klingler M (2008) Antenna and all gnathal appendages are similarly transformed by homothorax knock-down in the cricket Gryllus bimaculatus. Dev Biol 313:80–92. https://doi.org/10.1016/j.ydbio.2007.09.059

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Sucha MA, Huelsenbeck JP (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539–542. https://doi.org/10.1093/sysbio/sys029

Salichos L, Rokas A (2014) Inferring ancient divergences requires genes with strong phylogenetic signals. Nature 497:327–331. https://doi.org/10.1038/nature12130

Santibáñez-López CE, Krievel R, Sharma PP (2017) *eadem figura manet*: measuring morphological convergence in diplocentrid scorpions (Arachnida: Scorpiones: Diplocentridae) under a multilocus phylogenetic framework. Invertebr Syst 31:233–248. https://doi.org/10.1071/IS16078

Santibáñez-López CE, González-Santillán E, Monod L, Sharma PP (2019) 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. https://doi.org/10.1016/j.ympev.2019.02.021

Scholtz G, Kamenz C (2006) The book lungs of Scorpiones and Tetrapulmonata (Chelicerata, Arachnida): evidence for homology and a single terrestrialisation event of a common arachnid ancestor. 109:2–13. https://doi.org/10.1016/j.zool.2005.06.003

Schwager EE, Schoppmeier M, Pechmann M, Damen WG (2007) Duplicated Hox genes in the spider Cupiennius salei. Front Zool 4:10. https://doi.org/10.1186/1742-9994-4-10

Schwager EE, Sharma PP, Clarke T, Leite DJ, Wierschin T, Pechmann M, Akiyama-Oda Y, Esposito L, Bechsgaard J, Bilde T, Buffry AD, Chao H, Dinh H, Doddapaneni H, Dugan S, Eibner C, Extavour CG, Funch P, Garb J, Gonzalez LB, Gonzalez VL, Griffiths-Jones S, Han Y, Hayashi C, Hilbrant M, Hughes DST, Janssen R, Lee SL, Maeso I, Murali SC, Muzny DM, Nunes da Fonseca R, Paese CLB, Qu J, Ronshaugen M, Schomburg C, Schönauer A, Stollewerk A, Torres-Oliva M, Turetzek N, Vanthournout B, Werren JH, Wolff C, Worley KC, Bucher G, Gibbs RA, Coddington J, Oda H, Stanke M, Ayoub NA, Prpic NM, Flot JF, Posnien N, Richards S, McGregor A (2017) The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. BMC Biol 15:62. https://doi.org/10.1186/s12915-017-0399-x

Selden PA, Jeram AJ (1989) Palaeophysiology of terrestrialisation in the Chelicerata. Trans R Soc Edinb Earth Sci 80:303–310

Setton EVW, Sharma PP (2018) Cooption of an appendage-patterning gene cassette in the head segmentation of arachnids. Proc Natl Acad Sci 115:E3491–E3500. https://doi.org/10.1073/pnas.1720193115

Sharma PP (2017) Chelicerates and the conquest of land: a view of arachnid origins through an Evo-Devo spyglass. Integr Comp Biol 57: 510–522. https://doi.org/10.1093/icb/icx078

Sharma PP, Schwager EE, Extavour CG, Giribet G (2012a) Hox gene expression in the harvestman Phalangium opilio reveals divergent patterning of the chelicerate opisthosoma. Evol Dev 14:450–463. https://doi.org/10.1111/j.1525-142X.2012.00565.x

Sharma PP, Schwager EE, Extavour CG, Giribet G (2012b) Evolution of the chelicera: a dachshund domain is retained in the deutocerebral appendage of Opiliones (Arthropoda, Chelicerata). Evol Dev 14: 522–533. https://doi.org/10.1111/ede.12005

Sharma PP, Schwager EE, Giribet G, Jockusch EL, Extavour CG (2013) Distal-less and dachshund pattern both plesiomorphic and apomorphic structures in chelicerates: RNA interference in the harvestman Phalangium opilio (Opiliones). Evol Dev 15:228–242. https://doi.org/10.1111/ede.12029

Sharma PP, Kaluziak ST, Perez-Porro AR et al (2014a) Phylogenomic interrogation of Arachnida reveals systemic conflicts in phylogenetic signal. Mol Biol Evol 31:2963–2984. https://doi.org/10.1093/molbev/msu235

Sharma PP, Schwager EE, Extavour CG, Wheeler WC (2014b) Hox gene duplications correlate with posterior heteronomy in scorpions. Proc R Soc B Biol Sci 281:20140661–20140661. https://doi.org/10.1016/j.cub.2009.06.061

Sharma PP, Gupta T, Schwager EE, Wheeler WC, Extavour CG (2014c) Subdivision of arthropod cap-n-collar expression domains is restricted to Mandibulata. EvoDevo 5:3. https://doi.org/10.1186/2041-9139-5-3

Sharma PP, Santiago MA, González-Santillán E, Monod L, Wheeler WC (2015a) Evidence of duplicated Hox genes in the most recent common ancestor of extant scorpions. Evol Dev 17:347–355. https://doi.org/10.1111/ede.12166

Sharma PP, Tarazona OA, Lopez DH, Schwager EE, Cohn MJ, Wheeler WC, Extavour CG (2015b) A conserved genetic mechanism specifies deutocerebral appendage identity in insects and arachnids. Proc R Soc B Biol Sci 282:20150698–20150698. https://doi.org/10.1098/rspb.2015.0698

Sharma PP, Fernandez R, Esposito LA et al (2015c) Phylogenomic resolution of scorpions reveals multilevel discordance with morphological phylogenetic signal. Proc R Soc B Biol Sci 282: 20142953–20142953. https://doi.org/10.1093/molbev/mss208

Sharma PP, Baker CM, Cosgrove JG, Johnson JE, Oberski JT, Raven RJ, Harvey MS, Boyer SL, 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. https://doi.org/10.1016/j.ympev.2018.01.003

Shen X-X, Hittinger CT, Rokas A (2017) Contentious relationships in phylogenomic studies can be driven by a handful of genes. Nat Ecol Evol 1:1–10. https://doi.org/10.1038/s41559-017-0126

Shultz JW (1990) Evolutionary morphology and phylogeny of Arachnida. Cladistics 6:1–38. https://doi.org/10.1111/j.1096-0031.1990.tb00523.x

Shultz J (2007) A phylogenetic analysis of the arachnid orders based on morphological characters

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9):1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Turetzek N, Pechmann M, Schomburg C et al (2015) Neofunctionalization of a duplicate dachshund gene underlies the evolution of a novel leg segment in arachnids. Mol Biol Evol 33: 109–121. https://doi.org/10.1093/molbev/msv200

Turetzek N, Khadjeh S, Schomburg C, Prpic N-M (2017) Rapid diversification of homothorax expression patterns after gene duplication in spiders. BMC Evol Biol 17:168. https://doi.org/10.1186/s12862-017-1013-0

Waddington J, Rudkin DM, Dunlop JA (2015) A new mid-Silurian aquatic scorpion—one step closer to land? Biol Lett 11:20140815–20140815. https://doi.org/10.1016/j.palaeo.2008.05.008

Wang B, Dunlop JA, Selden PA, Garwood RJ, Shear WA, Müller P, Lei X (2018) Cretaceous arachnid *Chimerarachne yingi* gen. Et sp. nov. illuminates spider origins. Nat Ecol Evol 2:614–622. https://doi.org/10.1038/s41559-017-0449-3