Taxonomic Sampling and Rare Genomic Changes Overcome Long-Branch Attraction in the Phylogenetic Placement of Pseudoscorpions

Andrew Z. Ontano,[1] Guilherme Gainett [iD],[1] Shlomi Aharon [iD],[2] Jesús A. Ballesteros,[1] Ligia R. Benavides [iD],[3] Kevin F. Corbett,[1] Efrat Gavish-Regev [iD],[2] Mark S. Harvey [iD],[4] Scott Monsma,[5] Carlos E. Santibáñez-López [iD],[6] Emily V.W. Setton,[1] Jakob T. Zehms,[1] Jeanne A. Zeh [iD],[7] David W. Zeh [iD],[7] and Prashant P. Sharma [iD]*,[1]

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

[2]National Natural History Collections, The Hebrew University of Jerusalem, Jerusalem, Israel

[3]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA

[4]Collections & Research, Western Australian Museum, Welshpool, WA, Australia

[5]Lucigen Corporation, Middleton, WU, USA

[6]Department of Biology, Eastern Connecticut State University, Willimantic, CT, USA

[7]Department of Biology and Program in Ecology, Evolution & Conservation Biology, University of Nevada, Reno, NV, USA

*Corresponding author: E-mail: prashant.sharma@wisc.edu.
Associate editor: Tal Pupko

Abstract

Long-branch attraction is a systematic artifact that results in erroneous groupings of fast-evolving taxa. The combination of short, deep internodes in tandem with long-branch attraction artifacts has produced empirically intractable parts of the Tree of Life. One such group is the arthropod subphylum Chelicerata, whose backbone phylogeny has remained unstable despite improvements in phylogenetic methods and genome-scale data sets. Pseudoscorpion placement is particularly variable across data sets and analytical frameworks, with this group either clustering with other long-branch orders or with Arachnopulmonata (scorpions and tetrapulmonates). To surmount long-branch attraction, we investigated the effect of taxonomic sampling via sequential deletion of basally branching pseudoscorpion superfamilies, as well as varying gene occupancy thresholds in supermatrices. We show that concatenated supermatrices and coalescent-based summary species tree approaches support a sister group relationship of pseudoscorpions and scorpions, when more of the basally branching taxa are sampled. Matrix completeness had demonstrably less influence on tree topology. As an external arbiter of phylogenetic placement, we leveraged the recent discovery of an ancient genome duplication in the common ancestor of Arachnopulmonata as a litmus test for competing hypotheses of pseudoscorpion relationships. We generated a high-quality developmental transcriptome and the first genome for pseudoscorpions to assess the incidence of arachnopulmonate-specific duplications (e.g., homeobox genes and miRNAs). Our results support the inclusion of pseudoscorpions in Arachnopulmonata (new definition), as the sister group of scorpions. Panscorpiones (new name) is proposed for the clade uniting Scorpiones and Pseudoscorpiones.

Key words: arachnids, ohnologs, supermatrix, species tree reconciliation, microRNA.

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 the 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 and (2) accelerated evolution of multiple ingroup lineages, exacerbating long-branch attraction artifacts (Bergsten 2005; Rokas and Carroll 2006; King and Rokas 2017). The combination of these characteristics is difficult to overcome even with genome-scale data sets, due to homoplasy accrued over millions of years of evolutionary history, conflicting evolutionary signals in data partitions, systematic bias, and the lack of external arbiters to evaluate appropriateness of substitution and rate heterogeneity models. Within animals, examples of such problematic nodes include the base of Metazoa, Bilateria, the superclades

Open Access

Lophotrochozoa and Ecdysozoa, and internal relationships of many diverse phyla (Borner et al. 2014; Kocot et al. 2016; Feuda et al. 2017; Simion et al. 2017; Laumer et al. 2019; Marlétaz et al. 2019).

The basal phylogeny of the arthropod subphylum Chelicerata remains particularly recalcitrant to resolution despite the application of genome-scale phylogenomic data sets (Sharma, Kaluziak, et al. 2014; Ballesteros and Sharma 2019; Ballesteros et al. 2019; Lozano-Fernández et al. 2019). Initial diversification of this group and the crown age of many orders dates to the early Paleozoic (Lozano-Fernández et al. 2019). Within chelicerates, at least three orders exhibit the characteristics of long-branch taxa (Acariformes, Parasitiformes, and Pseudoscorpiones), with Solifugae and Palpigradi also prone to unstable placement, as inferred from taxon deletion experiments and assessments of topological stability (Sharma, Kaluziak, et al. 2014; Ballesteros and Sharma 2019; Ballesteros et al. 2019). Moreover, extinction has asymmetrically affected different branches in the chelicerate tree, resulting in both relictual orders such as horseshoe crabs, as well as several extinct orders. As a result, basic questions about the evolutionary history of Chelicerata remain controversial, namely, the monophyly of Arachnida (the terrestrial chelicerates; Ballesteros and Sharma 2019; Ballesteros et al. 2019; Lozano-Fernández et al. 2019). Even in data sets that support arachnid monophyly, relationships between chelicerate orders are highly unstable from one data set to the next, with the exception of the basal split between Pycnogonida (sea spiders) and the remaining chelicerates, Tetrapulmonata (a group of arachnid orders that bear four book lungs; Pepato et al. 2010; Regier et al. 2010; Sharma, Kaluziak, et al. 2014; Ballesteros and Sharma 2019; Ballesteros et al. 2019; Lozano-Fernández et al. 2019; Howard et al. 2020), and the robust recovery of Arachnopulmonata (Scorpiones + Tetrapulmonata; Sharma, Kaluziak, et al. 2014; Ballesteros and Sharma 2019; Ballesteros et al. 2019; Lozano-Fernández et al. 2019; Howard et al. 2020).

In addition to these phylogenomic analyses, Arachnopulmonata is also supported by analyses of genome architecture, as both spiders and scorpions share partial or whole-genome duplication (WGD). This inference is evidenced by retention of duplicated copies of numerous developmental patterning genes and microRNAs, to the exclusion of groups like Opiliones (harvestmen) and Acari (Schwager et al. 2007; Sharma, Schwager, et al. 2014; Leite et al. 2016; Sharma, Santiago, et al. 2015; Schwager et al. 2017; Leite et al. 2018). Moreover, exploratory analyses of gene trees and embryonic gene expression patterns in spiders, scorpions, and harvestmen have shown that the duplicated copies of arachnopulmonate leg-patterning genes also retain expression domains that reflect the evolutionary history of shared WGD (Gainett and Sharma 2020; Nolan et al. 2020). The systemic duplication of developmental patterning genes and gene expression patterns together constitute a highly complex character that unites Arachnopulmonata (Leite et al. 2016; Gainett and Sharma 2020; Gainett et al. 2020; Nolan et al. 2020), but the putative incidence of this phenomenon has not been assessed in many chelicerate orders, most of which lack genomic and functional genetic resources (Garb et al. 2018).

One potential solution to overcome long-branch attraction includes the expansion of taxonomic sampling, which serves to "break" long branches and improve the estimation of parameters of substitution models. Although recent efforts have targeted improving taxonomic representation of the acarine orders in phylogenetic data sets (Acariformes and Parasitiformes; Arribas et al. 2019; Charrier et al. 2019), only recently has phylogenomic sampling of Pseudoscorpiones successfully sampled all major extant lineages (Benavides et al. 2019). Intriguingly, in phylogenetic studies that have broadly sampled pseudoscorpions and scorpions, pseudoscorpions are frequently recovered as either sister group to Arachnopulmonata (Sharma, Fernández, et al. 2015) or as sister group to scorpions (Sharma et al. 2018; Benavides et al. 2019), although these works lacked complete representation of all chelicerate orders (fig. 1). In works assessing chelicerate phylogeny broadly, pseudoscorpion placement has proven unstable or unsupported, either clustering with the Acari or with arachnopulmonates (Sharma, Kaluziak, et al. 2014; Arribas et al. 2019; Ballesteros and Sharma 2019; Ballesteros et al. 2019; Lozano-Fernández et al. 2019) (fig. 1). In these works, taxonomic representation of Pseudoscorpiones has nevertheless been limited, often to a subset of derived lineages.

To evaluate these competing hypotheses for pseudoscorpion placement in the chelicerate tree of life, we established a phylogenomic data set of Chelicerata broadly sampling all major lineages of Pseudoscorpiones. We assessed the effect of an incomplete taxonomic sampling by sequentially pruning basally branching lineages of pseudoscorpions and gauged the effect on the inferred tree topology using different analytical approaches to phylogenetic reconstruction. Furthermore, we reasoned that if Pseudoscorpiones is nested within Arachnopulmonata, then they should share the systemic duplications of developmental patterning genes previously demonstrated for scorpions and spiders (Leite et al. 2016, 2018). The advantage of WGDs as phylogenetic characters is that even if an affected lineage exhibits broad scale loss of the resulting ohnologs (the duplicate gene copies resulting from WGD) over time, the signature of this event can be discerned using patterns of synteny across genomes as well via the ensuing gene trees (i.e., a retained single-copy homolog of an originally ohnologous pair is still expected to cluster with their orthologs of other taxa that share the genome duplication). WGD events shared across an array of taxa can be further discerned from lineage-specific duplications, using gene tree topologies (i.e., ohnologs clustering across shared WGD events vs. in-paralogs clustering by lineage). Such dynamics have been especially well studied at the base of the vertebrates (Dehal and Boore 2005; Putnam et al. 2008; Simakov et al. 2020).

Here we show that expanded taxonomic sampling of pseudoscorpions, systemic homeobox gene duplications, tree topologies of benchmarked ohnologs of developmental patterning genes, and duplications of miRNAs, all support the hypothesis that pseudoscorpions are nested within Arachnopulmonata as the sister group of scorpions.
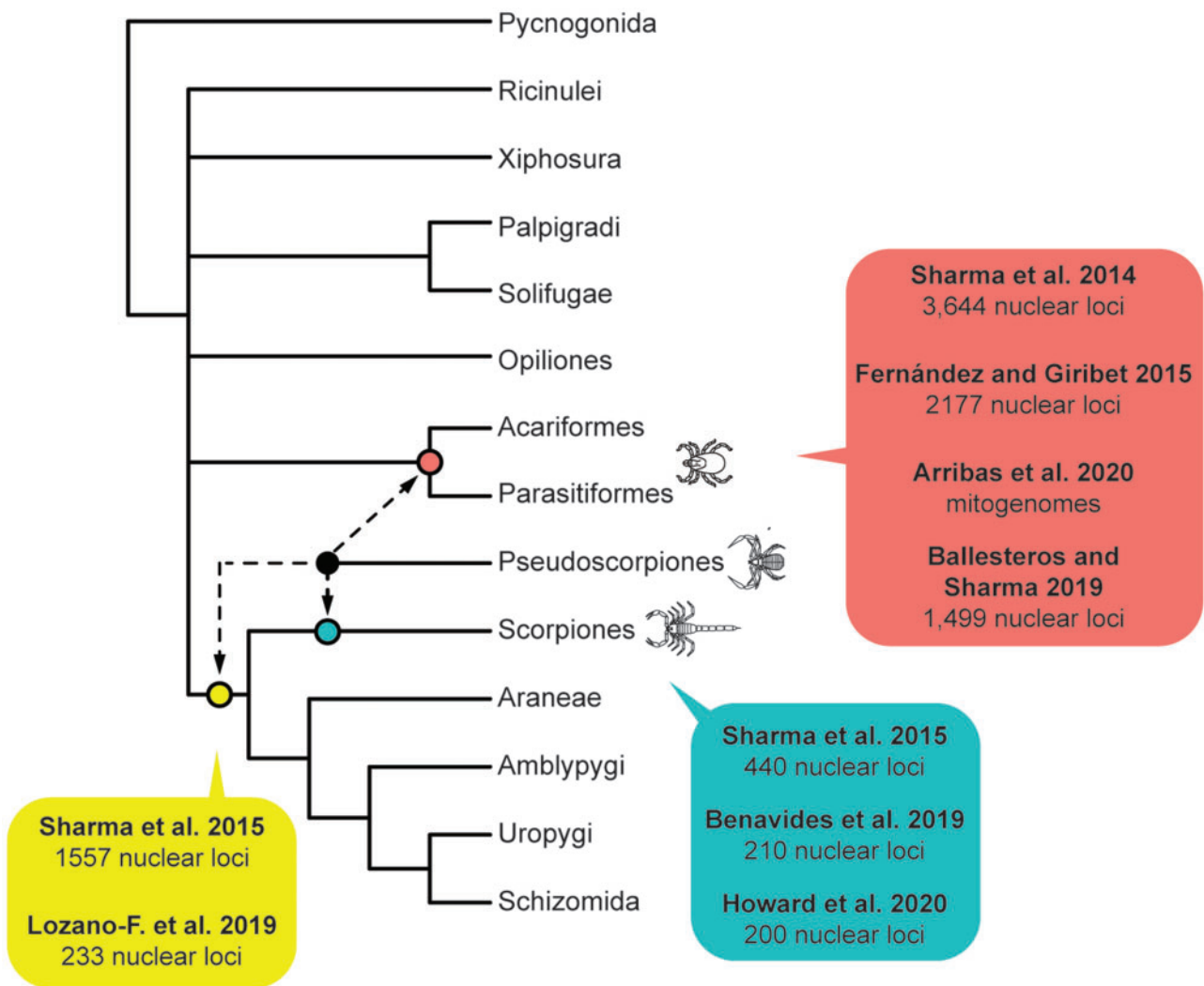
Fig. 1. Summary tree topology of Chelicerata showing relationships of orders. Phylogeny based on Ballesteros et al. (2019). Dotted lines for pseudoscorpiones show alternative placements of this order in selected historical phylogenetic analyses.

Results

Phylogenomics with Partitioned Models

To assess matrix completeness and denser taxonomic sampling as explanatory processes for the unstable phylogenetic placement of pseudoscorpions, we assembled a data set of 132 Panarthropoda, including 40 pseudoscorpion libraries previously generated by Benavides et al. (2019), which represent all pseudoscorpion superfamilies (fig. 2a; supplementary table S1, Supplementary Material online). Orthologs analyzed in this study consisted of the Benchmarked Universal Single Copy Orthologs of Arthropoda (BUSCO-Ar) derived from OrthoDB v.9.1 (Simão et al. 2015; Zdobnov et al. 2017; Waterhouse et al. 2018). Each library was analyzed with the OrthoDB pipeline to identify available homologs of 1,066 arthropod-specific BUSCO genes. Duplicated BUSCOs were discarded to retain only validated, single-copy loci. We constructed six matrices ranging in gene occupancy thresholds of 80% (248 BUSCO loci) to 55% (1002 BUSCO loci); we denote these as G_1 to G_6, in order of increasing matrix length (fig. 2b). For each of these six matrices, we additionally pruned basally

branches lineages within Pseudoscorpiones, with reference to Cheliferoidea. This superfamily was selected as the distal-most taxon, because it was represented by the most exemplars of any pseudoscorpion superfamily (12 transcriptomes), ensuring that the order would be well represented across all supermatrices, despite the pruning of other lineages (supplementary table S2, Supplementary Material online). For each matrix, we performed maximum likelihood (ML) searches and assessed phylogenetic placement of pseudoscorpions as sister group to scorpions, sister group to Arachnopulmonata (sensu Sharma, Kaluziak, et al. 2014) or sister group to one or both of the long-branch acarine orders (Acariformes and Parasitiformes). Six branches were sequentially pruned; we denote these data sets as T_{-1} to T_{-6}, in order of increasing branch pruning (fig. 2a and b).

Matrices retaining all superfamilies of pseudoscorpions (i.e., unpruned data sets) consistently recovered the relationship Pseudoscorpiones + Scorpiones, regardless of matrix completeness. ML tree topologies of pruned taxon subsets T_{-1} and T_{-2} similarly recovered the relationship
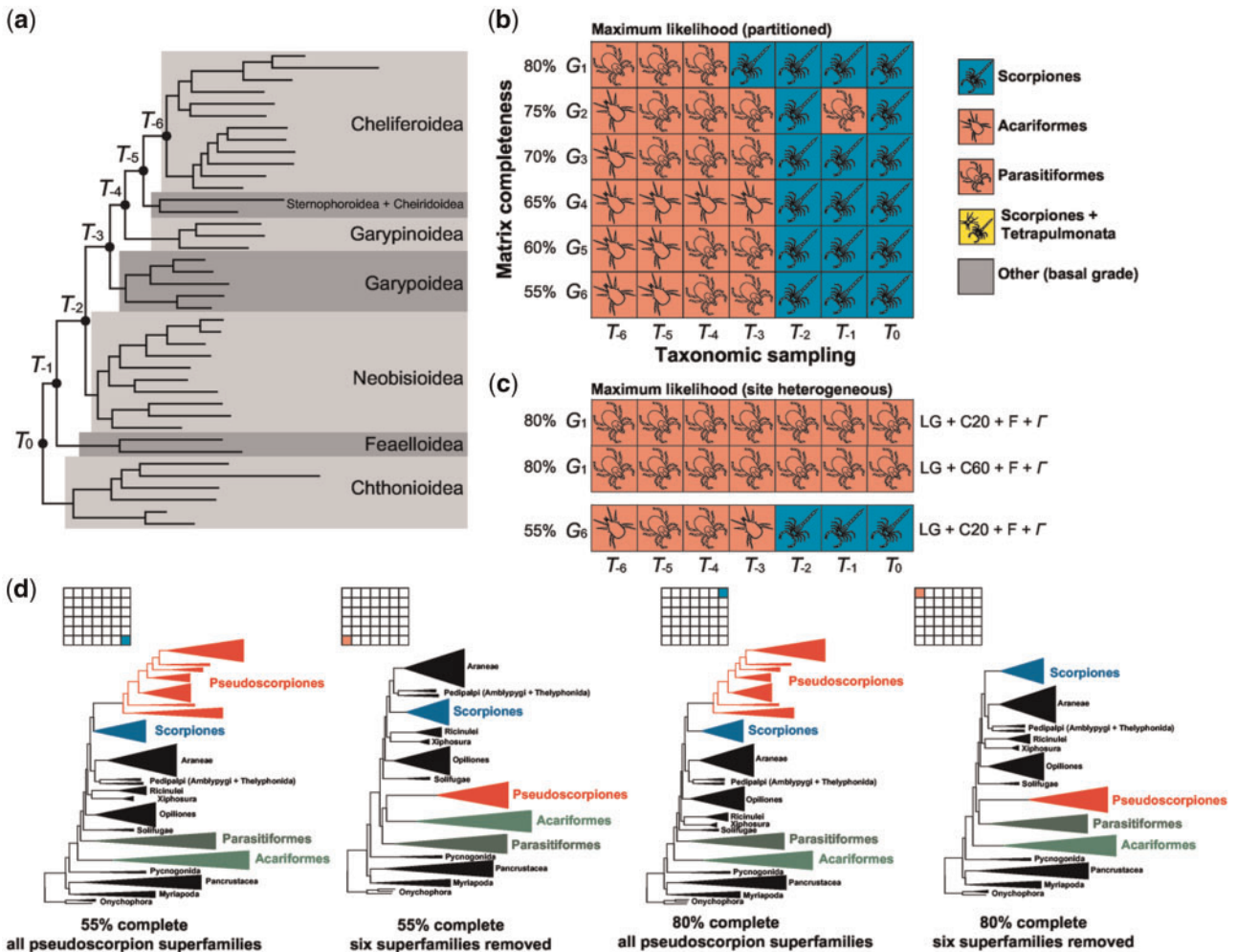
FIG. 2. Depth of taxonomic sampling is more influential than matrix completeness in supermatrix analyses of pseudoscorpion placement. (*a*) Internal phylogeny of Pseudoscorpiones showing major taxonomic groups. Notations on nodes indicate taxon subsets obtained by sequential pruning of branches. (*b*) Sensitivity plot of 42 phylogenomic matrices assembled by varying gene occupancy (*y* axis) and taxonomic sampling (*x* axis), with partitioned model-fitting. Colors of squares correspond to the sister group of Pseudoscorpiones obtained in each maximum likelihood analysis. (*c*) Analysis of largest (1002 loci) and densest (248 loci) matrices under variations of the PMSF model. (*d*) Selected tree topologies showing the dynamics of pseudoscorpion instability as a function of taxonomic sampling in partitioned-model analyses.

Pseudoscorpiones + Scorpiones, excepting matrix $G_2 \cdot T_{-1}$, which recovered an unsupported relationship of Pseudoscorpiones + Parasitiformes (ultrafast bootstrap resampling frequency [BS] = 43%).

Inversely, matrices exhibiting pruning of the three most basally branching pseudoscorpion lineages (Chthonioidea, Feaelloidea, and Neobisioidea) recovered ML tree topologies wherein pseudoscorpions were sister group to either Parasitiformes or Acariformes, regardless of matrix completeness (T_{-3} matrices). Further pruning of basally branching pseudoscorpions generally also incurred this tree topology (T_{-4} to T_{-6} matrices), with the exception of matrix $G_1 \cdot T_{-3}$ (fig. 2*b*). No matrix recovered the relationship of pseudoscorpions as sister group to Arachnopulmonata (*sensu* Sharma, Kaluziak, et al. 2014).

Relationships among other chelicerate taxa largely reflected the outcomes of previous works (Ballesteros and Sharma 2019; Ballesteros et al. 2019) and are not discussed in detail here (fig. 2*d*). Notably, we never recovered the monophyly of Acari or Arachnida.

Phylogenomics with Site Heterogeneous Models

Partitioned model ML analyses have sometimes been criticized as less accurate than site heterogeneous models, although these inferences have often been grounded in assumptions of true relationships based on traditional phylogenetic hypotheses (e.g., Wang et al. 2019). Simulations have previously shown that CAT+GTR and partitioned ML analyses are comparably accurate, with both of these outperforming CAT-F81 (sometimes referred to as CAT-Poisson) with respect to topological accuracy (Whelan and Halanych 2016). However, CAT+GTR models are notoriously difficult to implement in a Bayesian framework, due to excessive computational times for real data sets (i.e., >100 taxa, >500 genes), and numerous published analyses using PhyloBayes-mpi have exhibited failure to converge (defined as ESS >200; *maxdiff* <0.10), especially for chelicerate phylogeny (Sharma, Kaluziak, et al. 2014; Ballesteros and Sharma 2019; Ballesteros et al. 2019; Lozano-Fernández et al. 2019; Howard et al. 2020). As a workaround, we assessed the performance of the posterior mean site frequency (PMSF) model LG + C20+F+Γ, a mixture model alternative to the CAT

implementation. This model was implemented for the G_1 and G_6 family of matrices, which constitute the densest and the largest matrices we analyzed, respectively (248 and 1002 genes, respectively). For the G_6 matrices, the pattern of tree topologies recovered reflected the same outcome as the partitioned model analyses, with T_{-3} to T_{-6} matrices recovering Pseudoscorpiones as clustering with one of the acarine orders, and T_0 to T_{-2} matrices recovered Pseudoscorpiones + Scorpiones (fig. 2c).

Notably, all the G_1 matrices analyzed using the LG + C20 + F + Γ recovered Pseudoscorpiones + Parasitiformes with support (BS = 94–100%). To assess the impact of a more parameter-rich site heterogenous model on phylogenomic inference, we repeated the analyses of the G_1 matrices under the LG + C60 + F + Γ model. Despite the use of a model with additional rate categories, these analyses also uniformly recovered the relationship Pseudoscorpiones + Parasitiformes with high support (BS = 95–100%).

Analyses using site heterogeneous models never recovered the monophyly of Arachnida or Acari.

Nodal Support Dynamics

Ultrafast bootstrap resampling frequencies were used to estimate support for competing hypotheses for the phylogenetic placement of Pseudoscorpiones, across the 42 concatenated matrices analyzed with partitioned model-fitting (fig. 3a). Across all levels of matrix completeness, support for Pseudoscorpiones + Scorpiones was negligible (<10%) for T_{-3} to T_{-6} matrices, but increased dramatically upon including Neobisioidea (T_{-2} matrices). Increase in nodal support for Pseudoscorpiones + Scorpiones was not monotonic, as sampling of Feaelloidea and Chthonioidea resulted in some variability in bootstrap frequency (fig. 3a). The nodal support trajectories were identical for the hypotheses Pseudoscorpiones + Scorpiones and Pseudoscorpiones + Arachnopulmonata. This result reflects in part the nestedness of the two hypotheses (i.e., Scorpiones is nested within Arachnopulmonata).

By contrast, support for Pseudoscorpiones as the sister group of either Parasitiformes or Acariformes showed the opposite trend, with better representation of basally branching pseudoscorpion groups resulting in lower nodal support for pseudoscorpions clustering with either of these groups. For taxon subsets with the least representation of basally branching pseudoscorpions (T_{-4} to T_{-6} matrices), the most complete matrices recovered high support values for Pseudoscorpiones + Parasitiformes, whereas matrices with intermediate gene occupancy thresholds (60–70%, or G_3 to G_5 matrices) recovered high support values for Pseudoscorpiones + Acariformes.

Gene Trees, ΔGLS, and Species Tree Reconstruction

Approaches to inferring species tree using gene trees have been shown to be powerful predictors of phylogenetic accuracy, but these methods are predicated on the accuracy of the underlying gene tree set. To assess whether improving taxonomic sampling of a long-branch taxon also affects phylogenetic signal at the level of gene trees, we calculated gene-wise log-likelihood scores (ΔGLS) on gene trees corresponding to

each of the 42 matrices. ΔGLS assesses the likelihood of each gene given two competing tree topologies, across all genes in a data set (Shen et al. 2017). We generated ΔGLS distributions for the two competing hypotheses of pseudoscorpion placement (clustering with scorpions vs. clustering with either acarine order).

We observed minimal effects of taxon pruning in the largest matrices (G_5 and G_6), and no consistent trends in the distribution of genes favoring either competing hypothesis, across the ΔGLS distributions of 42 analyses (fig. 3b). Magnitudes of log likelihood favoring either hypothesis were also not consistently affected (supplementary fig. S1, Supplementary Material online). These results suggest that increasing taxonomic sampling of a long-branch lineage does not greatly alter the distribution of phylogenetic signal at the level of individual gene trees.

Although gene and site concordance factors were trialed (Minh et al. 2020), these were invariably low for all competing placements of pseudoscorpions, as well as interordinal relationships, reflecting well-known conflicting signal in basal chelicerate phylogeny (Sharma, Kaluziak, et al. 2014; Ballesteros and Sharma 2019; Ballesteros et al. 2019).

To assess whether the intransigence of ΔGLS distributions to taxonomic sampling has downstream effects on methods of phylogenetic reconstruction, especially those that use the multispecies coalescent model, we reconstructed species trees from gene trees using ASTRAL v.5.14.2 (Zhang et al. 2018). We discovered no clear difference between the performance of ASTRAL versus concatenation-based approaches, with respect to the tree topology recovered as a function of the number of basal branches pruned (fig. 4). Generally, T_0 to T_{-2} matrices recovered the relationship Pseudoscorpiones + Scorpiones, whereas T_{-3} to T_{-6} matrices again recovered Pseudoscorpiones as clustering with the acarine orders. The exceptions were matrices $G_2 \cdot T_{-1}$ and $G_4 \cdot T_0$, which recovered pseudoscorpions as the sister group of Parasitiformes or in a grade at the base of Chelicerata, respectively.

ASTRAL analyses never recovered the monophyly of Arachnida or Acari.

Filtering by Evolutionary Rate

It has been previously shown that support for some chelicerate relationships is strongly affected by evolutionary rate. As examples, support for Pseudoscorpiones + Scorpiones and Arachnida was initially shown to be restricted to slow-evolving genes by Sharma, Kaluziak, et al. (2014), a result partly corroborated by Howard et al. (2020), but reproduced with variable success in other analyses (Ballesteros and Sharma 2019; Ballesteros et al. 2019). To dissect the interaction of taxon sampling and evolutionary rate, we partitioned the G_2 and G_3 families of matrices into tertiles based on mean pairwise sequence identity (MPSI) of loci. The G_2 and G_3 matrices were selected for a tradeoff between high number of loci per tertile and low quantity of missing data.

For G_2, phylogenomic analyses based on maximum likelihood with partitioned model-fitting recovered Pseudoscorpiones + Scorpiones only for the slow-evolving tertile of T_0 to T_{-2} matrices, and the intermediate rate tertile of
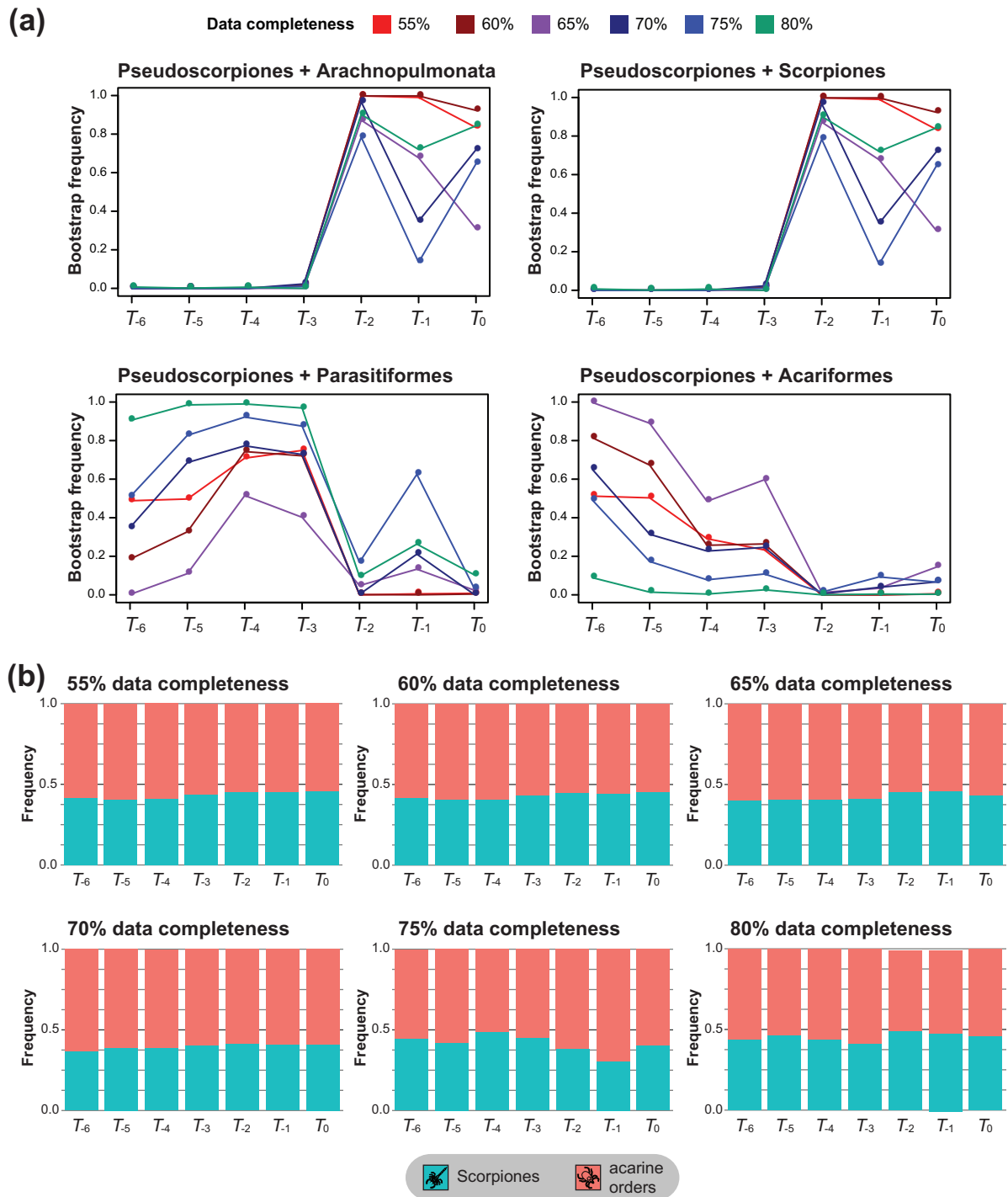
Fig. 3. Depth of taxonomic sampling affects supermatrix nodal support, but not per-locus support. (*a*) Nodal support frequency for competing hypotheses of pseudoscorpion placement as a function of taxonomic sampling and matrix completeness. (*b*) Proportion of loci favoring Pseudoscorpiones + Scorpiones versus Pseudoscorpiones + either acarine order under a ΔGLS framework, as a function of taxonomic sampling and matrix completeness.

the T_{-2} matrix (fig. 5a). Pseudoscorpiones were recovered as the sister group of Arachnopulmonata (*sensu* Sharma, Kaluziak, et al. 2014) by the intermediate rate tertiles of the T_0 and T_{-1} matrices. All other analyses of G_2 matrices recovered pseudoscorpions as sister group to an acarine order or in an unresolved position. For G_3, Pseudoscorpiones + Scorpiones was recovered for only the slow-evolving tertile

of T_0 and T_{-2} matrices, with all other analyses recovering pseudoscorpions as the sister group to an acarine order or in an unresolved position.

ASTRAL analyses never recovered Pseudoscorpiones + Scorpiones; pseudoscorpions were recovered as the sister group of Arachnopulmonata (*sensu* Sharma, Kaluziak, et al. 2014) for a subset of the slow-evolving tertiles for both G_1 and
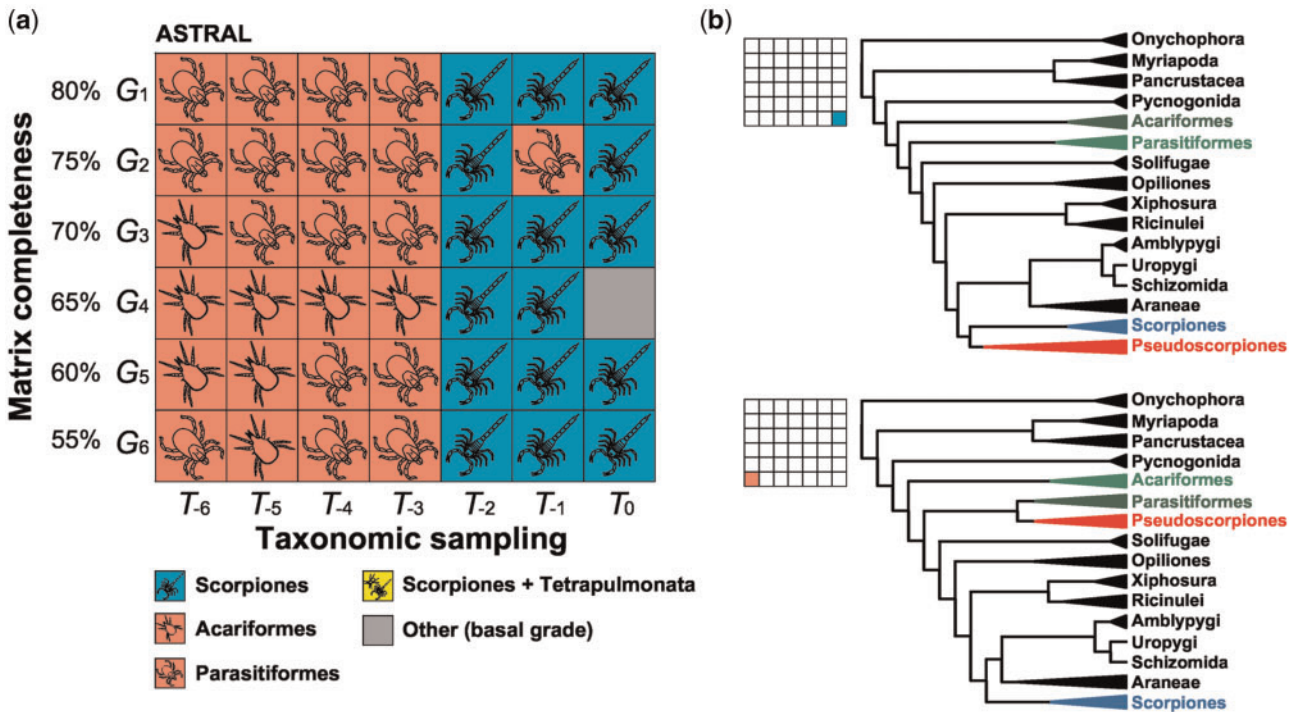
FIG. 4. Depth of taxonomic sampling is more influential than matrix completeness in ASTRAL analyses of pseudoscorpion placement. (a) Sensitivity plot of 42 phylogenomic matrices assembled by varying gene occupancy (y axis) and taxonomic sampling (x axis). (b) Selected tree topologies showing the dynamics of pseudoscorpion instability as a function of taxonomic sampling.
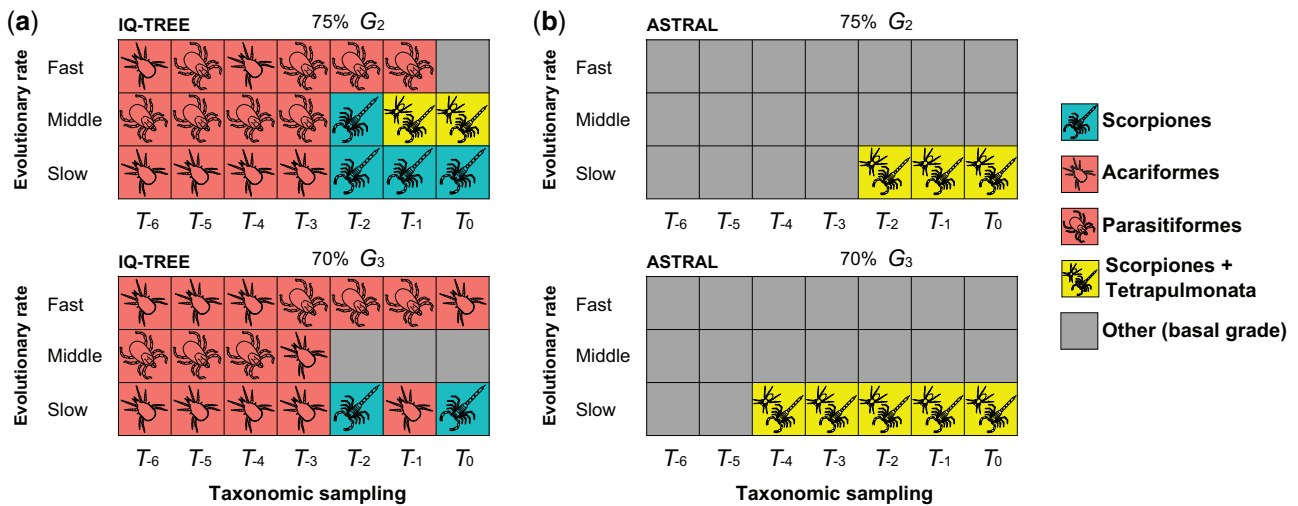
FIG. 5. Slow-evolving genes do not consistently recover Pseudoscorpiones + Scorpiones in the absence of dense taxon sampling. (a) Sensitivity plot of 21 phylogenomic matrices assembled by binning genes into tertiles of evolutionary rate (y axis) and taxonomic sampling (x axis), with partitioned model-fitting. Colors of squares correspond to the sister group of Pseudoscorpiones obtained in each maximum likelihood analysis. Above: Analyses based on the G_2 family of matrices (75% complete). Below: Analyses based on the G_3 family of matrices (70% complete). (b) Analyses of the corresponding data sets in (a) using ASTRAL.

G_2 data sets (fig. 5b). Taken together, these analyses suggest that slow-evolving genes alone cannot resolve long-branch taxa consistently in the absence of dense taxonomic sampling.

Analyses of data sets filtered for evolutionary rate never recovered the monophyly of Arachnida or Acari.

Duplications of Homeobox Genes

As an external arbiter of the two competing hypotheses of pseudoscorpion relationships, we generated a developmental transcriptome of the West Australian chernetid *Conicochernes crassus*. Homeobox gene surveys of developmental transcriptomes and/or genomes have previously been

shown to be faithful readouts of WGD in Chelicerata. WGDs are inferred to have occurred in the common ancestor of Arachnopulmonata (one event) and independently in the Xiphosura (2-fold or 3-fold WGD); groups like mites, ticks, and harvestmen do not exhibit these shared duplications (Sharma, Schwager, et al. 2014; Sharma, Santiago, et al. 2015; Kenny et al. 2016; Leite et al. 2016; Schwager et al. 2017; Leite et al. 2018; Shingate et al. 2020). A previous comprehensive analysis of homeobox genes by Leite et al. (2018) showed that the retention of duplicates is systemic in two arachnopulmonate lineages (spiders and scorpions), an inference subsequently supported by the first whip spider developmental transcriptomes (Gainett and Sharma 2020; Gainett et al. 2020) and by embryonic gene expression data (Gainett and Sharma 2020; Nolan et al. 2020). However, this survey of homeobox duplications omitted key groups, such as Xiphosura and Parasitiformes (Leite et al. 2018). Curiously, Leite et al. (2018) had indeed sampled two pseudoscorpion species, but recovered few homeobox genes for these taxa, likely owing to the sampling of postembryonic stages rather than embryos; in scorpions, developmental transcriptomes have been shown to recover far more duplicated homeobox genes than adult transcriptomes (Sharma, Schwager, et al. 2014; Sharma, Santiago, et al. 2015).

We therefore assembled a data set of 26 Panarthropoda, sampling genomes or developmental transcriptomes of all three major lineages of Arachnopulmonata sensu Sharma, Kaluziak, et al. (2014) (i.e., spiders, scorpions, and Pedipalpi [Amblypygi + Uropygi + Schizomida]), as well as mites, ticks, harvestmen, horseshoe crabs, and sea spiders. This data set leveraged recent developmental genetic resources generated by us for several non-model chelicerate groups, such as mygalomorph spiders, whip spiders, harvestmen, and sea spiders (Sharma et al. 2012; Setton et al. 2019; Ballesteros et al. 2020; Gainett and Sharma 2020; Gainett et al. 2020). We included in our analysis two adult transcriptomes of pseudoscorpions previously analyzed by Leite et al. (2018), which had been shown to harbor few homeobox genes and exhibited short contigs for many homeobox homologs. Outgroup data sets consisted of an onychophoran embryonic transcriptome and genomes of Mandibulata.

In contrast to the previous analyses of adult pseudoscorpion transcriptomes (Hesperochernes sp. and Neobisium carcinoides in Leite et al. 2018), our analysis of the first pseudoscorpion developmental transcriptome recovered homologs of 56 homeobox genes in C. crassus (fig. 6). Of these, 26 exhibited duplications in at least one of the three pseudoscorpion exemplars that were also found in at least one scorpion or one tetrapulmonate, with clear evidence of paralogy (i.e., overlapping peptide sequences exceeding 100 amino acids in length that exhibited multiple substitutions between duplicate pairs).

All ten Hox genes ancestral to Panarthropoda are known to be duplicated in scorpions and spiders, with embryonic expression patterns reflecting the shared duplication (Schwager et al. 2007; Sharma, Schwager, et al. 2014; Schwager et al. 2017). Recent work has shown that the common ancestor of Amblypygi (whip spiders) likely also

exhibited two copies of each Hox gene (Gainett and Sharma 2020). However, the previous homeobox survey of adult pseudoscorpion transcriptomes had only recovered five of the ten Hox genes, with none of these duplicated (Leite et al. 2018). By contrast, we discovered eight of the ten Hox homologs in the developmental transcriptome of C. crassus (all but Hox3 and Sex combs reduced). Of these eight, five exhibited duplications: labial, Deformed, fushi tarazu, Antennapedia, and abdominal-A.

Other well-characterized embryonic patterning genes among the homeobox family that were duplicated in both pseudoscorpions and arachnopulmonates included the Six gene family (e.g., sine oculis; Optix; Gainett et al. 2020), central nervous system patterning genes (e.g., empty spiracles; Pax3/7), appendage patterning genes (e.g., homothorax; extradenticle; Nolan et al. 2020), and segmentation cascade genes (e.g., engrailed; orthodenticle). Enumeration of the homeobox homologs across the 26 species is provided in supplementary table S3, Supplementary Material online.

By comparison to pseudoscorpions, we did not detect systemic duplications of homeobox genes (i.e., suggestive of shared WGD with arachnopulmonates) in Acariformes, Parasitiformes, Opiliones, or Pycnogonida. As a key example, among these groups of arachnids, duplicates of only two Hox genes were detected in the genome of the mite Tetranychus urticae (with these being tandem duplicates on a single Hox cluster; figure 4a of Grbić et al. 2011). By contrast, tetrapulmonate exemplars new to this analysis (the mygalomorph Aphonopelma hentzi; the three Amblypygi species) exhibited the expected trend of retention of homeobox duplicates. Taken together, this survey of homeobox genes suggests that pseudoscorpions were included in the shared WGD at the base of Arachnopulmonata.

Gene Tree Analysis of Benchmarked Embryonic Patterning Genes

Whereas embryonic expression data are abundant for spiders, and principally for the model system Parasteatoda tepidariorum, they are comparatively few for non-spider chelicerate groups (e.g., Blackburn et al. 2008; Jager et al. 2006; Grbic et al. 2007; Sharma et al. 2012; Sharma, Schwager, et al. 2014; Sharma, Tarazona, et al. 2015; Barnett and Thomas 2013; Gainett and Sharma 2020). In the recent comparative work, it was shown that four appendage patterning genes known to be duplicated in spiders and scorpions exhibited shared expression patterns that reflected the history of the species tree (i.e., ohnologs of P. tepidariorum and the scorpion C. sculpturatus exhibited shared, unique expression patterns, by comparison to the expression domains of their paralogs or of single-copy homologs of outgroups like harvestmen, mites, and mandibulates) (Nolan et al. 2020). These four genes (dachshund, homothorax, extradenticle, and optomotor blind) constitute benchmarked cases of arachnopulmonate ohnologs that have been validated via gene expression surveys, with additional and recent corroboration of this pattern in two of the four genes in the whip spider P. marginemaculatus (Gainett and Sharma 2020).
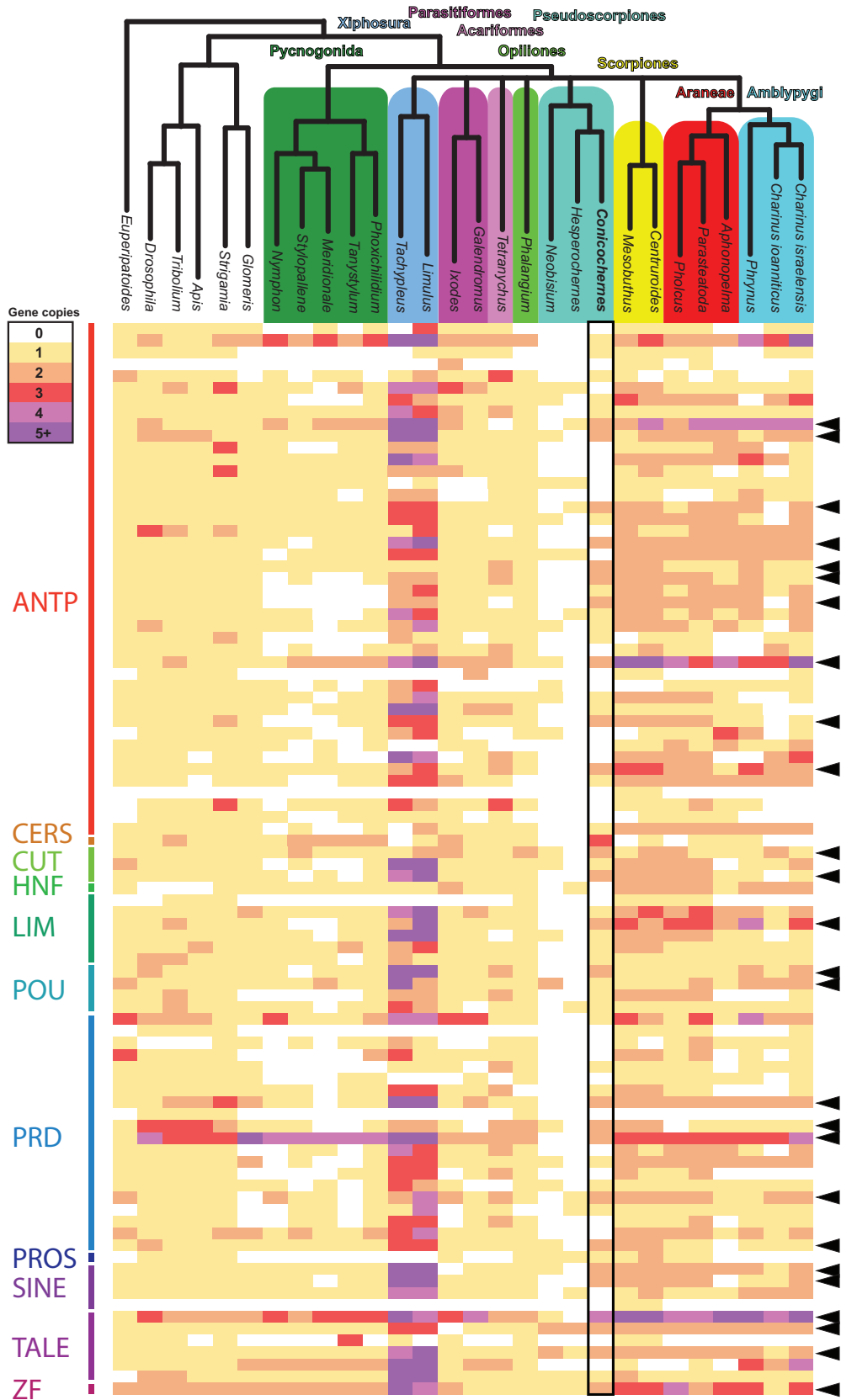
FIG. 6. Comparison of homeobox repertoires for 26 panarthropods supports retention of duplications in pseudoscorpions that are shared with arachnopulmonates. Rows correspond to individual homeobox genes. Colors correspond to numbers of paralogs. Black arrows to the right indicate duplications in at least one pseudoscorpion exemplar that is also shared by at least one arachnopulmonate.
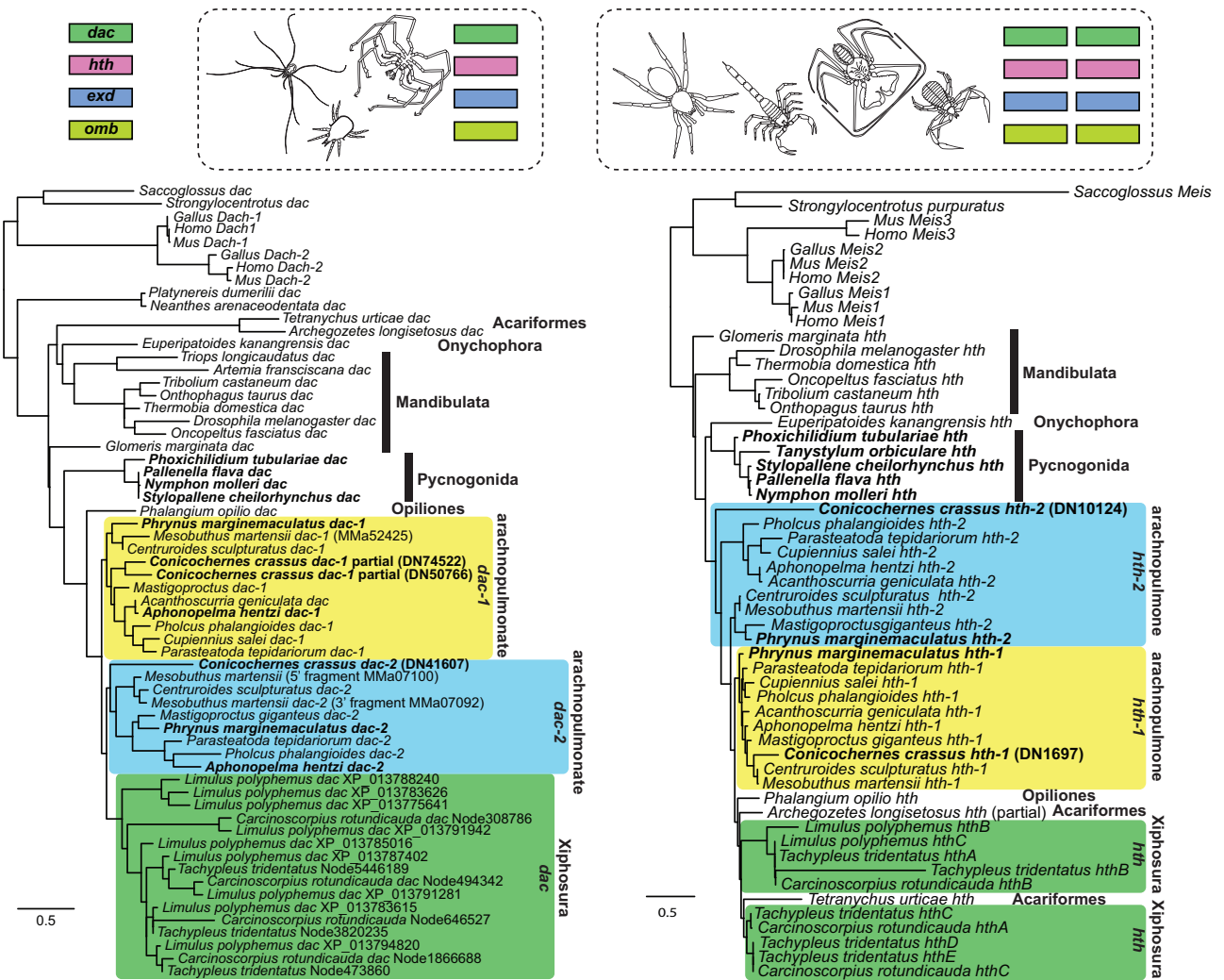
FIG. 7. Pseudoscorpions possess two copies of four developmental patterning genes known to exhibit paralogs with arachnopulmonate-specific spatiotemporal expression domains. Above: Single copy orthologs of *dac*, *hth*, *exd*, and *omb* were recovered from genomic resources for sea spiders, harvestmen, and Acariformes, whereas two copies of each gene were recovered for scorpions, tetrapulmonates, and pseudoscorpions. Below left: Maximum likelihood gene tree topology of the medial leg gap gene *dac*. Below right: Maximum likelihood gene tree topology of the proximal leg gap gene *hth*. Note the clustering of pseudoscorpion copies within arachnopulmonate clusters. *dac-1* of *C. crassus* was recovered as two non-overlapping fragments.

We therefore investigated whether duplicates of these four genes also occurred in the developmental transcriptome of *C. crassus*. To the surveys previously generated by Nolan et al. (2020), we searched for and added homologs of these genes from developmental transcriptomes of the pseudoscorpion, the whip spider species *Phrynus marginemaculatus* (Gainett and Sharma 2020), five sea spider species (Setton and Sharma 2018; Ballesteros et al. 2020), and the tarantula *A. hentzi* (Setton et al. 2019). We discovered two copies of all four genes in the developmental transcriptome of the pseudoscorpion, except for *dachshund*, wherein three putative homologs were discovered (fig. 7). However, two of these pseudoscorpion *dachshund* fragments were non-overlapping, suggesting that only two copies of *dachshund* are present in this transcriptome (comparable to the case of *Mesobuthus martensii*; Nolan et al. 2020). Similarly, we discovered two copies of these genes in the new arachnopulmonate data sets (whip spiders and the tarantula). By contrast, only one copy of these four

genes was discovered in the sea spiders, as with mites, ticks, and harvestmen.

Gene tree analysis of these four genes had previously shown sufficient signal to resolve monophyletic clusters of arachnopulmonate *dac* and *hth* ohnologs (Nolan et al. 2020). Upon reconstructing these two gene trees after adding the pseudoscorpion, the whip spiders, the tarantula, and the sea spiders, we observed each pseudoscorpion paralog clustering with an arachnopulmonate ohnolog, rather than with the single copy orthologs of acarine taxa. For *dac*, the arachnopulmonate (including pseudoscorpion) clusters were recovered as monophyletic; as previously reported, the horseshoe crab duplications are unrelated to those of Arachnopulmonata (Nolan et al. 2020; Shingate et al. 2020). For *hth*, one arachnopulmonate (including pseudoscorpion) ohnolog (*hth-1*, the ohnolog reflecting the ancestral expression pattern; Nolan et al. 2020) was recovered as monophyletic, whereas the other (*hth-2*, the copy with the derived
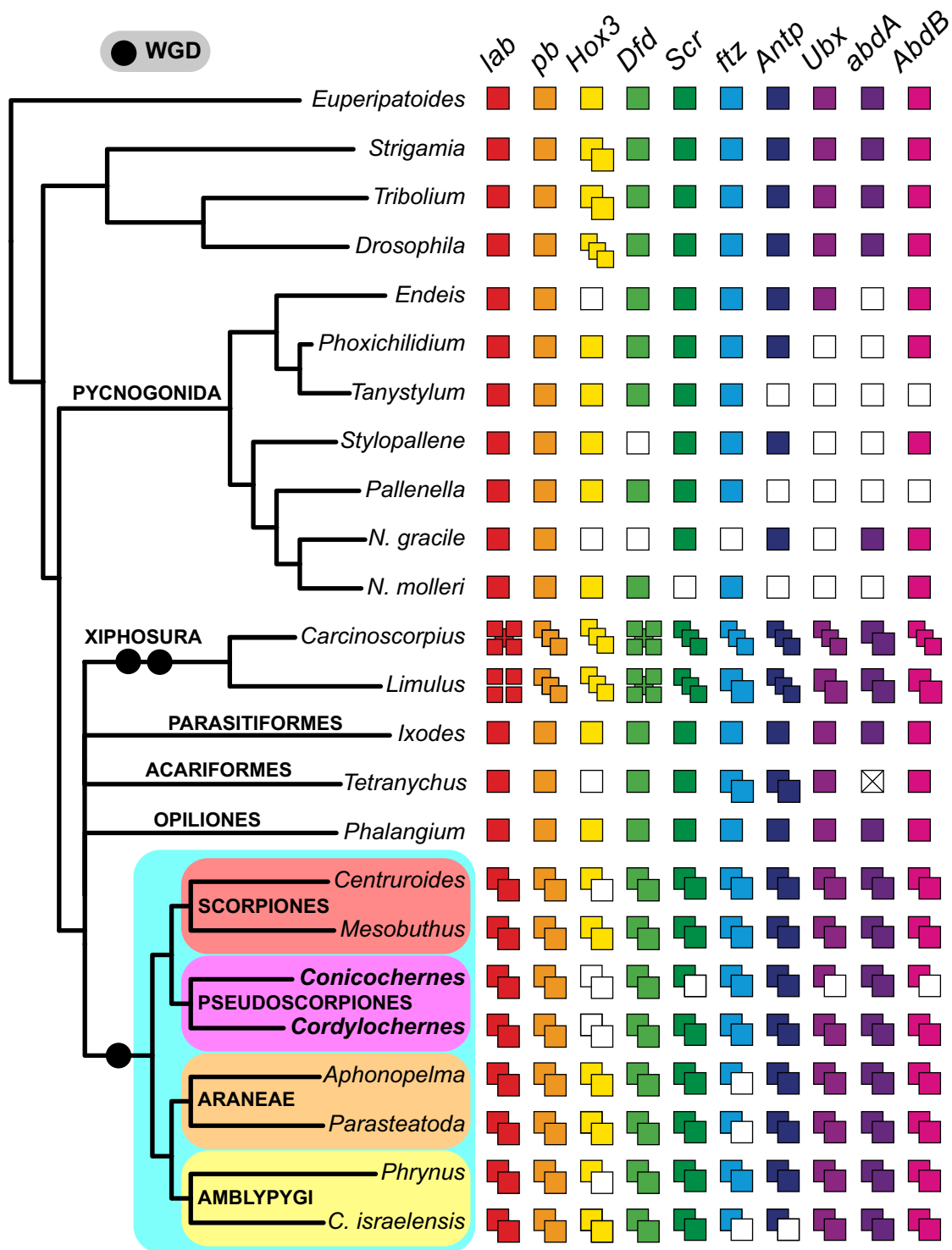
Fig. 8. Hox gene complement in the pseudoscorpion genome substantiates evidence of shared WGD with other Arachnopulmonata. Columns and colored squares correspond to each Hox gene. Unfilled squares correspond to absences, not losses. Cross through *abdA* in *T. urticae* indicates loss of this Hox gene in the mite genome. Note independent 2-fold WGD events in Xiphosura (Kenny et al. 2016; Shingate et al. 2020).

expression pattern; Nolan et al. 2020) was resolved as a grade (fig. 7). Gene trees of *extradenticle* and *optomotor blind* showed insufficient phylogenetic signal for testing phylogenetic placement, as previously reported (Nolan et al. 2020). These results corroborate the inference that systemic duplication unites pseudoscorpions with Arachnopulmonata.

Hox Genes and MicroRNA Duplications in the Pseudoscorpion Genome

Embryonic transcriptomes have proven useful for the inference of gene duplications, but are inferentially limited in that absence of gene copies cannot be distinguished as the result of gene loss versus absence of expression in the sequenced tissue and ontogenetic stage. As a separate validation of

systemic duplication in Pseudoscorpiones, we sequenced and analyzed the draft genome of the species *Cordylochernes scorpioides* for Hox gene clusters and miRNAs. Due to the fragmentation of the assembly, we were unable to recover more than one Hox gene per scaffold. Nevertheless, we discovered 18 Hox genes in the *C. scorpioides* genome, corresponding to two ohnologs of all Hox genes except for *Hox3* (fig. 8). Together with the homeobox duplications in *C. crassus*, these results are consistent with a shared genome duplication uniting arachnopulmonates and pseudoscorpions.

MicroRNAs (miRNAs) have been leveraged as rare genomic changes across the metazoan tree of life, with their effectiveness as phylogenetic markers being closely tied to the quality of genomic resources used for miRNA surveys (Tarver et al. 2013; Thomson et al. 2014; Tarver et al. 2018). In Chelicerata, Leite et al. (2016) previously surveyed miRNAs in the genomes of four spiders, a scorpion, a horseshoe crab, five Parasitiformes, and one Acariformes, as well as several outgroup taxa. This survey revealed lineage-specific duplications in *Limulus polyphemus* consistent with 2-fold WGD in Xiphosura; duplicated clusters of miRNAs in the spider *P. tepidariorum*, as well as tandem duplications; and a subset of duplicated miRNAs that were shared across spiders and scorpions.

To elucidate if pseudoscorpions exhibit miRNA duplications shared by arachnopulmonates, we expanded the survey of Leite et al. (2016) and searched for miRNAs in the draft genome of the pseudoscorpion, *C. scorpioides* and the genome of the scorpion, *Mesobuthus martensii*. Twenty-six conserved miRNA families were identified in the *C. scorpioides* genome, and another 35 in *M. martensii*. Among them, families iab-4, mir-71, and mir-276 had two or more ortholog copies in Arachnopulmonata, Pseudoscorpiones and Xiphosura (fig. 9). Similarly, we found two members of the families bantam and mir-1 in Scorpiones, Pseudoscorpiones, two spiders, and Xiphosura.

Two miRNAs, mir-190 and pte-bantam, were found duplicated only in Scorpiones and *C. scorpioides* (with inferred independent duplications in Xiphosura, fig. 9). Our survey did not recover the presence of miRNA sequences from the families mir-210, mir-275, mir-315, mir-981, mir-277, and mir-11960 (previously reported in genomes of spiders and scorpions). We cannot rule out that these absences are attributable to the incompleteness of the pseudoscorpion genome assembly.

We found no miRNAs unique to Arachnida, nor patterns of duplication consistent with arachnid monophyly.

Taken together, these surveys of miRNA duplication revealed four miRNA duplications supporting the inclusion of pseudoscorpions within arachnopulmonates, and two further duplications supporting the sister relationship of pseudoscorpions and scorpions.

Discussion

Consilience of Phylogenetic Data Classes in the Placement of Pseudoscorpions

Chelicerate higher-level phylogeny is plagued by topological uncertainty, with a subset of orders exhibiting long-branch attraction artifacts, as elucidated by taxon deletion experiments (Ballesteros and Sharma 2019; Ballesteros et al. 2019). Barring the monophyly of Euchelicerata (Xiphosura and arachnids), Arachnopulmonata (previously defined as Scorpiones + Tetrapulmonata), and relationships within Tetrapulmonata, ordinal relationships in the chelicerate tree of life are highly unstable across phylogenomic data sets. Here, we leveraged the previous discovery of a WGD subtending the common ancestor of spiders and scorpions to assess competing hypotheses for the placement of pseudoscorpions (Sharma, Schwager, et al. 2014; Schwager et al. 2017). Taxon-rich analyses of supermatrices as well as reconciliation of gene trees consistently recovered pseudoscorpions as the sister group of scorpions, the hypothesis supported by genome and miRNA duplication. Our taxon deletion experiments reveal that the sampling of basally branching lineages in the pseudoscorpion tree of life is key to overcoming long-branch attraction artifacts that draw pseudoscorpions together with the acarine orders.

Our results are also consistent with the variance of tree topologies in the previous chelicerate phylogenetics. Studies that have omitted basally branching pseudoscorpion families, or insufficiently sampled outgroup lineages, recovered Pseudoscorpiones as sister group to, or nested within, Acari (e.g., Sharma, Kaluziak, et al. 2014; Arribas et al. 2019). By contrast, phylogenomic works that sampled basal splits within Pseudoscorpiones have recovered support for their placement within Arachnopulmonata (e.g., Benavides et al. 2019; Howard et al. 2020). Our analyses further demonstrate that taxonomic sampling outweighs matrix completeness and analytical approach (supermatrix vs. gene tree reconciliation approaches) in achieving phylogenetic accuracy when long-branch attraction is incident.

To date, no morphological data matrix has ever recovered the monophyly of Arachnopulmonata (with or without pseudoscorpions), with both older and recent morphological cladistic studies continuing to recover the archaic grouping of Lipoctena (scorpions as the sister group to the remaining arachnid orders; Legg et al. 2013; Lamsdell 2016; Aria and Caron 2019; Bicknell et al. 2019; reviewed by Nolan et al. 2020). Shultz (1990, 2007) presented the first compelling cladistic analyses demonstrating that scorpions are derived within the arachnid tree, a result reflected in another body of recent paleontological investigations (e.g., Garwood and Dunlop 2014; Huang et al. 2018; Wang et al. 2018). In such works, pseudoscorpions have typically been recovered as the sister group of Solifugae (as the clade Haplocnemata), an another order exhibiting topological instability (Ballesteros et al. 2019). Nevertheless, a sister group relationship of scorpions and pseudoscorpions has previously been tenuously supported by some morphological analyses, namely, the cladistic analysis of Garwood and Dunlop (2014). Subsequent expansion and reuse of this matrix also recovered this relationship (Huang et al. 2018; Wang et al. 2018). However, the recovery of the clade Pseudoscorpiones + Scorpiones as a sister group of Opiliones in those studies is refuted by phylogenomic analyses, developmental gene expression, and genomic architecture (Sharma et al. 2014; Ballesteros et al. 2019;
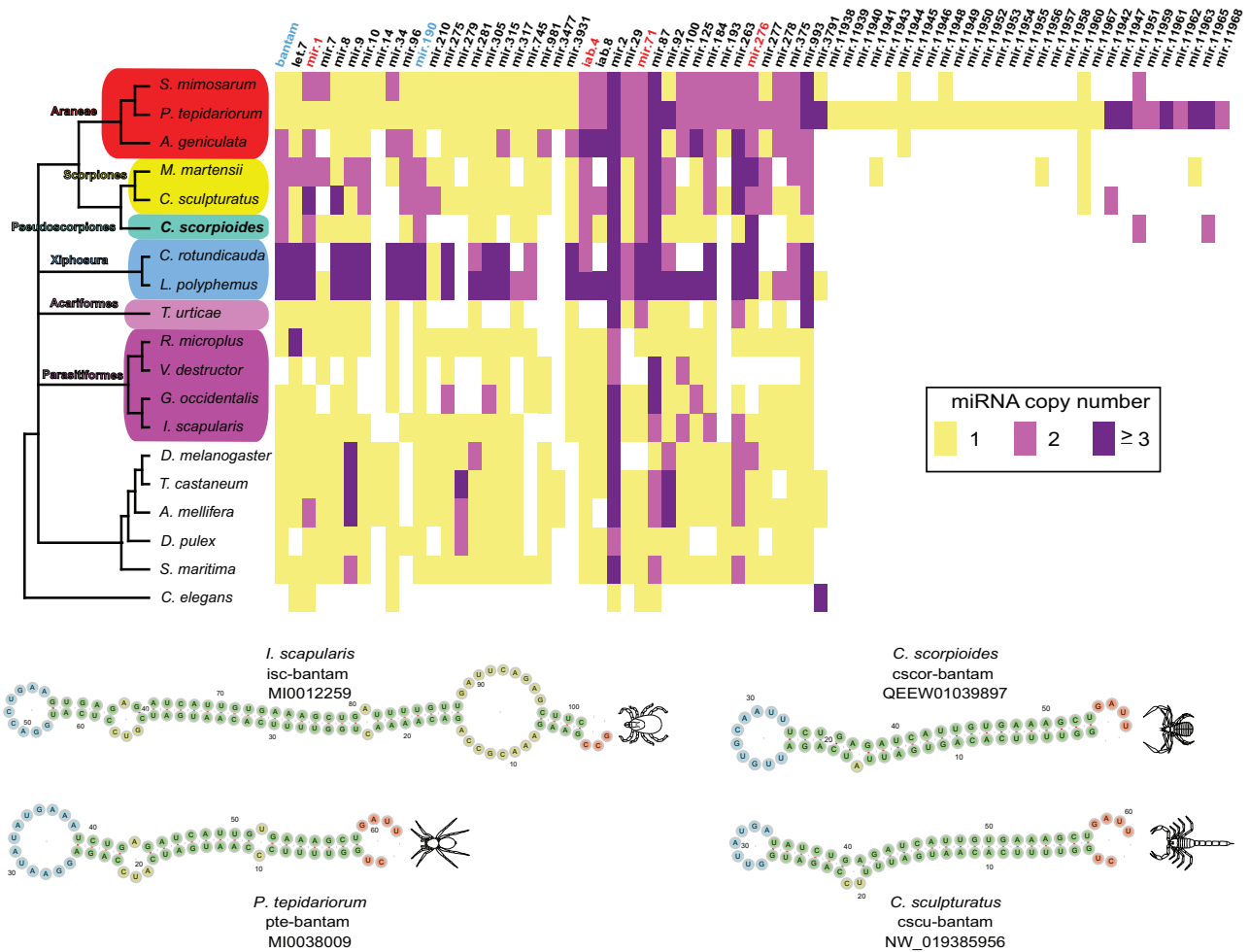
FIG. 9. Comparison of miRNA family copy number in *C. scorpioides* and other ecdysozoans supports retention of duplications in pseudoscorpions shared with scorpions and spiders. Columns correspond to individual miRNA families. Colors correspond to numbers of paralogs. miRNA families in red text indicate duplications in scorpions, pseudoscorpions and at least one spider. miRNA families in blue text are duplicated only in scorpions and pseudoscorpions.

Lozano-Fernández et al. 2019; Nolan et al. 2020). We therefore observe only partial concordance between our analyses and inferences based on morphological matrices.

By contrast to morphology, we identified clear and systemic evidence for a shared WGD in the first developmental transcriptome and genome of two pseudoscorpion exemplars, which is concordant with the hypothesis that pseudoscorpions are derived arachnopulmonates. Surveys of homeobox gene duplication, gene tree topologies of benchmarked arachnopulmonate-specific ohnologs with a known spatiotemporal subdivision of embryonic expression domains, and patterns of miRNA duplication all support the inclusion of Pseudoscorpiones within arachnopulmonates, with further evidence from two miRNA families for the clade Pseudoscorpiones + Scorpiones, a clade we term Panscorpiones (new name). Henceforth, we redefine Arachnopulmonata to include Pseudoscorpiones (new definition).

Due the unanticipated large size of the *C. scorpioides* genome (3.6 Gb), and the ensuing fragmentation of the assembly, we were not able to assess the number of Hox clusters in Pseudoscorpiones, which would constitute an independent

test of the hypothesized shared WGD (but see Hoy et al. 2016 for a case of atomized Hox clusters in a mite). A forthcoming long-read, proximity ligation-based genome assembly of this species is anticipated to inform the ancestral architecture of arachnopulmonate genomes. One additional line of evidence that would support this phylogenetic inference would be embryonic gene expression patterns of ohnologs known to exhibit shared spatiotemporal dynamics in developing appendages of spiders and scorpions (e.g., *dac*; *hth*; Nolan et al. 2020). More recently, evidence from whip spiders (Amblypygi) has additionally supported the inference of conserved expression domains of ohnologs that correspond to gene tree topologies (Gainett and Sharma 2020). Although we endeavored to generate expression data for the two copies of the appendage patterning transcription factors *dac*, *hth*, *exd*, and *omb* in *C. crassus*, we encountered technical challenges incurred by cuticle deposition early in pseudoscorpion development, as well as paucity of embryonic tissue. Whole mount in situ hybridization in pseudoscorpion embryos likely requires modified in situ hybridization protocols previously developed for highly sclerotized chelicerate embryos (e.g., sea spiders; Jager et al. 2006). Future efforts must establish a
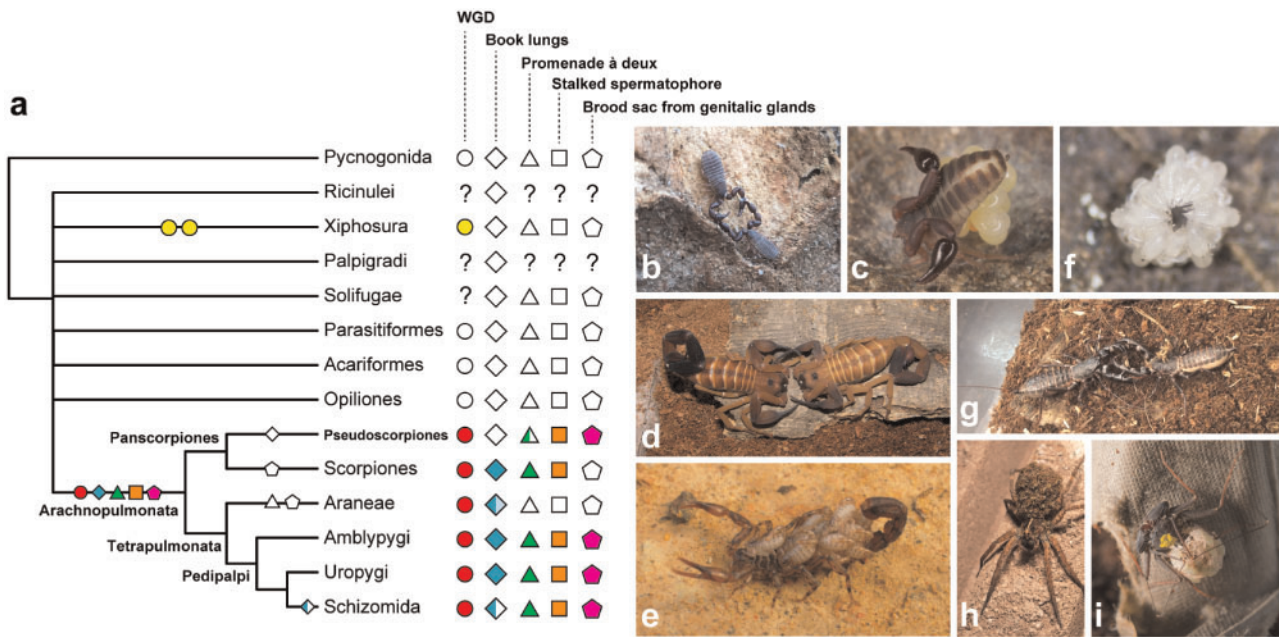
FIG. 10. Ancestral state reconstruction of shared genome duplication events and reproductive behaviors in Chelicerata, under accelerated transformation. (a) The revised placement of pseudoscorpions supports a shared origin of courtship behavior and maternal brood care across Arachnopulmonata. Empty symbols indicate absences. For book lungs, half-filled symbol for Schizomida reflects loss of posterior book lung pair; gradient in Araneae reflects transformation of posterior book lung pair to tracheal tubules in derived spiders. For *promenade á deux*, half-filled symbol for Pseudoscorpiones reflects retention only in Cheliferoidea. (b) A mating pair of *Conicochernes crassus* performing the *promenade á deux* (Denmark, Western Australia; photograph: A.Z. Ontano). (c) Maternal brood care in a chernetid (photograph: G. Giribet). (d) *Promenade á deux* behavior in the buthid scorpion *Babycurus gigas* (photograph: M. Cozijn). (e) Maternal care in the vaejovid scorpion *Vaejovis zapoteca*, with scorplings on the back of the female (photograph: C.E. Santibáñez-López). (f) Postembryos of an undescribed species of the schizomid genus *Rowlandius*; hatchling cluster removed from the female's back for image clarity (photograph: L. Carvalho). (g) A mating pair of the uropygid *Mastigoproctus giganteus* (photograph: A. Hochberg, R. Hochberg). (h) Female of the lycosid spider *Hogna* sp. with spiderlings on the back of the female (photograph: J.A. Ballesteros). (i) Female of the whip spider *Phrynus marginemaculatus* with postembryos on the back of the female; yellow marking is a biological paint used to distinguish individuals in a captive breeding colony (photograph: G. Gainett).

reliable pseudoscorpion model system for testing the downstream hypothesis that expression patterns of pseudoscorpion ohnolog pairs reflect arachnopulmonate-specific patterns. The establishment of a reliable pseudoscorpion model system would constitute a useful comparative data point for assessing the decay of ohnologs' expression patterns as a function of phylogenetic distance.

Ancient Origins of Courtship Behavior and Brood Care in Arachnopulmonata

The recovery of Pseudoscorpiones as the sister group of scorpions markedly alters the reconstruction of several key character in the chelicerate tree of life (fig. 10). Regarding their respiratory system, pseudoscorpions are reconstructed as arachnopulmonates that have secondarily lost their book lungs; instead, pseudoscorpions typically exhibit two pairs of tracheal tubules opening as spiracles on the third and the fourth opisthosomal segments. The evolutionary transition of book lungs to tracheal tubules is broadly associated with miniaturization in other arachnopulmonate orders (Dunlop 2019). For example, in derived spiders, the posterior pair of book lungs is replaced by openings of the tracheal tubules as well, which in turn have a complex evolutionary history within this order (Ramírez et al. 2021). In Schizomida,

the posterior pair of respiratory organs is lost altogether (Hansen and Sørensen 1905; Shultz 1990).

Separately, an arachnopulmonate affinity for pseudoscorpions suggests that both a courtship behavior and a mode of parental care are ancient across this group. Like scorpions, Amblypygi, Uropygi, and Schizomida, pseudoscorpions of the superfamily Cheliferoidea perform a characteristic courtship dance (the *promenade à deux*), wherein the male clasps the female using the pedipalps and the pair navigate over a substrate (fig. 10b, d, and g; Gravely 1915). The inferred purpose of this behavior is to guide the female to the spermatophore deposited by the male onto the substrate. The *promenade à deux* behavior is secondarily lost in spiders, which exhibit other, often complex, courtship behaviors. In addition, spiders do not produce an external spermatophore during mating; typically, sperm are passed to specialized copulatory bulbs on the distal palps, which are used for internal fertilization. Given the tree topology supported by analyses (reciprocally monophyletic Panscorpiones + Tetrapulmonata), and under accelerated transformation of character states (fig. 10), the *promenade à deux* appears to be a possible synapomorphy of Arachnopulmonata that was secondarily lost in spiders as well as in the common ancestor of Pseudoscorpiones, with a secondary regain in Cheliferoidea, or its retention in Cheliferoidea may represent a plesiomorphy

that reflects arachnopulmonate affinity. An equally parsimonious scenario (under delayed transformation; not shown) constitutes independent gains in Pedipalpi and Panscorpiones, with the same sequence of loss and regain of this character within Pseudoscorpiones. A less ambiguous reconstruction is the presence of a stalked spermatophore attached to the substrate is found across pseudoscorpion superfamilies, as well as in scorpions, Amblypygi, Uropygi, and Schizomida (Shultz 2007). A similarity of spermatophore structure in scorpions and pseudoscorpions has previously been noted as well (Francke 1979).

Many pseudoscorpion superfamilies will produce a brood sac on the underside of the female's opisthosoma that is secreted by gonoporal glands, wherein embryos develop until hatching (fig. 10c). A condition unique to brooding pseudoscorpion lineages is that developing embryos are additionally provisioned by nutritive secretions of the female (Weygoldt 1969). The production of a brood sac from genitalic glands is shared by Amblypygi, Uropygi, and Schizomida, which also brood embryos on the underside of the opisthosoma (Gravely 1915; Rowland 1972). The incidence of this mode of development in pseudoscorpions was previously thought to represent a morphological convergence (Shultz 1990). Scorpions exhibit a derived state in this regard, with all extant Scorpiones bearing live young (fig. 10d). Upon birth or hatching from the egg, postembryos of scorpions, Amblypygi, Uropygi, and Schizomida will climb onto the female's back until they advance to additional instar stages (fig. 10e, f, h, and i). A similar form of brood care (carrying of the eggs) occurs in some acarine groups as well, as exemplified by argasid ticks (Pienaar et al. 2018), though the hatchlings are not known to be carried by the adult females.

Pseudoscorpion postembryonic care is variable across this order, but can take the form of females forming brood chambers and cohabiting these with offspring (Weygoldt 1969). As with insemination, spiders again bear a derived form of brood care within arachnopulmonates, with the female typically enveloping egg masses in silk. Brood care in spiders is variable; egg sacs may be guarded by females in burrows until juveniles achieve a later instar and disperse (e.g., mesotheles and mygalomorphs), attached to the substrate (e.g., some Ctenidae, Corinnidae, Selenopidae, and Hersiliidae), attached to webs (most araneomorphs), or carried on the female's back (e.g., Lycosoidea; fig. 10h). In addition, brood care consisting of egg guarding has independently evolved in Solifugae and several times within laniatorean Opiliones (Punzo 1998; Machado and Macías-Ordóñez 2007).

Given the distribution of the *promenade à deux*, the stalked spermatophore, the production of the maternal brood sac from gonoporal glands, and comparable forms of maternal brood care across Chelicerata, we infer these four characters to be ancestral to Arachnopulmonata. As the the oldest known arachnopulmonate, *Parioscorpio venator*, is Silurian in age (439 Ma; Wendruff et al. 2020), the *promenade à deux* may constitute the oldest known courtship behavior.

The recovery of Panscorpiones precipitates reevaluation of other characters, whose homology in now in question. Key among these are venoms of Iocheirata (a clade of venomous

pseudoscorpions, which excludes Chthonioidea and Feaelloidea), scorpions, and spiders. As the venom glands of each of these groups do not share positional homology (pedipalpal fingers in pseudoscorpions; posterior-most somite in scorpions; chelicerae in spiders), it is most likely that each group has undergone independent recruitment of housekeeping genes to serve as venom peptides, though striking similarities exist in some toxins of these three groups and may constitute a deep homology (Santibáñez-López et al. 2018; Krämer et al. 2019). On the other hand, the evolution of silks, which occur in spiders, some pseudoscorpions, and some Acariformes (once again, with no shared positional homology of silk-producing organs), is most likely to reflect independent evolutionary gains.

Prospects for a Resolved "Arachnid" Phylogeny

Topological uncertainty in chelicerate phylogeny extends to the traditionally accepted monophyly of Arachnida, with an array of phylogenomic analyses recovering the derived placement of Xiphosura as the sister group of Ricinulei (Ballesteros and Sharma 2019; Ballesteros et al. 2019). In this study, Xiphosura was recovered as the sister group of Ricinulei (118/189 analyses), as part of a clade with Ricinulei and Solifugae (50/189 analyses), or sister group to a larger clade of derived arachnids, such as Arachnopulmonata (21/189), as previously reported (Ballesteros and Sharma 2019; Ballesteros et al. 2019). Once again, we found no support for arachnid monophyly.

This result has been challenged by another suite of phylogenomic studies (Lozano-Fernández et al. 2019; Howard et al. 2020) that have suggested three potential solutions to recovering arachnid monophyly: denser taxonomic sampling (Lozano-Fernández et al. 2019), the use of the site heterogeneous CAT model (Lozano-Fernández et al. 2019; Howard et al. 2020), and the use of slowly evolving (and/or less saturated) loci (Lozano-Fernández et al. 2019; Howard et al. 2020). Given the unstable support for an arachnid monophyly across phylogenomic data sets, it has been contended that the morphological result of arachnid monophyly should be accepted as the most likely evolutionary scenario (Howard et al. 2020).

As we have previously shown, the most taxon-rich phylogenomic data set of chelicerates—and the sole analysis sampling all extant chelicerate orders—does not support arachnid monophyly, including under the CAT model (Ballesteros et al. 2019). Recent reanalyses of data sets that had previously recovered arachnid monophyly under certain models (e.g., Regier et al. 2010; 500-slowest evolving genes in Sharma, Kaluziak, et al. 2014), showed that higher support for Arachnida could be obtained if these were analyzed under site heterogenenous models (Howard et al. 2020). Howard et al. ignored the observation that previous analyses computed under the CAT + GTR + Γ_4 model (as well as the PMSF model) do not consistently recover arachnid monophyly, including for data sets restricted to slowly evolving genes (figure 7 of Sharma, Kaluziak, et al. 2014; Ballesteros and Sharma 2019; Ballesteros et al. 2019). They also ignored emendations of the Lozano-Fernández et al. (2019) matrices that were subsequently augmented to include all chelicerate orders,

which incurred the collapse of arachnid monophyly despite use of site heterogeneous models (figure 3 of Ballesteros et al. 2019).

Howard et al. (2020) introduced a new matrix from subsampling the 200-slowest evolving genes of another data set produced by Lozano-Fernández et al. (2019; "Matrix B"); they suggested a trend of increasing support for arachnid monophyly with increasing model complexity, with maximal support for arachnid monophyly under the CAT-Poisson model. Their choice of the CAT model in those reanalyses is peculiar, given that CAT + Poisson is demonstrably less accurate than CAT + GTR + Γ_4 (Whelan and Halanych 2016). In any case, as a parenthetical test of the validity of the claims of Howard et al. (2020), we augmented their 200-slowest evolving gene data set to include two phylogenetically significant lineages previously sequenced by us: the palpigrade *Eukoenenia spelaea* (131/200 loci) and the opilioacariform *Adenacarus* sp. (180/200 loci), whose omission in the analyses of Howard et al. (2020) struck us as odd. Methods were identical to our approach in Ballesteros et al. 2019; the augmented alignments are available on the Dryad Digital Repository). These modifications achieved the sampling of all extant chelicerate orders (Palpigradi), as well as the slowly evolving putative sister group of Parasitiformes (Opilioacariformes). We then computed topologies under the same three likelihood models implemented by Howard et al. (2020) (LG + F + Γ; LG + R5; LG + C20 + R5).

Separately, we performed this same family of analyses, after removing 10 loci that represent duplicated genes in the 200-locus data set. Duplicates in this context refers to identical or nearly identical alignments that recur in the same supermatrix. These are typically the result of failing to reduce input transcriptomes to single isoforms per Trinity gene, prior to analysis with OMA (Altenhoff et al. 2013). A list of these erroneously duplicated alignments is provided in the Dryad Digital Repository.

As shown in supplementary figures S2 and S3, Supplementary Material online, the inclusion of just two phylogenetically significant lineages (with or without the removal of the duplicated loci) to the analyses of Howard et al. (2020) is sufficient to break arachnid monophyly, as well as the monophyly of Acari, with significant nodal support (90–99% ultrafast bootstrap resampling frequency), under all three substitution models. The consistent recovery of a non-monophyletic Acari in matrices that sample the basally branching parasitiform lineage Opilioacariformes (e.g., Ballesteros et al. 2019; this study) suggests that Acari monophyly is an another long-branch attraction artifact in chelicerate phylogeny. Taken together with the analyses of Ballesteros et al. (2019), as well as our analyses of BUSCO genes (this study), our reanalyses of OMA-inferred orthologs from Howard et al. (2020) suggest that neither Arachnida nor Acari are substantiated by dense taxonomic sampling, slowly evolving genes, site heterogeneous models, approach to orthology inference, or various combinations thereof.

Across Chelicerata, a subset of genes supporting arachnid monophyly, as identified by a ΔGLS framework, were previously shown to be statistically indistinguishable from the majority (which supported Xiphosura as derived), with respect to 70 parameters, including evolutionary rate, compositional heterogeneity, and alignment length (figure 3 of Ballesteros and Sharma 2019). In the present study, of the 189 phylogenetic analyses we performed using an independent orthology criterion for locus selection (BUSCO genes), not one analysis recovered arachnid monophyly. In addition, surveys of miRNAs revealed no support for Arachnida, either in the form of miRNAs unique to arachnids, or evidence of an arachnid-specific duplication (note that although not all chelicerate orders are represented by genomes, this should not hinder the recovery of putative arachnid-specific miRNAs in our analysis; Garb et al. 2018). Recovering arachnid monophyly in molecular data sets appears to require a concerted, and largely contrived, effort to circumscribe taxa, loci, models, and algorithms that will recover this preconceived relationship. As we have previously shown, this practice is questionable (if not outright unscientific) because it can be used to justify nonsensical groupings (figure 8 of Ballesteros and Sharma 2019). The attribution of arachnid non-monophyly to unspecified systematic biases or artifacts remains an unsubstantiated notion.

Strong arguments in favor of arachnid monophyly remain the domain of morphological and paleontological data sets; these span the nature of mouthparts, eyes, respiratory systems, and stratigraphic distributions of marine versus terrestrial lineages, among others (reviewed by Howard et al. 2020). Such discussions eerily echo arguments once advanced in support of Tracheata (Myriapoda + Hexapoda, or the terrestrial mandibulates), a group revealed by molecular phylogenetics to be an artifact of morphological convergence in another subset of terrestrial arthropods. As the history of hypotheses like Pulmonata (Gastropoda) and Tracheata has repeatedly shown, terrestrial lineages are highly prone to convergence, often to an astonishing degree (Friedrich and Tautz 1995; Shultz and Regier 2000; Giribet et al. 2001; Jörger et al. 2010). Shared reduction of the appendage-less intercalary segment (third head segment), the incidence of uniramous appendages, the gnathobasic architecture of the mandible, and the organization of the tracheal tubules in hexapods and myriapods serve as powerful examples of how parallel adaptations to life on land can confound interpretations of synapomorphies. More generally, the "holistic" approach of Howard et al. (2020) simply fails to reconcile its dependence upon the validity of morphology to support one questionable node (Arachnida) with its simultaneous dismissal of morphological data sets' consistent inability to recover the only higher-level chelicerate relationships that are robustly and independently supported by other data classes (Arachnopulmonata and Euchelicerata; figure 1 of Nolan et al. 2020).

We submit that an objective approach to testing phylogenetic hypotheses of terrestrialization in arthropods must regard traditional groupings with skepticism, rather than querying molecular sequence data for genes and data sets supporting preconceived relationships. Such investigations must also account for new neurophylogenetic characters that have recently suggested morphological support for a

closer relationship of Xiphosura to Arachnopulmonata (Lehmann and Melzer 2019a, 2019b). Due to the lack of genomes for Ricinulei (the putative Xiphosura sister group in some phylogenies) as well as other poorly studied arachnid groups (e.g., Palpigradi and Solifugae), we were not able to assess miRNAs or other rare genomic changes to test the competing hypothesis of Ricinulei + Xiphosura. However, the incidence of WGDs in horseshoe crabs proffers the tantalizing possibility of applying the approaches used herein to assess this competing hypothesis, as at least one of the two WGD events in Xiphosura is thought to be ancient (Roelofs et al. 2020). The discovery of shared duplications of gene families, miRNAs, and syntenic blocks between different sets of chelicerate orders could be used to evaluate independently the monophyly of Arachnida, as well as the placement of the unstable apulmonate orders. Future efforts should therefore target the generation of genomic resources for Ricinulei, Palpigradi, and Solifugae to reevaluate such hypotheses as Haplocnemata (Solifugae + Pseudoscorpiones), Megoperculata (Palpigradi + Tetrapulmonata), and Arachnida itself.

Conclusions

Consilience in phylogenetics is the outcome of multiple, independent topological tests recovering support for the same hypothesis (e.g., Rota-Stabelli et al. 2011; Fröbius and Funch 2017; Marlétaz et al. 2019). Here, we demonstrated that analyses of sequence data, gene family duplications, gene tree topologies of arachnopulmonate-specific paralogs, and miRNA duplications independently support a nested placement of pseudoscorpions within Arachnopulmonata. Our results reinforce that topological accuracy in the placement of long-branch taxa is most affected by dense sampling of basally branching lineages, rather than algorithmic approach (supermatrix vs. coalescent-based summary methods), matrix completeness, evolutionary rate, or model choice alone. Improvements to chelicerate phylogeny must therefore focus on the identification of basally branching groups within orders whose internal relationships remain poorly understood, such as Solifugae, Amblypygi, Uropygi, and Schizomida. Leveraging rare genomic changes stemming from the genome duplications exhibited by a subset of chelicerate orders may be a key to resolving some of the most obdurate nodes in the chelicerate tree of life.

Materials and Methods

Species Sampling

For phylogenetic reconstruction, we generated a data set of 117 chelicerates (40 pseudoscorpions, 12 scorpions, 17 spiders, 4 Pedipalpi, 13 Opiliones, 5 Ricinulei, 3 Xiphosura, 2 Solifugae, 9 Parasitiformes, 10 Acariformes, 2 Pycnogonida) and 15 outgroups (3 Onychophora, 4 Myriapoda, 8 Pancrustacea) . Taxon selection prioritized the representation of basal splits in all major groups (Sharma, Fernández, et al. 2015; Fernández et al. 2017, 2018; Ballesteros et al. 2019, 2020; Benavides et al. 2019; Santibáñez-López et al. 2019, 2020). Libraries of high quality were additionally selected

such that all chelicerate orders were represented in >95% of loci by at least one terminal, in all matrices constructed. Although we trialed the inclusion of a palpigrade library recently generated by us (Ballesteros et al. 2019), the low representation of BUSCO genes for this taxon across data sets (46–70%) prohibited the inclusion of this order in downstream analyses. A list of taxa and sequence accession data is provided in supplementary table S1, Supplementary Material online.

Orthology Inference and Phylogenomic Methods

Candidate ORFs were identified in transcripts using TransDecoder (Haas et al. 2013). Loci selected for phylogenomic analysis consisted of the subset of 1066 Benchmarked Universal Single Copy Orthologs identified for Arthropoda (BUSCO-Ar). For each library, these were discovered using a hidden Markov model approach, following the procedure detailed in Leite et al. (2018). Multiple sequence alignment was performed using MAFFT 7.3.8 (–anysymbol –auto; Katoh and Standley 2013). Gap-rich regions were masked with trimAl 1.2 (–gappyout; Capella-Gutiérrez et al. 2009) and alignment coverage verified and sanitized with Al2Phylo (-m 50 -p 0.25 -t 20; Ballesteros and Hormiga 2016).

To assess the tradeoff between data completeness and the number of loci per data set, six matrices were constructed by setting taxon occupancy thresholds to 55% (1002 loci), 60% (945 loci), 65% (846 loci), 70% (693 loci), 75% (480 loci), and 80% (248 loci) of total taxa. These thresholds were selected to represent broadly commonly occurring values for matrix completeness in phylotranscriptomic studies of metazoans. Representation of each terminal and ordinal lineage per matrix is provided in supplementary table S2, Supplementary Material online.

To assess the effect of denser taxonomic sampling on the placement of Pseudoscorpiones, basally branching lineages of pseudoscorpions (corresponding to superfamilies or families) were sequentially pruned until only Cheliferoidea (Cheliferidae + Chernetidae) was retained. Thus, six additional matrices were constructed, with sequential pruning of Chthonioidea (six terminals), Feaelloidea (two terminals), Neobisioidea (ten terminals), Garypoidea (five terminals), Garypinoidea (three terminals), and Cheridoidea + Sternophoroidea (two terminals). Pruning was performed for each of the six matrices constructed according to taxon occupancy thresholds, resulting in 42 matrices in total.

Tree topologies for individual loci and for concatenated data sets were computed with IQ-TREE 1.6.8 (Nguyen et al. 2015; Chernomor et al. 2016), coupled with model selection of substitution and rate heterogeneity based on the Bayesian Information Criterion (Kalyaanamoorthy et al. 2017) and 1000 ultrafast bootstraps to assess branch support (-m MFP -mset LG, JTT, WAG -st AA -bb 1000; Hoang et al. 2018). For the subset of the least complete matrices (55% taxon occupancy), we additionally performed model selection under the posterior mean site frequency (PMSF), a mixture model that approximates the CAT model in a maximum likelihood framework (Lartillot and Philippe 2004; Wang et al. 2018).

Analyses were performed using the LG + C20 + F + Γ and LG + C60 + F + Γ models.

To assess the interaction between evolutionary rate and taxon sampling, we selected the 70% complete (693 loci) and 75% complete (480 loci) matrices to optimize the tradeoff between sufficient sampling of genes and low quantity of missing data. These matrices were divided into tertiles of slow-, intermediate- and fast-evolving genes using mean percent pairwise identity as a metric of evolutionary rate, following the approach of Sharma, Fernández, et al. (2015). Subsequent pruning of basally branching pseudoscorpion taxa was performed as in other analyses. Tree inference was performed with a partitioned model-fitting and ASTRAL.

For phylogenetic analyses using multispecies coalescent methods, species trees were estimated with ASTRAL v. 5.14.2 (Mirarab and Warnow 2015; Zhang et al. 2018), using gene trees from IQ-TREE analyses as inputs. Phylogenetic signal at the level of individual genes was quantified using the gene-wise log-likelihood score (ΔGLS) for the unconstrained tree versus a competing hypothesis (Pseudoscorpiones + Acariformes; Pseudoscorpiones + Parasitiformes; Pseudoscorpiones + Scorpiones) (Shen et al. 2017). This metric maps the relative support for each of two competing hypotheses, for every locus in the data set; the amplitude of the log-likelihood indicates the degree of support for either hypothesis.

Embryo Collection, Sequencing, and Mapping of Homeodomains

Given that transcriptomes of adult tissues have been shown to sample poorly transcription factors relevant for developmental patterning in arachnids (Sharma, Santiago, et al. 2015), assessment of homeodomain duplications was performed only for genomes and developmental transcriptomes. The genome of *Cordylochernes scorpioides* was excluded from this analysis, due to the fragmentation of the assembly.

Conicochernes crassus (Pseudoscorpiones: Chernetidae) were hand collected from underneath the bark of karri trees in Denmark, Western Australia (−34.963640, 117.359720). Individuals were reared in plastic containers containing damp paper towels at room temperate to simulate living conditions between bark and sapwood. Adult pseudoscorpions were fed a combination of cricket nymphs and $ap^{-/-}$ fruit flies. Females of *C. crassus* carry developing embryos in a brood sac on the underside of the opisthosoma; individuals were checked for the presence of embryos. Females carrying embryos were separated from the colony for 12–72 h to prevent cannibalism and allow embryos to mature to a range of developmental time points. Entire brood sacs were then separated from the opisthosoma using forceps wetted with distilled water to prevent damage to the females before being returned to the colony.

Establishment of *Phrynus marginemaculatus* (Amblypygi: Phrynidae) for the study of developmental genetics and the comparative development was previously described by Gainett and Sharma (2020). Embryos of the whip spiders *Charinus ioanniticus* and *Charinus israelensis* were obtained by hand collecting brooding females from two cave sites in Israel, Hribet Hruba (31.913280, 34.960830) and Mimlach (32.858150, 35.44410). Two stages of deutembryos were obtained and sequenced for each species. Further details are provided in Gainett et al. (2020).

Field collection of embryos of the tarantula *Aphonopelma hentzi* (Araneae: Theraphosidae) for developmental genetics and transcriptomics was previously described by Setton et al. (2019).

Field collection of embryos and larvae was performed for five species of Pycnogonida: *Nymphon moelleri* (Nymphonidae), *Pallenella flava* (Callipallenidae), Stylopallene cheilorhynchus (Callipallenidae), *Phoxichilidium femoratum* (Phoxichilidiidae), and *Tanystylum orbiculare* (Ammotheidae). The details of collection and sequencing are provided in Ballesteros et al. (2020).

Embryos were transferred to Trizol Tri-reagent (Ambion Life Technologies, Waltham, MA, USA) for RNA extraction, following manufacturer's protocols. Library preparation and stranded mRNA sequencing were performed at the University of Wisconsin-Madison Biotechnology Center on an Illumina HiSeq 2500 platform (paired-end reads of 125 bp). Raw sequence reads are deposited in NCBI Sequence Read Archive. Filtering of raw reads and strand-specific assembly using Trinity v. 2.8.3 followed our previous approaches (Sharma et al. 2014; Ballesteros et al. 2019).

Discovery of homeobox genes followed the approach previously outlined by Leite et al. (2018). Briefly, homeodomain sequences were identified from genomes and embryonic transcriptomes using BLAST v. 2.9.0 or v. 2.10.0 (tblastn) (Altschul et al. 1990). Queries consisted of amino acid homeodomain sequences from outgroup arthropod species in HomeoDB (Zhong and Holland 2011) combined with homeodomain sequences from *Parasteatoda tepidariorum* (Schwager et al. 2017), *Centruroides sculpturatus* (Schwager et al. 2017), *Mesobuthus martensii* (Cao et al. 2013), and *Strigamia maritima* (Chipman et al. 2014). As additional chelicerate ingroup taxa, we included the genome of the horseshoe crabs *Limulus polyphemus* (Kenny et al. 2016) and *Carcinoscorpius rotundicauda* (Shingate et al. 2020), the genomes of the mites *Tetranychus urticae* (Grbić et al. 2011) and *Galendromus occidentalis* (Hoy et al. 2016), and a recently re-sequenced embryonic transcriptome of the harvestman *Phalangium opilio* (Sharma et al. 2012; Ballesteros and Sharma 2019). As additional outgroup taxa, we included the embryonic transcriptomes of the millipede *Glomeris marginata* and the onychophoran *Euperipatoides kanangrensis* (Janssen and Budd 2013). We thus assessed homeobox gene duplication for 26 panarthropod species.

All initial BLAST hits were retained. Next, the full protein sequences of the BLAST hits were predicted with TransDecoder v. 5.5.0 (Haas et al. 2013) with default parameters (-m 100; predicted transcripts with less than 100 amino acids were not retained) and thereafter analyzed using the Conserved Domain Database (CDD) (Marchler-Bauer et al. 2015) to confirm the presence of homeodomains and annotate other functional domains. BLAST hits that did not have homeodomains identified by CDD were removed. Transcripts within a species that had identical protein sequences

predicted to encode homeodomains were manually checked. Because this approach conservatively emphasized retention of complete homeobox genes with conserved sequences, we cannot rule out the exclusion of partial transcripts of homeobox genes that lack homeodomains or orthologs with highly divergent sequences. Multiple sequence alignment, trimming to retain only the homeodomain, and classification of verified homologs followed procedures described by Leite et al. (2018).

Analysis of Appendage Patterning Ohnologs

Homologs of four appendage patterning genes were retrieved from the *C. crassus* transcriptome using approaches described above. Multiple sequence alignment of peptide sequences and alignment trimming followed the approach of Nolan et al. (2020). Maximum likelihood inference of tree topologies was performed using IQ-TREE under an $LG + I + \Gamma$ substitution model. Nodal support was estimated using ultrafast bootstrapping.

Cordylochernes scorpioides Genome Sequencing

Illumina fragment libraries (insert sizes 270 and 420 bp) and mate-pair libraries (insert sizes 2, 4, and 8 kb) were constructed by Lucigen Corporation (Middleton, WI, USA). Fragment libraries were constructed from genomic DNA extracted from single individual inbred males; to meet DNA input requirements for mate-pair library construction, genomic DNA from 12 fourth generation inbred individuals was pooled. Fragment libraries were sequenced on HiSeq X with 150 b paired-end sequencing (Hudson Alpha Genomic Services Lab, Huntsville AL), and mate-pair libraries were sequenced on MiSeq with 150 b paired-end sequencing at Lucigen Corporation. The read data was assembled *de novo* at 125X coverage using MaSuRCA v. 3.2.3 (Zimin et al. 2013), with additional scaffolding using SSPACE Standard v 3.0 (BaseClear BV, Netherlands) followed by gap-filling using GapFiller v1.12 (BaseClear). The draft *C. scorpioides* genome assembly was submitted to GenBank (GenBank: QEEW00000000.1) and read data were deposited in NCBI SRA (SRA: SRP144365; BioProject: PRJNA449764). Global statistics for assessment of draft genome quality and completeness are provided on NCBI (https://www.ncbi.nlm.nih.gov/assembly/GCA_003123905.1).

MicroRNA and Hox Genes Orthology Search

Previous work on miRNA occurrence in the genome of the house spider *Parasteatoda tepidariorum* identified 40 miRNA families shared across Arthropoda, and a further 31 either unique to spiders (n = 30) or unique to arachnopulmonates (n = 1) (Leite et al. 2016). To extend this survey to new taxa, we searched for miRNA families in the draft genome assembly of *C. scorpioides* (GCA_003123905.1), as well as the genome of *Mesobuthus martensii* (GCA_000484575.1). All miRNA reported from *P. tepidariorum* were retrieved from the miRBASE and used as query sequences (Kozomara et al. 2019). An initial BLAST search was performed (*blastn –word_size 4 –reward 2 –penalty –3 –evalue 0.05*) and sequences with e-value <0.05 and percentage identity >70% were

retained. To accommodate the fragmentation of the *C. scorpioides* genome, as well as heterozygosity, putative hits were retained only if both the ELEKEF and KIWFQN motifs were discovered in the peptide translation, and peptide sequences were unique (i.e., pairs of sequences with only synonymous substitutions were considered putative alleles). Putative homologs were verified by multiple sequence alignment using MAFFT v. 7.407 (Katoh and Standley 2013). The structure and the minimum free energy of these selected miRNAs were analyzed with RNAfold v. 2.4.13 (as part of the ViennaRNA Package 2.0; Lorenz et al. 2011) and with The Vienna RNA WebServer (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) using default settings. Regarding the previous survey of miRNA families in 16 ecdysozoan taxa by Leite et al. (2016), we corroborated all reported results, except for the discovery that the mygalomorph spider *A. hentzi* exhibits only a single copy of the miRNA *pte-bantam*.

Permitting

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

Author Contributions

A.Z.O. and P.P.S. conceived of the study. A.Z.O., P.P.S., and M.S.H. collected specimens of pseudoscorpions in the field.

M.S.H. performed taxonomic identification. L.R.B. provided unpublished transcriptome assemblies. A.Z.O. cultivated pseudoscorpion embryos, performed the sequencing, and implemented phylogenomic analyses. Fieldwork and tissue collection was performed by S.A., J.AB., G.G., E.G.R., and P.P.S. for whip spider embryos; by G.B. and P.P.S. for sea spider embryos; and by E.V.W.S. and P.P.S. for mygalomorph embryos. Analysis of homeobox duplications was performed by A.Z.O., G.G., K.F.C., J.AB., E.V.W.S., J.T.Z., and P.P.S. Analysis of miRNAs was performed by C.E.S.L. J.A.Z. and D.W.Z. collected *C. scorpioides* in Panamá, established a laboratory population of the pseudoscorpion, and conceived of the genome sequencing of *C. scorpioides*. S.M., J.A.Z. and D.W.Z. were responsible for *C. scorpioides* genome sequencing and assembly. A.Z.O. and P.P.S. wrote the manuscript, and all authors edited and approved the final content.

Data Availability

The complete data set, including sequence alignments, tree files, miRNA alignments, and embryonic transcriptomic assemblies, have been deposited in the Dryad Digital Repository under URL https://datadryad.org/stash/dataset/doi:10.5061/dryad.9ghx3ffg9. Raw read data for the developmental transcriptome are available in NCBI Sequence Read Archive (PRJNA675208). The *C. scorpioides* genome assembly and associated SRA are available in GenBank under the WGS master record QEEW00000000.1.

References

Altenhoff AM, Gil M, Gonnet GH, Dessimoz C. 2013. Inferring hierarchical orthologous groups from orthologous gene pairs. *PLoS One* 8(1):e53786.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol*. 215(3):403–410.

Aria C, Caron J-B. 2019. A middle Cambrian arthropod with chelicerae and proto-book gills. *Nature* 573(7775):586–584.

Arribas P, Andújar C, Lourdes MM, Linard B, Emerson BC, Vogler AP. 2020. Mitochondrial metagenomics reveals the ancient origin and phylodiversity of soil mites and provides a phylogeny of the Acari. *Mol Biol Evol*. 37(3):683–694.

Ballesteros JA, Hormiga G. 2016. A new orthology assessment method for phylogenomic data: unrooted Phylogenetic Orthology. *Mol Biol Evol*. 33(8):2117–2134.

Ballesteros JA, Santibáñez López CE, Kováč Ľ, 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 B*. 286(1917):20192426.

Ballesteros JA, Sharma PP. 2019. A critical appraisal of the placement of Xiphosura (Chelicerata) with account of known sources of phylogenetic error. *Syst Biol*. 68(6):896–917.

Ballesteros JA, Setton EVW, Santibáñez-López CE, Arango CP, Brenneis G, Brix S, Corbett KF, Cano-Sánchez E, Dandouch M, Dilly GF, et al. 2020. Phylogenomic resolution of sea spider diversification through integration of multiple data classes. *Mol Biol Evol*. doi:10.1093/molbev/msaa228

Barnett AA, Thomas RH. 2013. The expression of limb gap genes in the mite *Archegozetes longisetosus* reveals differential patterning mechanisms in chelicerates. *Evol Dev*. 15(4):280–292.

Benavides LR, Cosgrove JG, Harvey MS, Giribet G. 2019. Phylogenomic interrogation resolves the backbone of the Pseudoscorpiones tree of life. *Mol Phylogenet Evol*. 139:106509.

Bergsten J. 2005. A review of long-branch attraction. *Cladistics* 21(2):163–193.

Bicknell RDC, Lustri L, Brougham T. 2019. Revision of "*Bellinurus*" *carteri* (Chelicerata: xiphosura) from the Late Devonian of Pennsylvania, USA. *Comptes Rendus Palevol*. 18(8):967–976.

Blackburn DC, Conley KW, Plachetzki DC, Kempler K, Battelle B-A, Brown NL. 2008. Isolation and expression of *Pax6* and *atonal* homologues in the American horseshoe crab, *Limulus polyphemus*. *Dev Dyn*. 237(8):2209–2219.

Borner J, Rehm P, Schill RO, Ebersberger I, Burmester T. 2014. A transcriptome approach to ecdysozoan phylogeny. *Mol Phylogenet Evol*. 80:79–87.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15):1972–1973.

Charrier NP, Hermouet A, Hervet C, Agoulon A, Barker SC, Heylen D, Toty C, McCoy KD, Plantard O, Rispe C. 2019. A transcriptome-based phylogenetic study of hard ticks (Ixodidae). *Sci Rep*. 9(1):12923.

Chernomor O, von Haeseler A, Minh BQ. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst Biol*. 65(6):997–1008.

Chipman AD, Ferrier DEK, Brena C, Qu J, Hughes DST, Schröder R, Torres-Oliva M, Znassi N, Jiang H, Almeida FC, et al. 2014. The first myriapod genome sequence reveals conservative arthropod gene content and genome organisation in the centipede *Strigamia maritima*. *PLoS Biol*. 12(11):e1002005.

Cao Z, Yu Y, Wu Y, Hao P, Di Z, He Y, Chen Z, Yang W, Shen Z, He X, et al. 2013. The genome of *Mesobuthus martensii* reveals a unique adaptation model of arthropods. *Nat Commun*. 4:2602.

Dehal P, Boore JL. 2005. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol*. 3(10):e314.

Dunlop JA. 2019. Miniaturisation in Chelicerata. *Arthropod Struct Dev*. 48:20–34.

Fernández R, Giribet G. 2015. Unnoticed in the tropics: phylogenomic resolution of the poorly known arachnid order Ricinulei (Arachnida). *R Soc Open Sci*. 2(6):150065.

Fernández R, Sharma PP, Tourinho AL, Giribet G. 2017. The Opiliones tree of life: shedding light on harvestmen relationships through transcriptomics. *Proc R Soc B*. 284(1849):20162340.

Fernández R, Kallal RJ, Dimitrov D, Ballesteros JA, Arnedo MA, Giribet G, Hormiga G. 2018. Phylogenomics, diversification dynamics, and comparative transcriptomics across the spider tree of life. *Curr Biol*. 28(9):1489–1497.

Feuda R, Dohrmann M, Pett W, Philippe H, Rota-Stabelli O, Lartillot N, Wörheide G, Pisani D. 2017. Improved modeling of compositional heterogeneity supports sponges as sister to all other animals. *Curr Biol*. 27(24):3864–3870.

Francke OF. 1979. Spermatophores of some North American scorpions (Arachnida, Scorpiones). *J Arachnol*. 7:19–32.

Friedrich M, Tautz D. 1995. Ribosomal DNA phylogeny of the major extant arthropod classes and the evolution of myriapods. *Nature* 376(6536):165–167.

Fröbius AC, Funch P. 2017. Rotiferan Hox genes give new insights into the evolution of metazoan bodyplans. *Nature Commun*. 8:9.

Gainett G, Ballesteros JA, Kanzler CR, Zehms JT, Zern JM, Aharon S, Gavish-Regev E, Sharma PP. 2020. Systemic paralogy and function of retinal determination network homologs in arachnids. *BMC Genom*. 21:811.

Gainett G, Sharma PP. 2020. Genomic resources and toolkits for developmental study of whip spiders (Amblypygi) provide insights into arachnid genome evolution and antenniform leg patterning. *EvoDevo*. 11(1):18.

Garb JE, Sharma PP, Ayoub NA. 2018. Recent progress and prospects for advancing arachnid genomics. *Curr Opin Insect Sci*. 25:51–57.

Garwood RJ, Dunlop J. 2014. Three-dimensional reconstruction and the phylogeny of extinct chelicerate orders. *PeerJ*. 2:e641.

Giribet G, Edgecombe GD, Wheeler WC. 2001. Arthropod phylogeny based on eight molecular loci and morphology. *Nature* 413(6852):157–161.

Gravely FH. 1915. Notes on the habits of Indian insects, myriapods and arachnids. *Records Indian Mus.* 11:483–539.

Grbic M, Khila A, Lee K-Z, Bjelica A, Grbic V, Whistlecraft J, Verdon L, Navajas M, Nagy L. 2007. Mity model: *Tetranychus urticae*, a candidate for chelicerate model organism. *BioEssays* 29(5):489–496.

Grbić M, Leeuwen TV, Clark RM, Rombauts S, Rouzé P, Grbić V, Osborne EJ, Dermauw W, Ngoc PCT, Ortego F, et al. 2011. The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. *Nature* 479(7374):487–492.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc.* 8(8):1494–1512.

Hansen HJ, Sørensen W. 1905. The Tartarides, a tribe of the order Pedipalpi. *Ark Zool.* 2(8):1–78.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LE. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol.* 35(2):518–522.

Howard RJ, Puttick MN, Edgecombe GD, Lozano-Fernández J. 2020. Arachnid monophyly: morphological, palaeontological and molecular support for a single terrestrialization within Chelicerata. *Arthropod Struct Dev.* 59:100997.

Hoy MA, Waterhouse RM, Wu K, Estep AS, Ioannidis P, Palmer WJ, Pomerantz AF, Simão FA, Thomas J, Jiggins FM, et al. 2016. Genome sequencing of the phytoseiid predatory mite *Metaseiulus occidentalis* reveals completely atomized *Hox* genes and superdynamic intron evolution. *Genome Biol Evol.* 8(6):1762–1775.

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(4):623–627.

Jager M, Murienne J, Clabaut C, Deutsch J, Guyader HL, Manuel M. 2006. Homology of arthropod anterior appendages revealed by Hox gene expression in a sea spider. *Nature* 441(7092):506–508.

Janssen R, Budd GE. 2013. Deciphering the onychophoran 'segmentation gene cascade': gene expression reveals limited involvement of pair rule gene orthologs in segmentation, but a highly conserved segment polarity gene network. *Dev Biol.* 382(1):224–234.

Jörger KM, Stöger I, Kano Y, Fukuda H, Knebelsberger T, Schrödl M. 2010. On the origin of Acochlidia and other enigmatic euthyneuran gastropods, with implications for the systematics of Heterobranchia. *BMC Evol Biol.* 10:323.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 14(6):587–589.

Katoh K, Standley DM. 2013. MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30(4):772–780.

Kenny NJ, Chan KW, Nong W, Qu Z, Maeso I, Yip HY, Chan TF, Kwan HS, Holland PWH, Chu KH, et al. 2016. Ancestral whole-genome duplication in the marine chelicerate horseshoe crabs. *Heredity* 116(2):190–199.

King N, Rokas A. 2017. Embracing uncertainty in reconstructing early animal evolution. *Curr Biol.* 27(19):R1081–R1088.

Kocot KM, Struck TH, Merkel J, Waits DS, Todt C, Brannock PM, Weese DA, Cannon JT, Moroz LL, Lieb B, et al. 2016. Phylogenomics of Lophotrochozoa with consideration of systematic error. *Syst Biol.* 66(2):256–282.

Kozomara A, Birgaoanu M, Griffiths-Jones S. 2019. miRBase: from microRNA sequences to function. *Nucleic Acids Res.* 47(D1):D155–D162.

Krämer J, Pohl H, Predel H. 2019. Venom collection and analysis in the pseudoscorpion *Chelifer cancroides* (Pseudoscorpiones: Cheliferidae). *Toxicon* 162:15–23.

Lamsdell JC. 2016. Horseshoe crab phylogeny and independent colonizations of fresh water: ecological invasion as a driver for morphological innovation. *Palaeontology* 59(2):181–194.

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.

Laumer CE, Fernández R, Lemer S, Combosch D, Kocot KM, Riesgo A, Andrade SCS, Sterrer W, Sørensen MV, Giribet G. 2019. Revisiting metazoan phylogeny with genomic sampling of all phyla. *Proc R Soc B.* 286(1906):20190831.

Legg DA, Sutton MD, Edgecombe GD. 2013. Arthropod fossil data increase congruence of morphological and molecular phylogenies. *Nat Commun.* 4:2485.

Lehmann T, Melzer RR. 2019a. The visual system of Thelyphonida (whip scorpions): support for Arachnopulmonata. *Arthropod Struct Dev.* 51:23–31.

Lehmann T, Melzer RR. 2019b. Also looking like *Limulus*? – Retinula axons and visual neuropils of Amblypygi (whip spiders). *Front Zool.* 15:52.

Leite DJ, Ninova M, Hilbrant M, Arif S, Griffiths-Jones S, Ronshaugen M, McGregor AP. 2016. Pervasive microRNA duplication in Chelicerates: insights from the embryonic microRNA repertoire of the spider *Parasteatoda tepidariorum*. *Genome Biol Evol.* 8(7):2133–2144.

Leite DJ, Baudouin-Gonzalez L, Iwasaki-Yokozawa S, Lozano-Fernández J, Turetzek N, Akiyama-Oda Y, Prpic N-M, Pisani D, Oda H, Sharma PP, et al. 2018. Homeobox gene duplication and divergence in arachnids. *Mol Biol Evol.* 35(9):2240–2253.

Lorenz R, Bernhart SH, Höner zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA Package 2.0. *Algorithms Mol Biol.* 6:26.

Lozano-Fernández J, Tanner AR, Giacomelli M, Carton M, 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.* 10:2295.

Machado G, Macías-Ordóñez R. 2007. Reproduction. In: Pinto-da-Rocha R, Machado G, Giribet G, editors. Harvestmen: the biology of opiliones. Cambridge (MA): Harvard University Press. p. 414–454.

Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, et al. 2015. CDD: NCBI's conserved domain database. *Nucleic Acids Res.* 43(D1):D222–D226.

Marlétaz F, Peijnenburg KTCA, Goto T, Satoh N, Rokhsar DS. 2019. A new spiralian phylogeny places the enigmatic arrow worms among gnathiferans. *Curr Biol.* 21(2):312–318.

Minh BQ, Hahn MW, Lanfear R. 2020. New methods to calculate concordance factors for phylogenomic datasets. *Mol Biol Evol.* 37(9):2727–2733.

Mirarab S, Warnow T. 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* 31(12):i44–i52.

Nguyen L-T, Schmidt HA, Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32(1):268–274.

Nolan ED, Santibáñez-López CE, Sharma PP. 2020. Developmental gene expression as a phylogenetic data class: support for the monophyly of Arachnopulmonata. *Dev Genes Evol.* 230(2):137–153.

Pepato AR, da Rocha CEF, Dunlop JA. 2010. Phylogenetic position of the acariform mites: sensitivity to homology assessment under total evidence. *BMC Evol Biol.* 10(1):235.

Pienaar R, de Klerk DG, Putterill JF, Mans BJ. 2018. Notes on maternal behaviour in soft ticks: specifically observed in *Argas (Argas) striatus* Bedford, 1932 and *Argas (Secretargas) transgariepinus* White, 1846. *Ticks Tick Borne Dis.* 9(4):889–895.

Punzo F. 1998. The biology of camel-spiders. Norwell (MA): Kluwer Academic Publishers.

Putnam NH, Butts T, Ferrier D, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu J-K, et al. 2008. The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453(7198):1064–1071.

Ramírez MJ, Magalhaes ILF, Derkarabetian S, Ledford J, Griswold CE, Wood HM, Hedin M. 2021. Sequence capture phylogenomics of true spiders reveals convergent evolution of respiratory systems. *Syst Biol.* 70(1):14–20.

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(7284):1079–1083.

Roelofs D, Zwaenepoel A, Sistermans T, Nap J, Kampfraath AA, Van de Peer Y, Ellers J, Kraaijeveld K. 2020. Multi-faceted analysis provides little evidence for recurrent whole-genome duplications during hexapod evolution. *BMC Biol*. 18(1):57.

Rokas A, Carroll SB. 2006. Bushes in the tree of life. *PLoS Biol*. 4(11):e352–e1904.

Rota-Stabelli O, Campbell L, Brinkmann H, Edgecombe GD, Longhorn SJ, Peterson KJ, Pisani D, Philippe H, Telford MJ. 2011. A congruent solution to arthropod phylogeny: phylogenomics, microRNAs and morphology support monophyletic Mandibulata. *Proc R Soc B*. 278(1703):298–306.

Rowland JM. 1972. The brooding habits and early development of *Trithyreus pentapeltis* (Cook) (Arachnida: Schizomida). *Entomol News*. 83:69–74.

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.

Santibáñez-López CE, Ojanguren-Affilastro AA, Sharma PP. 2020. Another one bites the dust: taxonomic sampling of a key genus in phylogenomic datasets reveals more non-monophyletic groups in traditional scorpion classification. *Invert Syst*. 34(2):133–143.

Santibáñez-López CE, Ontano AZ, Harvey MS, Sharma PP. 2018. Transcriptomic analysis of pseudoscorpion venom reveals a unique cocktail dominated by enzymes and protease inhibitors. *Toxins* 10(5):207.

Schwager EE, Schoppmeier M, Pechmann M, Damen WG. 2007. Duplicated Hox genes in the spider *Cupiennius salei*. *Front Zool*. 4(1):10.

Schwager EE, Sharma PP, Clarke T, Leite DJ, Wierschin T, Pechmann M, Akiyama-Oda Y, Esposito L, Bechsgaard J, Bilde T, et al. 2017. The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. *BMC Biol*. 15(1):62.

Setton EVW, Hendrixson BE, Sharma PP. 2019. Embryogenesis in a Colorado population of *Aphonopelma hentzi* (Girard, 1852) (Araneae: Mygalomorphae: Theraphosidae): establishing a promising system for the study of mygalomorph development. *J Arachnol*. 47(2):209–216.

Setton EVW, Sharma PP. 2018. Cooption of an appendage-patterning gene cassette in the head segmentation of arachnids. *Proc Natl Acad Sci USA*. 115(15):E3491–E3500.

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.

Sharma PP, Fernández R, Esposito LA, González-Santillán E, Monod L. 2015. Phylogenomic resolution of scorpions reveals multilevel discordance with morphological phylogenetic signal. *Proc R Soc B*. 282(1804):20142953.

Sharma PP, Kaluziak ST, Pérez-Porro AR, González VL, Hormiga G, Wheeler WC, Giribet G. 2014. Phylogenomic interrogation of Arachnida reveals systemic conflicts in phylogenetic signal. *Mol Biol Evol*. 31(11):2963–2984.

Sharma PP, Santiago MA, González-Santillán E, Monod L, Wheeler WC. 2015. Evidence of duplicated Hox genes in the most recent common ancestor of extant scorpions. *Evol Dev*. 17(6):347–355.

Sharma PP, Schwager EE, Extavour CG, Giribet G. 2012. Hox gene expression in the harvestman *Phalangium opilio* reveals divergent patterning of the chelicerate opisthosoma. *Evol Dev*. 14(5):450–463.

Sharma PP, Schwager EE, Extavour CG, Wheeler WC. 2014. Hox gene duplications correlate with posterior heteronomy in scorpions. *Proc R Soc B*. 281(1792):20140661.

Sharma PP, Tarazona OA, Lopez DH, Schwager EE, Cohn MJ, Wheeler WC, Extavour CG. 2015 A conserved genetic mechanism specifies deutocerebral appendage identity in insects and arachnids. *Proc R Soc B*. 282(1808):20150698–20150698.

Shen X-X, Hittinger CT, Rokas A. 2017. Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nat Ecol Evol*. 1:0126.

Shingate P, Ravi V, Prasad A, Tay B-H, Garg KM, Chattopadhyay B, Yap L-M, Rheindt FE, Venkatesh B. 2020. Chromosome-level assembly of the horseshoe crab genome provides insights into its genome evolution. *Nat Commun*. 11(1):2322.

Shultz JW, Regier JC. 2000. Phylogenetic analysis of arthropods using two nuclear protein–encoding genes supports a crustacean + hexapod clade. *Proc R Soc Lond B*. 267(1447):1011–1019.

Shultz JW. 1990. Evolutionary morphology and phylogeny of Arachnida. *Cladistics* 6(1):1–31.

Shultz JW. 2007. A phylogenetic analysis of the arachnid orders based on morphological characters. *Zool J Linn Soc*. 150(2):221–265.

Simakov O, Marlétaz F, Yue J-X, O'Connell B, Jenkins J, Brandt A, Calef R, Tung C-H, Huang T-K, Schmutz J, et al. 2020. Deeply conserved synteny resolves early events in vertebrate evolution. *Nat Ecol Evol*. 4(6):820–830.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 31(19):3210–3112.

Simion P, Philippe H, Baurain D, Jager M, Richter DJ, Di Franco A, Roure B, Satoh N, Quéinnec É, Ereskovsky A, et al. 2017. A large and consistent phylogenomic dataset supports sponges as the sister group to all other animals. *Curr Biol*. 27(7):958–967.

Tarver JE, Sperling EA, Nailor A, Heimberg AM, Robinson JM, King BL, Pisani D, Donoghue PCJ, Peterson PJ. 2013. miRNAs: small genes with big potential in metazoan phylogenetics. *Mol Biol Evol*. 30(11):2369–2382.

Tarver JE, Taylor RS, Puttick MN, Lloyd GT, Pett W, Fromm B, Schirrmeister BE, Pisani D, Peterson KJ, Donoghue PCJ. 2018. Well-annotated microRNAomes do not evidence pervasive miRNA loss. *Genome Biol Evol*. 10(6):1457–1470.

Thomson RC, Plachetzki DC, Mahler DL, Moore BR. 2014. A critical appraisal of the use of microRNA data in phylogenetics. *Proc Natl Acad Sci USA*. 111(35):E3659–E3668.

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(4):614–622.

Wang HC, Minh BQ, Susko S, Roger AJ. 2018. Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Syst Biol*. 67(2):216–235.

Wang H-C, Susko E, Roger AJ. 2019. The relative importance of modeling site pattern heterogeneity versus partition-wise heterotachy in phylogenomic inference. *Syst Biol*. 68(6):1003–1019.

Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol*. 35(3):543–548.

Wendruff AJ, Babcock LE, Wirkner CS, Kluessendorf J, Mikulic DG. 2020. A Silurian ancestral scorpion with fossilised internal anatomy illustrating a pathway to arachnid terrestrialisation. *Sci Rep*. 10(1):14.

Weygoldt P. 1969. The biology of pseudoscorpions. Cambridge (MA): Harvard University Press.

Whelan NV, Halanych KM. 2016. Who let the CAT out of the bag? Accurately dealing with substitutional heterogeneity in phylogenomic analyses. *Syst Biol*. 66(2):232–255.

Zdobnov EM, Tegenfeldt F, Kuznetsov D, Waterhouse RM, Simão FA, Ioannidis P, Seppey M, Loetscher A, Kriventseva EV. 2017. OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res*. 45(D1):D744–D749.

Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*. 19(S6):153.

Zhong Y, Holland PWH. 2011. HomeoDB2: functional expansion of a comparative homeobox gene database for evolutionary developmental biology. *Evol Dev*. 13(6):567–568.

Zimin AV, Marcai G, Puiu D, Roberts M, Salzberg SL, Yorke JA. 2013. The MaSuRCA genome assembler. *Bioinformatics* 29(21):2669–2677.