The Diagnosis and Resolution of Long Branch Taxa Among Chelicerata

By

Andrew Z. Ontano

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy
(Integrative Biology)

at the
University of Wisconsin-Madison
2021

Date of final oral examination: 08/18/2021

The dissertation is approved by the following members of the Final Oral Committee:
- Prashant P. Sharma, Assistant Professor, Integrative Biology
- Seth S. Blair, Professor, Integrative Biology
- Jeff D. Hardin, Professor, Integrative Biology
- Yevgenya Grinblat, Associate Professor, Integrative Biology
- Sean D. Schoville, Associate Professor, Entomology
- Mark S. Harvey Senior Curator of Arachnology, Western Australian Museum

Acknowledgements

I owe many people a great deal of gratitude for their help, guidance, and support during the completion of this PhD.

Firstly, thank you to my advisor Prashant Sharma for guidance these past five years. Your brilliance and excitement for EvoDevo is a contagious source for inspiration. I am forever indebted for your mentorship, and I can only hope to tell a story through my work a fraction as captivatingly as you.

Thank you to my committee: Jeff Hardin, Seth Blair, Yevgenya Grinblat, Sean Schoville, and Mark Harvey. Your insights and perspectives have been immeasurably helpful in broadening my perspectives towards this work. To Mark Harvey, I am forever thankful for your expertise in the field and for putting up with the International Date Line. Thank you also to the many UW-Madison colleagues and facilities that made this work possible.

Everyone in the Sharma Lab and those that have passed through have helped make this an amazing experience. Thank you to Emily Setton and Guilherme Gainett for your constant motivation, and for your support when the going was slow. I'll always be free if you need to do some field collecting. Thank you to Carlos Santibáñez López for your guidance and help with much of this work. I'll forever owe you a round of any boardgame you choose. Thanks to Jesús Ballesteros for your expertise, and for making sure I never settle for the easy answers. Thank you also to Caitlin, Erik, Audrey, Calvin, Logan, Kevin, Stephanie, Jakob, Hugh, Luciano, Benjamin, Pola, James, Keyton, Luis, Shlomi, and Nancy for helping make this lab a family even if only passing through for a short time. A special thanks to Jill for always being up for a bit of trivia.

I am grateful for many friends that have supported me in making this work possible. Thank you to Brett, Stephanie, Michelle, and Kiki who have kept me motivated and excited about science for nearly a decade and for our Biology Education Enrichment Coffee Hour. Thank you to everyone in November Project for being an endless source of positivity, energy, and high fives when that was needed most. Thank you to John, Logan, Blake, Mike, Eileen, and Chris for being so much more than the best housemates during my time in Madison.

Finally, thank you to Stephaney. You’ve been a constant pillar of support and my best friend during this endeavor. Thank you for putting up with listening to me rant about biology and when I point out every arthropod we see on hikes.

Table of Contents

Abstract

The Diagnosis and Resolution of Long Branch Taxa Among Chelicerata

Andrew Z. Ontano

Under the supervision of Prashant P. Sharma

At the University of Wisconsin – Madison

The arthropod subphylum Chelicerata underwent a history of early and rapid diversification that has proven difficult to resolve through phylogenomic analyses. Different rates of evolution among chelicerates drives this difficulty through long branch attraction, a systematic artifact whereby fast-evolving lineages are erroneously recovered clustered together towards the base of a phylogeny. To overcome issues created by long branch attraction, there is a present need for diagnosis of which lineages are susceptible to this error, and for the use of phylogenomic methods that alleviate these artefacts. This dissertation combines diagnosis and solutions to systematic errors for an intractable part of the tree of life.

In a combined work, I employed broad taxonomic sampling as a solution to the pseudoscorpion long branch attraction artifact (Chapter 1) and validated their placement using evidence of an ancient whole genome duplication through surveys of developmental patterning genes and microRNAs (Chapter 2). By generating and analyzing the first pseudoscorpion developmental transcriptome (as well as the first draft genome) I showed that pseudoscorpions underwent a whole genome duplication event that supports the clade Panscorpiones uniting pseudoscorpions and scorpions. Pseudoscorpions are nested within the book lung-bearing arachnids, suggesting the loss of these iconic respiratory structures as a result of miniaturization. I then diagnosed which chelicerate lineages are prone to long branch attraction through subsampling of lineages (Chapter 3). I showed that Palpigradi are prone to long branch attraction errors driving the instability of other fast-evolving lineages. Improved taxonomic sampling and use of complex evolutionary models refutes the monophyly of Acari. This research offers approaches for phylogenomic resolution to problematic areas of the tree of life where ancient radiations mask relationships and provides genomic and transcriptomic tools for validating the relationships between diverse taxa.

Introduction

The arthropod subphylum Chelicerata is a diverse clade of >120,000 described species (Sharma, 2018) found across a vast range of habitats. Named for the unique pair of anterior-most appendages, the chelicerae, chelicerates possess a wealth of body plan variations and morphological innovations that have facilitated their diversification. However, the evolution of this group is cast in shadow by our poor understanding of the basal phylogeny underlying the chelicerates. Traditionally, the group has been represented by three extant clades: the marine lineages Pycnogonida (sea spiders) and Xiphosura (horseshoe crabs), and the terrestrial lineage Arachnida (mites, ticks, spiders, scorpions, etc.). A sister group relationship between Pycnogonida and Euchelicerata (Xiphosura + Arachnida) has long been supported by phylogenetic analyses incorporating both morphological and molecular data (Legg et al., 2013; Wheeler & Hayashi, 1998) and genome-scale datasets (Ballesteros & Sharma, 2019; Borner et al., 2014; Regier et al., 2010; Sharma et al., 2014), but the monophyly of Arachnida and the hypothesis for a single terrestrialization event is frequently complicated by the nested placement of Xiphosura within the arachnids (Ballesteros & Sharma, 2019; Borner et al., 2014; Sharma et al., 2014).

Difficulties in resolving the phylogeny is not a problem unique to chelicerates. The basal relationships among Metazoa (animals) is plagued with similar issues of sparsely-represented lineages, deep radiations, and long branch taxa (Simion et al., 2017). The inclusion of increasing amounts of genomic and transcriptomic data in analyses has not ultimately solved the issue, where incorporation of poorly-fitting evolutionary models further reinforces errors like long branch attraction artefacts (Laumer et al., 2019; Simion et al., 2017). Overcoming phylogenetic inconsistencies and artefacts within chelicerates may serve as a stepping stone for resolving the broader tree of life and other neglected or poorly represented lineages.

The extant chelicerate lineages have undergone a range of evolutionary phenomena that complicate phylogenetic analyses. Chelicerates arose in the fossil record as far back as the Cambrian and subsequently underwent rapid diversification by the Silurian-Devonian boundary (Jago et al., 2016; Legg, 2014). Despite an extant representation by only four taxa from one derived family (Lamsdell, 2020), Xiphosura arose by the Late Ordovician (Rudkin et al., 2008). Arachnopulmonata, the arachnids bearing book lungs, diversified by the Early Silurian for primitive scorpions (Wendruff et al., 2020) and by the Early Devonian for Tetrapulmonata (Araneae, Amblypygi, and Uropygi) (Wang et al., 2018). Clades undergoing ancient, rapid diversification events appear as "bushes" in the tree of life, where the accumulation of homoplastic characters following diversification masks the true relationships among lineages (Rokas & Carroll, 2006). Another issue is apparent in the relative rates at which orders evolve. Lineages including Acariformes, Parasitiformes (ticks), and Pseudoscorpiones (pseudoscorpions) exhibit higher evolutionary rates than other arachnid orders (Ballesteros & Sharma, 2019; Dabert et al., 2010; Lozano-Fernandez et al., 2019; Pepato et al., 2010; Sharma et al., 2014), which makes these groups susceptible to a systematic artifact called long branch attraction (Bergsten, 2005). Understanding the relationships of lineages among the chelicerates and the downstream inference of the history of terrestrialization is hindered by the lack of a resolved basal phylogeny.

Phylogenetic frameworks incorporating increased taxonomic sampling, selection of slower-evolving genes, and utilization of computationally intensive evolutionary models have been employed for resolving chelicerate phylogeny with mixed results. The monophyly of Tetrapulmonata, and the sister group relationship of Araneae (spiders) and Pedipalpi

(Amblypygi, Schizomida and Thelyphonida) within the tetrapulmonates, are robustly supported nodes across phylogenomic analyses (Ballesteros et al., 2019; Ballesteros & Sharma, 2019; Borner et al., 2014; Howard et al., 2020; Lozano-Fernandez et al., 2019), in agreement with morphology (Shultz, 1990). Beyond Tetrapulmonata, the ordinal relationships are more problematic, save for the placement of scorpions. Scorpions had previously been proposed as a transitional state between the extinct Eurypterida (sea scorpions) and terrestrial arachnids (Weygoldt & Paulus, 1979), or as the sister group of Opiliones (harvestmen), with whom they share a preoral chamber formed by protuberances on the coxae of both pedipalps and first pair of walking legs (the stomotheca; Shultz, 2007). Phylogenomic studies recover scorpions as the sister group of Tetrapulmonata, consistent with the presence of book lungs and a whole genome duplication in their common ancestor (Schwager et al., 2017). Few other interordinal relationships have been successfully resolved.

Relationships of the long-branch orders have been especially elusive. Pseudoscorpiones have traditionally been viewed as the sister group to Solifugae (camel spiders), based on morphological characters including the two-segmented chelicerae and the elongated patellae resembling the femora (Shultz, 2007), but this relationship is never recovered in phylogenomic analyses. Phylogenomic analyses will instead recover pseudoscorpions as sister group to (1) Acariformes; (2) in a placement towards the base of the chelicerates (Ballesteros & Sharma, 2019; Fernández & Giribet, 2015; Sharma et al., 2014); or (3) as the sister group to Parasitiformes (Regier et al., 2010). Intriguingly, a placement of Pseudoscorpiones as the sister group to Scorpiones has been recovered by analyses incorporating a subsampling of slowly-evolving genes (Benavides et al., 2019; Howard et al., 2020; Sharma et al., 2014). Pseudoscorpion representation in those studies had previously been limited to taxa representing only a derived lineage (i.e., omitting a broad sampling of superfamilies that spans the common ancestor of extant pseudoscorpions). More recently, a phylogenomic study of pseudoscorpion internal phylogeny was able to include all major lineages (superfamilies) that advance understanding of the relationships within this order (Benavides et al., 2019), but these taxa have not been broadly represented in higher-level chelicerate phylogeny. Similarly, Acariformes and Parasitiformes have previously been suggested to be sister taxa (as the clade Acari) based upon anatomical structures, though an abundance of alternative relationships has been proposed (Dunlop & Alberti, 2008). Yet only a subset of phylogenomic analyses that have utilized slowly-evolving genes and computationally intensive evolutionary models recover Acari as monophyletic (Howard et al., 2020; Lozano-Fernandez et al., 2019). More commonly these orders are recovered as a grade towards the base of chelicerates in an affiliation of long branch orders. The inconsistent placements of the acarine orders may be driven by taxonomic sampling, wherein Opilioacariformes have been suggested as sister to the remaining Parasitiformes or to all Acari based on a presumptive plesiomorphic acarine body plan (Pepato et al., 2010; Shultz, 2007). A further issue lies in the taxonomic representation of several lineages. The small enigmatic order Palpigradi (microwhip scorpions) has previously been represented in one phylogenomic analyses by a single species; in that study, Palpigradi is recovered as sister group to Solifugae (with matrices optimized for high taxon occupancy) or as sister group to Parasitiformes (with matrices constructed for relaxed taxon occupancy; Ballesteros et al., 2019). Yet this single existing palpigrade data set exhibited low quality and presented its own challenges of high missing data. Such inconsistencies across phylogenomic analyses and discordance with morphological evidence make paramount the need for alternative approaches.

This dissertation ameliorates chelicerate phylogenetics through a focus on diagnosing problematic long branch taxa and assessing their placement. First, I employed broadened taxonomic sampling of Pseudoscorpiones to test their placement among chelicerates (Chapter 1), focusing on the effect of long branch attraction through sequential taxonomic subsampling. Next, I leveraged a whole-genome duplication event ancestral to Arachnopulmonata as a molecular character that unites Pseudoscorpiones and Arachnopulmonata (Chapter 2), investigating duplications of a developmentally relevant gene family and microRNAs. I defined Panscorpiones as the clade uniting pseudoscorpions and scorpions, and redefined Arachnopulmonata to include pseudoscorpions. My first two thesis chapters were jointly published as a single work this year in the journal *Molecular Biology and Evolution*. Finally, I diagnose how many long branch orders exist among Chelicerata (Chapter 3), focusing on improved genomic representation of the enigmatic order Palpigradi, and incorporating both phylogenomic subsampling and taxon deletion experiments to elucidate the effects of long branch attraction by problematic taxa. Complementary to the chapters outlined here are two appendices of this thesis, which constitute manuscripts to which I contributed as a junior author. First, we investigated the molecular components of a venomous pseudoscorpion and compared them to components from other venomous arachnids (Appendix 1). Then, we refuted the monophyly of Arachnida using comprehensive species sampling and a total evidence approach (Appendix 2).

References

Ballesteros, J. A., López, C. E. S., Kováč, L., Gavish-Regev, E., & Sharma, P. P. (2019). Ordered phylogenomic subsampling enables diagnosis of systematic errors in the placement of the enigmatic arachnid order Palpigradi. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1917). https://doi.org/10.1098/rspb.2019.2426

Ballesteros, J. A., & Sharma, P. P. (2019). A Critical Appraisal of the Placement of Xiphosura (Chelicerata) with Account of Known Sources of Phylogenetic Error. *Systematic Biology*, *0*(0), 1–14. https://doi.org/10.1093/sysbio/syz011

Benavides, L. R., Cosgrove, J. G., Harvey, M. S., & Giribet, G. (2019). Phylogenomic interrogation resolves the backbone of the Pseudoscorpiones tree of life. *Molecular Phylogenetics and Evolution*, *139*, 106509. https://doi.org/10.1016/j.ympev.2019.05.023

Bergsten, J. (2005). A review of long-branch attraction Introduction to long-branch attraction. *Cladistics*, *21*, 163–193. http://materiais.dbio.uevora.pt/Evo/A_review_of_long-branch_attraction.pdf

Borner, J., Rehm, P., Schill, R. O., Ebersberger, I., & Burmester, T. (2014). A transcriptome approach to ecdysozoan phylogeny. *Molecular Phylogenetics and Evolution*, *80*(1), 79–87. https://doi.org/10.1016/j.ympev.2014.08.001

Dabert, M., Witalinski, W., Kazmierski, A., Olszanowski, Z., & Dabert, J. (2010). Molecular phylogeny of acariform mites (Acari, Arachnida): Strong conflict between phylogenetic signal and long-branch attraction artifacts. *Molecular Phylogenetics and Evolution*, *56*(1), 222–241. https://doi.org/10.1016/j.ympev.2009.12.020

Dunlop, J. A., & Alberti, G. (2008). The affinities of mites and ticks: A review. *Journal of Zoological Systematics and Evolutionary Research*, *46*(1), 1–18. https://doi.org/10.1111/j.1439-0469.2007.00429.x

Fernández, R., & Giribet, G. (2015). Unnoticed in the tropics: Phylogenomic resolution of the poorly known arachnid order Ricinulei (Arachnida). *Royal Society Open Science*, *2*(6). https://doi.org/10.1098/rsos.150065

Howard, R. J., Puttick, M. N., Edgecombe, G. D., & Lozano-Fernandez, J. (2020). Arachnid monophyly: Morphological, palaeontological and molecular support for a single terrestrialization within Chelicerata. *Arthropod Structure and Development*, *59*, 100997. https://doi.org/10.1016/j.asd.2020.100997

Jago, J. B., García-Bellido, D. C., & Gehling, J. G. (2016). An early Cambrian chelicerate from the Emu Bay Shale, South Australia. *Palaeontology*, *59*(4), 549–562. https://doi.org/10.1111/pala.12243

Lamsdell, J. C. (2020). The phylogeny and systematics of Xiphosura. *PeerJ*, *8*. https://doi.org/10.7717/peerj.10431

Laumer, C. E., Fernández, R., Lemer, S., Combosch, D., Kocot, K. M., Riesgo, A., Andrade, S. C. S., Sterrer, W., Sørensen, M. V., Giribet, G., Laumer, C. E., Ferna, R., Andrade, C. S., Combosch, D., Kocot, K. M., Riesgo, A., Sterrer, W., Sørensen, M. V., & Giribet, G. (2019). Revisiting metazoan phylogeny with genomic sampling of all phyla. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1906). https://doi.org/10.1098/rspb.2019.0831

Legg, D. A. (2014). Sanctacaris uncata: the oldest chelicerate (Arthropoda). *Naturwissenschaften*, *101*(12), 1065–1073. https://doi.org/10.1007/s00114-014-1245-4

Legg, D. A., Sutton, M. D., & Edgecombe, G. D. (2013). Arthropod fossil data increase congruence of morphological and molecular phylogenies. *Nature Communications*, *4*, 1–7. https://doi.org/10.1038/ncomms3485

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

Pepato, A. R., da Rocha, C. E. F., & Dunlop, J. A. (2010). Phylogenetic position of the acariform mites: sensitivity to homology assessment under total evidence. *BMC Evolutionary Biology*, *10*, 235. https://doi.org/10.1186/1471-2148-10-235

Regier, J. C., Shultz, J. W., Zwick, A., Hussey, A., Ball, B., Wetzer, R., Martin, J. W., & Cunningham, C. W. (2010). Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature*, *463*(7284), 1079–1083. https://doi.org/10.1038/nature08742

Rokas, A., & Carroll, S. B. (2006). Bushes in the tree of life. *PLoS Biology*, *4*(11), 1899–1904. https://doi.org/10.1371/journal.pbio.0040352

Rudkin, D. M., Young, G. A., & Nowlan, G. S. (2008). The oldest horseshoe crab: A new xiphosurid from late ordovician konservat-lagerstätten deposits, Manitoba, Canada. *Palaeontology*, *51*(1), 1–9. https://doi.org/10.1111/j.1475-4983.2007.00746.x

Schwager, E. E., Sharma, P. P., Clarke, T., Leite, D. J., Wierschin, T., Pechmann, M., Akiyama-Oda, Y., Esposito, L., Bechsgaard, J., Bilde, T., Buffry, A. D., Chao, H., Dinh, H., Doddapaneni, H. V., Dugan, S., Eibner, C., Extavour, C. G., Funch, P., Garb, J., … McGregor, A. P. (2017). The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. *BMC Biology*, *15*(1), 1–27. https://doi.org/10.1186/s12915-017-0399-x

Sharma, P. P. (2018). Chelicerates. *Current Biology*, *28*(14), R774–R778. https://doi.org/10.1016/j.cub.2018.05.036

Sharma, P. P., Kaluziak, S. T., Pérez-Porro, A. R., González, V. L., Hormiga, G., Wheeler, W. C., & Giribet, G. (2014). Phylogenomic interrogation of arachnida reveals systemic conflicts in phylogenetic signal. *Molecular Biology and Evolution*, *31*(11), 2963–2984. https://doi.org/10.1093/molbev/msu235

Shultz, J. W. (1990). Evolutionary morphology and phylogeny of Arachnida. *Cladistics*, *6*, 1–38.

Shultz, J. W. (2007). A phylogenetic analysis of the arachnid orders based on morphological characters. In *Zoological Journal of the Linnean Society* (Vol. 150, Issue 2). https://doi.org/10.1111/j.1096-3642.2007.00284.x

Simion, P., Philippe, H., Baurain, D., Jager, M., Richter, D. J., Di Franco, A., Roure, B., Satoh, N., Quéinnec, É., Ereskovsky, A., Lapébie, P., Corre, E., Delsuc, F., King, N., Wörheide, G., & Manuel, M. (2017). A Large and Consistent Phylogenomic Dataset Supports Sponges as the Sister Group to All Other Animals. *Current Biology*, 1–10. https://doi.org/10.1016/j.cub.2017.02.031

Wang, B., Dunlop, J. A., Selden, P. A., Garwood, R. J., Shear, W. A., Müller, P., & Lei, X. (2018). Cretaceous arachnid Chimerarachne yingi gen. et sp. nov. illuminates spider origins. *Nature Ecology & Evolution*. https://doi.org/10.1038/s41559-017-0449-3

Wendruff, A. J., Babcock, L. E., Wirkner, C. S., Kluessendorf, J., & Mikulic, D. G. (2020). A Silurian ancestral scorpion with fossilised internal anatomy illustrating a pathway to arachnid terrestrialisation. *Scientific Reports*, *10*(1), 1–6. https://doi.org/10.1038/s41598-019-56010-z

Weygoldt, P., & Paulus, H. F. (1979). Untersuchungen zur Morphologie, Taxonomie und Phylogenie der Chelicerata II. Cladogramme und die Entfaltung der Chelicerata. *Journal of Zoological Systematics and Evolutionary Research*, *17*(3), 177–200. https://doi.org/10.1111/j.1439-0469.1979.tb00699.x

Wheeler, W. C., & Hayashi, C. Y. (1998). The Phylogeny of the Extant Chelicerate Orders. *Cladistics*, *14*(2), 173–192. https://doi.org/10.1111/j.1096-0031.1998.tb00331.x

Chapters 1 & 2

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

**Andrew Z Ontano, Guilherme Gainett, Shlomi Aharon, Jesús A Ballesteros, Ligia R Benavides, Kevin F Corbett, Efrat Gavish-Regev, Mark S Harvey, Scott Monsma, Carlos E Santibáñez-López, Emily V W Setton, Jakob T Zehms, Jeanne A Zeh, David W Zeh, Prashant P Sharma*

Published June 2021 in Molecular Biology and Evolution

Volume 38, Issue 6, Pages 2446–2467; doi:10.1093/molbev/msab038

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

---

*Contribution: I conceived the study, cultivated pseudoscorpion embryos, performed sequencing, implemented phylogenomic analyses, analyzed homeobox duplications, and wrote the manuscript.

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.

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 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-

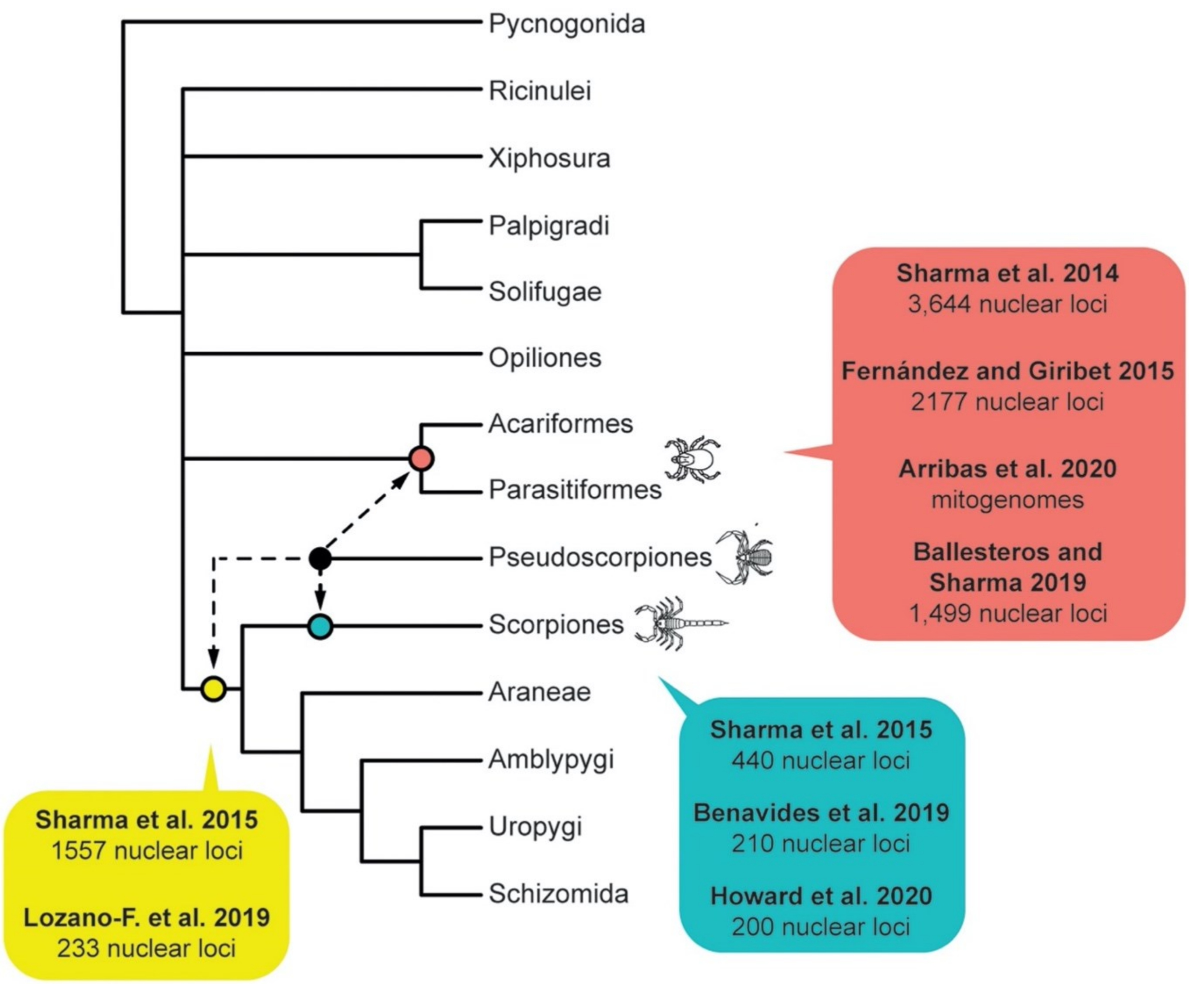


Figure 1 Summary tree topology of Chelicerata showing relationships of orders. Phylogeny based on Ballesteros et al. (2019). Dotted lines for pseudoscorpiones show alternative placements of this order in selected historical phylogenetic analyses.

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.

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.

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.

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

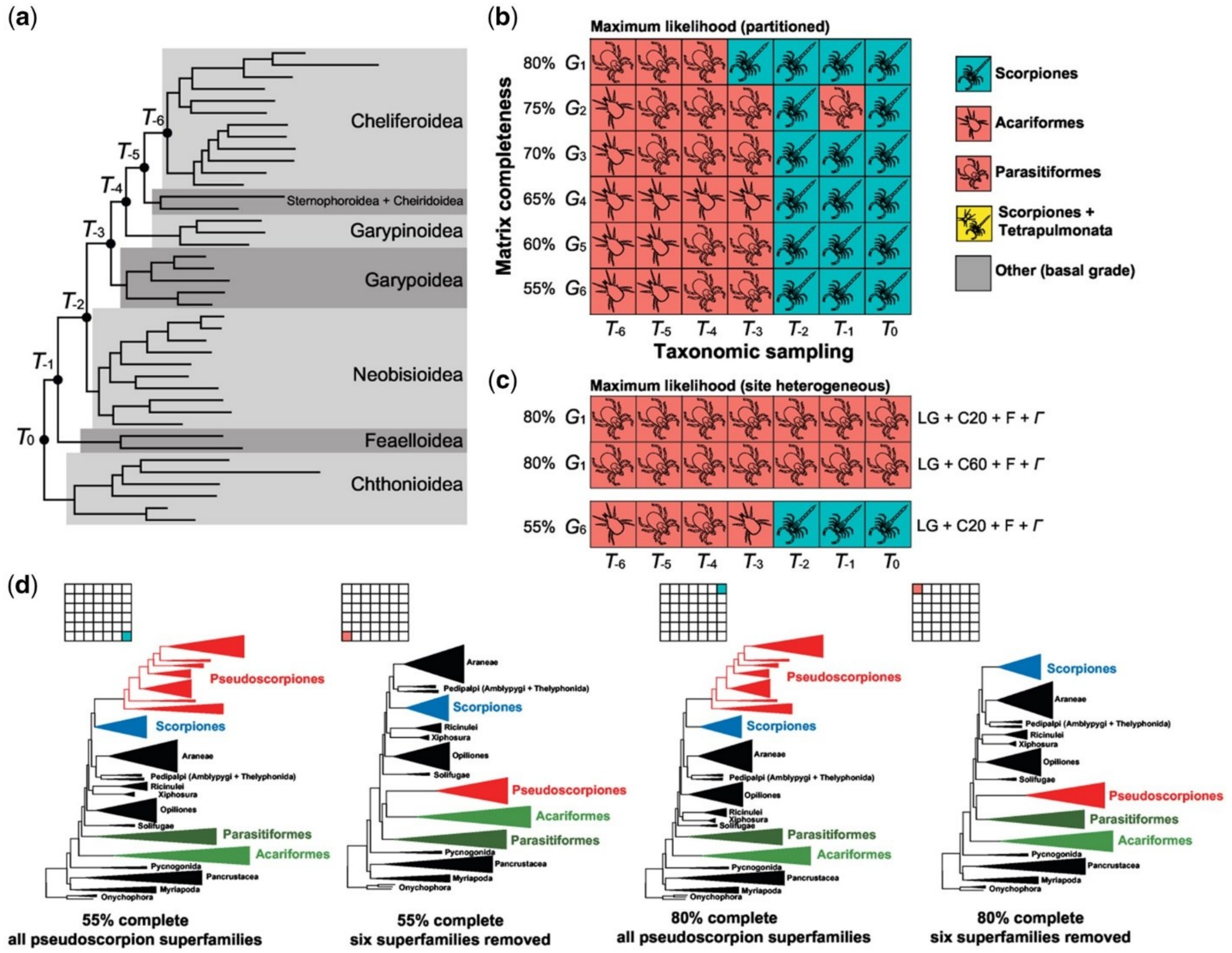


Figure 2 Depth of taxonomic sampling is more influential than matrix completeness in supermatrix analyses of pseudoscorpion placement. (a) Internal phylogeny of Pseudoscorpiones showing major taxonomic groups. Notations on nodes indicate taxon subsets obtained by sequential pruning of branches. (b) Sensitivity 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.

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).

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.

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 Pseudoscorpiones + Scorpiones, excepting matrix $G_2 \bullet T_{-1}$, 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 $G_1 \bullet T_{-3}$ (fig. 2b). 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. 2d). 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).

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.

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

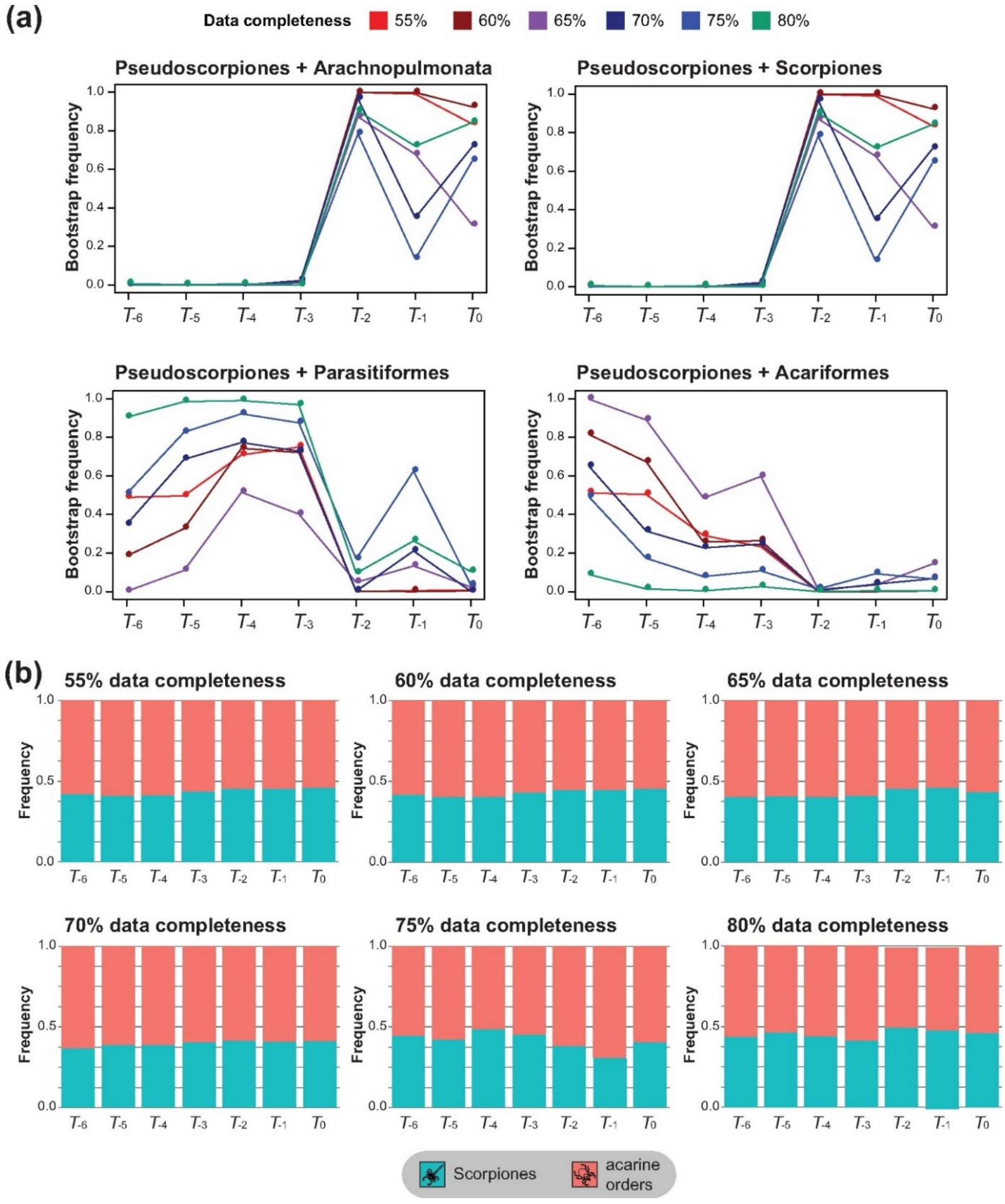


Figure 3 Depth of taxonomic sampling affects supermatrix nodal support, but not per-locus support. (a) Nodal support frequency for competing hypotheses of pseudoscorpion placement as a function of taxonomic sampling and matrix completeness. (b) Proportion of loci favoring Pseudoscorpiones + Scorpiones versus Pseudoscorpiones + either acarine order under a ΔGLS framework, as a function of taxonomic sampling and matrix completeness.

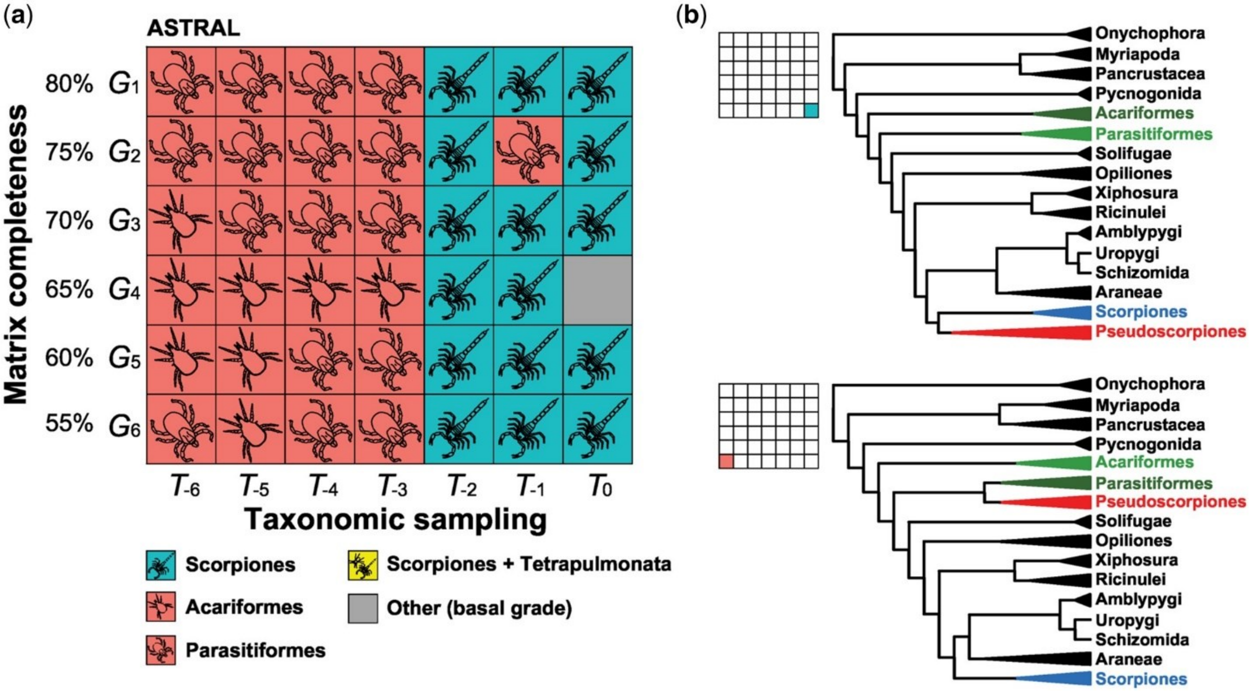


Figure 4 Depth of taxonomic sampling is more influential than matrix completeness in ASTRAL analyses of pseudoscorpion placement. (a) Sensitivity plot of 42 phylogenomic matrices assembled by varying gene occupancy (y axis) and taxonomic sampling (x axis). (b) Selected tree topologies showing the dynamics of pseudoscorpion instability as a function of taxonomic sampling.

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 $G_2 \bullet T_{-1}$ and $G_4 \bullet T_0$, which recovered pseudoscorpions as the sister group of Parasitiformes or in a grade at the base of Chelicerata, respectively.

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.

ASTRAL analyses never recovered the monophyly of Arachnida or Acari.

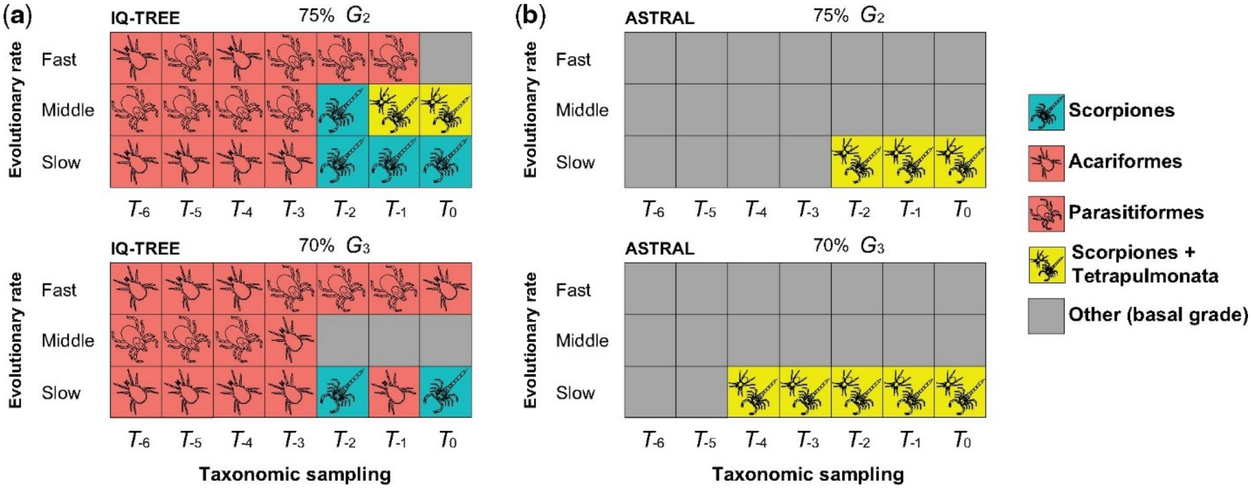


Figure 5 Slow-evolving genes do not consistently recover Pseudoscorpiones + Scorpiones in the absence of dense taxon sampling. (a) Sensitivity plot of 21 phylogenomic matrices assembled by binning genes into tertiles of evolutionary rate (y axis) and taxonomic sampling (x axis), with partitioned model-fitting. 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.

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 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.

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.

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 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).

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.

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).

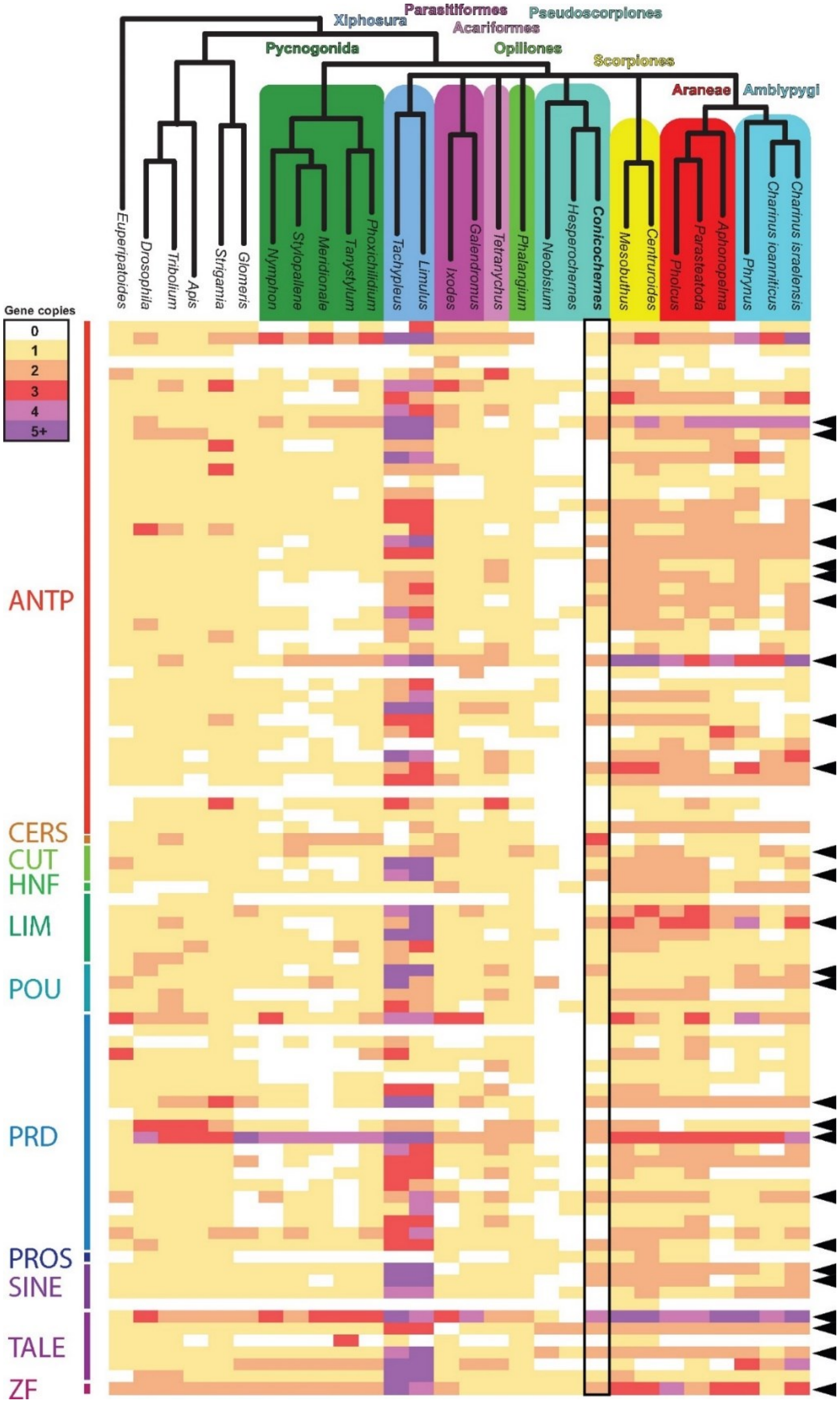


Figure 6 Comparison of homeobox repertoires for 26 panarthropods supports retention of duplications in pseudoscorpions that are shared with arachnopulmonates. Rows correspond to individual homeobox genes. Colors correspond to numbers of paralogs. Black arrows to the right indicate duplications in at least one pseudoscorpion exemplar that is also shared by at least one arachnopulmonate.

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).

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

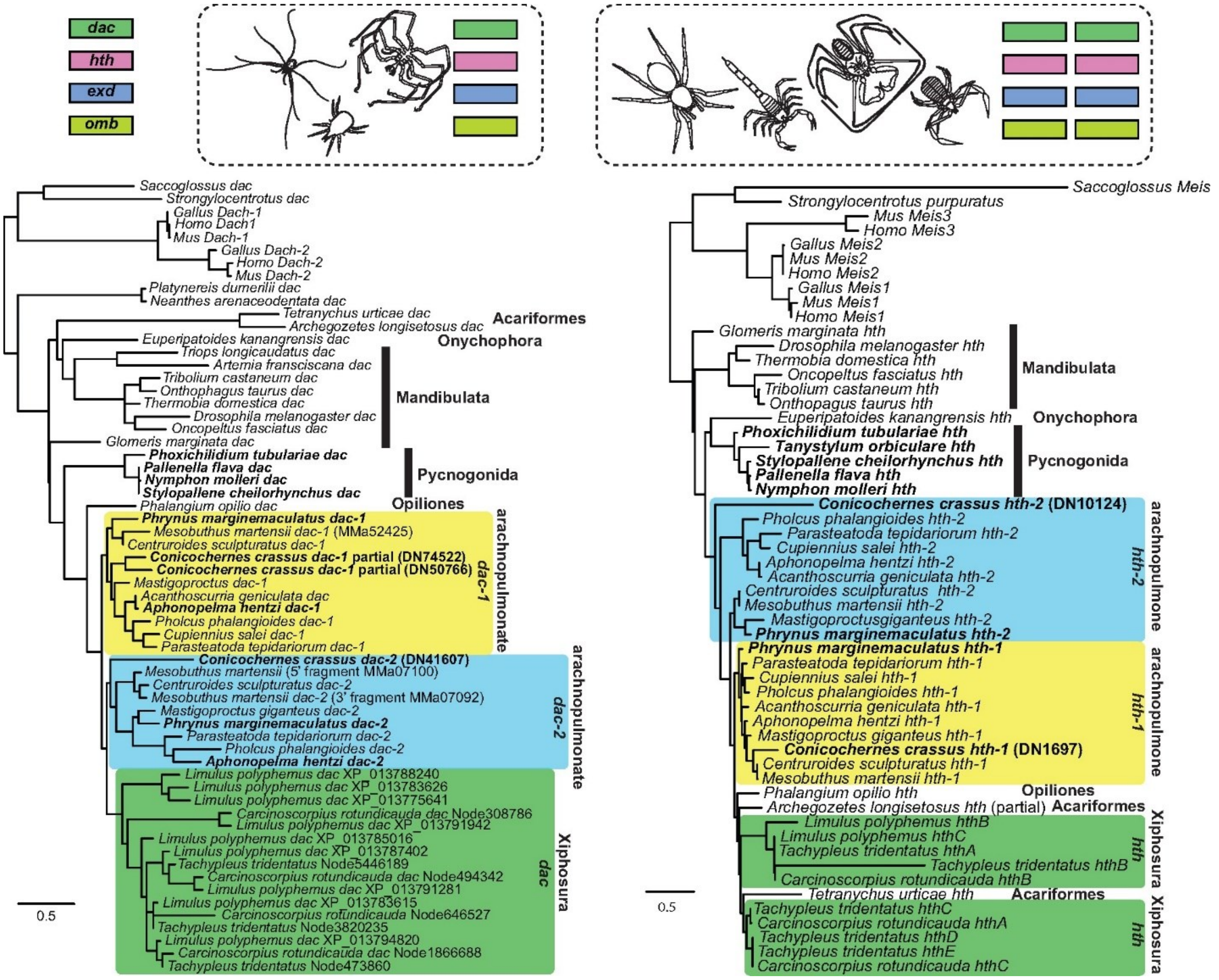


Figure 7 Pseudoscorpions possess two copies of four developmental patterning genes known to exhibit paralogs with arachnopulmonate-specific spatiotemporal expression domains. Above: Single copy orthologs of dac, hth, exd, and omb were recovered from genomic resources for sea spiders, harvestmen, and 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 nonoverlapping fragments.

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.

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 nonoverlapping fragments.

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 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.

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).

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.

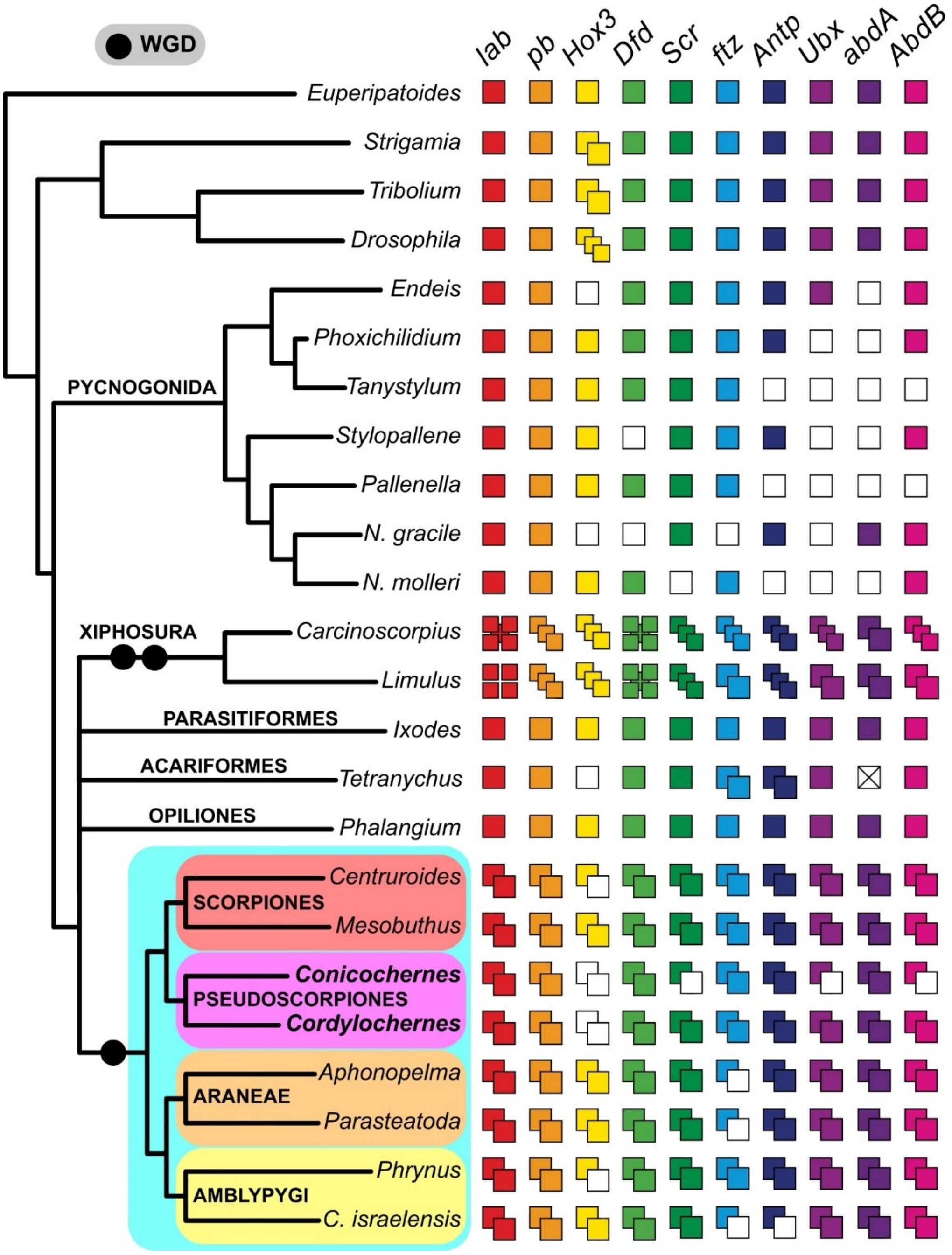


Figure 8 Hox gene complement in the pseudoscorpion genome substantiates evidence of shared WGD with other Arachnopulmonata. Columns and colored squares correspond to each Hox gene. Unfilled squares correspond to absences, not losses. Cross through abdA in T. urticae indicates loss of this Hox gene in the mite genome. Note independent 2-fold WGD events in Xiphosura (Kenny et al. 2016; Shingate et al. 2020).

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

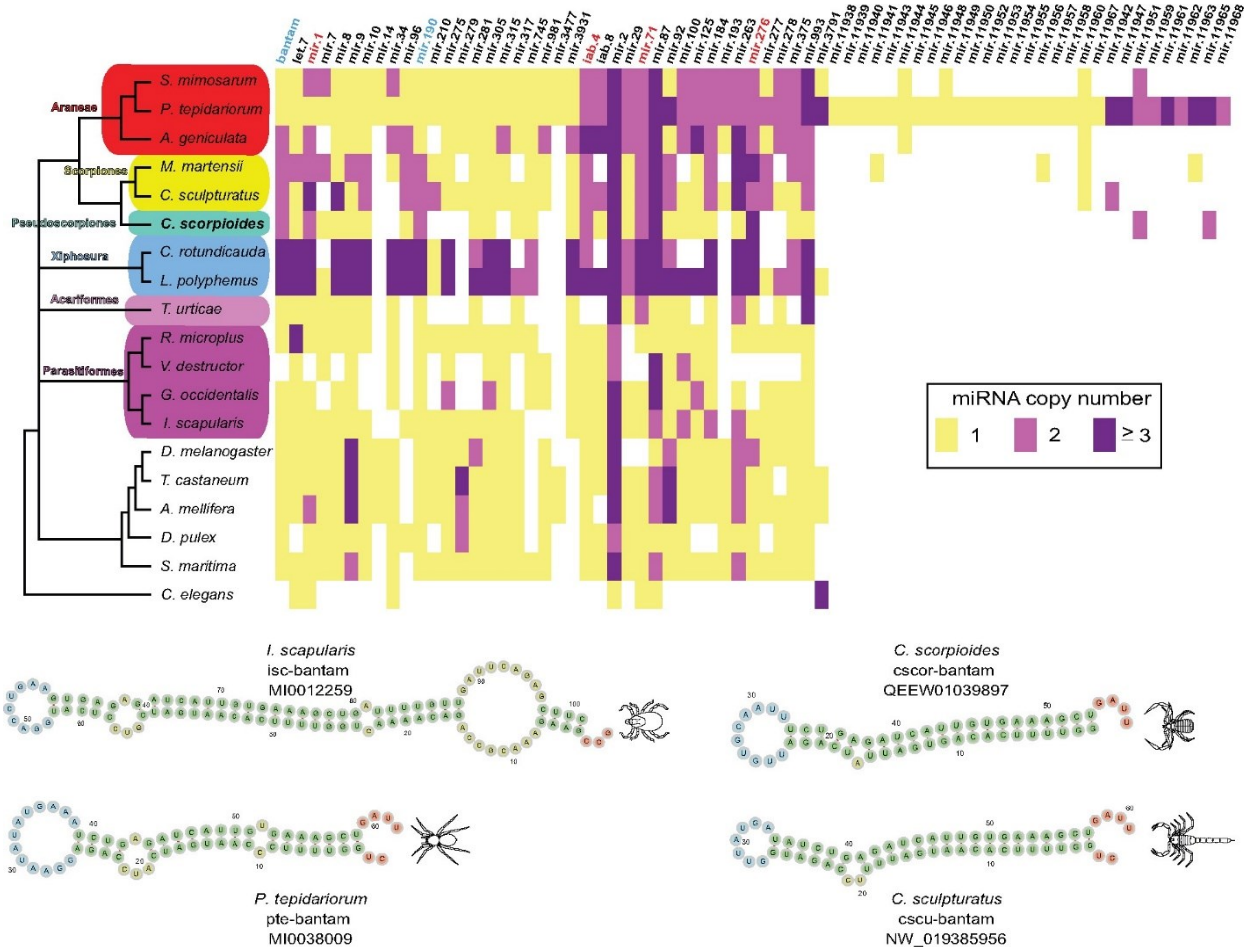


Figure 9 Comparison of miRNA family copy number in C. scorpioides and other ecdysozoans supports retention of duplications in pseudoscorpions shared with scorpions and spiders. Columns correspond to individual miRNA families. Colors correspond to numbers of paralogs. miRNA families in red text indicate duplications in scorpions, pseudoscorpions and at least one spider. miRNA families in blue text are duplicated only in scorpions and pseudoscorpions.

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; 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 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

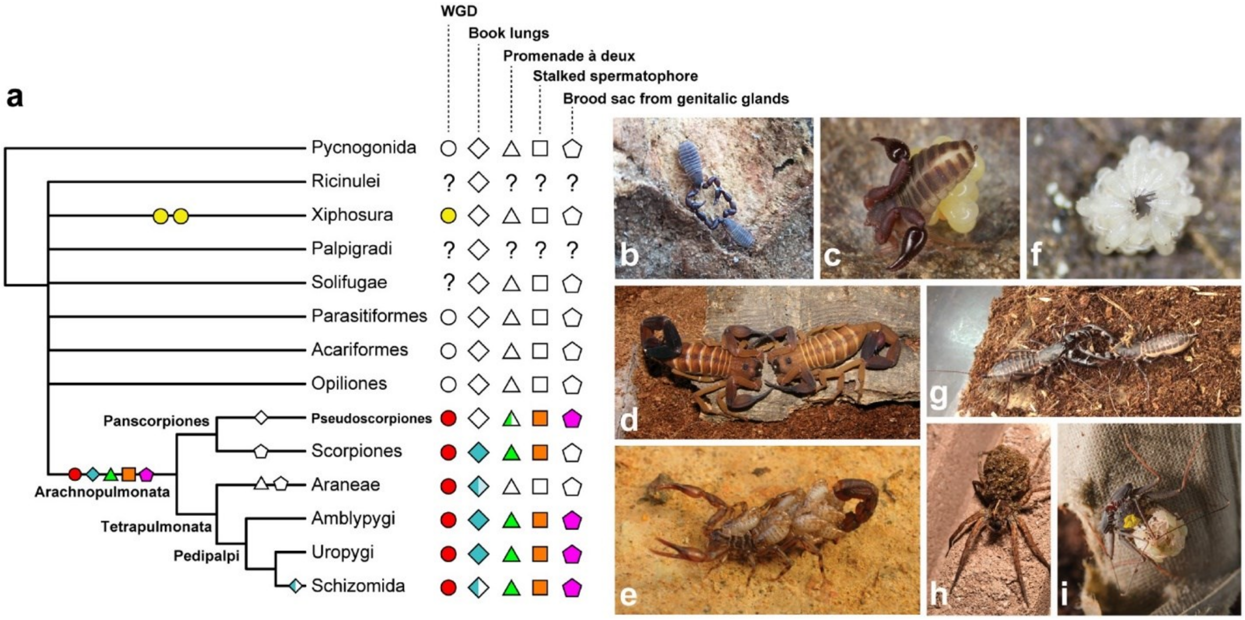


Figure 10 Ancestral state reconstruction of shared genome duplication events and reproductive behaviors in Chelicerata, under accelerated transformation. (a) The revised placement of pseudoscorpions supports a shared origin of courtship behavior and maternal brood care across Arachnopulmonata. Empty 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).

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 strandspecific 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 + Γ 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

Specimens of *C. crassus* were collected in Western Australia under permit number 08–000214–6 from the Department of Parks and Wildlife. Specimens of *C. scorpioides* were collected in Panamá under permits SE/A-92-05 (collecting) and SEX/A-142-05 (export), from the Autoridad Nacional del Ambiente, República de Panamá; and permit number 68818 (quarantine) from the Ministerio de Desarrollo Agropecuario, República de Panamá.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online under URL https://academic.oup.com/mbe/article/38/6/2446/6132263#248368421

Acknowledgments

We are indebted to Leonardo Carvalho, Adele Hochberg, Rick Hochberg, and Gonzalo Giribet for contributing photographs of Schizomida, Uropygi, and Pseudoscorpiones. Sequencing of *C. crassus* was performed at BioTechnology Center at UW-Madison. Access to computing nodes for intensive tasks was provided by the Center for High Throughput Computing (CHTC) and the Bioinformatics Resource Center (BRC) of the University of Wisconsin–Madison. We thank La Autoridad Nacional del Ambiente for permission to collect *C. scorpioides* in the República de Panamá and the Smithsonian Tropical Research Institute for logistical support. Comments from two anonymous reviewers improved a previous draft of the manuscript. Funding for the *C. scorpioides* genome sequencing project was provided by the National Science Foundation (Grant No. IOS-1656670) to J.A.Z and D.W.Z. Support for the generation of developmental genetic resources for *C. crassus* and other chelicerate species was supported by the National Science Foundation (Grant Nos. IOS-1552610 and IOS-2016141) to P.P.S.

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.A.B., 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, A. M., Gil, M., Gonnet, G. H., & Dessimoz, C. (2013). Inferring hierarchical orthologous groups from orthologous gene pairs. *PloS One*, *8*(1), e53786.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology*, *215*(3), 403–410.

Aria, C., & Caron, J.-B. (2019). A middle Cambrian arthropod with chelicerae and proto-book gills. *Nature*, *573*(7775), 586–589.

Arribas, P., Linard, B., Emerson, B. C., & Vogler, A. P. (2019). Mitochondrial metagenomics reveals the ancient origin and phylodiversity of soil mites and provides a phylogeny of the Acari. *Molecular Biology and Evolution*, 1–31.

Ballesteros, J. A., & Hormiga, G. (2016). A New Orthology Assessment Method for Phylogenomic Data: Unrooted Phylogenetic Orthology. *Molecular Biology and Evolution*, *33*(8), 2117–2134. https://doi.org/10.1093/molbev/msw069

Ballesteros, J. A., López, C. E. S., Kováč, L., Gavish-Regev, E., & Sharma, P. P. (2019). Ordered phylogenomic subsampling enables diagnosis of systematic errors in the placement of the enigmatic arachnid order Palpigradi. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1917). https://doi.org/10.1098/rspb.2019.2426

Ballesteros, J. A., Setton, E. V. W., Santibáñez-López, C. E., Arango, C. P., Brenneis, G., Brix, S., Corbett, K. F., Cano-Sánchez, E., Dandouch, M., Dilly, G. F., Eleaume, M. P., Gainett, G., Gallut, C., McAtee, S., McIntyre, L., Moran, A. L., Moran, R., López-González, P. J., Scholtz, G., … Sharma, P. P. (2021). Phylogenomic resolution of sea spider diversification through integration of multiple data classes. *Molecular Biology and Evolution*, *38*(2), 686–701. https://doi.org/10.1093/molbev/msaa228

Ballesteros, J. A., & Sharma, P. P. (2019). A Critical Appraisal of the Placement of Xiphosura (Chelicerata) with Account of Known Sources of Phylogenetic Error. *Systematic Biology*, *68*(6), 896–917. https://doi.org/10.1093/sysbio/syz011

Barnett, A. A., & Thomas, R. H. (2013). The expression of limb gap genes in the mite Archegozetes longisetosus reveals differential patterning mechanisms in chelicerates. *Evolution & Development*, *15*(4), 280–292.

Benavides, L. R., Cosgrove, J. G., Harvey, M. S., & Giribet, G. (2019). Phylogenomic interrogation resolves the backbone of the Pseudoscorpiones tree of life. *Molecular Phylogenetics and Evolution*, *139*, 106509. https://doi.org/10.1016/j.ympev.2019.05.023

Bicknell, R. D. C., 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, D. C., Conley, K. W., Plachetzki, D. C., Kempler, K., Battelle, B., & Brown, N. L. (2008). Isolation and Expression of Pax6 and atonal Homologues in the American Horseshoe Crab , Limulus polyphemus. *Developmental Dynamics*, *237*(July), 2209–2219. https://doi.org/10.1002/dvdy.21634

Cao, Z., Yu, Y., Wu, Y., Hao, P., Di, Z., He, Y., Chen, Z., Yang, W., Shen, Z., & He, X. (2013). *The genome of Mesobuthus martensii reveals a unique adaptation model of arthropods. Nat Commun 4*. Nature Publishing Group.

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Charrier, N. P., Hermouet, A., Hervet, C., Agoulon, A., Barker, S. C., Heylen, D., Toty, C., McCoy, K. D., Plantard, O., & Rispe, C. (2019). *A transcriptome-based phylogenetic study of hard ticks (Ixodidae). Sci. Rep. 9, 12923*.

Chipman, A. D., Ferrier, D. E. K., Brena, C., Qu, J., Hughes, D. S. T., Schröder, R., Torres-Oliva, M., Znassi, N., Jiang, H., & Almeida, F. C. (2014). The first myriapod genome sequence reveals conservative arthropod gene content and genome organisation in the centipede Strigamia maritima. *PLoS Biology*, *12*(11), e1002005.

Dehal, P., & Boore, J. L. (2005). Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biology*, *3*(10). https://doi.org/10.1371/journal.pbio.0030314

Dunlop, J. A. (2019). Miniaturisation in Chelicerata. *Arthropod Structure & Development*, *48*, 20–34.

Fernández, R., & Giribet, G. (2015). Unnoticed in the tropics: Phylogenomic resolution of the poorly known arachnid order Ricinulei (Arachnida). *Royal Society Open Science*, *2*(6). https://doi.org/10.1098/rsos.150065

Fernández, R., Kallal, R. J., Dimitrov, D., Ballesteros, J. A., Arnedo, M. A., Giribet, G., & Hormiga, G. (2018). Phylogenomics, Diversification Dynamics, and Comparative Transcriptomics across the Spider Tree of Life. *Current Biology*, *28*(9), 1489-1497.e5. https://doi.org/10.1016/j.cub.2018.03.064

Fernández, R., Sharma, P. P., Tourinho, A. L., & Giribet, G. (2017). The opiliones tree of life:

Shedding light on harvestmen relationships through transcriptomics. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1849). https://doi.org/10.1098/rspb.2016.2340

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. *Current Biology*, *27*(24), 3864–3870.

Francke, O. F. (1979). Spermatophores of some north American scorpions (Arachnida, Scorpiones). *Journal of Arachnology*, 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, A. C., & Funch, P. (2017). Rotiferan Hox genes give new insights into the evolution of metazoan bodyplans. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-00020-w

Gainett, G., Ballesteros, J. A., Kanzler, C. R., Zehms, J. T., Zern, J. M., Aharon, S., Gavish-Regev, E., & Sharma, P. P. (2020). Systemic paralogy and function of retinal determination network homologs in arachnids. *BMC Genomics*, *21*(1), 1–17.

Gainett, G., & Sharma, P. P. (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), 1–18.

Garb, J. E., Sharma, P. P., & Ayoub, N. A. (2018). Recent progress and prospects for advancing arachnid genomics. *Current Opinion in Insect Science*, *25*, 51–57.

Garwood, R. J., & Dunlop, J. (2014). Three-dimensional reconstruction and the phylogeny of extinct chelicerate orders. *PeerJ*, *2*, e641.

Giribet, G. (2007). *Harvestmenthe biology of opiliones* (Issue 595.43 H3).

Giribet, G., Edgecombe, G. D., & Wheeler, W. C. (2001). Arthropod phylogeny based on eight loci. *Nature*, *413*(September), 157–161.

Gravely, F. H. (1915). Notes on the habits of Indian insects, myriapods and arachnids. *Records of the Indian Museum*, *11*, 483–539.

Grbic, M., Khila, A., Lee, K., 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., Van Leeuwen, T., Clark, R. M., Rombauts, S., Rouzé, P., Grbić, V., Osborne, E. J., Dermauw, W., Ngoc, P. C. T., Ortego, F., Hernández-Crespo, P., Diaz, I., Martinez, M., Navajas, M., Sucena, É., Magalhães, S., Nagy, L., Pace, R. M., Djuranović, S., … Van De Peer, Y. (2011). The genome of Tetranychus urticae reveals herbivorous pest adaptations.pdf. *Nature*, *479*(7374), 487–492. https://doi.org/10.1038/nature10640

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Philip, D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Macmanes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., … Regev, A. (2013). De novo transcript sequence recostruction from RNA-Seq: reference generation and analysis with Trinity. *Nature Protocols*, *8*(8), 1–43. https://doi.org/10.1038/nprot.2013.084.De

Hansen, H. J., & Soerensen, W. E. (1905). *The Tartarides: a tribe of the order Pedipalpi*. Almqvist & Wiksell.

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, *35*(2), 518–522. https://doi.org/10.1093/molbev/msx281

Howard, R. J., Puttick, M. N., Edgecombe, G. D., & Lozano-Fernandez, J. (2020). Arachnid monophyly: Morphological, palaeontological and molecular support for a single terrestrialization within Chelicerata. *Arthropod Structure and Development*, *59*, 100997. https://doi.org/10.1016/j.asd.2020.100997

Hoy, M. A., Waterhouse, R. M., Wu, K., Estep, A. S., Ioannidis, P., Palmer, W. J., Pomerantz, A. F., Simão, F. A., Thomas, J., Jiggins, F. M., Murphy, T. D., Pritham, E. J., Robertson, H. M., Zdobnov, E. M., Gibbs, R. A., & Richards, S. (2016). Genome Sequencing of the Phytoseiid Predatory Mite Metaseiulus occidentalis Reveals Completely Atomized Hox Genes and Superdynamic Intron Evolution. *Genome Biology and Evolution*, *8*(6), 1762–1775. https://doi.org/10.1093/gbe/evw048

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. *Nature Ecology & Evolution*, *2*(4), 623–627.

Jager, M., Murienne, J., Clabaut, C., Deutsch, J., Le Guyader, H., & 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, G. E. (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. *Developmental Biology*, *382*(1), 224–234.

Jörger, K. M., 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 Evolutionary Biology*, *10*(1), 1–20.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Katoh, K., & Standley, D. M. (2013). *MAFFT Multiple Sequence Alignment Software Version 7 : Improvements in Performance and Usability Article Fast Track*. *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kenny, N. J., Chan, K. W., Nong, W., Qu, Z., Maeso, I., Yip, H. Y., Chan, T. F., Kwan, H. S., Holland, P. W. H., Chu, K. H., & Hui, J. H. L. (2016). Ancestral whole-genome duplication in the marine chelicerate horseshoe crabs. *Heredity*, *116*(2), 190–199. https://doi.org/10.1038/hdy.2015.89

King, N., & Rokas, A. (2017). Embracing uncertainty in reconstructing early animal evolution. *Current Biology*, *27*(19), R1081–R1088.

Kocot, K. M., Struck, T. H., Merkel, J., Waits, D. S., Todt, C., Brannock, P. M., Weese, D. A., Cannon, J. T., Moroz, L. L., & Lieb, B. (2017). Phylogenomics of Lophotrochozoa with consideration of systematic error. *Systematic Biology*, *66*(2), 256–282.

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

Krämer, J., Pohl, H., & Predel, R. (2019). Venom collection and analysis in the pseudoscorpion Chelifer cancroides (Pseudoscorpiones: Cheliferidae). *Toxicon*, *162*(February), 15–23. https://doi.org/10.1016/j.toxicon.2019.02.009

Lamsdell, J. C. (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. *Molecular Biology and Evolution*, *21*(6), 1095–1109. https://doi.org/10.1093/molbev/msh112

Laumer, C. E., Fernández, R., Lemer, S., Combosch, D., Kocot, K. M., Riesgo, A., Andrade, S. C. S., Sterrer, W., Sørensen, M. V., Giribet, G., Laumer, C. E., Ferna, R., Andrade, C. S., Combosch, D., Kocot, K. M., Riesgo, A., Sterrer, W., Sørensen, M. V., & Giribet, G. (2019). Revisiting metazoan phylogeny with genomic sampling of all phyla. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1906). https://doi.org/10.1098/rspb.2019.0831

Legg, D. A., Sutton, M. D., & Edgecombe, G. D. (2013). Arthropod fossil data increase congruence of morphological and molecular phylogenies. *Nature Communications*, *4*, 1–7. https://doi.org/10.1038/ncomms3485

Lehmann, T., & Melzer, R. R. (2018). Also looking like Limulus?–retinula axons and visual neuropils of Amblypygi (whip spiders). *Frontiers in Zoology*, *15*(1), 1–14.

Lehmann, T., & Melzer, R. R. (2019). The visual system of Thelyphonida (whip scorpions): support for Arachnopulmonata. *Arthropod Structure & Development*, *51*, 23–31.

Leite, D. J., Baudouin-Gonzalez, L., Iwasaki-Yokozawa, S., Lozano-Fernandez, J., Turetzek, N., Akiyama-Oda, Y., Prpic, N. M., Pisani, D., Oda, H., Sharma, P. P., & McGregor, A. P. (2018). Homeobox gene duplication and divergence in arachnids. *Molecular Biology and Evolution*, *35*(9), 2240–2253. https://doi.org/10.1093/molbev/msy125

Leite, D. J., Ninova, M., Hilbrant, M., Arif, S., Griffiths-Jones, S., Ronshaugen, M., & McGregor, A. P. (2016). Pervasive microRNA duplication in chelicerates: insights from the embryonic microRNA repertoire of the spider Parasteatoda tepidariorum. *Genome Biology and Evolution*, *8*(7), 2133–2144.

Lorenz, R., & Bernhart, S. H. (n.d.). Höner zu Siederdissen C et al (2011) ViennaRNA package 2.0. *Algorithms Mol Biol*, *6*, 26.

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

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., Geer, R. C., He, J., Gwadz, M., & Hurwitz, D. I. (2015). CDD: NCBI's conserved domain database. *Nucleic Acids Research*, *43*(D1), D222–D226.

Marlétaz, F., Peijnenburg, K. T. C. A., Goto, T., Satoh, N., & Rokhsar, D. S. (2019). A New Spiralian Phylogeny Places the Enigmatic Arrow Worms among Gnathiferans. *Current Biology*, *29*(2), 312-318.e3. https://doi.org/10.1016/j.cub.2018.11.042

Minh, B. Q., Hahn, M. W., & Lanfear, R. (2020). New methods to calculate concordance factors for phylogenomic datasets. *Molecular Biology and Evolution*, *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, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Nolan, E. D., Santibáñez-López, C. E., & Sharma, P. P. (2020). Developmental gene expression as a phylogenetic data class: support for the monophyly of Arachnopulmonata. *Development Genes and Evolution*, *2*. https://doi.org/10.1007/s00427-019-00644-6

Pepato, A. R., da Rocha, C. E. F., & Dunlop, J. A. (2010). Phylogenetic position of the acariform mites: sensitivity to homology assessment under total evidence. *BMC Evolutionary Biology*, *10*, 235. https://doi.org/10.1186/1471-2148-10-235

Pienaar, R., de Klerk, D. G., Putterill, J. F., & Mans, B. J. (2018). Notes on maternal behaviour in soft ticks: specifically observed in Argas (Argas) striatus Bedford, 1932 and Argas (Secretargas) transgariepinus White, 1846. *Ticks and Tick-Borne Diseases*, *9*(4), 889–895.

Punzo, F. (2012). *The biology of camel-spiders: Arachnida, Solifugae*. Springer Science & Business Media.

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

Ramírez, M. J., Magalhaes, I. L. F., Derkarabetian, S., Ledford, J., Griswold, C. E., Wood, H. M., & Hedin, M. (2021). Sequence capture phylogenomics of true spiders reveals convergent evolution of respiratory systems. *Systematic Biology*, *70*(1), 14–20.

Regier, J. C., Shultz, J. W., Zwick, A., Hussey, A., Ball, B., Wetzer, R., Martin, J. W., & Cunningham, C. W. (2010). Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature*, *463*(7284), 1079–1083. https://doi.org/10.1038/nature08742

Roelofs, D., Zwaenepoel, A., Sistermans, T., Nap, J., Kampfraath, A. A., Van de Peer, Y., Ellers, J., & Kraaijeveld, K. (2020). Multi-faceted analysis provides little evidence for recurrent whole-genome duplications during hexapod evolution. *BMC Biology*, *18*, 1–13.

Rokas, A., & Carroll, S. B. (2006). Bushes in the tree of life. *PLoS Biology*, *4*(11), 1899–1904. https://doi.org/10.1371/journal.pbio.0040352

Rota-Stabelli, O., Campbell, L., Brinkmann, H., Edgecombe, G. D., Longhorn, S. J., Peterson, K. J., Pisani, D., Philippe, H., & Telford, M. J. (2011). A congruent solution to arthropod phylogeny: phylogenomics, microRNAs and morphology support monophyletic Mandibulata. *Proceedings. Biological Sciences / The Royal Society*, *278*(1703), 298–306. https://doi.org/10.1098/rspb.2010.0590

Rowland, J. M. (1972). brooding habits and early development of Trithyreus pentapeltis (Cook),(Arachnida: Schizomida). *Entomol News Philadelphia*.

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

Santibáñez-López, C. E., Ojanguren-Affilastro, A. A., & Sharma, P. P. (2020). Another one bites the dust: Taxonomic sampling of a key genus in phylogenomic datasets reveals more non-monophyletic groups in traditional scorpion classification. *Invertebrate Systematics*, *34*(2), 133–143. https://doi.org/10.1071/IS19033

Santibáñez-López, C. E., Ontano, A. Z., Harvey, M. S., & Sharma, P. P. (2018). Transcriptomic analysis of pseudoscorpion venom reveals a unique cocktail dominated by enzymes and protease inhibitors. *Toxins*, *10*(5). https://doi.org/10.3390/toxins10050207

Schwager, E. E., Schoppmeier, M., Pechmann, M., & Damen, W. G. M. (2007). Duplicated Hox genes in the spider Cupiennius salei. *Frontiers in Zoology*, *4*(1), 1–11.

Schwager, E. E., Sharma, P. P., Clarke, T., Leite, D. J., Wierschin, T., Pechmann, M., Akiyama-Oda, Y., Esposito, L., Bechsgaard, J., Bilde, T., Buffry, A. D., Chao, H., Dinh, H., Doddapaneni, H. V., Dugan, S., Eibner, C., Extavour, C. G., Funch, P., Garb, J., … McGregor, A. P. (2017). The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. *BMC Biology*, *15*(1), 1–27. https://doi.org/10.1186/s12915-017-0399-x

Setton, E. V. W., Hendrixson, B. E., & Sharma, P. P. (2019). Embryogenesis in a Colorado population of Aphonopelma hentzi (Girard, 1852)(Araneae: Mygalomorphae: Theraphosidae): establishing a promising system for the study of mygalomorph development. *The Journal of Arachnology*, *47*(2), 209–216.

Setton, E. V. W., & Sharma, P. P. (2018). Cooption of an appendage-patterning gene cassette in the head segmentation of arachnids. *Proceedings of the National Academy of Sciences*, 201720193. https://doi.org/10.1073/pnas.1720193115

Sharma, P. P., Baker, C. M., Cosgrove, J. G., Johnson, J. E., Oberski, J. T., Raven, R. J., Harvey, M. S., Boyer, S. L., & Giribet, G. (2018). A revised dated phylogeny of scorpions: Phylogenomic support for ancient divergence of the temperate Gondwanan family

Bothriuridae. *Molecular Phylogenetics and Evolution*, *122*, 37–45.

Sharma, P. P., Fernández, R., Esposito, L. A., Gonzalez-Santillan, E., & Monod, L. (2015). Phylogenomic resolution of scorpions reveals multilevel discordance with morphological phylogenetic signal. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1804), 20142953–20142953. https://doi.org/10.1098/rspb.2014.2953

Sharma, P. P., Kaluziak, S. T., Pérez-Porro, A. R., González, V. L., Hormiga, G., Wheeler, W. C., & Giribet, G. (2014). Phylogenomic interrogation of arachnida reveals systemic conflicts in phylogenetic signal. *Molecular Biology and Evolution*, *31*(11), 2963–2984. https://doi.org/10.1093/molbev/msu235

Sharma, P. P., Santiago, M. A., González-Santillán, E., Monod, L., & Wheeler, W. C. (2015). Evidence of duplicated Hox genes in the most recent common ancestor of extant scorpions. *Evolution & Development*, *17*(6), 347–355.

Sharma, P. P., Schwager, E. E., Extavour, C. G., & Giribet, G. (2012). Hox gene expression in the harvestman Phalangium opilio reveals divergent patterning of the chelicerate opisthosoma. *Evolution & Development*, *14*(5), 450–463.

Sharma, P. P., Schwager, E. E., Extavour, C. G., & Wheeler, W. C. (2014). Hox gene duplications correlate with posterior heteronomy in scorpions. *Proceedings. Biological Sciences / The Royal Society*, *281*(1792), 20140661-. https://doi.org/10.1098/rspb.2014.0661

Sharma, P. P., Tarazona, O. A., Lopez, D. H., Schwager, E. E., Cohn, M. J., Wheeler, W. C., & Extavour, C. G. (2015). A conserved genetic mechanism specifies deutocerebral appendage identity in insects and arachnids. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1808), 20150698.

Shen, X., Hittinger, C. T., & Rokas, A. (2017). Contentious relationships in phylogenomic studies can be driven By a Handful of Genes. *Nat Ecol Evol.*, *1*(5), 1–23. https://doi.org/10.6084/m9.figshare.3792189.Shen

Shingate, P., Ravi, V., Prasad, A., Tay, B.-H., Garg, K. M., Chattopadhyay, B., Yap, L.-M., Rheindt, F. E., & Venkatesh, B. (2020). Chromosome-level assembly of the horseshoe crab genome provides insights into its genome evolution. *Nature Communications*, *11*(1), 1–13.

Shultz, J. W. (1990). Evolutionary morphology and phylogeny of Arachnida. *Cladistics*, *6*, 1–38.

Shultz, J. W. (2007). A phylogenetic analysis of the arachnid orders based on morphological characters. In *Zoological Journal of the Linnean Society* (Vol. 150, Issue 2). https://doi.org/10.1111/j.1096-3642.2007.00284.x

Shultz, J. W., & Regier, J. C. (2000). Phylogenetic analysis of arthropods using two nuclear protein–encoding genes supports a crustacean+ hexapod clade. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *267*(1447), 1011–1019.

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

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Simion, P., Philippe, H., Baurain, D., Jager, M., Richter, D. J., Di Franco, A., Roure, B., Satoh, N., Quéinnec, É., Ereskovsky, A., Lapébie, P., Corre, E., Delsuc, F., King, N., Wörheide, G., & Manuel, M. (2017). A Large and Consistent Phylogenomic Dataset Supports Sponges as the Sister Group to All Other Animals. *Current Biology*, 1–10. https://doi.org/10.1016/j.cub.2017.02.031

Tarver, J. E., Sperling, E. A., Nailor, A., Heimberg, A. M., Robinson, J. M., King, B. L., Pisani, D., Donoghue, P. C. J., & Peterson, K. J. (2013). miRNAs: small genes with big potential in metazoan phylogenetics. *Molecular Biology and Evolution*, *30*(11), 2369–2382.

Tarver, J. E., Taylor, R. S., Puttick, M. N., Lloyd, G. T., Pett, W., Fromm, B., Schirrmeister, B. E., Pisani, D., Peterson, K. J., & Donoghue, P. C. J. (2018). Well-annotated microRNAomes do not evidence pervasive miRNA loss. *Genome Biology and Evolution*, *10*(6), 1457–1470.

Thomson, R. C., Plachetzki, D. C., Mahler, D. L., & Moore, B. R. (2014). A critical appraisal of the use of microRNA data in phylogenetics. *Proceedings of the National Academy of Sciences*, *111*(35), E3659–E3668.

Wang, B., Dunlop, J. A., Selden, P. A., Garwood, R. J., Shear, W. A., Müller, P., & Lei, X. (2018). Cretaceous arachnid Chimerarachne yingi gen. et sp. nov. illuminates spider origins. *Nature Ecology & Evolution*, *2*(4), 614–622. https://doi.org/10.1038/s41559-017-0449-3

Wang, H.-C., Susko, E., & Roger, A. J. (2019). The relative importance of modeling site pattern heterogeneity versus partition-wise heterotachy in phylogenomic inference. *Systematic Biology*, *68*(6), 1003–1019.

Wang, H. C., Minh, B. Q., Susko, E., & Roger, A. J. (2018). Modeling Site Heterogeneity with Posterior Mean Site Frequency Profiles Accelerates Accurate Phylogenomic Estimation. *Systematic Biology*, *67*(2), 216–235. https://doi.org/10.1093/sysbio/syx068

Waterhouse, R. M., Seppey, M., Simao, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V., & Zdobnov, E. M. (2018). BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution*, *35*(3), 543–548. https://doi.org/10.1093/molbev/msx319

Wendruff, A. J., Babcock, L. E., Wirkner, C. S., Kluessendorf, J., & Mikulic, D. G. (2020). A Silurian ancestral scorpion with fossilised internal anatomy illustrating a pathway to arachnid terrestrialisation. *Scientific Reports*, *10*(1), 1–6. https://doi.org/10.1038/s41598-019-56010-z

Whelan, N. V, & Halanych, K. M. (2017). Who let the CAT out of the bag? Accurately dealing with substitutional heterogeneity in phylogenomic analyses. *Systematic Biology*, *66*(2), 232–255.

Zdobnov, E. M., Tegenfeldt, F., Kuznetsov, D., Waterhouse, R. M., Simao, F. A., Ioannidis, P., Seppey, M., Loetscher, A., & Kriventseva, E. V. (2017). OrthoDB v9. 1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Research*, *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*(6), 15–30.

Zhong, Y. F., & Holland, P. W. H. (2011). HomeoDB2: Functional expansion of a comparative homeobox gene database for evolutionary developmental biology. *Evolution and Development*, *13*(6), 567–568. https://doi.org/10.1111/j.1525-142X.2011.00513.x

Zimin, A. V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S. L., & Yorke, J. A. (2013). The MaSuRCA genome assembler. *Bioinformatics*, *29*(21), 2669–2677. https://doi.org/10.1093/bioinformatics/btt476

Chapter 3

How many long-branch orders occur in Chelicerata? Opposing effects of Palpigradi and Opilioacariformes on phylogenetic stability

Andrew Z. Ontano, Hugh G. Steiner, Prashant P. Sharma

This work is intended for submission to *Molecular Phylogenetics and Evolution*

Abstract

Excepting a handful of nodes, phylogenetic relationships among chelicerate orders remain poorly resolved, in part due to the incidence of long-branch attraction artifacts and limited sampling of key lineages. It has recently been shown that increasing representation of basal nodes plays an outsized role in resolving the placement of long-branch chelicerate orders such as pseudoscorpions. Two lineages have been consistently undersampled in the chelicerate phylogeny. First, sampling of the miniaturized order Palpigradi has been restricted to the partial transcriptome of a single species. Second, sampling of Opilioacariformes, the putative sister group to the remaining Parasitiformes, is similarly restricted to a single exemplar. These two lineages exhibit dissimilar properties with respect to branch length; Opilioacariformes shows relatively slow evolutionary rate, whereas Palpigradi possibly acts as a long-branch order (an effect that may be conflated with the degree of missing data). To assess these properties more rigorously, we constructed a phylogenomic dataset of Chelicerata wherein both lineages were sampled with three terminals, increasing the representation of these lineages per locus. We examined the effect of subsampling phylogenomic matrices using (1) taxon occupancy, (2) evolutionary rate, and (3) a principal components-based approach. We further explored the impact of taxon deletion experiments that mitigate the effect of long branches. Here, we show that Palpigradi constitutes a fourth long-branch chelicerate order (together with Acariformes, Parasitiformes, and Pseudoscorpiones), which further destabilizes the chelicerate backbone

---

* Contribution: I contributed data from my thesis chapters and performed a subset of the analyses.

topology. By contrast, the slow-evolving Opilioacariformes were consistently recovered within Parasitiformes, with certain subsampling practices recovering their traditional placement as the sister group to the remaining Parasitiformes. Whereas the inclusion of Opilioacariformes always resulted in the non-monophyly of Acari with support, deletion of Opilioacariformes from datasets consistently incurred the monophyly of Acari except in matrices constructed on the basis of evolutionary rate. Our results strongly suggest that Acari is an artifact of long-branch attraction.

Keywords: phylogenomic subsampling | long-branch attraction

Introduction

Long-branch attraction (LBA) refers to a well-characterized phylogenetic artifact in which rapidly evolving lineages (branches) that are distantly related are incorrectly resolved as sister taxa (Bergsten, 2005; Felsenstein, 1978). The symptoms of this artifact include the consistent recovery of counterintuitive groupings that are united only by accelerated evolutionary rate; the placement of one or more rapidly evolving ingroup lineages at the base of the tree near outgroups, which tend to be comparatively poorly sampled; or a combination of both of these. LBA is a form of statistical inconsistency, which cannot be mitigated by quantity of data alone. Several approaches have been proposed to mitigate LBA, such as the use of model-based phylogenetic reconstructions, the implementation of more sophisticated evolutionary models (e.g., site heterogeneous models) in molecular phylogenetics, intensive taxonomic sampling to “break” long branches, data recoding strategies, omission of rapidly evolving genes or sites, omission of fast-evolving lineages (or substitution with slow-evolving exemplars), and various criteria for data curation. Strategies to mitigating LBA seem to vary in effectiveness from one taxonomic group to another, as a function of species richness, the disparity of evolutionary rates of the taxa in question, and the phylogenetic depth of the splits that are targeted for resolution.

A prerequisite to mitigating LBA is to diagnose the ingroup lineages in which it occurs. Ideally, the number of long-branch taxa is limited to two, as fewer fast-evolving lineages translates to fewer potential attractants in the content of this artifact. But one of the Gordian knots of the animal tree of life is the basal phylogeny of Chelicerata, the subdivision of arthropods that includes the sea spiders (Pycnogonida), the horseshoe crabs (Xiphosura), and an assemblage of 12 terrestrial orders (Arachnida). The last of these includes at least three long-branch lineages: Pseudoscorpiones (pseudoscorpions or “book scorpions”), Parasitiformes (ticks and allies), and Acariformes (mites and allies). Together with an ancient rapid radiation at the base of Euchelicerata (=Arachnida + Xiphosura), these long-branch taxa have consistently clustered together as basally branching groups in molecular phylogenies of Chelicerata. It is traditionally thought that Parasitiformes and Acariformes together form the clade Acari, and indeed, they are sometimes recovered as sister groups in molecular phylogenies (Howard et al., 2020; Lozano-Fernandez et al., 2019). But the long branch lengths subtending these taxa, together with inconsistent support for this result across phylogenomic studies (Ballesteros et al., 2019; Ballesteros & Sharma, 2019) disfavors the monophyly of Acari as a definitive grouping. As a result, the higher-level relationships within Euchelicerata are poorly understood, with recent datasets disputing even the monophyly of Arachnida (Ballesteros et al., 2019; Ballesteros & Sharma, 2019; Howard et al., 2020; Nolan et al., 2020; Sharma et al., 2014).

In a recent work examining the placement of pseudoscorpions (Ontano et al., 2021), it was shown that pseudoscorpions are reliably placed as the sister group of scorpions in phylogenomic analyses based on BUSCO genes, provided that pseudoscorpions are well-sampled (i.e., that the branch length subtending this group is broken by the sampling of basally branching superfamilies); omitting the representation of basal nodes of this long-branch order resulted in their placement with one or the other long-branch acarine orders instead. The resolution of Pseudoscorpiones + Scorpiones was validated by the discovery that pseudoscorpions share a whole genome duplication with the remaining arachnopulmonates (scorpions, spiders, and three other orders that bear book lungs; Ontano et al. 2021), a result that cannot be reconciled with the alternative placement of pseudoscorpions with the other long-branch orders (as Parasitiformes and Acariformes exhibit no evidence of whole genome duplication; Leite et al. 2016, 2018; Gainett et al. 2021; Ontano et al. 2021). Thus, Ontano et al. (2021) were able to show that taxonomic sampling outperformed other strategies like filtering for dataset occupancy, filtering for evolutionary rate, use of site heterogeneous models, and algorithmic approaches to tree reconstruction in the context of mitigating LBA with respect to the placement of pseudoscorpions.

Regrettably, one lineage missing from that previous work was Palpigradi, the most enigmatic of the arachnid orders. This miniaturized group of arachnids exhibits a mysterious combination of morphological characters that has confounded efforts to place it reliably in arachnid phylogenies on the basis of morphological data, they lack eyes, a sternum, a coxal gland similar in anatomy to that of sun spiders (Solifugae), and possess a multi-articled flagellum (Shultz, 1990, 2007; Weygoldt & Paulus, 1979; ref. Ballesteros et al. 2019). Only one transcriptome of a palpigrade (the species *Eukoenenia spelaea*) has ever been produced for analysis in chelicerate phylogenetics (Ballesteros et al. 2019), but due to the quality of that dataset, Palpigradi was not well represented across phylogenomic matrices in that study, and its placement was inconclusive (partial support was obtained for the sister group relationship of palpigrades to Solifugae). Due to the level of missing data for this lineage, it could not be included in the main analyses of Ontano et al. (2021), as these required a specific minimum data completeness threshold. At present, it is therefore not known whether the instability of Palpigradi was attributable to missing data in the study of Ballesteros et al. (2019) or if Palpigradi constitutes a long-branch taxon.

Drawing upon the lessons of Ontano et al. (2021), a separate concern for resolving the chelicerate tree of life is the sampling of basal nodes within fast-evolving orders, and ideally, the inclusion of slowly-evolving lineages within such orders. One key candidate in this regard is Opilioacariformes, the putative sister group to the remaining Parasitiformes. The inclusion of the first opilioacariform in the study of Ballesteros et al. (2019) showed that this group exhibited the smallest patristic distance within the Parasitiformes. Intriguingly, Ontano et al. (2021) were able to show that previously published datasets supporting Acari monophyly (Howard et al. 2020) would instead support Acari diphyly upon the inclusion of a single opilioacariform exemplar.

To understand how the inclusion of Palpigradi and Opilioacariformes impacts chelicerate relationships (viz. LBA), we endeavored to increase the sampling of these groups in phylogenomic studies. With transcriptomes available for the palpigrade family Eukoeneniidae, we focused our efforts on developing a high-quality transcriptome of, Prokoeneniidae. In addition, we generated a new high-quality transcriptome for the genus *Opilioacarus*, bringing the total number of exemplars of each group to three datasets. Here, we show that Palpigradi constitutes a fourth long-branch order of Chelicerata, whose inclusion further destabilizes chelicerate relationships. By contrast, Opilioacariformes has the opposite effect, with its inclusion breaking up the monophyly of Acari, which we show to be an LBA artifact.

Materials and Methods

Field collection and sequencing

Specimens of the palpigrade *Prokoenenia wheeleri* and the opilioacariform *Opilioacarus texanus* were hand collected from Cypress Creek Park, Travis County, Austin, Texas, United States (30.438459, -97.874670) on 8-9 January 2020 by A.Z.O., P.P.S., Emily V.W. Setton, and Jesús A. Ballesteros. 20 individuals of *P. wheeleri* and one individual of *O. texanus* were transferred directly to Trizol TriReagent using paintbrushes and stored on ice. RNA extraction, mRNA purification, and library preparation followed our previously published protocols (Ballesteros et al. 2019). Sequencing was performed on an Illumina NovaSeq 6000 platform with a 2 x 150 bp paired end sequencing strategy. Completeness of single-copy BUSCOs (Benchmarking Universal Single Copy Orthologs) were 90.3% for *P. wheeleri* and 96.1% for *O. texanus*, as inferred using the BUSCO-Arach dataset for arachnids derived from OrthoDB v.10 (Kriventseva et al., 2019; Simão et al., 2015; Waterhouse et al., 2018).

The small body size of *P. wheeleri* incurs high risks of environmental contamination in pooled samples. As an additional validation of on-target sequencing, we performed BLASTn searches in the *P. wheeleri* transcriptome for a set of genes previously Sanger-sequenced for this species in the 62-locus dataset of Regier et al. (2010). For all genes recovered in this search, the best BLAST hit was invariably to the *P. wheeleri* data previously generated in the Regier et al. (2010) study.

Species sampling and orthology inference

We compiled a dataset of 126 chelicerates (3 Palpigradi, 38 Pseudoscorpiones, 12 Scorpiones, 18 Araneae, 6 Pedipalpi, 13 Opiliones, 7 Ricinulei, 3 Xiphosura, 2 Solifugae, 12

Parasitiformes, 10 Acariformes, 2 Pycnogonida) and 15 outgroups (3 Onychophora, 5 Myriapoda, 7 Pancrustacea). This dataset (Table S1) overlaps closely with our recent work (Ontano et al. 2021), with the following modifications: we added transcriptomes for three palpigrades, two Ricinulei, and three parasitiform mites, including two opilioacariform mites (*Neacarus* sp. and *Opilioacarus* sp.). Taxon selection prioritized libraries of high quality and the representation of basal splits in all major groups (Ballesteros et al., 2019, 2021; Benavides et al., 2019; Fernández et al., 2017, 2018; Santibáñez-López et al., 2019, 2020; Sharma et al., 2015).

Candidate ORFs were identified in transcripts using TransDecoder (Haas et al., 2013). Loci selected for phylogenomic analysis consisted of the subset of 2,934 Benchmarked Universal Single Copy Orthologs identified for Arachnida (BUSCO-Arach) derived from OrthoDB v.10 (Kriventseva et al., 2019; Simão et al., 2015; Waterhouse et al., 2018). Each dataset was analyzed with the OrthoDB pipeline to identify available homologs of 2,934 arachnid-specific BUSCO genes. Duplicated BUSCOs were discarded to retain only validated, single-copy loci. We further filtered loci using a taxon decisiveness criterion, such that every retained gene had to include at least one member of each higher-level lineage (chelicerate orders; Mandibulata; Pancrustacea; Onychophora) to be retained in the dataset. This filtering resulted in an initial set of 1,024 BUSCO genes. Multiple sequence alignment was performed using MAFFT 7.3.8 (-*anysymbol –auto*; Katoh & 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 & Hormiga, 2016).

Phylogenomic subsampling

We investigated three approaches to ordered phylogenomic subsampling. First, we generated matrices of 200, 400, and 600 loci based on taxon occupancy (i.e., allowing more loci with greater amounts of missing data). Second, we generated three matrices of the 200, 400, and 600 slowest-evolving loci based on mean percent pairwise sequence identity (MPSI), a proxy for evolutionary rate. Third, we applied a recently developed principal components-based method that accounts for multiple metrics of phylogenetic usefulness (*sortR*; Mongiardino Koch, 2021).

sortR requires a resolved species tree *a priori* for the computation of Robinson-Foulds distances for each gene tree. However, given the marked conflict at the base of the chelicerate tree of life, within some chelicerate orders, as well as within some outgroups (e.g., Myriapoda), we endeavored to limit the influence of the species tree on the ranking of phylogenetically useful genes. We therefore supplied a species tree wherein all nodes were collapsed except for higher-level splits that are robustly supported across analyses (e.g., monophyly of chelicerate orders, Tetrapulmonata, Pedipalpi, Euchelicerata, Chelicerata, Mandibulata, Pancrustacea, and Arthropoda). We supplied tree topologies for individual loci *a priori*, which were computed with IQ-TREE v.1.6.11 (Chernomor et al., 2016; Nguyen et al., 2015), 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). This pipeline recovered an axis of phylogenetic usefulness (PC2); as before, matrices of the 200, 400, and 600 loci most useful were generated from the ranking.

Phylogenomic inference

Tree topologies for concatenated datasets were computed with IQ-TREE v.1.6.11 (Chernomor et al., 2016; Nguyen et al., 2015), 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). We additionally performed model selection under the posterior mean site frequency (PMSF), a mixture model that approximates the Bayesian CAT model in a maximum likelihood framework (Lartillot & Philippe, 2004; Wang et al., 2018). Analyses were performed using the LG + C20 + F + Γ model.

Taxon deletion experiments

For each of the nine matrices previously generated, we performed the following deletions of taxa and recomputed the tree topology under models for the corresponding full dataset. First, to assess the impact of Opilioacariformes (the slow evolving putative sister group of the remaining Parasitiformes) in the phylogeny, we removed Opilioacariformes from each dataset. Second, to evaluate the topological stability of four putative long-branch taxa (Acariformes, Parasitiformes, Palpigradi, and Pseudoscorpiones), we deleted all but Palpigradi and computed the resulting trees. Lastly, we treated Opilioacariformes as separate from Parasitiformes, and deleted all but Palpigradi and Opilioacariformes and computed the resulting trees.

Results

Phylogenetic analyses under varying matrix construction criteria

Data matrices constructed using taxon occupancy thresholds (200-, 400-, and 600-most complete loci) resulted in the 83%, 78%, and 73% representation of the 141 terminals, respectively. Data matrices constructed using evolutionary rate thresholds (200-, 400-, and 600-

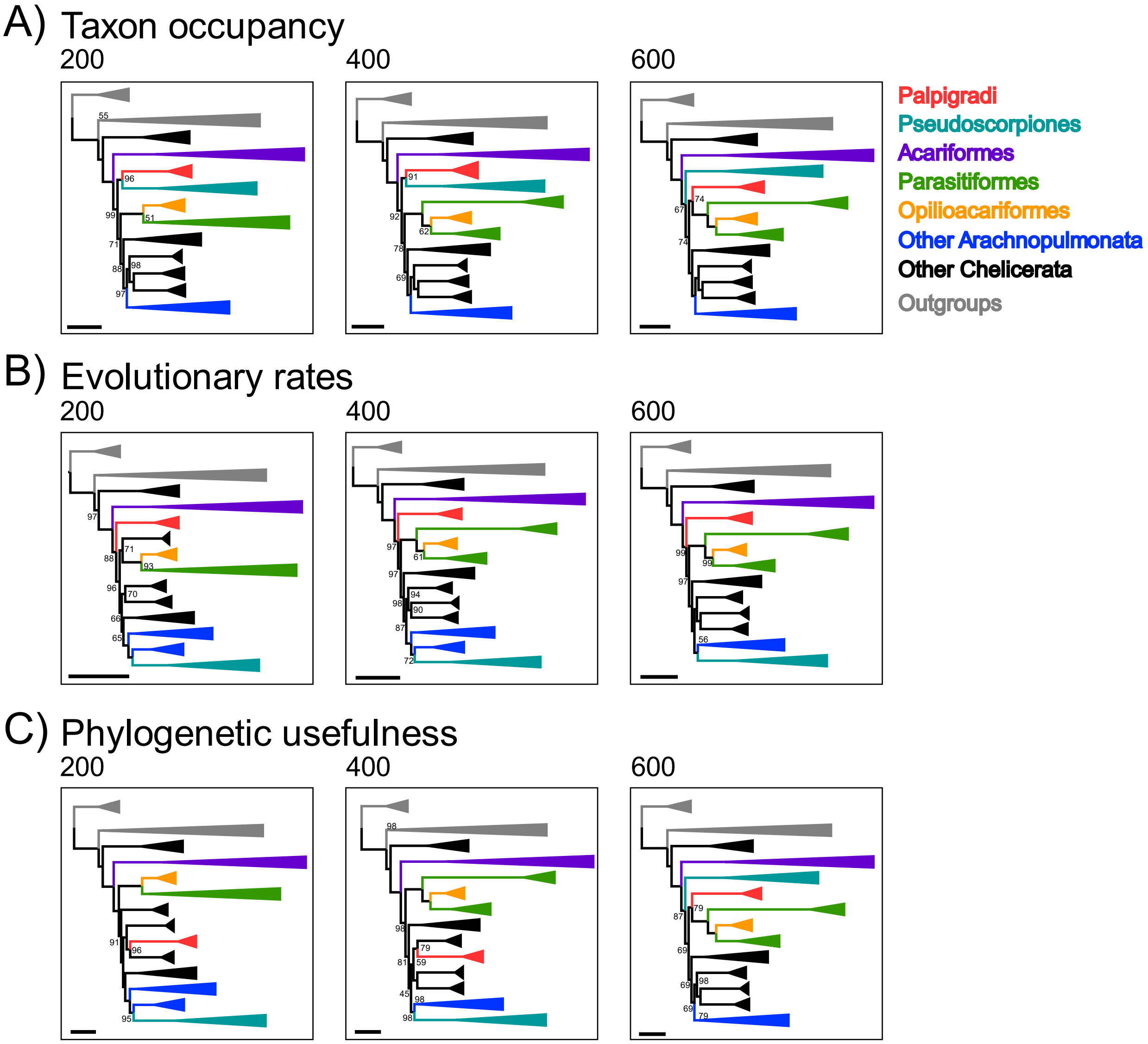


Figure 2 Maximum likelihood trees constructed from matrices of 200, 400, and 600 loci, selected under alternative optimality criteria. (a) Ordered by taxon occupancy. (b) Ordered by evolutionary rate. (c) Ordered by phylogenetic usefulness (output of sortR). Colors correspond to legend on the right. Ultrafast bootstrap support values under 100% are labeled next to their corresponding node; unlabeled nodes are maximally supported. Scale bar: 0.2. Note the placement of Opilioacariformes within Parasitiformes as a function of matrix composition.

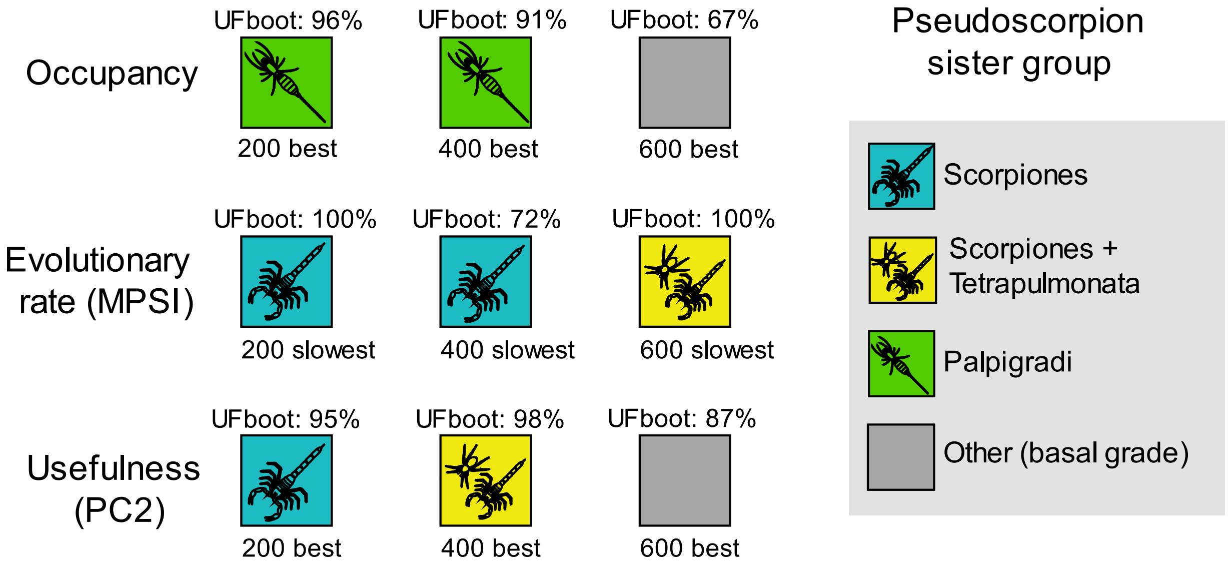


Figure 2 Sensitivity plot of pseudoscorpion placement in maximum likelihood analyses shown in Figure 1, with corresponding ultrafast bootstrap support values.

slowest evolving loci) resulted in mean sequence identity values of 73.7%, 65.0%, and 58.3% (across the entire matrix), respectively.

Partitioned maximum likelihood (ML) tree searches of these nine matrices recovered inconsistent phylogenetic placement of Palpigradi across analyses (Fig. 1). Matrices constructed based on taxon occupancy recovered the relationship Palpigradi + Pseudoscorpiones (ultrafast bootstraps [BS]=91-96%) or the relationship Palpigradi + Parasitiformes with poor support (BS=74%) (Fig. 1a). In all matrices built using taxon occupancy, pseudoscorpions were recovered as part of a grade toward the base of Euchelicerata, rather than as part of a monophyletic Arachnopulmonata. Of these three matrices, only the most complete matrix (200 loci) recovered Opilioacariformes as the sister group to the remaining Parasitiformes. The inclusion of more loci recovered a nested position with Opilioacariformes + Ixodida as sister group to Mesostigmata.

Matrices constructed based on evolutionary rate (Fig. 1b) recovered Palpigradi in a grade with Acariformes at the base of Euchelicerata (BS=88-99%). The 200- and 400-slowest evolving locus matrices were able to recover the monophyly of Arachnopulmonata and Panscorpiones, albeit with limited support for Panscorpiones (BS=72%) in the 400-slowest evolving locus matrix. In the 600-slowest evolving locus matrix, pseudoscorpions were recovered as the sister group of the remaining arachnopulmonates, though this relationship was weakly supported (BS=56%). In this family of matrices, only the 200-slowest evolving locus matrix was able to recover the traditional placement of Opilioacariformes as the sister group of the remaining Parasitiformes (BS=93%), with the inclusion of noisier genes resulting in the nested placement of Opilioacariformes as the sister group of Ixodida to the exclusion of Mesostigmata.

Matrices constructed based on PC2 of the *sortR* pipeline recovered the relationships Palpigradi + Ricinulei (BS=96%), Palpigradi + Solifugae (BS=79%), or Palpigradi + Parasitiformes (BS=79%), as a function of increasing matrix size. ML analysis of the 200 most "useful" locus matrix resulted in the recovery of Panscorpiones (BS=95%) and Arachnopulmonata (BS=100%), as well as the placement of Opilioacariformes as the sister group of the remaining Parasitiformes (BS=100%). The addition of less useful (noisier) genes resulted in the nested placement of Opilioacariformes, as well as loss of support for Panscorpiones, with eventual dissolution of Arachnopulmonata in the 600 most "useful" locus matrix.

Given the rare genomic change uniting Arachnopulmonata (Ontano et al. 2021), matrix construction criteria were compared directly on the basis of pseudoscorpion placement (Fig. 2).

The monophyly of Acari or Arachnida was not recovered in any partitioned model analysis.

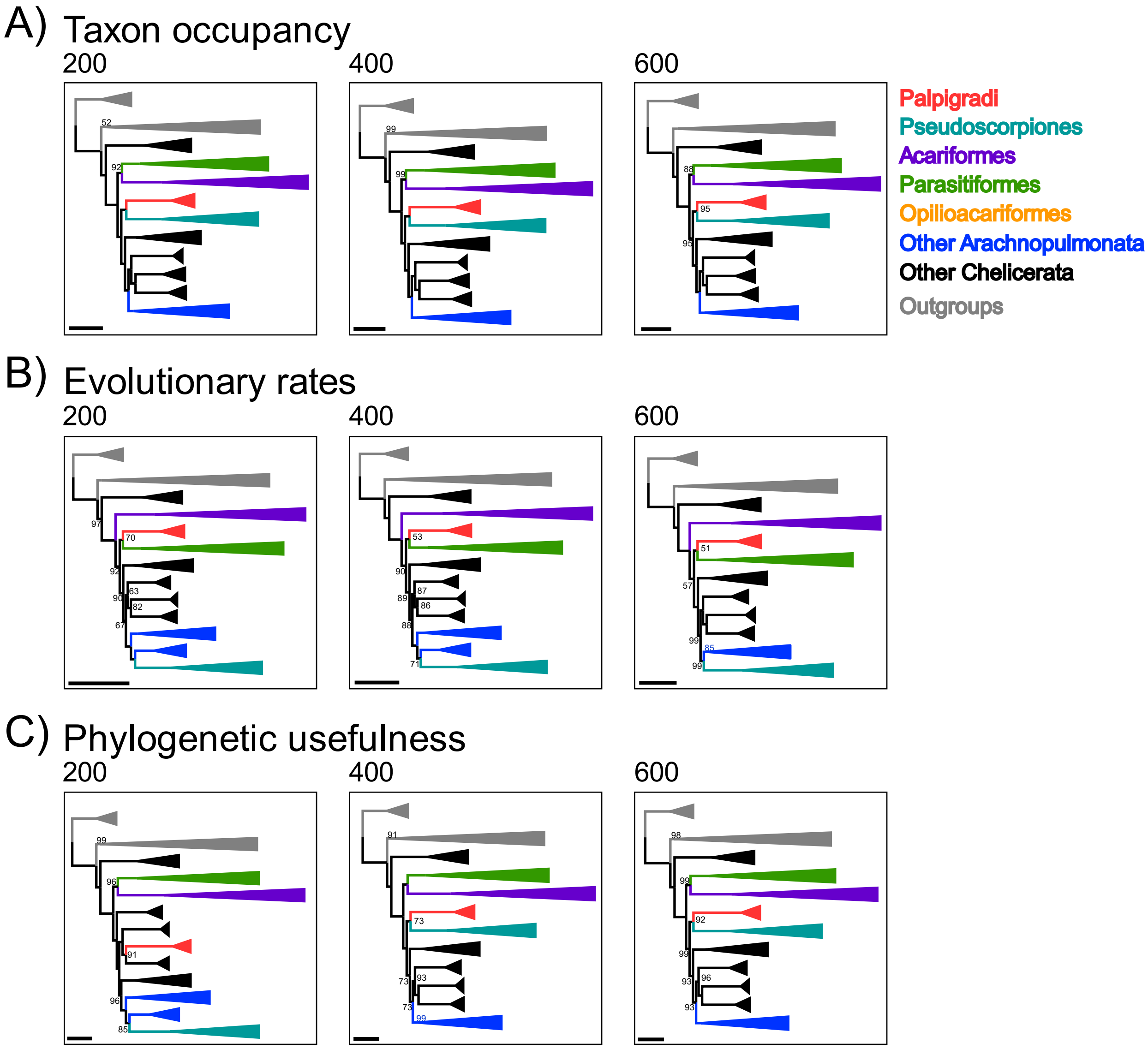


Figure 3 Taxon deletion experiments that remove Opilioacariformes result in the artifactual monophyly of Acari. Maximum likelihood trees are shown with same arrangement of matrix optimality criteria as in Figure 1. Colors correspond to legend on the right. Ultrafast bootstrap support values under 100% are labeled next to their corresponding node; unlabeled nodes are maximally supported. Scale bar: 0.2.

Taxon deletion experiments

Taxon deletion experiments followed one of three schemes: removal of Opilioacariformes (Fig. 3), removal of all long-branch chelicerate orders except Palpigradi (Fig. 4), or removal of all long-branch orders except both Palpigradi and Opilioacariformes (Fig. 5).

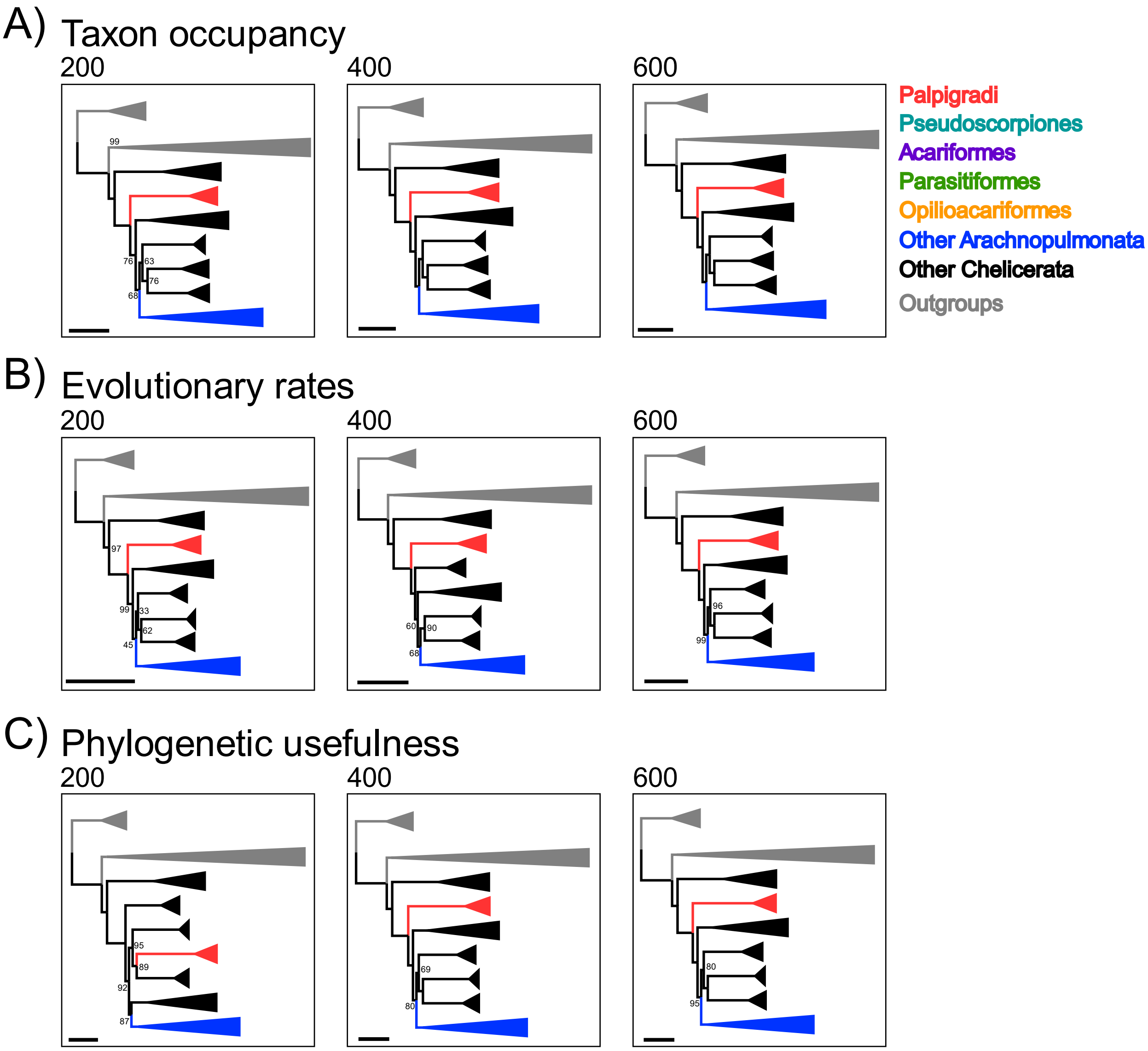


Figure 4 Taxon deletion experiments that remove all long-branch orders except Palpigradi do not mitigate phylogenetic instability of this lineage. Maximum likelihood trees are shown with same arrangement of matrix optimality criteria as in Figure 1. Colors correspond to legend on the right. Ultrafast bootstrap support values under 100% are labeled next to their corresponding node; unlabeled nodes are maximally supported. Scale bar: 0.2.

With the removal of Opilioacariformes (Fig. 3), matrices constructed under taxon occupancy and phylogenetic usefulness criteria (PC2) recovered Acari as monophyletic (BS=88-99% for occupancy; BS=96-100% for PC2). Matrices constructed based on rate (through MPSI) consistently recovered Acari as diphyletic with Parasitiformes recovered as the sister group of Palpigradi with low support (BS=51-70%; Fig. 3b). The removal of Opilioacariformes further

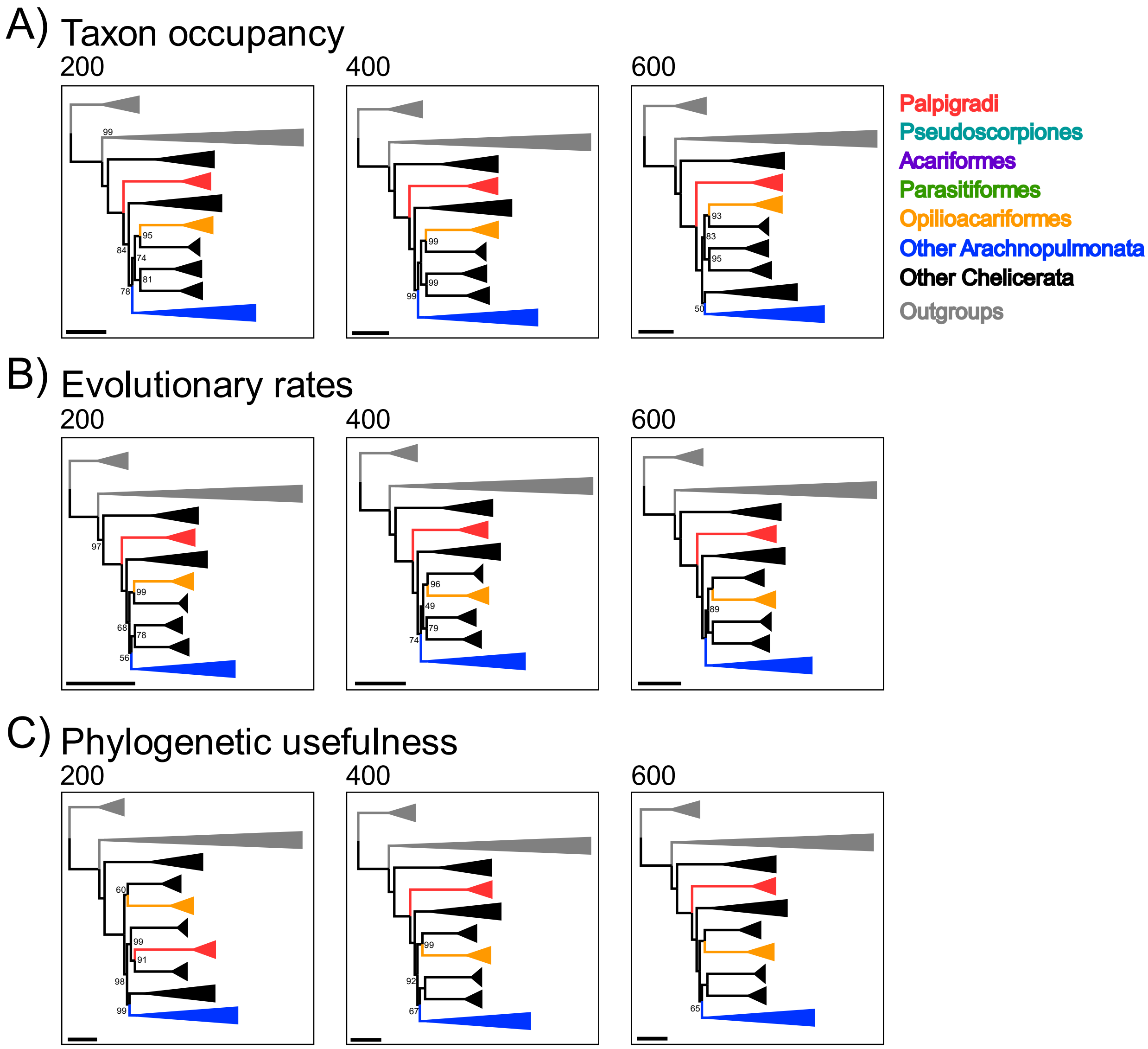


Figure 5 Taxon deletion experiments retaining only Opilioacariformes and Palpigradi, but no other long-branch taxa. Maximum likelihood trees are shown with same arrangement of matrix optimality criteria as in Figure 1. Colors correspond to legend on the right. Ultrafast bootstrap support values under 100% are labeled next to their corresponding node; unlabeled nodes are maximally supported. Scale bar: 0.2.

destabilized the placement of Pseudoscorpiones within Arachnouplmonata, but did not significantly affect Arachnopulmonata monophyly (compare support values to Fig. 1c). Matrices constructed under the criterion of evolutionary rate (Fig. 3b) consistently recovered Acari as diphyletic, with Parasitiformes recovered as the sister group of Palpigradi with low support (BS=51-70%).

The removal of all long-branch orders except Palpigradi (Fig. 4) did not have a strong effect on the placement of Palpigradi. In the absence of other long-branch orders, Palpigradi are recovered as sister to all other euchelicerates, except in the matrix constructed using the 200 most useful loci (i.e., highest values for PC2; Fig. 4c), which recovered the relationship Palpigradi + Ricinulei (BS=89%).

In analyses that included Opilioacariformes and Palpigradi in the absence of the remaining long-branch taxa (Fig. 5), matrices constructed using taxon occupancy and evolutionary rate recovered the relationship Parasitiformes + Xiphosura (BS=93-99%), except in the matrix incorporating the 600-slowest evolving loci. This analysis, along with all matrices constructed on the criterion of phylogenetic usefulness (Fig. 5c), recovered the relationship Parasitiformes + Solifugae (BS=100% for evolutionary rate-based matrices; BS=59-100% for phylogenetic usefulness-based matrices).

Analyses with site heterogeneous models

We performed the same family of analyses above, but with site heterogeneous model implementations for maximum likelihood. Tree topologies were therefore computed under the PMSF model (LG + C20 + F + Γ). Results differed minimally from partitioned analyses and are not discussed in detail here. In analyses with all taxa, Palpigradi were recovered more frequently in a basal grade in comparison to partitioned model approaches (Figure S1). Arachnopulmonata was maximally supported across all matrices constructed based on evolutionary rate (Fig. S1b) and matrices incorporating the 200 and 400 most "useful" loci (Fig. S1c), but pseudoscorpions were recovered as sister to all other arachnopulmonates in all these analyses except for the 200-slowest evolving locus matrix, which recovered Panscorpiones (BS=86%; Fig. S1b).

Taxon deletion experiments under the PMSF model also reflected similar results (Figs. S2-S4). In analyses with all long-branch taxa except Palpigradi removed (Fig. S3), we recovered a placement of Palpigradi in a basal grade, except for the matrix constructed from the 200 most useful (i.e., highest PC2) locus matrix, where they were recovered as the sister group to the clade Ricinulei + Xiphosura (BS=98%). In analyses with Opilioacariformes removed, the monophyly of Acari was never recovered (Fig. S2).

Discussion

Palpigradi: the fourth long-branch order of Chelicerata

Previous analyses of chelicerate phylogenetic relationships applied distributions of patristic distances and taxon deletion experiments to explore the stability of the arachnid orders (Sharma et al. 2014; Ballesteros and Sharma 2019). These works were unable to sample either Palpigradi or Opilioacariformes, and thus concluded that at least three orders exhibited clearly problematic trends of accelerated molecular evolution: Acariformes, Parasitiformes, and Pseudoscorpiones. Due to the quality of the only available palpigrade transcriptome, Ballesteros et al. (2019) were not able to infer whether the instability of *Eukoenenia spelaea* in a phylotranscriptomic dataset was attributable to poor taxonomic sampling, missing data, or other systematic artifacts.

Here, we generated the highest quality dataset of Palpigradi to date, adding this to the existing pair of transcriptomes for two *Eukoenenia*. The sampling of both Prokoeneniidae and Eukoeneniidae ensures the representation of the basal-most node in crown-group Palpigradi, given that only these two extant families are known and previous work has shown them to be reciprocally monophyletic (Giribet et al., 2014). In addition, the high quality of the *P. wheeleri* library, together with our decisiveness criterion (i.e., retaining only genes that sampled at least one exemplar of each chelicerate order, as well as major outgroup lineages), limited the impact of missing data or uninformative orthogroups for inferring higher-level relationships.

Our analyses of chelicerate relationships with these augmented matrices revealed clear evidence of instability in palpigrade placement, as a function of matrix assembly criterion (Figs. 1, S1). Palpigradi typically clustered with pseudoscorpions (Fig. 1a, S1a), formed a grade with Acariformes at the base of Euchelicerata (Fig. 1b, S1b), or was recovered as the sister group to other unstable groups like Parasitiformes or Ricinulei (Fig. 1c, S1c). The instability exhibited by this taxon, together with its clustering near the base of the tree with other long-branch orders, are strongly suggestive of an LBA artifact.

To corroborate this instability of Palpigradi across datasets, we undertook a separate analyses wherein we added the three palpigrade datasets to the G_3 matrix of Ontano et al. (2021), which consisted of 693 loci (70% taxon occupancy threshold) assembled with an older generation of arthropod-specific BUSCO genes. This matrix, which sampled all pseudoscorpion superfamilies, was previously shown to recover Panscorpiones (BS=81%) and Arachnopulmonata (BS=81%) under either partitioned or site heterogeneous model approaches. Upon addition of Palpigradi to this dataset, Pseudoscorpiones were drawn to the base of the tree as the sister group of Acariformes (BS=85%), with the four long-branch orders (Acariformes, Pseudoscorpiones, Palpigradi, Parasitiformes) forming a grade at the base of Euchelicerata (Fig. S5).

These results suggest that the instability of Palpigradi is not attributable to subsampling matrices smaller than ca. 700 genes, nor to properties of the BUSCO-Arach loci. We infer that palpigrades constitute a fourth long-branch arachnid order, whose inclusion in chelicerate phylogenomic datasets only further destabilizes the basal euchelicerate topology. This inference is consistent with the observation that the four long-branch chelicerate taxa all exhibit different degrees of miniaturization, and, in the case of many groups of Acariformes and Parasitiformes, a parasitic lifestyle. Both of these evolutionary phenomena are associated with rapid evolutionary rates and long-branch artifacts throughout Metazoa, as epitomized by LBA artifacts surrounding the relationships of Tardigrada and Nematoda (Borner et al., 2014; Laumer et al., 2019).

While taxonomic sampling has been shown to outperform other strategies to resolving LBA artifacts in Pseudoscorpiones, this strategy may have limited effectiveness for Palpigradi, because the basal-most node in crown-group palpigrades has already been sampled in the present matrix. Adding more exemplars of either palpigrade family may have little effect in breaking the branch subtending this node, if these families are indeed systematically valid and reciprocally monophyletic. New insights from rare genomic changes and the incidence of shared genome duplications may inform the placement of this enigmatic lineage, with particular emphasis on testing the older notion that Palpigradi are closely related to Tetrapulmonata (but see Seiter et al., 2021).

Inclusion of the slowly-evolving Opilioacariformes refutes Acari monophyly

The monophyly of Acari is another controversial topic in arthropod phylogenetics. Various analyses of morphological data have supported the sister group relationship of Acariformes + Parasitiformes (Shultz, 1990, 2007; but see Dunlop et al., 2012; Pepato et al., 2010). Molecular phylogenies have recovered variable support for this relationship (Ballesteros et al., 2019; Ballesteros & Sharma, 2019; Giribet et al., 2001; Howard et al., 2020; Lozano-Fernandez et al., 2019; Masta et al., 2009; Regier et al., 2010; Sharma et al., 2014; Wheeler &

Hayashi, 1998), which suspiciously resembles an LBA artifact, owing to the long patristic distances exhibited by commonly studied exemplars of both groups. One exception to this trend is Opilioacariformes, the rarely-encountered putative sister group to the remaining Parasitiformes; the sole opilioacariform transcriptome sequenced to date exhibited a comparatively short patristic distance across datasets, suggesting that this free-living (non-parasitic) taxon did not share the rapid evolutionary rates observed in many parasitiform genomes (e.g., Hoy et al., 2016).

Upon expanding the sampling of opilioacariform datasets to three genera, we never recovered the monophyly of Acari across our analyses (Fig. 1, S1). In taxon deletion experiments under partitioned model analyses, the removal of Opilioacariformes alone was sufficient to recover Acari monophyly in most datasets (Fig. 3a, 3c). These results closely parallel a previous analysis by Ontano et al. (2021), who added one opilioacariform and one palpigrade library to the analyses of Howard et al. (2020), in order to test the claims that these datasets could recover monophyly of Arachnida and Acari. Ontano et al. (2021) were able to show that the addition of just two phylogenetically significant taxa to those datasets was sufficient to collapse support for both arachnid and acarine monophyly (Figs. S2 and S3 of Ontano et al. 2021). Moreover, in taxon deletion experiments under site heterogeneous models, even the removal of Opilioacariformes did not render Acari monophyletic (Fig. S2).

These results strongly suggest that the monophyly of Acari reflects another LBA artifact in chelicerate phylogeny. The correspondences in the mouthparts and body plans of Acariformes and Parasitiformes therefore likely reflect morphological convergence rather than homologies. Future efforts to resolve the placement of these diverse groups must focus on expanding the sampling of phylogenetically significant groups that break long branches and potentially exhibit lower evolutionary rates. Examples of key targets for future interrogation of acarine relationships in phylogenomic studies include Holothyrida (Parasitiformes) and several groups of the "endeostigmatan" mites (Acariformes), such as the basally branching families Nanorchestidae and Alycidae (Klimov et al., 2018).

Phylogenetic usefulness versus *evolutionary rate in chelicerate phylogeny*

Beyond assessing the effects of sampling Palpigradi and Opilioacariformes in chelicerate phylogeny, we assessed competing strategies for locus selection as antidotes to LBA artifacts. Reducing missing data, with emphasis on clade-specific patterns of missing genes (i.e., rows in phylogenomic matrices), has been argued to be important for phylogenetic accuracy (Roure et al., 2013). In the specific case of LBA, matrix construction using slowly-evolving genes (either through filtering out noisy loci, saturated sites, or recoding strategies) has been argued to be an effective solution for reducing artifactual grouping of fast-evolving taxa. A more comprehensive strategy to subsampling genes for high phylogenetic signal and low noise was recently proposed by Mongiardino Koch (2021); *sortR* makes use of a principal components-based approach that aims to maximize metrics of phylogenetic signal (e.g., Robinson-Foulds distance from a species tree; bootstrap values on gene trees; see also Salichos and Rokas 2013), while minimizing metrics of noise (e.g., root-to-tip variance; saturation; compositional heterogeneity). This promising approach offers a more reliable and reproducible means of subsampling loci to construct matrices of reasonable size for computationally demanding approaches like phylogenomic dating. For well-behaved datasets, the composite metric of phylogenetic usefulness (principal component 2) has been shown to outperform subsampling by evolutionary rate (principal component 1), with exceptions in the cases of ancient and complex radiations (e.g., Hexapoda; Phasmatodea; Mongiardino Koch 2021).

We brought all three strategies to bear on the higher-level relationships of Chelicerata, one of two nodes explicitly mentioned by Mongiardino Koch (2021) as an undesirable test case for reason of controversial relationships (*sortR* ideally requires a resolved species tree for calculation of RF distances, though this requirement can be circumvented by collapsing controversial nodes, as performed herein). Given the extensive discordance of signal across datasets at the base of Euchelicerata, we used the placement of Pseudoscorpiones as our benchmark for phylogenetic accuracy (Fig. 2), as the membership of this long-branch order within Arachnopulmonata is strongly substantiated by rare genomic changes (Ontano et al. 2021).

Of the three strategies, we found that subsampling loci for taxon occupancy to be the least effective strategy in overcoming LBA *viz.* the placement of pseudoscorpions (Fig. 1a). No matrix constructed for optimizing taxon occupancy was able to recover Panscorpiones or Arachnopulmonata, either under a partitioned model or a site heterogeneous model analysis (Fig. S1a). Subsampling by evolutionary rate was the most effective strategy for recovering pseudoscorpions within Arachnopulmonata, with addition of noisier (i.e., faster evolving) genes causing pseudoscorpions to be pulled out of Panscorpiones and towards the root of Arachnopulmonata (Figs. 1b, S1b). Nevertheless, analyses of all matrices constructed on the basis of evolutionary rate were able to recover Arachnopulmonata (with pseudoscorpions either sister group to scorpions or to the remaining arachnopulmonates).

Subsampling by phylogenetic usefulness was of intermediate effectiveness for recovering Arachnopulmonata. For both partitioned model and site heterogeneous model analyses, only the

200 most “useful” genes were able to recover Panscorpiones and Arachnopulmonata; the 400 most “useful” genes recovered Arachnopulmonata, but not Panscorpiones; and the 600 most “useful” genes recovered Pseudoscorpiones in a basally branching position near the root of Euchelicerata (Figs. 1c, S1c). These results echo the conclusion of Mongiardino Koch (2021) that subsampling by phylogenetic usefulness may not be a universally effective strategy for ancient rapid radiations that include taxa with high heterogeneity of evolutionary rates.

We additionally observed the phenomenon of LBA artifacts within long-branch taxa. Within Parasitiformes, Opilioacariformes were resolved as the sister group to the remaining Parasitiformes (i.e., the traditional placement, based on morphology) only in the smallest matrices (the 200 most complete loci; the 200 slowest-evolving loci; and the 200 most useful loci; Figs. 1, S1). The addition of noisier genes, under any of the three criteria for matrix construction, destabilized this topology and recovered the long-branch Mesostigmata as the sister group to the remaining Parasitiformes. We postulate that future efforts to address the relationships within Acariformes and Parasitiformes consider the possibility that asymmetric rates of evolution in nested lineages may further exacerbate LBA artifacts in these groups.

Conclusion

The new datasets we generated show that Palpigradi and Opilioacariformes have opposing effects on chelicerate phylogeny. Palpigradi are demonstrably an unstable taxon that destabilize interordinal chelicerate relationships, despite high data occupancy and the sampling of the deepest node within the palpigrade crown-group. Opilioacariformes are a slowly evolving group of Parasitiformes and their inclusion invariably drives the dissolution of Acari, suggesting that Acari monophyly reflects another LBA artifact in chelicerate phylogeny. Aside from expanding taxonomic sampling, subsampling with slowly evolving genes may be an effective solution to mitigating LBA artifacts in chelicerate phylogeny.

Author Contributions

A.Z.O and P.P.S conceived of the study. A.Z.O and P.P.S collected specimens of Opilioacariformes and Palpigradi in the field, and performed the sequencing. A.Z.O and H.G.S implemented phylogenomic analyses. A.Z.O and P.P.S wrote the manuscript, and all authors edited and approved the final content.

Acknowledgements

We are indebted to Emily V.W. Setton and Jesús A. Ballesteros for assistance with collecting and identifying Palpigradi and Opilioacariformes. Discussions with Nicolás Mongiardino Koch greatly facilitated the design of analyses with *sortR.* Hugh Steiner was supported by a UW-Madison Letters and Sciences Honors Sophomore Summer Apprenticeship. This material is based on work supported by the National Science Foundation under Grants IOS-1552610 and IOS-2019141.

References

Ballesteros, J. A., & Hormiga, G. (2016). A New Orthology Assessment Method for Phylogenomic Data: Unrooted Phylogenetic Orthology. *Molecular Biology and Evolution*, *33*(8), 2117–2134. https://doi.org/10.1093/molbev/msw069

Ballesteros, J. A., López, C. E. S., Kováč, L., Gavish-Regev, E., & Sharma, P. P. (2019). Ordered phylogenomic subsampling enables diagnosis of systematic errors in the placement of the enigmatic arachnid order Palpigradi. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1917). https://doi.org/10.1098/rspb.2019.2426

Ballesteros, J. A., Setton, E. V. W., Santibáñez-López, C. E., Arango, C. P., Brenneis, G., Brix, S., Corbett, K. F., Cano-Sánchez, E., Dandouch, M., Dilly, G. F., Eleaume, M. P., Gainett, G., Gallut, C., McAtee, S., McIntyre, L., Moran, A. L., Moran, R., López-González, P. J., Scholtz, G., … Sharma, P. P. (2021). Phylogenomic resolution of sea spider diversification through integration of multiple data classes. *Molecular Biology and Evolution*, *38*(2), 686–701. https://doi.org/10.1093/molbev/msaa228

Ballesteros, J. A., & Sharma, P. P. (2019). A Critical Appraisal of the Placement of Xiphosura (Chelicerata) with Account of Known Sources of Phylogenetic Error. *Systematic Biology*, *0*(0), 1–14. https://doi.org/10.1093/sysbio/syz011

Benavides, L. R., Cosgrove, J. G., Harvey, M. S., & Giribet, G. (2019). Phylogenomic interrogation resolves the backbone of the Pseudoscorpiones tree of life. *Molecular Phylogenetics and Evolution*, *139*, 106509. https://doi.org/10.1016/j.ympev.2019.05.023

Bergsten, J. (2005). A review of long-branch attraction Introduction to long-branch attraction. *Cladistics*, *21*, 163–193. http://materiais.dbio.uevora.pt/Evo/A_review_of_long-branch_attraction.pdf

Borner, J., Rehm, P., Schill, R. O., Ebersberger, I., & Burmester, T. (2014). A transcriptome approach to ecdysozoan phylogeny. *Molecular Phylogenetics and Evolution*, *80*(1), 79–87. https://doi.org/10.1016/j.ympev.2014.08.001

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Chernomor, O., Von Haeseler, A., & Minh, B. Q. (2016). Terrace Aware Data Structure for Phylogenomic Inference from Supermatrices. *Systematic Biology*, *65*(6), 997–1008. https://doi.org/10.1093/sysbio/syw037

Dunlop, J. A., Krüger, J., & Alberti, G. (2012). The sejugal furrow in camel spiders and acariform mites. *Arachnologische Mitteilungen*, *43*, 29–36. https://doi.org/10.5431/aramit4303

Felsenstein, J. (1978). Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Biology*, *27*(4), 401–410. https://doi.org/10.1093/sysbio/syp096

Fernández, R., Kallal, R. J., Dimitrov, D., Ballesteros, J. A., Arnedo, M. A., Giribet, G., & Hormiga, G. (2018). Phylogenomics, Diversification Dynamics, and Comparative Transcriptomics across the Spider Tree of Life. *Current Biology*, *28*(9), 1489-1497.e5. https://doi.org/10.1016/j.cub.2018.03.064

Fernández, R., Sharma, P. P., Tourinho, A. L., & Giribet, G. (2017). The opiliones tree of life: Shedding light on harvestmen relationships through transcriptomics. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1849). https://doi.org/10.1098/rspb.2016.2340

Giribet, G., Edgecombe, G. D., & Wheeler, W. C. (2001). Arthropod phylogeny based on eight loci. *Nature*, *413*(September), 157–161.

Giribet, G., McIntyre, E., Christian, E., Espinasa, L., Ferreira, R. L., Francke, Ó. F., Harvey, M. S., Isaia, M., Kováč, L., McCutchen, L., Souza, M. F. V. R., & Zagmajster, M. (2014). The first phylogenetic analysis of Palpigradi (Arachnida) - The most enigmatic arthropod order. *Invertebrate Systematics*, *28*(4), 350–360. https://doi.org/10.1071/IS13057

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Philip, D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Macmanes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., … Regev, A. (2013). De novo transcript sequence recostruction from RNA-Seq: reference generation and analysis with Trinity. In *Nature protocols* (Vol. 8, Issue 8). https://doi.org/10.1038/nprot.2013.084.De

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, *35*(2), 518–522. https://doi.org/10.1093/molbev/msx281

Howard, R. J., Puttick, M. N., Edgecombe, G. D., & Lozano-Fernandez, J. (2020). Arachnid monophyly: Morphological, palaeontological and molecular support for a single terrestrialization within Chelicerata. *Arthropod Structure and Development*, *59*, 100997. https://doi.org/10.1016/j.asd.2020.100997

Hoy, M. A., Waterhouse, R. M., Wu, K., Estep, A. S., Ioannidis, P., Palmer, W. J., Pomerantz, A. F., Simão, F. A., Thomas, J., Jiggins, F. M., Murphy, T. D., Pritham, E. J., Robertson, H. M., Zdobnov, E. M., Gibbs, R. A., & Richards, S. (2016). Genome Sequencing of the Phytoseiid Predatory Mite Metaseiulus occidentalis Reveals Completely Atomized Hox Genes and Superdynamic Intron Evolution. *Genome Biology and Evolution*, *8*(6), 1762–1775. https://doi.org/10.1093/gbe/evw048

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Klimov, P. B., Oconnor, B. M., Chetverikov, P. E., Bolton, S. J., Pepato, A. R., Mortazavi, A. L., Tolstikov, A. V, Bauchan, G. R., & Ochoa, R. (2018). Molecular Phylogenetics and Evolution Comprehensive phylogeny of acariform mites (Acariformes) provides insights on the origin of the four-legged mites (Eriophyoidea), a long branch. *Molecular Phylogenetics and Evolution*, *119*(October 2017), 105–117. https://doi.org/10.1016/j.ympev.2017.10.017

Kriventseva, E. V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F. A., & Zdobnov, E. M. (2019). OrthoDB v10: Sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Research*, *47*(D1), D807–D811. https://doi.org/10.1093/nar/gky1053

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

Laumer, C. E., Fernández, R., Lemer, S., Combosch, D., Kocot, K. M., Riesgo, A., Andrade, S. C. S., Sterrer, W., Sørensen, M. V., Giribet, G., Laumer, C. E., Ferna, R., Andrade, C. S., Combosch, D., Kocot, K. M., Riesgo, A., Sterrer, W., Sørensen, M. V., & Giribet, G.

(2019). Revisiting metazoan phylogeny with genomic sampling of all phyla. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1906). https://doi.org/10.1098/rspb.2019.0831

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

Masta, S. E., Longhorn, S. J., & Boore, J. L. (2009). Arachnid relationships based on mitochondrial genomes: Asymmetric nucleotide and amino acid bias affects phylogenetic analyses. *Molecular Phylogenetics and Evolution*, *50*(1), 117–128. https://doi.org/10.1016/j.ympev.2008.10.010

Mongiardino Koch, N. (2021). Phylogenomic Subsampling and the Search for Phylogenetically Reliable Loci. *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msab151

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Nolan, E. D., Santibáñez-López, C. E., & Sharma, P. P. (2020). Developmental gene expression as a phylogenetic data class: support for the monophyly of Arachnopulmonata. *Development Genes and Evolution*, *2*. https://doi.org/10.1007/s00427-019-00644-6

Ontano, A. Z., Gainett, G., Aharon, S., Ballesteros, J. A., Benavides, L. R., Corbett, K. F., Gavish-Regev, E., Harvey, M. S., Monsma, S., Santibáñez-López, C. E., Setton, E. V. W., Zehms, J. T., Zeh, J. A., Zeh, D. W., & Sharma, P. P. (2021). Taxonomic Sampling and Rare Genomic Changes Overcome Long-Branch Attraction in the Phylogenetic Placement of Pseudoscorpions. *Molecular Biology and Evolution*, *38*(6), 2446–2467. https://doi.org/10.1093/molbev/msab038

Pepato, A. R., da Rocha, C. E. F., & Dunlop, J. A. (2010). Phylogenetic position of the acariform mites: sensitivity to homology assessment under total evidence. *BMC Evolutionary Biology*, *10*, 235. https://doi.org/10.1186/1471-2148-10-235

Regier, J. C., Shultz, J. W., Zwick, A., Hussey, A., Ball, B., Wetzer, R., Martin, J. W., & Cunningham, C. W. (2010). Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature*, *463*(7284), 1079–1083. https://doi.org/10.1038/nature08742

Roure, B., Baurain, D., & Philippe, H. (2013). Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. *Molecular Biology and Evolution*, *30*(1), 197–214. https://doi.org/10.1093/molbev/mss208

Salichos, L., & Rokas, A. (2013). Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature*, *497*(7449), 327–331. https://doi.org/10.1038/nature12130

Santibáñez-López, C. E., González-Santillán, E., Monod, L., & Sharma, P. P. (2019). Phylogenomics facilitates stable scorpion systematics: Reassessing the relationships of Vaejovidae and a new higher-level classification of Scorpiones (Arachnida). *Molecular*

Phylogenetics and Evolution, *135*, 22–30. https://doi.org/10.1016/j.ympev.2019.02.021

Santibáñez-López, C. E., Ojanguren-Affilastro, A. A., & Sharma, P. P. (2020). Another one bites the dust: Taxonomic sampling of a key genus in phylogenomic datasets reveals more non-monophyletic groups in traditional scorpion classification. *Invertebrate Systematics*, *34*(2), 133–143. https://doi.org/10.1071/IS19033

Seiter, M., Schwaha, T., Ferreira, R. L., Prendini, L., & Wolff, J. O. (2021). Fine structure of the epicuticular secretion coat and associated glands of Pedipalpi and Palpigradi (Arachnida). *Journal of Morphology*, *282*(8), 1158–1169. https://doi.org/10.1002/jmor.21360

Sharma, P. P., Fernández, R., Esposito, L. A., Gonzalez-Santillan, E., & Monod, L. (2015). Phylogenomic resolution of scorpions reveals multilevel discordance with morphological phylogenetic signal. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1804), 20142953–20142953. https://doi.org/10.1098/rspb.2014.2953

Sharma, P. P., Kaluziak, S. T., Pérez-Porro, A. R., González, V. L., Hormiga, G., Wheeler, W. C., & Giribet, G. (2014). Phylogenomic interrogation of arachnida reveals systemic conflicts in phylogenetic signal. *Molecular Biology and Evolution*, *31*(11), 2963–2984. https://doi.org/10.1093/molbev/msu235

Shultz, J. W. (1990). Evolutionary morphology and phylogeny of Arachnida. *Cladistics*, *6*, 1–38.

Shultz, J. W. (2007). A phylogenetic analysis of the arachnid orders based on morphological characters. In *Zoological Journal of the Linnean Society* (Vol. 150, Issue 2). https://doi.org/10.1111/j.1096-3642.2007.00284.x

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Wang, H. C., Minh, B. Q., Susko, E., & Roger, A. J. (2018). Modeling Site Heterogeneity with Posterior Mean Site Frequency Profiles Accelerates Accurate Phylogenomic Estimation. *Systematic Biology*, *67*(2), 216–235. https://doi.org/10.1093/sysbio/syx068

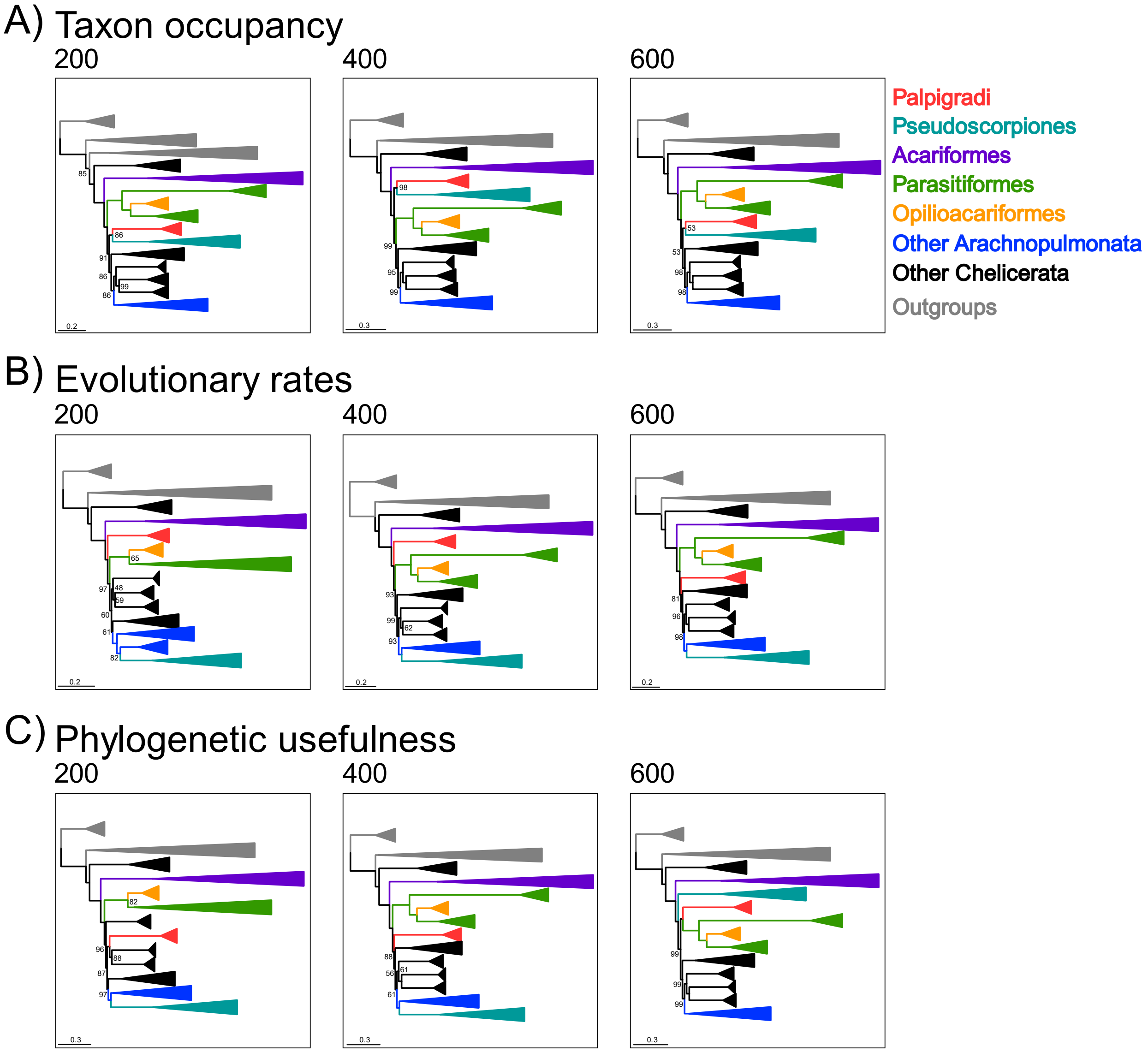
Waterhouse, R. M., Seppey, M., Simao, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V., & Zdobnov, E. M. (2018). BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution*, *35*(3), 543–548. https://doi.org/10.1093/molbev/msx319

Weygoldt, P., & Paulus, H. F. (1979). Untersuchungen zur Morphologie, Taxonomie und Phylogenie der Chelicerata II. Cladogramme und die Entfaltung der Chelicerata. *Journal of Zoological Systematics and Evolutionary Research*, *17*(3), 177–200. https://doi.org/10.1111/j.1439-0469.1979.tb00699.x

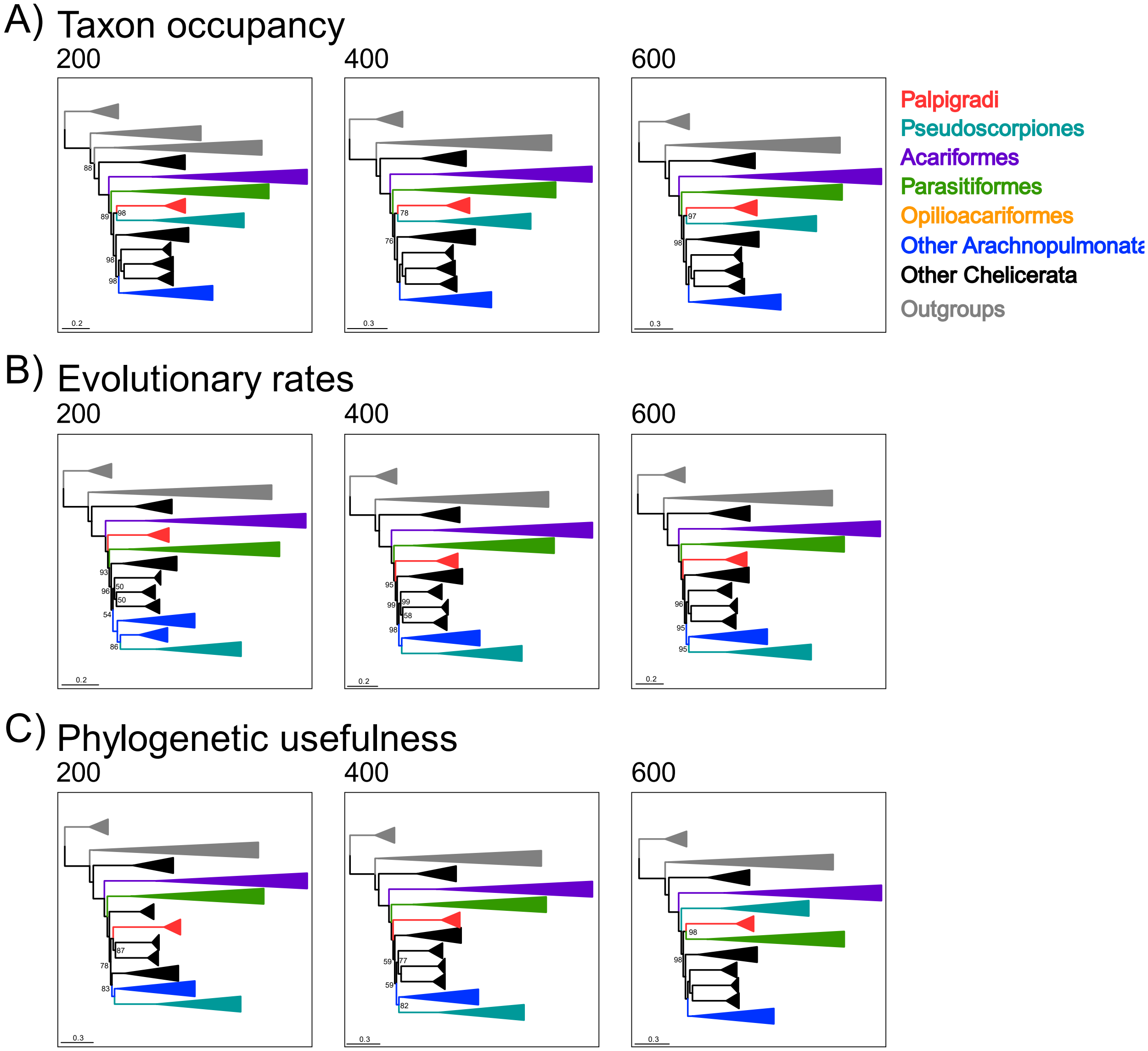
Wheeler, W. C., & Hayashi, C. Y. (1998). The Phylogeny of the Extant Chelicerate Orders. *Cladistics*, *14*(2), 173–192. https://doi.org/10.1111/j.1096-0031.1998.tb00331.x

Supplementary Figure and table legends

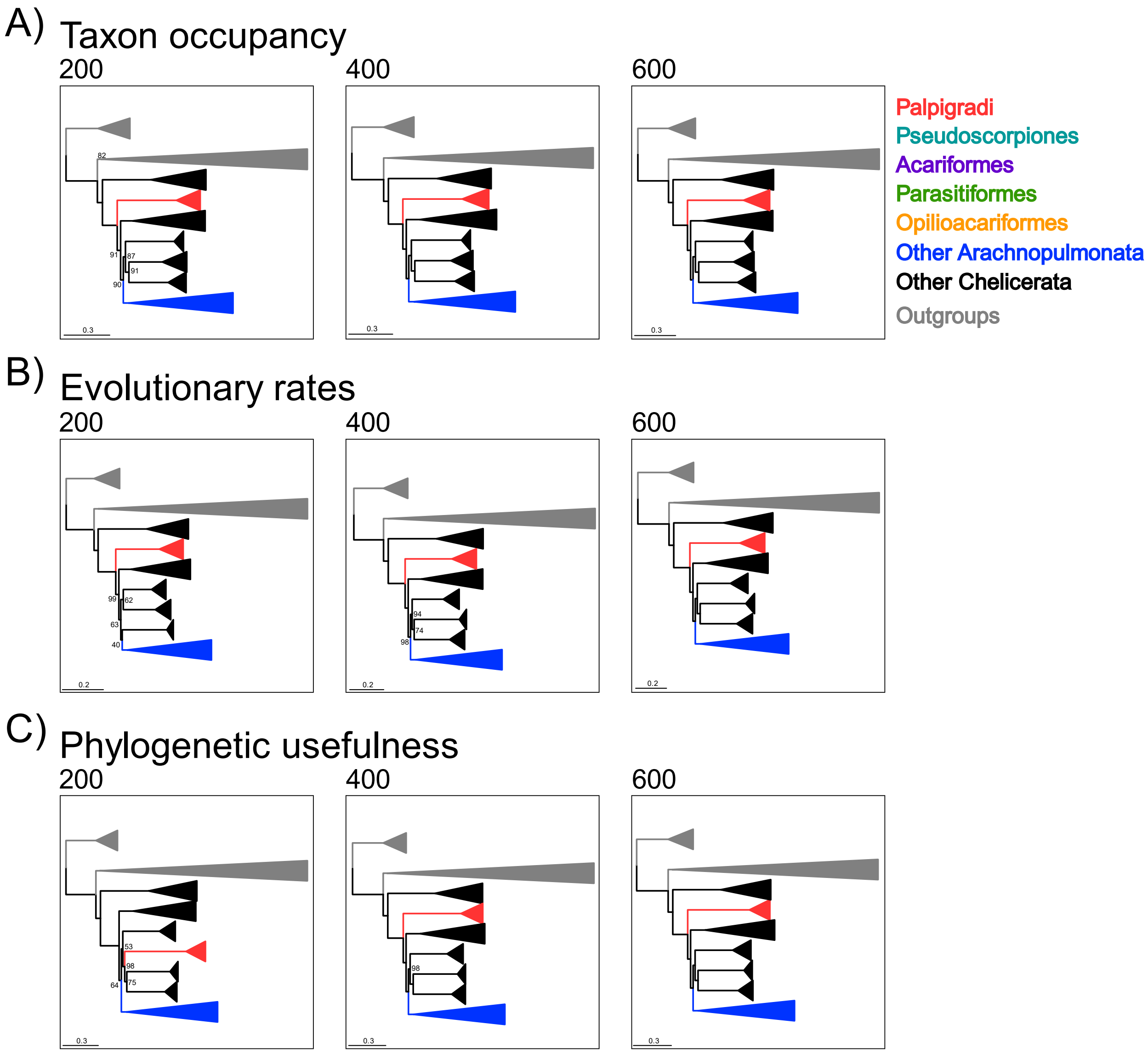
Supplementary Figure S1. Maximum likelihood analyses under site heterogeneous models (LG + C20 + F + Γ), selected under alternative optimality criteria. (a) Ordered by taxon occupancy. (b) Ordered by evolutionary rate. (c) Ordered by phylogenetic usefulness (output of *sortR*). Colors correspond to legend on the right. Ultrafast bootstrap support values under 100% are labeled next to their corresponding node; unlabeled nodes are maximally supported. Scale bar: 0.2.



Supplementary Figure S2. Taxon deletion experiments with Opilioacariformes removed, under the LG + C20 + F + Γ model. Note that this family of analyses also does not recover Acari monophyly (compare to Figure 3). Maximum likelihood trees are shown with same arrangement of matrix optimality criteria as in Figure S1. Colors correspond to legend on the right. Ultrafast bootstrap support values under 100% are labeled next to their corresponding node; unlabeled nodes are maximally supported. Scale bar: 0.2.

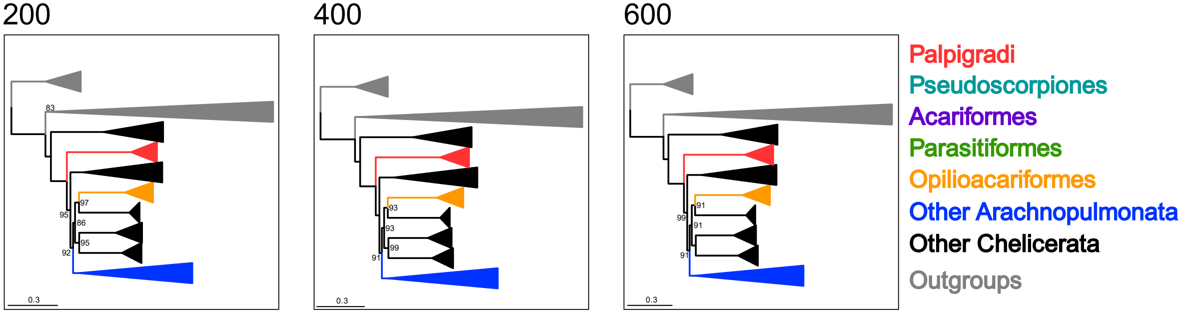


Supplementary Figure S3. Taxon deletion experiments that remove all long-branch orders except Palpigradi, under the LG + C20 + F + Γ model. Maximum likelihood trees are shown with same arrangement of matrix optimality criteria as in Figure S1. Colors correspond to legend on the right. Ultrafast bootstrap support values under 100% supported. Scale bar: 0.2.

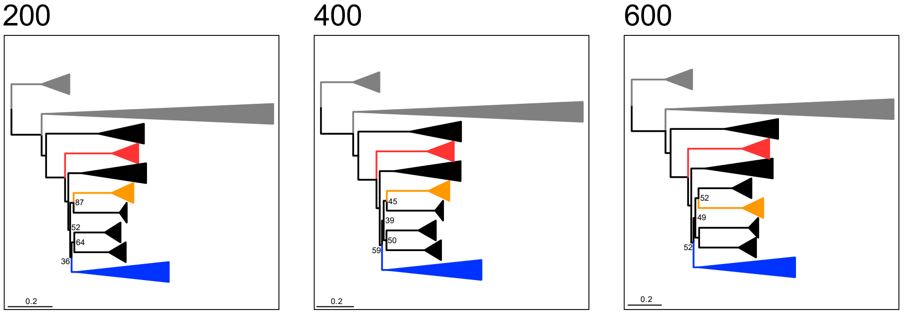


Supplementary Figure S4. Maximum likelihood trees with a deletion of all long-branch taxa except Palpigradi and Opilioacariformes, under the LG + C20 + F + Γ model. Maximum likelihood trees are shown with same arrangement of matrix optimality criteria as in Figure S1. Colors correspond to legend on the right. Ultrafast bootstrap support values under 100% are labeled next to their corresponding node; unlabeled nodes are maximally supported. Scale bar: 0.2.

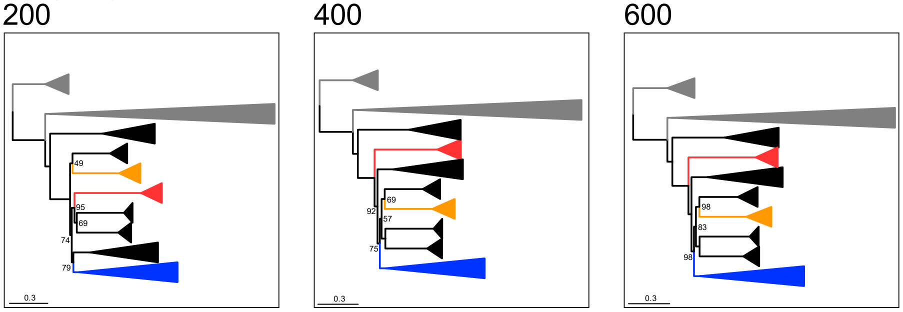
A) Taxon occupancy



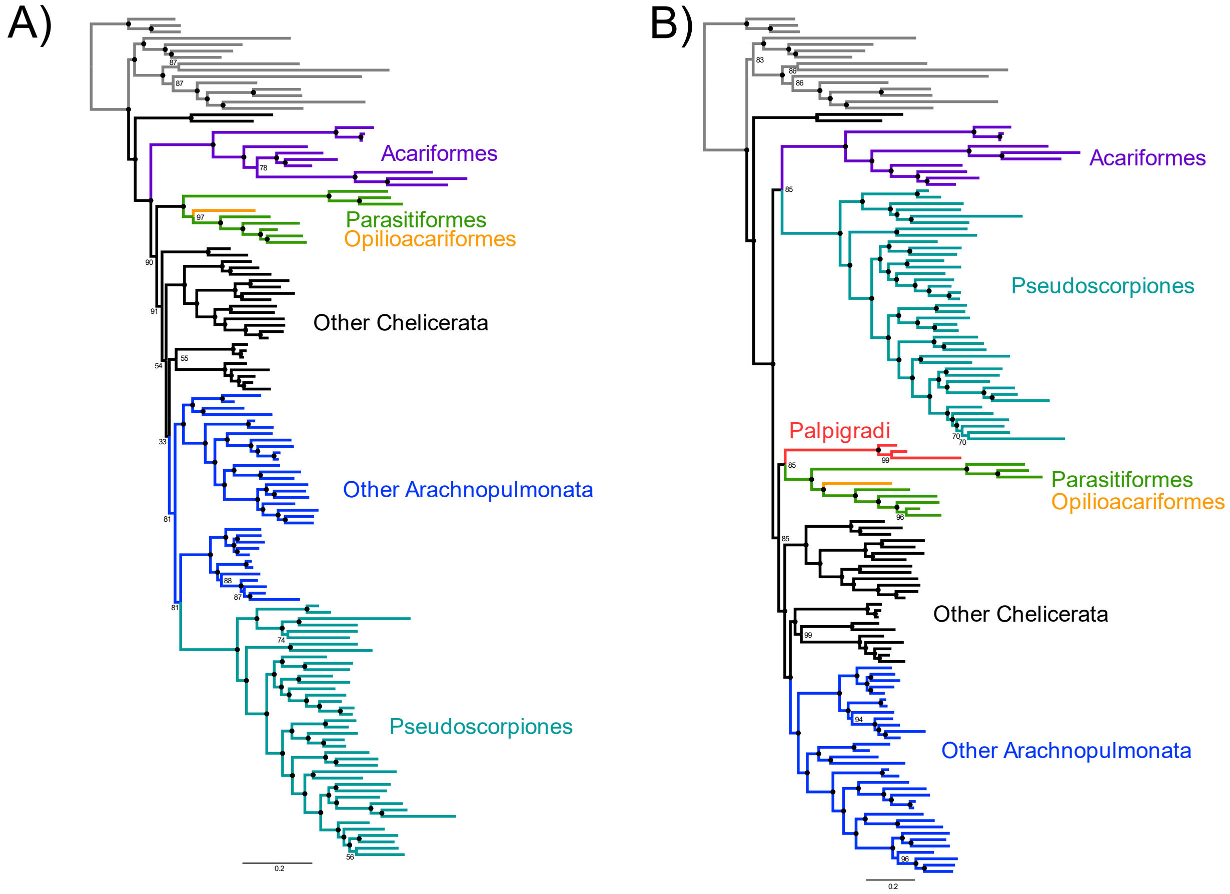
B) Evolutionary rates



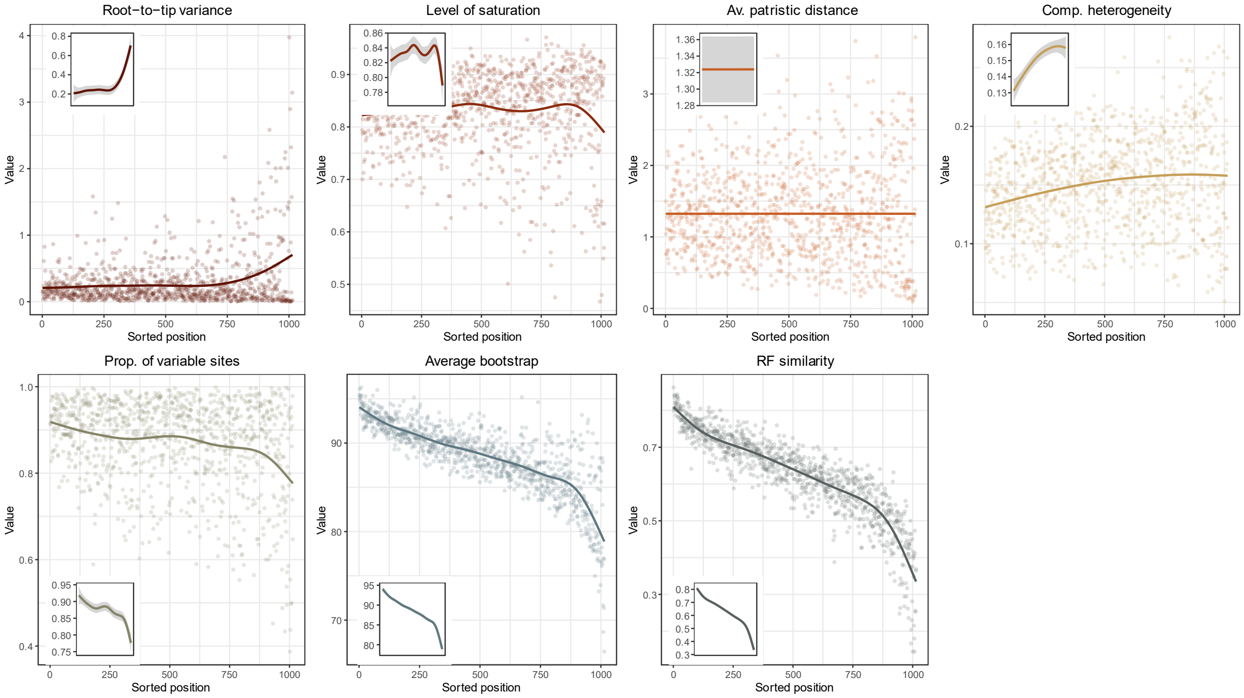
C) Phylogenetic usefulness



Supplementary Figure S5. Maximum likelihood tree of the $G_3 \bullet T_0$ matrix of Ontano et al. (2021), with the addition of three Palpigradi libraries from this study, under a partitioned model approach. Note that the addition of palpigrades in maximum likelihood analyses reinforces the long-branch attraction artifact of pseudoscorpions, breaking the monophyly of Panscorpiones and Arachnopulmonata. Ultrafast bootstrap support values below 100% are labeled next to their corresponding node; unlabeled nodes are maximally supported.



Supplementary Figure S6. Graphical output of *sortR* pipeline for ranking of gene loci, based on a phylogenetic usefulness criteria.



Supplementary Table S1. List of species sampled and GenBank accession data. Locality data for newly collected specimens are provided in text.

Higher taxon	Family	Species	NCBI Accession
Acariformes	Acaridae	Rhizoglyphus robini	SRX330812
Acariformes	Achipteriidae	Achipteria coleoptrata	SRR4039018
Acariformes	Hermanniidae	Hermannia gibba	SRX2030131
Acariformes	Hypchthoniidae	Hypochthonius rufulus	SRR4039020
Acariformes	Nothridae	Nothrus palustris	SRR4039021
Acariformes	Pyroglyphidae	Dermatophagoides farinae	SRR1016494
Acariformes	Sarcoptidae	Sarcoptes scabiei	GCA_000828355.1
Acariformes	Tetranychidae	Panonychus citri	SRR341928
Acariformes	Tetranychidae	Tetranychus cinnabarinus	SRR519097
Acariformes	Tetranychidae	Tetranychus urticae	GCA_000239435.1
Amblypygi	Charinidae	Charinus ioanniticus	PRJNA649577
Amblypygi	Charinidae	Charinus israelensis	PRJNA649577
Amblypygi	Phrynichidae	Damon variegatus	SRX450910
Amblypygi	Phrynidae	Phrynus marginemaculatus	PRJNA646521
Amphipdoa	Hyallellidae	Hyalella azteca	GCA_000764305.2
Araneae	Anthrodiaetidae	Aliatypus sp.	SRX652492
Araneae	Araneidae	Neoscona arabesca	SRX4510071
Araneae	Cybaeidae	Calymmaria persica	SRX1560159
Araneae	Eresidae	Stegodyphus mimosarum	ERX048741
Araneae	Filistatidae	Kukulcania sp.	SRX652494
Araneae	Idiopidea	Idiops bersebaensis	SRX652523
Araneae	Linyphiidae	Frontinella communis	SRX652520
Araneae	Liphistiidae	Liphistius malayanus	SRX450965
Araneae	Liphistiidae	Liphistius sp.	SRX652489
Araneae	Microhexuridae	Microhexura montivaga	SRX652506
Araneae	Oxyopidae	Peucetia longipalpis	SRX652514
Araneae	Pholcidae	Pholcus phalangoides	SRR3144082
Araneae	Salticidae	Habronattus ustulatus	SRX763246
Araneae	Scytodidae	Scytodes thoracica	SRX652514
Araneae	Sicariidae	Loxosceles deserta	SRX1560147
Araneae	Theraphosidae	Acanthoscurria geniculata	SRX373163
Araneae	Theraphosidae	Aphonopelma hentzi	SRX10000055
Araneae	Theridiidae	Parasteatoda tepidariorum	GCA_000365465.3
Cladocera	Daphniidae	Daphnia pulex	GCA_900092285.2
Coleoptra	Tenebrionidae	Tribolium castaneum	GCA_000002335.3
Diptera	Drosophilidae	Drosophila melanogaster	GCA_002310775.1
Entomobryomorpha	Isotomidae	Folsomia candida	GCA_002217175.1
Geophilomorpha	Linotaeniidae	Strigamia maritima	GCA_000239455.1
Glomeria	Glomeridae	Glomeris marginata	SRX1638914
Hymenoptera	Pteromalidae	Nasonia vitripennis	GCA_004768525.1
Isoptera	Archotermopsidae	Zootermopsis nevadensis	GCA_000696155.1
Lithobiomorpha	Lithobiidae	Lithobius atkinsoni	SRX4717996
Onychophora	Peripatidae	Epiperipatus sp.	SRX5426511
Onychophora	Peripatopsidae	Opisthopatus kwazululandi	SRX5131543
Onychophora	Peripatopsidae	Peripatopsis overbergiensis	SRX451023
Opiliones	Caddidae	Caddo agilis	SAMN06309539

Higher taxon	Family	Species	NCBI Accession
Opiliones	Ceratolasmatidae	Hesperonemastoma modestum	SRX450937
Opiliones	Neogoveidae	Metasiro savannahensis	SRX205297
Opiliones	Pettalidae	Aoraki denticulata	SRX6374211
Opiliones	Phalangiidae	Phalangium opilio	SRX450969
Opiliones	Protolophidae	Protolophus singularis	SRX450934
Opiliones	Sclerisimatidae	Leiobunum verrucosum	SRX450936
Opiliones	Sironidae	Siro boyerae	SRX450933
Opiliones	Stygnopsidae	Karos barbarikos	PRJNA556673
Opiliones	Triaenonychidae	Larifuga capensis	SRX451010
Opiliones	Triaenonychidae	Sclerobunus nondimorphicus	SRX647445
Opiliones	Trogulidae	Trogulus martensi	SRX450964
Opiliones	Zalmoxidae	Pachylicus acutus	SRX451775
Palpigradi	Eukoeneniidae	Eukonenia hanseni	pending
Palpigradi	Eukoeneniidae	Eukonenia spelaea	SRX4907783
Palpigradi	Prokoeneniidae	Prokoenenia wheeleri	pending
Pantopoda	Phoxichilidae	Anoplodactylus insignis	SRX2544807
Pantopoda	Pycnogonidae	Pycnogonum litorale	SRX4717995
Parasitiformes	Opilioacaridae	Neacarus sp.	pending
Parasitiformes	Opilioacaridae	Opilioacarus sp.	pending
Parasitiformes	Opilioacaridae	Adenacarus sp.	SRX4907784
Parasitiformes	Argasidae	Ornithodoros rostratus	SRR1732011
Parasitiformes	Ixodidae	Amblyomma americanum	SRX1710180
Parasitiformes	Ixodidae	Hyalomma excavatum	SRR3157672
Parasitiformes	Ixodidae	Ixodes scapularis	GCA_000208615.1
Parasitiformes	Ixodidae	Rhipicephalus microplus	GCA_002176555.1
Parasitiformes	Laelapidae	Pneumolaelaps niutirani	SRX6983067
Parasitiformes	Laelapidae	Tropilaelaps mercedesae	GCA_002081605.1
Parasitiformes	Phytoseiidae	Galendromus occidentalis	GCA_000255335.1
Parasitiformes	Varroidae	Varroa destructor	SRR3927486
Pseudoscorpiones	Atemnidae	Oratemnus curtus	SRX6098520
Pseudoscorpiones	Bochicidae	Bochica withi	SRX6098509
Pseudoscorpiones	Cheiridiidae	Cheiridiidae sp.	SRX6098503
Pseudoscorpiones	Cheliferidae	Chelifer cancroides	SRX5086902
Pseudoscorpiones	Cheliferidae	Cheliferidae sp.	SRX6098504
Pseudoscorpiones	Cheliferidae	Parachelifer persimilis	SRX6098521
Pseudoscorpiones	Cheliferidae	Protochelifer sp.	SRX6098515
Pseudoscorpiones	Chernetidae	Chernetidae sp. A	SRX6098511
Pseudoscorpiones	Chernetidae	Conicochernes crassus	SRX6098529
Pseudoscorpiones	Chernetidae	Hesperochernes sp.	SRX652493
Pseudoscorpiones	Chernetidae	Lamprochernes savignyi	SRX6098494
Pseudoscorpiones	Chthoniidae	Chthoniidae sp.	SRX6098512
Pseudoscorpiones	Chthoniidae	Ephippiochthonius tetrachelatus	SRX6098528
Pseudoscorpiones	Chthoniidae	Lagynochthonius australicus	SRX6098495
Pseudoscorpiones	Chthoniidae	Lechytia hoffi	SRX6098500
Pseudoscorpiones	Feaellidae	Feaella (Tetrafeaella) capensis	SRX6098525
Pseudoscorpiones	Garypidae	Anchigarypus californicus	SRX6098526

Higher taxon	Family	Species	NCBI Accession
Pseudoscorpiones	Garypidae	Synsphyronus apimelus	SRX450967
Pseudoscorpiones	Garypinidae	Protogarypinus giganteus	SRX6098516
Pseudoscorpiones	Garypinidae	Pseudogarypinus frontalis	SRX6098517
Pseudoscorpiones	Geogarypidae	Afrogarypus subimpressus	SRX6098507
Pseudoscorpiones	Geogarypidae	Geogarypus maculatus	SRX6098523
Pseudoscorpiones	Gymnobisiidae	Gymnobisium sp.	SRX6098524
Pseudoscorpiones	Hesperolpiidae	Apolpium sp.	SRX6098508
Pseudoscorpiones	Hyidae	Indohya sp.	SRX6098496
Pseudoscorpiones	Ideoroncidae	Dhanus sumatranus	SRX6098527
Pseudoscorpiones	Larcidae	Larca granulata	SRX6098493
Pseudoscorpiones	Neobisiidae	Microbisium brunneum	SRX6098498
Pseudoscorpiones	Neobisiidae	Microbisium sp.	SRX6098497
Pseudoscorpiones	Neobisiidae	Microcreagrinae sp.	SRX6098502
Pseudoscorpiones	Neobisiidae	Novobisium sp.	SRX6098519
Pseudoscorpiones	Parahyidae	Parahya submersa	SRX6098522
Pseudoscorpiones	Pseudogarypidae	Pseudogarypus banksi	SRX6098518
Pseudoscorpiones	Pseudotyrannochthoniidae	Pseudotyrannochthonius sp. A	SRX6098513
Pseudoscorpiones	Pseudotyrannochthoniidae	Pseudotyrannochthonius sp. B	SRX6098514
Pseudoscorpiones	Sternophoridae	Sternophoridae sp.	SRX6098491
Pseudoscorpiones	Syarinidae	Ideobisium crassimanum	SRX6098532
Pseudoscorpiones	Withiidae	Cacodemonius segmentidentatus	SRX6098510
Ricinulei	Ricinoididae	Cryptocellus becki	SRX998571
Ricinulei	Ricinoididae	Cryptocellus sp. A	SRX1001105
Ricinulei	Ricinoididae	Cryptocellus sp. B	SRX1001105
Ricinulei	Ricinoididae	Pseudocellus pearsei	SRX451793
Ricinulei	Ricinoididae	Pseudocellus sp.	pending
Ricinulei	Ricinoididae	Ricinoides atewa	SRX451011
Ricinulei	Ricinoididae	Ricinoides karschii	pending
Schizomida	Hubbardiidae	Stenochrus portoricensis	SRX3932195
Scolopendromorpha	Scolopendridae	Alipes grandidieri	SRX205685
Scolopendromorpha	Scolopendridae	Symphyllela vulgaris	SRR6144316
Scorpiones	Buthidae	Androctonus australis	SRX815858
Scorpiones	Buthidae	Centruroides sculpturatus	GCA_000671375.2
Scorpiones	Buthidae	Parabuthus transvaalicus	SRX815875
Scorpiones	Buthidae	Tityus serrulatus	SRX1639539
Scorpiones	Caraboctonidae	Hadrurus arizonensis	SRX815831
Scorpiones	Chaerilidae	Chaerilus celebensis	SRX815898
Scorpiones	Euscorpiidae	Megacormus gertschi	SRX1837205
Scorpiones	Euscorpiidae	Scorpiops sp.	SRX848924
Scorpiones	Iuridae	Iurus dekanum	SRX815834
Scorpiones	Pseudochactidae	Troglokhammouanus steineri	SRX815856
Scorpiones	Pseudochactidae	Vietbocap lao	SRX815857
Scorpiones	Scorpionidae	Pandinus imperator	SRX815750
Solifugae	Galeodidae	Galeodes sp.	SRX4907785
Solifugae	Eremobatidae	Eremobates sp.	SRX451779
Thelyphonida	Thelyphonidae	Mastigoproctus giganteus	SRX450914

Higher taxon	Family	Species	NCBI Accession
Xiphosura	Limulidae	Carcinoscorpius rotundicauda	GCA_011833715.1
Xiphosura	Limulidae	Limulus polyphemus	GCA_000517525.1
Xiphosura	Limulidae	Tachypleus tridentatus	GCA_004210375.1

Appendix 1

Transcriptomic Analysis of Pseudoscorpion Venom Reveals a Unique Cocktail Dominated by Enzymes and Protease Inhibitors

*Carlos E. Santibáñez-López, *Andrew Z. Ontano, Mark S. Harvey and Prashant P. Sharma*

Published on April 18th, 2018 in Toxins

See attached publication.

* Contribution: I performed the fieldwork and contributed to the preparation of the manuscript.

toxins

Article

Transcriptomic Analysis of Pseudoscorpion Venom Reveals a Unique Cocktail Dominated by Enzymes and Protease Inhibitors

Carlos E. Santibáñez-López [1,2,*], Andrew Z. Ontano [1], Mark S. Harvey [3] and Prashant P. Sharma [1]

[1] Department of Integrative Biology, University of Wisconsin-Madison, 430 Lincoln Drive, Madison, WI 53706, USA; ontano@wisc.edu (A.Z.O.); prashant.sharma@wisc.edu (P.P.S.)
[2] Posgrado en Ciencias Biológicas, Universidad Nacional Autónoma de México, Av. Universidad 3000, Coyoacán, Ciudad de México C.P. 04510, Mexico
[3] Department of Terrestrial Zoology, Western Australian Museum, Locked Bag 49, Welshpool DC, Western Australia 6986, Australia; Mark.Harvey@museum.wa.gov.au
* Correspondence: santibanezlo@wisc.edu

Received: 18 April 2018; Accepted: 16 May 2018; Published: 18 May 2018

Abstract: Transcriptomic and genomic analyses have illuminated the diversity of venoms in three of the four venomous arachnid orders (scorpions, spiders, and ticks). To date, no venom gland transcriptome analysis has been available for pseudoscorpions, the fourth venomous arachnid lineage. To redress this gap, we sequenced an mRNA library generated from the venom glands of the species *Synsphyronus apimelus* (Garypidae). High-throughput sequencing by the Illumina protocol, followed by de novo assembly, resulted in a total of 238,331 transcripts. From those, we annotated 131 transcripts, which code for putative peptides/proteins with similar sequences to previously reported venom components available from different arachnid species in protein databases. Transcripts putatively coding for enzymes showed the richest diversity, followed by other venom components such as peptidase inhibitors, cysteine-rich peptides, and thyroglobulin 1-like peptides. Only 11 transcripts were found that code for putatively low molecular mass spider toxins. This study constitutes the first report of the diversity of components within pseudoscorpion venom.

Keywords: Arachnida; enzymes; kunitz-type inhibitors

Key Contribution: The first report of pseudoscorpion venom components and their importance in venom evolution within arachnids.

1. Introduction

Pseudoscorpions, commonly known as false scorpions or book scorpions, are small arachnids (0.5 mm to 5 mm) that are similar to scorpions in that they bear a pair of chelate pedipalps (pincers), but lack the characteristic stinger-bearing metasoma (tail) [1]. These animals live in almost all terrestrial habitats, commonly in leaf litter or soil, but also in caves or littoral habitats [2]. Like many arachnid orders, pseudoscorpions appeared in the fossil record of the Devonian, with the oldest crown group fossils dating back to 390 Ma [3]. Their phylogenetic position remains controversial. Early studies (e.g., [4–6]) suggested pseudoscorpions were a sister group to either mites [5] or solifugids [6]. Comparatively recent phylogenomic analyses have revealed an array of unstable placements for this order: as a problematic long-branch taxon at the base of Arachnida, as the sister group to Arachnopulmonata (Scorpiones + Tetrapulmonata), or as the sister group to scorpions [7–9].

A large clade of pseudoscorpions (Iocheirata) possess one or two venom glands within the pedipalpal fingers (used to immobilize their prey); venom glands are missing in the less diverse superfamilies Feaelloidea and Chthonioidea [2,10]. They therefore represent one of the four venomous arachnid orders (together with Acari [ticks], Araneae [spiders], and Scorpiones [scorpions]). Surprisingly, and in contrast to the remaining venomous arachnid groups, the composition of pseudoscorpion venom remains unknown. Santos et al. [11] studied the effect of the crude venom from *Paratemnoides elongatus* on a rat cerebral cortex. Their findings were suggestive of the presence of selective compounds (e.g., neurotoxins) acting in L-glu and GABA dynamics, but no specific compounds were reported.

With the advent of high-throughput sequencing, studies on the diversity of peptidic components in scorpion and spider venom have become abundant. Through this approach, scorpion and spider venoms (see References [12,13]) have been discovered to contain many toxins that modulate the gating of ion channels, and also other components such as enzymes with phospholipase and hyaluronidase activities (e.g., [14–18]). Parallel inquiries via transcriptomic analysis of the salivary glands of ticks have revealed that these animals bear a great diversity of enzymes and protease inhibitors, but a low diversity of toxins [19–21]. As a first step toward discovering the diversity of venom components of Pseudoscorpiones, we present herein the first transcriptome analysis of the venom glands of the Western Australian species *Synsphyronus apimelus* (Garypidae; Figure 1). In addition, we selected transcripts coding for putative venom peptides and searched for orthologous sequences in two existing pseudoscorpion libraries (exemplars of the family Chernetidae) to assess evolutionary conservation of venom composition within the order.

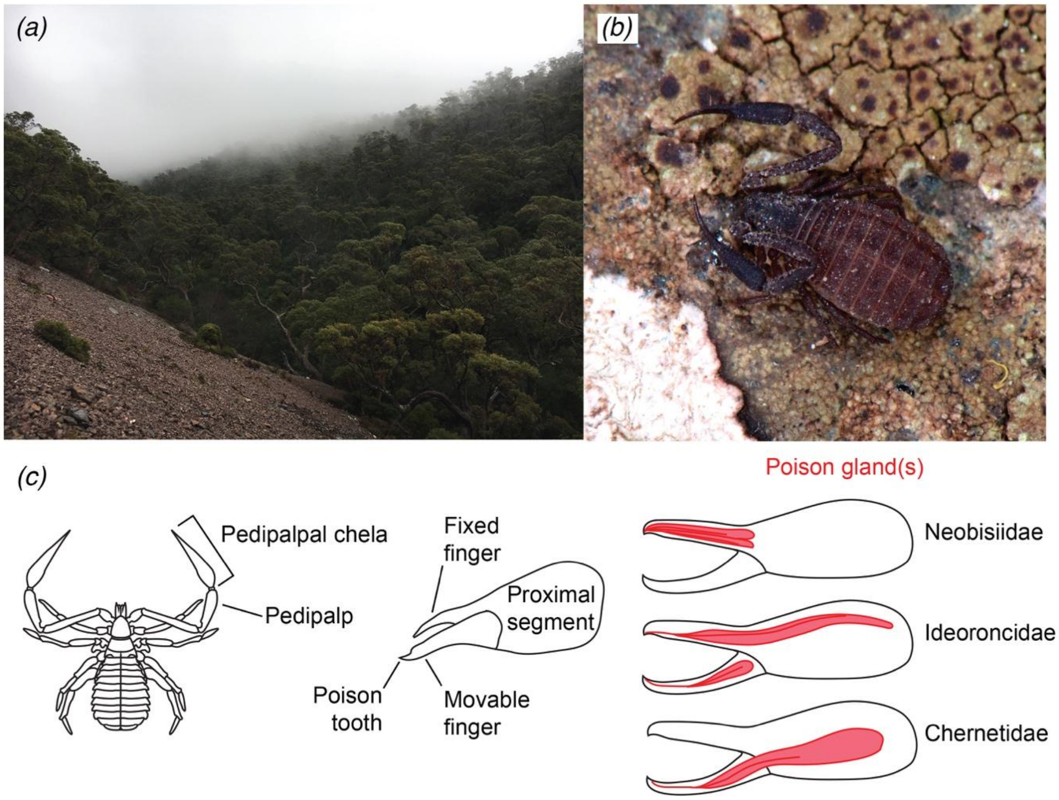


Figure 1. (**a**) Habitat of *Synsphyronus apimelus* in the Stirling Range National Park, Western Australia (photo by A.Z.O.). (**b**) Live habitus of adult *Synsphyronus apimelus* (photograph by G. Giribet, MCZ Database at https://mczbase.mcz.harvard.edu). (**c**) Schematic drawings showing the position of the venom glands in the pedipalpal chela of selected families of Iocheirata (the venomous pseudoscorpions), after References [22,23].

2. Results

The extraction of RNA from the pedipalpal chelae of *S. apimelus* yielded 3.367 µg of total RNA. After sequencing, assembly, and cleaning, 38,593,919 reads were obtained corresponding to 238,331 transcripts, 152,705 genes, and 53,483 peptides, with an N50 of 599 bp. From the transcripts, 37,148 were identified matching annotated genes listed in databases. Remarkably, only 54 were identified as matching arachnid sequences. This low number partly reflects the lack of annotated sequences in databases for arachnids, and especially so for pseudoscorpions [24]. In addition, 33,841 annotated genes were classified based on the Gene Ontology categories (GO-terms) [25,26]; the most abundant genes were those with molecular function (Figure S1). Finally, we detected 131 sequences (86 genes) which putatively code for venom components based on sequence similarity from UniProt, PFAM, or available literature (Figure 2a).

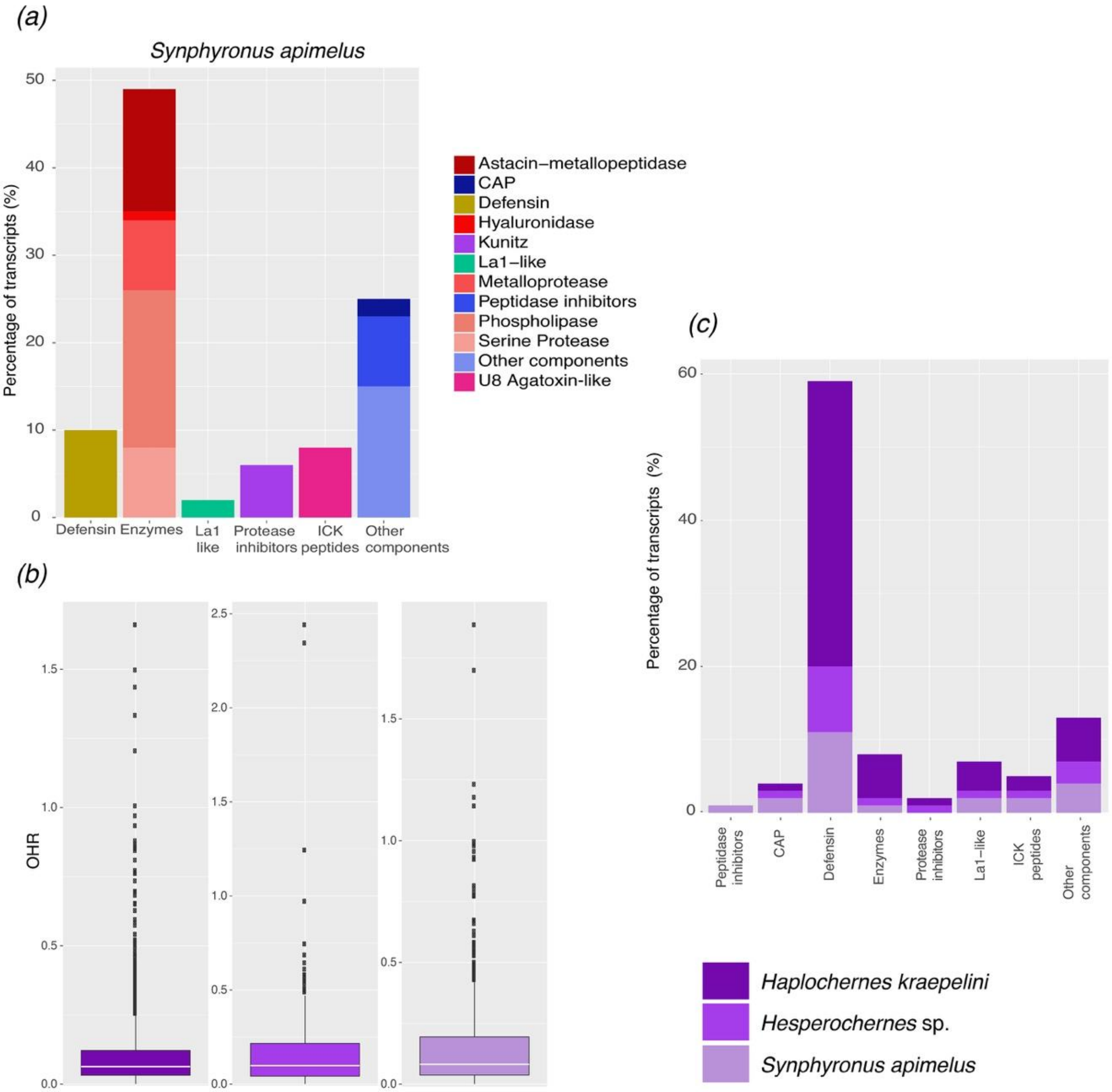


Figure 2. (**a**) Distribution of the annotated transcripts from the venom gland transcriptome of *S. apimelus* according to protein families and subfamilies. (**b**) Ortholog hit ratio (OHR) analysis showing the median (white line), and quartiles for three pseudoscorpion species. (**c**) Comparative distribution of the annotated transcripts from the transcriptomes of *S. apimelus*, *H. kraepelini*, and *Hesperochernes* sp.

2.1. Transcriptomic Analysis

2.1.1. ICK-Like Spider Venom Peptides

Toxins, generally the most widely studied venom fraction in all animals, are proteins classified according to their chemical class, biological origin, or target organ/ion channel [27]. Arachnid venoms are rich in toxins that modulate the opening of different ion channels in arthropods (mainly insects) and mammals. While high molecular mass toxins are more diverse in spider and tick venoms, low molecular mass toxins are far more diverse in scorpion venom. Here, we only found transcripts potentially coding for low molecular mass toxins in the pseudoscorpion. However, these were poorly represented in terms of sequence diversity, comprising only 11 transcripts (out of 131, 8%; Figure 2a). Within these transcripts, we discovered three sequences with 62–72% identity to the precursor of U8-agatoxin-like deduced from the genome of the spider *Parasteatoda tepidariorum*, seven sequences with 56–82% identity to the precursor of U8-agatoxin-like deduced from the genomic analysis of the scorpion *Centruroides sculpturatus*, and one sequence with 30% identity to the precursor of the U33-theraphotoxin-Cg1b deduced from cDNA cloned from the tarantula *Chilobrachys jingzhao*.

2.1.2. Protease Inhibitors

Protease inhibitors, proteins capable of inhibiting the activity of proteolytic enzymes, may play an important role in the protection of toxins from unwanted degradation [28,29]. Kunitz-type inhibitors are frequently found in arthropod venoms. In the scorpion and spider venoms, these peptides have dual functions (see also Reference [30]) as protease inhibitors and potassium channel blockers (e.g., [31]). However, in ticks, mites, and insects, these peptides only act as serine protease inhibitors. In *S. apimelus*, we discovered eight sequences (6% of the total transcripts, Figure 2a) with different percentages of similarity (ranging from 46 to 64%) to five different precursors of Kunitz-type serine proteases reported from three spiders, one scorpion, and one insect.

2.1.3. Enzymes

The most common enzymes in arachnid venom (i.e., mites, ticks, scorpions, and spiders) are hyaluronidases, metalloproteases, phospholipases, and serine proteases. Here, we report 62 sequences (48% of the total transcripts) coding putatively for the following enzymes: (a) one sequence with 34% identity to the precursor of a hyaluronidase deduced from cDNA cloned from the venom of the spider *Cupiennius salei*; (b) 17 sequences with different percentages of similarity (ranging from 42 to 73%) to six different precursors of Astacin-like metalloproteases reported from three spider, one scorpion, and one tick species; (c) seven sequences with identities ranging from 46 to 60% to two different precursors of Astacin-like metallopeptidases deduced from cDNA cloned from the venom gland of *Tityus serrulatus*; (d) two sequences of two different precursors of metalloproteinases reported from one spider and one insect; (e) 23 sequences with identities to four types of phospholipases (A2, D1, D2, and D3) reported from the venom of three spiders and two scorpions; and (f) 10 sequences with identities to six different precursors of serine proteases reported from the venom of one tick and two scorpions.

2.1.4. Single Domain von Willebrand Factor Type C Peptides (La1-Like Peptides)

La1-like peptides have been found in scorpion venom, and recently in spider venom, but their function remains unclear [32–34]. In *S. apimelus*, we found two sequences with 40–46% identity to the toxin-like protein 14 isolated and deduced from cDNA cloned from the venom of the scorpion *Urodacus yaschenkoi* (Figure 3).

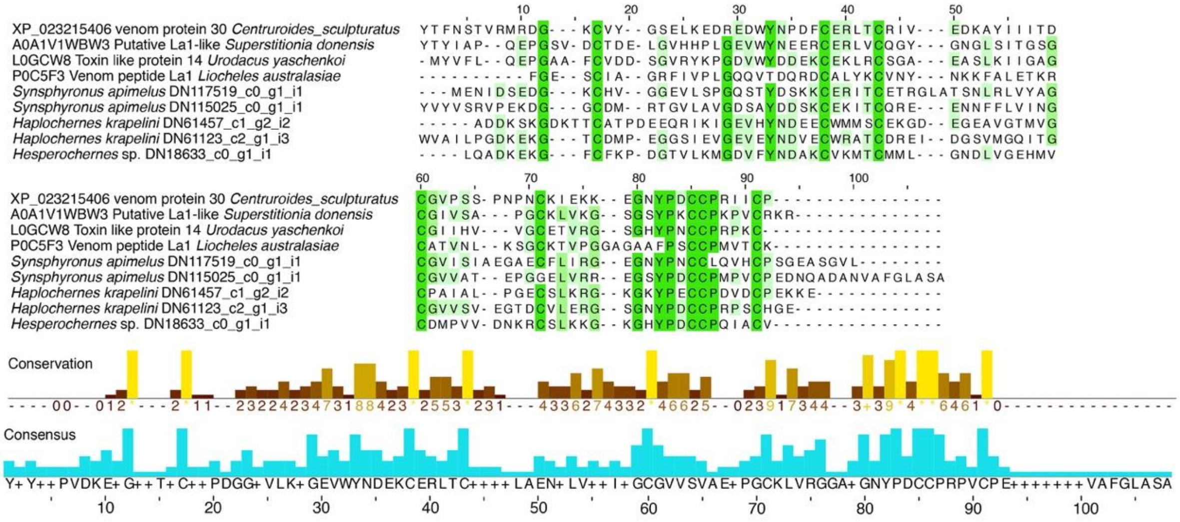


Figure 3. Multiple sequence alignment (MSA) of the peptide components with similarity to the single domain von Willebrand factor type C peptides (La1-like peptides) found in the transcriptome analysis of the venom gland of *S. apimelus*, *H. kraepelini*, and *Hesperochernes* sp. UniProt or GenBank numbers precede the peptide names. Percentage of identity between the MSA are highlighted in green. Below, histograms of the conservation and consensus of the MSA.

2.1.5. Defensins

Defensins are peptides widely distributed throughout vertebrates, invertebrates, plants, and fungi, whose functions are determined by the displayed inter-cysteine loops or the residues in the core [35]. For example, arthropod defensins have antimicrobial activity [36]. However, recent studies have suggested that scorpion defensins share a common ancestor with scorpion ion channel toxins [37,38]. We discovered 13 sequences with identities corresponding to four different precursors of defensins reported from two ticks, one scorpion, and one spider.

2.1.6. Other Components

Other transcripts potentially coding for venom proteins, including insulin-like growth factor binding protein, cysteine-rich secretory proteins, and peptidase inhibitors (not covered in the categories above) represent 26% of the transcripts annotated in this transcriptome analysis (Figure 2a). Among these, we found 15 with sequence similarity to the precursor of the U24 ctenitoxin Pn1 like that from *C. sculpturatus* and *P. tepidadorium*. Four sequences had similarity to a venom toxin peptide deduced from cDNA cloned from the venom of the scorpion *Hemiscorpius lepturus*. Additionally, we reported three sequences with less than 45% identity to a putative secreted salivary protein deduced from cDNA cloned from the tick *Ixodes scapularis*. Lastly, we found nine sequences with identities to two peptidase inhibitors reported from *C. sculpturatus* and *Stegodyphus mimosarum*.

2.2. Comparative Analysis of the Repertoire of Venom-Specific Transcripts in S. apimelus

The ortholog hit ratio (OHR) provides a proxy for the completeness of a transcriptome in terms of assembly coverage, with values above one suggesting insertions in the query sequence relative to the reference BLAST hit. Generally, most of the transcripts had a low OHR value (Figure 2b), suggesting that many of these transcripts contain relatively poorly conserved and/or unknown regions. Alternatively, the low OHR values could reflect low sequence coverage stemming from a large genome (to date, the size of a typical pseudoscorpion genome is unknown [24]). All venom categories reported here from the transcriptome of *S. apimelus* were found in the other two transcriptomes studied (Figure 2c). The number of genes coding for putatively venom proteins was slightly higher in the transcriptome of *Haplochernes kraepelini* (96 genes) but lower in the transcriptome of *Hesperochernes* sp. (73 genes). In all transcriptomes, enzymes were the most abundant proteins along with other venom components, such as the cysteine-rich secretory proteins and protease inhibitors (Figure 2c). Transcripts coding hyaluronidases were not found in the library of *Hesperochernes* sp. On the other hand, Kunitz-type inhibitors were poorly represented (5 to 10 genes). However, these transcripts correspond to several precursors of protease inhibitors reported from different arthropods (including insects and arachnids).

Low molecular mass spider toxins were poorly represented (in terms of diversity) in the three transcriptome libraries. Our phylogenetic analyses of U8-agatoxin-like peptides (ML and BI; Figure 4a) show the presence of three orthogroups, consisting of sequences from the three pseudoscorpion species. From these, one pseudoscorpion orthogroup was recovered with a U8-agatoxin-like homolog peptide, originally reported from the genomic analysis of *P. tepidariorum* with low nodal support (green clade in Figure 4b). The other two groups were pseudoscorpion-specific (orange and gray clades in Figure 4b). Finally, five transcripts from *S. apimelus* (representing one gene and five isoforms) clustered with the U8-agatoxin-like peptide from *C. sculpturatus* and another sequence reported from *Hemiscorpius lepturus*. No specific transcripts from any of the pseudoscorpion libraries clustered with peptides reported from tick venom (light blue clade in Figure 4b).

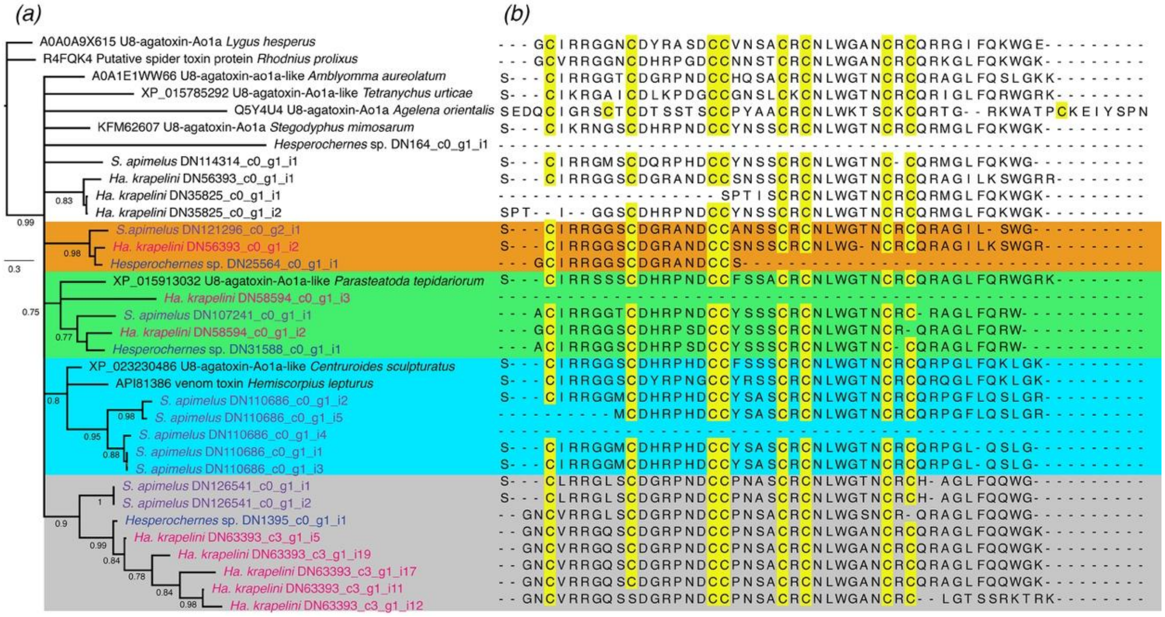


Figure 4. (**a**) Evolutionary tree of the U8-agatoxin-like peptides from a Bayesian analysis of 34 sequences reported from insects, and arachnids (including the sequences reported here). Posterior probabilities are indicated below nodes. (**b**) Multiple sequence alignment of the mature peptide predicted from the sequences used in the phylogenetic analysis. Those in blank had no mature peptide predicted and were represented only by the flanking region(s). Cysteine positions highlighted in yellow. Four orthogroup sequences from the three pseudoscorpion species are highlighted in colors (see text).

Our results suggest pseudoscorpion venom contains similar active peptides to those reported from spiders, ticks, and scorpions. To trace the evolutionary origin of the diversity of these components across arachnids, we mapped (using parsimony) eight categories of peptides with known function in the venom of spiders, scorpions, and ticks in the latest arachnid phylogeny [9] (Figure 5a–c). Enzymes (such as hyaluronidases and phospholipases), defensins, protease inhibitors, low molecular spider toxins (see below), and other venom components were shared by the four venomous arachnids. Within Pseudoscorpiones, all peptide categories were shared by the three libraries, except for the hyaluronidases (missing in the library of *Hesperochernes* sp., Figure 5d).

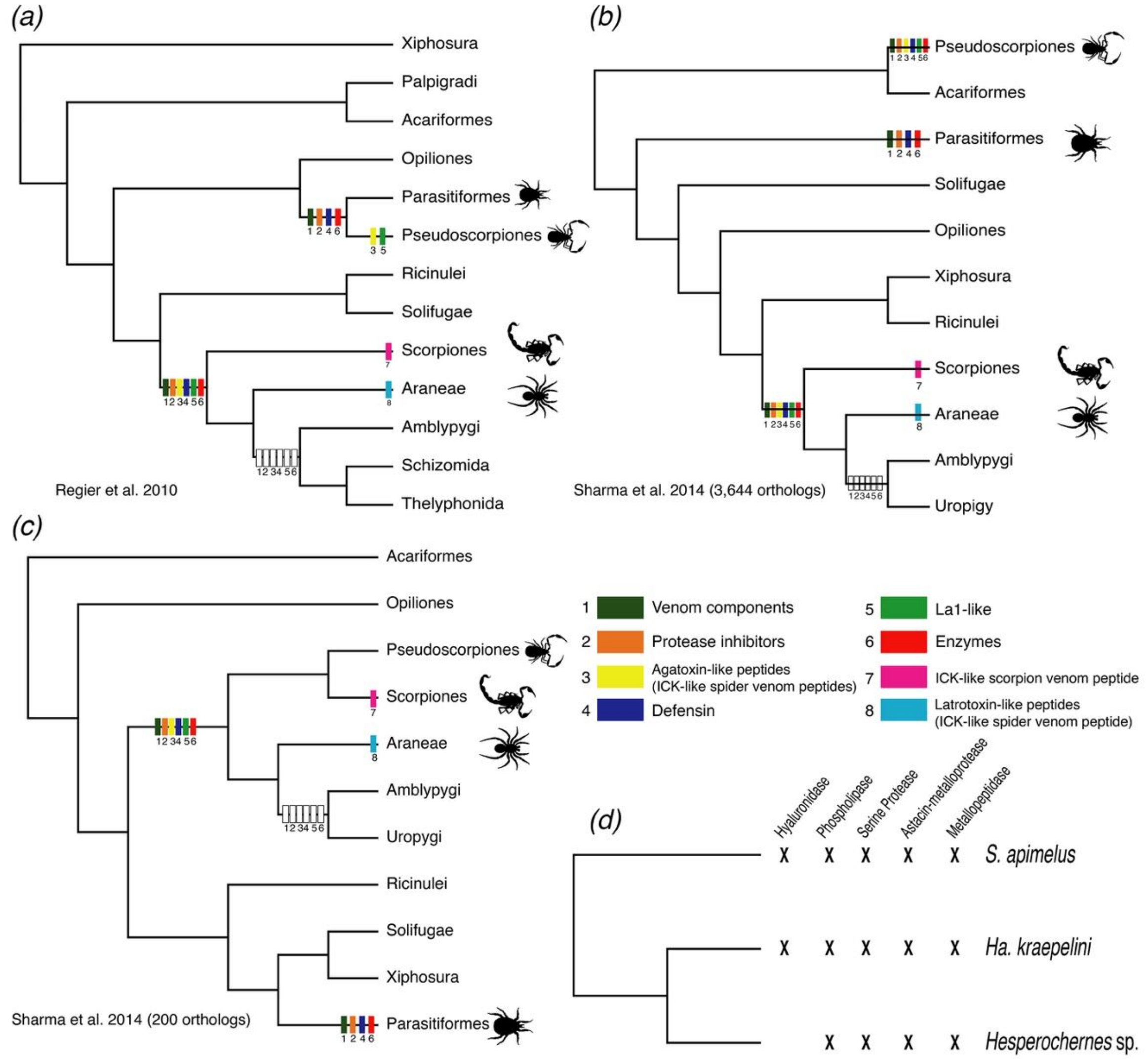


Figure 5. (**a**) Evolutionary hypotheses of the origins of venom components within Arachnida. (**a**–**c**) Tree topologies compiled from published sources [7,9] mapping each venom component by color. Blank squares indicate inferred secondary losses. (**d**) Phylogenetic relationships of the three pseudoscorpion species (compiled from References [1,2]) with the reported enzyme categories.

3. Discussion

Our high-quality pedipalpal transcriptome of *S. apimelus*, supported also by the analyses of two transcriptomes of other species, revealed for the first time the composition of pseudoscorpion venom. We found evidence for several components shared by the four venomous arachnid lineages, such as phospholipases, protease, and peptidase inhibitors. The presence of known peptidic toxins, such as those found in scorpion and spider venoms, were lowly represented. Our phylogenetic analyses of the low molecular mass spider toxins showed the presence of two different components unique to pseudoscorpions, and two components similar to the U8-agatoxin-like peptides from *P. tepidadorium* (spider) and from *C. sculpturatus* (scorpion). The function of the U8-agatoxin peptide (cloned from the spider *Agelena orientalis*) remains unknown [39]. However, its similarity to other agatoxins, a family of peptides including low mass molecular toxins with affinity to the sodium or calcium ion channels [40], suggest they might share similar functions. The only previous study on the effects of the crude venom of pseudoscorpion suggested the presence of putative neurotoxins of peptidic and nonpeptidic nature [11]. Whether the three groups found in pseudoscorpion venom are the culprit for neurotoxicity in rat brain is uncertain, because studies in spider venom have also shown the presence of polyamines targeting ionotropic glutamate receptors (e.g., [41,42]).

The composition of transcripts/genes in pseudoscorpion venom sheds light on the diversification of arachnid venom, both at the level of morphological sites of synthesis and molecular diversity. Spiders and ticks have their venom glands located anteriorly, injecting venom through the chelicerae (spiders) or salivary glands [19,43]. Scorpions on the other hand, possess venom glands located in the telson (the posterior-most part of the tail). Pseudoscorpions inject venom through the tips of the pedipalpal chela, with their venom glands located in the pedipalpal fingers or sometimes extending into the base of the chelal hand (Figure 1c). The evolution of venom glands within the four venomous arachnid orders is thus most likely the result of multiple independent evolutionary gains. The homology of the venom glands across the arachnids remains largely unexplored and may constitute an opportune target for cross-disciplinary studies of venom synthesis, evolution, and developmental genetics.

Resolving the phylogenetic position of Pseudoscorpiones would greatly refine the evolutionary context for arachnid venom. The origin of the venom peptide fraction has been suggested to be the recruitment of housekeeping genes into venom, followed by diversification and neofunctionalization (e.g., [44]). Following this reasoning, the phylogenetic position of Pseudoscorpiones is crucial to establishing the evolutionary relationship among venom components. Currently, the alternative phylogenetic positions of pseudoscorpions as (a) somehow related to the acarine orders (mites and ticks) or (b) more closely related to spiders and scorpions (Figure 5c), could be compatible with multiple scenarios. First, the most recent common ancestor (MRCA) of ticks and pseudoscorpions may have had these components, with separate gains at the base of Arachnopulmonata and secondary losses in non-venomous arachnopulmonate orders (Figure 5a). Alternatively, the similarities of pseudoscorpion and arachnopulmonate venom composition may be consistent with their closer phylogenetic relationship and a shared origin of venoms at the base of Pseudoscorpiones + Arachnopulmonata, a relationship supported in some phylogenomic analyses (Figure 5c). We also cannot rule out scenarios of multiple, independent gains of venom components in ticks, pseudoscorpions, and arachnopulmonates (Figure 5b). A more nuanced understanding of venom evolution within Arachnida is dependent upon resolving the phylogenetic position of Pseudoscorpiones and the constituent lineages of Acari, the mites and ticks.

4. Materials and Methods

Pseudoscorpion specimens were hand collected under stones in Stirling Range National Park, Western Australia (34°23′24″ S, 118°03′17″ E; 629 m elevation) on 18 August 2017 by A.Z.O., M.S.H., and P.P.S. (Figure 1). The pedipalpal chelae from 46 adult female and male specimens were dissected and transferred to 1.5 mL microcentrifuge tubes. Total RNA was extracted using the Trizol reagent (Ambion Life Technologies, Waltham, MA, USA). Library preparation and stranded mRNA sequencing followed protocols from the Biotechnology Center at the University of Wisconsin-Madison. Samples were run using an Illumina HiSeq2500 High Throughput platform with paired-end reads of 125 bp. Raw sequence reads can be found in the SRA database under the accession number SRR7062201 and the BioProject PRJNA453454. Adaptors were removed using Trimmomatic v. 0.36 [45] and the quality of cleaned raw reads was assessed with FastQC v. 0.11.5. [46]. Reads were assembled into contigs in a *de novo* fashion with Trinity v. 2.5 [47]. The quality of the assembly and basic statistics for the transcripts, genes, and isoforms were obtained using the *TrinityStats.pl* script. Assembled contigs were used as queries to search the UniProt database with the blastx and blastp algorithms; protein domains were identified with HMMER; and contigs were analyzed using Trinotate [47]. Additionally, to address the coverage of our transcripts, we calculated the ortholog hit ratio (OHR, [48–50]).

Selected transcripts with sequence similarity to venom components (e.g., from ticks, scorpions, spiders, or other arthropods) were used as queries to search UniProt and GenBank. From these new searches, matching sequences with lower expected (e) values, higher query cover values, or higher percentages of identity were selected as definitive matches. To contrast the venom components found in *S. apimelus* to other pseudoscorpion libraries (previously published), we assembled de novo the libraries of *Haplochernes kraepelini* and *Hesperochernes* sp. from raw reads downloaded from NBCI (accession numbers SRR1767661 and SRR1514877, respectively) following the same procedure as above. We calculated the OHR for their transcripts and used the selected venom transcripts from the library of *S. apimelus* as queries to search for orthologs in the other two libraries, using a phylogenetically informed orthology criterion, as implemented in UPhO [51]. The signal peptides and propeptides of these selected transcripts were determined with SpiderP from the Arachnoserver [52]. Multiple sequence alignments (MSA) of the relevant *S. apimelus* transcript-derived sequences with the corresponding input sets were obtained using MAFFT v. 7.0 [53]. Visualizations, conservations, and consensuses of MSA were obtain using Jalview v 2.10 [54].

To gain insights on the phylogenetic relationships of the transcripts with similarity to the low molecular mass spider toxins, we retrieved nine sequences which code, or putatively code, for U8 agatoxin and U8 agatoxin-like peptides from GenBank and UniProt. These sequences included two from cDNA cloned from two insect species; and seven deduced from cDNA cloned from seven arachnid species. Multiple sequence alignment for the full precursor was generated using MAFFT, resulting in a matrix consisting of 34 terminals and 208 amino acid sites. Maximum likelihood (ML) tree topologies were inferred in IQtree v 1.5.5 [55] using the PMB + Γ4 model, detected with ModelFinder [56] in IQtree, and by implementing 1000 ultrafast bootstrap resampling [57]. Bayesian inference (BI) analysis was performed with MrBayes 3.2.2 [58] using the JTT + Γ + I model, selected under the Bayesian information criterion using ProTest 3 [59]. Four runs, each with four Markov chains and a default distribution of chain temperatures, were implemented for 5×10^6 generations. Convergence of each chain was assessed using Tracer v. 1.6 with 5×10^5 generations discarded as burn-in.

Supplementary Materials: The following are available online at http://www.mdpi.com/2072-6651/10/5/207/s1, Sequences reported from the library of *Synphyronus apimelus* are in fasta format.

Author Contributions: C.E.S.-L. and P.P.S. conceived and designed the experiments. P.P.S., A.Z.O., and M.S.H. collected the specimens. C.E.S.-L performed the experiments and analyzed the data. P.P.S. contributed reagents. C.E.S.-L and P.P.S. wrote the paper, with all authors contributing edits.

Funding: This research was funded by the National Science Foundation under grant no. IOS-1552610. C.E.S.-L. was supported by a CONACYT postdoctoral fellowship (reg. 207146/454834).

Acknowledgments: Jesús Ballesteros assisted in the experiments and bioinformatics. Edits from Jesús Ballesteros and Greta Binford were incorporated into the manuscript. Comments from three anonymous reviewers refined the manuscript and visual presentation of the data.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Harvey, M.S. *Pseudoscorpions of the World, version 3.0*; Western Australian Museum: Perth, Australia, 2013. Available online: http://www.museum.wa.gov.au/catalogues-beta/pseudoscorpions (accessed on 18 April 2018).
2. Murienne, J.; Harvey, M.S.; Giribet, G. First molecular phylogeny of the major clades of Pseudoscorpiones (Arthropoda: Chelicerata). *Mol. Phylogenet. Evol.* **2008**, *49*, 170–184. [CrossRef] [PubMed]
3. Harms, D.; Dunlop, J.A. The fossil history of Pseudoscorpions (Arachnida: Pseudoscorpiones). *Foss. Rec.* **2017**, *20*, 215–238. [CrossRef]
4. Wheeler, W.C.; Hayashi, C.Y. The phylogeny of the extant chelicerate orders. *Cladistics* **1998**, *14*, 173–192. [CrossRef]
5. Giribet, G.; Edgecombe, G.D.; Wheeler, W.C.; Babbitt, C. Phylogeny and Systematic Position of Opiliones: A Combined Analysis of Chelicerate Relationships Using Morphological and Molecular Data. *Cladistics* **2002**, *18*, 5–70. [CrossRef] [PubMed]
6. Shultz, J.W. A phylogenetic analysis of the arachnid orders based on morphological characters. *Zool. J. Linn. Soc.* **2007**, *150*, 221–265. [CrossRef]

7. Regier, J.C.; Shultz, J.W.; Zwick, A.; Hussey, A.; Ball, B.; Wetzer, R.; Martin, J.W.; Cunningham, C.W. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature* **2010**, *463*, 1079–1083. [CrossRef] [PubMed]
8. Borner, J.; Rehm, P.; Schill, R.O.; Ebersberger, I.; Burmester, T. A transcriptome approach to ecdysozoan phylogeny. *Mol. Phylogenet. Evol.* **2014**, *80*, 79–87. [CrossRef] [PubMed]
9. Sharma, P.P.; Kaluziak, S.T.; Pérez-Porro, A.R.; González, V.L.; Hormiga, G.; Wheeler, W.C.; Giribet, G. Phylogenomic interrogation of Arachnida reveals systemic conflicts in phylogenetic signal. *Mol. Biol. Evol.* **2014**, *31*, 2963–2984. [CrossRef] [PubMed]
10. Harvey, M.S. The phylogeny and classification of the Pseudoscorpionida (Chelicerata: Arachnida). *Invertebr. Syst.* **1992**, *6*, 1373–1435. [CrossRef]
11. Santos dos, W.F.; Coutinho-Netto, J. Effects of the *Paratemnus elongatus* pseudoscorpion venom in the uptake and binding of the L-glutamate and GABA from rat cerebral cortex. *J. Biochem. Mol. Toxicol.* **2006**, *20*, 27–34. [CrossRef] [PubMed]
12. Gopalakrishnakone, P.; Possani, L.D.; Schwartz, E.F.; Rodriguez de la Vega, R.C. *Scorpion Venoms*; Springer: Berlin, Germany, 2015.
13. Gopalakrishnakone, P.; Corzo, G.; de Lima, M.E.; Diego-Garcia, E. *Spider Venoms*; Springer: Berlin, Germany, 2016.
14. Pessini, A.C.; Takao, T.T.; Cavalheiro, E.C.; Vichnewski, W.; Sampaio, S.V.; Giglio, J.R.; Arantes, E.C. A hyaluronidase from *Tityus serrulatus* scorpion venom: Isolation, characterization and inhibition by flavonoids. *Toxicon* **2001**, *39*, 1495–1504. [CrossRef]
15. Ferrer, V.P.; de Mari, T.L.; Gremski, L.H.; Trevisan Silva, D.; da Silveira, R.B.; Gremski, W.; Chaim, O.M.; Senff-Ribeiro, A.; Nader, H.B.; Veiga, S.S. A Novel Hyaluronidase from Brown Spider (*Loxosceles intermedia*) Venom (Dietrich's Hyaluronidase): From Cloning to Functional Characterization. *PLoS Negl. Trop. Dis.* **2013**, *7*, e2206. [CrossRef] [PubMed]
16. Bordon, K.C.F.; Wiezel, G.A.; Amorim, F.G.; Arantes, E.C. Arthropod venom Hyaluronidases: Biochemical properties and potential applications in medicine and biotechnology. *J. Venom. Anim. Toxins Incl. Trop. Dis.* **2015**, *21*, 43. [CrossRef] [PubMed]
17. Lajoie, D.M.; Zobel-Thropp, P.A.; Kumirov, V.K.; Bandarian, V.; Binford, G.J.; Cordes, M.H.J. Phospholipase D Toxins of Brown Spider Venom Convert Lysophosphatidylcholine and Sphingomyelin to Cyclic Phosphates. *PLoS ONE* **2013**, *8*, e72372. [CrossRef] [PubMed]
18. Incamnoi, P.; Patramanon, R.; Thammasirirak, S.; Chaveerach, A.; Uawonggul, N.; Sukprasert, S.; Rungsa, P.; Daduang, J.; Daduang, S. Heteromtoxin (HmTx), a novel heterodimeric phospholipase A2 from *Heterometrus laoticus* scorpion venom. *Toxicon* **2013**, *61*, 62–71. [CrossRef] [PubMed]
19. Cabezas-Cruz, A.; Valdés, J.J. Are ticks venomous animals? *Front. Zool.* **2014**, *11*, 47. [CrossRef] [PubMed]
20. Perner, J.; Provazník, J.; Schrenková, J.; Urbanová, V.; Ribeiro, J.M.C.; Kopáček, P. RNA-seq analyses of the midgut from blood- and serum-fed *Ixodes ricinus* ticks. *Sci. Rep.* **2016**, *6*, 36695. [CrossRef] [PubMed]
21. Rodriguez-Valle, M.; Moolhuijzen, P.; Barrero, R.A.; Ong, C.T.; Busch, G.; Karbanowicz, T.; Booth, M.; Clark, R.; Koehbach, J.; Ijaz, H.; et al. Transcriptome and toxin family analysis of the paralysis tick, *Ixodes holocyclus*. *Int. J. Parasitol.* **2018**, *48*, 71–82. [CrossRef] [PubMed]
22. Vachon, M. Ordre des pseudoscorpions. In *Grassé, Traité de Zoologie*; Masson: Paris, France, 1949; Volume 6, pp. 437–481.
23. Weygoldt, P. *The Biology of Pseudoscorpions*; Harvard University Press: Cambridge, MA, USA, 1969.
24. Garb, J.E.; Sharma, P.P.; Ayoub, N.A. Recent progress and prospects for advancing arachnid genomics. *Curr. Opin. Insect Sci.* **2018**, *25*, 51–57. [CrossRef] [PubMed]
25. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene Ontology: Tool for the unification of biology. *Nat. Genet.* **2000**, *25*, 25–29. [CrossRef] [PubMed]
26. Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* **2004**, *32*, 258D–261D.
27. Hodgson, E. Toxins and Venoms. *Prog. Mol. Biol. Transl. Sci.* **2012**, *112*, 373–415. [PubMed]
28. Chen, Z.; Wang, B.; Hu, J.; Yang, W.; Cao, Z.; Zhuo, R.; Li, W.; Wu, Y. SjAPI, the first functionally characterized Ascaris-type protease inhibitor from animal venoms. *PLoS ONE* **2013**, *8*, e57529. [CrossRef] [PubMed]
29. Mourão, C.; Schwartz, E. Protease inhibitors from marine venomous animals and their counterparts in terrestrial venomous animals. *Mar. Drugs* **2013**, *11*, 2069–2112. [CrossRef] [PubMed]

30. Wan, H.; Lee, K.S.; Kim, B.Y.; Zou, F.M.; Yoon, H.J.; Je, Y.H.; Li, J.; Jin, B.R. A Spider-derived Kunitz-Type Serine Protease Inhibitor that acts as a Plasmin Inhibitor and an Elastase Inhibitor. *PLoS ONE* **2013**, *8*, e53343. [CrossRef] [PubMed]
31. Chen, Z.-Y.; Hu, Y.-T.; Yang, W.-S.; He, Y.-W.; Feng, J.; Wang, B.; Zhao, R.-M.; Ding, J.-P.; Cao, Z.-J.; Li, W.-X.; et al. Hg1, Novel Peptide Inhibitor specific for Kv1.3 channels from first scorpion Kunitz-type Potassium channel toxin family. *J. Biol. Chem.* **2012**, *287*, 13813–13821. [CrossRef] [PubMed]
32. Miyashita, M.; Otsuki, J.; Hanai, Y.; Nakagawa, Y.; Miyagawa, H. Characterization of peptide components in the venom of the scorpion *Liocheles australasiae* (Hemiscorpiidae). *Toxicon* **2007**, *50*, 428–437. [CrossRef] [PubMed]
33. Zeng, X.-C.; Nie, Y.; Luo, X.; Wu, S.; Shi, W.; Zhang, L.; Liu, Y.; Cao, H.; Yang, Y.; Zhou, J. Molecular and Bioinformatical characterization of a novel superfamily of cysteine-rich peptides from arthropods. *Peptides* **2013**, *41*, 45–58. [CrossRef] [PubMed]
34. Santibáñez-López, C.; Cid-Uribe, J.; Batista, C.; Ortiz, E.; Possani, L. Venom gland transcriptomic and proteomic analyses of the enigmatic scorpion *Superstitionia donensis* (Scorpiones: Superstitioniidae), with insights on the evolution of its venom components. *Toxins* **2016**, *8*, 367. [CrossRef] [PubMed]
35. Shafee, T.M.A.; Lay, F.T.; Hulett, M.D.; Anderson, M.A. The Defensins consist of two independent, convergent protein superfamilies. *Mol. Biol. Evol.* **2016**, *33*, 2345–2356. [CrossRef] [PubMed]
36. White, S.H.; Wimley, W.C.; Selsted, M.E. Structure, function, and membrane integration of defensins. *Curr. Opin. Struct. Biol.* **1995**, *5*, 521–527. [CrossRef]
37. Froy, O.; Gurevitz, M. Arthropod defensins illuminate the divergence of scorpion neurotoxins. *J. Pept. Sci.* **2004**, *10*, 714–718. [CrossRef] [PubMed]
38. Zhu, S.; Peigneur, S.; Gao, B.; Umetsu, Y.; Ohki, S.; Tytgat, J. Experimental conversion of a defensin into a neurotoxin: Implications for origin of toxic function. *Mol. Biol. Evol.* **2014**, *31*, 546–559. [CrossRef] [PubMed]
39. Kozlov, S.; Malyavka, A.; McCutchen, B.; Lu, A.; Schepers, E.; Herrmann, R.; Grishin, E. A novel strategy for the identification of toxinlike structures in spider venom. *Proteins* **2005**, *59*, 131–140. [CrossRef] [PubMed]
40. Adams, M.E. Agatoxins: Ion channel specific toxins from the American funnel web spider, *Agelenopsis aperta*. *Toxicon* **2004**, *43*, 509–525. [CrossRef] [PubMed]
41. Strømgaard, K.; Jensen, L.S.; Vogensen, S.B. Polyamine toxins: Development of selective ligands for ionotropic receptors. *Toxicon* **2005**, *45*, 249–254. [CrossRef] [PubMed]
42. Quistad, G.B.; Suwanrumpha, S.; Jarema, M.A.; Shapiro, M.J.; Skinner, W.S.; Jamieson, G.C.; Lui, A.; Fu, E.W. Structure of paralytic acylpolyamines from the spider *Agelenopsis aperta*. *Biochem. Biophys. Res. Commun.* **1990**, *169*, 51–56. [CrossRef]
43. Foelix, R. *Biology of Spiders*; OUP: New York, NY, USA, 2011.
44. Dowell, N.L.; Giorgianni, M.W.; Kassner, V.A.; Selegue, J.E.; Sanchez, E.E.; Carroll, S.B. The deep origin and recent loss of venom toxin genes in rattlesnakes. *Curr. Biol.* **2016**, *26*, 2434–2445. [CrossRef] [PubMed]
45. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef] [PubMed]
46. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed on 12 December 2017).
47. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644–652. [CrossRef] [PubMed]
48. O'Neil, S.T.; Dzurisin, J.D.K.; Carmichael, R.D.; Lobo, N.F.; Emrich, S.J.; Hellmann, J.J. Population-level transcriptome sequencing of nonmodel organisms *Erynnis propertius* and *Papilio zelicaon*. *BMC Genom.* **2010**, *11*, 310. [CrossRef] [PubMed]
49. Ewen-Campen, B.; Shaner, N.; Panfilio, K.A.; Suzuki, Y.; Roth, S.; Extavour, C.G. The maternal and early embryonic transcriptome of the milkweed bug *Oncopeltus fasciatus*. *BMC Genom.* **2011**, *12*, 61. [CrossRef] [PubMed]
50. Riesgo, A.; Andrade, S.C.S.; Sharma, P.P.; Novo, M.; Pérez-Porro, A.R.; Vahtera, V.; González, V.L.; Kawauchi, G.Y.; Giribet, G. Comparative description of ten transcriptomes of newly sequenced invertebrates and efficiency estimation of genomic sampling in non-model taxa. *Front. Zool.* **2012**, *9*, 33. [CrossRef] [PubMed]

51. Ballesteros, J.A.; Hormiga, G. A new orthology assessment method for phylogenomic data: Unrooted Phylogenetic Orthology. *Mol. Biol. Evol.* **2016**, *33*, 2117–2134. [CrossRef] [PubMed]
52. Herzig, V.; Wood, D.L.A.; Newell, F.; Chaumeil, P.A.; Kaas, Q.; Binford, G.J.; Nicholson, G.M.; Gorse, D.; King, G.F. ArachnoServer 2.0, an updated online resource for spider toxin sequences and structures. *Nucleic Acids Res.* **2010**, *39*, D653–D657. [CrossRef] [PubMed]
53. Katoh, K.; Standley, D.M. MAFFT Multiple sequence alignment software Version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef] [PubMed]
54. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.A.; Clamp, M.; Barton, G.J. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **2009**, *25*, 1189–1191. [CrossRef] [PubMed]
55. Nguyen, L.-T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A fast and effective stochastic algorithm for estimating Maximum-Likelihood phylogenies. *Mol. Biol. Evol.* **2014**, *32*, 268–274. [CrossRef] [PubMed]
56. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; von Haeseler, A.; Jermiin, L.S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **2017**, *14*, 587–589. [CrossRef] [PubMed]
57. Minh, B.Q.; Nguyen, M.A.T.; von Haeseler, A. Ultrafast Approximation for Phylogenetic Bootstrap. *Mol. Biol. Evol.* **2013**, *30*, 1188–1195. [CrossRef] [PubMed]
58. Ronquist, F.; Teslenko, M.; van der Mark, P.; Ayres, D.L.; Darling, A.; Höhna, S.; Larget, B.; Liu, L.; Suchard, M.A.; Huelsenbeck, J.P. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Syst. Biol.* **2012**, *61*, 539–542. [CrossRef] [PubMed]
59. Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. ProtTest 3: Fast selection of best-fit models of protein evolution. *Bioinformatics* **2011**, *27*, 1164–1165. [CrossRef] [PubMed]

Appendix 2

Comprehensive species sampling and sophisticated algorithmic approaches refute the monophyly of Arachnida

*Jesús A. Ballesteros, Carlos E. Santibáñez-López, Caitlin M. Baker, Ligia R. Benavides, Tauana J. Cunha, Guilherme Gainett, *Andrew Z. Ontano, Emily V.W. Setton, Claudia P. Arango, Efrat Gavish-Regev, Mark S. Harvey, Ward C. Wheeler, Gustavo Hormiga, Gonzalo Giribet, Prashant P. Sharma*

This work is under review at *Molecular Biology and Evolution*

Abstract

Deciphering the evolutionary relationships of Chelicerata (arachnids, horseshoe crabs, and allied taxa) has proven notoriously difficult, due to their ancient rapid radiation and the incidence of elevated evolutionary rates in several lineages. While conflicting hypotheses prevail in morphological and molecular datasets alike, the monophyly of Arachnida is nearly universally accepted. Though a small number of phylotranscriptomic analyses have recovered arachnid monophyly, these did not sample all living chelicerate orders. We generated a dataset of 506 high-quality genomes and transcriptomes, sampling all living orders of Chelicerata with high occupancy and rigorous approaches to orthology inference. Our analyses consistently recovered the nested placement of horseshoe crabs within a paraphyletic Arachnida. This result was insensitive to analysis using slowly-evolving genes, site heterogeneous substitution models, and algorithmic approach to species tree inference. Investigation of systematic bias showed that genes and sites that recover arachnid monophyly are enriched in noise and exhibit low information content. To test the effect of morphological data, we generated a 514-taxon morphological data matrix of extant and fossil Chelicerata, analyzed in tandem with the molecular matrix. Combined analyses recovered the clade Merostomata (the marine orders

---

* Contribution: I contributed data from my thesis chapters and performed a subset of the analyses.

Xiphosura, Eurypterida, and Chasmataspidida), but nested within Arachnida. Our results suggest that morphological convergence resulting from adaptations to life in terrestrial habitats has driven the historical perception of arachnid monophyly, paralleling the history of numerous other invertebrate terrestrial groups.

Introduction

Chelicerates are a diverse group of arthropods that have played a major role as predators in ancient and recent ecosystems. United by the eponymous pincer-like appendages (the chelicerae/chelifores), chelicerates comprise the sister group to the remaining arthropods. The most familiar chelicerate orders are members of Arachnida, an assemblage of 12 orders (Acariformes and Parasitiformes are treated as single orders in this study) of terrestrial arthropods (e.g., spiders, scorpions, mites). Chelicerates also include two wholly marine clades—the sea spiders (Pycnogonida) and the horseshoe crabs (Xiphosura)—as well as considerable diversity of derived aquatic lineages within mites (1). The fossil record of chelicerates also attests to a broader aquatic diversity that includes freshwater horseshoe crabs (2), sea scorpions (Eurypterida) and chasmataspidids (3, 4).

Whereas most higher-level phylogenetic relationships of arthropods have been resolved by the advent of phylogenomic approaches (5, 6), the internal phylogeny of chelicerates has remained elusive. The traditional paradigm of chelicerate evolution postulates a single colonization of land by the common ancestor of a monophyletic Arachnida. In this scenario, extinct lineages such as the chasmataspidids and sea-scorpions are thought to represent stepping-stones in between horseshoe crabs and the origin of arachnids. Phylogenomic studies have recovered weak support for this scenario, with a considerable majority of analyses supporting a nested placement of Xiphosura as derived arachnids (7–10), a result also recovered in earlier Sanger-based molecular analyses (11–21).

A handful of phylogenomic matrices has recovered arachnid monophyly, attributing this result to (a) the use of slowly-evolving (i.e., less saturated) genes that are less prone to long-branch attraction (LBA) artifacts, or (b) expanded taxonomic sampling (7, 22, 23). However, the matrices of these works have been shown to be highly sensitive to model choice and algorithmic approach, and ironically lack representation of all extant chelicerate orders. Upon addition of libraries representing those missing arachnid orders to these same datasets, support for arachnid monophyly collapses (9, 10). Nevertheless, data quality and quantity remain limited for some groups in phylotranscriptomic datasets, specifically orders like Palpigradi and Solifugae.

Beyond arachnid monophyly, internal relationships within Chelicerata are unstable across phylogenomic analyses, which is in part attributable to the incidence of multiple fast-evolving lineages that incur long branch attraction artefacts, such as Acariformes, Parasitiformes, and Pseudoscorpiones. Well-resolved parts of the chelicerate phylogeny include the reciprocal monophyly of Pycnogonida, Euchelicerata (the remaining chelicerate orders), the monophyly and internal relationships of Tetrapulmonata (spiders and three other orders that plesiomorphically bear four book lungs), and the monophyly of each chelicerate order. More recently, phylogenomic analyses, together with rare genomic changes, have supported the clade Panscorpiones (Scorpiones + Pseudoscorpiones), in turn sister group to Tetrapulmonata (forming the clade Arachnopulmonata) (7, 10) (Fig. 1).

Towards a comprehensive chelicerate phylogeny that can inform the question of arachnid monophyly, we assembled a 506-taxon phylogenomic dataset representing the major lineages of all extant chelicerate orders and densely representing species-rich groups such as Araneae, Scorpiones, Pseudoscorpiones, and Opiliones. Our analyses examined sophisticated strategies to mitigate LBA, such as subsampling to minimize saturation, the use of infinite mixture site-heterogeneous models (CAT-GTR), and recently proposed recoding strategies in tandem with site-heterogeneous models applied to partitioned phylogenomics.

A common feature of phylogenomic studies is the omission of morphological taxa in an analytical framework, which has been shown to be detrimental to phylogenetic reconstruction (24). Combined analyses of phylogenomics and morphology have been proposed as a means to improve resolution and evaluate congruence between data classes (25, 26). While morphological datasets focusing on relationships among fossil taxa typically recover arachnid monophyly (2, 27; but see 28), most of these matrices have historically suffered from limited sampling of arachnid lineages, exhibit marked character conflict, and fail to recover the few relationships that are consistently supported by molecular phylogenies and genomics (e.g., Tetrapulmonata (27); Arachnopulmonata (2, 27, 28); Euchelicerata (29)). Therefore, toward assessing the impact of fossil taxa and morphological characters on phylogenomic analyses, we assembled a morphological dataset for Chelicerata to complement the phylogenomic dataset. The morphological dataset included extinct taxa (e.g., Chasmataspida, Eurypterida, Haptopoda, Phalangiotarbida, Synziphosurina, Trigonotarbida, and Uraraneida) as well as key fossils of extant orders.

We show that analyses of molecular datasets alone, as well as combined analyses of morphology and molecules, consistently recovered horseshoe crabs as nested within Arachnida. Interrogation of phylogenetic signal across loci showed that genes and sites supporting arachnid monophyly are more prone to systematic error than the remaining loci, suggesting that arachnid monophyly in molecular phylogenies reflects an analytical artifact.

Results

Partitioned analyses of phylogenomic datasets

We compiled 506 high-quality transcriptomes or genomes (>95% of libraries generated by us; 80 transcriptomes newly sequenced for this study focused on improving representation of scorpions, palpigrades, and opilioacariforms), sampling 24 outgroup and 482 chelicerate taxa (*SI Appendix*, Table S1). Phylogenetically informed inference of orthologs leveraged a recent *de novo* computation of orthologous genes for Chelicerata (3564 loci identified previously (8)) using the Unrooted Phylogenetic Orthology (UPhO) pipeline (30). As a separate, independent approach to orthology inference, orthologs were drawn from the Benchmarking Universal Single Copy Orthologs loci set for arthropods (BUSCO-Ar) (31, 32). Initial sets of orthologs were filtered based on maximal taxon decisiveness (33); we retained only loci that had at least one terminal for all the following clades: Araneae, Pedipalpi (= Thelyphonida + Schizomida + Amblypygi), Scorpiones, Ricinulei, Xiphosura, Solifugae, Opiliones, Palpigradi, Parasitiformes, Acariformes, Pseudoscorpiones, Pycnogonida, Pancrustacea, Myriapoda, and Onychophora. Applying this criterion, we reduced the UPhO ortholog set to 676 loci (Matrix 1) and the BUSCO set to 399 loci (Matrix 2). Thus, every major lineage (i.e., orders or closely related orders [e.g., Pedipalpi; Acariformes; Parasitiformes]) of chelicerates was represented by at least one terminal for every locus, in all analyses. For both matrices, we implemented (a) the site heterogeneous PMSF model for maximum likelihood search (34) (LG+C20+F+Γ_4), (b) traditional partitioned model maximum likelihood, and (c) gene tree summary (ASTRAL) approaches.

In all six analyses, we recovered the nested placement of Xiphosura within a paraphyletic Arachnida with support (bootstrap frequency [BS] > 95%; posterior probability > 0.95) and with significance in tests of monophyly (Fig. 2; *SI Appendix* Table S2). While relationships of apulmonate arachnid orders varied across topologies, all analyses invariably recovered the monophyly of Tetrapulmonata, Pedipalpi, Euchelicerata, and each chelicerate order. Scorpiones were consistently recovered as the sister group of Tetrapulmonata, whereas Pseudoscorpiones grouped with other long-branched orders (Acariformes and Parasitiformes).

Analyses of slowly-evolving matrices

In the case of Pseudoscorpiones, an external and independent phylogenetic data class informs the placement of this long-branched order. Specifically, a shared whole genome duplication unites the clade Arachnopulmonata, as evidenced by duplications of Hox clusters, systemic paralogy of developmental patterning genes, and enrichment of microRNA families (10) (Fig. 1). As our analyses of Matrices 1 and 2 did not recover a monophyletic Panscorpiones (with pseudoscorpions clustering with other long-branch orders), we reasoned that these datasets remained exposed to LBA.

Several strategies have been proposed to mitigate LBA in arachnid phylogeny, such as the use of site heterogeneous models, the use of slowly-evolving genes, or both (albeit with mixed results across datasets). To mitigate the impact of fast-evolving loci, we generated saturation plots for each locus and isolated a subset of 152 loci with high values for slope (≥ 0.4) and r^2 (≥ 0.95); these loci were concatenated to form Matrix 3 and analyzed using the same approaches as Matrices 1 and 2. Analyses of Matrix 3 with partitioned models, site heterogeneous models, and ASTRAL all recovered the monophyly of Arachnopulmonata (*sensu*

(10)) with maximal nodal support (Fig. 2*B*). Maximum likelihood inference under either partitioned or site heterogeneous models also recovered Panscorpiones (BS = 96% and 90%, respectively). All analyses of Matrix 3 rejected arachnid monophyly with support and with significance in tests of monophyly (Fig. 2*D*).

Bayesian inference analysis with CAT-GTR

Some of the most recalcitrant nodes in the tree of life that are impacted by LBA have been argued to be effectively resolved using analyses under the computationally intensive CAT-GTR infinite mixture model, as implemented in PhyloBayes-mpi. Examples of such nodes include the placement of Chaetognatha, Xenoturbellida, and Porifera (but see (35)). The PhyloBayes-mpi approach is notoriously difficult to implement for taxon-rich datasets due to the low probability of convergence. We therefore selected 56 representative terminals from the slow-evolving dataset (Matrix 3) such that major taxonomic groups (defined in *SI Appendix* Table S3) were each represented by three to five terminals, major basal splits were represented in each lineage, and the selected taxa exhibited the highest possible data completeness. This dataset was further filtered with BMGE v 1.12 (36) to remove heteropecillous sites, which violate the assumptions of the CAT model (37). The resulting matrix (Matrix 4) was comprised of 14,753 sites. Bayesian inference analysis was run on 8 independent chains for >20,000 cycles. To assess the impact of the starting tree on the analysis, two chains (C1 and C2) used the maximum likelihood tree computed for Matrix 4 as starting point (which recovered horseshoe crabs in a derived position). Another two chains (C3 and C4) were started on a maximum likelihood tree for Matrix 4, but constrained to recover arachnid monophyly. Four chains (C5–C8) used random starting trees.

Examination of ESS values and *a posteriori* tree distribution across all eight chains showed that summary statistics broadly exhibited convergence (*SI Appendix*, Tables S4, S5). A high value of the maximum split difference (*maxdiff*) was driven by a soft polytomy at the base of Euchelicerata. We examined estimates both from combined chains as well as summary topologies resulting from each starting tree type (*SI Appendix* Fig. S1, S2). None of the topologies in the 95% HPD interval of the *a posteriori* distribution supported the monophyly of Arachnida (PP = 0.02814) (Fig. 3*A*, 3*B*). Notably, Bayesian analysis using CAT-GTR rejected the monophyly of Acari (Acariformes + Parasitiformes) in favor of Poecilophysidea (Acariformes + Solifugae; PP=1.00) and Cephalosomata (Palpigradi + Poecilophysidea; PP=0.99) (*SI Appendix* Table S2). Moreover, Acari monophyly was supported in 0% of *a posteriori* tree space across the eight chains. These results suggest that Acari reflects another long branch attraction artifact. Notably, PhyloBayes-mpi was able to recover both the monophyly of Arachnopulmonata (PP=1.00) and Panscorpiones (PP≥0.99), regardless of the starting tree topology.

Partitioned analysis with mixture models and recoding

A recently proposed method for reconciling divergent results in partitioned *versus* mixture model studies of recalcitrant nodes makes use of a tiered approach to introduce site-heterogeneous models in tandem with SR4 recoding (RL2, *sensu* (38)). This approach has been shown to recover consistently the traditional placements of groups like Porifera in empirical datasets.

Upon applying the RL2 strategy to Matrix 3, we recovered yet another tree topology with a nested placement of Xiphosura, as well as Poecilophysidea (BS=73%), Panscorpiones

(BS=99%), and Arachnopulmonata (BS=100%) (Fig. 3*C*). The backbone of Euchelicerata exhibit negligible support, a result attributable to the loss of information via reducing the peptide alphabet to four states in SR4 recoding. Paralleling this result, previous applications of Dayhoff 6-state recoding to chelicerate datasets have rendered a basal polytomy at the root of Euchelicerata (22). These results are consistent with recent critiques of recoding strategies as solutions to saturation and compositional heterogeneity (39).

Tests of monophyly and concordance factors

Tests of monophyly were performed using the Approximately Unbiased (AU) test (40). The different topologies obtained from Matrices 1–3 were constrained to assess support for the monophyly of Arachnida, Acari, Poecilophysidea (Solifugae + Acariformes), and Panscorpiones (Pseudoscorpiones + Scorpiones). AU tests consistently rejected the monophyly of arachnids over the hypothesis of a derived Xiphosura (Fig. 2*D*, *SI Appendix* Table S2).

Traditional measures of nodal support are prone to inflation in phylogenomic datasets. Gene and site concordance factors (gCF and sCF) have been shown to measure phylogenetic signal irrespective of dataset size. We therefore computed values of gCF and both for unconstrained topologies under Matrices 1–3, as well as their counterparts when constrained to recover the monophyly of Arachnida. gCF and sCF values were consistently lower for Arachnida when compared to the hypothesis of a derived Xiphosura (Fig. 2*C*).

Interrogation of phylogenetic signal and systematic bias

To examine whether the derived placement of Xiphosura stemmed from a systematic artifact, we explored phylogenetic signal and properties of genes and sites, as a function of support for competing tree topologies (41). We found that loci favoring arachnid monophyly were consistently in the minority (39–41%) of genes across our datasets, irrespective of orthology criterion (Fig. 4*A*). Proportions of genes supporting arachnid monophyly are comparable to those supporting archaic groupings that have been debunked by phylogenomics and rare genomic changes, such as Dromopoda (=Scorpiones + Opiliones + Solifugae + Pseudoscorpiones; 34–36%) (*SI Appendix*, Fig. S3; Table S2). Across all matrices, genes exhibited the same distribution of saturation, evolutionary rate and missing data, regardless of support for a monophyletic Arachnida or for Xiphosura nested in Arachnida (Fig. 4*B*).

Furthermore, we discovered that genes supporting arachnid monophyly were shorter and exhibited fewer parsimony informative sites than genes supporting the unconstrained topology, across all matrices. Short genes with low informativeness have been linked to systematic error across an array of phylogenomic datasets, suggesting that arachnid monophyly may reflect noise rather than true phylogenetic signal. Consistent with this interpretation, we found that sites supporting arachnid monophyly exhibited higher Shannon entropy than sites supporting a nested Xiphosura (Fig. 4*C*). Sites supporting arachnid monophyly were fewer in number and had higher Shannon entropy even when compared to sites supporting a debunked grouping that has been falsified by rare genomic changes (Dromopoda; *SI Appendix*, Fig. S4).

Combined analyses of morphology and molecules

To assess the impact of morphological data, we began with the character matrix of Huang *et al.* (42), the most comprehensively coded morphological matrix of extant chelicerates to date, including recently discovered arachnid fossils that have impacted reconstruction of ancestral states. To this matrix, we added the sea spider *Flagellopantopus* and Phalangiotarbida from codings in the literature, as well as all extant chelicerates in the molecular matrix. Errors previously entered in the character coding were corrected. We added new characters from the recent literature pertaining to the neuroanatomy of Xiphosura and several arachnid orders, as well as previously overlooked character systems.

To overcome artefacts stemming from missing and inapplicable character partitions, non-chelicerate outgroup taxa (Onychophora, Mandibulata) were removed from this analysis. For the same reason, we excluded putative chelicerate stem-groups of questionable and controversial placement for which molecular sequence data are inapplicable. Pycnogonida was used to root the Euchelicerata.

When analyzed alone under equal weights parsimony, the morphological dataset yielded little basal resolution (Fig. 5*A*). A strict consensus of equally parsimonious trees recovered a basal polytomy of Euchelicerata. Various interordinal relationships received negligible nodal support, although they accorded closely with recent morphological analyses *viz.* the recovery of Tetrapulmonata (including Trigonotarbida and Haptopoda) and Acaromorpha (Ricinulei + Acari). Under a Bayesian inference approach (Fig. 5*B*), the morphological dataset showed monophyly of Arachnida, Panscorpiones, Acaromorpha, Acari, and Tetrapulmonata, albeit without support (PP<0.95). Both approaches recovered the monophyly of Merostomata (BS=90%; PP=1.00).

Maximum likelihood analysis of the combined matrix (Fig. 5*C*) recovered monophyly of Merostomata (a grouping of the marine taxa Xiphosura, Synziphosurina, Eurypterida, and Chasmataspidida). Merostomata in turn is sister group to Ricinulei. We recovered Trigonotarbida as part of the tetrapulmonates, consistent with the presence of two pairs of book lungs in these groups. Phalangiotarbida is the sister group to Opiliones. Key fossil taxa were recovered in expected placements, such as the harvestman suborder Tetraophthalmi, and the orders Uraraneida and Haptopoda. Palpigradi appears as the sister group to the remaining Euchelicerata, with moderate support, paralleling the result of the RL2 recoding strategy. Compared to molecular analyses, support values were lowered by the inclusion of morphological data in the combined analysis, a result attributable to the instability incurred by data-poor fossil taxa.

Discussion

Arachnid monophyly is not supported by modern phylogenomic approaches

Molecular results that recover non-traditional groupings are often labeled as artifacts, especially when morphological patterns and long-held evolutionary scenarios come under question. Like the basal topology of groups like Metazoa, birds, and angiosperms, the basal topology of Euchelicerata has long defied stability in molecular datasets. Proposals to “correct” the tree and recover arachnid monophyly using molecular datasets have included restricting analyses to slowly-evolving genes (or less saturated genes, a correlate of evolutionary rate) (7, 23), expansion of taxonomic sampling (23), the use of site heterogeneous models (23, 24), or some combination thereof. As our analyses show, the derived placement of Xiphosura (possibly with the other merostomate orders) is consistently recovered despite concomitant application of all these putative solutions. This outcome is consistent with reexaminations of datasets that were previously used to justify arachnid monophyly (23, 24); when reanalyzed with missing groups to achieve the sampling of all extant chelicerate orders, every one of these datasets rejects arachnid monophyly with support (9, 10).

Why have some recent molecular datasets been able to recover arachnid monophyly (albeit with incomplete sampling of arachnid orders)? As previously shown, the matrices of

Lozano *et al.* (22) and Howard *et al.* (23) exhibit a number of bioinformatic and analytical errors in matrix assembly, rendering those matrices flawed and, in one case, unreproducible (refs. 9, 10). Upon further reexamining those datasets, we additionally found an unexpectedly high number of outliers in root-to-tip distances across gene trees. Using an annotation strategy based on the *Drosophila melanogaster* proteome, we discovered that the cause of this noise was the widescale inclusion of paralogs across these datasets. Specifically, 29% (68/233) of loci in the Lozano *et al.* Matrix A, and 41% (82/200) of loci in the Howard *et al.* matrix (algorithm-based orthology inference strategy) included paralogs, often from distantly related multigene families (*SI Appendix*, Tables S6, S7).

Could the properties of genes that are able to recover arachnid monophyly inform the selection of "better" loci for chelicerate phylogenomics? To address this, we examined the distribution of phylogenetic signal in our datasets for genes and sites supporting arachnid monophyly, versus the unconstrained topology, using ΔGLS and ΔSLS approaches (41). Genes supporting the nested placement of Xiphosura exhibited no evidence of systematic biases compared to the minority, which supported arachnid monophyly (39–41%). Instead, we discovered the opposite trend: genes supporting arachnid monophyly tended to have shorter alignment lengths and fewer informative sites than genes supporting a nested Xiphosura. Short genes and low informativeness are closely associated with phylogenetic error. Consistent with this interpretation, sites supporting arachnid monophyly exhibited higher Shannon entropy and low structure (i.e., greater randomness). For context, the proportions of genes supporting a grouping that has been clearly discredited by genome architecture (i.e., Dromopoda, which historically united two arachnopulmonate orders with two apulmonate orders) are nearly identical to those supporting arachnid monophyly (*SI Appendix*, Fig. S4). Moreover, the number of sites supporting this debunked grouping is higher than those supporting arachnid monophyly (*SI Appendix*, Fig. S5).

These analyses suggest that support for arachnid monophyly does not reflect hidden signal, so much as noise and error in the datasets that have putatively supported this grouping. We submit that the sum of our analyses, however counterintuitive, may reflect a phylogenetically accurate relationship—Xiphosura (and possibly the other merostomates) may simply constitute derived arachnids.

Slowly evolving genes and site heterogeneous models overcome LBA artifacts in chelicerate phylogeny

As anticipated, several groups in our phylogeny reflected long root-to-tip distances, constituting lineages prone to LBA artifacts. The inclusion of Opilioacariformes, the slowly-evolving sister group of the remaining Parasitiformes, was previously shown to break up the grouping of Acariformes and Parasitiformes, suggesting that Acari is a long branch artifact (10). In this study, we increased the sampling of Opilioacariformes to three libraries, and concordantly, never obtained the monophyly of Acari, particularly when pursuing approaches best suiting to mitigating LBA. This outcome suggests that the correspondences of mite and tick bauplans represent a case of morphological convergence in chelicerates. Similar convergence of mouthparts occurs in the gnathobasic preoral chambers of Opiliones and Scorpiones, which were previously grouped by a subset of morphological analyses (43).

Phylogenomic subsampling for slowly evolving genes did show Panscorpiones within Arachnopulmonata (Fig. 2*B*), a result that is attributable to a marked shift in the proportion of genes supporting this group as a function of evolutionary rate (*SI Appendix*, Fig. S5). However, even in maximum likelihood analyses that prioritized slowly evolving genes, we recovered Acariformes and Parasitiformes clustered near the base of the euchelicerate tree, placements that we regarded as possible LBA artifacts. Upon analyzing the slowly evolving matrix with site heterogeneous models in a Bayesian framework (CAT-GTR in PhyloBayes-mpi), not only were Panscorpiones and Arachnopulmonata recovered, but this approach also resolved Acariformes as the sister group of Solifugae (=Poecilophysidea), with Poecilophysidea in turn sister group to Palpigradi (=Cephalosomata) (Fig. 3*B*). Four-state recoding in tandem with site heterogeneous models eroded all support from the base of Euchelicerata, but this analysis did recover Poecilophysidea as well (Fig. 3*C*).

Intriguingly, these groupings (Poecilophysidea and Cephalosomata, respectively) were previously supported by a minority of phylogenetic analyses and were proposed on the basis of patterns of anterior sclerotization in these orders (19, 44, 45). A proximate relationship of Palpigradi and Solifugae is also supported by the anatomy of the coxal gland (9). Given the species richness of both Acariformes and Parasitiformes, future efforts to clarify the relative placements of these groups must focus on increasing the representation of basal nodes, a strategy that has been shown to outperform algorithmic and data trimming solutions to resolving the placement of pseudoscorpions. Balanced consideration of alternative and overlooked morphological groupings is also warranted in reexaminations of chelicerate phylogeny.

Morphology is confounded by convergence in chelicerate phylogeny

Unlike in other animal clades (e.g., 24), the addition of morphological data to this mix does not ameliorate the discordance with the traditional phylogeny of chelicerates; we found that combining morphological and molecular datasets using model-based approaches recovers

Merostomata (the marine group that includes horseshoe crabs) as nested within Arachnida. The notion that morphological synapomorphies of Arachnida can outweigh the dissonance found in molecular data found no support in this study. Furthermore, only in combination with molecular data was morphology able to recover clades supported by rare genomic characters (Panscorpiones and Arachnopulmonata); by itself, morphology has not recovered this arrangement of Arachnopulmonata, either in this analysis or in historical efforts (27–29, 42, 43).

One caveat of the combined analysis is that outgroups like putative stem-groups of Chelicerata (e.g., megacheirans) were not included, as their phylogenetic position is controversial even in morphological datasets. The exclusion of these groups may prevent character states from being optimized correctly, such as biramous appendages (the presence of exopods), faceted eyes, and gnathobasic mouthparts in marine groups. To assess this possibility, we trialed fusing our molecular dataset (Matrix 3) to morphological matrices from the literature with widely different taxon sets (total group Chelicerata (46); Panarthropoda (47)). These supplementary datasets featured minimal sampling of extant arachnid fauna (typically, one exemplar per order), as well as greater proportions of missing data. We found that combining data classes destabilized the traditional relationships previously predicted by those studies, either incurring the non-monophyly of Euchelicerata (*SI Appendix*, Fig. S6A, S6B) or of Chelicerata (*SI Appendix*, Fig. S6C, S6D). Within Euchelicerata, datasets that broadly represented panarthropod diversity (fossil and extant) recovered a nested placement of Merostomata within Euchelicerata when combined with molecular data (*SI Appendix*, Fig. S6C, S6D), paralleling our results. These outcomes suggest that morphological data partitions seeking to capture deep chelicerate relationships may feature far less robustness of phylogenetic signal than commonly portrayed, particularly in a total evidence framework. Concordantly, a recent paleontological study failed to recover even Tetrapulmonata (27), the only higher-level group that is consistently recovered by most morphological and molecular datasets. Another recent paleontological study that recovered arachnid non-monophyly took the step of constraining Arachnida *a priori* to ensure the recovery of the traditional topology (28).

Admittedly, the scenario of a nested Xiphosura invites entrenched skepticism, particularly from adherents of paleontology. In addition to an extensive fossil record, horseshoe crabs exhibit an array of putatively plesiomorphic traits that are suggestive of a basally branching placement. The fossil record of merostomates is rich with Xiphosura and Eurypterida species, thought to represent a stepwise colonization of land via internalization of the book gill of these marine groups (for this reason, the position of scorpions at the base of the Arachnida was a central tenet of this evolutionary transformation series). Recent arguments in favor of arachnid monophyly have thus focused on the faceted eye, which is thought to reflect the ancestral condition; the gnathobasic (enditic) mouthparts of merostomates; the biramous condition of merostomate appendages; and the anatomy of the book gill, which shares correspondences with the book lung of large-bodied arachnids (e.g., scorpions; basally branching spiders) (23). Moreover, arachnid monophyly has historically been defended based on a series of characters stemming from the musculoskeletal system (48).

However, a comparison with the history of mandibulate arthropod phylogeny offers compelling reasons to doubt the linearity of morphological evolutionary scenarios. Within Chilopoda, only one order of centipedes (Scutigeromorpha) has retained the faceted eye found in fossil outgroups, whereas all other centipedes bear ocelli or are blind, suggesting that faceted eyes are highly prone to discretization and loss in terrestrial habitats in a group at least Devonian in age (5). Paralleling this trend, various fossil arachnid groups (e.g., fossil scorpions, Trigonotarbida, and fossil Ricinulei) exhibit “semi-compound” eyes (aggregations of ocelli) in head regions positionally homologous to the faceted eyes of Xiphosura and Eurypterida (4). The faceted eyes of merostomates may reflect a plesiomorphic condition retained deep in the euchelicerate tree, like the faceted eye of scutigeromorphs within centipedes.

Similarly, discussions of the gnathobasic mouthparts of merostomates echo historical debates over the nature of the gnathobasic mandible of terrestrial mandibulates, as well as other correspondences of head appendages. It was previously thought that Hexapoda and Myriapoda constituted sister groups (the clade Tracheata), a relationship supported by their putatively shared gnathobasic mandible, appendage-free intercalary segment, uniramous appendages, and arrangement of the respiratory organs (tubular tracheae, typically opening as paired spiracles on pleural territories of trunk segments). The gradual overturning of this relationship by molecular phylogenies in favor of the Pancrustacea hypothesis revealed that striking morphological convergences could occur in distantly related taxa as a result of common selection pressures in terrestrial environments (5, 6, 14, 16, 17). In this light, the reduction of gnathobasic mouthparts in terrestrial chelicerate orders could also reflect parallel losses as adaptations to life on land, as evidenced by the mandibular architecture of Hexapoda and Myriapoda. Parallel losses of secondary rami and simplification of appendages are also broadly observed in terrestrial arthropods, such as arachnids, myriapods, hexapods, and terrestrial malacostracans (e.g., Isopoda, Amphipoda). We submit that the morphology of merostomate appendages is closely tied to evolution in marine habitats and may reflect retention of plesiomorphies; the absence of these structures in terrestrial arthropod groups does not offer compelling evidence uniting Arachnida.

Convergent evolution of tracheal tubules in other terrestrial groups, such as Onychophora, Hexapoda, and Myriapoda, falsifies the interpretation that a lung-like organ is a necessary stepping stone to the acquisition of tracheal tubules in chelicerates. The conventional and simplistic evolutionary transformation series of book gill to book lung to tracheal tubule is deeply undermined by the complexity of respiratory organ evolution in Chelicerata. This point is underscored by the recent discovery of a eurypterid with trabeculate book gills well after the appearance of arachnids in the fossil record (340 Mya; (49)), as well as secondarily marine scorpions with lamellate gills (*Waeringoscorpio*; 4), and the diversity of modern aquatic mites (1). The recent recovery of Pseudoscorpiones as a derived member of Arachnopumonata, as well as investigations of respiratory structures across spiders, reveals that book lungs have been frequently lost and repeatedly transformed into tracheal tubules, with loss of book lungs observed in multiple miniaturized arachnopumonate groups (e.g., the posterior book lung pair of Schizomida and most araneomorph spiders; complete loss of book lungs in miniaturized spiders and pseudoscorpions) (10, 50). There is no compelling evidence that evolutionary transitions of respiratory organs have followed a simple, linear series at the base of Arachnida, nor that water-to-land (or the reverse) transitions are rare or irreversible in the arthropod fossil record.

As for the putative musculoskeletal synapomorphies established for Arachnida (musculature and patterns of appendage joints (48)), we submit that the evolution of this entire character system may be closely tied to the selective pressures of a terrestrial life style. Arthropod appendages are highly adaptive structures, and biomechanical demands on locomotory appendages differ greatly between aquatic and terrestrial organisms (51). As with the correspondences of insect and myriapod musculoskeletal anatomy, there is no evident reason why the musculoskeletal system would constitute a homoplasy-free data source for arachnids.

Taken together, morphological character systems that putatively support arachnid monophyly tend to exhibit high levels of homoplasy upon closer examination, especially when examining their counterparts in Mandibulata. Given the remarkable morphological convergence exhibited by Hexapoda and Myriapoda, we postulate that parallel evolution in terrestrial chelicerate orders may confound inferences of homology in morphological datasets. While no morphological characters overtly support a closer relationship of Xiphosura to any subset of arachnid orders (but see (52, 53)), the absence of morphological support for numerous, robustly recovered molecular clades is a common feature of ancient invertebrate relationships, as exemplified by the modern higher-level phylogeny of groups like Annelida, Mollusca, and Nematoda (54, 55, 56). Indeed, the discovery of a particular well-supported relationship in molecular datasets has often served as the catalyst for revitalized morphological study and reinterpretation of previous homology statements, as in the case of Pancrustacea, Ecdysozoa, and Arachnopulmonata (53). Given the recovery of Poecilophysidea and Cephalosomata in some analyses, reexamination of previously overlooked interordinal groupings may provide a better understanding of hidden phylogenetic signal in specific chelicerate morphological character systems. A derived placement of merostomates as a group more proximal to Arachnopulmonata could also reconcile the morphology of extinct marine groups like eurypterids with the unambiguously nested position of Scorpiones, a hypothesis that could be tested through functional genetic approaches to understanding the developmental basis for respiratory organ patterning in horseshoe crabs, arachnopulmonates, and apulmonate arachnids.

The nested placement of Xiphosura, together with the reconstruction of multiple terrestrialization events across a grade of arachnid diversity, must be treated as a valid competing hypothesis. Future efforts to integrate new phylogenetic data classes and rare genomic characters (e.g., (10)) may offer clearer resolution of relationships among the apulmonate arachnid orders.

Conclusion

Analyses of molecular data and total evidence phylogenetic approaches do not support arachnid monophyly. The concept of Arachnida may reflect the antiquated notion that terrestrialization is rare or costly in evolutionary history. As revealed by the history of groups like mandibulate arthropods, nematodes, and Pulmonata (gastropods), terrestrialization has not only evolved many times independently within such taxa, but is also the cause of remarkable and misleading cases of morphological convergence in terrestrial taxa.

The strongest evidence that morphological datasets of Chelicerata may be prone to misinterpretation of homologies is provided by the positions of scorpions and pseudoscorpions, which are united with tetrapulmonates by a rare genomic change (an ancient whole genome duplication event). Morphological datasets, including the dataset we generated, have consistently failed to recover this grouping (with or without the miniaturized Pseudoscorpiones). Recent paleontological phylogenies and evolutionary reconstructions continue to support the 19th-century idea that scorpions constitute the sister group of the remaining Arachnida (2, 46, 47).

If morphological datasets can falter in the recovery of such robustly resolved chelicerate groups, it stands to reason that phylogenetic signal in morphological datasets may not be sufficiently robust to adjudicate other nodes in chelicerate interordinal phylogeny. The traditional placements of Xiphosura, Eurypterida, and various stem-group fossils must then also be regarded as suspect. Given the history of erstwhile morphological groupings like Tracheata, Uniramia, Articulata, Polychaeta, Pulmonata, Opisthobranchia, and numerous others, we postulate that phylogenomic approaches to deep metazoan relationships should treat morphological interpretations with skepticism *prima facie*, especially in the context of selective pressures like terrestrialization that promote morphological convergence.

Materials and Methods

Details of the methods below are provided in the *SI Appendix*.

Taxon sampling and orthology inference

Taxon selection consisted of 24 outgroup and 482 ingroup terminals; these 506 transcriptomes and genomes (*SI Appendix*, Table S1) sampled all extant chelicerate orders with multiple terminals. Eighty new libraries were generated following previously published protocols. Proteomes and peptide sequences were used as inputs. Phylogenetically-informed inference of orthologs leveraged a recent *de novo* computation of orthologous genes for Chelicerata using UPhO (3564 loci identified previously (8)). For validation, these collections of putative orthologous sequences were BLASTed (blastp v. 2.9.0+ (57)) against the *Drosophila melanogaster* proteome for annotation using the best hit. Sequences not matching the most common annotation were discarded. Separately, orthologs benchmarked using BUSCO-Arthropoda database were analyzed independently. The set of complete, single-copy BUSCOs >100 amino acids in length was retained from each library.

Matrix construction

Initial sets of orthologs were filtered based on taxon decisiveness (33). We retained only loci that had at least one terminal for all the following clades: Araneae, Pedipalpi (Uropygi + Schizomida + Amblypygi), Scorpiones, Ricinulei, Xiphosura, Solifugae, Opiliones, Palpigradi, Parasitiformes (treated here as the order uniting Holothyrida, Ixodida, Mesostigmata, and Opilioacariformes), Acariformes (treated here as the order uniting Sarcoptiformes and Trombidiformes), Pseudoscorpiones, Pycnogonida, Pancrustacea, Myriapoda, and Onychophora. Applying this criterion, we reduced the UPhO ortholog set to 676 loci (Matrix 1) and the BUSCO set to 399 loci (Matrix 2).

To assess the impact of saturation, we generated saturation plots for each locus (see Saturation Plots, below) and isolated a subset of 152 loci with slope ≥ 0.4 and $r^2 \geq 0.95$; these loci were concatenated to form Matrix 3. To operate PhyloBayes-mpi with the computationally demanding CAT+GTR+Γ model, we selected 56 representative terminals from the slow-evolving dataset (Matrix 2) such that major taxonomic groups (*SI Appendix*, Table S3) were each represented by three to five terminals, major basal splits were represented in each lineage, and the selected taxa exhibited high data completeness. This dataset was further filtered with BMGE v 1.12 (36) to remove heteropecillous sites to form Matrix 4 for analysis with PhyloBayes-mpi.

Partitioned analyses

Gene trees were inferred using IQ-TREE v. 1.6.10 (58) with model-fitting using ModelFinder (59) and nodal support estimation using the ultrafast bootstrap (60) as follows: *iqtree -mset LG,WAG,JTT,Dayhoff,JTTDCMut,DCMut,PMB -m MFP -bb 1000*. Maximum likelihood analyses of concatenated datasets (Matrices 1–3) were run using a gene partitioning strategy implementing the best substitution models identified during gene tree reconstruction. Tree topologies were inferred using IQ-TREE, with nodal support estimated using ultrafast bootstrapping. Summary coalescent estimates of the species phylogenies were estimated from the individual gene trees using ASTRAL v 3.14.2 (61).

Mixture model analyses

We computed maximum likelihood analyses with the posterior mean site frequency model (34) for Matrices 1-3, using the LG+C20+F+G implementation. The use of more site categories (e.g., C60) proved prohibitive for a dataset of this size, with the C20 model demanding 1.1 Tb of RAM to compute site-specific model parameters. Analyses were computed using IQ-TREE v. 1.6.10. Nodal support was estimated using ultrafast bootstrapping. Bayesian inference analysis was performed using PhyloBayes-mpi v 1.8 (62) and the CAT+GTR+Γ model on Matrix 4, which was optimized for this purpose. Bayesian inference analysis was run on 8 independent chains for >20,000 cycles. Convergence of parameters and topologies was assessed using Tracer 1.7.1 (63) and native PhyloBayes-mpi summary programs. Summary statistics and chain lengths are provided in the *SI Appendix*, respectively. Trace files of parameters and tree files from each run are provided in FigShare. Convergence parameters exhibited differences as a function of combining different chains. Different combinations of chains produced varying maximum split differences. Examination of ESS values and *a posteriori* tree distribution across all eight chains showed that summary statistics broadly exhibited convergence; the high value of the maximum split difference is driven by a soft polytomy at the base of Euchelicerata.

Recoded mixture model analyses

We implemented a site heterogeneous mixture model approach to partitioned phylogenomics using the RL2 strategy recently proposed (38), which implements 4-state recoding of amino acid data (SR4) (64). Analyses were performed in IQ-TREE v. 1.6.10, following the original implementation (38).

Morphological analysis

We developed a morphological matrix of 291 characters coded for 482 extant and 39 fossil chelicerates. Fossil taxa are listed in *SI Appendix* Table S5. Given the unambiguous recovery of Pycnogonida as the sister group to the remaining chelicerates, Pycnogonida were used to root the tree. Character codings were drawn from previous higher-level analyses of sea spiders (65), harvestmen (15, 66), scorpions (67), and arachnids (15, 29, 42, 43). Errors and discrepancies with previous character codings were modified and we additionally coded new characters informed by recent investigations (52, 53). Fossil taxa were coded using original descriptions from the literature.

Model-fitting for the morphological dataset was performed in IQ-TREE v. 1.6.10 from the dataset initially partitioned based on the number of character states. Bayesian analyses using the same partitioning scheme were performed in MrBayes v 3.2.7a (68) using the Mk1 model with unlinked rate and state frequency parameters per partition. The analyses consisted of four independent runs of 50 M cycles. Equal weights and implied weights parsimony analyses were performed using TNT v. 1.5 (69).

We additionally performed total evidence analyses using two recently published morphological matrices with differing representations of stem-group chelicerate taxa, (46, 47), complemented by molecular data from Matrix 3. In cases of non-overlapping taxa, a chimeric terminal was constructed using the closest related species to a given terminal in the morphological datasets. For each chimeric terminal, character codings were checked to ensure their applicability to their morphological counterpart; no coding changes were required for chimeras. Due to the degree of missing data in these matrices, analyses were only performed using parsimony (equal and implied weights); model-based analyses consistently failed to converge for these supplementary datasets.

Supplementary Material

Supplementary data are available at bioRxiv under URL https://www.biorxiv.org/content/10.1101/2021.08.16.456573v1.supplementary-material

Data Availability

Raw sequence data that support the findings of this study have been deposited in NCBI Sequence Read Archive. All transcriptomic assemblies, gene alignments, gene trees, supermatrices, phylogenomic trees, morphological data matrix, PhyloBayes-mpi trace files and scripts have been deposited in FigShare.

Author contributions

J.A.B. and P.P.S. designed the project. J.A.B., C.E.S.L., G. Giribet, and P.P.S. designed the taxon sampling strategy. J.A.B., C.E.S.L., C.M.B., L.R.B., T.J.C., G. Gainett, A.Z.O., E.V.W.S., E.G.R., M.S.H., G. Giribet, and P.P.S. performed fieldwork. J.A.B., C.E.S.L., C.M.B., L.R.B., T.J.C., G. Gainett, A.Z.O., E.V.W.S., and P.P.S performed sequencing. J.A.B., C.E.S.L., and P.P.S. coded morphological data, with input from C.P.A., M.S.H., W.C.W., G.H., and G. Giribet. J.A.B. performed bioinformatics and phylogenomic analyses. J.A.B. and P.P.S wrote the manuscript. P.P.S. supervised the work. All authors edited the manuscript.

Acknowledgments

Sequencing was performed at the Bauer Core Facility (Harvard University) and the BioTechnology Center (UW-Madison). High-throughput computational analyses were performed through the Center for High Throughput Computing (CHTC) and the Bioinformatics Resource Center (BRC) of UW-Madison. J.A.B. was supported by a Guyer postdoctoral fellowship. C.E.S.L. was supported by a CONACYT postdoctoral fellowship. This material is based on work supported by the National Science Foundation under Grants IOS-1552610 and IOS-2016141 to P.P.S; DEB-1457539 to G. Giribet; and DEB-1457300 to G.H.

References

M. Dabert, H. Proctor, J. Dabert, Higher-level molecular phylogeny of the water mites (Acariformes: Prostigmata: Parasitengonina: Hydrachnidiae). *Mol. Phylogenet. Evol.* **101**, 75–90 (2016).

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

J. A. Dunlop, C. Kamenz, G. Scholtz, Reinterpreting the morphology of the Jurassic scorpion *Liassoscorpionides*. *Arthropod Struct. Dev.* **36**, 245–252 (2007).

J. A. Dunlop, Geological history and phylogeny of Chelicerata. *Arthropod Struct. Dev.* **39**, 124–142 (2010).

G. Giribet, G. D. Edgecombe, The phylogeny and evolutionary history of arthropods. *Curr. Biol.* **29**, R592–R602 (2019).

G. D. Edgecombe, Arthropod origins: Integrating paleontological and molecular evidence. *Ann. Rev. Ecol. Evol. Syst.* **51**, 1–25 (2020).

P. P. Sharma, *et al.*, Phylogenomic interrogation of Arachnida reveals systemic conflicts in phylogenetic signal. *Mol. Biol. Evol.* **31**, 2963–2984 (2014).

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

J. A. Ballesteros, C. E. Santibáñez López, Ľ. Kováč, E. Gavish-Regev, P. P. Sharma, Ordered phylogenomic subsampling enables diagnosis of systematic errors in the placement of the enigmatic arachnid order Palpigradi. *Proc. R. Soc. Lond. B.* **286**, 20192426 (2019).

A. Z. Ontano, *et al.*, Taxonomic sampling and rare genomic changes overcome long-branch attraction in the phylogenetic placement of pseudoscorpions. *Mol. Biol. Evol.* **38**, 2446–2467, (2020).

W. C. Wheeler, C. Y. Hayashi, The phylogeny of the extant chelicerate orders. **14**, 173–192 (1998).

D. J. Colgan, *et al.*, Histone H3 and U2 snRNA DNA sequences and arthropod molecular evolution. *Aust. J. Zool.* 46, 419–437 (1998).

G. D. Edgecombe, G. D. F. Wilson, D. J. Colgan, M. R. Gray, G. Cassis, Arthropod cladistics: Combined analysis of histone H3 and U2 snRNA sequences and morphology. *Cladistics* **16**, 155–203 (2000).

G. Giribet, G. D. Edgecombe, W. C. Wheeler, Arthropod phylogeny based on eight molecular loci and morphology. *Nature* **413**, 157–161 (2001).

G. Giribet, G. D. Edgecombe, W. C. Wheeler, C. Babbitt, Phylogeny and systematic position of opiliones: A combined analysis of chelicerate relationships using morphological and molecular data. *Cladistics* **18**, 5–70 (2002).

J. M. Mallatt, J. R. Garey, J. W. Shultz, Ecdysozoan phylogeny and Bayesian inference: first use of nearly complete 28S and 18S rRNA gene sequences to classify the arthropods and their kin. *Mol. Phylogenet Evol.* **31**, 178–191 (2004).

J. Mallatt, G. Giribet, Further use of nearly complete 28S and 18S rRNA genes to classify Ecdysozoa: 37 more arthropods and a kinorhynch. *Mol. Phylogenet Evol.* **40**, 772–794 (2006).

S. E. Masta, S. J. Longhorn, J. L. Boore, Arachnid relationships based on mitochondrial genomes: Asymmetric nucleotide and amino acid bias affects phylogenetic analyses. *Mol. Phylogenet. Evol.* **50**, 117–128 (2009).

A. R. Pepato, C. E. da Rocha, J. A. Dunlop, Phylogenetic position of the acariform mites: sensitivity to homology assessment under total evidence. *BMC Evol. Biol.* **10**, 235 (2010).

K. L. Sanders, M. S. Y. Lee, Arthropod molecular divergence times and the Cambrian origin of pentastomids. *System. Biodivers.* **8**, 63–74 (2010).

J. Arabi, *et al.*, Nucleotide Composition of CO1 Sequences in Chelicerata (Arthropoda): Detecting New Mitogenomic Rearrangements. *J. Mol. Evol.* **74**, 81–95 (2012).

J. Lozano-Fernandez, *et al.*, Increasing species sampling in chelicerate genomic-scale datasets provides support for monophyly of Acari and Arachnida. *Nat. Commun.* **10**, 2295 (2019).

R. J. Howard, M. N. Puttick, G. D. Edgecombe, J. Lozano-Fernandez, Arachnid monophyly: Morphological, palaeontological and molecular support for a single terrestrialization within Chelicerata. *Arthropod Struct. Dev.* **59**, 100997 (2020).

N. M. Koch, L. A. Parry, Death is in our side: Paleontological data drastically modify phylogenetic hypotheses. *Syst. Biol.* **69**, 1052–1067 (2020).

N. Mongiardino Koch, J. R. Thompson, A total-evidence dated phylogeny of Echinoidea combining phylogenomic and paleontological data. *Syst. Biol.*, **70**, 421–439 (2021).

J. S. Neumann, R. Desalle, A. Narechania, B. Schierwater, M. Tessler, Morphological characters can strongly influence early animal relationships inferred from phylogenomic data sets. *Syst. Biol.* **70**, 360–375 (2021).

J. M. Wolfe, Metamorphosis is ancestral for crown euarthropods, and evolved in the Cambrian or earlier. *Integr. Comp. Biol.* **57**, 499–509 (2017).

C. Aria, J.-B. Caron, A middle Cambrian arthropod with chelicerae and proto-book gills. *Nature* **573**, 586–589 (2019).

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

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

F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).

R. M. Waterhouse, *et al.*, BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* **35**, 543–548 (2018).

M. Steel, M. J. Sanderson, Characterizing phylogenetically decisive taxon coverage. *Appl. Math. Lett.* **23**, 82–86 (2010).

H.-C. Wang, B. Q. Minh, E. Susko, A. J. Roger, Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Syst. Biol.* **67**, 216–235 (2018).

N. V. Whelan, K. M. Halanych, Who let the CAT out of the bag? Accurately dealing with substitutional heterogeneity in phylogenomic analyses. *Syst. Biol.* **66**, 232–255 (2016).

A. Criscuolo, S. Gribaldo, BMGE (Block Mapping and Gathering with Entropy): A new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* **10**, 210 (2010).

P. Simion, *et al.*, A large and consistent phylogenomic dataset supports sponges as the sister group to all other animals. *Curr. Biol.* **27**, 958–967 (2017).

A. K. Redmond, A. McLysaght, Evidence for sponges as sister to all other animals from partitioned phylogenomics with mixture models and recoding. *Nat. Commun.* **12**, 1783 (2021).

A. M. Hernandez, J. F. Ryan, Six-state amino acid recoding is not an effective strategy to offset compositional heterogeneity and saturation in phylogenetic analyses. *Syst. Biol.* (2021) https:/doi.org/10.1093/sysbio/syab027 (April 30, 2021).

H. Shimodaira, An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* **51**, 492–508 (2002).

X.-X. Shen, C. Hittinger, A. Rokas, Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nature Ecol. Evol.* **1**, 0126 (2017).

D. Huang, *et al.*, Origin of spiders and their spinning organs illuminated by mid-Cretaceous amber fossils. *Nature Ecol. Evol.* **2**, 623–627 (2018).

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

G. Alberti, A. V. Peretti, Fine structure of male genital system and sperm in Solifugae does not support a sister-group relationship with Pseudoscorpiones (Arachnida). *J. Arachnol.* **30**, 268–274 (2002).

J. A. Dunlop, J. Krüger, G. Alberti, The sejugal furrow in camel spiders and acariform mites. *Arach. Mitteil.* **43**, 29–36 (2012).

R. D. C. Bicknell, L. Lustri, T. Brougham, Revision of "*Bellinurus*" *carteri* (Chelicerata: Xiphosura) from the Late Devonian of Pennsylvania, USA. *C. R. Palevol.* **18**, 967–976 (2019).

D. J. Siveter, D. E. G. Briggs, M. D. Sutton, D. A. Legg, A new crustacean from the Herefordshire (Silurian) Lagerstätte, UK, and its significance in malacostracan evolution. *Proc. R. Soc. Lond. B* **284**, 20170279 (2017).

Shultz JW. 2001. Gross muscular anatomy of *Limulus polyphemus* (Chelicerata, Xiphosura) and its bearing on evolution in the Arachnida. *J. Arachnol.* **29**, 283–303.

J. C. Lamsdell, V. E. McCoy, O. A. Perron-Feller, M. J. Hopkins. Air breathing in an exceptionally preserved 340-million-year-old sea scorpion. *Curr. Biol.* **30**, P4316–4321 (2020).

M. J. Ramírez, I. L. F. Magalhaes, S. Derkarabetian, J. Ledford, C. E. Griswold, H. M. Wood, M. Hedin. Sequence capture phylogenomics of true spiders reveals convergent evolution of respiratory systems. *Syst. Biol.* **70**, 14–20 (2021).

G. A. Boxshall. The evolution of arthropod limbs. *Biol. Rev.* **79**, 253–300 (2004).

T. Lehmann, R. R. Melzer, Looking like *Limulus*? – Retinula axons and visual neuropils of the median and lateral eyes of scorpions. *Front. Zool.* **10**, 40 (2013).

T. Lehmann, R. R. Melzer, Also looking like *Limulus*? – Retinula axons and visual neuropils of Amblypygi (whip spiders). *Front. Zool.* **15**, 52 (2018).

T. H. Struck, A. Golombek, A. Weigert, F. A. Franke, W. Westheide, G. Purschke, C. Bleidorn, K. M. Halanych. The evolution of annelids reveals two adaptive routes to the interstitial realm. *Curr. Biol.* **25**, 1993–1999 (2015).

K. M. Kocot, A. J. Poustka, I. Stöger, K. M. Halanych, M. Schrödl. New data from Monoplacophora and a carefully-curated dataset resolve molluscan relationships. *Sci. Rep.* **10**, 101 (2020).

A. B. Smythe, O. Holovachov, K. M. Kocot, Improved phylogenomic sampling of free-living nematodes enhances resolution of higher-level nematode phylogeny. *BMC Evol. Biol.* **19**, 121 (2019).

C. Camacho, *et al.*, BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

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

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

D. T. Hoang, O. Chernomor, A. von Haeseler, B. Q. Minh, L. S. Vinh, UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).

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

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

A. Rambaut, A. J. Drummond, D. Xie, G. Baele, M. A. Suchard, Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).

E. Susko, A. J. Roger, On reduced amino acid alphabets for phylogenetic inference. *Mol. Biol. Evol.* **24**, 2139–2150 (2007).

C. P. Arango, Morphological phylogenetics of the sea spiders (Arthropoda: Pycnogonida). *Org. Divers. Evol.* **2**, 107–125 (2002).

R. J. Garwood, P. P. Sharma, J. A. Dunlop, G. Giribet, A Paleozoic stem group to mite harvestmen revealed through integration of phylogenetics and development. *Curr. Biol.* **24**, 1017–1023 (2014).

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

F. Ronquist, *et al.*, MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).

P. A. Goloboff, S. A. Catalano, TNT version 1.5, including a full implementation of phylogenetic morphometrics. *Cladistics* **32**, 221–238 (2016).

Fig. 1. Higher-level phylogeny of Chelicerata showing well-resolved groups (boldface text), based on (10). Circles indicate whole genome duplication (WGD) events subtending specific taxa. Branch lengths are not to scale.

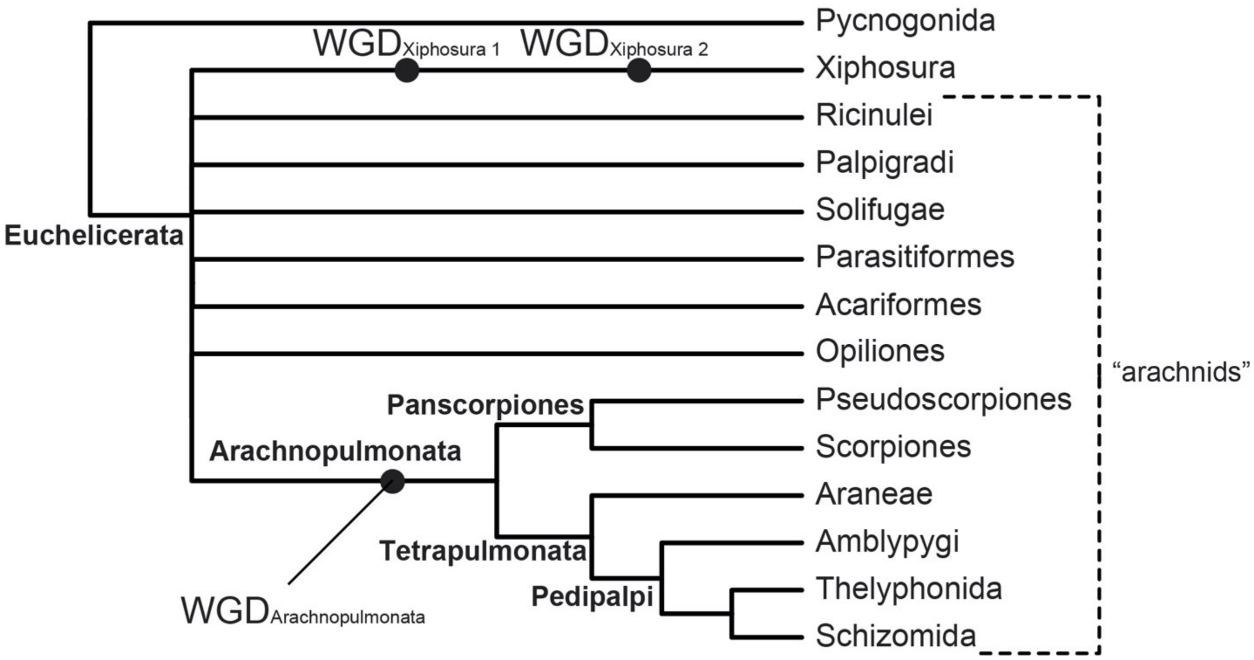


Fig. 2. Comprehensive phylogenomic sampling of all extant chelicerate orders recovers horseshoe crabs as derived arachnids. (*A*) Phylogenomic relationships of 506 chelicerate datasets based on maximum likelihood analysis of slowly evolving loci (Matrix 3) and site heterogeneous evolutionary models. Colors correspond to orders; note that Acariformes and Parasitiformes are each treated as separate orders in this study. (*B*) Summary of relationships inferred under site heterogeneous models by three matrices, under varying algorithmic approaches and orthology criteria. Numbers on nodes correspond to bootstrap resampling frequencies below 100%; all unlabeled nodes are maximally supported. (*C*), Gene (gCF) and site (sCF) concordance factors exhibit higher support for the derived placement of Xiphosura under all three 506-taxon matrices. (*D*) Tests of monophyly consistently rejected the monophyly of Arachnida and Acari over the unconstrained topology for Matrices 1–3. Non-significant result for Matrix 3 results from the unconstrained recovery of Panscorpiones in this analysis.

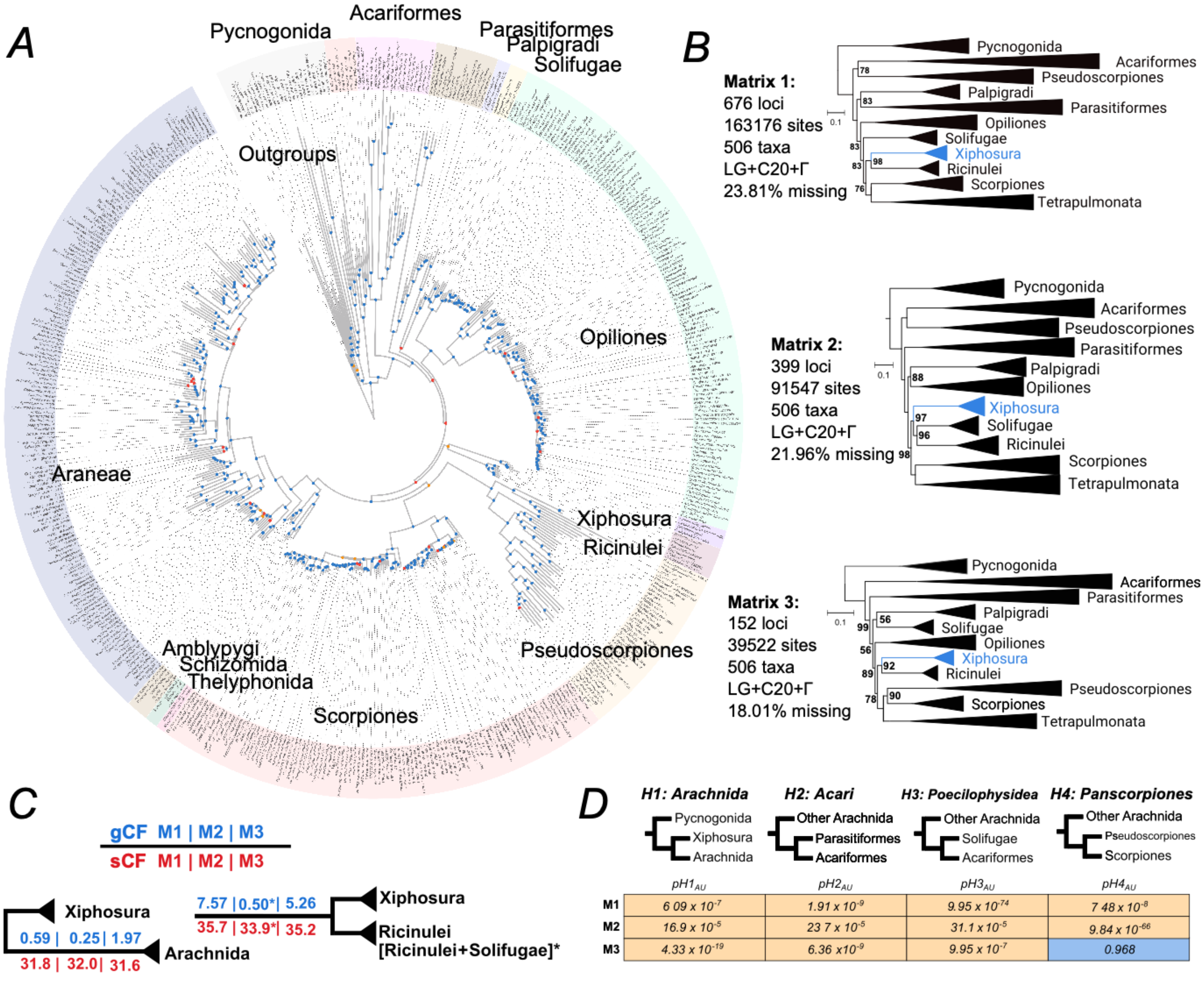


Fig. 3. Site heterogeneous model-based approaches using CAT+GTR+Γ and SR4 recoding refute the monophyly of Arachnida. (*A*) Summary tree of eight chains from PhyloBayes-mpi analysis of Matrix 4. Numbers on nodes correspond to posterior probabilities below 1.00; all unlabeled nodes are maximally supported. Lower right: Distribution of support across *a posteriori* trees for arachnid monophyly (yellow) versus nested placement of Xiphosura. (*B*) Summary trees from PhyloBayes-mpi analysis separated by starting tree topology. Top: Chains started on maximum likelihood tree topology for Matrix 4 (Xiphosura nested). Middle: Chains started on maximum likelihood tree topology for Matrix 4 with a constraint for arachnid monophyly. Bottom: Chains started on random tree topologies. Nodal support values and pie charts for each summary tree reflect the conventions for (*A*). (*C*) Maximum likelihood tree topology based on SR4 recoding and multi-profile tiered site heterogeneous models (RL2 approach; 38). Numbers on nodes correspond to bootstrap resampling frequencies below 100%; all unlabeled nodes are maximally supported.

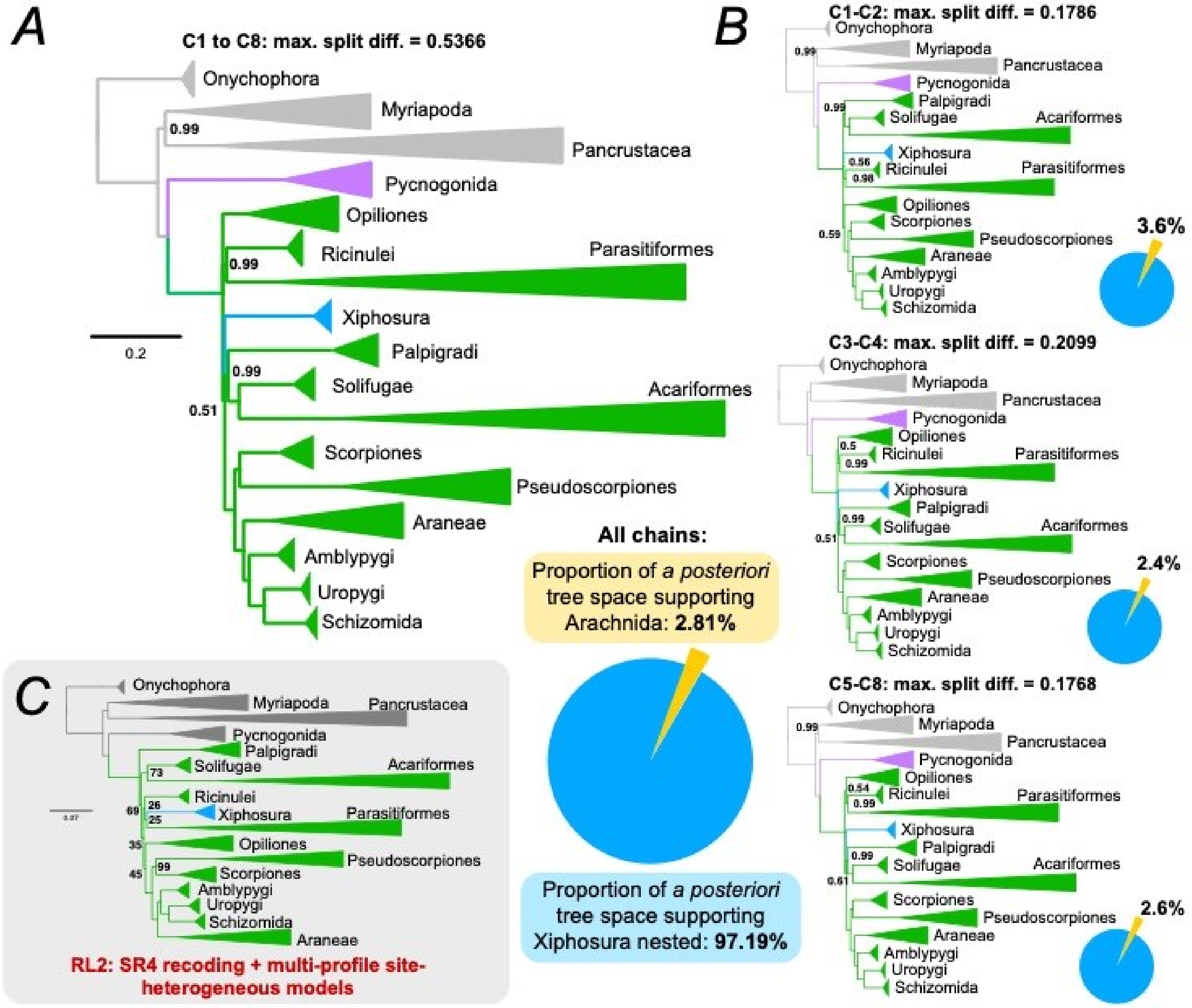


Fig. 4. Dissection of phylogenetic signal shows that a minority of artifact-prone genes support arachnid monophyly. (*A*) ΔGLS distributions mapping phylogenetic support for competing hypotheses reveal that a minority of genes (39-41%) support arachnid monophyly, regardless of orthology criterion (Matrices 1 and 2) and filtering of fast-evolving genes (Matrix 3). These proportions are similar to the proportions of genes supporting spurious groupings (*SI Appendix*, Fig. S3). (*B*) Genes supporting the derived placement of Xiphosura exhibit comparable or better metrics of systematic bias (e.g., saturation, evolutionary rate, missing data) than genes supporting Arachnida. (C) ΔSLS distributions reveal that the majority of sites (68%) support a derived placement of Xiphosura. Whereas the two categories of sites are similar with respect to missing data, sites supporting arachnid monophyly exhibit high levels of Shannon entropy (exceeding entropy values for spurious groupings; *SI Appendix*, Fig. S4).

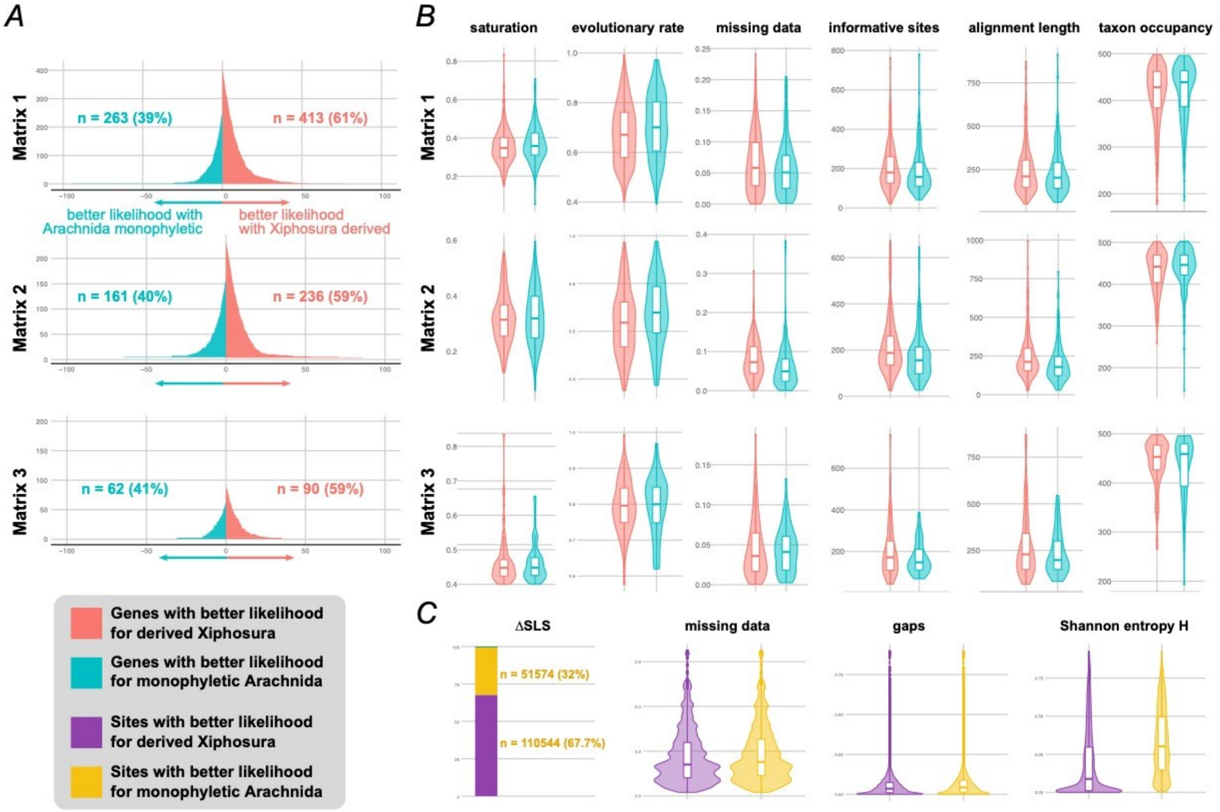


Fig. 5. Inclusion of morphology does not rescue arachnid monophyly in total evidence analyses. (*A*) Strict consensus of 1000 equally parsimonious trees inferred for a morphological matrix of 291 morphological characters. Merostomata comprises extinct groups Eurypterida (sea scorpions), Chasmataspidida, and Synziphosurina, as well as Xiphosurida, including the extant Limulidae. Numbers on nodes indicate bootstrap resampling frequencies. (*B*) Summary tree from Bayesian inference analysis of morphological matrix. Numbers on nodes indicate posterior probabilities below 1.00; unlabeled nodes are maximally supported. (*C*) Maximum likelihood total evidence topology based on 152 slowly-evolving genes and morphological characters. Numbers on nodes represent bootstrap resampling frequencies; unlabeled nodes are maximally supported. Note that the timing of WGD events in Xiphosura cannot be pinpointed on the branches subtending this group.