Phylogenomic Interrogation of Arachnida Reveals Systemic Conflicts in Phylogenetic Signal

Prashant P. Sharma,*[1] Stefan T. Kaluziak,[2] Alicia R. Pérez-Porro,[3] Vanessa L. González,[3] Gustavo Hormiga,[4] Ward C. Wheeler,[1] and Gonzalo Giribet[3]

[1]Division of Invertebrate Zoology, American Museum of Natural History, New York, NY

[2]Marine Science Center, Northeastern University

[3]Department of Organismic and Evolutionary Biology, Museum of Comparative Zoology, Harvard University

[4]Department of Biological Sciences, George Washington University

***Corresponding author:** E-mail: psharma@amnh.org.

Associate editor: Nicolas Vidal

Abstract

Chelicerata represents one of the oldest groups of arthropods, with a fossil record extending to the Cambrian, and is sister group to the remaining extant arthropods, the mandibulates. Attempts to resolve the internal phylogeny of chelicerates have achieved little consensus, due to marked discord in both morphological and molecular hypotheses of chelicerate phylogeny. The monophyly of Arachnida, the terrestrial chelicerates, is generally accepted, but has garnered little support from molecular data, which have been limited either in breadth of taxonomic sampling or in depth of sequencing. To address the internal phylogeny of this group, we employed a phylogenomic approach, generating transcriptomic data for 17 species in combination with existing data, including two complete genomes. We analyzed multiple data sets containing up to 1,235,912 sites across 3,644 loci, using alternative approaches to optimization of matrix composition. Here, we show that phylogenetic signal for the monophyly of Arachnida is restricted to the 500 slowest-evolving genes in the data set. Accelerated evolutionary rates in Acariformes, Pseudoscorpiones, and Parasitiformes potentially engender long-branch attraction artifacts, yielding nonmonophyly of Arachnida with increasing support upon incrementing the number of concatenated genes. Mutually exclusive hypotheses are supported by locus groups of variable evolutionary rate, revealing significant conflicts in phylogenetic signal. Analyses of gene-tree discordance indicate marked incongruence in relationships among chelicerate orders, whereas derived relationships are demonstrably robust. Consistently recovered and supported relationships include the monophyly of Chelicerata, Euchelicerata, Tetrapulmonata, and all orders represented by multiple terminals. Relationships supported by subsets of slow-evolving genes include Ricinulei + Solifugae; a clade comprised of Ricinulei, Opiliones, and Solifugae; and a clade comprised of Tetrapulmonata, Scorpiones, and Pseudoscorpiones. We demonstrate that outgroup selection without regard for branch length distribution exacerbates long-branch attraction artifacts and does not mitigate gene-tree discordance, regardless of high gene representation for outgroups that are model organisms. Arachnopulmonata (new name) is proposed for the clade comprising Scorpiones + Tetrapulmonata (previously named Pulmonata).

Key words: **Arthropoda, arachnids, concatenation, topological conflict, transcriptomics, orthology prediction, ancient rapid radiation.**

Introduction

Subsequent to decades of pitched contention over the interrelationships of Arthropoda, a consensus has begun to emerge regarding the shape of the arthropod tree of life. With the advent of transcriptome- and genome-scale data, the monophyly of Arthropoda and its three major rami—Chelicerata, Myriapoda, and Pancrustacea (alternatively, "Tetraconata")—is strongly supported, together with a sister group relationship of arthropods and velvet worms (e.g., Hejnol et al. 2009; Campbell et al. 2011; Rota-Stabelli et al. 2011). Multiple phylogenomic efforts have been directed toward resolving internal relationships of Pancrustacea, which unite hexapods and the paraphyletic "crustaceans," and have contributed to the identification of the hexapod sister group (Regier et al. 2010; Meusemann et al. 2010; Simon et al. 2012; von Reumont et al. 2012). The interrelationships of Chelicerata remain among the last major challenges for arthropod systematics. Chelicerates constitute the second largest branch of the arthropod tree of life and include several iconic lineages, such as horseshoe crabs, spiders, and scorpions.

Morphological phylogenetic analyses have divided extant Chelicerata into three groups: Pycnogonida (sea spiders), Xiphosura (horseshoe crabs), and Arachnida (terrestrial chelicerates) (e.g., Weygoldt and Paulus 1979; Shultz 1990, 2007). Implicit in this schema is the minimization of terrestrialization events, namely by the arachnid common ancestor. The mutual monophyly of Pycnogonida and Euchelicerata (=Xiphosura + Arachnida) is presently supported by a number of morphological characters, embryological evidence,

Article

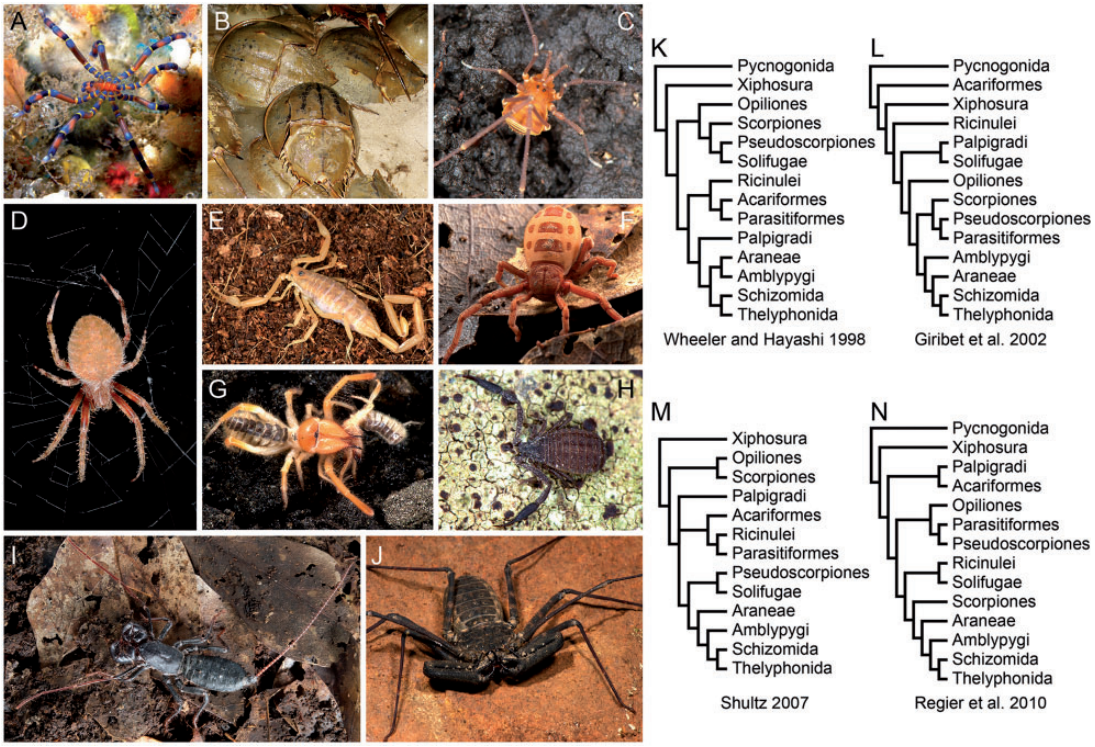

Fig. 1. (*A–J*) Exemplars of chelicerate diversity. (*A*) *Anoplodactylus evansi* (Pycnogonida), (*B*) *Limulus polyphemus* (Xiphosura) spawning, (*C*) *Pachylicus acutus* (Opiliones), (*D*) *Neoscona arabesca* (Araneae), (*E*) *Centruroides sculpturatus* (Scorpiones), (*F*) *Ricinoides atewa* (Ricinulei), (*G*) *Eremobates* sp. (Solifugae) consuming a smaller conspecific, (*H*) *Synsphyronus apimelus* (Pseudoscorpiones), (I) *Mastigoproctus giganteus* (Thelyphonida), (*J*) *Damon variegatus* (Amblypygi). (*K–N*) Hypotheses of chelicerate ordinal relationships, from (*K*) Wheeler and Hayashi (1998) (morphology and ribosomal sequence data); (*L*) Giribet et al. (2002) (two ribosomal genes); (*M*) Shultz (2007) (morphology); and (*N*) Regier et al. (2010) (62 nuclear genes, analysis with degenerated codons). Image of *Anoplodactylus evansi* courtesy of Mick Harris. Image of *Limulus polyphemus* courtesy of Peter Funch.

and molecular sequence data (Wheeler et al. 1993; Wheeler and Hayashi 1998; Mallatt et al. 2004; Regier et al. 2010), including phylogenomic approaches (e.g., Roeding et al. 2009; Meusemann et al. 2010). In contrast, due to extensive morphological character conflict, basal relationships of a putatively monophyletic Arachnida have remained recalcitrant to resolution, and many hypotheses have been proposed for these interrelationships (e.g., Shultz 1990, 2007; Wheeler and Hayashi 1998; Giribet et al. 2002) (fig. 1). For example, mites (Acariformes) and ticks (Parasitiformes) are traditionally united as Acari by such characters as the subcapitulum and the "hexapod larva" (a six-legged postembryonic stage, which also occurs in Ricinulei); this relationship conflicts with the presence of a prominent sejugal furrow in camel spiders (Solifugae) and mites (Shultz 2007; Dunlop et al. 2012), as well as the fundamentally divergent morphology of the two constituent lineages of Acari (Van der Hammen 1989). Some authors disputed even the monophyly of Arachnida, on the basis of "primitive" characters occurring in scorpions and Opiliones (harvestmen) (e.g., Van der Hammen 1989).

Paralleling morphology, molecular data have not resolved relationships within Arachnida either, much less affirmed the validity of this group (fig. 1). Analyses of nucleotide sequence data infrequently recover the monophyly of Arachnida, although very few studies have comprehensively sampled all extant arachnid orders (Wheeler and Hayashi 1998; Giribet et al. 2002; Masta et al. 2009, 2010; Roeding et al. 2009; Meusemann et al. 2010; Pepato et al. 2010; Regier et al. 2010; Arabi et al. 2012; von Reumont et al. 2012). The largest phylogenetic effort sampling all extant arachnid lineages, based on 62 Sanger-sequenced loci, recovered strong support only for the monophyly of Euchelicerata, Tetrapulmonata (Araneae, Amblypygi, and Uropygi), and relationships within the tetrapulmonates; Arachnida was modestly supported in that study by only two of four analyses, and basal relationships of arachnids were highly unstable (Regier et al. 2010).

Multiple aspects of Arachnida make the internal resolution of this group prone to recalcitrance. First, the origins of chelicerates are presumed to be ancient, occurring during or prior to the Cambrian, with rapid diversification of extant arachnid orders (Dunlop 2010; Rota-Stabelli et al. 2013). A considerable number of arachnid lineages equal in rank to orders has gone extinct, such as trigonotarbids (early arachnids present by the Silurian) and phalangiotarbids (arachnids of unknown affinity present in the Devonian and Carboniferous), in addition to non-arachnid chelicerate groups such as eurypterids and chasmataspidids (sea scorpions), and synziphosurines (stem-group horseshoe crabs or stem-group euchelicerates) (Dunlop 2010; Briggs et al. 2012; Lamsdell 2013). Many extant orders also demonstrate the signature of prolonged historic extinction, that is, marked spans of time between origin and diversification, and the retention of few extant species. This phenomenon is epitomized by horseshoe crabs, which

appeared in the fossil record in the Ordovician and are survived by just four extant species that diverged in the Cretaceous (Obst et al. 2012). The combination of ancient diversification and subsequent extinctions has presumably produced short internodes in extant chelicerate phylogeny, a documented affliction for phylogenetic reconstruction (Rokas and Carroll 2006; Salichos and Rokas 2013).

Second, difficulty of phylogenetic resolution is compounded by accelerated evolution in certain arachnid orders. Specifically, mitochondrial genes of Pseudoscorpiones, Acariformes, and Parasitiformes (or at least Mesostigmata) have demonstrably higher rates of evolution, and larger variability in nucleotide composition, than other Euchelicerata (Arabi et al. 2012). Similarly, branch lengths inferred from nuclear ribosomal gene alignments suggest heterotachy in the same groups (as well as Palpigradi and some araneomorph spiders), even upon accounting for secondary structure (Pepato et al. 2010). Significantly accelerated rates of evolution in certain lineages underlie long-branch attraction (LBA), the artifactual close relationship of lineages with long-branch lengths (Felsenstein 1978; Huelsenbeck 1997; Bergsten 2005; Philippe et al. 2005). Because outgroup taxa are often undersampled, the placement of long-branch taxa sister to, or part of a grade with, outgroup terminals is symptomatic of LBA artifacts. Application of sequence data at genomic scales, compounded by inclusion of relatively few terminals, can result in maximal support values for spurious placement of long-branch terminals (reviewed by Bergsten 2005).

A third and final aspect of arachnid biology hindering phylogenetic resolution may be variability of genome size. At one extreme, the genome of the mite *Tetranychus urticae* is among the smallest arthropod genomes, comprising approximately 18,224 coding genes (Grbic et al. 2011). In contrast, the genome of the scorpion *Mesobuthus martensii* contains approximately 32,016 genes, more than any other animal genome (Cao et al. 2013). For reference, approximately 15,771 genes occur in the fruit fly *Drosophila melanogaster* (FlyBase Consortium, Release 5), and approximately 21,000–25,000 in *Homo sapiens* (Genome Reference Consortium). Furthermore, whereas genome compaction in mites is associated with gene loss (Grbic et al. 2011), scorpions appear to have undergone high gene family turnover, as well as gene duplications and neofunctionalization of resulting paralogs (Cao et al. 2013; Sharma et al. 2014). Such marked historical disparity in the evolution of gene families across arachnids may have hindered accurate assessments of orthology in data sets composed of Sanger-sequenced nuclear genes, particularly some commonly utilized phylogenetic markers assumed to be single-copy orthologs throughout arthropods (e.g., Hedin et al. 2010), but subsequently shown to contain numerous paralogs using cloning techniques and/or transcriptomic resources (Riesgo et al. 2012; Clouse et al. 2013).

Toward improving resolution within Arachnida, we compiled a data set of 48 taxa, including new transcriptomic libraries for 17 species published herein, two complete genomes (the mite *T. urticae* and the tick *Ixodes scapularis*), and previously published expressed sequence tag (EST) and Illumina platform libraries. Sampled outgroup taxa to Arachnida consisted of Xiphosura and Pycnogonida, as well as basally branching Mandibulata (myriapods) and Onychophora. We used recently developed tools for orthology prediction and optimization of matrix composition. The largest data set of 3,644 genes and greater than 1.2 million amino acid sites (at 25% gene occupancy) included exemplars of all extant arachnid orders except for Schizomida (the sister group of Thelyphonida and part of Uropygi). Here we show that phylogenetic analysis of the 3,644-ortholog data set yields the nonmonophyly of Arachnida with 100% bootstrap support, due to suspect placement of Pseudoscorpiones, Acariformes, and Parasitiformes. We demonstrate that this result is an LBA artifact by examining nodal support as a function of concatenation by evolutionary rate. Intriguingly, a set of genes supporting the monophyly of Arachnida strongly conflicts with other such sets, which support other hypotheses also consistent with morphology (e.g., the placement of Pseudoscorpiones), indicating nonuniformity of phylogenetic signal. Significant incongruence among gene trees corroborates the pervasiveness of phylogenetic conflict at the base of Arachnida. Nevertheless, we demonstrate consistent support for the monophyly of Chelicerata, Euchelicerata, and several derived clades of arachnids.

Results

Orthology Assignment and Supermatrix Assembly

We sequenced cDNA from 16 arachnid and one xiphosuran species using the Illumina GAII and HiSeq 2500 platforms, and combined these with published transcriptomes and genomes from 30 other species, including some libraries sequenced in our laboratories (Riesgo et al. 2012; Fernández et al. 2014). Assembly statistics for new libraries, as well as published transcriptomes assembled from raw reads, are provided in table 1. Subsequent to open reading frame (ORF) prediction and selection of the longest ORFs per unigene, 3,055–18,814 peptide sequences were retained for Illumina libraries and complete genomes, 815–1,127 for 454 Life Science platform libraries, and 54–10,646 for Sanger-sequenced EST libraries. Orthology assessment of this 48-taxon data set with the Orthologous MAtrix (OMA) stand-alone algorithm recovered 77,774 orthogroups. Subsequent to alignment and culling of ambiguously aligned positions, a supermatrix was constructed by retaining a minimum gene occupancy threshold of 13 taxa (i.e., occupancy of >25% of the data set), resulting in 3,644 concatenated orthologs (1,235,912 aligned sites; 36.23% occupied). The number of orthologs per taxon varied from 6 (the palpigrade *Prokoenenia wheeleri*; based on published Sanger data) to 3,215 (the harvestman *Sitalcina lobata*). Due to the paucity of data for the palpigrade, its phylogenetic placement was analyzed separately.

Maximum-Likelihood Analyses

Phylogenetic analysis of the 3,644-ortholog matrix under maximum likelihood (ML) recovered a fully resolved tree topology with 100% bootstrap support for the monophyly of Chelicerata and Euchelicerata (fig. 2). Contrary to the

Table 1. Species Sequenced and Analyzed in This Study.

	Order	Source	Raw Reads	Contigs/ESTs/Genes	Peptides	Filtered Peptides
Onychophora						
Epiperipatus sp.		GenBank (Sanger ESTs)	—	1,868	698	480
Peripatopsis sedgwicki		GenBank (454)	—	10,476	5,489	2,552
Peripatopsis capensis		De novo (Illumina HiSeq)	40,774,042	92,516	17,452	12,846
Myriapoda						
Archispirostreptus gigas	Diplopoda	GenBank (454)	—	4,008	1,237	815
Alipes grandidieri	Chilopoda	GenBank (Illumina GAII)	32,332,034	147,732	34,505	18,814
Scutigera coleoptrata	Chilopoda	GenBank (Sanger ESTs)	—	2,400	961	570
Pycnogonida						
Anoplodactylus eroticus	Pycnogonida	GenBank (454)	—	3,744	2,291	1,127
Endeis spinosa	Pycnogonida	GenBank (454)	—	4,069	3,302	2,656
Xiphosura						
Limulus polyphemus	Xiphosura	De novo (Illumina HiSeq)	65,099,444	110,362	33,041	17,824
Carcinoscorpius rotundicauda	Xiphosura	GenBank (Sanger ESTs)	—	575	202	199
Arachnida						
Damon variegatus	Amplypygi	De novo (Illumina GAII)	64,733,221	64,715	19,686	11,823
Acanthoscurria gomesiana	Araneae	GenBank (Sanger ESTs)	—	6,790	2,203	1,567
Liphistius malayanus	Araneae	De novo (Illumina HiSeq)	62,897,982	61,775	16,526	11,221
Frontinella communis	Araneae	De novo (Illumina HiSeq)	66,183,160	155,930	53,936	18,978
Leucauge venusta	Araneae	De novo (Illumina HiSeq)	49,301,974	189,630	62,318	17,591
Neoscona arabesca	Araneae	De novo (Illumina HiSeq)	57,103,328	150,856	43,252	16,594
Blomia tropicalis	Acariformes	GenBank (Sanger ESTs)	—	1,432	925	888
Glycyphagus domesticus	Acariformes	GenBank (Sanger ESTs)	—	2,589	1,309	1,303
Suidasia medadensis	Acariformes	GenBank (Sanger ESTs)	—	3,585	2,323	1,766
Tetranychus urticae	Acariformes	GenBank (whole genome)	—	18,313	18,313	14,276
Amblyomma americanum	Parasitiformes	GenBank (Sanger ESTs)	—	6,502	1,266	971
Amblyomma variegatum	Parasitiformes	GenBank (Sanger ESTs)	—	3,992	2,248	1,419
Dermacentor andersoni	Parasitiformes	GenBank (Sanger ESTs)	—	3,994	614	452
Ixodes scapularis	Parasitiformes	GenBank (whole genome)	—	20,473	20,473	15,288
Rhipicephalus appendiculatus	Parasitiformes	GenBank (Sanger ESTs)	—	19,123	10,778	6,303
Rhipicephalus microplus	Parasitiformes	GenBank (Sanger ESTs)	—	52,902	20,939	10,646
Larifuga sp.	Opiliones	De novo (Illumina HiSeq)	38,354,659	115,648	23,750	10,510
Metabiantes sp.	Opiliones	De novo (Illumina HiSeq)	72,266,554	85,203	30,754	12,843
Metasiro americanus	Opiliones	GenBank (Illumina GAII)	—	83,792	30,041	16,556
Pachylicus acutus	Opiliones	De novo (Illumina HiSeq)	18,681,026	69,803	23,379	14,202
Phalangium opilio	Opiliones	De novo (Illumina GAII)	16,225,145	131,889	40,234	15,277
Vonones ornata	Opiliones	De novo (Illumina HiSeq)	66,096,788	114,248	37,076	19,208
Leiobunum verrucosum	Opiliones	GenBank (Illumina HiSeq)	—	42,636	26,407	13,329
Protolophus singularis	Opiliones	GenBank (Illumina HiSeq)	—	287,150	63,477	13,987
Ortholasma coronadense	Opiliones	GenBank (Illumina HiSeq)	—	45,807	23,846	10,780
Trogulus martensi	Opiliones	GenBank (Illumina HiSeq)	—	55,077	23,233	12,765
Hesperonemastoma modestum	Opiliones	GenBank (Illumina HiSeq)	—	55,565	32,409	8,845
Sclerobunus nondimorphicus	Opiliones	GenBank (Illumina HiSeq)	—	56,727	26,599	11,518
Sitalcina lobata	Opiliones	GenBank (Illumina HiSeq)	—	64,104	26,410	14,828
Siro boyerae	Opiliones	GenBank (Illumina HiSeq)	—	38,618	18,657	11,387
Prokoenenia wheeleri	Palpigradi	GenBank (Sanger ESTs)	—	56	54	54
Synsphyronus apimelus	Pseudoscorpiones	De novo (Illumina HiSeq)	65,443,655	103,556	31,567	17,820
Pseudocellus pearsei	Ricinulei	De novo (Illumina HiSeq)	91,403,481	29,077	7,687	5,922
Ricinoides atewa	Ricinulei	De novo (Illumina GAII)	52,766,585	97,327	31,331	14,324
Centruroides vittatus	Scorpiones	De novo (Illumina HiSeq)	45,691,843	14,215	3,859	3,055
Pandinus imperator	Scorpiones	GenBank (454)	—	17,253	2,219	2,209
Eremobates sp.	Solifugae	De novo (Illumina GAII)	86,794,767	94,687	25,474	11,765
Mastigoproctus giganteus	Uropygi	De novo (Illumina GAII)	25,983,006	116,600	32,827	17,674

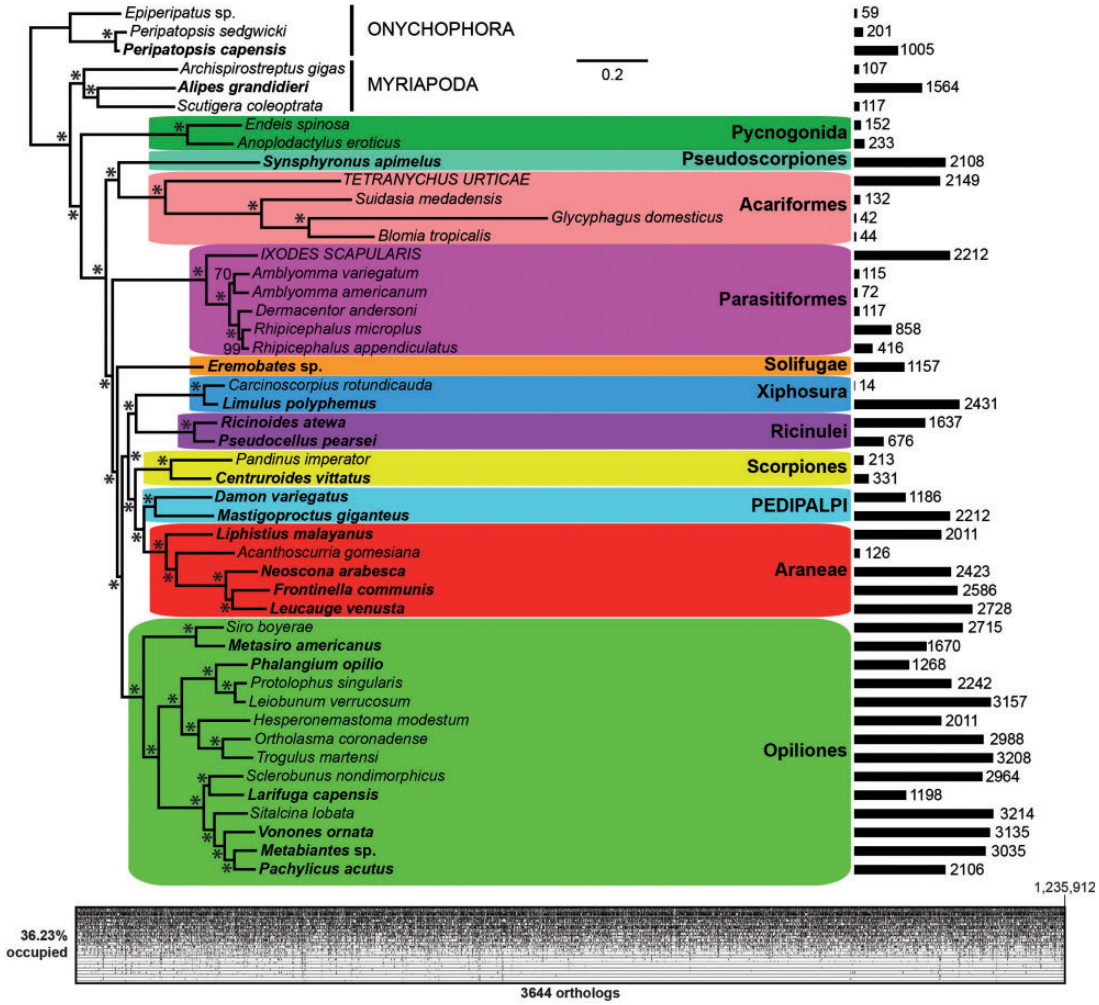

FIG. 2. Relationships of Arachnida inferred from ML analysis of 3,644 orthologs. Numbers on nodes indicate bootstrap resampling frequencies (asterisks indicate a value of 100%). Uppercase text indicates terminals represented by complete genomes; boldface lowercase text indicates terminals with novel transcriptomic data from this study. Bars on right represent number of genes available for each terminal. Below: Global view of the sequence alignment.

traditional basal split within Euchelicerata between Xiphosura and the terrestrial Arachnida, the ML analysis supported the paraphyly of Arachnida, with Xiphosura sister to Ricinulei. The clade Pseudoscorpiones + Acariformes was recovered sister to the remaining Euchelicerata; this clade in turn formed part of a grade with Acariformes with respect to the remaining Euchelicerata and Xiphosura. All orders represented by multiple terminals were recovered as monophyletic with 100% bootstrap support, often in spite of marked disparities in gene representation. The greatest such disparity is exemplified by the Illumina-sequenced library of *Limulus polyphemus*, which has over 2 orders of magnitude greater representation than the Sanger-sequenced EST library of *Carcinoscorpius rotundicauda* (2,431 vs. 14 orthologs, respectively); support for Xiphosura was nevertheless maximal. All nodes in this topology were fully supported, excepting the two corresponding to sister relationships of two sets of congeneric ticks (*Amblyomma* and *Rhipicephalus*). Longer branch lengths were observed for the orders Acariformes, Parasitiformes, and Pseudoscorpiones, with respect to the remaining taxa.

Matrix Reduction and Concatenation by Gene Occupancy

To assess the impact of heterogeneous levels of missing data, we generated three supermatrices. Two of these were generated using the MARE algorithm, retaining either all 47 terminals (516 orthologs, 122,436 sites, 39.99% occupied, call handle "MARE1") or only the 30 libraries comprising Illumina data sets and whole genomes (1,237 orthologs, 373,349 sites, 62.49% occupied; call handle "MARE2"). A third matrix was generated by removing from our original 3,644-ortholog alignment all terminals with smaller libraries, that is, the same number of taxa as MARE2 (3,644 orthologs, 1,235,912 sites, 55.34% occupied, call handle "IL-GEN").

In spite of limited topological differences, none of these three matrices recovered the monophyly of Arachnida (fig. 3 and supplementary fig. S1, Supplementary Material online). In the two 30-taxon matrices (MARE2 and IL-GEN), *L. polyphemus* was consistently recovered as nested within a paraphyletic Arachnida with strong support (bootstrap resampling frequency [BS] $\geq$ 98%).

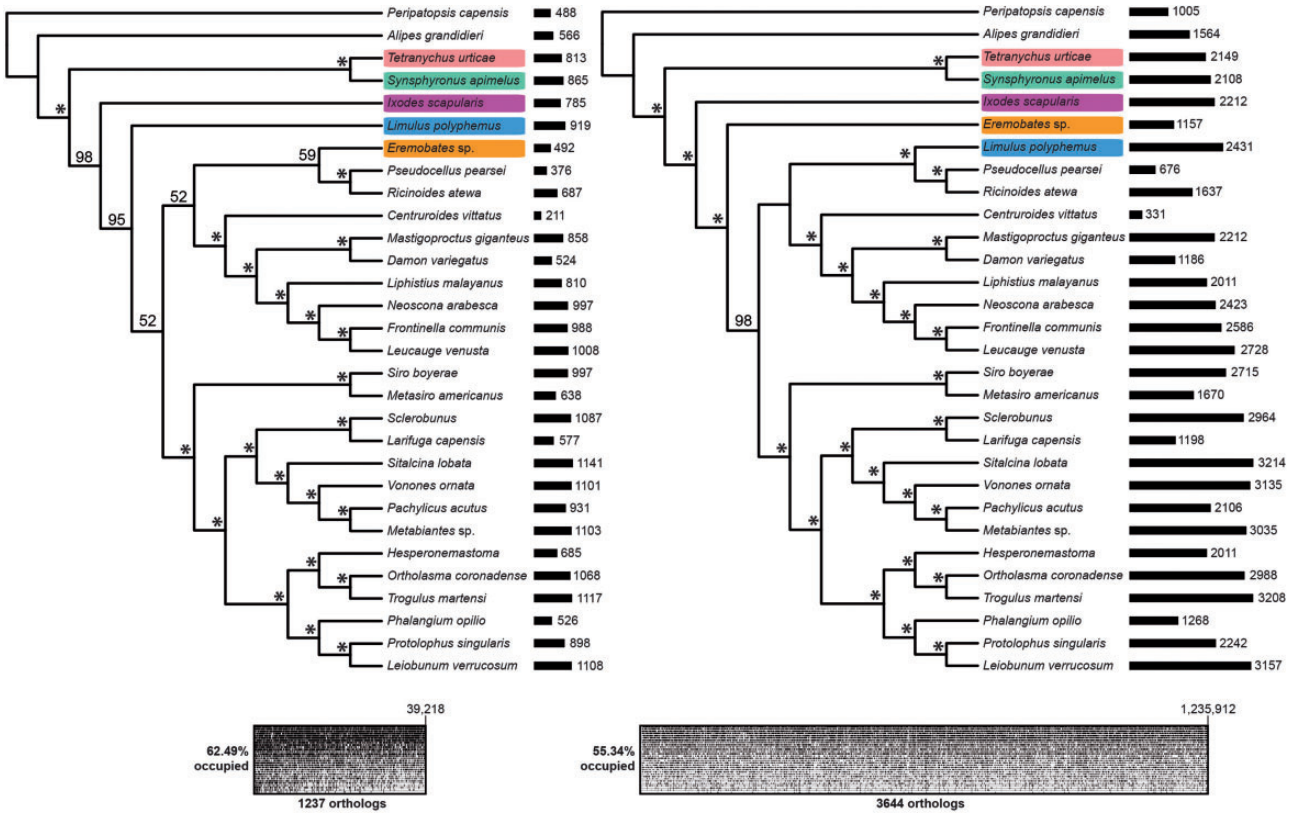

Fig. 3. Left: Cladogram of arachnid relationships inferred from ML analysis post-matrix reduction, retaining only terminals with Illumina-sequenced transcriptomes or complete genomes. Right: Cladogram of chelicerate relationships inferred from ML analysis upon manual exclusion of Sanger-sequenced EST and 454 library data from the 3,644-ortholog supermatrix. Orders of interest are indicated with shading. Numbers on nodes indicate bootstrap resampling frequencies. Bars on right represent number of genes available for each terminal. Below: Global view of the sequence alignment.

In all three topologies, Parasitiformes and the clade Pseudoscorpiones + Acariformes formed a grade sister to the remaining Euchelicerata, and together with Solifugae in IL-GEN (a topology congruent with the supermatrix topology in fig. 2).

To infer the impact of gene occupancy on phylogenetic signal, another six matrices were constructed wherein the threshold of gene occupancy was serially increased by five taxa, up to representation for a minimum of 35 out of 47 taxa. A summary of matrix composition and amount of missing data for these six matrices is provided in table 2. Bootstrap resampling frequency for various topological hypotheses of interest showed no consistent effect of missing data (fig. 4). Support for Arachnida was equal to or near zero for all matrices except for one, which enforced gene occupancy as representation in at least 30 taxa (table 2, fig. 4). The ML tree topology based on this exceptional matrix (98 orthologs, 16,594 sites, 68.20% occupancy, call handle "AL30") recovered the monophyly of Arachnida with moderate support (BS = 79%; supplementary fig. S2, Supplementary Material online).

Concatenation by Percent Pairwise Identity

To assess how nodal support for alternative hypotheses of chelicerate relationships is affected by rate of molecular evolution, a series of 15 matrices was assembled wherein

Table 2. Nodal Support for Three Ingroup Clades as a Function of Gene Occupancy Threshold.

Gene Occupancy Threshold	Proportion Missing/Gaps	Number of Orthologs (sites)	Bootstrap Resampling Frequency		
			Chelicerata	Euchelicerata	Arachnida
At least 13 taxa	0.638	3,644 (1,235,912)	100	100	0
At least 15 taxa	0.606	2,806 (896,210)	100	100	0
At least 20 taxa	0.520	1,152 (312,646)	100	100	0
At least 25 taxa	0.448	352 (77,516)	100	100	0
At least 30 taxa	0.318	98 (16,594)	99	100	79
At least 35 taxa	0.249	23 (4,021)	72	100	1

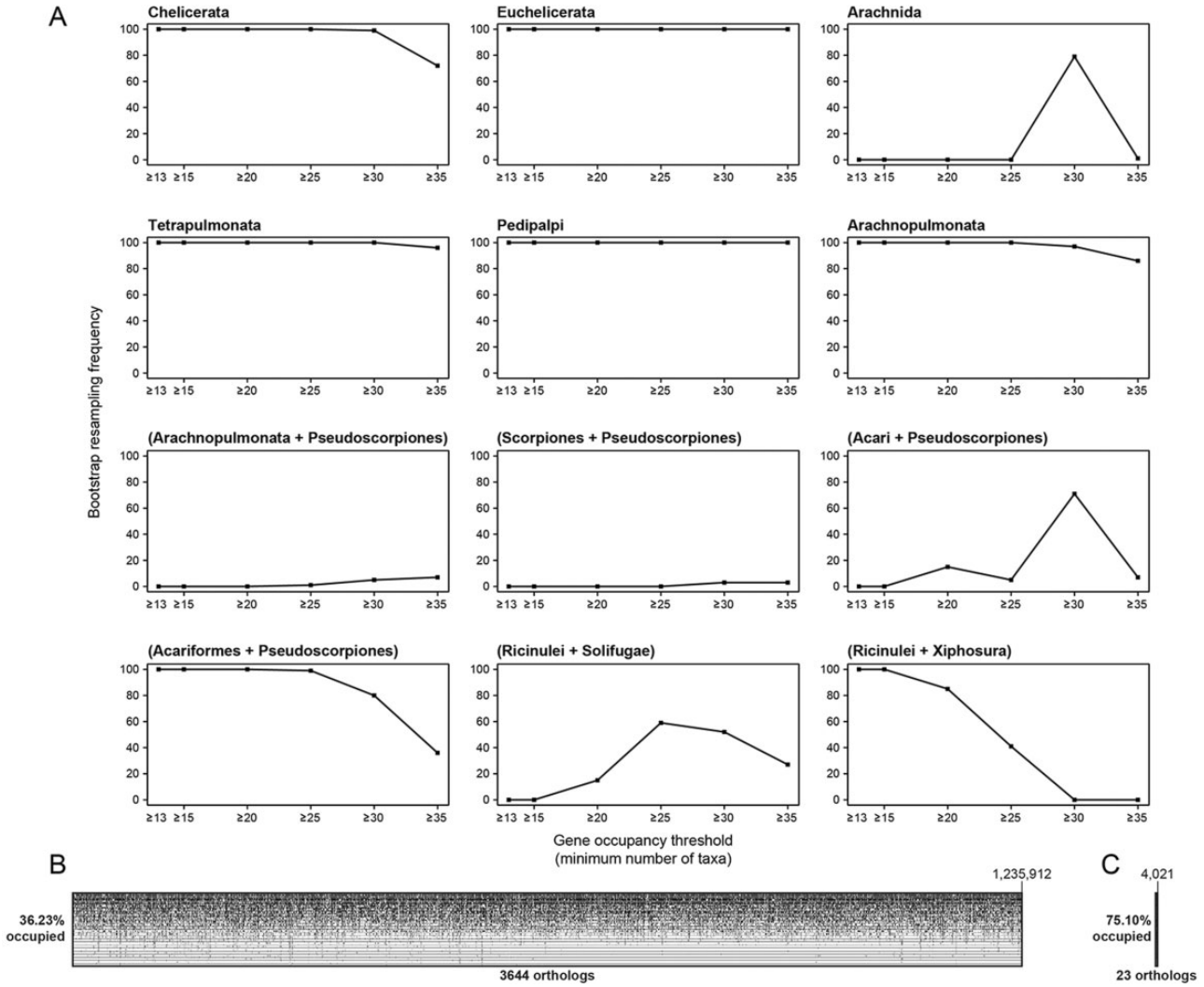

FIG. 4. (*A*) Bootstrap resampling frequency for phylogenetic hypotheses inferred from concatenation of orthologs by gene occupancy. (*B*) Global view of sequence alignment of largest matrix (≥13 taxa). (*C*) Global view of sequence alignment of densest matrix (≥35 taxa).

orthologs were concatenated in order of percent pairwise identity, beginning with the most conserved (i.e., similar). We added genes at increments of 100 until 1,000 genes had been concatenated, and increments of 500 subsequently until 3,500 genes had been concatenated. We scored the bootstrap resampling frequency for various topological hypotheses. For such nodes as Chelicerata and Eucheliceratа, we observed a trajectory of nodal support theoretically expected with accruing data: monotonic or nearly monotonic increase in bootstrap resampling frequency until fixation at the maximum value (fig. 5). A similar result was obtained for some derived relationships that are consistent with morphology, such as Pedipalpi (Amblypygi + Uropygi), Tetrapulmonata (Araneae + Pedipalpi), and a clade we term Arachnopulmonata **new name**, which comprises Scorpiones + Tetrapulmonata. Albeit recovered with moderate support in the majority of their analyses, Regier et al. (2010) used the term "Pulmonata" for this clade, a name introduced by Firstman (1973) for a clade of pulmonate arachnids, but which had already been in use for decades for a group of gastropods. To facilitate systematic discourse more broadly, we rename this group Arachnopul-monata.

In contrast, the trajectory corresponding to Arachnida showed initial increase to maximum bootstrap frequency upon concatenation of the 500 slowest-evolving orthologs. Thereafter, nodal support for Arachnida decreased to zero by the concatenation of the 1,000th slowest-evolving ortholog. As illustrated in figure 6, the topology inferred from the 500 slowest-evolving orthologs recovers three separate clades within Arachnida: Arachnopulmonata; a clade comprised of Opiliones, Solifugae, and Ricinulei (a trio of orders with a completely segmented opisthosoma and a tracheal tubule system for respiration); and a clade comprised of Pseudoscorpiones, Acariformes, and Parasitiformes.

Like Arachnida, a peak of support is observed for the placement of Pseudoscorpiones as part of a clade with Arachnopulmonata. Specifically, much of that nodal support places Pseudoscorpiones as sister group to Scorpiones (figs. 5 and 6). Unlike Arachnida, nodal support for these hypotheses is more restricted to the 200 slowest-evolving orthologs,

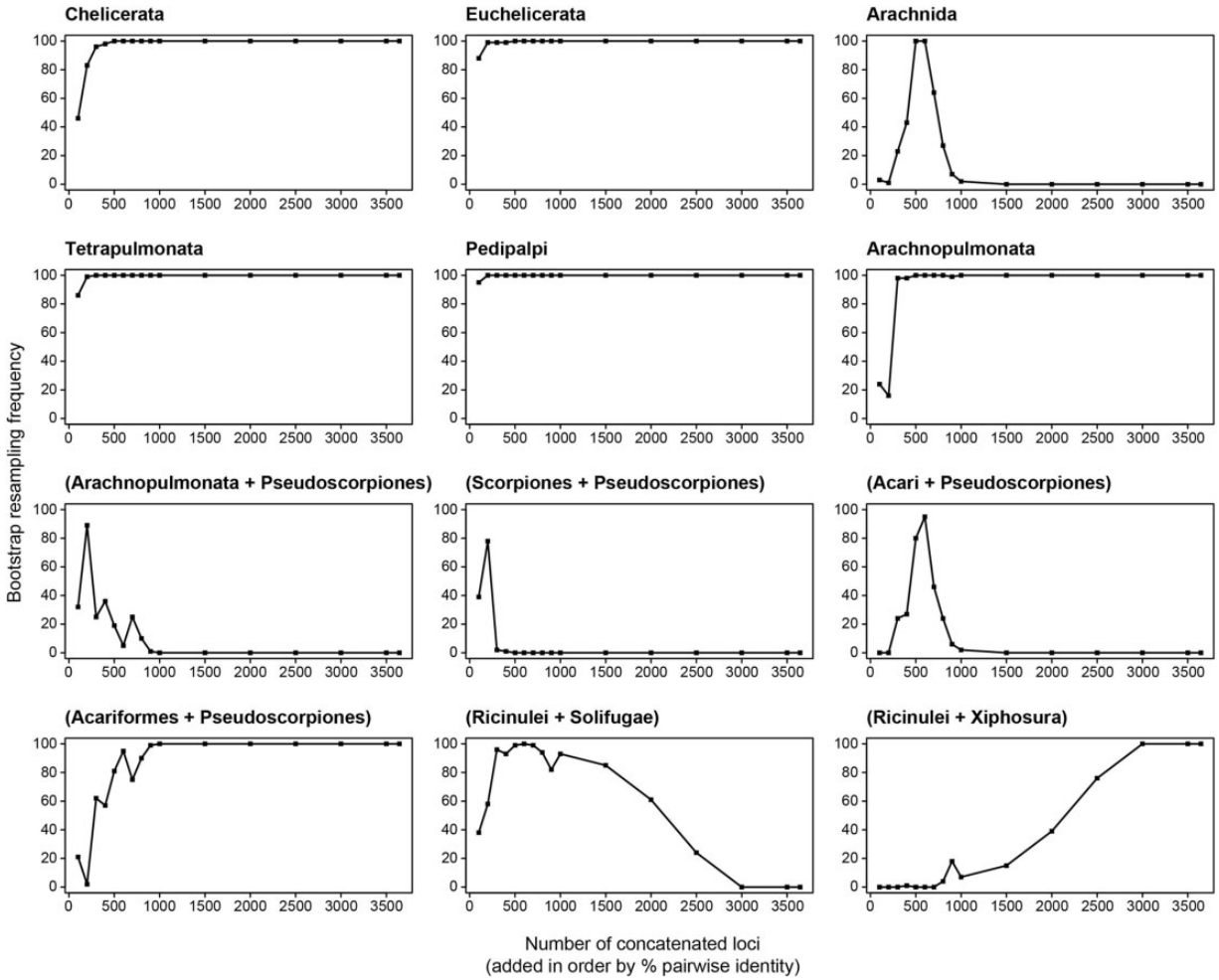

FIG. 5. Bootstrap resampling frequency for phylogenetic hypotheses inferred from sequential concatenation of orthologs in order of percent pairwise identity (most to least conserved).

which are insufficient to support the monophyly of Arachnida (fig. 6).

To ensure that results from this method of concatenation were not driven by missing data, we ranked the 3,644 orthologs in order of percent pairwise identity and mapped the oldest most recent common ancestor (MRCA) recoverable for each orthogroup. Dense representation was observed for the MRCAs of Arachnida, Euchelicerata and Arthropoda, and the root of the tree (supplementary fig. S3, Supplementary Material online). Considering the composition of three matrices of significance (200-slowest evolving orthologs, 500-slowest evolving orthologs, complete matrix), the proportion of orthologs wherein at least the MRCA of Arachnida was represented exceeded 98% for all matrices. Only 72 (1.98%) of orthologs in the complete matrix were restricted to sampling of derived arachnids groups, and these were evenly distributed with respect to ordination of evolutionary rate. These results indicate that missing data affected the sampling of basal nodes in a nonsignificant and random way.

Bayesian Inference Analyses

Runs of PhyloBayes implementing infinite mixture models (GTR + CAT) were implemented for the 500 slowest-evolving ortholog data set and for MARE1 did not achieve convergence ($0.3 <$ maximum discrepancy < 1) after 7 months of computation on 500 parallelized processors. We examined posterior distributions under the four independent runs for both data sets and observed that all chains had recovered the same consensus topology for each data set. Under either data set, internodes of infinitesimal length separated major clades within Euchelicerata; the result is indistinguishable from a basal polytomy for both data sets (fig. 7). Bayesian inference analysis of the 500 slowest-evolving ortholog data set recovered the clade (Pseudoscorpiones + Scorpiones), but not arachnid monophyly (fig. 7 compare with fig. 6). The two Bayesian inference topologies were almost identical due to extensive overlap of orthologs between the two data sets; of the 516 genes in MARE1, 301 are shared with the 500 slowest-evolving ortholog set.

Discussion

Conflict, Resolution

The phylogenetic data set amassed in this study is the largest deployed to test basal relationships of Arachnida to date. Nevertheless, the tree topology recovered by ML analysis of the 3,644 supermatrix (fig. 2) recapitulates a recurring

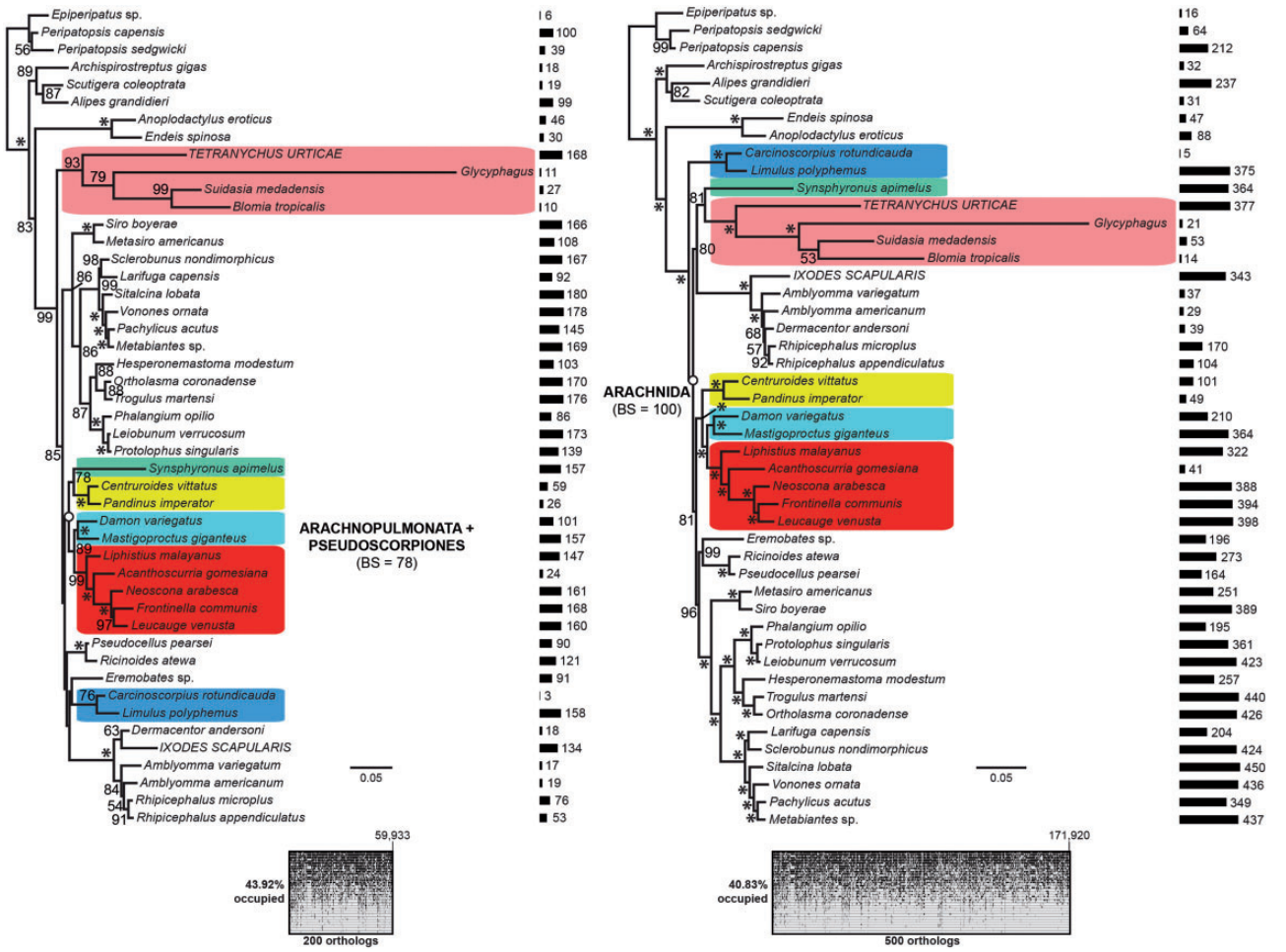

FIG. 6. Left: Tree topology inferred from ML analysis of the 200 slowest-evolving genes. Right: Tree topology inferred from ML analysis of the 500 slowest-evolving genes. Orders of interest are indicated with shading. Numbers on nodes indicate bootstrap resampling frequencies. Bars on right represent number of genes available for each terminal. Below: Global views of the sequence alignments.

phenomenon in molecular phylogenetic studies of Chelicerata: The putative nonmonophyly of Arachnida supported by this analysis is a counterintuitive result, overturning a key relationship supported by morphological cladistic studies (e.g., Wheeler and Hayashi 1998; Giribet et al. 2002; Shultz 2007). Specifically, the nested placement of Xiphosura has been recovered in other phylogenomic assessments, albeit with incomplete sampling of arachnid orders (Roeding et al. 2009; Meusemann et al. 2010). Shultz (2007) observed that arachnids are putatively united by a series of synapomorphies, such as aerial respiration, an anteriorly/anteroventrally directed mouth, and the loss of some characters observed in marine chelicerates (e.g., walking leg gnathobases, cardiac lobe). Counterarguments to morphological support for arachnid monophyly, which are grounded in convergence of characters driven by independent episodes of terrestrialization, are difficult to falsify, due to the fragmentary nature of the fossil record and limitations in parameterizing the historical probability of terrestrialization events (Shultz 2007; von Reumont et al. 2012).

However, we regarded with suspicion the placement of three orders (Pseudoscorpiones, Acariformes, and Parasitiformes) with known higher rates of molecular evolution as a grade at the base of Eucheliceratа with maximum bootstrap support, whether based on transcriptomic data (Roeding et al. 2009), nuclear ribosomal genes (Pepato et al. 2010), or mitochondrial genes (Arabi et al. 2012). High or even optimal nodal support in genomic-scale analyses can belie inaccuracy of the tree topology, interpartition conflicts, and other biases (Felsenstein 1978; Rokas et al. 2003; Salichos and Rokas 2013) and may be symptomatic of an LBA phenomenon (reviewed by Bergsten 2005). Furthermore, the inclusion of smaller (454 and Sanger-sequenced EST) libraries in our data set engenders the possibility that the nonmonophyly of Arachnida is an artifact stemming from missing data.

The effect of missing data in phylogenetic analysis is controversial. Simulations have shown that highly incomplete terminals can be reliably placed in phylogenies, and often confer advantages over excluding such terminals, but may engender LBA artifacts in some cases (Wiens 2006). Other analyses of empirical data sets have implicated missing data as misleading, specifically with respect to inflating nodal support values for problematic nodes, but data sets selected to balance representation of sequence data have not been

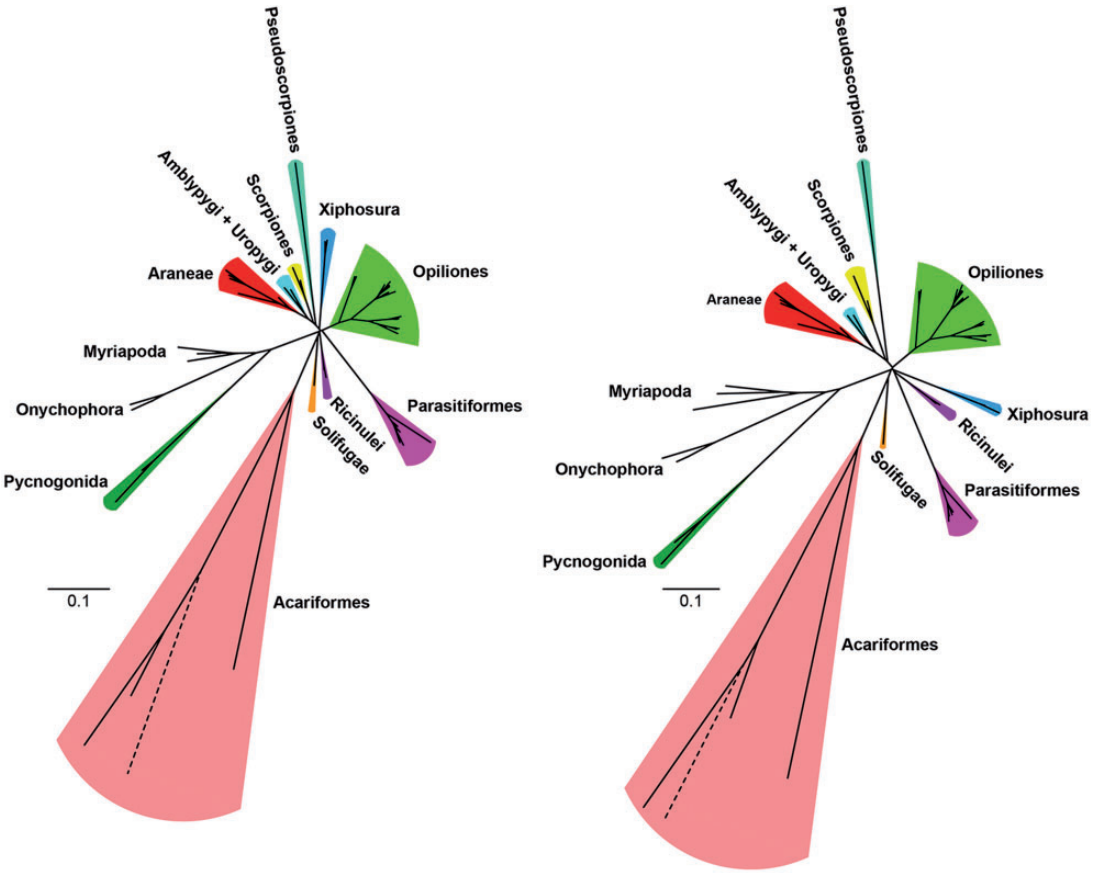

Fig. 7. Consensus of PhyloBayes tree topologies across four independent runs. Left: Tree topology inferred using the 500 slowest-evolving genes. Right: Tree topology inferred using the MARE1 matrix. Dotted line represented abbreviated branch length for *Glycyphagus domesticus* (lengths of 1.545 and 1.076 for left and right topologies, respectively).

shown to resolve ambiguous nodes either (Dell'Ampio et al. 2014; Fernández et al. 2014).

To address this possibility, we combined concatenation approaches with ML inference of gene-tree topologies. In the 3,644-ortholog data set, all arachnid orders were represented by at least one Illumina data set or complete genome, and we thus calculated large numbers of potentially informative gene trees (i.e., trees containing at least one member of each descendant branch and two distinct outgroups) for all nodes corresponding to interordinal relationships, in the range of 701–3,116 for nodes corresponding to basal arachnid relationships (fig. 8). However, the proportion of those potentially informative gene trees that were congruent with a given node varied markedly. Generally, internodes corresponding to the monophyly of derived clades (e.g., orders sampled with multiple terminals; Tetrapulmonata) were characterized by a higher proportion of congruent gene trees, in contrast to basal nodes within Euchelicerata. These data suggest that the gene tree incongruence observed is not entirely attributable to missing data.

We conducted additional analyses to investigate whether nonmonophyly of Arachnida was attributable to missing data. Neither manual exclusion of 454 and Sanger-sequenced libraries nor matrix reduction with MARE yielded arachnid monophyly for their respective 30-terminal data sets (fig. 3), although the measures of matrix completeness for these reduced data sets are comparable to phylogenomic counterparts among pancrustaceans and myriapods (e.g., Meusemann et al. 2010; von Reumont et al. 2012; Fernández et al. 2014). Matrix reduction with MARE while retaining all 47 terminals yielded nonmonophyly of Arachnida as well, with nonsignificant nodal support at the base of Euchelicerata (supplementary fig. S1, Supplementary Material online).

Sequential concatenation of six matrices with varying level of gene occupancy did not suggest a uniform effect of missing data either for several relationships of interest (fig. 4). The smallest matrix with gene occupancy of over 35 taxa per gene (23 orthologs; 75.10% sequence occupancy) recovered some relationships with slightly lower node support that were robustly supported elsewhere (e.g., Chelicerata, Tetrapulmonata, Arachnopulmonata), consistent with the theoretical expectation of increased nodal support with accruing data for robust nodes. Support for Arachnida was consistently low, excepting its recovery with moderate support in the AL30 matrix (gene occupancy of at least 30 taxa, 98 orthologs, 68.20% sequence occupancy; supplementary fig. S2, Supplementary Material online). This recovery of a monophyletic Arachnida through exclusive use of molecular data, albeit with limited support, is a rare result (Regier et al. 2010).

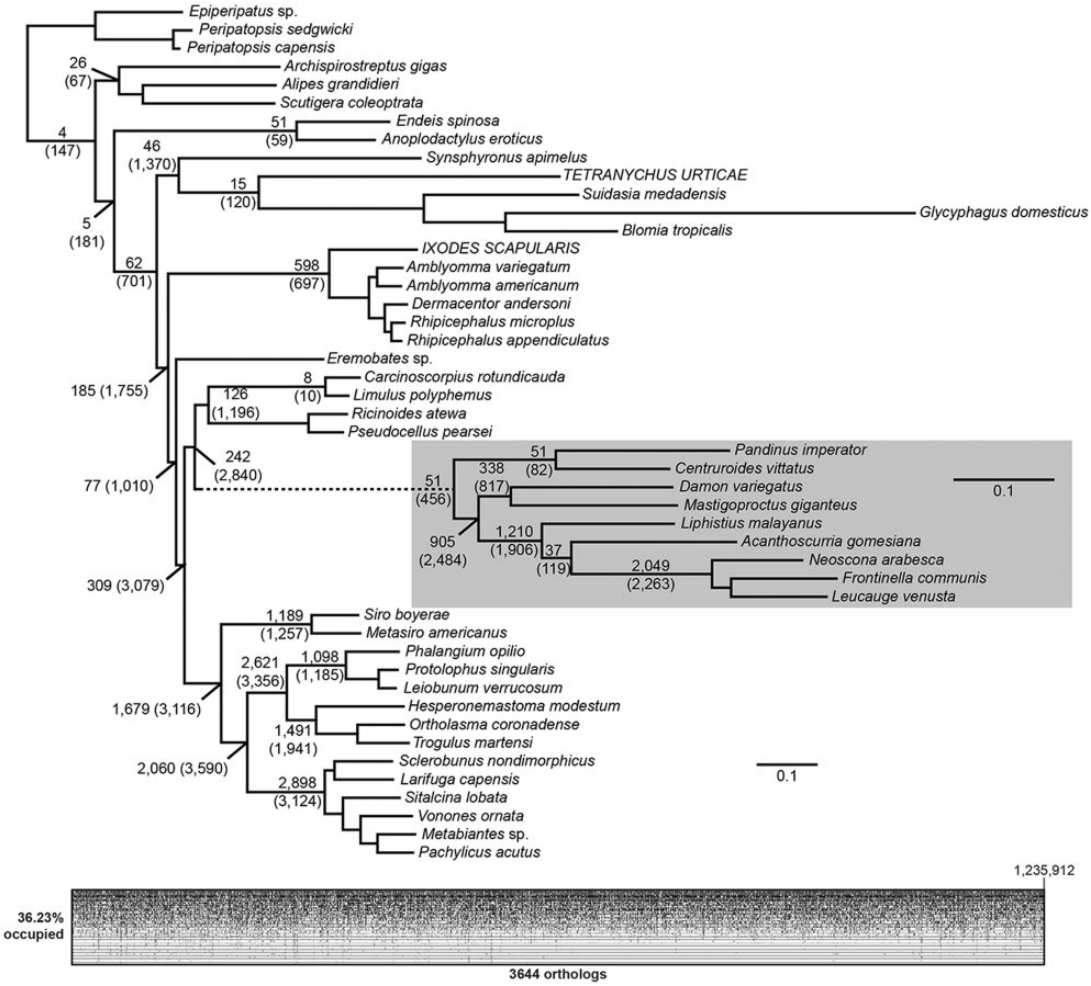

FIG. 8. Relationships of Arachnida inferred from ML analysis of 3,644 orthologs (as in fig. 2). Numbers on nodes indicate number of gene trees congruent with nodes; numbers in parentheses indicate total number of potentially informative genes (for nodes of interest). Area shaded in gray indicates Tetrapulmonata, with branch lengths extended for clarity.

Nevertheless, as this threshold of gene occupancy (representation for at least 30 taxa) is at neither extreme of the parameter's spectrum of values, we considered this result insufficient evidence for a systemic and pervasive effect of missing data on tree topology inference. We therefore investigated another aspect of our data set that might reveal phylogenetic support for Arachnida, namely, rate of molecular evolution.

Serial concatenation by evolutionary rate was conducted for 15 matrices wherein orthologs were concatenated in order of percent pairwise identity, beginning with the most conserved. In contrast to concatenation by gene occupancy, we observed distinct patterns in trajectories of nodal support upon addition of genes, enabling discernment of robustly supported nodes that were insensitive to evolutionary rate (e.g., Chelicerata, Euchelicerata, Tetrapulmonata) and nodes that were supported by a subset of genes with slower rates (e.g., Arachnida, Pseudoscorpiones + Scorpiones) (fig. 5). Intriguingly, the matrix of the 200- and 500-slowest evolving orthologs recovered interrelationships between arachnid clades characterized by very short internodes in spite of relatively large numbers of potentially informative genes at those internodes (supplementary figs. S4 and S5, Supplementary Material online), corroborating previous inferences of ancient and rapid radiation inferred from the fossil record and molecular phylogenetic efforts (Giribet et al. 2002; Dunlop 2010; Regier et al. 2010).

We also considered the possibility that evolutionary rate and degree of missing data were conflated. We dissected the relationship between these two variables and found that the matrices concatenated by evolutionary rate had approximately the same amount of gene occupancy (36.23–43.92%), with slightly higher occupancy in the slowest-evolving data sets (supplementary fig. S6, Supplementary Material online). In contrast, matrices concatenated by gene occupancy exhibit a clear negative relationship with evolutionary rate, with denser matrices including a larger proportion of slowly evolving genes (supplementary fig. S6, Supplementary Material online)—genes that are more conserved across taxa. The latter correlation suggests that evolutionary rate is a stronger explanatory variable than amount of missing data, not only for phylogenetic signal but also for the amount of missing data itself. In spite of this trend, concatenation in order of percent pairwise identity did not affect

representation of basal nodes (supplementary fig. S3, Supplementary Material online), with the ingroup (MRCA of Arachnida) represented in over 98% of orthologs. The negative relationship between evolutionary rate and gene occupancy could reflect biological pattern (e.g., orthologs with conserved function and sequence may be expressed in many taxa), the algorithmic operation of OMA (Altenhoff et al. 2013), or a combination of the two.

The number of different ML topologies obtained with various matrices is strongly suggestive of systemic incongruence of phylogenetic signal, attributable in part to LBA artifacts; in all topologies we obtained from the incremental concatenation series with nonmonophyly of Arachnida, arachnid paraphyly was caused by the suspect placement of Pseudoscorpiones, Acariformes, and/or Parasitiformes outside of the clade uniting the remaining arachnids and Xiphosura (figs. 2, 3, 6, and 7 and supplementary fig. S1, Supplementary Material online). Consistent with this prediction as well, two separate Bayesian inference analyses (the 500-slowest evolving orthologs and MARE1) yielded a polytomy or infinitesimal branch lengths at the base of Arachnida (fig. 7), with long branches subtending the problematic orders. The LG4X + F and CAT + GTR models utilized in ML and Bayesian inference analyses, respectively, were chosen for their superior ability to accommodate site-heterogeneous patterns of molecular evolution, particularly in the event of LBA artifacts (Lartillot and Philippe 2004, 2008; Le et al. 2012). But upon examination of branch lengths across postburn-in samples from the PhyloBayes runs (based on the 500 slowest-evolving ortholog data set), we continued to observe significantly longer branch lengths under a Bayesian relative rates test (Wilcox et al. 2004) for all three problematic orders (Acariformes, Parasitiformes, and Pseudoscorpiones) with respect to the MRCA of Euchelicerata, in comparison to the remaining euchelicerate taxa ($P << 0.0001$ for all comparisons; fig. 9). Together with the ML topologies, these analyses reinforce the inference of short internodes in the euchelicerate tree of life, precipitated by rapid divergence of basal lineages, and marked heterotachy in problematic orders.

However, concatenation methods have been critiqued for their tendency to mask phylogenetic conflict when strong gene tree incongruence is incident (Jeffroy et al. 2006; Kubatko and Degnan 2007; Nosenko et al. 2013; Salichos and Rokas 2013), and some authors have advocated against the use of bootstrap resampling in phylogenomic analyses (Salichos and Rokas 2013; Dell'Ampio et al. 2014). Several newly developed metrics for quantifying gene tree incongruence are inapplicable to partial trees (e.g., internode certainty; Salichos and Rokas 2013). Therefore, we visualized the dominant bipartitions among the ML gene tree topologies by constructing supernetworks using the SuperQ method (Grünewald et al. 2013), which decomposes all gene trees into quartets to build supernetworks where edge lengths correspond to quartet frequencies (see Fernández et al. 2014). We inferred supernetworks for four data sets: 1) 3,644 orthologs, 2) the 200 slowest-evolving orthologs, 3) the 500 slowest-evolving orthologs, and 4) the 1,237 orthologs of the MARE2 data set, which retain only the 30 data-rich taxa. Invariably, we observed an unresolved star topology, with numerous reticulations corresponding to basal divergences of Arachnida (fig. 10). The retention of gene tree incongruence irrespective of various approaches to matrix optimization (e.g., matrix reduction, gene occupancy, evolutionary rate) indicates that conflicts at the base of Arachnida are not solely attributable to stochastic sampling error, but are systemic. The greater amount of reticulation in supernetworks of slowly evolving genes corroborates previous reports of greater gene tree incongruence in slowly evolving partitions (Salichos and Rokas 2013; but see Betancur-R. et al. 2014)

The sum of these analyses suggests that phylogenetic resolution of Arachnida is beleaguered by incongruent signal in different genomic regions, which in turn is a function of evolutionary rate and exacerbated by LBA artifacts in rapidly evolving lineages. The combination of long-branch taxa, incident heterotachy, and discordant results stemming from consideration of evolutionary rate closely reflects historical disputes over the root of Metazoa, and tradeoffs between amount and parameterization of sequence data (Dunn et al. 2008; Philippe et al. 2009, 2011; Schierwater et al. 2009; Nosenko et al. 2013). Beyond demonstrating concealed phylogenetic support for arachnid monophyly, these results are illuminative in understanding why a morphologically well-circumscribed group has historically proven difficult to recover in molecular phylogenetic analyses. We note that concatenation by order of evolutionary rate proved an effective approach of identifying this phylogenetic incongruence. Serial concatenation strategies that rely upon random sampling of orthologs (e.g., RADICAL; Narechania et al. 2012; Simon et al. 2012) were trialed in our data set, but did not identify incongruence when it was highly localized. For example, unless a random concatenation series captures a significant proportion of that subset supporting Pseudoscorpiones + Arachnopulmonata (200 out of 3,644 orthologs), this incongruence will not be identified. Apropos, of the 98 orthologs in the AL30 matrix, which supported the monophyly of Arachnida, 47 are among the 500 slowest evolving orthologs, further indicating that rate of evolution has greater explanatory power than missing data toward accounting for arachnid nonmonophyly.

Evaluating Historical Hypotheses Based on Morphology: Pseudoscorpions and Solifugae

We observed a nodal support profile similar to that of Arachnida for the clades (Pseudoscorpiones + Acari) and (Ricinulei + Solifugae) (fig. 5), neither historically supported by analyses of morphological characters (see Shultz 2007 for the most recent analysis and summary of hypotheses). Monophyly of Acari (Acariformes + Parasitiformes) is contentious, as molecular sequence data do not uniformly support this clade (Giribet et al. 2002; Pepato et al. 2010; Regier et al. 2010; but see Meusemann et al. 2010) and relevant morphological characters may have undergone extensive convergence. For example, a six-legged postembryonic stage occurs not only in Acariformes and Parasitiformes but also in the distantly related Ricinulei. The embryology and genome

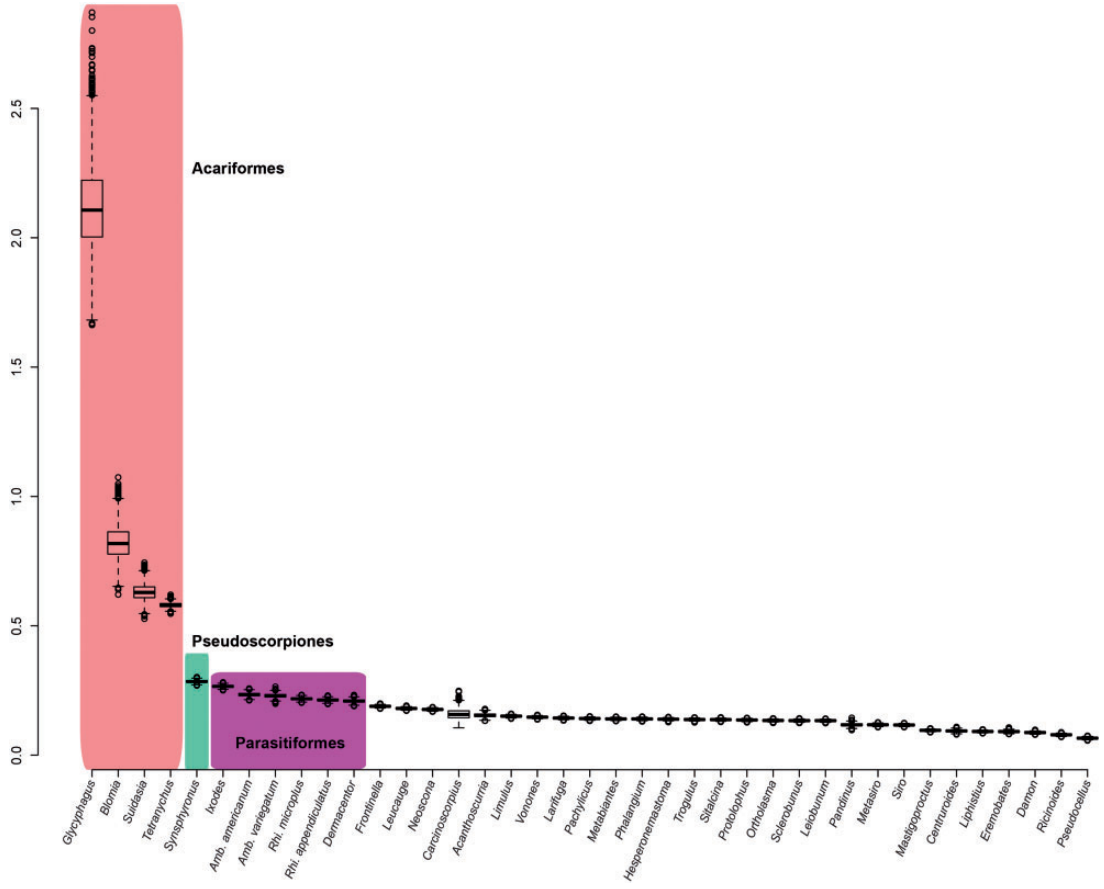

Fig. 9. Bayesian relative rates test, indicating patristic distances from MRCA of Eucheliceratа to constituent terminals. Samples are drawn from 4,000 postburn-in tree topologies (1,000 randomly sampled trees per chain) from analyses of the 500 slowest-evolving ortholog data set in Phylobayes.

of *T. urticae* have also demonstrated the loss of opisthosomal segmentation and such conserved genes as the Hox transcription factor *abdominal-A*, found in all other surveyed arachnids (Hughes and Kaufman 2002; Grbic et al. 2011; Sharma et al. 2012b).

A greater measure of confidence was once placed in the sister relationship of pseudoscorpions and solifuges (Weygoldt and Paulus 1979; Shultz 1990, 2007; but see Dunlop et al. 2012), with these forming part of a clade with scorpions and harvestmen (fig. 1) (Weygoldt and Paulus 1978; Shultz 1990, 2007; Giribet et al. 2002). The sister relationship of Pseudoscorpiones and Solifugae is consistent with a number of morphological characters, such as two-segmented chelicerae, the shape of the epistome (a structure covering the mouth), and similarity in walking leg segments (Shultz 2007). However, recent advances in comparative developmental genetics have elucidated a simple mechanism for the transition from the three- to the two-segmented chelicerael state: Loss of the expression domain of the transcription factor *dachshund*, which in three-segmented chelicerae is uniquely expressed in the proximal segment (Prpic and Damen 2004; Sharma et al. 2012a, 2013; Barnett and Thomas 2013). Similarly, the highly variable gnathobasic mouthparts of spiders (the pedipalpal "maxilla"), mites (the rutellum), and harvestmen (the stomotheca) all result from the same embryonic tissues, whose growth is disrupted by abrogation of the appendage-patterning transcription factor *Distal-less* (Schoppmeier and Damen 2001; Khila and Grbic 2007; Sharma et al. 2013). These data underscore the evolutionary lability of cheliceral segment number and enditic mouthparts, and disfavor their utility as phylogenetic character systems.

Few morphological characters could potentially unite Pseudoscorpiones and Acari. Nevertheless, the relationship of Pseudoscorpiones and Acariformes is consistent with the presence of prosomal silk glands in these orders (produced from cheliceral glands in the former and salivary glands in some lineages of the latter; in contrast, spider silk is produced from dedicated opisthosomal organs). A reciprocal BLAST (Basic Local Alignment Search Tool) search against spider and *Tetranychus* silk protein sequences revealed no hits in the transcriptome of the pseudoscorpion *Synsphyronus apimelus*, and thus we are presently unable to evaluate the homology of mite and pseudoscorpion silks based on their protein structure.

Interestingly, in contrast to Arachnida, nodal support for the clade (Ricinulei + Solifugae) is more recalcitrant to concatenation of fast-evolving genes. Only after concatenation of 3,000 orthologs does nodal support for this relationship become fixed at zero (fig. 5). Neither this relationship, nor the placement of Opiliones as sister group of (Ricinulei + Solifugae), has been supported previously in the

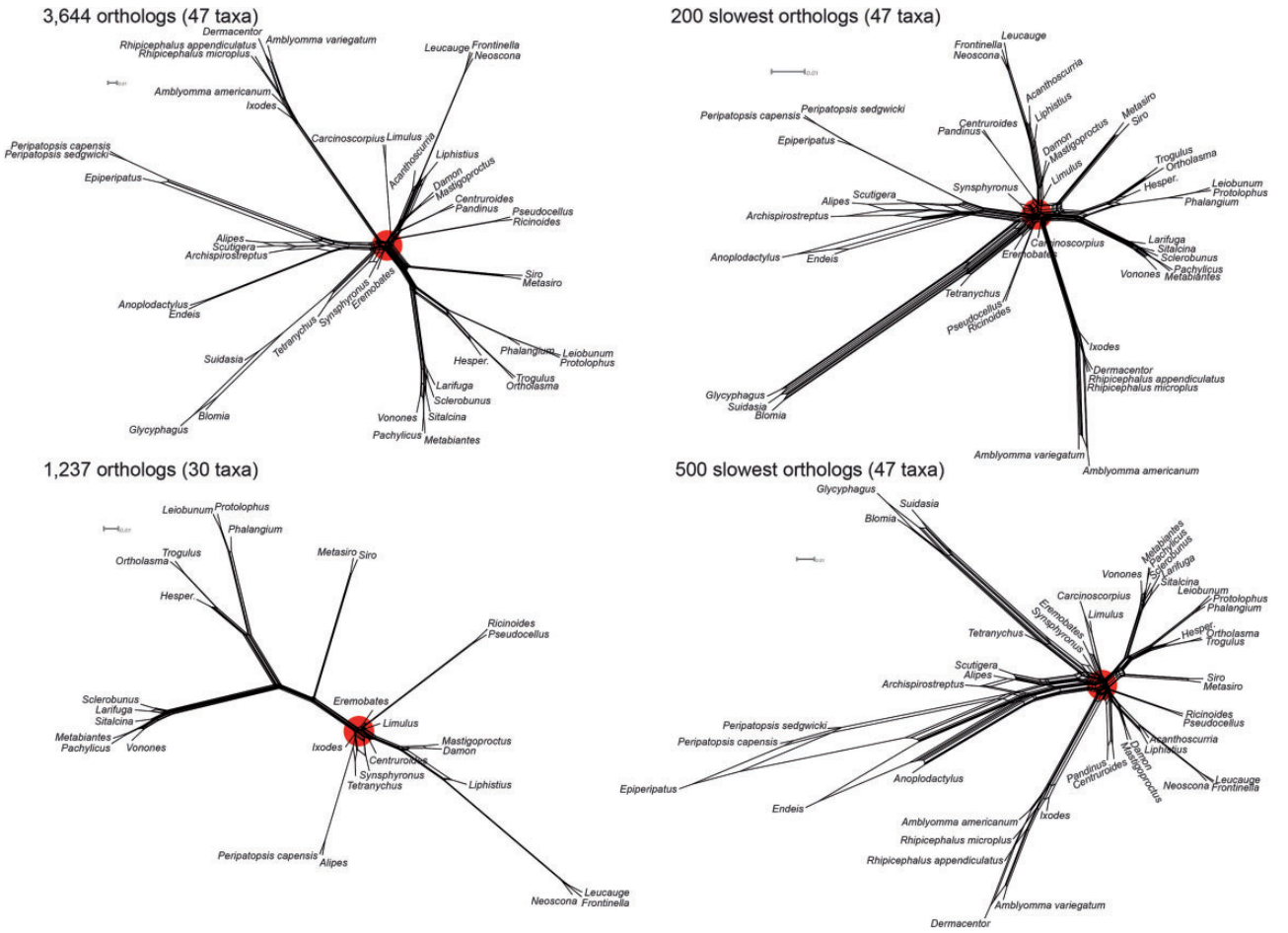

Fig. 10. Supernetwork representation of quartets derived from individual ML gene trees, for four different data sets. Phylogenetic conflict is represented by reticulations. Edge lengths correspond to quartet frequencies. Shading indicates base of arachnid radiation.

literature (very limited support was obtained for Ricinulei + Solifugae by Regier et al. 2010), and both warrant further anatomical and embryological investigation.

The Placement of Palpigradi

The 47 total species in the supermatrix represented all but two orders of Chelicerata: Schizomida and Palpigradi. Schizomida is a lineage consistently recovered as sister group of Thelyphonida, and the pair of orders comprises the clade Uropygi (Shultz 1990, 2007; Giribet et al. 2002; Pepato et al. 2010; Regier et al. 2010). Several characters unite Schizomida and Thelyphonida, including a unique mating behavior (Shultz 1990), and the mutual monophyly of the two orders has never been tested (i.e., one may constitute a nested lineage of the other).

In contrast, the placement of Palpigradi, an order of minuscule (typically < 1 mm in length) chelicerates (see Giribet et al. 2014), has long been in dispute. Various analyses of different data classes have suggested that palpigrades are sister to Acariformes (Van der Hammen 1989; Regier et al. 2010), Acari (Pepato et al. 2010), Tetrapulmonata (Shultz 1990; Wheeler and Hayashi 1998), Solifugae (Giribet et al. 2002), or all remaining arachnids (Shultz 2007). At the heart of the contention is the interpretation of palpigrade opisthosomal "sacs," three pairs of putatively respiratory structures that appear in opisthosomal sternites 4–6. These structures were historically inferred to be either the precursors or the remnants of book lungs, but a respiratory function of these structures has never been conclusively demonstrated. As we were unable to include genome-scale data for Palpigradi, we separately assessed the placement of this order using the 56 Sanger-sequenced genes for *P. wheeleri* from Regier et al. (2010). These were analyzed as part of the matrix with the 500 slowest-evolving orthologs.

This analysis recovered *P. wheeleri* as sister group to Parasitiformes, with little support (BS = 50%), and modest reduction of nodal support elsewhere in the tree; relationships among and within other orders were unchanged (supplementary fig. S7, Supplementary Material online). Moreover, given the placement of Palpigradi within a group of lineages with suspect placement, as well as its infinitesimal representation in the matrix (three orthologs out of 500), we consider the placement of this order insufficiently addressed in this study.

Outgroup Sampling and LBA Artifacts

One of the commonly invoked solutions for redressing topological instability is depth of taxonomic sampling, and

particularly of closely related outgroup taxa. This is a practice that a subset of the authors has historically advocated in arthropod molecular phylogenies (Wheeler et al. 1993; Wheeler and Hayashi 1998; Giribet et al. 2001). However, for the majority of the analyses conducted in this study, sampling of Mandibulata was restricted to three myriapod species, out of a total of ten outgroups (three onychophorans, three myriapods, two pycnogonids, and two xiphosurans). This deliberate selection of mandibulate exemplars was driven by the topologically stable position of Myriapoda in the arthropod tree of life (Regier et al. 2010; Rehm et al. 2014) and their relatively short-branch lengths in comparison to all of Pancrustacea, as inferred by recent phylogenomic analyses (Regier et al. 2010). An argument favoring selection of outgroup taxa that have short patristic distances from their MRCA with the ingroup taxa (i.e., low substitution rates) has been previously articulated by several workers as a solution for overcoming LBA artifacts (Bergsten 2005; Rota-Stabelli and Telford 2008), which constitute a documented affliction of arachnid phylogenies (Pepato et al. 2010; Arabi et al. 2012). Although the restriction of mandibulate exemplars to myriapods means that nodal support for Myriapoda may be overestimated in our analyses (due to the conflation of mandibulate synapomorphies with support for myriapod monophyly), our focus was on arachnid internal phylogeny, and the inclusion of Pycnogonida and Xiphosura as the most closely related outgroup taxa to Arachnida was the greater imperative for this study. Comparable considerations in outgroup sampling among arthropods are exemplified by the phylogenomic analyses of Regier et al. (2010), who sampled just five outgroups for their analysis, but specifically chose the closest relatives of Arthropoda (three Onychophora and two Tardigrada), due to their focus on arthropod internal relationships. Brewer and Bond (2013) used a tick, a water flea (pancrustacean), and a centipede to root their internal millipede phylogeny. In an analysis of Opiliones, Hedin et al. (2012) used one outgroup (in one analysis, two): a tick or a chimeric scorpion terminal, the latter constructed from libraries of three different families of Scorpiones. These last two cases represent a much sparser outgroup sampling than the one conducted here.

Nevertheless, to demonstrate that sampling of outgroups without regard for branch length distributions can in fact exacerbate topological conflict when LBA artifacts already plague the ingroup, we conducted a separate family of analyses wherein we added four pancrustacean taxa, retaining the criterion of comparable data (i.e., Illumina libraries or complete, well-annotated genomes): the dipteran *Drosophila melanogaster* (complete genome), the coleopteran *Tribolium castaneum* (complete genome), the hemipteran *Oncopeltus fasciatus* (Illumina transcriptome; Ewen-Campen et al. 2011), and the amphipod *Parhyale hawaiensis* (Zeng et al. 2011). Due to incident concerns about missing data in this data set (see above), 454 libraries for pancrustaceans were not analyzed. A fifth published pancrustacean data set (*Daphnia pulex*, complete genome) was added in a preliminary analysis, but proved greatly problematic, due to a branch length significantly longer than all other pancrustacean taxa analyzed and ensuing LBA artifacts (see also Fernández et al. 2014). Results from the data set including *Daphnia* (a 52-taxon data set) are therefore not presented here.

As shown in figure 11, ML topologies of identical data partitions augmented with pancrustacean outgroups demonstrate significantly longer patristic distances for Pancrustacea in comparison to Myriapoda (i.e., with respect to the MRCA of arthropods; see also Regier et al. 2010), corroborating the preferential selection of myriapods as outgroups in the main analyses of this study. The branch lengths of Pancrustacea as a whole are comparable to those of Acariformes, and the inclusion of Pancrustacea renders Chelicerata nonmonophyletic in some analyses, due to the phylogenetic proximity of Pancrustacea and Pycnogonida. Analysis of an augmented MARE2 matrix (which lacks pycnogonids) did not yield mandibulate monophyly either, with euchelicerates nested within a paraphyletic Mandibulata (not shown). As demonstrated by visualizations of gene tree conflict (fig. 12), these placements of Pancrustacea are directly attributable to LBA, with a large number of genes recovering a close relationship of *Parhyale* and Acariformes, particularly in slower-evolving gene partitions that are more susceptible to topological incongruence (Salichos and Rokas 2013). Although larger data sets do mitigate this problem (3,644-ortholog and MARE2 matrices; fig. 12), the lack of resolution at the base of Arachnida is entirely unaffected by the addition of pancrustacean outgroups, regardless of outgroup data set size (fig. 12, compare with fig. 10), and Arachnida was not recovered as monophyletic by the augmented data sets.

These analyses reinforce the importance of outgroup choice when addressing lineages with demonstrable long-branch ingroup taxa and justify the limitation of mandibulate sampling to myriapods in our central analyses. The addition of outgroup data sets with nearly complete gene sampling (e.g., *Drosophila*, *Tribolium*) cannot surmount LBA artifacts when a large swath of basally branching pancrustacean taxa is not represented. Admittedly, the basal placement of Pancrustacea may be greatly improved by extensively sampling basally branching lineages throughout Mandibulata, but this is demonstrably the case even using a minuscule fraction of the genes deployed herein (e.g., eight-gene analysis of Giribet et al. 2001; the 62-gene analysis of Regier et al. 2010). For the present analysis, pancrustacean genomic resources comparable to those of model organisms like the four species we analyzed (in addition to the chelicerate data we generated) remain limited, and particularly for such key lineages as copepods, ostracods, remipedes, cephalocarids, and collembolans.

Terrestrialization and Morphological Convergence

Although a number of aforementioned problems plague inference of chelicerate relationships, perhaps the most haunting concern is that the pursuit and identification of phylogenetic signal supporting the monophyly of Arachnida is itself an idiosyncratic objective. A feature common to early hypotheses of arthropod phylogeny, largely driven by morphological data, was the monophyly of Atelocerata, a clade

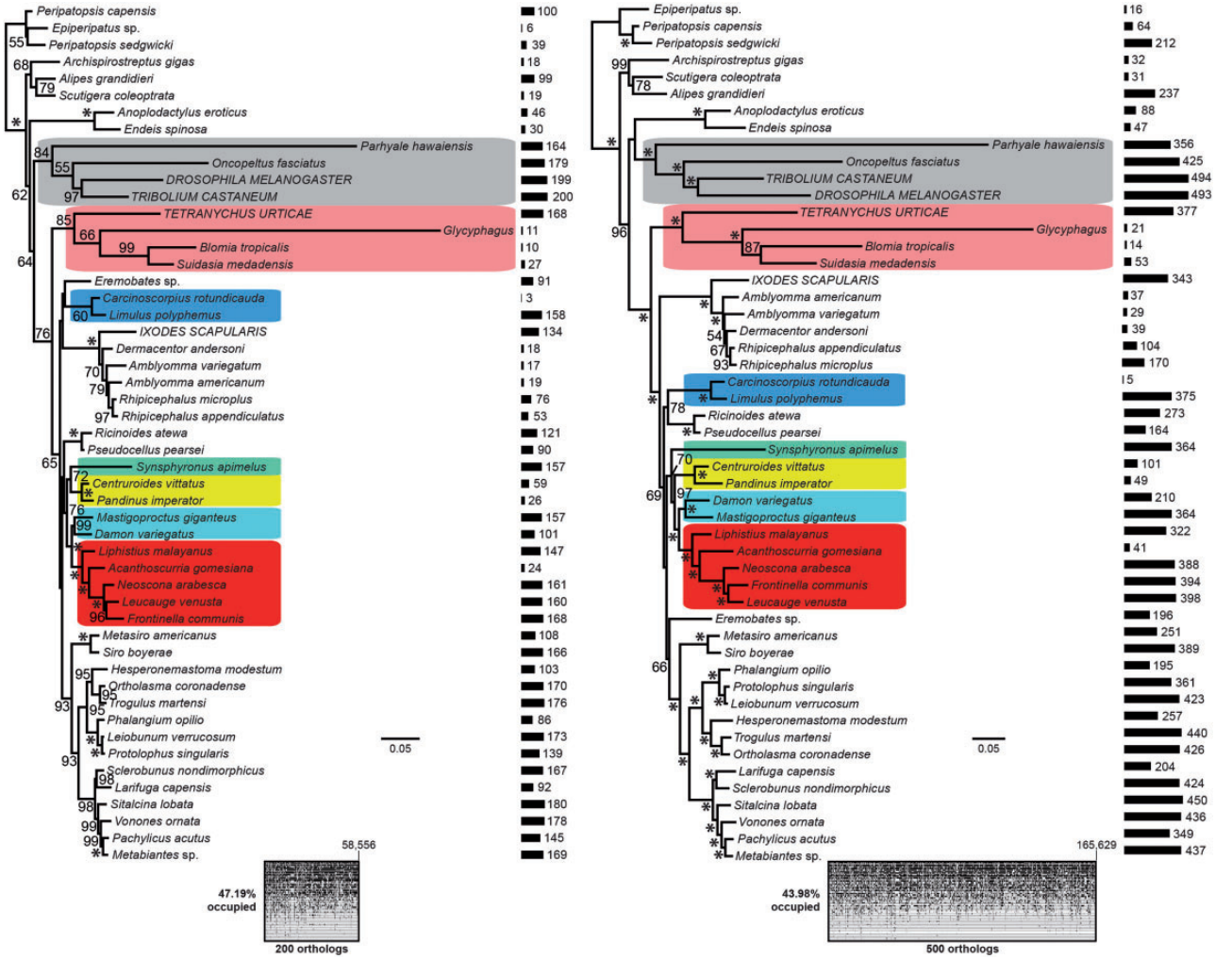

Fig. 11. Left: Tree topology inferred from ML analysis of the 200 slowest-evolving genes, with four pancrustacean outgroups added. Right: Tree topology inferred from ML analysis of the 500 slowest-evolving genes, with four pancrustacean outgroups added. Orders of interest are indicated with shading. Numbers on nodes indicate bootstrap resampling frequencies. Bars on right represent number of genes available for each terminal. Below: Global views of the sequence alignments.

comprising myriapods and hexapods (e.g., Snodgrass 1938; Schram and Emerson 1991; Wheeler et al. 1993; Fortey et al. 1997; Emerson and Schram 1998; Kraus 1998; Wheeler 1998). Numerous aspects of anatomy united the two terrestrial mandibulate clades, such as the absence of tritocerebral appendages, the uniramy of present appendages, and characters pertaining to tracheal respiration. This hypothesis was progressively overturned in favor of Pancrustacea (a clade uniting hexapods with a paraphyletic assemblage collectively termed "crustaceans") by accruing molecular data (e.g., Turbeville et al. 1991; Friedrich and Tautz 1995; Giribet et al. 1996, 2001; Giribet and Ribera 1998; Zrzavý et al. 1998; Koenemann et al. 2010; Meusemann et al. 2010; Campbell et al. 2011; Rota-Stabelli et al. 2011). The diphyly of Atelocerata rendered a large number of its putative synapomorphies homoplastic, the mechanism of which is inferred to be convergence due to terrestrial habitat (Legg et al. 2013).

If terrestrialization has driven morphological convergence in the sister group of Chelicerata, then any epistemological justification for minimizing the number of inferred terrestrialization events in the chelicerate branch of the tree is greatly diminished. Integral to the reconstruction of terrestrialization is the phylogenetic placement of scorpions, which have an impressive Paleozoic fossil record demonstrating marine (or at least aquatic) origins (Jeram 1998; Dunlop et al. 2007; Dunlop 2010). Some workers therefore considered scorpions sister to the remaining arachnids (Weygoldt and Paulus 1979) or separate from arachnids and sister to Xiphosura (Van der Hammen 1989; ref. Dunlop and Webster 1999 for a review of these hypotheses). In contrast, others have inferred a single terrestrialization event in the ancestor of Arachnida, based on comparative morphology and putative homology of scorpion and tetrapulmonate respiratory systems (Firstman 1973; Scholtz and Kamenz 2006; Kamenz et al. 2008) and circulatory systems (Wirkner et al. 2013); implicit in this inference is the assumption that the MRCA of scorpions and tetrapulmonates was the ancestral arachnid (i.e., the topology of Shultz 1990, 2007). Given the strongly supported placement of Scorpiones sister to Tetrapulmonata—a result recovered elsewhere with

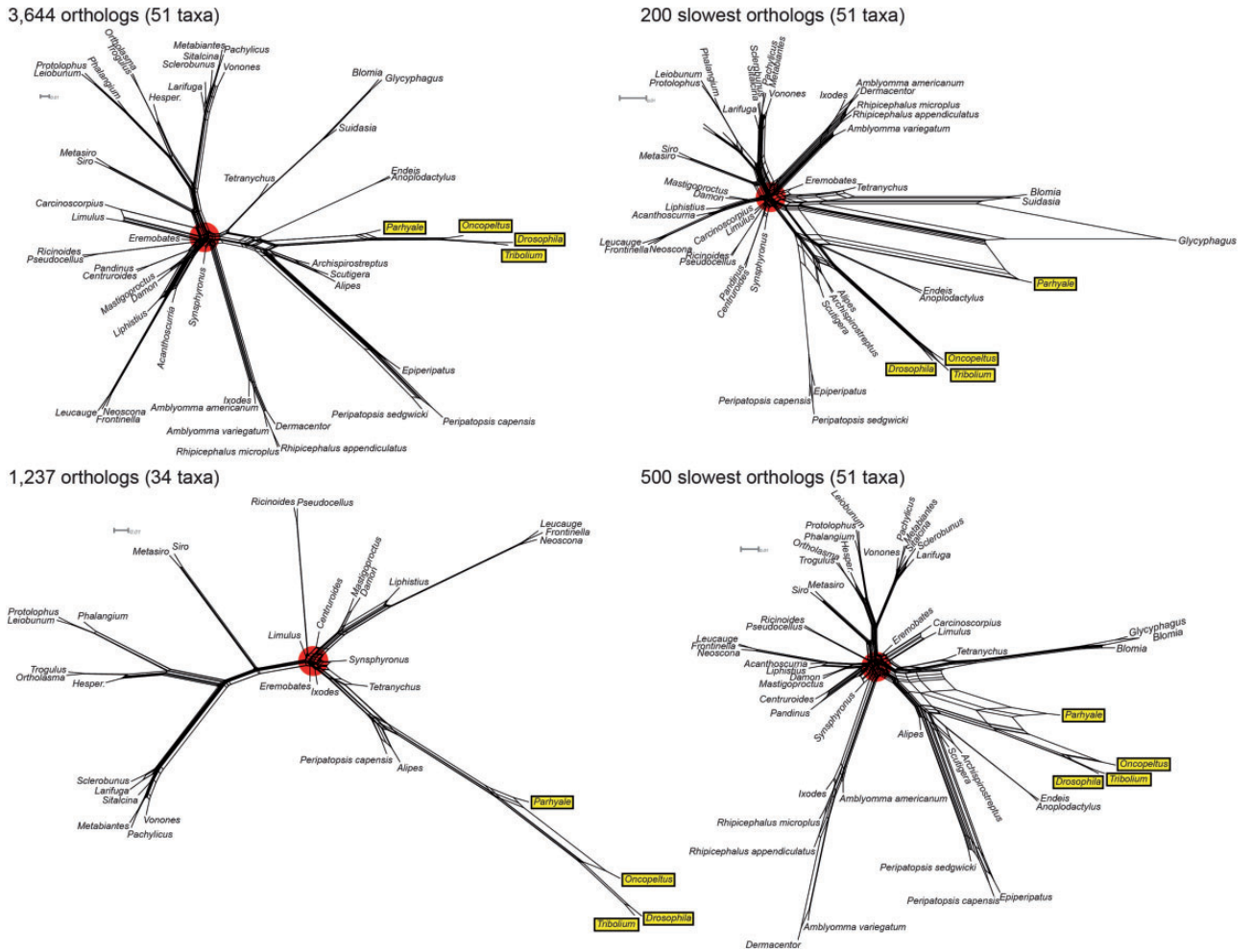

Fig. 12. Supernetwork representation of quartets derived from individual ML gene trees, for four different data sets, with four pancrustacean outgroups added (indicated by boxes). Phylogenetic conflict is represented by reticulations. Edge lengths correspond to quartet frequencies. Shading indicates base of arachnid radiation.

moderate support (Regier et al. 2010)—or in some analyses, the sister relationship of (Scorpiones + Pseudoscorpiones) to Tetrapulmonata (fig. 6), the tree topology at hand obviates inference of a terrestrial arachnid ancestor. The phylogenetic proximity of Scorpiones to Tetrapulmonata is certainly consistent with the homology of their respiratory and circulatory systems (i.e., implies a single origin of the arachnid book lung), but it implies that a single terrestrialization event is not a viable reconstruction of chelicerate evolutionary history, even if Arachnida is recovered as monophyletic. Chelicerata therefore could encompass an evolutionary history with multiple terrestrialization events, in a manner comparable to Mandibulata (von Reumont et al. 2012).

Due to the incidence of heterotachy for some chelicerate orders and concealed phylogenetic signal for several plausible phylogenetic hypotheses, we tentatively favor elements of the tree topology recovered by the slower-evolving orthologs that are consistent with morphological analyses, such as arachnid monophyly, the placement of Pseudoscorpiones within or sister to Arachnopulmonata, and the relationship of (Opiliones + (Ricinulei + Solifugae)). We add the caveat that these relationships should be rigorously tested using broader sampling of constituent terminals (particularly of basally branching exemplars of the problematic orders). We also note that derived relationships, particularly within the well-sampled orders Opiliones and Araneae, unambiguously support traditional systematics within these groups (Shultz 1998; Coddington et al. 2004; Giribet et al. 2010; Sharma and Giribet 2011; Hedin et al. 2012).

Conclusion

We assessed the phylogeny of Arachnida using the largest phylogenomic assessment to date and conducted 40 separate analyses, varying different parameters of matrix assembly and algorithmic approach. We identified Pseudoscorpiones, Acariformes, and Parasitiformes as problematic long-branch taxa, even upon inclusion of full genomes and use of sophisticated substitution models. We consistently observed short internodes in basal relationships attributable to ancient rapid radiation. Serial concatenation in order of percent pairwise identity demonstrates systemic conflicts in phylogenetic signal. Topological conflict at the base of Arachnida is retained among gene trees across data sets, regardless of evolutionary rate or minimization of missing data. Slow-evolving

orthologs recovered the monophyly of Arachnida, the clade (Pseudoscorpiones + Arachnopulmonata), and the relationship of (Opiliones + (Ricinulei + Solifugae)). Addition of outgroups without regard for evolutionary rate is shown to exacerbate LBA artifacts. A number of relationships was unambiguously supported in all analyses. Future investigations of chelicerate phylogeny should emphasize taxonomic sampling efforts for breaking long branches by identifying basally branching and putatively slowly evolving exemplars of the orders Pseudoscorpiones, Acariformes, and Parasitiformes. Improving matrix representation of Palpigradi is mandatory for assessing its placement within Chelicerata.

Materials and Methods

Species Sampling and Molecular Techniques

Taxa were sampled toward maximizing phylogenetic representation within Chelicerata. Given the availability of whole genomes for both Parasitiformes and Acariformes, we focused on generating Illumina paired-end read data for other chelicerate orders. New transcriptomic data were thus generated for 17 euchelicerate taxa. Collecting locality information and statistics on sequencing yields are provided in table 1 and supplementary table S1, Supplementary Material online. All tissues were collected fresh and prepared immediately for RNA extraction. Total RNA was extracted using TRI Reagent (Invitrogen). mRNA samples were prepared with the Dynabeads mRNA Purification Kit (Invitrogen), following the manufacturer's protocol. Quality of mRNA was measured with picoRNA assays in an Agilent 2100 Bioanalyzer (Agilent Technologies); quantity was measured with a RNA assay in Qubit fluorometer (Life Technologies).

Almost all cDNA libraries were constructed with the TruSeq for RNA Sample Preparation kit (Illumina) following the manufacturer's instructions; three spider libraries (*Frontinella communis*, *Leucauge venusta*, and *Neoscona arabesca*) were prepared using the PrepX mRNA kit for Apollo 324 (Integen). Samples were barcoded with TruSeq adaptors, facilitating multiplexed sequencing runs, with up to four specimens per lane sequenced on the Illumina Hi-Seq 2500 platform. Concentration of cDNA was measured using dsDNA High Sensitivity assays in an Agilent 2100 Bioanalyzer (Agilent Technologies), as well as a Qubit fluorometer (Invitrogen). Samples were run using a Illumina Genetic Analyzer II or Illumina HiSeq 2500 platform with paired-end reads of 100–150 bp at the FAS Center for Systems Biology at Harvard University, excepting the *Mastigoproctus giganteus* library, which was sequenced using single-end reads of 100 bp.

Sequenced results were quality filtered accordingly to a threshold average quality Phred score of 30 and adaptors trimmed. Retained reads shorter than 25 bp were discarded. Ribosomal RNA (rRNA) was filtered out using Bowtie v.1.0.0 (Langmead et al. 2009), by constructing a set of metazoan rRNA sequences from GenBank with the command, "bowtie-build." Each sample was sequentially aligned to the index allowing up to two mismatches through Bowtie v.1.0.0, retaining all unaligned reads in FASTQ format (command line: "-a –best"). Unaligned results were paired using read header information and exported as separate files for left and right mate pairs. This process was repeated for each sample and reads corresponding to rRNA sequences were removed.

De novo assemblies were conducted using Trinity (Grabherr et al. 2011; Haas et al. 2013) using paired read files, with the exception of hardware specifications (numbers of dedicated CPUs). For the single-end read library of *M. giganteus*, a path reinforcement distance of 25 was enforced; for paired-end libraries, a value of 75 was enforced. Raw reads have been deposited in the NCBI (National Center for Biotechnology Information) Sequence Read Archive database. All assemblies, alignments, matrices, and tree topologies generated in this study are deposited in the Dryad Digital Repository.

We added 30 other taxa to the data set of 18 Illumina-sequenced chelicerates (table 1). Peptide sequences from whole genomes of *I. scapularis* and *T. urticae* were accessed from the public databases Vectorbase and Ensembl, respectively. A pair of outgroup species, *Alipes grandidieri* (Myriapoda) and *Peripatopsis capensis* (Onychophora), were sequenced by a subset of the authors using the same methods and platforms for a separate study (Fernández et al. 2014). A subset of smaller, assembled 454 libraries and Sanger-sequenced ESTs were accessed from GenBank (table 1). Finally, we obtained reads for eight harvestman libraries published by Hedin et al. (2012), and sanitized and assembled these de novo in Trinity as described above.

Orthology Assignment

Redundancy reduction was done with CD-HIT (Fu et al. 2012) for all Trinity assemblies (95% similarity). Resulting culled assemblies were processed in TransDecoder (Haas et al. 2013) in order to identify candidate ORFs within the transcripts. Predicted peptides were then processed with a further filter to select only one peptide per putative unigene, by choosing the longest ORF per Trinity subcomponent with a python script, thus removing the variation in the coding regions of Trinity assemblies due to alternative splicing, closely related paralogs, and allelic diversity. Peptide sequences with all final candidate ORFs were retained as fasta files. We assigned predicted ORFs into orthologous groups across all samples using OMA stand-alone v.0.99t (Altenhoff et al. 2011, 2013). All input files were single line multi-fasta files and the parameters.drw file specified retained all default settings. We parallelized all-by-all local alignments across 200 CPUs. In a separate family of analyses, additional outgroups were added to the output from the principal OMA run using best reciprocal BLAST hits, implemented in a customized script.

Phylogenomic Analyses

We constructed an amino acid supermatrix by concatenating the set of OMA groups containing 13 or more taxa (gene occupancy of 25%), which were extracted using a Unix script. These 3,644 orthogroups were aligned individually using MUSCLE v.3.6 (Edgar 2004). Ambiguously aligned positions were culled using GBlocks v.0.91b (Castresana 2000), as well as regions where more than 50% of the columns constituted

missing data (using the command –b5 = h). Trimmed orthogroups were concatenated using Phyutility 2.6 (Smith and Dunn 2008) for this and other phylogenomic matrices (see below).

Due to the degree of missing data for the palpigrade *P. wheeleri*, represented by 56 Sanger-sequenced genes, we principally analyzed our data set without this species, adding the palpigrade terminal in a final analysis to infer the placement of this enigmatic order. Attempts to obtain new transcriptomes of *Eukoenenia spelaea* (from Slovakia) and *Eukoenenia* sp. (from Mexico) were unsuccessful due to the diminutive size and scarcity of these animals.

To assess the impact of gene occupancy on nodal support and tree topology, we constructed five other matrices at different thresholds of gene occupancy. An additional 15 matrices were constructed based on evolutionary rate, for which percent pairwise identity was employed as a proxy. This method was chosen to approximate evolutionary rate because it is agnostic to tree topology. Percent pairwise identity was calculated for each orthogroup alignment in Geneious v.6.1.6, and is computed by taking all possible pairs of bases at the same column and scoring a hit (one) when a pair is identical; this value is divided by the total number of pairs for each column, and all column values are then averaged over the length of the alignment. Cells with indels ("-") and missing taxa do not contribute to pairwise calculations. The percent pairwise identity was calculated for each of the 3,644 orthogroup alignments subsequent to treatment with GBlocks v. 0.91b. This ensured that regions with large indels did not affect the calculation of percent pairwise identity.

As an algorithmic alternative to optimization of matrix construction, we used the software MARE (MAtrix REduction; http://mare.zfmk.de, last accessed November 11, 2013), which estimates informativeness of every orthogroup based on weighted geometry quartet mapping (Nieselt-Struwe and von Haeseler 2001). We analyzed the 3,644 orthogroups with gene occupancy over 25% using MARE, and thereby constructed two matrices. First, we implemented MARE, enforcing retention of all libraries. Second, we repeated algorithmic matrix reduction, but enforced retention of only those 30 libraries constituting Illumina libraries or whole genomes (i.e., keeping only terminals with significantly more data than 454 or Sanger-sequenced EST counterparts). As a separate test of influence of data quantity, we constructed a manual reduction of the 3,644-ortholog supermatrix, retaining only the 30 libraries constituting Illumina libraries or whole genomes.

ML analyses were conducted using RAxML v.7.7.5 (Berger et al. 2011) and Bayesian inference using PhyloBayes MPI 1.4e (Lartillot et al. 2013). For RAxML v.7.7.5, we implemented a unique LG4X + F model for each gene (Le et al. 2012). Bootstrap resampling frequencies were estimated with 500 replicates using a rapid bootstrapping algorithm (Stamatakis et al. 2008). Analyses with PhyloBayes MPI 1.4e were limited to smaller data sets, as implementation of PhyloBayes for large matrices requires intractable amounts of time for convergence. We implemented PhyloBayes MPI 1.4e using the site-heterogeneous CAT + GTR model of evolution (Lartillot and Philippe 2004). Four independent chains were run for 6,965–10,145 cycles, and the initial 5,000 cycles were discarded as burn-in, with convergence assessed using the maximum bipartition discrepancies across chains.

Supplementary Material

Supplementary figures S1–S7 and table S1 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

Mark S. Harvey, Lionel Monod, Peter J. Schwendinger, and Piotr Naskrecki kindly provided samples of *Synsphyronus apimelus*, *Liphistius malayanus*, and *Ricinoides atewa*. Antje Fischer and Woods Hole Marine Biological Laboratory provided specimens of *Limulus polyphemus*. Sónia C.S. Andrade, Jerome Murienne, and Ana Riesgo prepared some of the Illumina libraries. Thorsten Burmester provided the 454 transcriptomic assembly of *Pandinus imperator*. Marshal Hedin contributed raw read data for the eight Opiliones libraries we reassembled. Adrian Altenhoff, James Cuff, Geoffrey F. Dilly, Paul Edmon, Mike Ethier, and Rosa Fernández facilitated computing operations. Discussions with Casey Dunn and Antonis Rokas refined some of the ideas presented. This research was supported by internal funds from the Museum of Comparative Zoology to G.G., by National Science Foundation grants DEB-1144492 and DEB-114417 to G.H. and G.G., and by the National Science Foundation Postdoctoral Research Fellowship in Biology under Grant No. DBI-1202751 to P.P.S. The authors are indebted to Editor-in-Chief Sudhir Kumar, Associate Editor Nicolas Vidal, and three anonymous reviewers for reviewing this manuscript.

References

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

Altenhoff AM, Schneider A, Gonnet GH, Dessimoz C. 2011. OMA 2011: orthology inference among 1000 complete genomes. *Nucleic Acids Res.* 39:D289–D294.

Arabi J, Judson ML, Deharveng L, Lourenço WR, Cruaud C, Hassanin A. 2012. Nucleotide composition of CO1 sequences in Chelicerata (Arthropoda): detecting new mitogenomic rearrangements. *J Mol Evol.* 74:81–95.

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

Berger SA, Krompass D, Stamatakis A. 2011. Performance, accuracy, and Web server for evolutionary placement of short sequence reads under maximum likelihood. *Syst Biol.* 60:291–302.

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

Betancur-RR, Naylor G, Ortí G. 2014. Conserved genes, sampling error, and phylogenomic inference. *Syst Biol.* 63:257–262.

Brewer MS, Bond JE. 2013. Ordinal-level phylogenomics of the arthropod class Diplopoda (millipedes) based on an analysis of 221 nuclear protein-coding loci generated using next-generation sequence analyses. *PLoS One* 8:e79935.

Briggs DEG, Siveter DJ, Sieveter DJ, Sutton MD, Garwood RJ, Legg D. 2012. Silurian horseshoe crab illuminates the evolution of arthropod limbs. *Proc Natl Acad Sci U S A.* 109:15702–15705.

Campbell LI, Rota-Stabelli O, Edgecombe GD, Marchioro T, Longhorn SJ, Telford MJ, Philippe H, Rebecchi L, Peterson KJ, Pisani D. 2011. MicroRNAs and phylogenomics resolve the relationships of Tardigrada and suggest that velvet worms are the sister group of Arthropoda. *Proc Natl Acad Sci U S A.* 108:15920–15924.

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

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

Clouse RM, Sharma PP, Giribet G, Wheeler WC. 2013. Elongation factor-1α, a putative single-copy nuclear gene, has divergent sets of paralogs in an arachnid. *Mol Phylogenet Evol.* 68:471–481.

Coddington JA, Giribet G, Harvey MS, Prendini L, Walter DE. 2004. Arachnida. In: Cracraft J, Donoghue MJ, editors. Assembling the tree of life. New York: Oxford University Press. p. 296–318.

Dell'Ampio E, Meusemann K, Szucsich NU, Peters RS, Meyer B, Borner J, Petersen M, Aberer AJ, Stamatakis A, Walzl MG, et al. 2014. Decisive data sets in phylogenomics: lessons from studies on the phylogenetic relationships of primarily wingless insects. *Mol Biol Evol.* 31: 239–249.

Dunlop JA. 2010. Geological history and phylogeny of Chelicerata. *Arthropod Struct Dev.* 39:124–142.

Dunlop JA, Kamenz C, Scholtz G. 2007. Reinterpreting the morphology of the Jurassic scorpion *Liassoscorpionides*. *Arthropod Struct Dev.* 36: 245–252.

Dunlop JA, Krüger J, Alberti G. 2012. The sejugal furrow in camel spiders and acariform mites. *Arachnol Mitt.* 43:8–15.

Dunlop JA, Webster M. 1999. Fossil evidence, terrestrialization, and arachnid phylogeny. *J Arachnol.* 27:86–93.

Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M, Edgecombe GD, et al. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 452:745–749.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.

Emerson MJ, Schram FR. 1998. Theories, patterns, and reality: game plan for arthropod phylogeny. In: Fortey RA, Thomas RH, editors. Arthropod relationships. London: Chapman & Hall. p. 67–86.

Ewen-Campen B, Shaner N, Panfilio K, Suzuki Y, Roth S, Extavour CG. 2011. The maternal and early embryonic transcriptome of the milkweed bug *Oncopeltus fasciatus*. *BMC Genomics* 12:61.

Felsenstein J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Syst Biol.* 27:401–410.

Fernández R, Laumer CE, Vahtera V, Libro S, Kaluziak ST, Sharma PP, Pérez-Porro AR, Edgecombe GD, Giribet G. 2014. Evaluating topological conflict in centipede phylogeny using transcriptomic data sets. *Mol Biol Evol.* 31:1500–1513.

Firstman B. 1973. The relationship of the chelicerate arterial system to the evolution of the endosternite. *J Arachnol.* 1:1–54.

Fortey RA, Briggs DEG, Wills MA. 1997. The Cambrian evolutionary 'explosion' recalibrated. *BioEssays* 19:429–434.

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

Fu LM, Niu BF, Zhu ZW, Wu ST, Li WZ. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28: 3150–3152.

Giribet G, Carranza S, Baguñà J, Riutort M, Ribera C. 1996. First molecular evidence for the existence of a Tardigrada + Arthropoda clade. *Mol Biol Evol.* 13:76–84.

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

Giribet G, Edgecombe GD, Wheeler WC, Babbitt C. 2002. Phylogeny and systematic position of Opiliones: a combined analysis of chelicerate relationships using morphological and molecular data. *Cladistics* 18: 5–70.

Giribet G, McIntyre E, Christian E, Espinasa L, Ferreira RL, Francke ÓF, Harvey MS, Isaia M, Kováč Ľ, McCutchen L, et al. 2014. The first phylogenetic analysis of Palpigradi (Arachnida)—the most enigmatic arthropod order. *Invertebr Syst.* 28. doi:10.1071/IS13057.

Giribet G, Ribera C. 1998. The position of arthropods in the animal kingdom: a search of a reliable outgroup for internal arthropod phylogeny. *Mol Phylogenet Evol.* 9:481–488.

Giribet G, Vogt L, Pérez González A, Sharma P, Kury AB. 2010. A multilocus approach to harvestmen (Arachnida: Opiliones) phylogeny with emphasis on biogeography and the systematics of Laniatores. *Cladistics* 26:408–437.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 29:644–652.

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

Grünewald S, Spillner A, Bastkowski S, Bogershausen A, Moulton V. 2013. SuperQ: computing supernetworks from quartets. *IEEE/ACM Trans Comput Biol Bioinform.* 10:151–160.

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

Hedin M, Derkarabetian S, McCormack M, Richart C, Shultz JW. 2010. The phylogenetic utility of the nuclear protein-coding gene EF-1α for resolving recent divergences in Opiliones, emphasizing intron evolution. *J Arachnol.* 38:9–20.

Hedin M, Starett J, Akhter S, Schönhofer AL, Shultz JW. 2012. Phylogenomic resolution of Paleozoic divergences in harvestmen (Arachnida, Opiliones) via analysis of next-generation transcriptome data. *PLoS One* 7:e42888.

Hejnol A, Obst M, Stamatakis A, Ott M, Rouse GW, Edgecombe GD, Martinez P, Baguñà J, Bailly X, Jondelius U, et al. 2009. Assessing the root of bilaterian animals with scalable phylogenomic methods. *Proc R Soc Lond B Biol Sci.* 276:4261–4270.

Huelsenbeck JP. 1997. Is the Felsenstein zone a fly trap? *Syst Biol.* 46: 69–74.

Hughes CL, Kaufman TC. 2002. Hox genes and the evolution of the arthropod body plan. *Evol Dev.* 4:459–499.

Jeffroy O, Brinkmann H, Delsuc F, Philippe H. 2006. Phylogenomics: the beginning of incongruence? *Trends Genet.* 22:225–231.

Jeram AJ. 1998. Phylogeny, classifications and evolution of Silurian and Devonian scorpions. In: Selden PA, editor. In: Proceedings of the 17th European Colloquium of Arachnology. Edinburgh: British Arachnological Society, Burnham Beeches. p. 17–31.

Kamenz C, Dunlop JA, Scholtz G, Kerp H, Hass H. 2008. Microanatomy of Early Devonian book lungs. *Biol Lett.* 4:212–215.

Khila A, Grbic M. 2007. Gene silencing in the spider mite *Tetranychus urticae*: dsRNA and siRNA parental silencing of the *Distal-less* gene. *Dev Genes Evol.* 217:241–251.

Koenemann S, Jenner RA, Hoenemann M, Stemme T, von Reumont BM. 2010. Arthropod phylogeny revisited, with a focus on crustacean relationships. *Arthropod Struct Dev.* 39:88–110.

Kraus O. 1998. Phylogenetic relationships between higher taxa of tracheate arthropods. In: Fortey RA, Thomas RH, editors. Arthropod relationships. London: Chapman & Hall. p. 295–303.

Kubatko LS, Degnan JH. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Syst Biol.* 56:17–24.

Lamsdell JC. 2013. Revised systematics of Palaeozoic 'horseshoe crabs' and the myth of monophyletic Xiphosura. *Zool J Linn Soc.* 167:1–27.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25.

Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol.* 21:1095–1109.

Lartillot N, Philippe H. 2008. Improvement of molecular phylogenetic inference and the phylogeny of Bilateria. *Philos Trans R Soc Lond B Biol Sci.* 363:1463–1472.
Lartillot N, Rodrigue N, Stubbs D, Richer J. 2013. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst Biol.* 62:611–615.
Le SQ, Dang CC, Gascuel O. 2012. Modeling protein evolution with several amino acid replacement matrices depending on site rates. *Mol Biol Evol.* 29:2921–2936.
Legg DA, Sutton MD, Edgecombe GD. 2013. Arthropod fossil data increase congruence of morphological and molecular phylogenies. *Nat Commun.* 4:2485.
Mallatt JM, Garey JR, Shultz JW. 2004. 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.
Masta SE, Longhorn SJ, Boore JL. 2009. Arachnid relationships based on mitochondrial genomes: asymmetric nucleotide and amino acid bias affects phylogenetic analyses. *Mol Phylogenet Evol.* 50:117–128.
Masta SE, McCall A, Longhorn SJ. 2010. Rare genomic changes and mitochondrial sequences provide independent support for congruent relationships among the sea spiders (Arthropoda, Pycnogonida). *Mol Phylogenet Evol.* 57:59–70.
Meusemann K, von Reumont BM, Simon S, Roeding F, Strauss S, Kück P, Ebersberger I, Walzl M, Pass G, Breuers S, et al. 2010. A phylogenomic approach to resolve the arthropod tree of life. *Mol Biol Evol.* 27: 2451–2464.
Narechania A, Baker RH, Sit R, Kolokotronis S-O, DeSalle R, Planet PJ. 2012. Random addition concatenation analysis: a novel approach to the exploration of phylogenomic signal reveals strong agreement between core and shell genomic partitions in the cyanobacteria. *Genome Biol Evol.* 4:30–43.
Nieselt-Struwe K, von Haeseler A. 2001. Quartet-mapping, a generalization of the likelihood-mapping procedure. *Mol Biol Evol.* 18: 1204–1219.
Nosenko T, Schreiber F, Adamska M, Adamski M, Eitel M, Hammel J, Maldonado M, Müller WE, Nickel M, Schierwater B, et al. 2013. Deep metazoan phylogeny: when different genes tell different stories. *Mol Phylogenet Evol.* 67:223–233.
Obst M, Faurby S, Bussarawit S, Funch P. 2012. Molecular phylogeny of extant horseshoe crabs (Xiphosura, Limulidae) indicates Paleogene diversification of Asian species. *Mol Phylogenet Evol.* 62:21–26.
Pepato AR, da Rocha CEF, Dunlop JA. 2010. Phylogenetic position of the acariform mites: sensitivity to homology assessment under total evidence. *BMC Evol Biol.* 10:235.
Philippe H, Brinkmann H, Lavrov DV, Littlewood DTJ, Manuel M, Wörheide G, Baurain D. 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biol.* 9:e1000602.
Philippe H, Derelle R, Lopez P, Pick K, Borchiellini C, Boury-Esnault N, Vacelet J, Renard E, Houliston E, Quéinnec E, et al. 2009. Phylogenomics revives traditional views on deep animal relationships. *Curr Biol.* 19:706–712.
Philippe H, Zhao Y, Brinkmann H, Rodrigue N, Delsuc F. 2005. Heterotachy and long-branch attraction in phylogenetics. *BMC Evol Biol.* 5:50.
Prpic N-M, Damen WGM. 2004. Expression patterns of leg genes in the mouthparts of the spider *Cupiennius salei* (Chelicerata: Arachnida). *Dev Genes Evol.* 214:296–302.
Regier JC, Shultz JW, Zwick A, Hussey A, Ball B, Wetzer R, Martin JW, Cunningham CW. 2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature* 463: 1079–1083.
Rehm P, Meusemann K, Borner J, Misof B, Burmester T. 2014. Phylogenetic position of Myriapoda revealed by 454 transcriptomic sequencing. *Mol Phylogenet Evol.* 77:25–33.
Riesgo A, Andrade SCS, Sharma PP, Novo M, Pérez-Porro AR, Vahtera V, González VL, Kawauchi GY, Giribet G. 2012. Comparative description of ten transcriptomes of newly sequenced invertebrates and efficiency estimation of genomic sampling in non-model taxa. *Front Zool.* 9:33.
Roeding F, Borner J, Kube M, Klages S, Reinhardt R, Burmester T. 2009. A 454 sequencing approach for large scale phylogenomic analysis of the common emperor scorpion (*Pandinus imperator*). *Mol Phylogenet Evol.* 53:826–834.
Rokas A, Carroll SB. 2006. Bushes in the tree of life. *PLoS Biol.* 4:e352.
Rokas A, King N, Finnerty J, Carroll SB. 2003. Conflicting phylogenetic signals at the base of the metazoan tree. *Evol Dev.* 5:346–359.
Rota-Stabelli O, Campbell L, Brinkmann H, Edgecombe GD, Longhorn SJ, Peterson KJ, Pisani D, Philippe H, Telford MJ. 2011. A congruent solution to arthropod phylogeny: phylogenomics, microRNAs and morphology support monophyletic Mandibulata. *Proc R Soc Lond B Biol Sci.* 278:298–306.
Rota-Stabelli O, Daley AC, Pisani D. 2013. Molecular timetrees reveal a Cambrian colonization of land and a new scenario for ecdysozoan evolution. *Curr Biol.* 23:392–398.
Rota-Stabelli O, Telford MJ. 2008. A multi criterion approach for the selection of optimal outgroups in phylogeny: recovering some support for Mandibulata over Myriochelata using mitogenomics. *Mol Phylogenet Evol.* 48:103–111.
Salichos L, Rokas A. 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature* 497:327–331.
Schierwater B, Eitel M, Jakob W, Osigus HJ, Hadrys H, Dellaporta SL, Kolokotronis SO, Desalle R. 2009. Concatenated analysis sheds light on early metazoan evolution and fuels a modern "urmetazoon" hypothesis. *PLoS Biol.* 7:e1000020.
Scholtz G, Kamenz C. 2006. The book lungs of Scorpiones and Tetrapulmonata (Chelicerata, Arachnida): evidence for homology and a single terrestrialization event of a common arachnid ancestor. *Zoology* 109:2–13.
Schoppmeier M, Damen WGM. 2001. Double-stranded RNA interference in the spider *Cupiennius salei*: the role of *Distal-less* is evolutionarily conserved in arthropod appendage formation. *Dev Genes Evol.* 211:76–82.
Schram FR, Emerson MJ. 1991. Arthropod pattern theory: a new approach to arthropod phylogeny. *Mem Queensl Mus.* 31:1–18.
Sharma PP, Giribet G. 2011. The evolutionary and biogeographic history of the armoured harvestmen—Laniatores phylogeny based on ten molecular markers, with the description of two new families of Opiliones (Arachnida). *Invertebr Syst.* 25:106–142.
Sharma PP, Schwager EE, Extavour CG, Giribet G. 2012a. Evolution of the chelicera: a dachshund domain is retained in the deutocerebral appendage of Opiliones (Arthropoda, Chelicerata). *Evol Dev.* 14: 522–533.
Sharma PP, Schwager EE, Extavour CG, Giribet G. 2012b. Hox gene expression in the harvestman *Phalangium opilio* reveals divergent patterning of the chelicerate opisthosoma. *Evol Dev.* 14: 450–463.
Sharma PP, Schwager EE, Extavour CG, Wheeler WC. 2014. Hox gene duplications correlate with posterior heteronomy in scorpions. *Proc R Soc Lond B Biol Sci.* 281:20140661.
Sharma PP, Schwager EE, Giribet G, Jockusch EL, Extavour CG. 2013. *Distal-less* and *dachshund* pattern both plesiomorphic and apomorphic structures in chelicerates: RNA interference in the harvestman *Phalangium opilio* (Opiliones). *Evol Dev.* 15: 228–242.
Shultz JW. 1990. Evolutionary morphology and phylogeny of Arachnida. *Cladistics* 6:1–38.
Shultz JW. 1998. Phylogeny of Opiliones (Arachnida): an assessment of the "Cyphopalpatores" concept. *J Arachnol.* 26:257–272.
Shultz JW. 2007. A phylogenetic analysis of the arachnid orders based on morphological characters. *Zool J Linn Soc.* 150:221–265.
Simon S, Narechania A, DeSalle R, Hadrys H. 2012. Insect phylogenomics: exploring the source of incongruence using new transcriptomic data. *Genome Biol Evol.* 4:1295–1309.
Smith SA, Dunn CW. 2008. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics* 24: 715–716.

Snodgrass RE. 1938. Evolution of the Annelida, Onychophora and Arthropoda. *Smithson. Misc. Collect.* 97:1–159.

Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol.* 57:758–771.

Turbeville JM, Pfeifer DM, Field KG, Raff RA. 1991. The phylogenetic status of arthropods, as inferred from 18S rRNA sequences. *Mol Biol Evol.* 8:669–686.

Van der Hammen L. 1989. An introduction to comparative arachnology. Leiden (United Kingdom): SPB Academic Publishing.

von Reumont BM, Jenner RA, Wills MA, Dell'ampio E, Pass G, Ebersberger I, Meyer B, Koenemann S, Iliffe TM, Stamatakis A, et al. 2012. Pancrustacean phylogeny in the light of new phylogenomic data: support for Remipedia as the possible sister group of Hexapoda. *Mol Biol Evol.* 29:1031–1045.

Weygoldt P, Paulus HF. 1979. Untersuchungen zur Morphologie, Taxonomie und Phylogenie der Chelicerata. *Z Zool Syst Evol.* 17: 85–116, 177–200.

Wheeler WC. 1998. Sampling, groundplans, total evidence and the systematics of arthropods. In: Fortey RA, Thomas RH, editors. Arthropod relationships. London: Chapman & Hall. p. 87–96.

Wheeler WC, Cartwright P, Hayashi CY. 1993. Arthropod phylogeny: a combined approach. *Cladistics* 14:173–192.

Wheeler WC, Hayashi CY. 1998. The phylogeny of the extant chelicerate orders. *Cladistics* 14:173–192.

Wiens JJ. 2006. Missing data and the design of phylogenetic analyses. *J Biomed Inform.* 39:34–42.

Wilcox TP, García de León FJ, Hendrickson DA, Hillis DM. 2004. Convergence among cave catfishes: long-branch attraction and a Bayesian relative rates test. *Mol Phylogenet Evol.* 31: 1101–1103.

Wirkner CS, Tögel M, Pass G. 2013. The arthropod circulatory system. In: Minelli A, Boxshall G, Fusco G, editors. Arthropod biology and evolution: molecules, development, and morphology. Heidelberg (Germany): Springer Press. p. 343–391.

Zeng V, Villanueva KE, Ewen-Campen B, Alwes F, Browne WE, Extavour CG. 2011. *De novo* assembly and characterization of a maternal and developmental transcriptome for the emerging model crustacean *Parhyale hawaiensis*. *BMC Genomics* 12:581.

Zrzavý J, Hypša V, Vlášková M. 1998. Arthropod phylogeny: taxonomic congruence, total evidence and conditional combination approaches to morphological and molecular data sets. In: Fortey RA, Thomas RH, editors. Arthropod relationships. London: Chapman & Hall. p. 97–107.