Syst. Biol. 68(6):896–917, 2019

DOI:10.1093/sysbio/syz011
Advance Access publication March 27, 2019

A Critical Appraisal of the Placement of Xiphosura (Chelicerata) with Account of Known Sources of Phylogenetic Error

JESÚS A. BALLESTEROS* AND PRASHANT P. SHARMA

Department of Integrative Biology, University of Wisconsin-Madison, Madison, WI 53706, USA
**Correspondence to be sent to: Department of Integrative Biology, University of Wisconsin-Madison, 430 Lincoln Drive, Madison, WI 53706, USA; E-mail: ballesterosc@wisc.edu.*

Received 8 June 2018; reviews returned 20 December 2018; accepted 10 February 2019
Associate Editor: Ken Halanych

Abstract.—Horseshoe crabs (Xiphosura) are traditionally regarded as sister group to the clade of terrestrial chelicerates (Arachnida). This hypothesis has been challenged by recent phylogenomic analyses, but the non-monophyly of Arachnida has consistently been disregarded as artifactual. We re-evaluated the placement of Xiphosura among chelicerates using the most complete phylogenetic data set to date, expanding outgroup sampling, and including data from whole genome sequencing projects. In spite of uncertainty in the placement of some arachnid clades, all analyses show Xiphosura consistently nested within Arachnida as the sister group to Ricinulei (hooded tick spiders). It is apparent that the radiation of arachnids is an old one and occurred over a brief period of time, resulting in several consecutive short internodes, and thus is a potential case for the confounding effects of incomplete lineage sorting (ILS). We simulated coalescent gene trees to explore the effects of increasing levels of ILS on the placement of horseshoe crabs. In addition, common sources of systematic error were evaluated, as well as the effects of fast-evolving partitions and the dynamics of problematic long branch orders. Our results indicated that the placement of horseshoe crabs cannot be explained by missing data, compositional biases, saturation, or ILS. Interrogation of the phylogenetic signal showed that the majority of loci favor the derived placement of Xiphosura over a monophyletic Arachnida. Our analyses support the inference that horseshoe crabs represent a group of aquatic arachnids, comparable to aquatic mites, breaking a long-standing paradigm in chelicerate evolution and altering previous interpretations of the ancestral transition to the terrestrial habitat. Future studies testing chelicerate relationships should approach the task with a sampling strategy where the monophyly of Arachnida is not held as the premise. [Arthropods; Arachnids; composition bias; ILS; long branch attraction; phylogenetics; signal.]

"So close is this resemblance in structure to the Arachnida that many zoölogists, and among them some of those who have studied ***Limulus*** *most carefully, regard the Xiphosura as an order of Arachnida"* (Comstock 1912)

Systematic error in phylogenetics, resulting in inaccurate or spurious relationships, occurs when the assumptions of the methods of inference are not met by the data analyzed. These biases are not always overcome by sheer amount of data and affect many recalcitrant nodes of the tree of life (Rosenberg and Kumar 2003; Phillips et al. 2004; Jeffroy et al. 2006; Philippe et al. 2011; Salichos and Rokas 2013). Frequently, testing for systematic biases is triggered by the encounter of groupings that are deemed counterintuitive to the investigator, based on some external criteria for the preferred alternative hypothesis. Although counterintuitive phylogenetic results are often attributed to systematic biases, it is not always straightforward to assess how much a given result is due to systematic error, real signal, or simply the lack of phylogenetic signal (noise). Recent studies suggest that some controversial results can be caused by a few genes favoring a particular grouping regardless of phylogenetic accuracy, even though the majority of the data contributes little to support or reject the relationship of interest (Shen et al. 2017).

These concerns weigh heavily upon the phylogenetic relationships of chelicerate arthropods. Horseshoe crabs (Xiphosura) are one of the oldest lineages of extant arthropods. These marine chelicerates have persisted in the fossil record since the early Ordovician (Rudkin et al. 2008; Van Roy et al. 2010) and survived major mass extinction events (McGhee et al. 2012). The extant diversity of horseshoe crabs is limited to only four species classified in three genera: *Limulus*, *Carcinoscorpius*, and *Tachypleus*. Extinct xiphosurans were generally smaller in size than the living species, some of which reach 0.5 m in length. Nevertheless, the overall resemblance between extinct and living forms has led to these organisms being deemed "living fossils" (Briggs et al. 2012; Kin and Błażejowski 2014). While horseshoe crabs may appear to represent a relictual lineage (comparable to the tuatara or leiopelmatid frogs), the fossil record suggests that the group held only a modest number of species at any given time through most of their evolutionary history (Lamsdell 2013). All living representatives of Xiphosura are restricted to the marine habitat but recent evidence suggests past independent colonization events of freshwater environments (Lamsdell 2016).

As their common name suggests, horseshoe crabs were once thought to be closely related to crustaceans. The evolutionary affinity of Xiphosura with other Chelicerata (e.g., spiders, scorpions, and sea spiders) was established in the late 19th century (Lankester 1881) and since then, the dominant hypothesis has been that horseshoe crabs represent the sister lineage to the terrestrial chelicerates, the highly diverse Arachnida (Snodgrass 1938; Weygoldt and Paulus 1979; Shultz 1990, 2007). In this scenario, extinct marine chelicerate groups like Eurypterida (sea scorpions) and Chasmataspidida are inferred to constitute a grade subtending Arachnida

(Dunlop and Webster 1999). Implicit in this hypothesis of a monophyletic Arachnida is the notion of a single transition to the terrestrial environment by the common ancestor of arachnids. This hypothesis is supported in part by the morphological correspondence between the respiratory organs of horseshoe crabs (the book gills) and the counterparts of some arachnid groups such as spiders and scorpions (the book lungs, which resemble internalized gills; Scholtz and Kamenz, 2006; Kamenz et al., 2008).

In spite of the appeal of a terrestrial chelicerate monophylum, phylogenetic relationships within chelicerates have proven elusive, exhibiting wildly different tree topologies depending on the type of data analyzed (i.e., morphological; sequence data; a combination of the two; mitochondrial gene order), the comparative sampling of extant and fossil forms, and the number of characters included (e.g., Giribet et al., 2002; Shultz, 2007; Masta et al., 2009; Regier et al., 2010). Support for the monophyly of arachnids is stronger in studies based only on morphological evidence (Firstman 1973; Grasshoff 1978; Weygoldt and Paulus 1979; Legg et al. 2013; Garwood et al. 2016; Wolfe 2017). Some works on arachnid phylogeny typically failed to test the monophyly of this group by using Xiphosura as the sole outgroup for tree rooting and character polarity (Shultz 1990, 2007; Garwood et al. 2016; Starrett et al. 2017). While the archetypal morphological phylogeny of Arachnida that is typically depicted is that of Shultz (2007), a few earlier works based on morphological studies opposed the hypothesis of arachnid monophyly altogether (Van der Hammen 1977, 1985; Grasshoff 1978), and more recent assessments by paleontologists have questioned even the monophyly of Xiphosura (Lamsdell 2013) and the composition of Chelicerata (Legg et al. 2013; Wolfe 2017). The advent of molecular phylogenetics did little to ameliorate the conflict. A survey of tree topologies based on molecular matrices of increasing size from the past two decades reveals a variety of conflicting hypotheses. A small number of studies favors the traditional sister group relationship of Xiphosura and Arachnida, albeit with low or modest support (Regier et al. 2010; Zwick et al. 2012). The more common result in studies of chelicerate or arthropod relationships, based on few nuclear and mitochondrial loci, is a nested placement of horseshoe crabs within Arachnida (Wheeler and Hayashi 1998; Giribet et al. 2001, 2002; Mallatt et al. 2004; Mallatt and Giribet 2006; Masta et al. 2009; Pepato et al. 2010; Sanders and Lee 2010; Regier and Zwick 2011; Arabi et al. 2012). The controversy has not been resolved even with the application of genome-scale data. Analyses based on expressed sequence tag or pyrosequencing (454) data sets have also recovered the non-monophyly of Arachnida, albeit with limited sampling of arachnid orders (Roeding et al. 2009; Meusemann et al. 2010; von Reumont et al. 2011; Borner et al. 2014; Rehm et al. 2014). A recent investigation of chelicerate phylogeny using a broader sampling of arachnid transcriptomes and a handful of genomes showed pervasive conflict of phylogenetic signal, with data sets composed of slowly-evolving genes exhibiting maximum nodal support for the monophyly of Arachnida, whereas the majority of analyses supported the nested placement of Xiphosura within Arachnida (Sharma et al. 2014a).

Understanding the placement of these groups in the chelicerate phylogeny is relevant for contextualizing the evolution of genome architecture, with different chelicerate lineages exhibiting episodes of gene expansion, in some cases consistent with ancient whole genome duplication. These duplications have been documented in scorpions (Sharma et al. 2014b, 2015), spiders (Clarke et al. 2015; Schwager et al. 2017) and horseshoe crabs (Kenny et al. 2016). Evidence of these duplications is also supported by multiple copies of microRNAs in spiders, scorpions, and xiphosurans (Leite et al. 2016). More generally, a robust and stable chelicerate phylogeny is a fundamental prerequisite for deciphering the history of chelicerate terrestrialization, venom evolution, and the utilization of silk within arthropods (Sharma 2017; Santibáñez López et al. 2018).

At the root of the challenge of chelicerate phylogeny is the ancient age of the chelicerate radiation, together with the rapid diversification of the ordinal lineages (Rokas and Carroll 2006; Whitfield and Lockhart 2007; Dunlop 2010). These conditions can potentially engender spurious groupings due to incomplete lineage sorting (ILS). ILS, produced by stochastic coalescence leading to conflicting gene histories, is one of the best-documented sources of conflict between individual gene genealogies and the species tree (ILS, Maddison, 1997; Maddison and Knowles, 2006). This type of conflict has motivated the development of methods that account for individual coalescent gene histories (Edwards et al. 2007; Liu et al. 2009, 2010; Bryant et al. 2012; Chifman and Kubatko 2014; Edwards et al. 2016). In spite of these advances, theoretical studies have described conditions where successive short internodes will produce gene tree topologies that are more likely than the species tree. These gene trees that differ from the species tree by this process are deemed anomalous gene trees (AGT) and the condition of branch lengths where AGTs are more likely is known as the anomaly zone (Degnan and Rosenberg 2006). Further investigations of this phenomenon have suggested the potential deleterious effects of the anomaly zone in accurate species tree estimation even for coalescent-aware methods (Kubatko and Degnan 2007; Degnan and Rosenberg 2009; Degnan et al. 2012; Degnan 2013). The complexity of analytical estimation of gene trees probabilities in the anomaly zone increases with more taxa (Rosenberg and Tao 2008; Rosenberg 2013), thus obscuring its effects on the accuracy of phylogenetic estimation.

A separate hurdle for stable chelicerate relationships is that multiple orders of arachnids exhibit accelerated rates of evolution, such as mites, ticks, pseudoscorpions, and possibly palpigrades (Murienne et al. 2008; Dabert et al. 2010; Pepato et al. 2010). This condition can

engender the phenomenon of long branch attraction (LBA) and typically manifests as the long-branch orders clustering together and/or with the typically poorly sampled outgroup taxa. This signature of rapid evolution, genome rearrangements, and dynamic gene loss is readily apparent throughout some acarine genomes (Grbić et al. 2011; Hoy et al. 2016).

Here, we aimed to evaluate the position of Xiphosura within chelicerates using a denser sampling of loci via inclusion of recently sequenced genomes for three of the four extant species of Xiphosura, as well as new genomes and high-quality transcriptomes of various arachnid species. The increased availability of genomic sequencing allowed us to bridge a gap in sampling curated genomic data for pancrustacean (e.g., insects; crustacean) and myriapod (e.g., centipede; millipede) outgroups. Our data set also included new and high-quality transcriptomes of Pycnogonida (sea spiders, the putative sister group to the remaining Chelicerata). Our goal was to (1) evaluate the position of Xiphosura based on sampling of orthologous genes, identifying composition and rate biases; and (2) assess the impact of discordant gene histories (e.g., those caused by gene trees in the anomaly zone) on the placement of horseshoe crabs within Chelicerata.

Materials and Methods

Taxon Sampling and Orthology Assessment

The taxon sampling comprised 53 terminals, with Chelicerata represented by three Xiphosura, two Pycnogonida, and 34 arachnids. Exemplars of all extant arachnid orders were included except Schizomida (sister group to Uropygi; Clouse et al., 2017) and Palpigradi (*incertae sedis*; Shultz 2007; Sharma et al. 2014a). Outgroup taxa consisted of eight Pancrustacea, five Myriapoda, and one Onychophora. Taxonomic and accession data are listed in the Supplementary Table S1 available on Dryad at http://dx.doi.org/10.5061/dryad.2g1f4n5. Data from whole genome sequence (WGS) projects (gene models) and RNA sequence libraries (transcriptomes) were accessed from NCBI genome and TSA databases. For WGS projects where gene models were not available, genomic coding sequences were extracted from the scaffolds using the genomic feature file (gff) and bedtools vers. 2.26 (Quinlan and Hall 2010). For RNA sequence data, Illumina-sequenced libraries were preferred to optimize data completeness, but exceptions were made to accommodate phylogenetically significant lineages (e.g., the daesiid solifuge *Gluvia dorsalis*; the cheliferid pseudoscorpion *Chelifer cancroides*). Readily assembled transcriptomes were used when provided in the original sources (Sharma et al. 2014a; Kenny et al. 2016). Transcript and genome coding sequence files were processed with Transdecoder vers. 3.0.1 (Haas et al. 2013) to identify open reading frame sequences (ORF) and translated to amino acids. In the case of *Limulus polyphemus*, for which a reference genome and several transcriptomes are available, amino acid sequences from both sources were combined and redundancies resolved after validation by the orthology assessment.

Initial homology searches used an all-versus-all strategy using psi-BLAST (BLAST+ 2.6.0, Camacho et al., 2009) and then processed with mcl (i = 6) (Dongen 2000) to produce clusters of homologous sequences. Gene clusters were aligned and sanitized as described below. Gene family trees (GFT) where estimated with IQ-TREE v.1.5.5 (LG+R4, Nguyen et al., 2015). Groups of orthologous sequences were identified *de novo* from the GFT using the phylogenetic orthology criterion as implemented in UPhO (Ballesteros and Hormiga 2016), enforcing the presence of at least five or 15 species per orthogroup and a minimum branch support of the incident branch >95. In-paralogs, isoforms, and allelic variants were resolved in favor of the longest sequence; thus only one sequence per taxon was retained in each orthogroup; cases of ambiguous orthology membership were resolved in favor of the largest group.

Matrix Subsets, Composition and Pairwise Identity

General statistics from each gene partition, compositional homogeneity tests (χ^2), and compositional heterogeneity (RCFV) were computed using BaCoCa v.1.105.r (Kück and Struck 2014). Similarly, taxa failing χ^2 compositional homogeneity in each partition were identified from the IQ-TREE log files.

Mean pair-wise sequence identity (MPSI) was calculated for each alignment as described by Sharma et al. (2014a) with a custom script.

$$MPSI = \frac{\sum_{j=1}^{n} (I_j / P_j)}{n}$$

where I_j is the number or identical character pairs in the column j, and Pj the total number of pairwise comparison of characters (k) in the column j or $\binom{k_j}{2}$. Characters representing gaps or ambiguities (-, X, ?) were ignored for this calculation. This metric is used as a proxy of evolutionary change, independent of an explicit model and tree topology. We used this score to concatenate genes in progressively descending order of MPSI, to inspect the effect of adding increasingly faster evolving genes on phylogenetic inference.

Partitions and taxa potentially affected by LBA artifacts were identified from individual gene trees using the LB score metric as implemented in TreSpEx v1.1 (Struck 2014). A matrix using only the genes showing the more homogeneous distribution of branch lengths (i.e., lower LB heterogeneity score) was selected for concatenation and phylogenetic analysis.

More complete matrix subsets were produced to minimize missing data by (1) filtering partitions with less than 47 terminals, (2) the automatic matrix reduction criterion using MARE v0.1.2-rc (Meyer et al. 2011) with

default parameters, and (3) identifying partitions with decisive taxon compositions (Steel and Sanderson 2010; DellAmpio et al. 2013). The last of these were identified by enforcing the presence in each locus of at least one representative sequence in each of the following groups: Pycnogonida, Scorpiones, Xiphosura, Acari, Opiliones, Tetrapulmonata, and at least one outgroup taxon. Matrices were constructed by concatenating the selected individual alignments (all genes or subsets described below) using the geneStitcher.py script (https://github.com/ballesterus/Utensils)

Phylogenetic Methods

Sequences were aligned using MAFFT 7.3.8 (–anysymbol –auto, Katoh and Standley, 2013). Gap-rich regions were masked with trimAl 1.2 (-gappyout, Capella-Gutiérrez et al., 2009) and alignment coverage verified and sanitized with Al2Phylo (-m 50 -p 0.25 -t 20, Ballesteros and Hormiga, 2016). Phylogenetic inference of orthologous gene trees (OGT) was computed with IQ-TREE v.1.5.5 or 1.6.8 (Nguyen et al. 2015; Chernomor et al. 2016), coupled with model selection of substitution and rate heterogeneity based on the Bayesian Information Criterion (Kalyaanamoorthy et al. 2017) and 1000 ultrafast bootstraps to assess branch support (Hoang et al. 2018) (-m MFP -mset LG, JTT, WAG -st AA -bb 1000).

Maximum likelihood (ML) analyses of the concatenated matrices used the same parameters listed above, except that partitions and their precomputed best substitution models were explicitly declared (-spp partition.nex). Additional ML searches were conducted for the decisive data set with ExaML v. 3.0.19 (Kozlov et al. 2015) followed by 100 bootstrap replicates produced with RAxML v. 8.2.11 (Stamatakis 2014) and TNT v. 1.1 (Goloboff et al. 2008) was used to generate the corresponding starting trees.

The computationally demanding mixture models were explored only for the smallest data set (98 loci), using the posterior mean site frequency (PMSF) approach (Wang et al., 2018). Site frequencies were estimated under the $LG+C20+F+\Gamma$ mixture model using the best ML tree from the 3534-gene data set as the starting tree and followed by 100 bootstrap replicates (PMSF-T1). To rule out biases produced by the starting tree, an additional analysis was performed using site frequencies derived from the ML tree where the monophyly of Arachnida was constrained (PMSF-T2).

Finally, to account for the potentially deleterious effects of concatenating loci with conflicting gene genealogies, species trees were estimated with ASTRAL vers. 5.5.9 (Mirarab and Warnow 2015; Zhang et al. 2017) using the collection of OGT as the input.

Informativeness and Signal

The information content of the data set was evaluated using quartet likelihood mapping (Strimmer and Von Haeseler 1997) in IQ-TREE (-lm ALL). Clusters for the quartet mappings were defined for the clades Pycnogonida, Xiphosura, Opiliones, and Arachnopulmonta, effectively ignoring the rest of the terminals. The mapping was done for the concatenated (3534 loci) and the individual gene alignments, contrasting the power of resolution of the concatenated data set and the individual gene trees.

To detect the effects of conflicting signal among loci, we computed the difference in log-likelihood score (ΔGLS) as described in Shen et al. (2017). This metric compares the overall fitness of each locus between two competing tree topologies. The two trees contrasted correspond to the unconstrained (T_1) and constrained (T_2, enforcement of monophyletic Arachnida) ML trees estimated from the sparse (3534 loci) data set. Individual gene trees, isomorphic to the alternative topologies, were prepared by pruning terminals not represented in the locus, while preserving the tree structure and branch lengths (droptips.py). Gene-specific likelihood scores under the best fitting model were computed in RAxML (-f G -m PROTGAMMAAUTO). To account for differences in model adequacy, the scores were also estimated under a fixed $LG+\Gamma$ model. Finally, because the scores can be affected by the placement of taxa other than the clade of interest, ML gene trees were re-estimated with clade specific constraints on (1) the bipartition Eucheliceratа, (2) a clade comprised of Xiphosura and the non-Acari arachnids (thus representing the derived position of horseshoe crabs), and (3) the bipartition corresponding to the monophyly of Arachnida. Likelihood scores (ΔGLS) for contrasting these topologies were estimated under fixed $LG+\Gamma$ model.

Coalescent Simulation of Gene Trees

Coalescent gene tree simulations were conducted to investigate the effects of increasing amounts of AGT on phylogenetic accuracy to infer the underlying species tree (Degnan and Rosenberg 2006) and the recovery rate of the clade of interest.

Simulations were executed using DendroPy (Sukumaran and Holder 2010) under the coalescent model using two time-calibrated species trees, representing the alternative scenarios as the "true" species tree. In the first scenario, Xiphosura is placed within Arachnida (T_1) and in the alternative, Arachnida monophyly was enforced (T_2). The required ultrametric trees were produced from ML trees using the "chronos" (lambda=1, model=relaxed) function of the R package ape (Paradis et al. 2004; Paradis 2013). The fossil calibration points are based on Wolfe et al. (2016) and listed in the Supplementary Table S2 available on Dryad. The intensity of coalescent variance was adjusted by modifying the population size parameter (N_e) = 5, 10, 50, 100, 1000, 5000, 10,000, 100,000. Each simulation setting (parameter pair T, N_e) was used to generate 2300 gene trees and the process repeated for 50 replicates. The number of gene trees was chosen arbitrarily. Densitree

vers. 2.2.5 (Bouckaert 2010) was used to visualize the intensity of gene-tree discordance among treatments. The python implementation used to produce these simulations is provided in the Supplementary materials available on Dryad. Species trees derived from the collection of simulated gene trees were estimated using ASTRAL. The normalized quartet score reported by ASTRAL, which represents the proportion of quartets in the input gene trees that are congruent with the species tree, is reported as a quantitative proxy for the amount of discordance observed in the simulated gene trees. Additionally, the frequency of individual bipartitions in the collection of gene trees were computed with a custom script, allowing assessment of the relative frequency of bipartitions of interest in the simulated gene trees, e.g., those inducing the monophyly of Arachnida and the ones present in the template species tree.

Results

Orthology and Matrix Compositions

The composition and properties of the primary matrices are summarized in Table 1 and a Supplementary file available on Dryad (matrices.xlxs) details the properties of each locus, including alignment length, the best fitting evolutionary model, MPSI, and RCFV.

The phylogenetic orthology pipeline identified 10,514 orthogroups with a minimum of five species per orthogroup and 3565 when the threshold was set to 15 species. Preliminary ASTRAL trees of these two sets produced the same species tree with similar quartet scores, 0.751 and 0.747, respectively (Supplementary Figs. S1 and S2 available on Dryad). The collection of orthogroups found using the 15 species threshold resulted in 3,534 loci after re-alignment and cleaning; this data set was carried for downstream analyses and comprised the "sparse data set". The combination of genomes, high-quality transcriptomes (mostly Illumina) and phylogenetic orthology assessment produced a more compact data set with more homogeneous coverage of genes across the ingroup and outgroup taxa. The majority of the terminals was represented in 1000 or more loci in the sparse data set, with the exception of *Glomeris pustulata, Chelifer cancroides, Pseudocellus pearsei, Centruroides vittatus, Mesobuthus martensii*, and *Gluvia dorsalis*; of these the only one derived from WGS is *M. martensii*.

No locus in the sparse data set had all 53 species represented and nearly one-third of loci lacked horseshoe crab representatives altogether (see Table 1). By requiring the presence of Xiphosura and six additional lineages, a subset of 1499 loci was selected to comprise the "decisive data set". This sampling strategy guarantees that each locus has the taxon coverage to test the placement of horseshoe crabs explicitly. A more strict minimum species threshold using 47 taxa produced the compact matrix composed of 98 loci that was used to rule out effects of missing data and facilitate computationally intensive analyses.

Finally, the automatic matrix reduction method (MARE) retained 721 loci and removed the following nine taxa: *Sarcoptes scabei*, *Chelifer cancroides*, *Pseudocellus pearsei*, *Mesobuthus martensii*, *Centruroides vittatus*, *Gluvia dorsalis*, *Glomeris pustulata*, *Symphyllela vulgaris*, and *Peripatopsis capensis*. Additional downstream matrices were constructed to explore the effects of MPSI, compositional heterogeneity or biases in signal, as detailed in Table 1 and Supplementary materials available on Dryad.

Phylogenetic Results

The ML tree of the decisive data set found Xiphosura nested within Arachnida, as the sister group to Ricinulei (Fig. 1a); this topology is used as the basis for comparisons. Analyses based on the sparse data set recovered a similar topology except for the recovery of a monophyletic Acari (Acariformes + Parasitiformes) as the most basally branching chelicerates. Topology comparisons using the approximately unbiased test (Shimodaira 2002) rejected the monophyly of Arachnida ($p = 0.009$).

Across the different analyses, the placement of Xiphosura within Arachnida was not affected by the estimation method; summary coalescent (ASTRAL) and ML (IQ-TREE, EXaML) approaches consistently recovered horseshoe crabs nested within a paraphyletic Arachnida. The monophyletic status of well established groups (namely, all ordinal categories within Chelicerata, Tetrapulmonata, Myriapoda, Pancrustacea, and Mandibulata) received maximum support under all analytical conditions. These well-supported clades are collapsed in subsequent figures to facilitate readability but fully labeled trees are available in the Supplementary material available on Dryad along with NEWICK-formatted files (Supplementary Figs. S3–S10, S15–S17 available on Dryad). In all cases, horseshoe crabs were found within other arachnid lineages. The monophyly of Arachnida is only recovered in constrained analyses (T_2) or in those data sets that were enriched for loci predicated upon the constraint of arachnid monophyly (see ΔGLS results below). The internal relationships of the chelicerate orders remained unstable across the variety of data sets and analyses explored, save for the consistent recovery of Tetrapulmonata and its constituent ordinal relationships.

Particularly recalcitrant are the positions of Pseudoscorpiones, Solifugae, Parasitiformes, and Acariformes. Pseudoscorpiones jumped from a closer relationship with Acari to one closer to Arachnopulmonata between analyses. Solifugae (camel spiders) are found either allied with Acariformes (mites) or as sister group to Ricinulei + Xiphosura. The acarine lineages (Acariformes and Parasitiformes; mites and ticks, respectively) are more often found at the base of Eucheliceratа (the non-Pycnogonida chelicerates),

TABLE 1. Description and general properties of the main data sets analyzed

Data set	Number of taxa	Number of loci	With Xiphosura	Number of sites	Missing data (%)	Notes
Sparse	53	3534	2379	1,484,206	55.56	Standard phylogenomic data set using a relaxed taxon occupancy threshold (min. 15 spp.). The resulting unconstrained (T_1) and constrained (T_2) ML trees are the basis for ΔGLS calculations, PMSF, and gene tree simulations.
Decisive	53	1499	1499	596,510	43.10	Taxon decisive data set made of loci with representatives of Xiphosura, Pycnogonida, Acari, Opiliones, Tetrapulmonta, Scorpiones, and outgroups in each locus. Analyses based on Δ GLS category and selective taxon removal are based on this data set.
Compact	53	98	98	25,034	13.51	Stringent taxon occupancy threshold (min. 47 spp.) minimizing missing data.
MARE	44	721	548	217,014	32.33	Matrix constructed with the automated matrix reduction algorithm with default parameters.
$MPSI_{100-3K}$	53	100–3000	30–2004	38,294–1,209,812	53.03–61.94	Matrices produced concatenating loci with increasing more variable loci based on MPSI.
LB herogeneity	53	882	542	369,655	59.67	Matrix including only loci in the lower quartile of LB heterogeneity as estimated with TreSpEx.

Note: Additional analyses and their results are shown in the Supplementary material available on Dryad.

either as mutually monophyletic groups (the traditional Acari) or as a grade.

The reduced compact and MARE reduced matrices (Supplementary Figs. S6–S9 available on Dryad) showed similar topologies to the decisive data set and agreed on the derived placement of Xiphosura but differed on the placement of Solifugae. In both cases Solifugae was sister group to Xiphosura, while Ricinulei was resolved as sister group to Arachnopulmonata (compact data set) or sister to Solifugae + Xiphosura (MARE data set). The results from the compact (98 loci) data set using the mixture PMSF models produced alternative resolutions suggesting influence of the input tree on the result. The tree using frequencies derived from the constrained tree (T_2) show Xiphosura as sister to Solifugae instead of Ricinulei, the same pattern is shown by the ML tree under homogeneous partitioned model. From the same data set, ASTRAL and PMSF analyses using frequencies from T1, show the more common Ricinulei + Xiphosura resolution (Supplementary Fig. S7 available on Dryad).

Across MPSI and other *ad hoc* data sets, Xiphosura was consistently found within Arachnida. Most analyses resolved the horseshoe crabs as sister group to Ricinulei and more closely related to the Arachnopulmonata than Opiliones is to the latter. An alternative resolution, where Opiliones are recovered as more closely related to Arachnopulmonata than Xiphosura, was observed in three analyses: the 300 slowest-evolving loci based on MPSI (both ASTRAL and IQ-TREE, Supplementary Fig. S23 available on Dryad) and the 1200 slowest-evolving loci with ASTRAL only (Supplementary Fig. S27 available on Dryad).

Phylogenetic Signal

The cluster likelihood mapping from loci in the sparse data set found most quartets concentrated in the corners of the likelihood map, indicating strong resolution power of the data set in informing the likelihood of the three alternative quartet topologies (Fig. 2). None of the possible quartets fell in the uninformative region of the graph. The majority of the quartet mappings favor the placement of Xiphosura among arachnids (Q2 + Q3). The quartet grouping Xiphosura as the sister lineage of the arachnopulmonates was strongly supported in 387 of the 540 quartets (Q3), while only 39 quartets supported the sister group relationship of Xiphosura with Opiliones (Q2) and the arrangement where Opiliones and Arachnopulmonata are sister taxa (Q1) was supported in 109 quartets. Notably, Q1 is the only quartet (among the alternative quartet topologies tested) that corresponds to the possible monophyly of Arachnida, wherein the position of unstable taxa such as Acari, Pseudoscorpiones, and Solifugae is not considered. The likelihood mapping on the individual alignments found 397,889 quartets from 1692 genes. Genes where any of the clusters was not represented did not contribute in the analysis. In contrast to the concatenated analyses, the majority of the quartets fell in the uninformative region of the map (28.54%), suggesting lower power of individual loci to resolve the relationship of these four clusters. Of those in the informative regions, a slight majority supported quartet Q3 (20.72%), with 17.45% in support of Q1 and 14.09% supporting Q2.

The ΔGLS compares the fit of each locus between two competing topologies, allowing the classification of

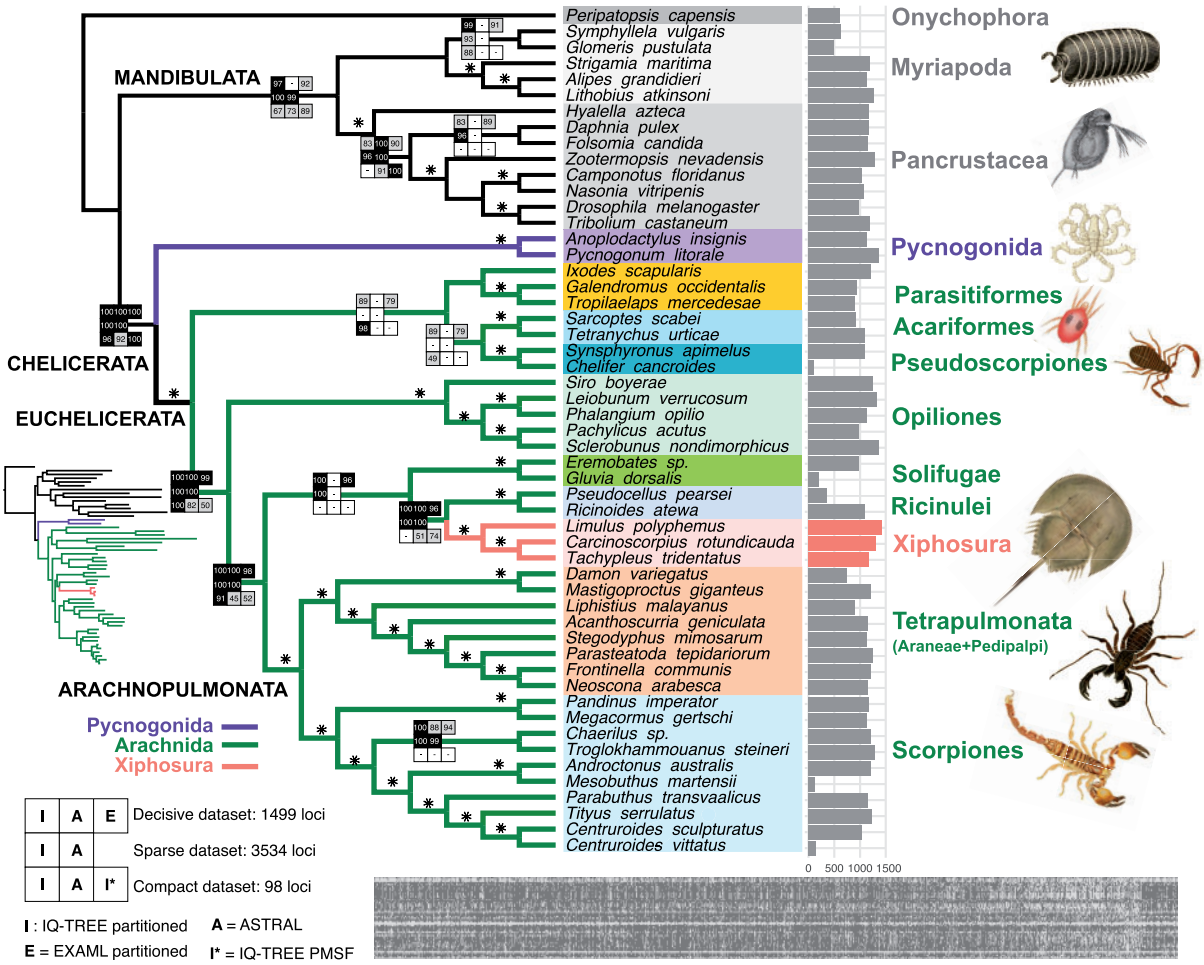

FIGURE 1. ML tree of the decisive data set (1499 loci). Differences in topology and support in alternative analyses is represented as navajo rugs (key on lower left corner); well supported (>95) bipartitions shown as black squares; lower supported and missing bipartitions shown in gray and white squares, respectively. A star above branches indicates bipartition well supported in all analyses. The gene occupancy of each taxon in the decisive data set is represented with bars and an overview of matrix density is shown in the bottom. Drawings of representative lineages in the tree, from top to bottom, are as follows: *Glomeris* (Myriapoda), *Daphnia* (Crustacea), *Pycnogonum* (Pycnogonida), *Tetranychus* (Acariformes), *Chelifer* (Pseudoscorpiones), *Tachypleus* (Xiphosura), *Mastigoproctus* (Uropygi), *Pandinus* (Scorpiones). All images sourced from works in the public domain (Blanchard, 1877; Claus, 1884; Ewing, 1914; Koch and Hahn, 1841–1843; Möbius, 1902, Naturalis Biodiversity Center / WikimediaCommons RMNH.ART.30).

loci favoring one topology over the other. A graphical representation of ΔGLS using $LG+\Gamma4$ for loci in the decisive data set is shown in Figure 3a. The graph shows 996 loci with favorable scores for the derived position of Xiphosura within Arachnida (T_1, max=55.08, mean=7.84) and 503 in favor of topology T_2, constrained on Arachnida (max = −27.53, mean = −4.85). A similar pattern of values was observed using automatic model selection, where 976 favored T_1 and 523 T_2. Using clade-specific contrasts, gene trees constrained to recover the Eucheliceratа bipartition were overwhelmingly favored over gene trees constrained to recoverer the Arachnida bipartition (Eucheliceratа: 1363; Arachnida: 97; neutral: 39). Likewise, gene trees constrained to recover the clade including Xiphosura and all non-acarine arachnids were favored in 1027 loci, whereas 472 loci favored the alternative gene trees constrained on Arachnida monophyly. These scores show the same trends and results for the sparse data set; ΔGLS scores for all partitions are available as Supplementary Figures S11–S13 available on Dryad.

Phylogenetic analyses of the 503 loci favoring T_2 (regardless of the substitution model used to score the metric) resulted in topologies where a monophyletic Arachnida was recovered with high support (Fig. 3c and Supplementary Figs. S11 available on Dryad). Notably, summary coalescent species trees (ASTRAL) using the gene trees of the same 503 loci favorable to T_2 still showed Xiphosura within Arachnida, albeit in a more basal position (Supplementary Fig. S11 available on Dryad). The summary coalescent method recovered arachnid monophyly only when analyzing gene trees of the 472 loci that were identified using clade-specific contrasts (Supplementary Fig. S11 available on Dryad).

Correlation of ΔGLS ($LG+\Gamma4$) with intrinsic properties of the individual loci in the decisive data

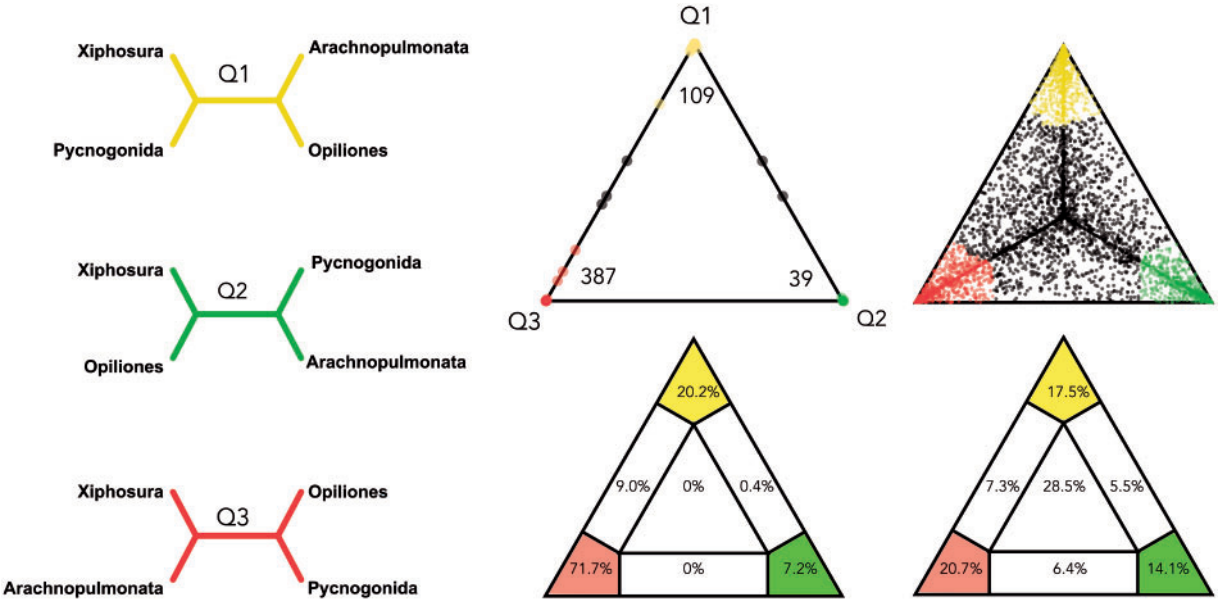

FIGURE 2. Likelihood maps of the three alternative quartet topologies shown on left column. Center column shows the results from the concatenated data set (3534 loci) with the mapping of the quartets (top) and their respective proportions in the informative regions of the map (bottom) data set. Right column aggregates the mapping of 5000 random quartets from 397,889 quartets in 1692 loci (top), and the summary distribution on the likelihood areas (bottom).

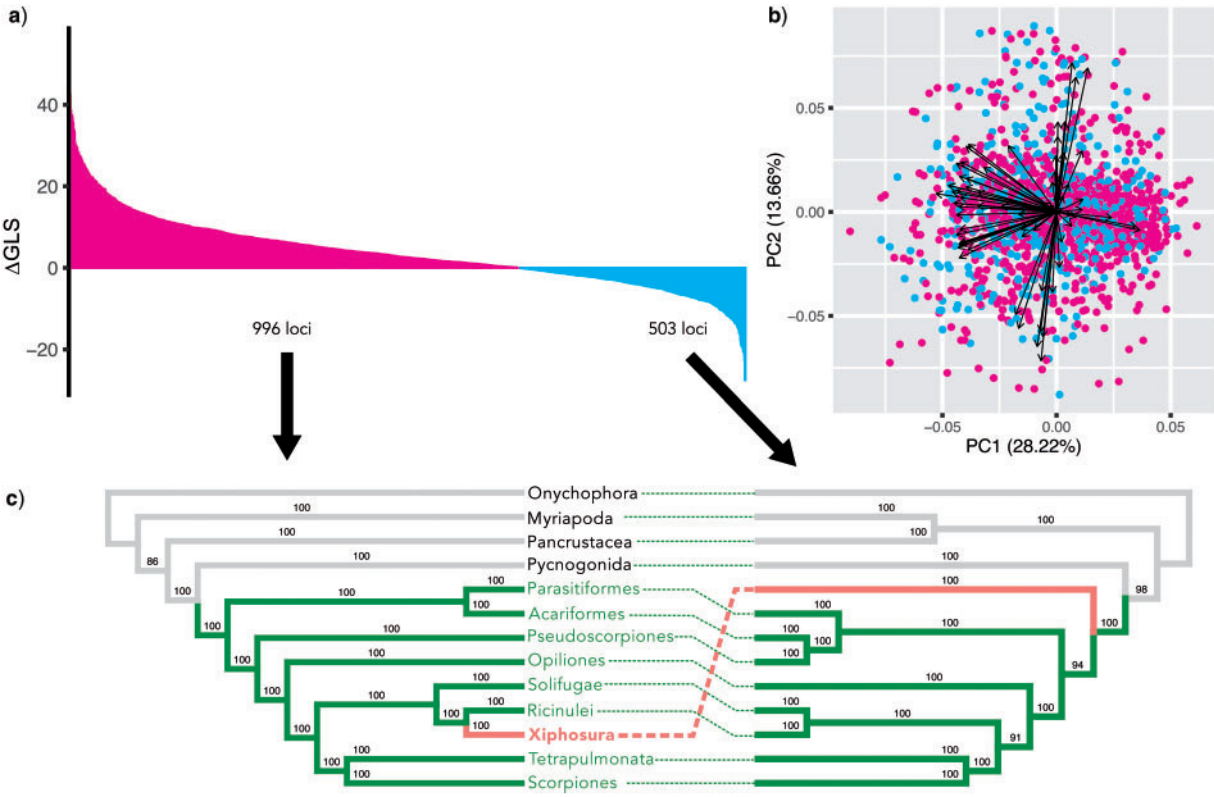

FIGURE 3. a) Ranked distribution of Δ*GLS* of 1499 loci under a fixed *LG*+Γ4 model. Loci with positive values favor the ML tree where Xiphosura is nested within Arachnida (T1); loci with negative values favor the monophyly of Arachnida (T2-constrained). b) Distribution of loci in the two first principal components from 70 variables. Loci colored by Δ*GLS* category; dark gray dots (fucsia) = T_1, light gray (cyan) = T_2. Vectors of the loadings of each variable shown as black arrows. No clustering of Δ*GLS* classes is observed and none of the variables correlate with support for either of the alternative topologies as indicated by Δ*GLS*. (c) Separate phylogenetic analyses of the loci classified by Δ*GLS* produce conflicting results with maximum branch support.

set was explored using principal components analysis (PCA, computed in R using prcomp; R Core Team, 2017). Variables for PCA, such amino acid composition, polarity, and RCFV, were extracted from BaCoCa's report and complemented with general alignment statistics for number of species, missing data, MPSI, and TreSpEx LB heterogeneity for a total of 70 variables. The dispersion of loci in the first two PCAs shows no apparent clustering or segregation trends based on ΔGLS, with loci in both categories interspersed across PC space (Fig. 3b). Similarly, PCA using all 3534 loci in the sparse data set showed the same scattered dispersion and, although not in a defined cluster, loci with neutral scores ($\Delta GLS = 0$) concentrated on the left side of the plot where the proportion of uninformative sites is among the variables with highest loading.

Directed pairwise comparisons (Table 2) of sequence length, composition heterogeneity (RCFV), number of species, and mean pairwise identity (MPSI) between loci in favor of T_1 or T_2 found significant differences in sequence length and MPSI (one-tailed *t*-test and Wilcoxon rank sum test $p < 0.05$) between loci favoring each of the contrasting topologies, indicating that loci in favor of the unconstrained (T_1) tend to be longer and faster evolving. No statistical differences were observed based on RCFV or the number of species represented in each locus.

Composition Biases

Only a small fraction of the partitions (106 out of 3534) failed the compositional homogeneity test ($p \leq 0.05$). A phylogenetic analysis of the partitions that failed the χ^2 test showed obvious topological anomalies, with the millipede *Glomeris pustulata* within Arachnida and a non-monophyletic Solifugae (Supplementary Fig. S19 available on Dryad). The analysis where serine positions were replaced by ambiguities to avoid non-homologous serine similarities (Rota-Stabelli et al. 2012; Zwick et al. 2012) was also congruent with the originally coded data set (Supplementary Fig. S18 available on Dryad).

Taxon-wise χ^2 tests in the individual alignments found the majority of the taxa displayed homogeneous amino acid compositions (Supplementary Fig. S14 available on Dryad). Among chelicerates, the taxa most frequently failing the homogeneity test were *Sarcoptes scabei* (364 of 1660 loci), *Ixodes scapularis* (362 of 2185 loci), and *Chelifer cancroides* (17 of 176 loci). For all other taxa, including Xiphosura, the hypothesis of compositional homogeneity was favored in more than 95% of the loci where the taxon is present.

Analyses of partitions with low RCFV (< 0.025) and high RCFV (> 0.15) loci both recovered the derived position of Xiphosura within arachnids. In the low RCFV case, Xiphosura was found as the sister group to Arachnopulmonata; under high RCFV, Xiphosura was recovered as sister group to Solifugae. Both cases displayed lower branch support but no other anomalies (Supplementary Fig. S17 available on Dryad).

Long Branch Attraction

The effect of evolutionary rate was explored by analyzing matrices of loci concatenated by decreasing order of MPSI, thus adding increasingly more variable partitions. The resulting best ML trees of these data sets and their corresponding ASTRAL species trees consistently found Xiphosura nested within Arachnida (Supplementary Figs. S22–S30 available on Dryad). The trajectories of bootstrap frequencies of selected clades across these data sets are shown in Figure 4. These graphs allow for visual inspection of the effect of adding increasingly more variable loci on the recovery of specific clades in the collection of bootstrapped trees. For example, the panel for Chelicerata shows that this clade occurs in less than half of the bootstrap replicates for the slowest-evolving loci up to 600 genes, whereas the Cormogonida hypothesis (Pycnogonida sister group to all remaining Arthropoda), is favored by this subset. The trend is reversed in favor of a monophyletic Chelicerata with the addition of faster-evolving sites. Uncontroversial groupings, such as Euchelicerata, exhibited maximum support across all data sets. Arachnopulmonata *sensu stricto*, comprising Scorpiones, Pedipalpi, and Araneae, also exhibited maximum support when including faster evolving loci; this pattern is seemingly inverse to the trend observed for the clade of Pseudoscorpiones and Arachnopulmonata, where only the slower-evolving matrices favor a closer relation of Pseudoscorpiones with Arachnopulmonata. Controversial groupings and conflicting groupings, such as a monophyletic Acari, or the clade Acariformes + Pseudoscorpiones, showed moderate to low frequencies. The bipartition inducing Arachnida *sensu stricto* was not found in any of the bootstrap replicates regardless of evolutionary rate. Rather, the clade Xiphosura + Ricinulei attained maximum support for matrices including fast-evolving genes.

The placement of Xiphosura within Arachnida was robust to the removal of outgroups and known long-branch taxa (Fig. 5a–d, Supplementary Fig. S10 available on Dryad). Differences between the taxon-reduced topologies and the complete analyses are observed in the relationship of the acarine orders, Pseudoscorpiones, and Solifugae. None of the taxon-reduced trees showed Acari and Pseudoscorpiones forming a clade; instead these groups were resolved as a paraphyletic assemblage (Fig. 5a,d) or, in the absence of Acari, with Pseudoscorpiones closely allied with scorpions (Fig. 5b). In addition, solifugids moved to a more basally-branching position upon the removal of Acari (Fig. 5b,c). Finally, we tested the possibility of horseshoe crabs being attracted to its derived position by Ricinulei or Solifugae. As before, the derived placement of Xiphosura within the arachnids remain unchanged upon the removal of both those taxa (Fig. 5d). These analyses indicate that the inclusion of distant outgroups has no observable effects on the nested placement of horseshoe crabs by way of attracting fast-evolving clades towards the base of the tree. On the other hand,

TABLE 2. Summary and comparison of some alignment statistics of loci in the sparse data set classified based on ΔGLS as supporting of T_1 orT_2 under best (AUTO) and fixed ($LG+\Gamma$) amino acid substitution models)

	Aln. length	Num. spp	MPSI	RCFV
T_1 AUTO $(\bar{x},\sigma)$	445.63, 372.19	31.22, 8.98	0.59, 0.09	0.08,0.03
T_2 AUTO $(\bar{x},\sigma)$	372.90, 323.31	31.22, 9.57	0.61, 0.09	0.08,0.03
p (T-test)	1.4e−07	0.99	3.6e−05	0.08
p (Wilcoxon test)	9.2e−11	0.774	2.9e−05	0.08
T_1 $LG+\Gamma$ $(\bar{x},\sigma)$	445.04, 371.07	31.32, 9.013	0.5981, 0.09	0.08, 0.03
T_2 $LG+\Gamma$ $(\bar{x},\sigma)$	373.0303, 324.813	31.03, 9.54	0.61, 0.10	0.08, 0.03
p (T-test)	2.1e−07	0.44	17.1e−5	0.31
p (Wilcoxon test)	1.5e−10	0.27	24.49e−5	0.23

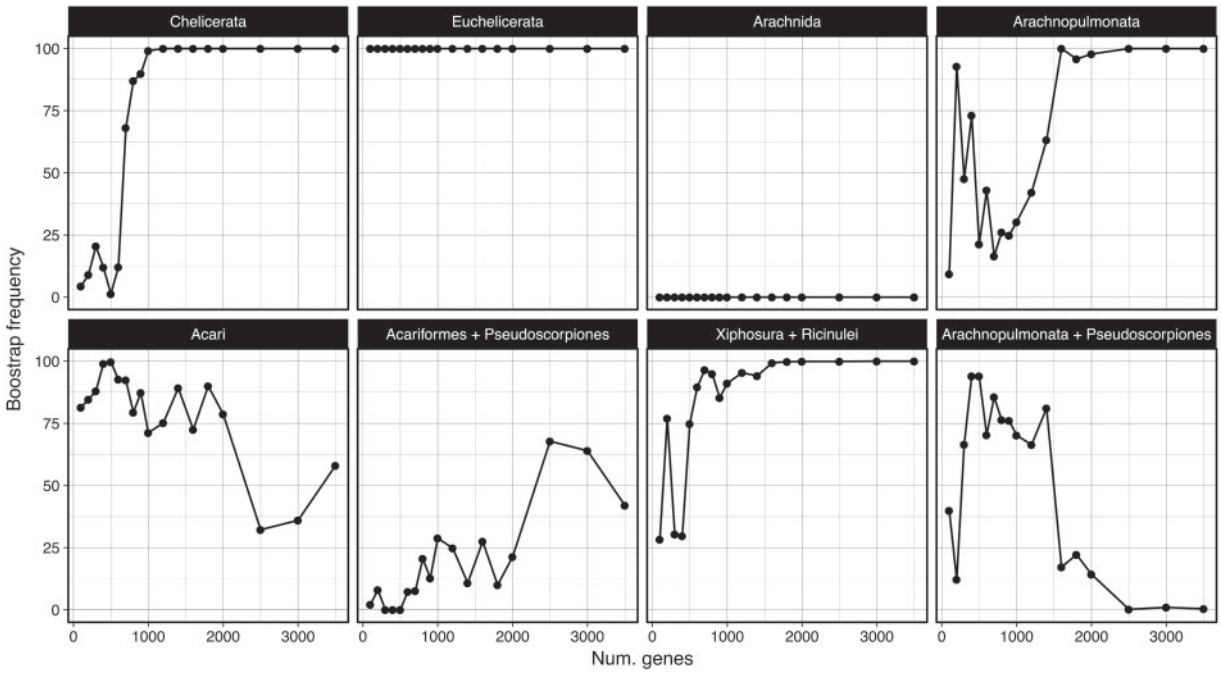

FIGURE 4. Bootstrap resampling frequency of specific clades on matrices constructed by adding increasingly fast-evolving loci, as ranked by MPSI score.

the placement of Pseudoscorpiones near the base of arachnids, observed in many of the analyses, may indeed be the result of LBA artifacts.

In addition, partitions potentially suffering from LBA artifacts were identified based on the LB score heterogeneity as implemented in TreSpEx (Struck 2014). This metric estimates differences in the patristic distances across all tips in individual gene trees, with lower scores indicating homogeneity in branch lengths and higher values suggesting disparity in branch length. Per partition scores ranged from 5.55 to 210.01 ($\bar{x}=$ 27.19). Phylogenetic analysis of 882 partitions in the lower quartile of LB score heterogeneity (≤ 18.91) showed no difference in the placement of Xiphosura within Arachnida, as the sister group of Ricinulei and this clade in turn sister group to Solifugae (Supplementary Fig. S19 available on Dryad). Relative to other taxa, horseshoe crabs showed shorter branch length scores ($\overline{LB}=$ $-14.92, -15.4, -15.7$, Supplementary Fig. S18 available on Dryad), similar in value and dispersion to other arachnids. By comparison, clearly fast-evolving taxa showed higher LB scores, such as Pseudoscorpiones ($\overline{LB}=16.2, 44.3$) and Acari (e.g., $\overline{LB}=53.9$ for *Sarcoptes scabei*).

Simulation of AGT

The time calibrated unconstrained (T_1) and constrained (T_2) species trees used as inputs in the simulations are shown in Supplementary Figure S31 available on Dryad. Both trees represent scenarios of successive cladogenesis over a short period of time at the base of the Euchelicerata, with potentially deleterious effects on the reconstruction of the topology.

The set of *Ne* values explored produced a wide range of coalescent variance in the simulated gene trees, thus providing disparate scenarios to explore the effects of increasing levels of ILS (see Fig. 6 and Table 3).

Relative frequencies of individual bipartitions in the simulated gene trees are shown in Figure 6 for three values of *Ne* and the constrained tree. Corresponding figures for the other *Ne* and species tree are shown

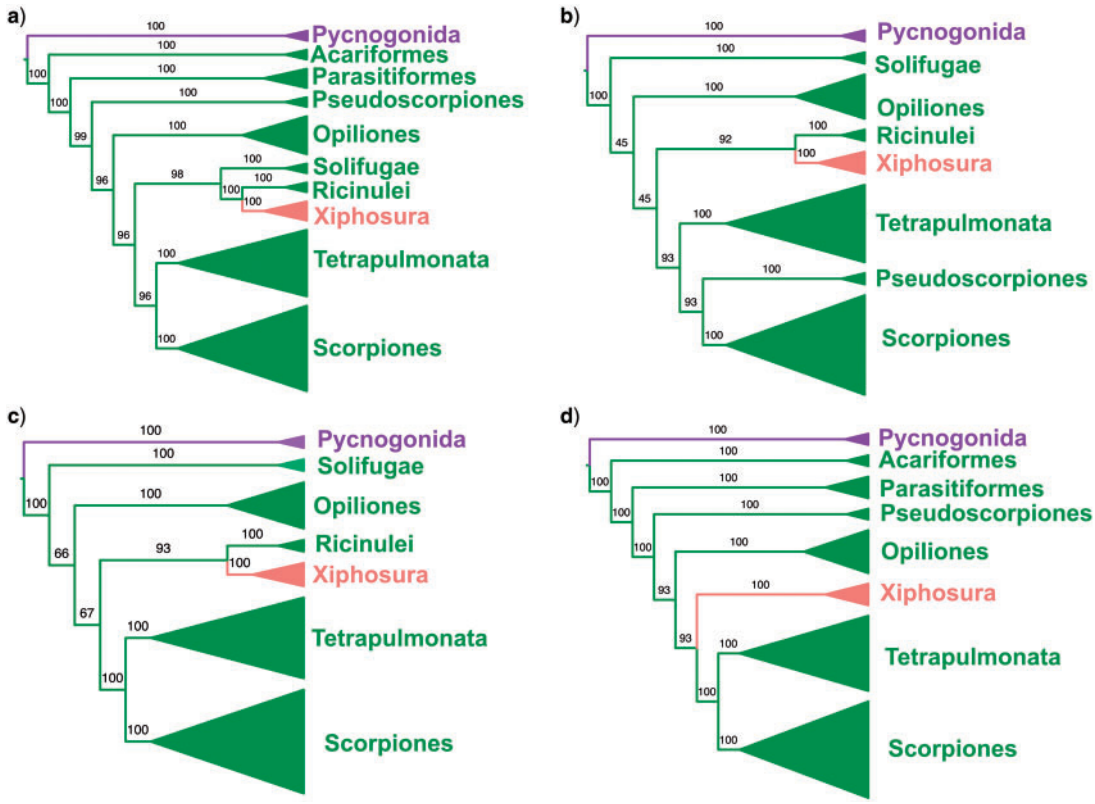

FIGURE 5. Phylogenetic effects of the selective removal of outgroups and long branch taxa from the decisive data set. a) Non-chelicerate outgroups removed. b) Acari orders (Parasitiformes and Acariformes) removed. c) Pseudoscorpiones and Acari removed. d) Acari and Pseudoscorpiones are included but Ricinulei and Solifugae removed.

in Supplementary Figure S32 available on Dryad. Independently of the template species tree, simulating smaller values of *Ne* (5, 10, 50) showed that the most frequent bipartitions in the simulated gene trees were the same as the corresponding template species trees. With moderate levels of ILS, ($Ne = 100$), some anomalous bipartitions were more common than the one in the template tree, but some "true" bipartitions occurred in most of the gene trees. At high levels of ILS (*Ne*=1000, 5000, and 10,000) anomalous bipartitions were the most frequent ones. With increasing levels of ILS, the proportional frequency of all bipartitions was considerably reduced; with *Ne*=5, 10, 100, there was at least one non-trivial bipartition present in all the gene trees (that is present in all gene trees); at *Ne*=1000, the most frequent bipartition occurred only in 17% of the gene trees. This frequency dropped to 1.5% with $Ne = 100,000$.

In spite of the gene tree conflict, ASTRAL was able to accurately infer the species tree under low and medium levels of ILS for both template trees ($Ne = 5 - 1000$). Majority rule consensus trees are shown in Supplementary Figures S33–S36 available on Dryad. Anomalous groupings were observed for trees estimated from simulated gene trees with $Ne > 1000$. Trees simulated under the constraint of arachnid monophyly recovered instances of Xiphosura breaking the monophyly of Arachnida, as well as obvious anomalous groupings, such as a polyphyletic Xiphosura, Opiliones, Pycnogonida, Pancrustacea, and Myriapoda.

Using the collection of AGTs simulated on T_2, ASTRAL recovered Arachnida in the vast majority of the 50 replicates under low to moderate levels of ILS (5, 10, 100, 1000). Exceptions were found on two ASTRAL trees with $Ne = 100$ and seven with $Ne = 1000$, where Xiphosura was recovered as the sister group to Acari (Fig. 7a). The success of ASTRAL in recovering this clade is remarkable even when the proportional frequency of a true species tree bipartition (e.g., Eucheliceratа) in the AGTs drops to zero. In other words, none of the simulated gene trees showed that bipartition, yet it was correctly recovered in the species tree. These cases are observed for $Ne = 1000$ and higher values. This counterintuitive result occurs because ASTRAL relies on frequencies of quartets, rather than bipartitions, to estimate the underlying species tree. Only AGTs under high levels of ILS resulted in ASTRAL trees consistently rejecting the monophyly of Arachnida. In a similar manner, the clade composed of Xiphosura and Ricinulei, found in T_1, showed similar resilience to increasing levels of ILS (Fig. 7b).

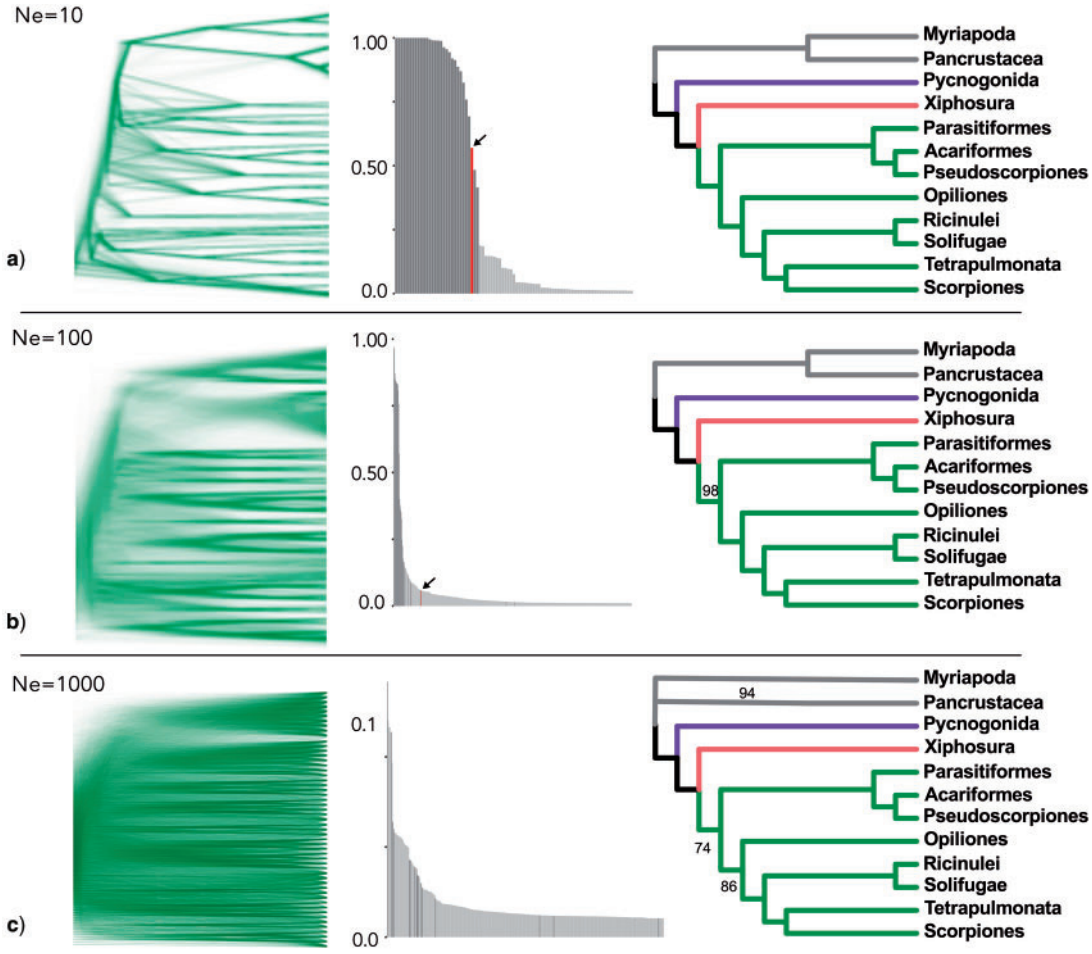

FIGURE 6. Using gene genealogies simulated on the constrained species tree, ASTRAL recovers the "correct" species tree under mild (a, *Ne* = 10), medium (b, *Ne* = 100), and high (c, *Ne* = 1000) levels of gene tree incongruence. The amount of conflict is represented visually by cloudograms of gene trees (left). Plots of ranked relative frequencies of individual bipartitions in the gene trees (center) show the effects of ILS on the proportion of "correct" (present in the template species tree, dark gray) and anomalous bipartitions (light gray). The bipartition underpinning the monophyly of Arachnida is shown with an arrow. Under low (a) and mild (b) levels of ILS, the correct bipartitions are more abundant than anomalous ones. With high ILS conditions, anomalous bipartitions are more frequent than correct ones; nonetheless, the overall frequency of any bipartition is low. In spite of the conflict in gene genealogies, the majority rule consensus of the 50 ASTRAL tree replicates (right) under each condition shows a high recovery rate for the "true" species tree.

DISCUSSION

Horseshoe Crabs are Aquatic Arachnids

The resolution of chelicerate phylogeny is obscured by its age, rapid diversification, and accelerated evolutionary rate in some orders. Phylogenomic approaches, including the present study, have corroborated many relationships (e.g., the monophyly of most chelicerate orders; Chelicerata; Euchelicerata; Arachnopulmonata; Tetrapulmonata). Other relationships between chelicerates remain highly contentious. Intriguingly, all unconstrained tree topologies inferred from the refined and updated phylogenomic data set analyzed here recovered horseshoe crabs as nested within the arachnids, most frequently in a clade with Ricinulei. The clade (Xiphosura + Ricinulei) in turn was frequently recovered as the sister group of the clade of arachnids that possess book lungs (Arachnopulmonata).

This result is surprising because traditionally, Ricinulei have been regarded as close relatives of Acari, in part due to the hexapodous larva found in all three of these orders (Weygoldt and Paulus 1979; Shultz 1990, 2007). From the perspective of morphology, there are not many obvious similarities between Xiphosura and Ricinulei, because the latter respires using tubular tracheae, whereas horseshoe crabs respire using external book gills. Previous hypotheses of these organs' evolutionary origin favored a scenario wherein the book gills are plesiomorphic and homologous to the book lungs of scorpions, and supported the position of Eurypterida (sea scorpions) as part of a grade between these states (Dunlop 1997; Braddy et al. 1999; Scholtz and Kamenz 2006). If the

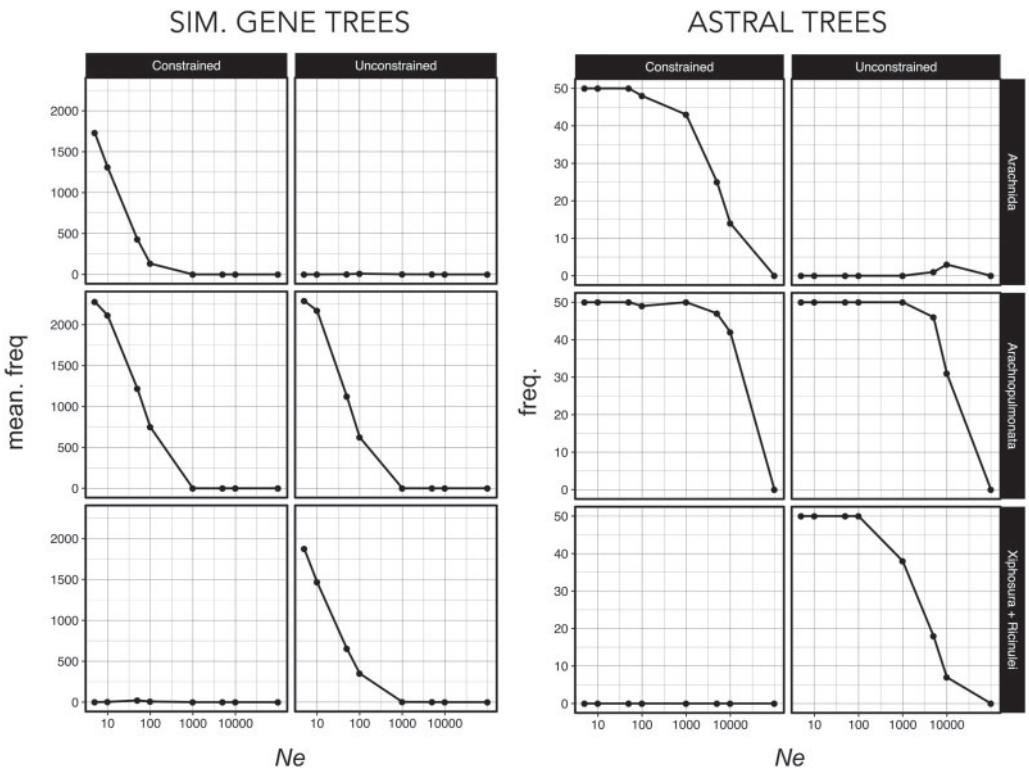

FIGURE 7. Recovery rate of specific clades under increasing *Ne* conditions in the collection of simulated gene trees (left) and their corresponding ASTRAL species trees (right). Trends are shown for simulations under the two alternative "true" species trees scenarios (unconstrained or with the constraint of monophyletic Arachnida). For the gene trees, frequency values are averaged across the 50 simulation replicates (max. frequency = 2300); for the ASTRAL trees, maximum recovery frequency is 50, one per simulation replicate. The bipartitions traced (top to bottom) are the monophyletic Arachnida, the uncontroversial Arachnopulmonata, and the clade formed by Ricinulei and horseshoe crabs.

relationship of Ricinulei + Xiphosura (and the placement of this clade as sister group to Arachnopulmonata) were phylogenetically accurate, then this tree topology could be consistent with a scenario of colonization of aquatic habitat by the ancestor of horseshoe crabs. Alternatively, the relationship could be interpreted to imply independent adaptations to the terrestrial habitat for Ricinulei and arachnopulmonates (i.e., more than one terrestrialization event), an interpretation supported by the incidence of putatively aquatic scorpion fossils from Paleozoic strata (Dunlop 2010; Waddington et al. 2015). Incidentally, Ricinulei has been proposed to be the sister group of the extinct order Trigonotarbida (which bore book lungs; Dunlop, 1996; Kamenz et al., 2008; Garwood and Dunlop, 2011) but the placement of trigonotarbids based on morphological data is tenuous as well (Garwood et al. 2016,?).

While solving chelicerate phylogeny is beyond the scope of this data set, the recurrent placement of Xiphosura as nested within arachnids, regardless of degree and manner of analytical perturbation, demands scrutiny. The acquisition of complete genomes for all three extant horseshoe crab genera disfavors a dismissal of this result based on such criticisms as insufficient data or incomplete taxonomic sampling. For this reason, we directed our attention to systematic biases and known pitfalls for phylogenomic investigations.

Long Branch Attraction

LBA is the most frequently invoked cause of anomalous groupings in phylogenetic analysis (Anderson and Swofford 2004; Bergsten 2005; Brinkmann et al. 2005; Philippe et al. 2005; Jeffroy et al. 2006). Although the topic has received substantial attention in theoretical and empirical studies alike, it remains difficult to assess the incidence, intensity, and effect of LBA on empirical tree inference (Lyons-Weiler and Hoelzer 1997; Rodríguez-Ezpeleta et al. 2007; Qu et al. 2017).

Common solutions to escape LBA artifacts include the use of more complex models of evolution and rate heterogeneity (Lartillot et al. 2007), removal of fast-evolving sites (Philippe and Roure 2011), and the selective removal of outgroups and long-branch taxa. In the case of chelicerates, some orders are known to display comparatively higher rates of molecular evolution, such as Acari (Dabert et al. 2010) and Pseudoscorpiones (Murienne et al. 2008); LBA is partially responsible for the unstable positions of these taxa. Admittedly, the LBA effects observed in these taxa may be aggravated by limited taxon sampling in these taxa as well, reflecting the paucity of genomic data available for the smaller chelicerate orders. In the case of the placement of Pseudoscorpiones, analyses of the slower-evolving genes (based on MPSI) and the decisive data set where

acarine taxa were removed, found pseudoscorpions closely allied to Arachnopulmonata, while still showing Xiphosura within Arachnida. Likewise, the derived position of horseshoe crabs was observed in analyses using only the loci less prone to LBA based on a branch length metric, but notably this tree still showed Pseudoscorpiones at the base of the tree in a close relationship with the acarine lineages, potentially due to fewer loci available for these long branch taxa, and thus confounding the effects of missing data and LBA.

The relation of LBA and taxon sampling is well documented (Bergsten, 2005), and expanding taxon sampling is one the most reliably invoked solutions to escape LBA artifacts. With almost all extant species of horseshoe crabs (three out of four) represented in out analyses, deep taxon sampling within Xiphosura is exhausted as solution to breaking long branches in this order. Moreover, in the case of Xiphosura and Ricinulei, we note that the characteristically large patristic distances associated with LBA were never observed for these taxa.

In spite of variation in the branching pattern, data sets composed of slower- and faster-evolving loci alike rejected the monophyly of Arachnida. Analyses of similarly constructed data sets in Sharma et al. (2014a) suggested that failure to recover Arachnida could be attributable to the effects of adding faster-evolving sites; they recovered Arachnida with maximal nodal support only in matrices composed of the 500 and 600 slowest-evolving loci. A reanalyses of that 500 locus data set from Sharma et al. (2014a) replicated the reported result of monophyletic Arachnida when using the same $LG+\Gamma$ model per partition. However, analyses using the best-fitting substitution and rate heterogeneity model resulted in Parasitiformes taking the place of Xiphosura as the most basally branching Eucheliceratа (Supplementary Figs. S20–S21 available on Dryad). These reanalyses demonstrate that even in data sets where the monophyly Arachnida is supported, this result is fragile to alternative analytical parameters. In a similar manner, the removal of loci with heterogeneous branch length score in our data set, and consequently of partitions more likely to induce LBA artifacts, had no effect in the placement of horseshoe crabs within Arachnida.

Separately, the use of complex infinite mixture models (Lartillot et al. 2007) has been suggested as an approach to overcome systematic errors generally and long branch attraction particularly. The debate over the use of mixture versus homogeneous models constitutes the fulcrum of some of the most controversial nodes in the animal tree of life (e.g., the placement of Porifera and Ctenophora; the basal phylogeny of Spiralia), with alternative resolutions obtained under competing analyses (Kocot et al. 2011; Moroz et al. 2014; Struck et al. 2014; Borowiec et al. 2015; Whelan et al. 2015; Simion et al. 2017; Laumer et al. 2017; Whelan et al. 2017). At the same time, some results suggest that in many cases, the use of CAT mixture models does not out-perform analyses using partitioned homogeneous models (Li et al. 2017; Whelan and Halanych 2017). Current implementation of the CAT model, particularly in the Bayesian framework, is computationally demanding (Pisani et al. 2015; Whelan and Halanych 2017; Whelan et al. 2017) and only feasible for reduced matrices (e.g., Simion et al., 2017), forcing a tradeoff between "realistic" models and the amount of empirical evidence that can be analyzed. Runs on the smallest (98 genes) data set using mpi-phylobayes with $CAT+GTR+\Gamma$ model failed to converge after three months of computation on 16-32 high-performance cores (max split. diff = 0.32851, ess = 156). While results from incomplete runs should be approached with skepticism, the summary tree of this run also recovered the nested position of Xiphosura within Arachnida in spite of using the ML tree with arachnid monophyly constrained (T_2) as the starting tree. An alternative approach to the mixture model has been recently described and implemented in IQTREE (Wang et al. 2018). This method first optimizes the parameters of a mixture model on an input tree, and uses the computed site frequencies to search for ML trees. Although noticeably faster than the phylobayes implementation of CAT+GTR, the fitting PMSF model still proved prohibitive for larger data sets, requiring more than 383 GB of RAM for the 1499-gene data set. The results from the PMSF analyses on the 98-gene data set from two different input trees agreed on the derived the placement of horseshoe crabs but one placed horseshoe crabs sister to Ricinulei, while the second favored Xiphosura as the sister taxon of Solifugae (cf. Supplementary Fig. S7 available on Dryad). The impact of the initial tree for PMSF model requires further exploration but the instability may be due to insufficient information in the reduced data set. Similar conflict was observed between ASTRAL and ML partitioned analyses of the same data set with similarly low bootstrap values. It must also be noted that the use these complex models, which require combined data from many genes to estimate model parameters accurately, is in direct opposition to the multispecies coalescent paradigm. In the face of incongruence in gene genealogies, researchers must decide if disagreements are better explained by coalescent variance (thus favoring coalescent-aware methods of inference) or by limited signal of individual loci (in which complex models of evolution may be a reasonable strategy).

In our data set, the use of either type of substitution model had no effect on the placement of Xiphosura as clearly nested within the arachnids, even when the starting tree topology was congruent with the hypothesis of monophyletic Arachnida. Similarly, the removal of outgroups and selected taxa had no effect on the placement on the recovery of Xiphosura + Ricinulei, and the removal of Ricinulei did not alter the placement of horseshoe crab or the remaining taxa. We, therefore, submit that LBA artifacts do not appear to affect either Xiphosura or Ricinulei; LBA is an insufficient explanation for the consistent recovery of horseshoe crabs as derived arachnids.

Sequence Compositional Bias

Although taxon- and partition-wise compositional biases are known to induce phylogenetic error, we did not find evidence of systemic skews in the amino acid composition in our data set. Consequently, removal of these partitions had no impact on the results.

Skews on sequence composition were detected with a χ^2 test and relative composition frequency variability (RCFV, Zhong et al., 2011). A clear advantage of the χ^2 test over RCFV is that the first provides a statistical framework to accept or reject homogeneity as a null hypothesis. A known limitation of the χ^2 test in assessing compositional bias in phylogenetics is that it incurs a higher type II error (deeming partitions homogeneous when they are not). The error arises due to the use of a null distribution that ignores phylogenetic relatedness (Foster and Hickey 1999; Foster 2004; Cox et al. 2014). On the other hand, the distribution of RCFV is used as a measure of the variability in composition, but it does not provide critical values. In practice, values of RCFV are used as relative guidelines to identify genes or partitions with greater or lower values of this index (Struck et al. 2014; Andrade et al. 2015; Struck et al. 2015; Whelan et al. 2015; Fernández et al. 2016, 2017). In our data sets, the use of loci with low or high RCFV had no effect on the placement of horseshoe crabs (Supplementary Fig. S17 available on Dryad), while those that failed χ^2 showed a few but obvious anomalous groupings: polyphyletic Solifugae, Pseudoscorpiones and the myriapod *Glomeris pustulata* as sister group to Xiphosura (Supplementary Fig. S19 available on Dryad).

Phylogenetic Information and Unbiased Sampling of Loci

As previous phylogenomic analyses have shown, the information content of individual loci, as revealed by the likelihood mapping, lacks the power to resolve all bipartitions for a given species tree (Salichos and Rokas 2013). It is only with the combined information of several loci that a general pattern emerges, either from the branching pattern found in individual gene trees or the combined analyses of sequence data.

The use of ΔGLS is attractive for its simplicity and intuitive interpretation. Nevertheless, its application is complicated when, as in our case, the problem cannot be reduced to only two alternative hypotheses or it involves unstable taxa. The likelihood scores of the competing trees may be swayed by unstable groupings, beyond the bipartition of interest. The use of constrained gene trees for these comparisons, where only one bipartition is enforced for the likelihood optimization instead of the whole species tree, is a possible alternative to untangle the fitness of the locus regarding only the partition of interest.

The influence of the model used for calculating the gene likelihood scores also requires further scrutiny; under a common model for all loci ($LG+\Gamma$) there are risks of model misspecification, while the use of best-fitting models could favor ones with more parameters. In our case, the differences caused by using one or the other model strategy were minimal, showing similar numbers of loci favoring each topology, but it is not clear if this result would hold in other cases.

The result from the ΔGLS analyses was also useful in explaining differences observed in the placement of Xiphosura. In general, data sets with higher proportions of loci favoring T_2 recovered the horseshoe crabs in a more basal position. Conversely, those enriched with loci favoring T_1 favored a derived placement of the group. In the case of the compact matrix, 42% of the loci favored the topology T_2, resulting in a higher proportion of loci in favor of the alternative hypothesis compared with the unfiltered data set (with 27% in support of T_2). The point is demonstrated to its extreme when analyzing only loci favoring T_2, resulting in a tree congruent with the monophyly of Arachnida. The question of which one is correct cannot be answered *a priori*.

Unlike cases characterized by overall lack of signal, such as the Neoaves example described by Shen et al. (2017), where the species tree topology is swayed by a handful of outlier loci with strong signal, our data set's favor for T_1 is found in the majority of loci and not attributable to outlier scores. Those in favor of T_2, however, cannot be discarded as *a priori* flawed in terms of compositional heterogeneity, species composition, or other obvious anomalies. At the moment, the species tree from the concatenated data set reflects the hypothesis favored by the majority of the loci, in spite of the proportion of loci in support of alternative solutions. If the proportion of these divided signals is natural, sampling of loci should aim to capture the natural distribution of these gene histories, instead of cherry-picking the loci that support preconceived hypothesis.

As a proof of concept, we identified loci favoring a species constraining the monophyly of Xiphosura + Pancrustacea, reflecting the archaic systematic notion that Xiphosura are allied to crustaceans (hence, horseshoe "crabs"). The resulting ML tree showed zero branch length for this implausible constraint, suggesting no substitutions supported that grouping. The constraint disrupted the mutual monophyly of Mandibulata and Chelicerata, resulting in the transposition of Pancrustacea into chelicerates. Nevertheless, the search for loci with ΔGLS favoring this anomalous tree found 662 genes; after concatenation, this matrix produced a ML tree mirroring the anomalous groupings seen in the constrained tree, with high bootstrap values. This thought experiment underscores that special care must be put to avoid choosing or discarding loci on the basis of an arbitrarily favored phylogenetic hypothesis (Fig. 8).

The Effect of Anomalous Gene Trees

The effects of AGT have been thoroughly studied on theoretical grounds (Kubatko and Degnan 2007; Degnan et al. 2012; Degnan 2013). Nevertheless, analytical

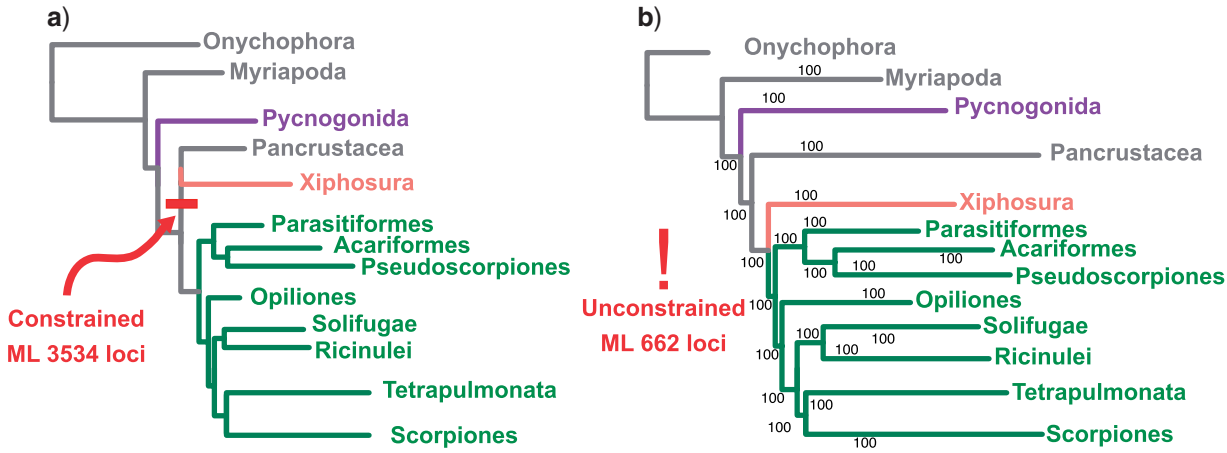

FIGURE 8. Cherry-picking loci will produce well-supported anomalous topologies. a) ML tree from 3534 loci partitioned by gene and constrained to recover the clade Xiphosura + Pancrustacea. b) Analysis of 662 loci with ΔGLS score in favor of the tree in (a) mirrors the anomalous groupings.

solution of gene trees probabilities becomes prohibitive for problems with more than nine species (Rosenberg and Tao 2008; Rosenberg 2013). In these cases simulation of gene trees under various coalescent conditions can provide reasonable approximations of the effects of stochastic coalescence of gene copies (Edwards et al. 2016).

The goal of our simulation was to generate increasing levels of gene tree incongruence to the point where the induced conflict would affect the estimation of the "true" species tree with ASTRAL. The values of *Ne* implemented were chosen arbitrarily but could be transformed into biologically meaningful units. The simulation implementation assumes branch lengths are provided in coalescent time units (CU), where $1CU = 2N \times t_g$ and t_g is the generation time. Therefore, if we equate 1 million years in the trees to 1 CU, the minimum and maximum *Ne* values used in our simulations (5 and 100,000) translate to population sizes of 2.5×10^6 and 50×10^9 individuals, respectively, assuming generation times of one year. For comparison, scaling of coalescent units to millions of years based on empirical estimates of *Ne* and generation times for a mammal species tree resulted in CU values of 0.8 to 2.75 my (Edwards et al. 2016). A few studies provide similar estimates of *Ne* and t_g for chelicerates. For example, Settepani et al. (2017) reports $Ne = 594 \times 10^3$ for the social spider *Stegodyphus mimosarum* and Lynch (2005) reports $Ne = 10^6$ for invertebrates, while generation times range from a few days in some mites (7.5 days for *T. urticae*, Shih et al., 1976) to several years (13–14 for *Limulus polyphemus*, Sweka et al., 2007). Evidently, the simulation imposed biologically unrealistic assumptions such as panmixia and constant population sizes across taxa and the evolutionary time scale. Nevertheless, the range of values used in the simulations clearly exceed plausible values of *Ne* or t_g in both directions, providing worst-case scenarios for the effects of gene tree incongruence.

TABLE 3. Congruence levels (ASTRAL normalized quartet score, nqs) of simulated gene trees under various population size (*Ne*) values

Ne	T1	T2
5	0.972	0.980
10	0.935	0.964
50	0.760	0.770
100	0.684	0.721
1000	0.419	0.433
5000	0.352	0.356
10,000	0.343	0.345
100,000	0.336	0.336

Note: Values closer to one indicate higher agreement between the simulated gene trees. Scores are shown for the gene trees collections simulated under unconstrained (T1) and constrained (T_2) template trees. For comparison, the average quartet score of the empirical T_1 ASTRAL tree is 0.75.

Species trees estimated from genes trees simulated under the assumption of arachnid monophyly recovered this clade even under relatively elevated levels of incongruence. Only under very high levels of ILS ($Ne >$ 5000) did ASTRAL fail to recover the monophyly of Arachnida. This could be interpreted as support for the plausibility of ILS as the cause of the derived placement of Xiphosura. However, the amount of gene tree disagreement under these conditions is more severe ($nqs \leq 0.352$) than that observed in the empirical data set (nqs=0.75, Table 3); the additional topological anomalies seen in those trees (e.g., the non-monopyly of Opiliones (harvestmen), Pycnogonida (sea spiders), Pancrustacea, and Myriapoda) additionally disfavor the plausibility of the parameter values at this range of the simulations.

Although we did not simulate sequences on the gene trees and their effect on concatenated analysis, the agreement between ASTRAL and concatenated results, as well as the magnitude of gene genealogy discordance, as indicated by the normalized quartet score, suggested that the placement of Xiphosura is not clearly attributable to ILS.

Our analyses corroborate previous results suggesting that the effects of ILS in empirical data sets is negligible

at these time scales (Gadagkar et al. 2005; Maddison and Knowles 2006; Huang and Knowles 2009; Tonini et al. 2015). At the same time, summary methods, such as ASTRAL, that account for coalescent gene tree discordance provided a prompt alternative to assess and ameliorate the potential negative effects of conflicting gene histories.

The species trees estimated using ASTRAL and ML were largely congruent under the variety of analytical conditions herein explored. In cases where differences were observed, these usually involved low supported relationships or unstable taxa. Although the impact of ILS seems limited in our data set, the ability of ASTRAL to overcome incongruence may prove useful to reconcile conflicting signal from sources other than stochastic coalescence of genes (Liu et al. 2014; Edwards et al. 2016). A formal comparison of the performance of these methods is beyond the scope of this contribution and further studies will help clarify the utility and limitations of summary coalescent methods for addressing deep phylogenetic questions (Lanier and Knowles 2015; Li et al. 2017).

The Presumption of Arachnid Monophyly

Skepticism for the non-monophyly of traditional groupings calls for the reexamination of morphological data. Upon reexamining the evidentiary body as it relates to the present case, we observe that the monophyly of Arachnida is infirm even in morphological data sets. An earlier generation of cladistic works, which heavily influenced later interpretations of arachnid phylogeny, did not actually test arachnid monophyly, as Xiphosura was used as the sole outgroup to root the arachnids, which were presumed monophyletic (Shultz 1990, 2007). Few morphological works have fully sampled all of the chelicerate orders, and these tend to recover markedly different tree topologies from those of Shultz (2007). A recent analysis of chelicerate relationships based on morphological data from fossils and extant taxa (Garwood et al. 2016) recovered a polyphyletic Xiphosura at the base of the chelicerate tree, with Pycnogonida as sister group to Arachnida; only when fossils were removed was the traditional grouping of (Pycnogonida + (Xiphosura + Arachnida)) recovered, albeit with dismally low levels of nodal support (0–5%). In this analysis, only two synapomorphies for Arachnida were established: (1) the loss of the first opisthosomal ("abdominal") appendages in postembryonic stages and (2) a tibial origin of the apotele depressor. At the same time, not all morphological studies support the monophyly of Arachnida. For example, the analysis of Strausfeld et al. (2006), based on 10 neuro-anatomical characters, found *Limulus polyphemus* nested within Arachnida, as sister group to a scorpion (*Centruroides*). There are in fact few anatomical characters outright supporting the monophyly of terrestrial arachnids.

These proposed synapomorphies of arachnids have also come under question as putative parallel adaptations to the terrestrial environment. As an analogous case, the phylogeny of Mandibulata (the non-chelicerate arthropods) was long held by the majority of morphologists to consist of (Crustacea + (Myriapoda + Hexapoda)). The terrestrial groups, Myriapoda and Hexapoda, share a surprising number of morphological characteristics, such as uniramous appendages, a gnathobasic mandible, a reduced and appendage-less third head segment, and a respiratory system composed of tracheal tubules that (generally) open into the body wall as pairs of spiracles on trunk segments. Apropos, they were placed together in the rank Tracheata (alternatively called Atelocerata). Molecular phylogenetics overturned this relationship, supporting instead the nested placement of Hexapoda within a paraphyletic Crustacea. While *post hoc* interpretations of morphology and fossils brought morphology into alignment with molecular phylogenetics over a decade later (Legg et al. 2013), this tree topology implies independent terrestrialization events in common ancestors of hexapods and myriapods, and demonstrates that a suite of morphological characters has exhibited remarkable convergence during the transition to a terrestrial habitat in at least one empirical case. Given that Chelicerata and Mandibulata are comparably old clades harboring multiple aquatic and terrestrial lineages, the evolutionary history of the mandibulates forewarns of the possibility of confounding and comparable morphological convergence in the terrestrial chelicerate orders.

The value of a new phylogenetic hypothesis extends far beyond the branching diagram. Alternative phylogenetic hypotheses provide researchers with an evolutionary framework to contextualize comparative biological data and test assumptions about traditionally held homologies. Once a phylogenetic hypothesis is under consideration, congruent patterns in morphological data sometimes emerge. As an example, the traditional position of scorpions close to the base of the arachnid tree (Weygoldt and Paulus 1979; Shultz 2007) is contrary to the phylogenomic placement of scorpions as the sister group of the tetrapulmonates (Regier et al., 2010; Sharma et al., 2014a, this study). This phylogenomic placement is finding support in recent studies of comparative morphology. Remarkable similarities occur in the vascular systems of scorpions, Uropygi and Amblypygi (Klußmann-Fricke and Wirkner 2016), as well as in the genome content of spiders and scorpions (to the exclusion of non-arachnopulmonate orders Leite et al., 2016; Schwager et al., 2017; Leite et al., 2018). In the same manner, the placement of Xiphosura within Arachnida invites a re-evaluation and exploration of diverse character systems and homology schema. Examples of similarities noted between horseshoe crabs, scorpions, and tetrapulmonates (to date, interpreted as symplesiomorphies) include the presence of the hemolymph vascular system (Göpel and Wirkner

2015), presence of hemocyanins (Rehm et al. 2012), the neuroanatomy of the pre-stomodeal commissure of the cheliceral ganglion (Mittmann and Scholtz 2003), and the presence of cuticular fluorescence under ultraviolet light observed in Xiphosura, Scorpiones, and Eurypterida (Rubin et al. 2017).

Conclusions

Our data set, similarly to previous analyses, shows uncertainty in the placement of several ordinal taxa. Nevertheless, it should be emphasized that this uncertainty does not involve the position of horseshoe crabs. The placement of Xiphosura within Arachnida, found in several independent data sets and supported under a variety of analytical conditions, must be interpreted as a potentially accurate estimate of the evolutionary history of horseshoe crabs. In spite of uncertainty in some parts of the chelicerate tree, the placement of horseshoe crabs does not seem to be caused by compositional bias, missing data, evolutionary rates, lack of signal, method of phylogenetic inference, LBA, or the stochastic effects of AGT due to ILS. In the light of the amount of evidence and no clear signs of phylogenetic error, the hypothesis of monophyletic Arachnida, and with it, the scenario of a single and irreversible colonization of the land by an arachnid ancestor, has become untenable.

"That the King crab [horseshoe crab] *is as closely related to the Scorpion as is the Spider has for years been an open secret, which has escaped notice by something like fatality."* (Ray Lankester, 1881)

Supplementary Material

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.2g1f4n5.

Funding

This work was supported by the National Science Foundation [IOS-1552610].

Acknowledgments

We are grateful to The Center for High Throughput Computing (CHTC) and Bioinformatics Resource Center (BRC) of the University of Wisconsin for facilitating computing resources. J.A.B. was supported by the University of Wisconsin-Madison's Department of Integrative Biology through a M. Guyer postdoctoral fellowship. Carlos E. Santibañez López, Guilherme Gainett, and members of the Sharma lab provided helpful comments and suggestions on early versions of this work. A previous draft of the manuscript was greatly improved by comments from the Associate Editor and two anonymous reviewers.

References

Anderson F.E., Swofford D.L. 2004. Should we be worried about long-branch attraction in real data sets? investigations using metazoan 18s rDN. Mol. Phylogenet. Evol. 33:440–451.
Andrade S.C., Novo M. Kawauchi G.Y., Worsaae K., Pleijel F., Giribet G., Rouse G.W. 2015. Articulating "archiannelids": phylogenomics and annelid relationships, with emphasis on meiofaunal taxa. Mol. Biol. Evol. 32:2860–2875.
Arabi J., Judson M.L., Deharveng L., Lourenço W.R., Cruaud C., Hassanin A. 2012. Nucleotide composition of CO1 sequences in Chelicerata (Arthropoda): detecting new mitogenomic rearrangements. J. Mol. Evol. 74:81–95.
Ballesteros J.A., Hormiga G. 2016. A new orthology assessment method for phylogenomic data: unrooted phylogenetic orthology. Mol. Biol. Evol. 33:2117–2134.
Bergsten J. 2005. A review of long-branch attraction. Cladistics. 21: 163–193.
Blanchard E. 1877. Métamorphoses moeurs et instincts des insectes: insectes, myriapodes, arachnides, crustacés. Paris: G. Baillière.
Borner J., Rehm P., Schill R.O., Ebersberger I., Burmester T. 2014. A transcriptome approach to ecdysozoan phylogeny. Mol. Phylogenet. Evol. 80:79–87.
Borowiec M.L., Lee E.K., Chiu J.C., Plachetzki D.C. 2015. Extracting phylogenetic signal and accounting for bias in whole-genome data sets supports the Ctenophora as sister to remaining Metazoa. BMC Genomics. 16:987.
Bouckaert R.R. 2010. Densitree: making sense of sets of phylogenetic trees. Bioinformatics. 26:1372–1373.
Braddy S.J., Aldridge R.J., Gabbott S.E., Theron J.N. 1999. Lamellate book-gills in a late Ordovician eurypterid from the soom shale, South Africa: support for a eurypterid-scorpion clade. Lethaia. 32:72–74.
Briggs D.E.G., Siveter D.J., Siveter D.J., Sutton M.D., Garwood R.J., Legg D. 2012. Silurian horseshoe crab illuminates the evolution of arthropod limbs. Proc. Natl. Acad. Sci USA. 109:15702–15705.
Brinkmann H., Van der Giezen M., Zhou Y., De Raucourt G.P., Philippe H. 2005. An empirical assessment of long-branch attraction artefacts in deep eukaryotic phylogenomics. Syst. Biol. 54: 743–757.
Bryant D., Bouckaert R., Felsenstein J., Rosenberg N.A., RoyChoudhury A. 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol. Biol. Evol. 29:1917–1932.
Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., Madden T.L. 2009. BLAST+: architecture and applications. BMC Bioinformatics. 10:421.
Capella-Gutiérrez S., Silla-Martínez J.M., Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 25:1972–1973.
Chernomor O., von Haeseler A., Minh B.Q. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. Syst. Biol. 65:997–1008.
Chifman J., Kubatko L. 2014. Quartet inference from snp data under the coalescent model. Bioinformatics. 30:3317–3324.
Clarke T.H., Garb J.E., Hayashi C.Y., Arensburger P., Ayoub N.A. 2015. Spider transcriptomes identify ancient large-scale gene duplication event potentially important in silk gland evolution. Genome Biol. Evol. 7:1856–1870.
Claus C. 1884. Elementary text-book of zoology, Vol. 1. New York: Macmillan.
Clouse R.M., Branstetter M.G., Buenavente P., Crowley L.M., Czekanski-Moir J., General D.E.M., Giribet G., Harvey M.S., Janies D.A., Mohagan A.B., Mohagan, D.P., Sharma, P.P. Wheeler W.C. 2017. First global molecular phylogeny and biogeographical analysis of two arachnid orders (Schizomida and Uropygi) supports a tropical pangean origin and mid-cretaceous diversification. J. Biogeogr. 44:2660–2672.
Comstock J.H. 1912. The spider book: a manual for the study of the spiders and their near relatives, the scorpions, pseudoscorpions, whip-scorpions, harvestmen, and other members of the class arachnida, found in America North of Mexico, with analytical keys

for their classification and popular accounts of their habits. Garden City: Doubleday, Page & Company.
Cox C.J., Li B., Foster P.G., Embley T.M., Civáň P. 2014. Conflicting phylogenies for early land plants are caused by composition biases among synonymous substitutions. Syst. Biol. 63:272–279.
Dabert M., Witalinski W., Kazmierski A., Olszanowski Z., Dabert J. 2010. Molecular phylogeny of acariform mites (Acari, Arachnida): strong conflict between phylogenetic signal and long-branch attraction artifacts. Mol. Phylogenet. Evol. 56:222–241.
Degnan J.H. 2013. Anomalous unrooted gene trees. Syst. Biol. 62:574–590.
Degnan J.H., Rosenberg N.A. 2006. Discordance of species trees with their most likely gene trees. PLoS Genet. 2:e68.
Degnan J.H., Rosenberg N.A. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24:332–340.
Degnan J.H., Rosenberg N.A., Stadler T. 2012. A characterization of the set of species trees that produce anomalous ranked gene trees. IEEE/ACM Trans. Comput. Biol. Bioinform. 9:1558–1568.
DellAmpio E., Meusemann K., Szucsich N.U., Peters R.S., Meyer B., Borner J., Petersen M., Aberer A.J., Stamatakis A., Walzl M.G., et al. 2013. Decisive data sets in phylogenomics: lessons from studies on the phylogenetic relationships of primarily wingless insects. Mol. Biol. Evol. 31:239–249.
Dongen S. 2000. A cluster algorithm for graphs. Technical Report. Amsterdam, The Netherlands: National Research Institute for Mathematics and Computer Science in the Netherlands.
Dunlop J.A. 1996. Evidence for a sister group relationship between Rininulei and Trigonotarbida. Bull. Br. Arachnol. Soc. 10: 193–204.
Dunlop J.A. 1997. The origins of tetrapulmonate book lungs and their significance for chelicerate phylogeny. In: Paul A. Selden, editor. Proceedings of the 17th European colloquium of arachnology, Edinburgh. Burnham Beeches, Bucks:British Arachnological Society. p. 9–16.
Dunlop J.A. 2010. Geological history and phylogeny of Chelicerata. Arthropod Struct. Dev. 39:124–142.
Dunlop J.A., Webster M. 1999. Fossil evidence, terrestrialization and arachnid phylogeny. J. Arachnol. 27:86–93.
Edwards S.V., Liu L., Pearl D.K. 2007. High-resolution species trees without concatenation. Proc. Natl. Acad. Sci. USA. 104:5936–5941.
Edwards S.V., Xi Z., Janke A., Faircloth B.C., McCormack J.E., Glenn T.C., Zhong B., Wu S., Lemmon E.M., Lemmon A.R., et al. 2016. Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. Mol. Phylogenet. Evol. 94:447–462.
Ewing H.E. 1914. The common red spider or spider mite, Vol. 121. Corvallis Oregon: Oregon Agricultural College Experiment Station.
Fernández R., Edgecombe G.D., Giribet G. 2016. Exploring phylogenetic relationships within Myriapoda and the effects of matrix composition and occupancy on phylogenomic reconstruction. Syst. Biol. 65:871–889.
Fernández R., Sharma P.P., Tourinho A.L., Giribet G. 2017. The Opiliones tree of life: shedding light on harvestmen relationships through transcriptomics. Proc. R. Soc. B. 284:20162340.
Firstman B. 1973. The relationship of the chelicerate arterial system to the evolution of the endosternite. J. Arachnol. 1:1–54.
Foster P.G. 2004. Modeling compositional heterogeneity. Syst. Biol. 53:485–495.
Foster P.G., Hickey D.A. 1999. Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. J. Mol. Evol. 48:284–290.
Gadagkar S.R., Rosenberg M.S., Kumar S. 2005. Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. J. Exp. Zool. B Mol. Dev. Evol. 304:64–74.
Garwood R.J., Dunlop J. 2016. Three-dimensional reconstruction and the phylogeny of extinct chelicerate orders. PeerJ. 2:e641.
Garwood R.J., Dunlop J.A. 2011. Morphology and systematics of Anthracomartidae (Arachnida: Trigonotarbida). Palaeontology. 54:145–161.
Garwood R.J., Dunlop J.A., Selden P.A., Spencer A.R., Atwood R.C., Vo N.T., Drakopoulos M. 2016. Almost a spider: a 305-million-year-old fossil arachnid and spider origins. Proc. R. Soc. B. 283: 20160125.
Giribet G., Edgecombe G.D., Wheeler W.C. 2001. Arthropod phylogeny based on eight molecular loci and morphology. Nature. 413:157.
Giribet G., Edgecombe G.D., Wheeler W.C., Babbitt C. 2002. Phylogeny and systematic position of opiliones: a combined analysis of chelicerate relationships using morphological and molecular data. Cladistics. 18:5–70.
Goloboff P.A., Farris J.S., Nixon K.C. 2008. TNT, a free program for phylogenetic analysis. Cladistics. 24:774–786.
Göpel T., Wirkner C.S. 2015. An ancient complexity? Evolutionary morphology of the circulatory system in Xiphosura. Zoology. 118:221–238.
Grasshoff M. 1978. A model of the evolution of the main chelicerate groups. In: P. Merret, editor. Arachnology, Vol. 42. p. 273–284. London: Academic Press.
Grbić M., Van Leeuwen T., Clark R.M. Rombauts S., Rouzé P., Grbić V., Osborne E.J., Dermauw W., Ngoc P.C.T., Ortego F., et al. 2011. The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. Nature. 479:487.
Haas B.J., Papanicolaou A., Yassour M., Grabherr M., Blood P.D., Bowden J., Couger M.B., 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.
Hoang D.T., Chernomor O., von Haeseler A., Minh B.Q., Vinh L.S. 2018. Ufboot2: Improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35:518–522.
Hoy M.A., Waterhouse R.M., Wu K., Estep A.S., Ioannidis P., Palmer W.J., Pomerantz A.F., SimÃ£o F.A., Thomas J., Jiggins F.M., Murphy T.D., Pritham E.J., Robertson H.M., Zdobnov E.M., Gibbs R.A., Richards S. 2016. Genome sequencing of the phytoseiid predatory mite *Metaseiulus occidentalis* reveals completely atomized Hox genes and superdynamic intron evolution. Genome Biol. Evol. 8:1762–1775.
Huang H., Knowles L.L. 2009. What is the danger of the anomaly zone for empirical phylogenetics? Syst. Biol. 58:527–536.
Jeffroy O., Brinkmann H., Delsuc F., Philippe H. 2006. Phylogenomics: the beginning of incongruence? Trends Genet. 22:225–231.
Kalyaanamoorthy S., Minh B.Q., Wong T.K.F., von Haeseler A., Jermiin L.S. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods. 14:587–589.
Kamenz C., Dunlop J.A., Scholtz G., Kerp H., Hass H. 2008. Microanatomy of Early Devonian book lungs. Biol. Lett. 4:212–215.
Katoh K., Standley D.M. 2013. Mafft multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30:772–780.
Kenny N.J., Chan K.W., Nong W., Qu Z., Maeso I., Yip H.Y., Chan T.F., Kwan H.S., Holland P.W.H., Chu K.H., Hui J.H.L. 2016. Ancestral whole-genome duplication in the marine chelicerate horseshoe crabs. Heredity. 116:190.
Kin A., Błażejowski B. 2014. The horseshoe crab of the genus *Limulus*: living fossil or stabilomorph? PLoS One. 9:e108036.
Klußmann-Fricke B.-J., Wirkner C. 2016. Comparative morphology of the hemolymph vascular system in Uropygi and Amblypygi (Arachnida): complex correspondences support Arachnopulmonata. J. Morphol. 277:1084–1103.
Koch C.L., Hahn C.W. 1841–1843. Die Arachniden: Getreu nach der Natur abgebildet und beschrieben, vol. bd. 8-10 (1841-1843) plates 253-360. Nurnberg:In der C. H. Zehschen Buchhandlung.
Kocot K.M., Cannon J.T., Todt C., Citarella M.R., Kohn A.B., Meyer A., Santos S.R., Schander C., Moroz L.L., Lieb B., et al. 2011. Phylogenomics reveals deep molluscan relationships. Nature. 477:452.
Kozlov A.M., Aberer A.J., Stamatakis A. 2015. Examl version 3: a tool for phylogenomic analyses on supercomputers. Bioinformatics. 31:2577–2579.
Kubatko L.S., Degnan J.H. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56:17–24.

Kück P., Struck T.H. 2014. BaCoCa heuristic software tool for the parallel assessment of sequence biases in hundreds of gene and taxon partitions. Mol. Phylogenet. Evol. 70:94–98.
Lamsdell J.C. 2013. Revised systematics of Palaeozoic "horseshoe crabs" and the myth of monophyletic Xiphosura. Zool. J. Linnean Soc. 167:1–27.
Lamsdell J.C. 2016. Horseshoe crab phylogeny and independent colonizations of fresh water: ecological invasion as a driver for morphological innovation. Palaeontology. 59:181–194.
Lanier H.C., Knowles L.L. 2015. Applying species-tree analyses to deep phylogenetic histories: challenges and potential suggested from a survey of empirical phylogenetic studies. Mol. Phylogenet. Evol. 83:191–199.
Lankester E.R. 1881. *Limulus* an arachnid. Q. J. Microsc. Sci. s2–21:504–548.
Lartillot N., Brinkmann H., Philippe H. 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. BMC Evol. Biol. 7:S4.
Laumer, Christopher E., Harald Gruber-Vodicka, Michael G. Hadfield, Vicki B. Pearse, Ana Riesgo, John C. Marioni, and Gonzalo Giribet. "Support for a clade of Placozoa and Cnidaria in genes with minimal compositional bias." *Elife* 7 (2018): e36278.
Legg D.A., Sutton M.D., Edgecombe G.D. 2013. Arthropod fossil data increase congruence of morphological and molecular phylogenies. Nat. Commun. 4:ncomms3485.
Leite D.J., Baudouin-Gonzalez L., Iwasaki-Yokozawa S., Lozano-Fernandez J., Turetzek N., Akiyama-Oda Y., Prpic N.-M., Pisani D., Oda H., Sharma P.P., et al. 2018. Homeobox gene duplication and divergence in arachnids. Mol. Biol. Evol.
Leite D.J., Ninova M., Hilbrant M., Arif S., Griffiths-Jones S., Ronshaugen M., McGregor A.P. 2016. Pervasive microRNA duplication in chelicerates: insights from the embryonic microRNA repertoire of the spider *Parasteatoda tepidariorum*. Genome Biol. Evol. 8:2133–2144.
Li Y., Kocot K.M., Whelan N.V., Santos S.R., Waits D.S., Thornhill, D.J., Halanych K.M. 2017. Phylogenomics of tubeworms (Siboglinidae, Annelida) and comparative performance of different reconstruction methods. Zool. Scr. 46:200–213.
Liu L., Xi Z., Davis C.C. 2014. Coalescent methods are robust to the simultaneous effects of long branches and incomplete lineage sorting. Mol. Biol. Evol. 32:791–805.
Liu L., Yu L., Edwards S.V. 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. BMC Evol. Biol. 10:302.
Liu L., Yu L., Pearl D.K., Edwards S.V. 2009. Estimating species phylogenies using coalescence times among sequences. Syst. Biol. 58:468–477.
Lynch M. 2005. The origins of eukaryotic gene structure. Mol. Biol. Evol. 23:450–468.
Lyons-Weiler J., Hoelzer G.A. 1997. Escaping from the Felsenstein zone by detecting long branches in phylogenetic data. Mol. Phylogenet. Evol. 8:375–384.
Maddison W.P. 1997. Gene trees in species trees. Syst. Biol. 46:523–536.
Maddison W.P., Knowles L.L. 2006. Inferring phylogeny despite incomplete lineage sorting. Syst. Biol. 55:21–30.
Mallatt J., Giribet G. 2006. Further use of nearly complete 28S and 18S rRNA genes to classify Ecdysozoa: 37 more arthropods and a kinorhynch. Mol. Phylogenet. Evol. 40:772–794.
Mallatt J.M., Garey J.R., Shultz J.W. 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 S.E., Longhorn S.J., Boore J.L. 2009. Arachnid relationships based on mitochondrial genomes: asymmetric nucleotide and amino acid bias affects phylogenetic analyses. Mol. Phylogenet. Evol. 50:117–128.
McGhee G.R., Sheehan P.M., Bottjer D.J., Droser M.L. 2012. Ecological ranking of Phanerozoic biodiversity crises: the Serpukhovian (early Carboniferous) crisis had a greater ecological impact than the end-Ordovician. Geology. 40:147–150.
Meusemann K., von Reumont B.M., Simon S., Roeding F., Strauss S., Kück P., Ebersberger I., Walzl M., Pass G., Breuers S., Achter V., von Haeseler A., Burmester T., Hadrys H., Wägele J.W., Misof B. 2010. A phylogenomic approach to resolve the arthropod tree of life. Mol. Biol. Evol. 27:2451–2464.
Meyer B., Meusemann K., Misof B. 2011. MARE: MAtrix REduction a tool to select optimized data subsets from supermatrices for phylogenetic inference. Bonn (Germany): Zentrum für molekulare Biodiversitätsforschung (zmb) am ZFMK .
Mirarab S., Warnow T. 2015. Astral-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics. 31:i44–i52.
Mittmann B., Scholtz G. 2003. Development of the nervous system in the "head" of *Limulus polyphemus* (Chelicerata: Xiphosura): morphological evidence for a correspondence between the segments of the chelicerae and of the (first) antennae of Mandibulata. Dev. Genes Evol. 213:9–17.
Möbius K.A. 1902. Die Pantopoden der deutschen Tiefsee-Expedition 1898-1899, Vol. 3. Jena: G. Fischer.
Moroz L.L., Kocot K.M., Citarella M.R., Dosung S., Norekian T.P., Povolotskaya I.S., Grigorenko A.P., Dailey C., Berezikov E., Buckley K.M., Ptitsyn A., Reshetov D., Mukherjee K., Moroz T.P., Bobkova Y., Yu F., Kapitonov V.V., Jurka J., Bobkov Y.V., Swore J.J., Girardo D.O., Fodor A., Gusev F., Sanford R., Bruders R., Kittler E., Mills C.E., Rast J.P., Derelle R., Solovyev V.V., Kondrashov F.A., Swalla B.J., Sweedler J.V., Rogaev E.I., Halanych K.M., Kohn A.B. 2014. The ctenophore genome and the evolutionary origins of neural systems. Nature. 510:109.
Murienne J., Harvey M.S., Giribet G. 2008. First molecular phylogeny of the major clades of Pseudoscorpiones (Arthropoda: Chelicerata). Mol. Phylogenet. Evol. 49:170–184.
Nguyen L.-T., Schmidt H.A., von Haeseler A., Minh B.Q. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32:268–274.
Paradis E. 2013. Molecular dating of phylogenies by likelihood methods: a comparison of models and a new information criterion. Mol. Phylogenet. Evol. 67:436–444.
Paradis E., Claude J., Strimmer K. 2004. APE: analyses of phylogenetics and evolution in r language. Bioinformatics. 20:289–290.
Pepato A.R., da Rocha C.E., Dunlop J.A. 2010. Phylogenetic position of the acariform mites: sensitivity to homology assessment under total evidence. BMC Evol. Biol. 10:235.
Philippe H., Brinkmann H., Lavrov D.V., Littlewood D.T.J., Manuel M. Wörheide G., Baurain D. 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol. 9:e1000602.
Philippe H., Roure B. 2011. Difficult phylogenetic questions: more data, maybe; better methods, certainly. BMC Biol. 9:91.
Philippe H., Zhou Y., Brinkmann H., Rodrigue N., Delsuc F. 2005. Heterotachy and long-branch attraction in phylogenetics. BMC Evol. Biol. 5:50.
Phillips M.J., Delsuc F., Penny D. 2004. Genome-scale phylogeny and the detection of systematic biases. Mol. Biol. Evol. 21:1455–1458.
Pisani D., Pett W., Dohrmann M., Feuda R., Rota-Stabelli O., Philippe H., Lartillot N., Wörheide G. 2015. Genomic data do not support comb jellies as the sister group to all other animals. Proc. Natl. Acad. Sci. USA. 112:15402–15407.
Qu X.-J., Jin J.-J., Chaw S.-M., Li D.-Z., Yi T.-S. 2017. Multiple measures could alleviate long-branch attraction in phylogenomic reconstruction of Cupressoideae (Cupressaceae). Sci. Rep. 7: 41005.
Quinlan A.R., Hall I.M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26:841–842.
R Core Team. 2017. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
Regier J.C., Shultz J.W., Zwick A., Hussey A., Ball B., Wetzer R., Martin J.W., Cunningham C.W. 2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. Nature. 463:1079.
Regier J.C., Zwick A. 2011. Sources of signal in 62 protein-coding nuclear genes for higher-level phylogenetics of arthropods. PLoS One. 6:e23408.
Rehm P., Meusemann K., Borner J., Misof B., Burmester T. 2014. Phylogenetic position of Myriapoda revealed by 454 transcriptome sequencing. Mol. Phylogenet. Evol. 77:25–33.

Rehm P., Pick C., Borner J., Markl J., Burmester T. 2012. The diversity and evolution of chelicerate hemocyanins. BMC Evol. Biol. 12: 19.
Rodríguez-Ezpeleta N., Brinkmann H., Roure B., Lartillot N., Lang B.F., Philippe H. 2007. Detecting and overcoming systematic errors in genome-scale phylogenies. Syst. Biol. 56:389–399.
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 S.B. 2006. Bushes in the tree of life. PLoS Biol. 4:e352.
Rosenberg M.S., Kumar S. 2003. Taxon sampling, bioinformatics, and phylogenomics. Syst. Biol. 52:119.
Rosenberg N.A. 2013. Discordance of species trees with their most likely gene trees: a unifying principle. Mol. Biol. Evol. 30:2709–2713.
Rosenberg N.A., Tao R. 2008. Discordance of species trees with their most likely gene trees: the case of five taxa. Syst. Biol. 57: 131–140.
Rota-Stabelli O., Lartillot N., Philippe H., Pisani D. 2012. Serine codon-usage bias in deep phylogenomics: pancrustacean relationships as a case study. Syst. Biol. 62:121–133.
Rubin M., Lamsdell J., Prendini L., Hopkins M. 2017. Exocuticular hyaline layer of sea scorpions and horseshoe crabs suggests cuticular fluorescence is plesiomorphic in chelicerates. J. Zool. 303: 245–253.
Rudkin D.M., Young G.A., Nowlan G.S. 2008. The oldest horseshoe crab: a new xiphosurid from Late Ordovician Konservat-Lagerstätten deposits. Manitoba, Canada. Palaeontology. 51:1–9.
Salichos L., Rokas A. 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. Nature. 497:327.
Sanders K.L., Lee M.S.Y. 2010. Arthropod molecular divergence times and the Cambrian origin of pentastomids. Syst. Biodivers. 8:63–74.
Santibáñez López C.E., Ontano A.Z., Harvey M.S., Sharma P.P. 2018. Transcriptomic analysis of pseudoscorpion venom reveals a unique cocktail dominated by enzymes and protease inhibitors. Toxins. 10:207.
Scholtz G., Kamenz C. 2006. The book lungs of Scorpiones and Tetrapulmonata (Chelicerata, Arachnida): evidence for homology and a single terrestrialisation event of a common arachnid ancestor. Zoology. 109:2–13.
Schwager E.E., Sharma P.P., Clarke T., Leite D.J., Wierschin T., Pechmann M., Oda Akiyama Y., Esposito L., Bechsgaard J., Bilde T., et al. 2017. The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. BMC Biol. 15:62.
Settepani V., F Schou M., Greve M., Grinsted L., Bechsgaard J., Bilde T. 2017. Evolution of sociality in spiders leads to depleted genomic diversity at both population and species level. Mol. Ecol. 16:4197–4210.
Sharma P.P. 2017. Chelicerates and the conquest of land: a view of arachnid origins through an evo-devo spyglass. Integr. Comp. Biol. 57:510–522.
Sharma P.P., Kaluziak S.T., Pérez-Porro A.R., González V.L., Hormiga G., Wheeler W.C., Giribet G. 2014a. Phylogenomic interrogation of Arachnida reveals systemic conflicts in phylogenetic signal. Mol. Biol. Evol. 31:2963–2984.
Sharma P.P., Santiago M.A., González-Santillán E., Monod L., Wheeler W.C. 2015. Evidence of duplicated Hox genes in the most recent common ancestor of extant scorpions. Evol. Dev. 17:347–355.
Sharma P.P., Schwager E.E., Extavour C.G., Wheeler W.C. 2014b. Hox gene duplications correlate with posterior heteronomy in scorpions. Proc. R. Soc. Lond. B Biol. Sci. 281:20140661.
Shen X.X., Hittinger C.T., Rokas A. 2017. Contentious relationships in phylogenomic studies can be driven by a handful of genes. Nat. Ecol. Evol. 1:126.
Shih C.-i.T., Poe S.L., Cromroy H.L. 1976. Biology, life table, and intrinsic rate of increase of *Tetranychus urticae*. Ann. Entomol. Soc. Am. 69:362–364.
Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51:492–508.
Shultz J.W. 1990. Evolutionary morphology and phylogeny of Arachnida. Cladistics. 6:1–38.
Shultz J.W. 2007. A phylogenetic analysis of the arachnid orders based on morphological characters. Zool. J. Linnean Soc. 150:221–265.
Simion P., Philippe H., Baurain D., Jager M., Richter D., Di Franco A., Roure B, Satoh N., Quéinnec E., Ereskovsky A., Lapébie P., Corre E., Delsuc F., King N., Wörheide G., Manuel M. 2017. A large and consistent phylogenomic dataset supports sponges as the sister group to all other animals. Curr. Biol. 27:958–967.
Snodgrass R.E. 1938. Evolution of the Annelida, Onychophora and Arthropoda, Smithsonian Miscellaneous Collections. 57:1–159.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30:1312–1313.
Starrett J., Derkarabetian S., Hedin M., Bryson R.W., McCormack J.E., Faircloth B.C. 2017. High phylogenetic utility of an ultraconserved element probe set designed for Arachnida. Mol. Ecol. Resour. 17:812–823.
Steel M., Sanderson M.J. 2010. Characterizing phylogenetically decisive taxon coverage. Appl. Math. Lett. 23:82–86.
Strausfeld N.J., Strausfeld C.M., Loesel R., Rowell D., Stowe S. 2006. Arthropod phylogeny: Onychophoran brain organization suggests an archaic relationship with a chelicerate stem lineage. Proc. R. Soc. Lond. B Biol. Sci. 273:1857–1866.
Strimmer K., Von Haeseler A. 1997. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. Proc. Natl. Acad. Sci. USA. 94:6815–6819.
Struck T.H. 2014. Trespex—detection of misleading signal in phylogenetic reconstructions based on tree information. Evol. Bioinform. 10:EBO–S14239.
Struck T.H., Golombek A., Weigert A., Franke F.A., Westheide W., Purschke G., Bleidorn C., Halanych K.M. 2015. The evolution of annelids reveals two adaptive routes to the interstitial realm. Curr. Biol. 25:1993–1999.
Struck T.H., Wey-Fabrizius A.R., Golombek A., Hering L., Weigert A., Bleidorn C., Klebow S., Iakovenko N., Hausdorf B., Petersen M., et al. 2014. Platyzoan paraphyly based on phylogenomic data supports a noncoelomate ancestry of spiralia. Mol. Biol. Evol. 31:1833–1849.
Sukumaran J., Holder M.T. 2010. Dendropy: a python library for phylogenetic computing. Bioinformatics. 26:1569–1571.
Sweka J.A., Smith D.R., Millard M.J. 2007. An age-structured population model for horseshoe crabs in the delaware bay area to assess harvest and egg availability for shorebirds. Estuaries Coasts. 30:277–286.
Tonini J., Moore A., Stern D., Shcheglovitova M., Ortí G. 2015. Concatenation and species tree methods exhibit statistically indistinguishable accuracy under a range of simulated conditions. PLoS Curr. 7.
Van der Hammen L. 1977. A new classification of Chelicerata. Rijksmuseum van natuurlijke historie.
Van der Hammen L. 1985. Functional morphology and affinities of extant chelicerata in evolutionary perspective. Earth Environ. Sci. Trans. R. Soc. Edinb. 76:137–146.
Van Roy P., Orr P.J., Botting J.P., Muir L.A., Vinther J., Lefebvre B., El Hariri K., Briggs D.E. 2010. Ordovician faunas of Burgess Shale type. Nature. 465:215.
von Reumont B.M., Jenner R.A., Wills M.A., DellAmpio E., Pass G., Ebersberger I., Meyer B., Koenemann S., Iliffe T.M., Stamatakis A., et al. 2011. 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.
Waddington J., Rudkin D.M., Dunlop J.A. 2015. A new mid-Silurian aquatic scorpion—one step closer to land? Biol. Lett. 11:20140815.
Wang H., Minh B., Susko E., Roger A. 2018. Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. Syst. Biol. 67:216.
Weygoldt P., Paulus H. 1979. Untersuchungen zur Morphologie, Taxonomie und Phylogenie der chelicerata ii. Cladogramme und die Entfaltung der Chelicerata. J. Zool. Syst. Evol. Res. 17:177–200.
Wheeler W.C., Hayashi, C.Y. 1998. The phylogeny of the extant chelicerate orders. Cladistics. 14:173–192.
Whelan N.V., Halanych K.M. 2017. Who let the CAT out of the bag? Accurately dealing with substitutional heterogeneity in phylogenomic analyses. Syst. Biol. 66:232–255.
Whelan N.V., Kocot K.M., Moroz L.L., Halanych K.M. 2015. Error, signal, and the placement of Ctenophora sister to all other animals. Proc. Natl. Acad. Sci. USA. 112:5773–5778.

Whelan N.V., Kocot K.M., Moroz T.P., Mukherjee K., Williams P., Paulay G., Moroz L.L., Halanych K.M. 2017. Ctenophore relationships and their placement as the sister group to all other animals. Nat. Ecol. Evol. 1:1737.

Whitfield J.B., Lockhart P.J. 2007. Deciphering ancient rapid radiations. Trends Ecol. Evol. 22:258–265.

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

Wolfe J.M., Daley A.C., Legg D.A., Edgecombe G.D. 2016. Fossil calibrations for the arthropod tree of life. Earth Sci. Rev. 160: 43–110.

Zhang C., Sayyari E., Mirarab S. 2017. ASTRAL-III: increased scalability and impacts of contracting low support branches. In: RECOMB International Workshop on Comparative Genomics. Springer. p. 53–75.

Zhong M., Hansen B., Nesnidal M., Golombek A., Halanych K.M., Struck T.H. 2011. Detecting the symplesiomorphy trap: a multigene phylogenetic analysis of terebelliform annelids. BMC Evol. Biol. 11:369.

Zwick A., Regier J.C., Zwickl D.J. 2012. Resolving discrepancy between nucleotides and amino acids in deep-level arthropod phylogenomics: differentiating serine codons in 21-amino-acid models. PLoS One. 7:e47450.