# Converging on the orb: denser taxon sampling elucidates spider phylogeny and new analytical methods support repeated evolution of the orb web

Robert J. Kallal[1, 2*], Siddharth S. Kulkarni[1,2], Dimitar Dimitrov[3], Ligia R. Benavides[4], Miquel A. Arnedo[5], Gonzalo Giribet[4], Gustavo Hormiga[1*]

[1]Department of Biological Sciences, The George Washington University, 2029 G St. NW, Washington, D.C. 20052, USA

[2]Department of Entomology, National Museum of Natural History, 10th & Constitution Ave. NW, Washington, DC, 20560, USA

[3]Department of Natural History, University Museum of Bergen, University of Bergen, P.O. Box 7800, 5020 Bergen, Norway

[4]Museum of Comparative Zoology, Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA

[5]Department of Evolutionary Biology, Ecology and Environmental Sciences, & Biodiversity Research Institute (IRBio) Universitat de Barcelona, Avinguda Diagonal 643, Barcelona, Spain

*Correspondence to be sent to: Department of Biological Sciences, The George Washington University, 2029 G St. NW, Washington, D.C. 20052, USA; E-mail: hormiga@gwu.edu and/or Department of Entomology, National Museum of Natural History, 10th & Constitution Ave. NW, Washington, DC, 20560, USA; E-mail: kallalr@si.edu

Running title: Phylogenomics of Spiders

**Abstract**

High throughput sequencing and phylogenomic analyses focusing on relationships among spiders have both reinforced and upturned long-standing hypotheses. Similarly, the evolution of spider webs – perhaps their most emblematic attribute – is being understood in new ways. With a matrix including 272 spider species and close arachnid relatives, we analyze and evaluate the relationships among these lineages using a variety of orthology assessment methods, occupancy thresholds, tree inference methods, and support metrics. Our analyses include families not previously sampled in transcriptomic analyses, such as Symphytognathidae, the only araneoid family absent in prior such works. We find support for the major established spider lineages, including Mygalomorphae, Araneomorphae, Synspermiata, Palpimanoidea, Araneoidea, and the RTA Clade, as well as the UDOH Grade. Resulting trees are evaluated using bootstrapping, SH-aLRT, local posterior probabilities, and concordance factors. Using structured Markov models to assess the evolution of spider webs while accounting for hierarchically nested traits, we find multiple convergent occurrences of the orb web across the spider tree of life. Overall, we provide the most comprehensive spider tree of life to date using transcriptomic data and use new methods to explore controversial issues of web evolution, including the origins and multiple losses of the orb web.

Keywords: Araneae, concordance factors, maximum likelihood, penalized likelihood, structured Markov models

Table of contents

Introduction 4

Material and methods 10

- Extraction and transcriptome sequencing 10
- Orthology and matrix variations 11
- Tree inference 13
- Divergence dating 13
- Web evolution 14

Results 17

- Matrix composition 17
- Phylogenetic analyses 19
- Divergence dating 26
- Web evolution 28

Discussion 31

- Spider phylogeny 31
- Methodology and support values 34
- Web evolution 37

Conclusions 41

Funding 41

Acknowledgments 42

References 43

Figure and table caption 58

## Introduction

Spiders (Araneae) are omnipresent predators and comprise one of the most diverse animal orders outside Hexapoda (Zhang, 2011). Over the course of several hundred million years, spiders have evolved into myriad shapes and sizes, filling niches in virtually all terrestrial (and some aquatic) habitats, except in Antarctica. Currently, over 48,000 species are described among 120 families (World Spider Catalog, 2020), with three families described in 2019 alone (Hedin et al., 2019; Ramírez et al., 2019). One of the most iconic traits of spiders is the production of silk, along with its many uses. A handful of gland types secrete silk used for producing egg sacs, bonding to substrates, wrapping prey, ballooning, and prey interception and capture. Of these, orb webs deserve special mention. The typical orb web's architecture consists of a frame holding radii that support a spiral sticky thread, but there are many architectural variations of this basic layout. The typical web needs to absorb the energy of the intercepted prey and retain them long enough to give the spider time to locate and subdue them. Geometrically similar orb webs are constructed by the members of the superfamily Araneoidea and two cribellate families: Uloboridae and Deinopidae. Araneoidea includes approximately a quarter of described spider species among 17 families (World Spider Catalog, 2020) and has an appropriately impressive variation in web architectures. They include the characteristic orb webs (e.g., Araneidae, Tetragnathidae), as well as cob-webs (Theridiidae, Nesticidae) and sheet webs (e.g., Linyphiidae, Cyatholipidae), among others, such as those that have secondarily lost capture webs entirely (e.g., Mimetidae, Arkyidae). Most spiders, however, do not construct a web to intercept prey and

many make no foraging web whatsoever (Shear, 1986; Dimitrov et al., 2017; Fernández et al., 2018a, 2018b; Coddington et al., 2019).

The prevailing understanding of the phylogenetic relationships among spider species has changed over the past decades, sometimes gradually and sometimes drastically, but it is converging on a more stable, supported pattern. As data matrix sizes have increased – both in terms of terminals and loci – a number of novel relationships have become apparent. One of the most important of these regards the former cribellate orb-weaving superfamily Deinopoidea (Uloboridae + Deinopidae). Based on a number of data types and analyses, it is both no longer considered a monophyletic group nor closely related to Araneoidea, the ecribellate orb weavers, and thus refuting the Orbiculariae hypothesis which suggested that cribellate and ecribellate orb-weavers formed a lineage (Bond et al., 2014; Fernández et al., 2014; Dimitrov et al., 2017; Wheeler et al., 2017). While the stickiness of araneoid orb webs is achieved by the use of a unique type of visicid, gluey silk, cribellate orb webs rely on a different type of sticky silk which is made of thousands of fine looped nanofibrils, and its adhesive properties are attained by a combination of mechanical interlock with the prey cuticle, adhesion to insect cuticular wax via capillary, hygroscopic, and van der Waals forces (Opell, 2013; Bott et al., 2017). Recent analyses have suggested that cribellate orb-weavers are more closely related to members of the Retrolateral Tibial Apophysis (RTA) Clade (a diverse clade of largely cursorial and ambush spiders; most RTA members do not rely on webs to intercept prey) and the families Oecobiidae and Hersiliidae than to ecribellate orb-weavers (Bond et al., 2014; Garrison et al., 2016; Fernández et al., 2018a), as it was generally accepted until 2014 (e.g., Hormiga and Griswold, 2014). The grade of families subtending the RTA Clade – uloborids, deinopids, oecobiids, and

hersiliids, together termed the UDOH Grade – has been broadly supported but the relationships therein are inconsistent (Fernández et al., 2018a). Other changes in Mygalomorphae (a clade including tarantulas and their kin), Synspermiata (a clade of ecribellate spiders with simple genitalia including spitting spiders, cellar spiders, and others), and within both Araneoidea and the RTA Clade have led to a critical eye being cast on relationships both old and new (Kallal et al., 2018; Kuntner et al., 2019; Michalik et al., 2019; Opatova et al., in press).

As the datasets have grown from a handful to hundreds or thousands of loci with the adoption of modern sequencing methods, the spider tree of life is coming into sharper focus. However, the long legacy of Sanger sequencing data has resulted in a taxon sampling which dwarfs that of high throughput sequencing methods, even at the family level, thus far resulting in trees missing numerous branches needed to answer long-standing questions about spider biology. Over the past five years, both the overall number of taxa and taxon specificity (e.g., family-level) of analyses of spider interfamilial relationships have increased (e.g., Garrison et al., 2016; Cheng and Piel, 2018; Fernández et al., 2018a; Hedin et al., 2018; Kallal et al., 2018; Shao and Li, 2018; Wood et al., 2018; Hedin et al., 2019; Kuntner et al., 2019; Michalik et al., 2019; Kulkarni et al., 2019; Opatova et al., in press). The increasing consensus has led to more resolved and robust trees at various phylogenetic scales, making comparative questions interpretable in new ways.

One of the perennial questions in spider biology involves the evolution and diversification of webs (Fig. 1) and has been the subject of some recent debate as new phylogenies revised our understanding of spider relationships (e.g., Fernández et al., 2018a, 2018b; Coddington et al.,

2019). While all spiders spin silk, how silk is used varies considerably across the spider tree of life. The types of capture webs have been coded and analyzed in increasingly sophisticated ways although a consensus on how to code and analyze these complex and interconnected type of data is lacking (Blackledge et al., 2009; Dimitrov et al., 2012; Garrison et al., 2016; Dimitrov et al., 2017; Fernández et al., 2018a, 2018b; Coddington et al., 2019; Dimitrov and Hormiga, submitted). This is not least because of imperfect comparative methods and simplification of a complex suite of behaviors into states based on resemblance of the final silken structure. While specific data on morphological and behavioral homologies may be ideal, they are absent for the vast majority of spider lineages. It is important to note that coding web architecture comes with a major caveat: a web is the result of a series of integrated behaviors, some of which can be homologized across species (e.g., the radius construction behavior in an orb web; Eberhard, 1982) or not (e.g., what would be the homolog of the radius construction behavior in a sheet web?). Up to this date, comparative biology analyses have treated web architecture as a phylogenetic character, with web types analyzed as alternative states, an approach that has been referred as a quantum leap of the concept of homology (Dimitrov and Hormiga, submitted) because webs are not, in and of themselves, homologous to each other (e.g., Eberhard, 2018). Accepting the treatment of variation in overall web architecture as a character brings us to additional difficulties, such as how to exactly code and analyze such variation into states with the goal of reconstructing ancestral webs. New methods using nested hidden states and structured Markov models (Tarasov, 2019) may be an important new tool to simultaneously account for the hierarchical and hidden processes addressing the absence or presence of webs and their diverse forms.

Questions regarding the phylogeny of spiders remain, mainly in the areas of sampling and analysis, and these must be resolved before tangling ourselves into further discussions about the origins and evolution of complex structures, such as webs. Approximately one quarter of spider families remain unsampled to date for transcriptomes, including one araneoid family which was not included in previous phylotranscriptomic works focusing on orb-weavers (Fernández et al., 2018a). Topologically, insufficient taxon sampling can lead to spurious relationships, sometimes stemming from long branch attraction (LBA) and other common artifacts in phylogenetic

inference. Such a phylogenetic hypothesis might not only have specious clades but also confound downstream comparative analyses related to ancestral trait reconstruction and diversification rate estimation. In the case of the former, conflicting signal made placement of Uloboridae difficult in Fernández et al. (2014), where the UDOH Grade was represented by only two terminals. Another example lies in speciation rate analyses, which rely on branch lengths and tree density. Garrison et al. (2016) indicated that multiple lineages, including the RTA Clade, Avicularioidea, and Araneoidea had elevated diversification rates, whereas increased sampling by Fernández et al. (2018a) produced a more nuanced picture by including more taxa. Rather than a basal speciation increase subtending Araneoidea, the families Theridiidae, Tetragnathidae, Linyphiidae, and Araneidae are specifically suggested to have a higher diversification rate relative to the other araneoid families. For these reasons, increasing taxon sampling is a fundamental aim.

The second assortment of issues to be examined relates to phylogenetic methods. The range of approaches used to analyze phylogenomic data is vast, with variations both subtle and substantial often unique to specific taxa or working groups. It is beyond the scope of this work to outline the numerous types of data and how they can be analyzed. Typical methodological differences may include orthology assessment, matrix occupancy, alignment trimming, tree inference methods, and model selection, all of which can have an effect on topologies with variable resemblance to each other. For instance, trimming has been shown to have deleterious effects on single locus trees (Tan et al., 2015) but this effect is believed to be overwhelmed by weight of signal in concatenated analyses (Philippe et al., 2017). In addition, some of the methods to calculate node support may be ill-equipped for matrices of phylogenomic scale; that is, are traditional measures

of support, such as bootstrapping, actually telling us as much when hundreds of loci and thousands of sites are analyzed? Some analyses suggest not (Kumar et al., 2012), and alternatives using concordance factors based on loci (Gadagkar et al., 2005; Ane et al., 2007) and sites (Minh et al., 2020) have been proposed.

Here, we infer a new phylogenetic hypothesis for Araneae with special emphasis on the ecribellate orb-weavers (Araneoidea), based on several hundred loci generated from transcriptome (RNA-Seq) data. Our taxon sampling greatly expands on previous works in an effort to understand more of the relationships among the main spider lineages. Analyses are conducted with varying orthology assessment methods, matrix occupancy, trimming, and tree inference approaches in order to explore their impact on the final hypothesis. We use the resulting topologies to reexamine two long-standing questions in spider evolution: the tempo and mode of divergences and the evolution and diversification of their webs.

## MATERIAL AND METHODS

### *Extraction and transcriptome sequencing*

New transcriptomic data were generated for 53 spider specimens, focusing on increasing taxon sampling in areas previously undersampled or unsampled. Voucher specimens and tissue for these animals are deposited at the Museum of Comparative Zoology at Harvard University. In addition to material sequenced previously (Bond et al., 2014; Fernández et al., 2014; French et

al., 2014; Sanggaard et al., 2014; Sharma et al., 2014; Zhao et al., 2014; Brewer et al., 2015; Hedin, 2015; Meng et al., 2015; Garrison et al., 2016; Rix et al., 2017; Cheng and Piel, 2018; Fernández et al., 2018a; Kallal et al., 2018; Shao and Li, 2018; Michalik et al., 2019), our available taxon sampling includes 272 terminals, of which 263 are spiders (Supplement 1). This sums to more than 100 additional taxa in comparison to the largest published dataset (Fernández et al., 2018a). This full taxon sample of 272 terminals is henceforth the *all* dataset. The *all* dataset includes representatives of 99 of 120 spider families (82.5%). Specifically, 13 of 21 mygalomorph families (61.9%) and 86 of 99 araneomorph families (86.9%) are represented. For the first time in a phylogenomic work, representatives of all 17 araneoid families are included in the analyses. To focus on resolving araneoid interfamilial relationships, we also analyzed a reduced matrix comprised of 94 araneoids and six outgroup lineages (eresids, nicodamoids, and the lycosid *Schizocosa rovneri*), which we call the *ara* dataset.

Extraction of mRNA and strand-specific cDNA library construction followed the protocols described in Fernández et al. (2018a). New RNA-Seq sequences were generated using Illumina HiSeq2500 (2 x 150 bp) technologies. Assembly, sanitation, and reading frame detection pipeline are as in Fernández et al. (2018a) with the addition of running the perl script Rcorrector (Song and Florea, 2015) for preassembly error correction and downstream efficiency.

*Orthology and matrix variations*

Orthology assessment was conducted using two methods: BUSCO (Simão et al., 2015) and UPhO (Ballesteros and Hormiga, 2016). Single copy loci retrieved using BUSCO were used for

the *all* and *ara* datasets due to its relative ease of use, and it follows Fernández et al. (2018a). UPhO delivered matrices with more missing data and lower locus counts in datasets with many terminals, and so its use is limited to the *ara* dataset.

Orthology assessment using BUSCO was conducted by querying a list of hidden Markov model profiles of putatively single copy arthropod amino acid loci. A maximum of 2,675 loci are retrievable. The pipeline used here follows Fernández et al. (2018a) and Kallal et al. (2018). The *all* dataset was tested at 1%, 50%, 67%, and 90% occupancy thresholds with and without trimming. Multiple sequence alignment was conducted using MAFFT v7 (Katoh and Standley, 2013) and trimmed, if relevant, using trimAl v1.2 (Capella-Gutiérrez et al., 2009) with default settings. For UPhO, the *ara* dataset was subjected to all-versus-all BLAST searches using an expectation value threshold of e = 1 x 10-3. Homolog clustering was performed using MCL in an inflation factor of 6 (van Dongen, 2000; Enright et al., 2002). Homolog groups were aligned and trimmed as above facilitated with the UPhO script paMATRAX+ (Ballesteros and Hormiga, 2016). Occupancy thresholds of 25%, 33%, and 50% were tested in UPhO, keeping in-paralogs and other variations of the same taxon per ortholog group. These were culled subsequently to include only the longest sequence when the ortholog groups were aligned and trimmed as described above.

To examine the effects of trimming on the resulting phylogenetic hypotheses, trees were inferred on matrices that had and had not been subjected to trimming via trimAl (Capella-Gutiérrez et al., 2009). In a few cases where entire terminals were trimmed, analyses were conducted with both the reduced taxon matrix and the full matrix.

*Tree inference*

Parsimony analyses were conducted using MPboot (Hoang et al. 2018b) with 1,000 bootstrap replicates in IQ-TREE v1.7-betaX (Minh et al., 2020). Accuracy and speed of bootstrap calculation in MPboot, which uses ultrafast bootstrapping (Hoang et al., 2018a) compared favorably to other methods. Maximum likelihood tree inference was conducted using IQ-TREE v1.6 (Nguyen et al., 2015), with the best-fit amino acid model of the supermatrix determined using ModelFinder, as implemented in IQ-TREE (Kalyaanamoorthy et al., 2017). Nodal support was estimated using ultrafast bootstrapping (Hoang et al. 2018a) and an SH-like approximate likelihood ratio test (Guindon et al., 2010). Individual gene trees were built using IQ-TREE v1.6, with each run 5–10 times with the highest likelihood tree kept, using the model JTT+G. The gene trees were then analyzed using ASTRAL-II v4.10.12 (Mirarab and Warnow, 2015) in a multispecies coalescence (MSC) framework, with quality and support determined by normalized quartet score and local posterior probabilities (Sayyari and Mirarab, 2016). Due to low sampling variance in traditional resampling resulting in inflated supports, we used gene concordance factors (gCF) and site concordance factors (sCF) as implemented in IQ-TREE v1.7-betaX (Minh et al., 2020). This metric determines the number of loci and sites that are reflected in the maximum likelihood topology. This was conducted on a subset of analyses that used both BUSCO and UPhO orthology assessment methods. Robinson-Foulds (RF) distances (Robinson and Foulds, 1981) were generated using IQ-TREE (Nguyen et al., 2015).

*Divergence dating*

Time calibration of large phylogenies can be a difficult prospect, with some coestimation methods scaling poorly and requiring subsampling or months (or years) of computing time (e.g., Laumer et al., 2019). Following Eberle et al. (2018) wherein a number of faster methods were tested, we selected treePL (Smith and O'Meara, 2012) to analyze these data. This non-parametric rate-smoothing penalized likelihood method performed favorably against MCMCtree (Yang, 2007) and RelTime (Tamura et al., 2012). The topology generated by Fernández et al. (2018a) was reevaluated using an expanded and revised fossil calibration in a recent review of spider fossils and their placement, wherein younger clade ages were found than in most phylogenies (Magalhães et al., 2020). These findings have been modified for use in this work using treePL, with fossil maxima using the lower bounds of the 95% confidence intervals determined by Magalhães et al. (2020) to prevent anomalously ancient divergences. A total of 29 fossils were used as calibration points on 26 nodes and are summarized in Supplement S18. 'Prime' and 'thorough' options were used to optimize the analyses, and cross validation was used to select the optimal smoothing parameter. The smoothing parameter penalizes rate heterogeneity across the tree; increase of smoothing value assumes lower rate heterogeneity and more clock-like mode of rate evolution. Following Eberle et al. (2018), penalized likelihood optimization iterations were increased from the default of 2 to 5, and the number of penalized likelihood simulated annealing was doubled from 5,000 to 10,000.

*Web evolution*

The coding of web architecture was modified from Fernández et al. (2018b) and Coddington et al. (2019); some of the mygalomorph entries were taken from Opatova et al. (in press) (see Supplemental File S16). Following the analyses of Fernández et al. (2018a, 2018b), we use two separate reconstructions to first address the origin of foraging webs (and orb webs in particular) and second, the diversification of web architectures. Our analyses follow an approach specifically developed to handle hierarchically nested characters/states using structured Markov models (SMM) and hidden states (Tarasov, 2019). This method was developed to handle cases where hierarchical dependency between phenotypic traits occurs, such as in the case of spider webs and their architecture: only when the web is present it can have architecture. Hierarchical dependencies result in inapplicable codings which cannot be analyzed properly using alternative approaches (Tarasov, 2019). To apply this method we scored two characters: one for the absence or presence of foraging webs, and a second for scoring whether the web is an orb or not. We then amalgamated these two characters following Tarasov (2019) and built several models with increasing complexity starting from a three-state one-rate model without hidden states to models with up to 14 hidden states and 15 transition rates. In this framework, hidden states within an observable state only imply that the evolution of the observable state is not Markovian. Thus, observable states should consist of two or more hidden states to describe trait evolution as a Markovian process and all hidden states (where present) should be interpreted as the corresponding observable states. Each model was run with a switch dependency on and off (see Tarasov, 2019) using the rayDISC function in the R package corHMM (Beaulieu et al., 2017; Paradis and Schliep, 2019). The “switch-on” type of dependencies arise from phenotypic dependencies between traits in which a hierarchically upstream trait switches on and off the downstream trait (e.g., if the web is absent, then the web architecture character is switched off

and does not evolve). For the purpose of comparison between different approximations to the scoring of web absence, we also designed a model where web absence is treated as a third state in a web architecture character as done in most previous analyses (e.g. Blackledge et al., 2009, Garrison et al., 2016, Dimitrov et al., 2017, Fernández et al., 2018a, 2018b). As an alternative Coddington et al. (2019) have scored web presence and web architecture in two different characters where "?" and reconstructed ancestral states for each of these characters independently and the two reconstructions were visually interpreted together. Their approach, however, does not jointly estimate the marginal likelihood for both characters at internal nodes (although they show a hierarchical relationship). In addition, scoring taxa where the web is absent is equivalent to polymorphic coding which implies that taxa with such scores are interpreted as having some web architecture while they do not build foraging webs at all. Because of these shortcomings and given that the SMM approach properly handles hierarchical dependencies we have not analyzed web diversity as two characters as proposed by Coddington et al. (2019). The second question – the diversification of web architectures – is addressed scoring web types in 13 states instead of just two: brush sheet (1), irregular aerial sheet (2), irregular ground sheet (3), stereotyped aerial sheet (4), cob-web (5), orb web (6), aerial silk tube (7), tubular silk-lined burrow with trap door(s) (8), irregular non-sheetlike tangle (9), terminal line (10), pseudo-orb (11), burrow with collar door (12) and open burrow (13). While many analyses essentially treat variation in webs as phylogenetic character(s) (i.e., a transformation series) to be optimized on a tree, this approach involves a highly questionable expansion of the concept of homology, albeit tacitly adopted by many authors (e.g., Blackledge et al. 2009; Dimitrov et al. 2012; Bond et al. 2014; Dimitrov et al. 2017; Fernández et al. 2018a) (see 'Discussion'). Using this 13-states scoring scheme for web architecture and the additional web

presence/absence character, we evaluated a set of SMM with rates that vary from a single rate up to 15 different rates, with and without switch character dependency, and with and without a hidden state associated to the orb web state. Because models with higher number of parameters generally result in better likelihoods, in order to compare models performance and avoid over parameterization, for each model we calculated the corresponding Akaike information criterion (AIC) (Beaulieu et al., 2017; Paradis and Schliep, 2019) and BIC (Bayesian Information Criterion) values (see supplementary materials). We used BIC for model comparison as a recent study shows that AIC may be biased in a phylogenetic context (Susko and Roger, 2020). We should also note that some web codings are open to different interpretations rather than simply being "correct" or "incorrect." For example, the sheet webs of *Physoglenes puyehue* (Physoglenidae) are extremely similar to the webs of many species of Linyphiidae (Dimitrov et al., 2017, fig. 7; Arnedo et al., 2009, fig. 2), and in absence of any data on the web building behavior of physoglenids, we code its web as we have coded linyphiid webs (a stereotyped aerial sheet), while Coddington et al. (2019) code *Physoglenes* as having an irregular aerial sheet and linyphiids as having stereotyped aerial sheets.

## RESULTS

### *Matrix composition*

Analyses were conducted on matrices ranging in size from 12 to 2,661 loci and between 4,491 and 1,270,722 sites depending on the orthology assessment method, occupancy threshold, and trimming (Table 1). For BUSCO analyses of all spiders, the most compact matrix (90%

occupancy) had 76 loci, whereas the maximum matrix including all loci represented in at least two terminals had 2,661 loci. For araneoids only, BUSCO analyses ranged from 12 (90% occupancy) to 2,040 (33% occupancy) loci, and for UPhO, from 162 (50% occupancy) to 1,263 (33% occupancy) loci. When implemented, trimming reduced the matrix size by between one-third to two-thirds (e.g., BUSCO on all spiders at 67% occupancy: 460,845 untrimmed, 221,014 trimmed).

| Dataset | Orthology assessment | Occupancy threshold | Loci | Sites |
|---|---|---|---|---|
| *all* | BUSCO | 1% | 2,665 | 1,270,722 |
| *all* | BUSCO | 50% | 1,409 | 526,007 |
| *all* | BUSCO | 67% | 598 | 221,014 / 460,845 |
| *all* | BUSCO | 90% | 76 | 33,187 / 66,518 |
| *ara* | BUSCO | 33% | 2,040 | 930,557 |
| *ara* | BUSCO | 50% | 1,458 | 624,653 |
| *ara* | BUSCO | 67% | 646 | 270,267 / 410,393 |
| *ara* | BUSCO | 90% | 12 | 4,491 |
| *ara* | UPhO | 25% | 1,263 | 438,670 / 837,626 |
| *ara* | UPhO | 33% | 589 | 184,895 / 386,958 |
| *ara* | UPhO | 50% | 162 | 37,043 / 70,599 |

*Phylogenetic analyses*

Model selection using ModelFinder selected 'JTT plus empirical frequencies' as the preferred model for all analyses, additionally selecting a free rate parameter of between seven and ten, which relaxes the assumption of the Gamma distribution in which four categories are insufficient for fitting the data. All major spider clades (e.g., Mygalomorphae, Araneomorphae, Synspermiata, Palpimanoidea, Entelegynae, Araneoidea, RTA Clade) were recovered by most of the 33 phylogenetic analyses conducted (Figs. 2, 3). The araneoid family Symphytognathidae, not represented in previous transcriptomic analyses, was supported as sister group to Anapidae, the sister lineage to all other araneoid families except theridiids, which are the sister group of all other Araneoidea. Within the UDOH Grade, lower occupancy matrices (1–67%) found Deinopidae as sister lineage to the RTA Clade, with Hersiliidae + Oecobiidae sister lineage to Deinopidae + the RTA clade, and Uloboridae sister lineage to all of these. In contrast, high occupancy analyses (90%) found Uloboridae to be the sister group to the RTA Clade and Deinopidae as the sister lineage to Hersiliidae + Oecobiidae.

Different topologies and their support values are summarized in Table 2. Lower occupancy matrices resulted in topologies more akin to each other based on RF distances than those derived from higher occupancy matrices. Trimming seemed to result in little difference compared to other variations of the matrix in concatenation and slightly more so for MSC. Furthermore, topologies were more similar (based on RF distances) for concatenation methods versus MSC methods regardless of occupancy at lower thresholds. Such differences included concatenation

analyses supporting Eresidae as the sister lineage to Nicodamoidea + Araneoidea, whereas MSC analyses supported Araneoidea as sister lineage of Eresidae + Nicodamoidea. Concatenation analyses supported Cyatholipidae as sister lineage to Synaphridae; MSC supported cyatholipids as sister lineage to linyphioids (Linyphiidae + Pimoidae). Another relevant difference within Araneoidea was related to the tetragnathoids (Tetragnathidae, Mimetidae and Arkyidae) and their relatives. MSC analyses found Mysmenidae as sister group to the clade including Malkaridae, Arkyidae, Mimetidae, and Tetragnathidae, but concatenation analyses determined Malkaridae as sister group to the remaining taxa, or Malkaridae + Mysmenidae sister clade to tetragnathoids (BUSCO and UPhO, respectively). In most cases, the alternatives had high local posterior probabilities and bootstrap scores.

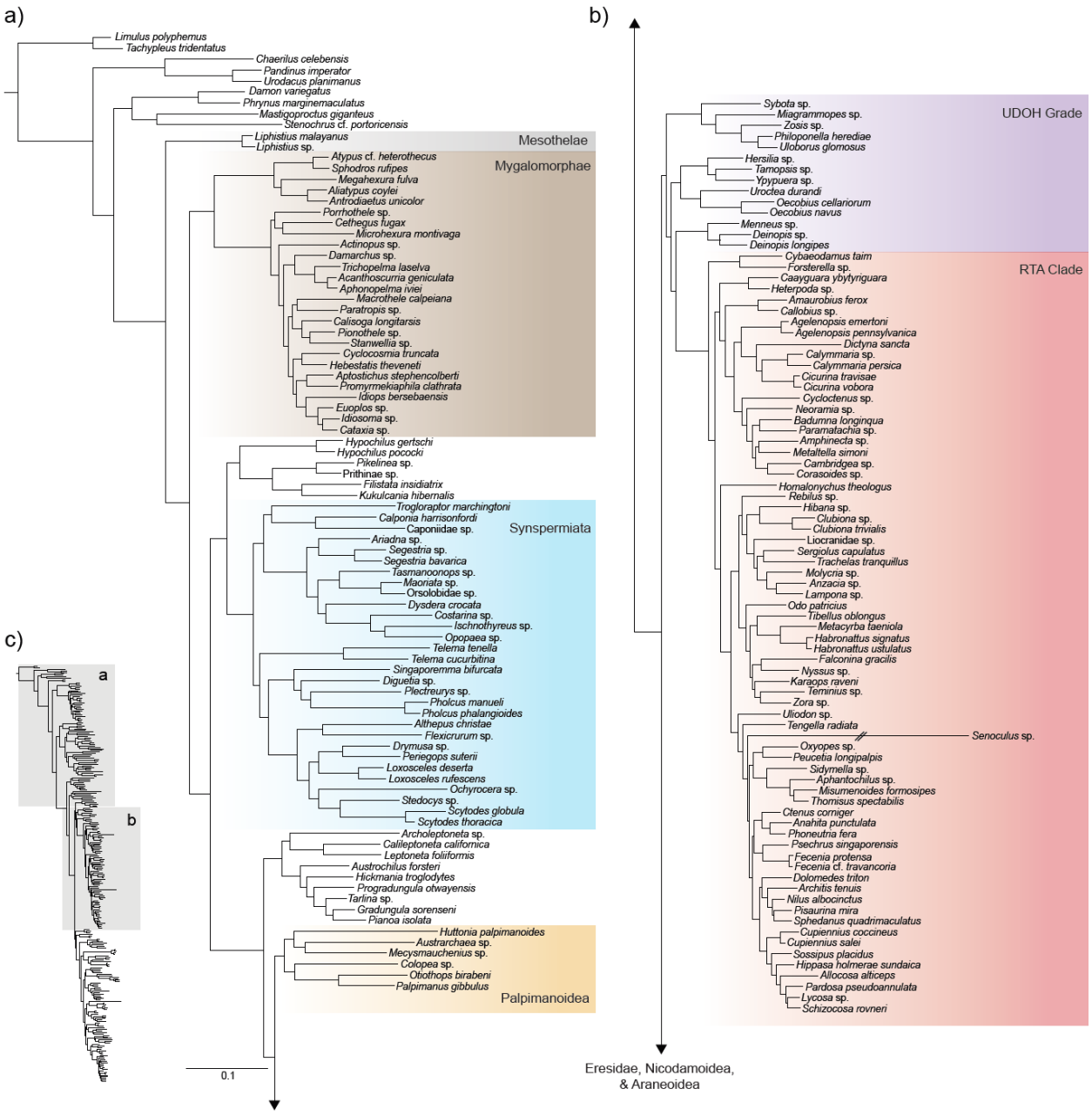
| Clade | Method | UFB (Tr/Un) | SH-aLRT | gCF | sCF | LPP (Tr/Un) |
|---|---|---|---|---|---|---|
| NIC + ARA | B | 100/100 | - | 5.74 | 33 | (Ere, NIC): 0.98/0.97 |
| Ana + Sym | B | 100/100 | - | 14.7 | 35.9 | 1/1 |
| Mal + Mys | B | 97/91 | - | 4.87 | 29.4 | Tr: (Mys, (Mal, Tet)): 0.96; Un: ((Mys, (Mal, Tet)): 0.5 |
| Syn + Phys | B | 100/100 | - | 4.41 | 33.3 | (Syn, (The, Ara)): 0.57/0.64 |
| The + Ara | B | 100/100 | - | 11.6 | 35.7 | 1/1 |
| NIC + ARA | U | 99/98 | 98.3/98.6 | 24.5 | 31.6 | (Ere, NIC): 0.83/0.87 |
| Ana + Sym | U | 100/100 | 100/100 | 2.93 | 35.6 | 1/1 |
| Mal + (Mys + Tet) | U | 100/100 | 100/100 | 4.41 | 34.9 | (Mys, (Mal, Tet)): 1/1 |
| Mys + Tet | U | 100/99 | 99.3/99.6 | 1.01 | 29.1 | (Mal, Tet): 0.93/0.99 |
| Syn + Nes | U | 100/100 | 100/99.8 | 11.2 | 30.4 | Tr: (Syn, (Nes, Phy)): 0.8; Un: (Syn, ((Nes,Phy), (The,Ara))): 0.61 |
| The + Ara | U | 100/100 | 100/100 | 24.5 | 36.5 | 1/1 |

Additional differences in tree topology were found using the different orthology assessment methods. In addition to the differences related to tetragnathoids (including tip-level differences within Tetragnathidae and Mysmenidae, Figs. 3a, 3c), the placement of Synotaxidae differed between analyses. Using BUSCO, *Synotaxus* was sister lineage to Physoglenidae (Fig. 3a), whereas UPhO determined *Synotaxus* to be the sister group of Nesticidae (Fig. 3d). Different topologies were also detected in the family Araneidae. Argiopinae (*Argiope* + *Cyrtophora*) was not monophyletic in BUSCO analyses but it was monophyletic using UPhO; additionally, there were various changes in the sister group to gasteracanthines (Figs. 3a, 3e). Additional topologies with full support are available at the Harvard Dataverse repository. Trimming did not have an appreciable effect on topology. The RF distances were lower between untrimmed and trimmed MSC analyses (RF = 32) than concatenation (RF = 44), less than half of the difference between MSC and concatenation regardless of trimming done (RF = 92–98).

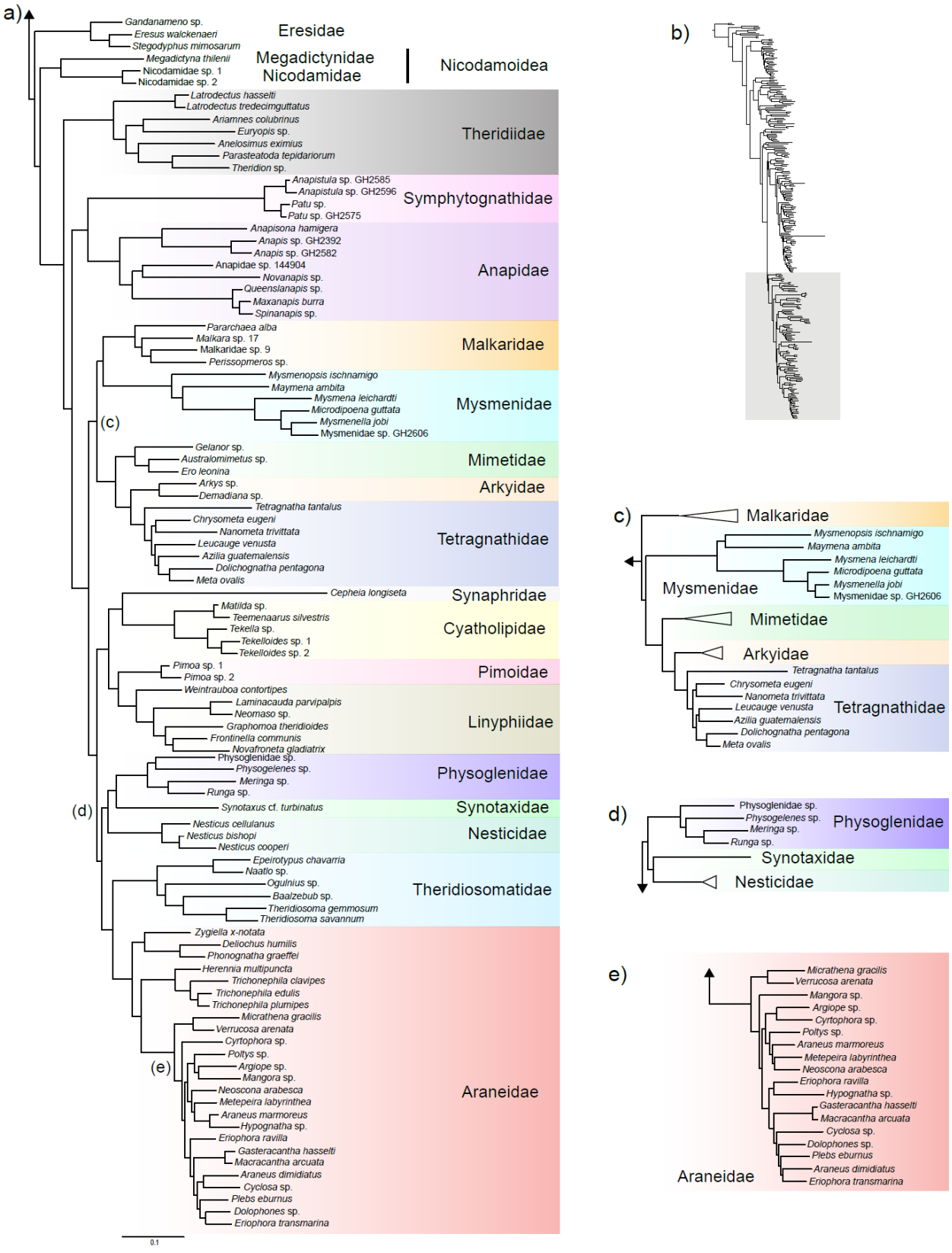
For most nodes, we found uniformly high ultrafast bootstrap support. Concordance factor values, however, varied widely, with gCFs lower than their respective sCFs. Where present, low bootstrap support coincided with low concordance factor values. Furthermore, approximately 90% of the nodes had full bootstrap support despite concordance factors ranging from zero to 100. Even in well established clades, concordance values do not approach the uniformly high bootstrap support (for gCF and sSCF, respectively, in BUSCO analysis with 67% occupancy; S14): Opisthothelae (49.6, 39.6), Mygalomorphae (56.8, 44.3), Araneomorphae (43.3, 38.7), Synspermiata (41.1, 42.7), Palpimanoidea (13.7, 37.7), RTA Clade (37.0, 50.4), and Araneoidea (26.5, 40.1). Within Araneoidea, there is a similar pattern, with even congenerics' concordance factor scores varying (S14, S15). For instance, *Trichonephila edulis* and *T. plumipes* were scored

(47.0, 63.6) and (55.0, 50.3) in the preferred BUSCO and UPhO analyses; *Theridiosoma gemmosum* and *T. savannum* scored (80.9, 75.2) and (97.7, 73.8).

Parsimony analyses differed minimally from model-based analyses, except in two key areas. First, in the analysis on the *all* dataset, Eresidae and Nicodamidae form a clade (UFBoot = 91), which is in turn sister group to the UDOH Grade + RTA Clade (UFBoot = 58) rather than Araneoidea (S16). Second, the symphytognathoid families Anapidae, Mysmenidae, Symphytognathidae, and Synaphridae form a clade, sister group to all other araneoid families except Theridiidae (S16, S17). Theridiosomatidae is a sister group to Araneidae as in model-based analyses (UFBoot = 98 and 95 on *all* and *ara* datasets, respectively).


a)
Limulus polyphemus
Tachypleus tridentatus
Chaerilus celebensis
Pandinus imperator
Urodacus planimanus
Damon variegatus
Phrynus marginemaculatus
Mastigoproctus giganteus
Stenochrus cf. portoricensis
Liphistius malayanus
Liphistius sp.
Mesothelae
Atypus cf. heterothecus
Sphodros rufipes
Megahexura fulva
Aliatypus coylei
Antrodiaetus unicolor
Porrhothele sp.
Cethegus fugax
Microhexura montivaga
Actinopus sp.
Damarchus sp.
Trichopelma laselva
Acanthoscurria geniculata
Aphonopelma iviei
Macrothele calpeiana
Paratropis sp.
Calisoga longitarsis
Pionothele sp.
Stanwellia sp.
Cyclocosmia truncata
Hebestatis theveneti
Aptostichus stephencolberti
Promyrmekiaphila clathrata
Idiops bersebaensis
Euoplos sp.
Idiosoma sp.
Cataxia sp.
Mygalomorphae
Hypochilus gertschi
Hypochilus pococki
Pikelinea sp.
Prithinae sp.
Filistata insidiatrix
Kukulcania hibernalis
Trogloraptor marchingtoni
Calponia harrisonfordi
Caponiidae sp.
Ariadna sp.
Segestria sp.
Segestria bavarica
Tasmanoonops sp.
Maoriata sp.
Orsolobidae sp.
Dysdera crocata
Costarina sp.
Ischnothyreus sp.
Opopaea sp.
Telema tenella
Telema cucurbitina
Singaporemma bifurcata
Diguetia sp.
Plectreurys sp.
Pholcus manueli
Pholcus phalangioides
Althepus christae
Flexicrurum sp.
Drymusa sp.
Periegops suterii
Loxosceles deserta
Loxosceles rufescens
Ochyrocera sp.
Stedocys sp.
Scytodes globula
Scytodes thoracica
Synspermiata
Archoleptoneta sp.
Calileptoneta californica
Leptoneta foliiformis
Austrochilus forsteri
Hickmania troglodytes
Progradungula otwayensis
Tarlina sp.
Gradungula sorenseni
Pianoa isolata
Huttonia palpimanoides
Austrarchaea sp.
Mecysmaucheniu sp.
Colopea sp.
Otiothops birabeni
Palpimanus gibbulus
Palpimanoidea
0.1
c)
a
b
b)
Sybota sp.
Miagrammopes sp.
Zosis sp.
Philoponella herediae
Uloborus glomosus
Hersilia sp.
Tamopsis sp.
Ypypuera sp.
Uroctea durandi
Oecobius cellariorum
Oecobius navus
Menneus sp.
Deinopis sp.
Deinopis longipes
UDOH Grade
Cybaeodamus taim
Forsterella sp.
Caayguara ybytyriguara
Heterpoda sp.
Amaurobius ferox
Callobius sp.
Agelenopsis emertoni
Agelenopsis pennsylvanica
Dictyna sancta
Calymmaria sp.
Calymmaria persica
Cicurina travisae
Cicurina vobora
Cycloctenus sp.
Neoramia sp.
Badumna longingua
Paramatachia sp.
Amphinecta sp.
Metaltella simoni
Cambridgea sp.
Corasoides sp.
Homalonychus theologus
Rebilus sp.
Hibana sp.
Clubiona sp.
Clubiona trivialis
Liocranidae sp.
Sergiolus capulatus
Trachelas tranquillus
Molycria sp.
Anzacia sp.
Lampona sp.
Odo patricius
Tibellus oblongus
Metacyrba taeniola
Habronattus signatus
Habronattus ustulatus
Falconina gracilis
Nyssus sp.
Karaops raveni
Terminus sp.
Zora sp.
Uliodon sp.
Tengella radiata
Senoculus sp.
Oxyopes sp.
Peucetia longipalpis
Sidymella sp.
Aphantochilus sp.
Misumenoides formosipes
Thomisus spectabilis
Ctenus corniger
Anahita punctulata
Phoneutria fera
Psechrus singaporensis
Fecenia protensa
Fecenia cf. travancoria
Dolomedes triton
Architis tenuis
Nilus albocinctus
Pisaurina mira
Sphedanus quadrimaculatus
Cupiennius coccineus
Cupiennius salei
Sossipus placidus
Hippasa holmerae sundaica
Allocosa alticeps
Pardosa pseudoannulata
Lycosa sp.
Schizocosa rovneri
RTA Clade
Eresidae, Nicodamoidea,
& Araneoidea

34

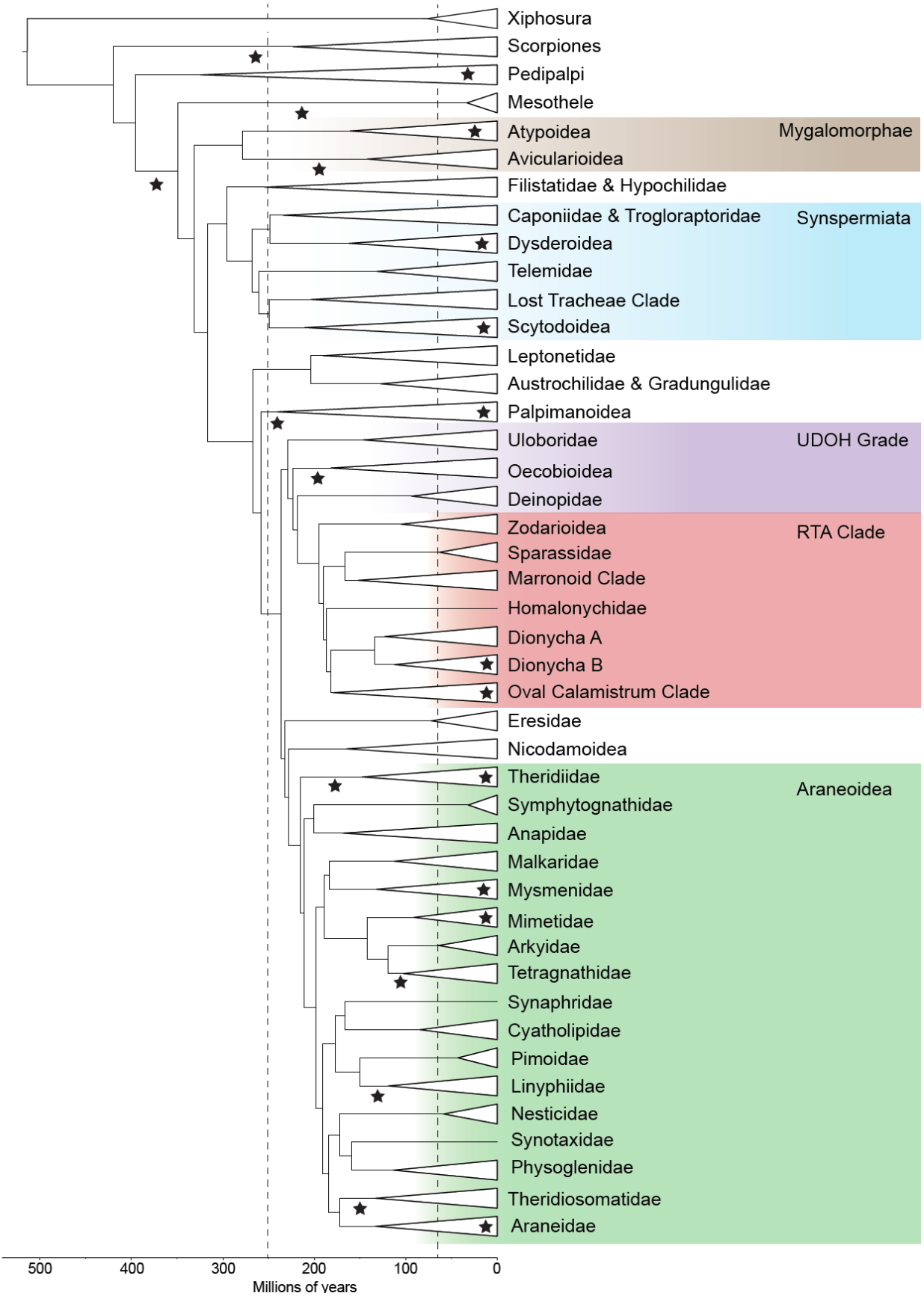

a)
Gandanameno sp.
Eresus walckenaeri
Stegodyphus mimosarum
Megadictyna thilenii
Nicodamidae sp. 1
Nicodamidae sp. 2
Eresidae
Megadictynidae
Nicodamidae
Nicodamoidea
Latrodectus hasselti
Latrodectus tredecimguttatus
Ariamnes colubrinus
Euryopis sp.
Anelosimus eximius
Parasteatoda tepidariorum
Theridion sp.
Theridiidae
Anapistula sp. GH2585
Anapistula sp. GH2596
Patu sp.
Patu sp. GH2575
Symphytognathidae
Anapisona hamigera
Anapis sp. GH2392
Anapis sp. GH2582
Anapidae sp. 144904
Novanapis sp.
Queenslanapis sp.
Maxanapis burra
Spinanapis sp.
Anapidae
Pararchaea alba
Malkara sp. 17
Malkaridae sp. 9
Perissopmeros sp.
Malkaridae
Mysmenopsis ischnamigo
Maymena ambita
Mysmena leichardti
Microdipoena guttata
Mysmenella jobi
Mysmenidae sp. GH2606
Mysmenidae
(c)
Gelanor sp.
Australomimetus sp.
Ero leonina
Mimetidae
Arkys sp.
Demadiana sp.
Arkyidae
Tetragnatha tantalus
Chrysometa eugeni
Nanometa trivittata
Leucauge venusta
Azilia guatemalensis
Dolichognatha pentagona
Meta ovalis
Tetragnathidae
Cepheia longiseta
Synaphridae
Matilda sp.
Teemenaarus silvestris
Tekella sp.
Tekelloides sp. 1
Tekelloides sp. 2
Cyatholipidae
Pimoa sp. 1
Pimoa sp. 2
Pimoidae
Weintrauboa contortipes
Laminacauda parvipalpis
Neomaso sp.
Graphomoa theridioides
Frontinella communis
Novafroneta gladiatrix
Linyphiidae
Physoglenidae sp.
Physogelenes sp.
Meringa sp.
Runga sp.
Physoglenidae
(d)
Synotaxus cf. turbinatus
Synotaxidae
Nesticus cellulanus
Nesticus bishopi
Nesticus cooperi
Nesticidae
Epeirotypus chavarria
Naatlo sp.
Ogulnius sp.
Baalzebub sp.
Theridiosoma gemmosum
Theridiosoma savannum
Theridiosomatidae
Zygiella x-notata
Deliochus humilis
Phonognatha graeffei
Herennia multipuncta
Trichonephila clavipes
Trichonephila edulis
Trichonephila plumipes
Micrathena gracilis
Verrucosa arenata
Cyrtophora sp.
Poltys sp.
(e)
Argiope sp.
Mangora sp.
Neoscona arabesca
Metepeira labyrinthea
Araneus marmoreus
Hypognatha sp.
Eriophora ravilla
Gasteracantha hasselti
Macracantha arcuata
Araneus dimidiatus
Cyclosa sp.
Plebs eburnus
Dolophones sp.
Eriophora transmarina
Araneidae
0.1
b)
c)
Malkaridae
Mysmenopsis ischnamigo
Maymena ambita
Mysmena leichardti
Microdipoena guttata
Mysmenella jobi
Mysmenidae sp. GH2606
Mysmenidae
Mimetidae
Arkyidae
Tetragnatha tantalus
Chrysometa eugeni
Nanometa trivittata
Leucauge venusta
Azilia guatemalensis
Dolichognatha pentagona
Meta ovalis
Tetragnathidae
d)
Physoglenidae sp.
Physogelenes sp.
Meringa sp.
Runga sp.
Physoglenidae
Synotaxidae
Nesticidae
e)
Micrathena gracilis
Verrucosa arenata
Mangora sp.
Argiope sp.
Cyrtophora sp.
Poltys sp.
Araneus marmoreus
Metepeira labyrinthea
Neoscona arabesca
Eriophora ravilla
Hypognatha sp.
Gasteracantha hasselti
Macracantha arcuata
Cyclosa sp.
Dolophones sp.
Plebs eburnus
Araneus dimidiatus
Eriophora transmarina
Araneidae

35

36

*Divergence dating*

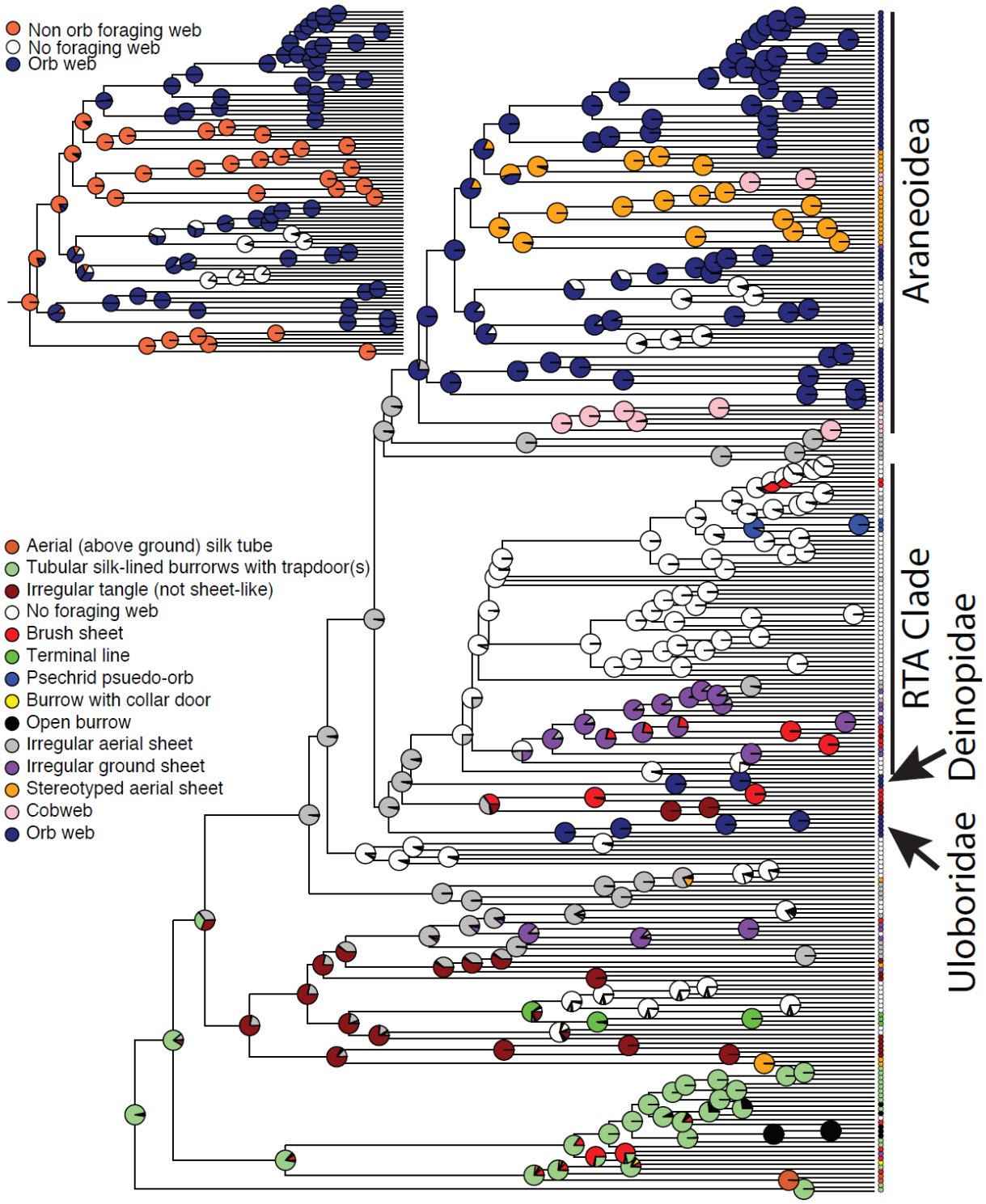
Cross-validation in treePL selected the default smoothing factor of 100. Divergence dates inferred using treePL are depicted in Fig. 4. The last common ancestor of spiders and Pedipalpi occurred approximately 396 Ma, and the last common ancestor of Mesothelae and Opisthothelae occurred approximately 350 Ma. Mygalomorphae and Araneomorphae last shared a common ancestor 331 Ma. Within Araneoidea, families diverged from their sister lineages 120–215 Ma, with Arkyidae + Tetragnathidae being the youngest and Theridiidae + other araneoids being the oldest.


Xiphosura
Scorpiones
Pedipalpi
Mesothele
Atypoidea
Avicularioidea
Mygalomorphae
Filistatidae & Hypochilidae
Caponiidae & Trogloraptoridae
Dysderoidea
Telemidae
Lost Tracheae Clade
Scytodoidea
Synspermiata
Leptonetidae
Austrochilidae & Gradungulidae
Palpimanoidea
Uloboridae
Oecobioidea
Deinopidae
UDOH Grade
Zodarioidea
Sparassidae
Marronoid Clade
Homalonychidae
Dionycha A
Dionycha B
Oval Calamistrum Clade
RTA Clade
Eresidae
Nicodamoidea
Theridiidae
Symphytognathidae
Anapidae
Malkaridae
Mysmenidae
Mimetidae
Arkyidae
Tetragnathidae
Synaphridae
Cyatholipidae
Pimoidae
Linyphiidae
Nesticidae
Synotaxidae
Physoglenidae
Theridiosomatidae
Araneidae
Araneoidea
500
400
300
200
100
0
Millions of years

*Web evolution*

The different SMM models used in the ancestral state reconstruction analyses and the relevant statistics are summarized in Table (S22). Several models scored closely when using reduced scoring for the web architecture. A model where we allowed for one hidden state for the orb web character and two rates scored marginally, but not significantly (BIC difference < 2), better than the second best model. This model, as well as most of the best scoring models from the analyses of the two states web architecture dataset, supports five independent origins of the orb webs: three in Araneoidea, once in Deinopidae, and once in Uloboridae (Figs. 5, S20). There are multiple instances of web loss within and outside araneoids and support for the presence of a web as the ancestral condition for spiders (as in Fernández et al., 2018a, 2018b). The analysis of the dataset with fine grain scoring of web architectures (13 states) also supports multiple origins of orb webs but it differs in that it suggests a single origin of orb webs in Araneoidea (Figs. 5, S21). Tubular silk-lined burrow with trap door(s) is inferred as the ancestral web type for spiders and multiple losses of web from different ancestral web types are inferred across the phylogeny. In the 13-state analyses we found two models that scored very closely with only a marginal difference in their BIC values, however the inferred evolution of web architecture under these two models does not differ and here we present the result of the model with the best BIC (Fig. 5). Analyses of both the two and the 13-state dataset support multiple instances of transitions from webless foraging to webs in the RTA Clade and a single loss of webs in the ancestor of this group. When we tested models with an increasing number of hidden states for the non-orb web state in the two state web architecture results converged to those of analyses of the 13-state

dataset when hidden character spaces was increased to 11. Only two models of all those tested inferred a single origin of orb webs as hypothesized by the ancient orb hypothesis but those were significantly worse than any of the models suggesting multiple origins (BIC differences were higher than 30 in favour of models resulting in multiple origins inference).


Non orb foraging web
No foraging web
Orb web
Aerial (above ground) silk tube
Tubular silk-lined burrorws with trapdoor(s)
Irregular tangle (not sheet-like)
No foraging web
Brush sheet
Terminal line
Psechrid psuedo-orb
Burrow with collar door
Open burrow
Irregular aerial sheet
Irregular ground sheet
Stereotyped aerial sheet
Cobweb
Orb web
Araneoidea
RTA Clade
Deinopidae
Uloboridae

77

78

79

80

## DISCUSSION

### *Spider phylogeny*

Our study, with strategically increased taxon sampling to maximize web diversity across lineages and a variety of analytical methods produced a well-resolved spider tree of life. For the overlapping taxa, results recovered most of the phylogenetic relationships established by previous works (Garrison et al., 2016; Dimitrov et al., 2017; Wheeler et al., 2017; Cheng and Piel, 2018; Fernández et al., 2018a; Hedin et al., 2019; Kallal et al., 2018; Shao and Li, 2018; Kulkarni et al., 2019; Michalik et al., 2019), reinforcing several previous topologies. For this reason, we limit our discussion of relationships to highlight places in which our results are novel.

Within Mygalmorphae, we found the typical atypoid – avicularioid split within Mygalomorphae. A notable difference compared to recent work on this clade by Fernández et al. (2018a) and Hedin et al. (2019) involves the placement of Dipluridae and Porrhothelidae; we find them as sister lineages, as the earliest diverging clade with Avicularioidea. The analyses of Fernández et al. (2018a) did not place *Porrhothele* with good support, whereas Hedin et al. (2018) placed this taxon more distally, sister to a clade including Macrothelidae, Nemesiidae, Halonoproctidae, Atracidae, and Actinopodidae. The more densely sampled mygalomorph analyses of Opatova et al. (in press) show porrhothelids as early diverging avicularioids and diplurids in the fraught clade including nemesiids and close relatives. Our analyses place *Actinopus* as an early diverging avicularioid near theraphosids whereas actinopodids are sister lineage to atracids in Hedin et al. (2018) and Opatova et al. (in press).

104

105 Within Synspermiata, relationships are similar to other works on the group (e.g., Michalik et al.,
106 2019), with monophyly for the major clades Dysderoidea, Scytodoidea, and the Lost Tracheae
107 Clade. Notably, Ochyroceratidae and Psilodercidae are not closely related, a proposal initially
108 based on morphology (Wunderlich 2008) but not supported conclusively by molecular data (the
109 analysis of Wheeler et al. 2017 places 'cf. *Psiloderces*' in Ochyroceratidae with weak support).
110 Plectreuridae is placed in the Lost Tracheae Clade with tetrablemmids and pholcids, following
111 Wheeler et al. (2017) and Shao and Li (2018). Additionally, Oonopidae rather than Orsolobidae
112 is the sister group of Dysderidae, *contra* the results of Fernández et al. (2018a). In the
113 superfamily Palpimanoidea (trapjaw spiders, pelican spiders, and their kin), all five families are
114 represented, but relationships differ from those of Fernández et al. (2018a) as well as
115 palpimanoid-specific analyses using Sanger markers and morphology (Wood et al. 2012) or
116 UCEs (Wood et al. 2018, Kulkarni et al. 2019). Our taxon sampling is more limited than that of
117 Wood et al. (2018), and the degree of difference suggests sampling, data type, and analysis type
118 are very important for resolving palpimanoid relationships.

119

120 Most analyses (see above) place Uloboridae as sister lineage to a clade that includes all other
121 UDOH families (Hersilidae, Oecobiidae and Deinopidae) and the RTA lineages. Hersiliidae and
122 Oecobiidae are always sister lineages (Oecobioidea). This is consistent with the results of
123 Fernández et al. (2018a) and contrary to the results of Shao and Li (2018), in which eresids were
124 placed in the UDOH Grade. Within the RTA Clade, additional taxa change little of the
125 relationships established by previous works focusing on this group (e.g., Cheng and Piel, 2018).
126 The non-monophyly of Ctenidae is supported, following other recent works (Wheeler et al.,

2017; Piacentini and Ramírez, 2019). In Fernández et al. (2018a), the zoropsids *Tengella* and *Uliodon* were sister groups but not strongly supported; here, they are not supported as sister lineages and Zoropsidae is not monophyletic. The senoculid branch is very long (as in Fernández et al. 2018a); this may be an artifact due to a relatively low BUSCO count, and its placement varies in our analyses.

The majority of analyses also corroborate Nicodamoidea as the sister lineage of Araneoidea as suggested by Dimitrov et al. (2012, 2017), Wheeler et al. (2017) and Fernández et al. (2018a). Some analyses place Eresidae as sister group to the nicodamoids and/or as sister lineage to UDOH Grade + RTA Clade, but most place eresids as sister lineage to Araneoidea + Nicodamoidea. All but parsimony analyses differed from the results of Kulkarni et al. (2019) in this way. The earliest diverging araneoid lineage is Theridiidae, differing from Fernández et al. (2018a) where it was sister group to Anapidae which were in turn sister lineage to all remaining araneoids, but consistent with the results of Dimitrov et al. (2012) using Sanger data and Kulkarni et al. (2019) using UCEs. The newly included araneoid family Symphytognathidae is sister group to Anapidae in virtually all analyses; symphytognathids were not represented in Fernández et al. (2018a), which may explain the placement of Anapidae as sister to Theridiidae. Anapidae and Symphytognathidae are the only two symphytognathoid families to form a clade based on model-based transcriptomic data. Parsimony analyses showed a monophyletic symphytognathoid group (sans Theridiosomatidae), a result strikingly similar to the UCE-based analyses by Kulkarni et al. (2019) where the symphytognoid families are resolved as a clade. The UCE datasets that support symphytognathoid monophyly happen to have low occupancy (<50%), with more loci but also more missing data than datasets with higher occupancy.

Malkaridae and Mysmenidae are placed near the base of the tetragnathoid clade, but their positions vary. Their phylogenetic placement has key implications for the evolution of the capture web given that malkarids, arkyids, and mimetids do not spin foraging webs (Framenau et al., 2010; Benavides et al., 2017; Hormiga and Scharff, 2020) whereas webs are found in mysmenids and tetragnathids (see web discussion below). Within Tetragnathidae, *Chrysometa* and *Nanometa* are closely related in many analyses, a relationship suggested by morphological data but lacking in Sanger-based molecular analyses (Álvarez-Padilla and Hormiga, 2011; Álvarez-Padilla, et al. 2020). The rogue taxon *Azilia* remains resistant to stable placement and is the sister taxon to either leucaugines or metaines depending on the analysis. Likewise, *Cepheia* (the sole representative of the family Synaphridae in our analyses) is typically placed as sister group to cyatholipids, but is found in a more basal location or near other symphytognathoids in other analyses, an issue perhaps solvable by including additional representatives of this small family (with only three genera and 13 described species). Physoglenids, nesticids, and synotaxids form a clade. Finally, all analyses find Theridiosomatidae and Araneidae as sister lineages including both model-based and parsimony analyses. Within Araneidae, the Phonognathinae lineage is sister to Nephilinae + remaining araneids. Short internodes in the araneid clade sister to gasteracanthines continue to evade resolution.

*Methodology and support values*

Notably, the more conserved loci from BUSCO, where they differed from UPhO, presented differences at more recent splits. For instance, morphology and Sanger-based data have nearly

always supported a close relationship between argiopine and cyrtophorine araneids (in fact, combined into Argiopinae in Scharff et al. 2020), but *Argiope* and *Cyrtophora* were not sister lineages in numerous BUSCO-based analyses. This suggests the utility of orthologs inferred with BUSCO, while functional at all levels, may be best at a deeper timescale. We also found the occupancy threshold could be raised higher using BUSCO than UPhO before returning anomalous results, but both were fairly consistent when provided with sparse matrices. The minimal occupancy BUSCO matrix was less different topologically from the 67% occupancy analyses than concatenation was from MSC analyses of the same occupancy (based on RF distances). This indicates that, for this dataset, the phylogenetic signal is more robust to the presence of missing data than to tree inference method. Furthermore, we found that robustness to trimming was comparable using both concatenation and MSC methods. This furthers Philippe et al.'s (2017) statement of signal overwhelming trimming effects using concatenation but also for coalescence-based methods, despite single locus tree issues found by Tan et al. (2015).

We found that bootstrap support and SH-aLRT support seemed to covary, but concordance factors differed considerably. The use of concordance factors (Minh et al. 2020) gives additional value to inflated bootstrap supports common in many phylogenomic analyses. Given that concordance factors do not have a generally recognized threshold of acceptable support (and indeed may never have such a threshold given how predicated they are on the number of sites and loci), they can be difficult to interpret. Interestingly, many relationships that are strongly supported by previous morphological and molecular works with high support found little corroboration in gCF and sCF analyses. For instance, only about one third of loci supported well-established hypotheses like the monophyly of Pimoidae + Linyphiidae (e.g., Arnedo et al., 2009;

196 Dimitrov et al. 2017) or the monophyly of Araneidae including nephilines (e.g., Kallal et al.
197 2018; Scharff et al. 2019). Perhaps worryingly, this suggests that many relationships supported
198 via bootstrapping could mask conflict with other nearly as frequent relationships.
199
200 Recent advances in understanding spider interfamilial relationships have come from two
201 different types of phylogenomic data. While transcriptomes are the more common data type in
202 spider phylogenetics (e.g., Garrison et al., 2016; Cheng and Piel, 2018; Fernández et al. 2018a;
203 Kallal et al., 2018), target capture methods are increasingly used (e.g., Hedin et al., 2018; Wood
204 et al., 2018, Hedin et al. 2019; Kulkarni et al., 2019; Opatova et al., in press). In a study with a
205 similar scope as this, Kulkarni et al. (2019) used UCEs to build a spider phylogeny with
206 considerable overlap but also concerning conflict in the results. The symphytognathoid
207 assemblage, which comprises Anapidae, Symphytognathidae, Mysmenidae, and
208 Theridiosomatidae (Griswold et al., 1998) and sometimes Synaphridae (Lopardo et al., 2010; not
209 sampled in Kulkarni et al. 2019) was found to be monophyletic in that UCE dataset – a result not
210 recovered in any other analysis using molecular data, but supported by morphology. Previous
211 studies using transcriptomes for three of these four families (analyzed as amino acids only) found
212 no support for close relationships among them (Fernández et al. 2018a). Another notable
213 incongruence with the UCE hypothesis is offered by the placement of Nicodamoidea, a small
214 clade that includes the families Nicodamidae and Megadictynidae, which is sister group to
215 Araneoidea with transcriptomes (Fernández et al., 2018a) whereas it is sister group to Eresidae
216 with UCEs (Kulkarni et al., 2019). The reason for this perplexing discrepancy remains to be fully
217 understood.
218

The loci revealed by transcriptomes are exclusively coding in nature and evolve at varying rates, while UCEs are ultraconserved regions composed of exons as well as introns. Hedin et al. (2019) showed that the UCEs targeted using the Arachnida probe set are mostly exonic and multiple UCE loci may target different regions of the same gene. This finding makes both data types comparable in a way that they both are coding, however UCEs can then be viewed as a subset of transcriptomes.

Transcriptomes are generally subjected to tree-building algorithms as amino acids whereas UCEs are analyzed as nucleotide data. The third nucleotide in synonymous codons might contain phylogenetic signal, however such information is masked in case of transcriptomes. Additionally, transcriptomic data are subjected to orthology assessment, however the same methods are not used with the UCE data prior to phylogenetic analysis. Instead, the duplicate removal step in the Phyluce pipeline (Faircloth, 2016) is assumed to filter out paralogs. Both data types are genome scale in size and the highly supported conflicting relationships are implausible since the bootstrap support is mostly >95 for large scale datasets; that is, one or both of such hypotheses must be erroneous. Future analytical comparison is therefore warranted to understand the conflict between these data types.

*Web evolution*

An increasing number of studies have relied on recent advances in spider phylogeny to reconstruct the evolution and diversification of webs across Araneae. Following Blackledge et al. (2009), recent analyses by Garrison et al. (2016) and Fernández et al. (2018a) coded nine and ten

types of webs respectively, with the latter study specifically coding capture webs plus a simplified variable with three states addressing orb web origins. Fernández et al. (2018a) specifically coded foraging webs, which are directly used to capture prey, although we lack a precise and universally accepted definition, if such a thing is feasible, of what exactly a "foraging web" is. Not surprisingly, both studies found webs to be ancestral for Araneae (Garrison et al., 2016: Fig 6; Fernández et al., 2018a: Fig. S2), but the former inferred a single origin for orb webs whereas the latter found multiple (ranging from two to six origins). Fernández et al. (2018a) used two different characters to first, reconstruct the origin of webs (including orb webs) and second, the diversification of web architectures, consequently it is not appropriate to use their second reconstruction to address the first question. The assertions that Fernández et al. (2018a) concluded that "the ancestral spider spun no foraging web" and that "spider webs evolved *de novo* 10–14 times", repeatedly made by Coddington et al. (2019), are inaccurate and easily refuted by the aforementioned study itself: Fernández et al.'s Figure S2 clearly shows the state "web (non-orbicular)", represented by a green circle, as ancestral for Araneae.

Using an alternative coding scheme to deal with the absence of foraging webs, along with a number of scoring changes correcting errors or providing alternative interpretations of web types, Coddington et al. (2019) reassessed the Fernández et al. (2018a) data, finding a single origin of the orb web (the "ancient origin hypothesis") but repeated independent losses of a foraging web in all lineages of the large RTA Clade and within Araneoidea. The RTA Clade is a large lineage of mostly cursorial spiders that includes such well-known groups as jumping and wolf spiders (Salticidae and Lycosidae, respectively). The most recent common ancestor of the

RTA Clade has been traditionally understood as lacking foraging webs given the vast majority of extant members of the clade do not have them. Types of webs traditionally used in phylogenetic analyses (that is, variation in web architecture treated as states of a character) may not make distinctions where superficial similarity masks important differences (e.g., brushed sheet-webs and others discussed by Eberhard and Hazzi, 2017). The confluence of different methods of ancestral character reconstruction, different web coding methods, different taxon samples and tree topologies, and indeed disagreement on whether webs are a character at all has led to misinterpretations and conflicting hypotheses on the evolution of webs.

Our analyses suggest that webs are ancestral for Araneae (see also Huang et al., 2018; Wang et al., 2018), were subsequently lost in several lineages and only in a few instances in the RTA Clade have evolved *de novo* in clades with cursorial ancestors. Orb webs also have a complex evolutionary history. Our results are consistent with those from Fernández et al. (2018a, 2018b) and Wolff et al. (2019) but not with those of Coddington et al. (2019). The latter found an ancestral orb web probable in the common ancestor of Entelegynae, then secondarily lost in numerous lineages. Coddington et al. (2019) also scored web presence and web architecture as two different characters and "?" was used to score web type in the web architecture character when the web was absent (i.e., treated as 'inapplicable'), but then they did not use SMMs and reconstructed ancestral states for each of these characters independently. The two reconstructions were displayed against each other and visually interpreted together. Their approach, however, does not jointly estimate the marginal likelihood for both characters at internal nodes (although they show a hierarchical relationship, as discussed above). In addition, scoring taxa where web is absent is equivalent to polymorphic coding which implies that taxa with such scores are

interpreted as having some web architecture while they do not build foraging webs at all ("?" notation is also used in the SMM approach of Tarasov (2019) but there it is handled specifically as a result of the character amalgamation). These critical shortcomings indicate that the method of Coddington et al. (2019) does not allow for proper treatment of hierarchically nested characters and thus we did not use it here.

Our results suggest that the orb web evolved twice in the UDOH Grade and one (in the analyses of the 13-state web architecture dataset) or up to three times (in the analyses of the two-state web architecture data) in Araneoidea. In all cases, there are examples of subsequent losses that are more granular than the scope of this work allows. In the latter, two of these events involved subsequent losses. Within Araneidae, multiple lineages have much reduced or absent capture webs, such as the genera *Chorizopes* and *Kaira*, while others have modified it beyond recognition, such as the bolas spiders (mastophorine araneids). The other more complicated case involves the tetragnathoid lineages, Mysmenidae, and Malkaridae. Most phylogenetic analyses of tetragnathoids indicate Arkyidae as sister lineage to Tetragnathidae, and Mimetidae as sister lineage to both. Arkyidae and Mimetidae both lack foraging webs (Framenau et al., 2010; Benavides et al., 2017), as do some lineages within Tetragnathidae (Gillespie, 1991) and all Malkaridae (Hormiga and Scharff, 2020). The topology of the closely related lineages Mysmenidae and Malkaridae are key to interpreting web evolution as well, given the former has orb webs and the latter has none. Previous analyses (e.g., some but not all found by Fernández et al. 2018a, 2018b) suggest a possible origin in the common ancestor of these five families, loss of the orb web, and a new origin in Tetragnathidae. Here, we find a single origin of the orb webs of

tetragnathids and mysmenids, and three independent losses of the orb web in Arkyidae, Mimetidae, and Malkaridae (Fig. 5).

Finally, it is interesting to note the differences in the inference of the number of origins of orb webs in Araneoidea when web architecture is scored in two versus 13 states. Lumping character states is expected to produce the same results if the scoring scheme with a higher number of states is adequately partitioning the observed phenotypic variability (e.g., Tarasov, 2019). Thus, our results support the idea expressed by us and others (see above) that we should continue to work on improving our understanding of web architecture and its variability.

*Conclusions*

More loci and more taxa have allowed us to considerably improve our understanding of the spider tree of life, corroborating established theories and sometimes proposing new relationships. Our analyses placed many lineages in a genomic-scale phylogenetic framework for the first time and continue to refute the single origin of orb webs within Araneae. The greater sophistication of phylogenetic and comparative methods is not without caveats. Historically robust node support measures seem to mask pervasively noisy signal, and coding and analytical variations in complex trait reconstructions can generate disparate outcomes. Nonetheless, we believe this work provides the foundation for the next steps in spider evolutionary studies.

FUNDING

This research was supported by US National Science Foundation grants DEB 1457300, 1457539 "Collaborative Proposal: Phylogeny and diversification of the orb weaving spiders (Orbiculariae, Araneae)" to GH and GG, and through multiple Putnam Expedition Grants from the Museum of Comparative Zoology to GG. DD received additional access to computational resources via the Norwegian Metacenter for Computational Science (NOTUR; project NN9601K), MA received additional support by the Spanish Ministry of Education, Culture, and Sports grant PRX15-00403 and the Catalan Government grant 2017SGR83. SK received support from the Harlan Fund of The George Washington University.

## ACKNOWLEDGEMENTS

Field and museum work in Australia were made possible through the generous help and support of Robert Raven, Barbara Baehr, Mark Harvey, Michael Rix, Jamie Seymour, Catherine Byrne, Simon Grove, Kirrily Moore and Graham Milledge. In New Zealand we were greatly helped by Cor Vink, Grace Hall, Phil Sirvid, and Barry Fitzgerald. We are grateful to Christopher K. Starr for his assistance in the fieldwork in Trinidad and Tobago. Carlos Víquez assisted us in the Costa Rican logistics and fieldwork. Abel Pérez-González collaborated during our fieldwork in Chile. We thank Fernando Álvarez-Padilla, Rosa Fernández, Caitlin Baker, and Julia Cosgrove for their aid in the field. We thank Akio Tanikawa, Marc Domènech, Jordi Moya, and Erika García for loaned tissues. We thank Sergei Tarasov for advice on running the SMM analyses. We acknowledge the computing resources granted by the Colonial One High Performance Computing Cluster at The George Washington University.

## REFERENCES

Álvarez-Padilla, F., Kallal, R. J., Hormiga, G. 2020. Taxonomy and phylogenetics of Nanometinae and other Australasian orb-weaving spiders (Araneae: Tetragnathidae). Bull. Am. Mus. Nat. Hist. 438, 1–107.

Ané, C., Larget, B., Baum, D. A., Dewitt Smith, S., Rokas, A. 2007. Bayesian estimation of concordance among gene trees. Mol. Biol. Evol. 24, 412–426. doi: 10.1093/molbev/msl170

Ballesteros, J. A., Hormiga, G. 2016. A new orthology assessment method for phylogenomic data: Unrooted Phylogenetic Orthology. Mol. Biol. Evol. 33, 2117–2134. doi: 10.1093/molbev/msw069

Beaulieu, J. M., Oliver, J. C., O'Meara, B. 2017. corHMM: Analysis of Binary Character Evolution. R package version 1.22. https://CRAN.R-project.org/package=corHMM

Benavides, L. R., Giribet, G., Hormiga, G. 2017. Molecular phylogenetic analysis of "pirate spiders" (Araneae, Mimetidae) with the description of a new African genus and the first report of maternal care in the family. Cladistics 33, 375–405. doi: 10.1111/cla.12174

Blackledge, T. A., Scharff, N., Coddington, J. A., Szüts, T., Wenzel, J. W., Hayashi, C. Y., Agnarsson, I. 2009. Reconstructing web evolution and spider diversification in the molecular era. Proc. Natl. Acad. Sci. USA. 106, 5529–5234. doi: 10.1073/pnas.0901377106

Bond, J. E., Garrison, N. L., Hamilton, C. A., Godwin, R. L., Hedin, M., Agnarsson, I. 2014. Phylogenomics resolves a spider backbone phylogeny and rejects a prevailing paradigm for orb web evolution. Curr. Biol. 24, 1765–1771. doi: 10.1016/j.cub.2014.06.034

Bott, R. A., Baumgartner, W., Bräunig, P., Menzel, F., Joel, A.-C. 2017. Adhesion enhancement of cribellate capture threads by epicuticular waxes of the insect prey sheds new light on spider web evolution. Proc. R. Soc. B 284, 20170363. doi: 10.1098/rspb.2017.0363

Brewer, M. S., Carter, R. A., Croucher, P. J., Gillespie, R. G. 2015. Shifting habitats, morphology, and selective pressures: developmental polyphenism in an adaptive radiation of Hawaiian spiders. Evolution 69, 162–178. doi: 10.1111/evo.12563.

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. doi: 10.1093/bioinformatics/btp348.

Cheng, D.-Q., Piel, H. W. 2018. The origins of the Psechridae: web-building lycosoid spiders. Mol. Phylogenet. Evol. 125, 213–219. doi: 10.1016/j.ympev.2018.03.035

Coddington, J. A., Agnarsson, I., Hamilton, C. A., Bond, J. E. 2019. Spiders did not repeatedly gain, but repeatedly lost, foraging webs. PeerJ 7, e6703. doi: 10.7717/peerj.6703

401 Dimitrov, D., Lopardo, L., Giribet, G., Arnedo, M. A., Álvarez-Padilla, F., Hormiga, G. 2012.
402 Tangled in a sparse spider web: single origin of orb weavers and their spinning work unravelled
403 by denser taxonomic sampling. Proc. R. Soc. B 279, 1341–1350. doi: 10.1098/rspb.2011.2011.
404
405 Dimitrov, D., Benavides, L. R., Arnedo, M. A., Giribet, G., Griswold, C. E., Scharff, N.,
406 Hormiga, G. 2017. Rounding up the usual suspects: a standard target-gene approach for
407 resolving the interfamilial relationships of ecribellate orb-weaving spiders with a new family-
408 rank. Cladistics 33, 221–250. doi: 10.1111/cla.12165.
409
410 Eberhard, W. G. 2018. Modular patterns in behavioural evolution: webs derived from orbs.
411 Behaviour 155, 1–35. doi: 10.1163/1568539X-00003502
412
413 Eberhard, W. G., Hazzi, N. A. 2017. Web building and prey wrapping behavior of *Aglaoctenus*
414 *castaneus* (Araneae: Lycosidae: Sosippinae). J. Arachnol. 45, 177–197. doi: 10.1636/JoA-S-16-
415 019.1
416
417 Eberle, J., Dimitrov, D., Valdez-Mondragón, A., Huber, B. A. 2018. Microhabitat change drives
418 diversification in pholcid spiders. BMC Evol. Biol. 18, 1–13. doi: 10.1186/s12862-018-1244-8
419
420 Enright, A. J., van Dongen, S., Ouzounis, C. A. 2002. An efficient algorithm for large-scale
421 detection of protein families. Nucleic Acids Res. 30, 1575–1584.
422
423 Faircloth, B. C. 2016. PHYLUCE is a software package for the analysis of conserved genomic
424 loci. Bioinformatics 32, 786–788.
425

Fernández, R., Hormiga, G., Giribet, G. 2014. Phylogenomic analysis of spiders reveals nonmonophyly of orb weavers. Curr. Biol. 24, 1772–1777. doi: 10.1016/j.cub.2014.06.03

Fernández, R., Kallal, R. J., Dimitrov, D., Ballesteros, J. A., Arnedo, M. A., Giribet, G., Hormiga, G. 2018a. Phylogenomics, diversification dynamics, and comparative transcriptomics across the Spider Tree of Life. Curr. Biol. 28, 1489–1497. doi: 10.1016/j.cub.2018.03.064.

Fernández, R., Kallal R. J., Dimitrov, D., Ballesteros, J. A., Arnedo, M. A., Giribet, G., Hormiga, G. 2018b. Phylogenomics, diversification dynamics, and comparative transcriptomics across the Spider Tree of Life (Correction). Curr. Biol. 28, 2190–2193. doi: 10.1016/j.cub.2018.06.018

Framenau, V. W., Scharff, N., Harvey, M. S. 2010. Systematics of the Australian orb-weaving spider genus *Demadiana* with comments on the generic classification of the Arkyinae (Araneae: Araneidae). Invertebr. Syst. 24, 139–171. doi: 10.1071/IS10005

French, A. S., Li, A. W., Meisner, S., Torkkel, P. H. 2014. Upstream open reading frames and Kozak regions of assembled transcriptome sequences from the spider *Cupiennius salei*. Selection or chance? Gene 539, 203–208. doi: 10.1016/j.gene.2014.01.079.

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. Part B 304b, 64–74.

Garrison, N. L., Rodriguez, J., Agnarsson, I., Coddington, J. A., Griswold, C. E., Hamilton, C. A., Hedin, M., Kocot, K. M., Ledford, J. M., Bond, J. E. 2016. Spider phylogenomics: untangling the Spider Tree of Life. PeerJ. 4, e1719. doi: 10.1016/j.cub.2018.03.064

Gillespie, R. G. 1991. Hawaiian spiders of the genus Tetragnatha: I. Spiny leg clade J Arachnol. 19, 174–209

Griswold, C., Coddington, J., Hormiga, G., Scharff, N. 1998. Phylogeny of the orb-web building spiders (Araneae, Orbiculariae: Deinopoidea, Araneoidea). Zool. J. Linn. Soc. Lond. 123, 1–99.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321. doi: 10.1093/sysbio/syq010.

Hedin, M. 2015. High-stakes species delimitation in eyeless cave spiders (*Cicurina*, Dictynidae, Araneae) from central Texas. Mol. Ecol. 24, 346–361. doi: 10.1111/mec.13036

Hedin, M., Derkarabetian, S., Ramírez, M. J., Vink, C., Bond, J. E. 2018. Phylogenomic reclassification of the world's most venomous spiders (Mygalomorphae, Atracinae), with implications for venom evolution. Sci. Rep. 8, 1636. doi: 10.1038/s41598-018-19946-2

Hedin, M., Derkarabetian, S., Alfaro, A., Ramírez, M. J., Bond, J. E. 2019. Phylogenomic analysis and revised classification of atypoid mygalomorph spiders (Araneae, Mygalomorphae), with notes on arachnid ultraconserved element loci. PeerJ 7, e6864. doi: 10.7717/peerj.6864

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B.-Q., Vinh, L. S. 2018a. UFBoot2: Improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35, 518–522. doi: 10.1093/molbe v/msx281

Hoang, D. T., Vinh, L. S., Flouri, T., Stamatakis, A., von Haeseler, A., Minh, B. Q. 2018b. MPBoot: fast phylogenetic maximum parsimony tree inference and bootstrap approximation. BMC Evol. Biol. 18, 11. doi: 10.1186/s12862-018-1131-3

Hormiga, G., Griswold, C. E. 2014. Systematics, phylogeny, and evolution of orb-weaving spiders. Annu. Rev. Entomol. 59, 487–512. doi: 10.1146/annurev-ento-011613-162046

Hormiga, G., Scharff, N. 2020. The malkarid spiders of New Zealand (Araneae, Malkaridae). Invertebr. Syst.s 34, 345–405. https://doi.org/10.1071/IS19073

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

491 Kalyaanamoorthy, S., Minh, B.-Q., Wong, T. K. F., von Haeseler, A., Jermiin, L. S. 2017.
492 ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–
493 589. doi: 10.1038/Nmeth.4285
494
495 Kallal, R. J., Fernández, R., Giribet, G., Hormiga, G. 2018. A phylotranscriptomic backbone of
496 the orb-weaving spider family Araneidae (Arachnida, Araneae) supported by multiple
497 methodological approaches. Mol. Phylogenet. Evol. 126, 129–140. doi:
498 10.1016/j.ympev.2018.04.007.
499
500 Katoh, K., Standley, D. M. 2013. MAFFT multiple sequence alignment software version 7:
501 Improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. doi:
502 10.1093/molbev/mst010
503
504 Kulkarni, S. S., Wood, H. M., Lloyd, M., Hormiga, G. 2019. Spider-specific probe set for
505 ultraconserved elements offers new perspectives on the evolutionary history of spiders
506 (Arachnida, Araneae). Mol. Ecol. Res. 20, 185-203. doi: 10.1111/1755-0998.13099
507
508 Kumar, S., Filipski, A. J., Battistuzzi, F. U., Kosakovksy Pond, S. L., Tamura, K. 2012. Statistics
509 and truth in phylogenomics. Mol. Biol. Evol. 29, 457–472. doi: 10.1093/molbev/msr202.
510
511 Kuntner, M., Hamilton, C. A., Cheng, R.-C., Gregorič, M., Lupse, N., Lokovsek, T., Lemmon, E.
512 M., Lemmon, A. R., Agnarsson, I., Coddington, J. A., Bond, J. E. 2019. Golden orbweavers

ignore biological rules: phylogenomic and comparative analyses unravel a complex evolution of sexual size dimorphism. Syst. Biol. 68, 555–572. doi: 10.1093/sysbio/syy082.

Laumer, C. E, Fernández, R., Lemer, S., Combosch, D., Kocot, K. M., Riesgo, A., Andrade, S. C. S., Sterrer, W., Sørensen, M. V., Giribet, G. 2019. Revisiting metazoan phylogeny with genomic sampling of all phyla. Proc. R. Soc. B 286, 20190831. doi: 10.1098/rspb.2019.0831

Lopardo, L., Giribet, G., Hormiga, G. 2010. Morphology to the rescue: molecular data and the signal of morphological characters in combined phylogenetic analyses—
a case study from mysmenid spiders (Araneae, Mysmenidae), with comments on the evolution of web architecture. Cladistics 26, 1–52. doi: 10.1111/j.1096-0031.2010.00332.x

Magalhães, I. L. F., Azevedo, G. H. F., Michalik, P., Ramírez, M. J. 2020. The fossil record of spiders revisited: implications for calibrating trees and evidence for a major faunal turnover since the Mesozoic. Biol. Rev. 95, 184–217. doi:10.1111/brv.12559

Meng, X., Zhang, Y., Bao, H., Liu, Z. 2015. Sequence analysis of insecticide action and detoxification-related genes in the insect pest natural enemy *Pardosa pseudoannulata*. PLoS One 10, e0125242. doi:10.1371/journal.pone.0125242

Michalik, P., Kallal, R. J., Dederichs, T. M., Labarque, F. M., Hormiga, G., Giribet, G., Ramírez, M. J. 2019. Phylogenomics and genital morphology of cave raptor spiders (Araneae,

535 Trogloraptoridae) reveal an independent origin of a flow-through genital system. J. Zool. Syst.
536 Evol. Res. 57, 737–747. doi: 10.1111/jzs.12315
537
538 Minh, B. Q., Hanh, M. W., Lanfear, R. 2020. New methods to calculate concordance factors for
539 phylogenomic datasets.. Mol. Biol. Evol. doi:10.1093/molbev/msaa106
540
541 Mirarab, S., Reaz, R., Bayzid, M. S., Zimmermann, T., Swenson, M. S., Warnow, T. 2014.
542 ASTRAL: Genome-scale coalescent-based species tree estimation. Bioinformatics 30, 1541–
543 1548. doi: 10.1093/bioinformatics/btu462
544
545 Nguyen, L. T., Schmidt, H. A., von Haeseler, A., Minh, B.-Q. 2015. IQTREE: a fast and
546 effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol.
547 32, 268–274. doi: 10.1093/molbe v/msu300
548
549 Opatova, V., Hamilton, C. A., Hedin, M., Montes de Oca, L., Král, J., Bond, J. E. in press.
550 Phylogenetic systematics and evolution of the spider infraorder Mygalomorphae using genomic
551 scale data. Syst. Biol. doi: 10.1093/sysbio/syz064
552
553 Opell, B,D. 2013. Cribellar thread. In *Spider Ecophysiology*, ed. W. Nentwig, pp. 303–18.
554 Berlin/Heidelberg: Springer.
555
556 Paradis, E., Schliep, K. 2019. ape 5.0: an environment for modern phylogenetics and
557 evolutionary analyses in R. Bioinformatics 35, 526–528. doi: 10.1093/bioinformatics/bty633

Philippe, H., de Vienne, D. M., Ranwez, V., Roure, B., Baurain, D., Delsuc, F. 2017. Pitfalls in supermatrix phylogenomics. Eur. J. Taxon. 283, 1–25. doi: 10.5852/ejt.2017.283

Piacentini, L. N., Ramírez, M. J. 2019. Hunting the wolf: a molecular phylogeny of the wolf spiders (Araneae, Lycosidae). Mol. Phylogenet. Evol. 136, 227–240. doi: 10.1016/j.ympev.2019.04.004

R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ramírez, M. J., Grismado, C. J., Ubick, D., Ovtcharenko, V. I., Cushing, P. E., Platnick, N. I., Wheeler, W. C., Prendini, L., Crowley, L. M., Horner, N. V. 2019. Myrmecicultoridae, a new family of myrmecophilic spiders from the Chihuahuan Desert (Araneae, Entelegynae). Am. Mus. Novit. 3930, 1–24. doi: 10.1206/3930.1

Rix, M. G., Cooper, S. J. B., Meusemann, K., Klopfstein, S., Harrison, S. E., Harvey, M. S., Austin, A. D. 2017. Post-Eocene climate change across continental Australia and the diversification of Australasian spiny trapdoor spiders (Idiopidae: Arbanitinae). Mol. Phylogenet. Evol. 109, 302–320. doi: 10.1016/j.ympev.2017.01.008

Robinson, D. F., Foulds, L. R. 1981. Comparison of phylogenetic trees. Math. Biosci. 53, 131–147. doi: 10.1016/0025-5564(81)90043-2

Sanggaard, K. W., Bechsgaard, J. S., Fang, X., Duan, J., Dyrlund, T. F., Gupta, V., Jiang, X., Cheng, L., Fan, D., Feng, Y., Han, L., Huang, Z., Wu, Z., Liao, L., Settepani, V., Thøgersen, I. B., Vanthournout, B., Wang, T., Zhu, Y., Funch, P., Enghild, J. J., Schauser, L., Andersen, S. U., Villesen, P., Schierup, M. H., Bilde, T., Wang, J. 2014. Spider genomes provide insight into composition and evolution of venom and silk. Nat. Comm. 5, 3765. doi: 10.1038/ncomms4765

Sayyari, E., Mirarab, S. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Mol. Biol. Evol. 33, 1654–1668. doi: 10.1093/molbev/msw079

Scharff. N., Coddington. J. A., Blackledge, T. A., Agnarsson, I., Framenau, V., Szüts, T., Hayashi, C. Y., Dimitrov, D. (2020). Phylogeny of the orb-weaving spider family Araneidae (Araneae, Araneoidea). Cladistics 36, 1–21. doi: 10.1111/cla.12382

Shao, L., Li, S. 2018. Early Cretaceous greenhouse pumped higher taxa diversification in spiders. Mol. Phylogenet. Evol. 127, 146–155. doi: 10.1016/j.ympev.2018.05.026

Sharma, P. P., Kaluziak, S. T., Pérez-Porro, A. R., González, V. L., Hormiga, G., Wheeler, W. C., Giribet, G. 2014. Phylogenomic interrogation of arachnida reveals systemic conflicts in phylogenetic signal. Mol. Biol. Evol. 31, 2963–2984. doi: 10.1093/molbev/msu235

Shear, W. A. 1986. Spiders: webs, behavior, and evolution. Stanford, California, USA: Stanford University Press.

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

Smith, S. A., O'Meara, B. C. 2012. treePL: divergence time estimation using penalized likelihood for large phylogenies. Bioinformatics 28, 2689–2690. doi: 10.1093/Bioinformatics/Bts492

Song, L., Florea, L. 2015. Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. GigaScience 4, 48. doi: 10.1186/s13742-015-0089-y

Susko, E., Roger, A. J. 2020. On the use of information criteria for model selection in phylogenetics. Mol. Biol. Evol. 37, 549–562. doi: 10.1093/molbev/msz228

Tamura, K., Battistuzzi, F. U., Billing-Ross, P., Murillo, O., Filipski, A., Kumar, S. 2012. Estimating divergence times in large molecular phylogenies. Proc. Natl. Acad. Sci. USA 109, 19333–19338. doi: 10.1073/pnas.1213199109

Tan, G., Muffato, M., Ledergerber, C., Herrero, J., Goldman, N., Gil, M., Dessimoz, C. 2015. Current methods for automated filtering of multiple sequence alignments frequently worsen single-gene phylogenetic inference. Syst. Biol. 64, 778–791. doi: 10.1093/sysbio/syv033

627 Tarasov, S. 2019. Integration of anatomy ontologies and evo-devo using structured Markov
628 models suggests a new framework for modeling discrete phenotypic traits. Syst. Biol. 68, 698–
629 716. doi: 10.1093/sysbio/syz005
630
631 van Dongen, S. 2000. Graphs clustering by flow simulation. Ph.D. thesis. University of Utrecht,
632 Utrecht.
633
634 Wang, B., Dunlop, J. A., Selden, P. A., Garwood, R. J., Shear, W. A., Müller, P., Lei, X. 2018.
635 Cretaceous arachnid *Chimerarachne yingi* gen. et. sp. nov. illuminates spider origins. Nat. Ecol.
636 Evol. 2, 614–622. doi: 10.1038/s41559-017-0449-3
637
638 Wheeler, W. C., Coddington, J. A., Crowley, L. M., Dimitrov, D., Goloboff, P. A., Griswold, C.
639 E., Hormiga, G., Prendini, L., Ramírez, M. J., Sierwald, P., Almeida-Silva, L., Álvarez-Padilla,
640 F., Arnedo, M. A., Benavides Silva, L. R., Benjamin, S. P., Bond, J. E., Grismado, C. J., Hasan,
641 E., Hedin, M., Izquierdo, M. A., Labarque, F. M., Ledford, J., Lopardo, L., Maddison, W. P.,
642 Miller, J. A., Piacentini, L. N., Platnick, N. I., Polotow, D., Silva-Dávila, D., Scharff, N., Szüts,
643 T., Ubick, D., Vink, C. J., Wood, H. M., Zhang, J. X. 2017. The spider tree of life: phylogeny of
644 Araneae based on target-gene analyses from an extensive taxon sampling. Cladistics 33, 574–
645 616. doi: 0.1111/cla.12182.
646
647 Wolff, J. O., Paterno, G. B., Liprandi, D., Ramírez, M. J., Bosia, F., van der Meijden, A.,
648 Michalik, P., Smith, H. M., Jones. B. R., Ravelo, A. M., Pugno, N., Herberstein, M. E. 2019.

649 Evolution of aerial spider webs coincided with repeated structural optimization of silk
650 anchorages. Evolution 73, 2122–2134. doi: 10.1111/evo.13834
651
652 Wood, H. M., González, V. L., Lloyd, M., Coddington, J. A., Scharff, N. 2018. Next-generation
653 museum genomics: phylogenetic relationships among palpimanoid spiders using sequence
654 capture techniques (Araneae: Palpimanoidea). Mol. Phylogenet. Evol. 127, 907–918. doi:
655 10.1016/j.ympev.2018.06.038.
656
657 Wood, H. M., Griswold, C. E., Gillespie, R. G. 2012. Phylogenetic placement of pelican spiders
658 (Archaeidae, Araneae), with insight into the evolution of the "neck" and predatory behaviours of
659 the superfamily Palpimanoidea. Cladistics 28, 598–626. doi: 10.1111/j.1096-0031.2012.00411.x
660
661 World Spider Catalog. 2020. World Spider Catalog. Version 20.5. Natural History Museum
662 Bern. Available from: URL http://wsc.nmbe.ch (accessed on 8 January 2020). doi: 10.24436/2.
663
664 Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24,
665 1586–1591. doi: 10.1093/molbev/msm088.
666
667 Zhang, Z.-Q. 2011 Animal biodiversity: an introduction to higher-level classification and
668 taxonomic richness. Zootaxa 12, 7–12.
669

Zhao, Y. J., Zeng, Y., Chen, L., Dong, Y., Wang, W. 2014. Analysis of transcriptomes of three orb-web spider species reveals gene profiles involved in silk and toxin. Insect Sci. 21, 687–698. doi: 10.1111/1744-7917.12068.

693 **Figure 1.** A sample of the diversity of webs and foraging strategies in araneomorph spiders. (a),
694 *Deinopis* sp. (Deinopidae) from Madagascar, a cribellate weaver with a highly modified orb
695 web. (b), The ecribellate orb web of *Ocrepeira darlingtoni* (Araneidae), from the Dominican
696 Republic. (c), The modular vertical web of *Synotaxus* sp. (Synotaxidae), from Brazil. (d), A
697 typical web of Linyphiidae, from Taiwan. (e), The cribellate orb web of an uloborid from
698 Australia. (f), The ecribellate orb web of *Maxanapis* sp. (Anapidae), from Australia. (g),
699 *Exechocentrus lancearius* (Araneidae), a bolas spider from Madagascar. (h), The sheet web of
700 *Runga* sp. (Physoglenidae) from New Zealand. (i), A typical aerial sheet web of a cyatholipid
701 from Australia. (j), The characteristic tent web of *Oecobius* sp. (Oecobiidae), from Tobago. (k),
702 The cribellate web of *Paramatachia* sp. (Desidae), from Australia, a member of the RTA Clade.
703 (l), The highly modified orb web of a mysmenid from Madagascar. (m), The webless
704 plapimanoid spider *Eriauchenius workmani* (Archaeidae) feeding on a theridiid spider in
705 Madagascar (all photos by G. Hormiga).
706
707 **Figure 2**. Spider tree of life, non-araneoid and close relatives. Derived from BUSCO orthology,
708 67% occupancy, trimmed using trimAl, and maximum likelihood analysis using IQ-TREE. (a),
709 arachnid outgroup taxa, Mesothelae, Mygalomorphae, and non-entelegyne araneomorphs. (b)
710 UDOH Grade and RTA Clade. (c ) full tree highlighting regions (a) and (b).
711
712 **Figure 3.** Phylogenetic relationships of araneoids and close relatives. Derived from BUSCO
713 orthology, 67% occupancy, trimmed using trimAl, and maximum likelihood analysis using IQ-
714 TREE. (a) Eresidae, Nicodamoidea, and Araneoidea, derived from BUSCO orthology, 67%
715 occupancy, trimmed using trimAl, and maximum likelihood analysis using IQ-TREE. (b) full

716 tree highlighting eresids, nicodamoids, and araneoids. (c-e), Different topologies generated by
717 UPhO orthology, 33% occupancy, trimmed using trimAl, and maximum likelihood analysis
718 using IQ-TREE. (c ), Malkarid, mysmenid, and tetragnathoid clade. (d), Physoglenid, synotaxid,
719 and nesticid clade. (e), Araneid lineage except Phonognathinae and Nephilinae.
720
721 **Figure 4.** Calibrated phylogram using treePL, with major groups highlighted. Stars represent
722 fossil placements. Stars inside collapsed clades represent at least one calibration point within that
723 clade.
724
725 **Figure 5.** Ancestral state reconstruction of foraging webs and their architecture using the
726 structured Markov models. The results shown here are based on the 13 states scoring scheme for
727 the web architecture and a 2-rates model which resulted in the best BIC score. Differences with
728 reconstructions based on 2 states scoring scheme for web architecture are shown in the top right
729 left corner (based on the best scoring model – a 2-rates model with one hidden state for the orb
730 web state; only the Araneoidea clade is depicted here). Full versions of these results with tip
731 labels are provided as supplementary materials on Harvard Dataverse.
732
733 **Table 1.** Matrix construction and analyses. Where the number of sites is presented as a single
734 value, it is the trimmed matrix; two values signify trimmed and untrimmed version of the matrix.
735
736 **Table 2**. Selected clades relevant to Araneoidea and their support values. Analytical
737 abbreviations: Method B, BUSCO orthology on the *all* dataset with 67% occupancy; Method U,
738 UPhO orthology on the *ara* dataset with 33% occupancy; UFB (Tr/Un), ultrafast bootstrap

739 support on trimmed and untrimmed matrices; SH-aLRT, Shimodaira-Hasegawa approximate
740 likelihood ratio test; gCF, gene concordance factor; sCF, site concordance factor; LPP (Tr/Un),
741 local posterior probability on trimmed or untrimmed matrix. In the latter, where the ASTRAL
742 topology differs from that of the maximum likelihood topology, that alternative is given with its
743 supports. Taxon abbreviations: Ana, Anapidae; ARA, araneoids; Ara, Araneidae; Ere, Eresidae;
744 Mal, Malkaridae; Mys, Mysmenidae; Nes, Nesticidae; NIC, nicodamoids; Phy, Physoglenidae;
745 Sym, Symphytognathidae; Syn, Synotaxidae; Tet, tetragnathoids; The, Theridiosomatidae.