MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2016) doi: 10.1111/1755-0998.12621

High phylogenetic utility of an ultraconserved element probe set designed for Arachnida

JAMES STARRETT,*[1,2] SHAHAN DERKARABETIAN,*†[2] MARSHAL HEDIN,* ROBERT W. BRYSON JR.,‡§ JOHN E. MCCORMACK§ and BRANT C. FAIRCLOTH¶
*Department of Biology, San Diego State University, 5500 Campanile Drive, San Diego CA 92182, USA, †Department of Biology, University of California, Riverside, 900 University Avenue, Riverside CA 92521, USA, ‡Department of Biology and Burke Museum of Natural History and Culture, University of Washington, 4331 Memorial Way Northeast, Seattle WA 98195, USA, §Moore Laboratory of Zoology, Occidental College, Los Angeles CA 90041, USA, ¶Department of Biological Sciences and Museum of Natural Science, Louisiana State University, Baton Rouge LA 70803, USA

Abstract

Arachnida is an ancient, diverse and ecologically important animal group that contains a number of species of interest for medical, agricultural and engineering applications. Despite their importance, many aspects of the arachnid tree of life remain unresolved, hindering comparative approaches to arachnid biology. Biologists have made considerable efforts to resolve the arachnid phylogeny; yet, limited and challenging morphological characters, as well as a dearth of genetic resources, have hindered progress. Here, we present a genomic toolkit for arachnids featuring hundreds of conserved DNA regions (ultraconserved elements or UCEs) that allow targeted sequencing of any species in the arachnid tree of life. We used recently developed capture probes designed from conserved regions of available arachnid genomes to enrich a sample of loci from 32 diverse arachnids. Sequence capture returned an average of 487 UCE loci for all species, with a range from 170 to 722. Phylogenetic analysis of these UCEs produced a highly resolved arachnid tree with relationships largely consistent with recent transcriptome-based phylogenies. We also tested the phylogenetic informativeness of UCE probes within the spider, scorpion and harvestman orders, demonstrating the utility of these markers at shallower taxonomic scales and suggesting that these loci will be useful for species-level differences. This probe set will open the door to phylogenomic and population genomic studies across the arachnid tree of life, enabling systematics, species delimitation, species discovery and conservation of these diverse arthropods.

Keywords: Araneae, arthropods, Opiliones, phylogenomics, population genomics, Scorpiones, target enrichment, ultraconserved elements

Received 28 July 2016; revision received 24 September 2016; accepted 4 October 2016

Introduction

Arachnida is an extremely ancient and diverse arthropod lineage, including conspicuous taxa such as spiders, scorpions, mites and ticks. The oldest known arachnid fossils consist mostly of scorpions and extinct trigonotarbids from the Silurian period, 443–416 MYA (million years ago; Laurie 1899; Jeram *et al*. 1990; Dunlop 1996; Dunlop *et al*. 2008). More than 110 000 species of arachnids have been described, with spiders and mites ranking among the most diverse of all animal orders (Harvey 2002; Zhang 2011). Yet, more than half of arachnid species diversity remains to be discovered (Chapman 2009). Arachnida also contains a number of medically important venomous or disease-vector species and many important agricultural mite pests, while spiders are of particular interest to biologists and engineers for the strong and elastic silk fibres they produce. Despite the attention arachnids have received for their ecological importance and practical utility to humans, phylogenetic relationships among and within many arachnid orders remain uncertain. At the root of this problem is that morphological characters are limited and difficult to interpret (Shultz 2007), and genomic resources for this group are sparse. Adding to the difficulty is the uncertainty in the rooting of the arachnid tree, with fossil,

Correspondence: Shahan Derkarabetian, Department of Biology, San Diego State University, 5500 Campanile Drive, San Diego CA 92182, USA; E-mail: sderkarabetian@gmail.com

[1]Present address: Department of Biological Sciences, Auburn University, 101 Rouse Life Sciences, Auburn AL 36849, USA
[2]Equal contribution.

morphological and molecular data recovering drastic discrepancies in early arachnid relationships and placing traditional nonarachnid chelicerates (e.g. *Limulus*) within Arachnida (Wheeler & Hayashi 1998; Masta *et al.* 2009; Regier *et al.* 2010; Sharma *et al.* 2014). Genomewide phylogenetic markers are essential for resolving deep relationships within Arachnida and for helping to uncover the numerous arachnid species that await discovery.

Ultraconserved elements (UCEs) provide one potential target for a universal set of genomic markers for arachnids and would allow researchers to collect genomic information from diverse taxa across the arachnid tree of life. UCEs are segments of DNA that are highly conserved across divergent taxa (Bejerano *et al.* 2004) and are thought to regulate and/or enhance gene expression (Alexander *et al.* 2010). UCEs can also be used as anchors to target and retrieve variable DNA sequence flanking the core UCE regions. This flanking DNA shows a trend of increasing genetic variability as distance from the core UCE increases (Faircloth *et al.* 2012). UCEs are ideal markers for molecular systematics for several reasons. Whereas transcriptomes require high-quality RNA as input, the UCE protocol only requires DNA, and enrichments can be performed using relatively low starting DNA concentrations. This allows the method to be extended to small-bodied taxa, large collections of specimens preserved for Sanger-based DNA research and even to 'standard' museum specimens with varying levels of DNA degradation (McCormack *et al.* 2015; Blaimer *et al.* 2016). Homology between UCE loci across divergent taxa is also easy to assess because the core UCE region often displays >95% sequence similarity, and UCE cores are rarely duplicated (Derti *et al.* 2006). Although core UCE regions show reduced sequence variation, UCE-flanking DNA shows levels of phylogenetic informativeness equal to or exceeding that of traditionally used protein-coding markers (Gilbert *et al.* 2015).

Ultraconserved elements have been successfully used in phylogenetic studies at multiple taxonomic levels, including recent divergences (~5 MYA, within species; Smith *et al.* 2014; Harvey *et al.* 2016; Manthey *et al.* 2016) and ancient divergences (e.g. all amniotes; Faircloth *et al.* 2012). Originally developed for use in tetrapods, the vast majority of UCE phylogenetic studies have been conducted on vertebrate taxa, including Amniota (Faircloth *et al.* 2012), mammals (McCormack *et al.* 2012), birds (McCormack *et al.* 2013; Smith *et al.* 2014; Meiklejohn *et al.* 2016), reptiles (Crawford *et al.* 2012, 2015; Streicher *et al.* 2016) and fish (Faircloth *et al.* 2013). More recently, UCE probe sets have been developed for use in insect taxa, including Hymenoptera (Faircloth *et al.* 2015), Coleoptera, Diptera, Lepidoptera and Hemiptera (Faircloth 2016).

Here, we performed an in vitro test of a recently developed bait set targeting arachnid UCEs (Faircloth 2016). We then demonstrated the phylogenetic utility of this probe set at multiple evolutionary timescales, from those spanning hundreds of millions of years to species-level divergences of less than 10 million years. We reconstructed well-supported phylogenies between and within arachnid orders and demonstrated that these UCE baits may also be useful when reconstructing species-level relationships.

Materials and methods

Ultraconserved element library construction and enrichment

Ultraconserved element capture was tested using the principal arachnid bait set (Faircloth 2016) on 32 arachnid samples representing six orders (Table 1). The orders selected span the root of Arachnida (Regier *et al.* 2010; Sharma *et al.* 2014). Within three orders (Araneae, Opiliones and Scorpiones), taxa were selected from major lineages, including across the internal root for each (Hedin *et al.* 2012; Bond *et al.* 2014; Sharma *et al.* 2015). To assess UCE variability between closely related taxa, we included two turret spiders from the *Antrodiaetus riversi* complex (Hedin *et al.* 2013), two congeneric harvestmen (*Briggsus pacificus* and *B. bilobatus*) and a second bark scorpion *Centruroides sculpturatus* to complement the published *C. exilicauda* genome. Voucher specimens are deposited in the San Diego State University Terrestrial Arthropod Collection (SDSU_TAC). In the case of small arachnids (e.g. mites and ticks), whole specimens were used in extractions, but vouchers from the same locality are deposited in the SDSU_TAC.

Genomic DNA was extracted from legs or whole specimens using the DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA) following the manufacturer's protocol. DNA concentrations were determined with a Qubit fluorometer (Life Technologies, Inc) and run out on a 1% agarose gel to assess quality. The lowest starting DNA quantity for processing was 108 ng (*Siro acaroides*, a small-bodied harvestmen typically <2 mm in length), while most samples started with approximately 500 ng (Table 1). Samples with high molecular weight DNA were fragmented with a QSonica Q800R sonicator for 10–12 cycles of 20 s on/20 s off, resulting in fragments predominantly in the range of 300–1000 bp. DNA from sample AR014 (*Neomolgus littoralis*) was partially degraded and not sonicated.

Libraries were prepared with the KAPA Hyper Prep Kit (Kapa Biosystems), with a generic SPRI substitute used for bead clean-up steps (Rohland and Reich 2012; Glenn *et al.* 2016). Universal adapters were ligated onto end-repaired and A-tailed DNA fragments. Each adapter-ligated library was amplified in a 50-μL total reaction

Table 1 Sample information and sequencing statistics

Order	Species	Voucher	DNA (ng)	Pool	Raw read pairs	Reads pass QC	% Pass QC	Contigs	UCE loci
Acariformes	*Neomolgus littoralis*	AR014	500	G	633 368	595 582	94.03	26 192	170
Amblypygi	*Damon diadema*	AR010	500	F	1 623 520	1 484 780	91.45	26 761	442
Araneae	*Antrodiaetus riversi*	AR17	494	A	1 277 026	1 216 925	95.29	39 879	576
Araneae	*Antrodiaetus riversi*	MY1253	360	A	1 240 653	1 171 503	94.43	28 151	571
Araneae	*Calisoga sp.*	MY4085	500	A	1 283 059	1 209 706	94.28	37 800	664
Araneae	*Diguetia signata*	G0580	500	C	2 109 935	1 988 677	94.25	76 172	504
Araneae	*Euagrus chisoseus*	MY4295	500	A	1 672 688	1 592 368	95.20	46 272	674
Araneae	*Habronattus tarsalis*	HA0955	500	C	1 016 446	957 507	94.20	30 683	484
Araneae	*Hebestatis theveneti*	MY4392	500	A	1 327 914	1 257 588	94.70	34 651	722
Araneae	*Hypochilus pococki*	H0595	500	C	1 143 368	1 081 293	94.57	32 241	552
Araneae	*Kukulcania* sp.	G0551	500	C	1 032 516	984 284	95.33	32 448	466
Araneae	*Liphistius malayanus*	MIS1	500	A	832 843	781 359	93.82	26 169	557
Araneae	*Megahexura fulva*	MY4378	500	A	1 680 701	1 602 923	95.37	60 357	565
Araneae	*Nesticus gertschi*	N0806	456	B	1 354 456	1 238 198	91.42	19 918	543
Araneae	*Oecobius navus*	GAR4	280	C	1 596 815	1 523 543	95.41	39 711	539
Araneae	*Porrhothele* sp.	MY0857	500	A	1 121 499	1 046 279	93.29	33 797	628
Opiliones	*Bishopella laciniosa*	OP0569	500	D	1 486 862	1 385 611	93.19	42 527	329
Opiliones	*Briggsus bilobatus*	OP3559	216	E	1 148 872	1 093 882	95.21	26 325	406
Opiliones	*Briggsus pacificus*	OP3625	320	D	1 355 200	1 267 279	93.51	50 366	399
Opiliones	*Fumontana deprehendor*	OP0623	432	D	1 265 798	1 174 121	92.76	28 841	356
Opiliones	*Leiobunum calcar*	OP1089	500	D	1 315 511	1 242 038	94.41	47 403	346
Opiliones	*Metanonychus n. nigricans*	OP3704	247	D	1 313 727	1 227 411	93.43	43 275	377
Opiliones	*Sabacon cavicolens*	OP1518	500	D	990 843	907 152	91.55	18 060	293
Opiliones	*Siro acaroides*	OP3383	108	D	1 260 358	1 179 969	93.62	33 373	369
Scorpiones	*Bothriurus keyserlingi*	AR021	500	I	1 238 117	1 183 837	95.62	23 438	500
Scorpiones	*Centruroides sculpturatus*	AR015	500	H	3 913 560	3 750 267	95.83	100 622	706
Scorpiones	*Diplocentrus peloncillensis*	AR016	500	H	533 530	507 517	95.12	14 225	402
Scorpiones	*Hadrurus arizonensis*	AR017	500	H	734 122	700 561	95.43	17 206	565
Scorpiones	*Paravaejovis spinigerus*	AR018	500	H	1 035 158	990 302	95.67	22 872	478
Thelyphonida	*Mastigoproctus giganteus*	AR011	500	F	926 004	871 445	94.11	22 195	431
				Mean	**1 315 482.3**	**1 240 463.57**	**94.22**	**36 064.33**	**487.13**

DNA refers to starting quantity that was processed. Voucher specimens are housed in the San Diego State University Terrestrial Arthropod Collection (SDSU_TAC).

volume, which consisted of 15 μL of adapter-ligated DNA, 1X KAPA HiFi HotStart ReadyMix and 5 μM of each Illumina TruSeq dual-indexed primer (i5 and i7) with modified 8-bp indexes (Glenn *et al.* 2016). Amplification conditions were 98 °C for 45 s, followed by 10–12 cycles of 98 °C for 15 s, 60 °C for 30 s and 72 °C for 60 s and then a final extension of 72 °C for 60 s. After bead clean-up of amplified libraries, equimolar amounts of libraries were combined into 1000 ng total pools. Pool combinations ranged from 1 to 8 individual libraries (Table 1). Due to a low concentration following library preparation and amplification, the pool containing sample AR014 (*Neomolgus littoralis*) included a lower amount of library (108.9 ng) for this specimen compared with the other two libraries that we included (333 ng each).

Target enrichment of libraries was performed using a MYbaits custom kit (MYcroarray, Inc.) following the Target Enrichment of Illumina Libraries v. 1.5 protocol (http://ultraconserved.org/#protocols). Custom TruSeq adapter blockers (Glenn *et al.* 2016) and standard MYbaits blockers were annealed to 147 ng/uL library pools, followed by hybridization to the master arachnid bait set (Faircloth 2016). Hybridizations were performed at 65 °C for 24 h. After hybridization, library pools were bound to Dynabeads MyOne Streptavidin C1 magnetic beads (Life Technologies) for enrichment. We performed with-bead PCR recovery of the posthybridization enrichments in a 50-μL reaction volume consisting of 15 μL enriched DNA, 1X KAPA HiFi HotStart ReadyMix and 5 μM each of TruSeq forward and reverse primers. Amplification conditions were 98 °C for 45 s, followed by 18 cycles of 98 °C for 15 s, 60 °C for 30 s and 72 °C for 60 s and then a final extension of 72 °C for 5 min. Following PCR recovery, libraries were quantified using

a Qubit fluorometer and diluted to 5 ng/μL. We performed qPCR quantification of enriched library pools, and prepared a 10 μM mix of each pool at equimolar ratios. We sequenced the library pool using a partial run of paired-end 150-bp sequencing on an Illumina NextSeq (Georgia Genomics Facility).

Read processing, contig assembly and matrix creation

Raw read data were processed using the PHYLUCE pipeline (Faircloth 2015). Adapter removal and quality control trimming were conducted using the ILLUMIPROCESSOR wrapper (Faircloth 2013) using default values. Reads were assembled using TRINITY version r2013-02-25 (Grabherr *et al.* 2011). Contigs from all samples were matched to probes using minimum coverage and minimum identity values of 65. We additionally extracted UCE loci in silico from available arachnid genomes and *Limulus polyphemus*. UCE loci were aligned using MAFFT (Katoh & Standley 2013) and trimmed with GBLOCKS (Castresana 2000; Talavera & Castresana 2007) as implemented in the PHYLUCE pipeline. Multiple data sets were created for downstream analyses. First, a '2perArachnid' data set was created that contained up to two representative samples from each arachnid order sequenced, plus *Limulus* as an outgroup. This included UCE data extracted from both novel samples and previously deposited genomes. Representative samples were chosen to span the root node of the relevant arachnid orders and based on the number of UCE loci recovered. Because the purpose of our study was not to reconstruct arachnid phylogeny, we did not include all arachnid orders. Second, we assembled a 'UCEsample' data set that included only samples newly sequenced for this study. Third, three individual data sets were created that included all samples from within the orders Araneae, Opiliones and Scorpiones. For each of these individual-order data sets, two matrices were created, one including the amblypygid *Damon* as the outgroup and a second without an outgroup. Data matrices without outgroups were used for determining matrix statistics, with the number of parsimony-informative characters computed using PAUP* 4.0 (Sinauer Associates, Inc.). Finally, to assess species-level utility of UCEs, three congeneric data sets were created for *Antrodiaetus* (turret spiders), *Briggsus* (Briggs' harvestmen) and *Centruroides* (bark scorpions).

Phylogenetic analysis

Data matrices including an outgroup taxon were also subjected to phylogenetic analyses using RAXML HPC v8.0 (Stamatakis 2014), implementing the rapid bootstrap algorithm (Stamatakis *et al.* 2008) plus ML tree search option (-f a), 200 bootstrap replicates and the GTRGAMMA model. ML analyses were conducted on three matrices for each concatenated data set, which consisted of taxon coverages of 90%, 70% and 50% for each locus. Topologies and support scores based on 70% and 90% taxon coverage were nearly identical to those from 50% taxon coverage, and we show matrix statistics for only the 50% data sets and trees for the 50% and 90% data sets. For Araneae, Opiliones and Scorpiones, we conducted coalescent-based analyses with ASTRAL version 4.10.8 (Mirarab *et al.* 2014; Mirarab & Warnow 2015). Here, trees for each locus were generated in RAXML with 500 bootstrap replicates from the 90%, 70% and 50% taxon coverage data sets. All analyses were conducted on a late 2015 iMac with a 4-GHz Intel i7 processor, with the exception of contig assembly with TRINITY, which was run on a 12 core CentOS linux machine with 48 GB of RAM.

Results

Sequencing results, assembly statistics and the number of UCE loci recovered are presented in Table 1 for samples sequenced in vitro in this study and in Table S1 (Supporting information) for published genomes used in silico. On average, we produced 1 315 482 raw reads per sample, with an average of 1 240 464 reads (94.2%) passing quality control. Assemblies resulted in an average of 35 884 contigs per sample. The number of UCE loci recovered from newly sequenced samples varied between 170 and 722 (average = 487), while the average recovered in silico from published genomes was 675, including 555 UCEs from the outgroup *Limulus*. The average number of recovered UCE loci differed among orders (unpaired two-tailed *t*-test, $t = 6.4308$, df = 38, *P*-value <0.0001), being higher in groups from which the probes were designed (Araneae, Parasitiformes, Scorpiones: 605 loci) versus those that were not used in probe design (Acariformes, Amblypygi, Opiliones, Thelyphonida: 376 loci). We tested whether the number of samples included in a hybridization pool influenced the number of UCE loci recovered and found no correlation ($R^2 = 0.03$), although the Opiliones-only pools recovered the fewest loci (Fig. S1, Supporting information). Matrix statistics are presented for a 50% taxon coverage data set in Table 2. We recovered identical topologies and nearly identical support scores for matrices with 70% taxon coverage (not shown) despite an approximate 37% decrease in locus number. Across the final matrices, an average of 589 UCE loci were included in the 50% matrices, the lowest being the Opiliones + OUT data sets. Matrix lengths varied from 105 337 bp in the 'UCEsample' matrix to 452 309 bp in the Scorpiones matrix. The average percentage of parsimony-informative characters was 28.99%.

For shallow timescale comparisons, *Briggsus* (Briggs' harvestmen) resulted in the lowest number of UCE loci

Table 2 Matrix statistics

		50% taxon coverage					
Data set	*N*	Loci	Length	Mean length	Min–max length	PI	% PI
2perArachnid	11	602	199 810	331.91	113–853	60 468	30.3
UCEsample	32	510	105 337	206.54	80–746	49 362	46.9
Araneae + OUT	20	686	169 134	246.55	92–833	67 644	40.0
Araneae	19	724	181 915	251.26	92–833	71 786	39.5
Opiliones + OUT	9	435	158 314	363.94	97–893	33 263	21.0
Opiliones	8	381	153 187	402.07	97–914	29 807	19.5
Scorpiones + OUT	8	627	318 178	507.46	145–1136	65 227	20.5
Scorpiones	7	749	452 309	603.88	99–1141	64 892	14.4
Mean		**589.3**	**217 273**	**364.2**		**55 306.1**	**29.0**

PI, parsimony-informative sites.

recovered (Table 3; 292 loci, total length of 172 562, average locus length 591), but contained the highest number of variable sites (25 524—14.8%). Conversely, the *Centruroides* (bark scorpions) comparison recovered the highest number of UCE loci (585 loci, total length of 583 454, average locus length 997.4), but the lowest number of variable sites (7660—1.3%). The proportion of variable loci for all comparisons was above 0.9, with an average number of variable sites per locus ranging from 14 to 87 (Table 3).

Maximum-likelihood analyses of concatenated data sets demonstrated the utility of the arachnid-specific UCE loci in resolving relationships among and within arachnid orders (Figs 1–3). For the limited deep-level sample data set (i.e. '2perArachnid'), bootstrap support was 100% for all nodes except the node uniting Scorpiones to Araneae + Pedipalpi (=Amblypygi + Thelyphonida) (Fig. 1). For the Araneae data set, most nodes were fully supported (BS = 100%), while two nodes within Entelegyne spiders had bootstrap values of 97% (Fig. 2). Within Opiliones, all nodes received bootstrap support ≥99% (Fig. 3), and within Scorpiones, all nodes were fully supported except for one node with bootstrap support of 69% (Fig. 3). For the ASTRAL analyses of Araneae, Opiliones and Scorpiones, most relationships were consistent across analyses of different matrix completeness, and with concatenated results (Fig. S3, Supporting information). We elaborate on notable discrepancies in the discussion.

Discussion

Arachnid UCEs

The development of genetic markers for arachnids has lagged behind that of many groups due, in part, to the ancient divergences within the group (at least 400 MYA; Rehm *et al.* 2012; Rota-Stabelli *et al.* 2013) and the relatively few published arachnid genomes. Here, we demonstrate the utility of a UCE probe set targeting around 1000 loci that works for all arachnid species tested (and likely also with chelicerate outgroups). Across all samples, which include the deepest divergence within Arachnida (Regier *et al.* 2010; Sharma *et al.* 2014), we recovered 510 UCE loci with 49 362 parsimony-informative sites. For data sets within scorpions, spiders and harvestmen, which each span root nodes estimated at hundreds of millions of years of divergence (Hedin *et al.* 2012; Bond *et al.* 2014; Sharma and Wheeler 2014), we recovered 749, 724 and 381 loci, respectively. We did not find a significant correlation between library pool size and average number of loci recovered (Fig. S1, Supporting information), and thus, the lower number of loci recovered for Opiliones likely reflects the fewer genomic resources for this group and their distant relationship to the taxa from which the probes were developed.

Phylogenetic utility

Recent arachnid phylogenomic data sets based on transcriptomes have provided better-resolved and well-supported phylogenies compared with prior morphological and genetic work using few loci (Hedin *et al.* 2012; Bond *et al.* 2014; Fernández *et al.* 2014; Sharma *et al.* 2014, 2015; Garrison *et al.* 2016). However, obtaining sequence coverage across taxa from transcriptomes requires high-quality RNA and a consistent expression pattern, which may be difficult to obtain for many nonmodel taxa. Additionally, UCEs can be obtained from museum specimens or other potentially degraded samples (McCormack *et al.* 2015; Blaimer *et al.* 2016).

Our reconstructed phylogenies based on UCEs (Figs 1–3, S2 and S3, Supporting information) demonstrate the utility of this probe set for resolving relationships at a wide range of divergence levels, from the deepest bifurcations in the arachnid tree to shallower

Table 3 Congeneric matrix statistics

Data set	Loci	Total length	Mean length	Min–max length	Variable sites	% variable	# Polymorphic loci	Proportion of polymorphic	Average variable sites per variable locus
Antrodiaetus	480	329 110	685.65	129–1248	8389	2.55	476	0.99	17.62
Briggsus	292	172 562	590.97	144–1190	25 524	14.79	292	1.00	87.41
Centruroides	585	583 454	997.36	243–1164	7660	1.31	527	0.90	14.54

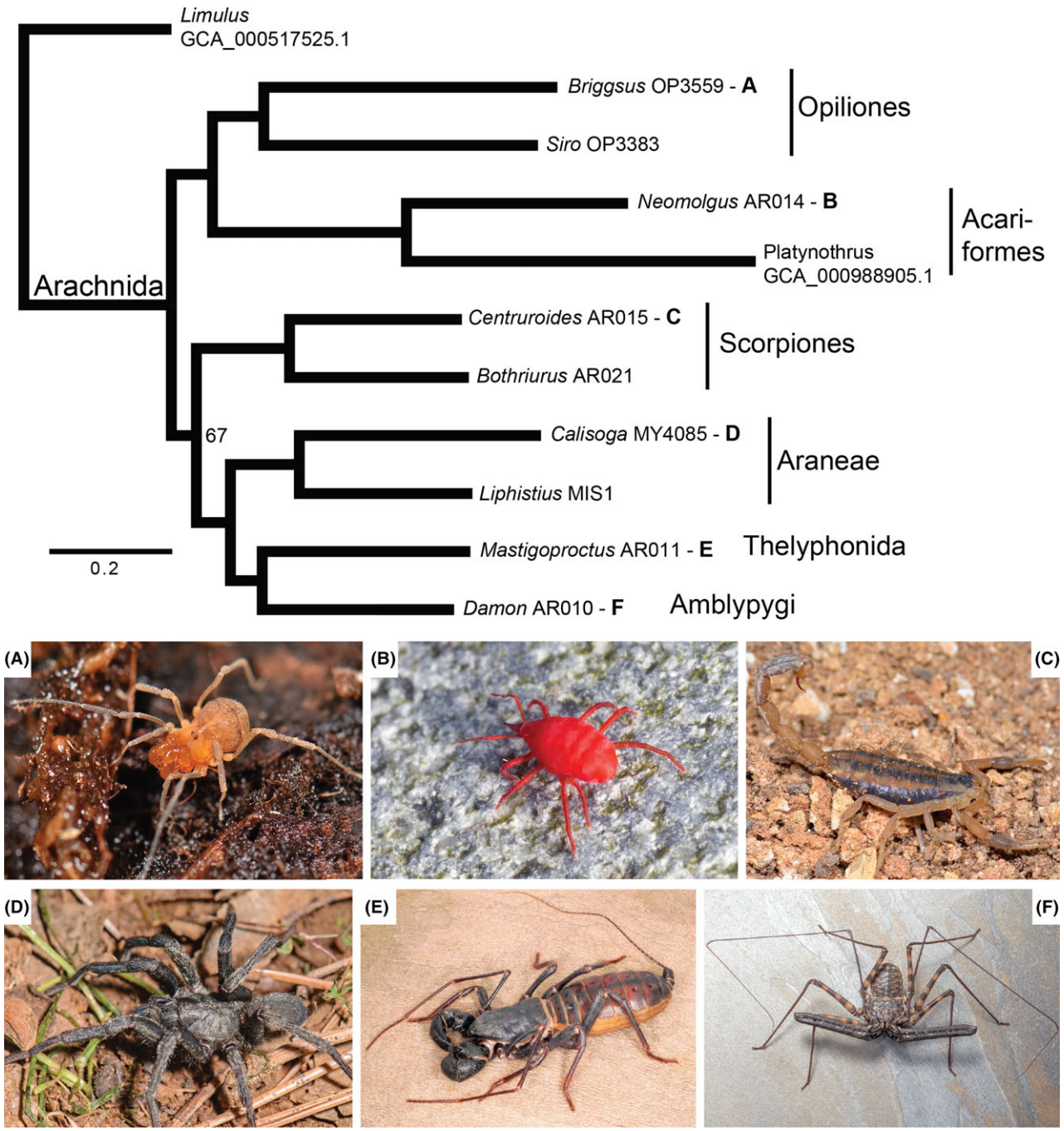

Fig. 1 Maximum-likelihood phylogeny based on 50% minimum taxon coverage '2perArachnid' data set, including representative images of included arachnid orders. Nodes have 100% bootstrap support, unless otherwise indicated. Letters A–F correspond to representative images of included arachnid orders. Photograph credit: M Hedin (A, C, D), M Erbland (B), WE Savary (E, F). [Colour figure can be viewed at wileyonlinelibrary.com].

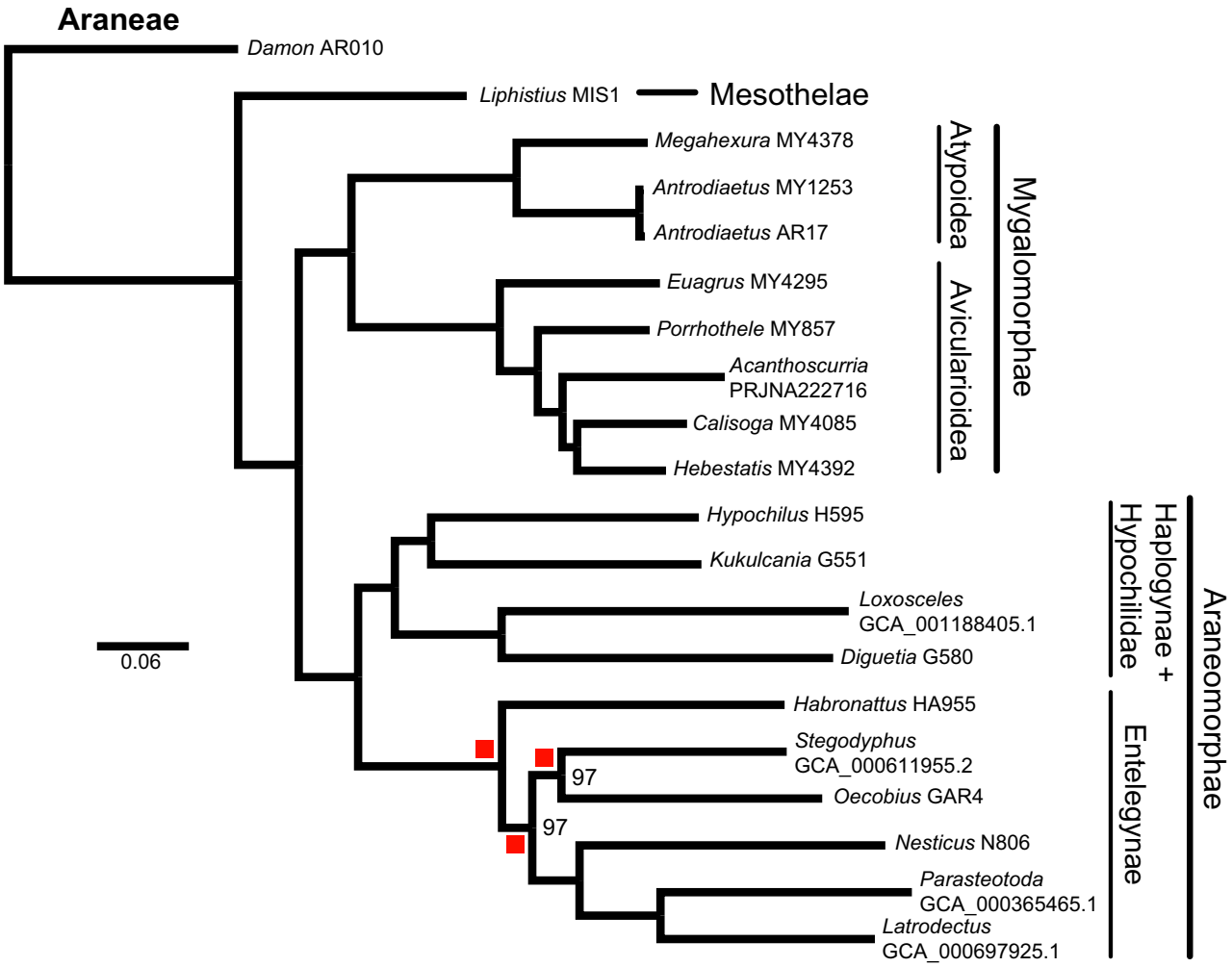

Fig. 2 Maximum-likelihood phylogeny based on 50% minimum taxon coverage Araneae + OUT data set. Nodes have 100% bootstrap support, unless otherwise indicated. Boxes at nodes indicate nodes that disagree with the Garrsion *et al.* (2016) transcriptome study. [Colour figure can be viewed at wileyonlinelibrary.com].

divergences within genera. In our limited, deep-level sample data set (i.e. '2perArachnid'), we recovered well-supported monophyletic groups for all orders for which more than one individual was included (Fig. 1). Additionally, we obtained well-supported sister relationships between Thelyphonida and Amblypygi (Pedipalpi), and a Pedipalpi sister group to Araneae, consistent with relationships based on morphology, Sanger sequence data and transcriptome data (Giribet *et al.* 2002; Shultz 2007; Regier *et al.* 2010; Sharma *et al.* 2014). The low support in our phylogeny for the placement of Scorpiones may be a result of sparse sampling, or it may reflect the difficulty in determining the position of this group due to a possible rapid radiation deep in the arachnid tree of life.

Phylogenetic analyses of the UCE data for more densely sampled groups (Araneae, Opiliones, Scorpiones; Figs 2 and 3) also recover topologies that are mostly congruent with recent transcriptome-based phylogenies (Hedin *et al.* 2012; Bond *et al.* 2014; Fernández *et al.* 2014; Sharma *et al.* 2015; Garrsion *et al.* 2016) and a traditional PCR approach (Dimitrov *et al.* 2016). Within Araneae, relationships that are well supported with transcriptome data (see Fig. 2 of Garrsion *et al.* 2016) are also recovered with UCE data in both concatenated and coalescent analyses under different missing data schemes (Fig. 2, S2 and S3, Supporting information). These relationships and recovered clades include a split between the Atypoidea and Avicularioidea within mygalomorphs, and a split between Haplogynae and Entelegynae within araneomorphs. We also recover a well-supported Paleocribellate (*Hypochilus*) + Filistatidae (*Kukulcania*) sister group, a relationship only recently hypothesized based on transcriptome data (Bond *et al.* 2014; Garrsion *et al.* 2016). Opisthothelae (Mygalomorphae + Araneomorphae) is recovered in all analyses except the ASTRAL analysis of the 90% complete data set,

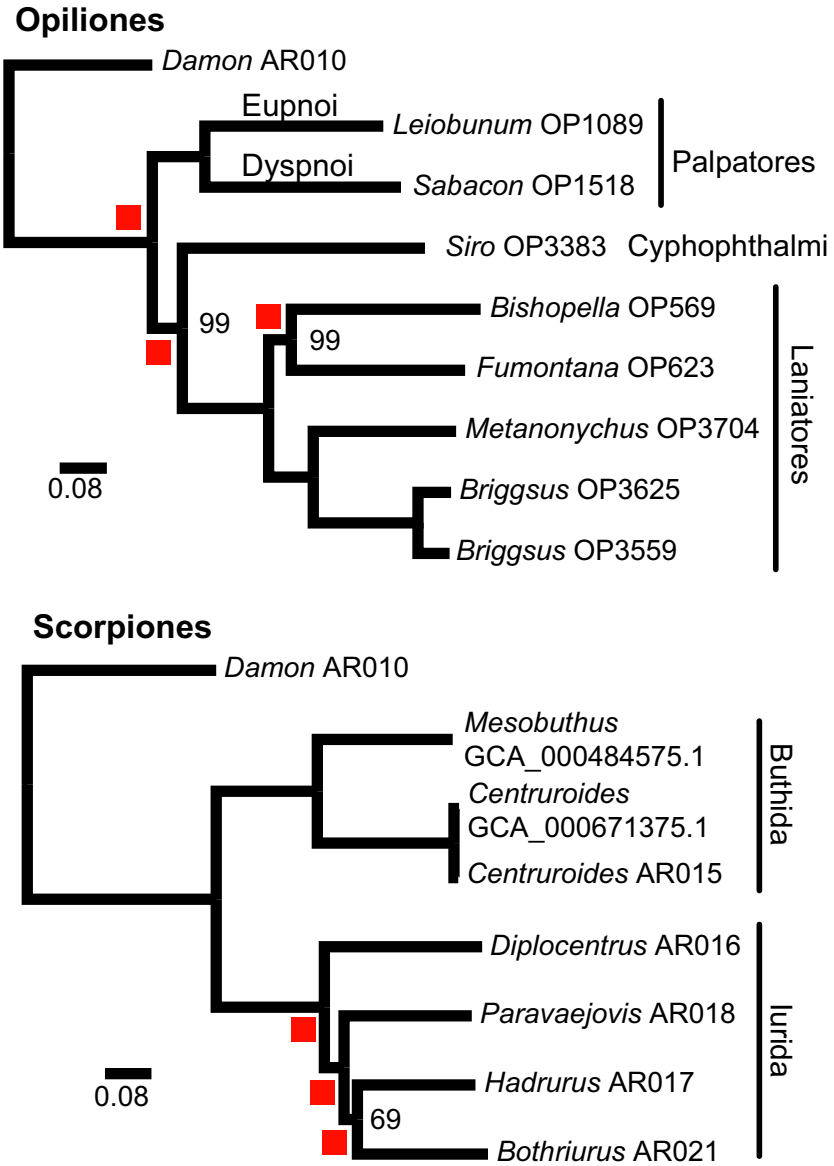

Fig. 3 Maximum-likelihood phylogenies based on 50% minimum taxon Opiliones + OUT and Scorpiones + OUT data sets. Nodes have 100% bootstrap support, unless otherwise indicated. Boxes at nodes indicate nodes that disagree with recent transcriptome studies (Hedin *et al.* 2012; Sharma et al. 2015). [Colour figure can be viewed at wileyonlinelibrary.com].

which resulted in a Mesothelae + Mygalomorphae sister relationship with low support (Fig. S3, Supporting information). Entelegyne relationships were stable and highly supported across our concatenated analyses under different missing data schemes (Figs 2, S2, Supporting information). These relationships were not consistent with those obtained from coalescent analyses where relationships had low support and were not consistent across the 50/70 and 90% complete data sets (Fig. S3, Supporting information). Support is generally lower for deep nodes within Entelegynae than that observed for much of the rest of the spider phylogeny based on transcriptome data (Garrison *et al.* 2016). The conflict in relationships and low support in both UCEs and transcriptomes indicate that more comprehensive taxon sampling is needed before meaningful comparisons can be made between these different data classes.

Within Opiliones, analyses of concatenated data under all missing data schemes resulted in strong support for a Palpatores clade (Eupnoi + Dyspnoi) and a Laniatores clade (Figs 3, S2 and S3, Supporting information). These clades have also received strong support in analyses of transcriptome data (Hedin *et al.* 2012; Sharma *et al.* 2014) and a five-gene data set combined with morphological characters (Sharma & Giribet 2014). With a single exception, all missing data schemes for both concatenated and coalescent analyses of UCEs resulted in strong support for a sister relationship between Cyphophthalmi (*Siro*) and Laniatores. This relationship conflicts with the well-accepted Phalangida clade

(Palpatores + Laniatores), which has received strong support based on transcriptome data (Hedin *et al.* 2012; Sharma *et al.* 2014). However, the 90% ASTRAL analysis does recover Phalangida with low support (Fig. S3, Supporting information). The misplacement of *Siro* does not appear to be an artefact of missing data given the consistency and strong support across our missing data schemes. We suggest that the sparse taxonomic sampling and the distant available outgroup are likely culprits for the conflict in relationships.

Within Scorpiones (Figs 3, S2 and S3, Supporting information), phylogenetic analyses of UCEs recover the same split between parvorders Buthida and Iurida as inferred from transcriptome data (Sharma *et al.* 2015). However, the well-supported position of *Bothriurus* deep within Iurida in our phylogeny conflicts with its more basal position in Iurida based on transcriptome analyses (Sharma *et al.* 2015). Notably, the 90% and 50% concatenated trees yield identical topologies and similar support, indicating that the relationships and support are robust to missing data when using concatenation (or at least not an artefact of missing data). All ASTRAL analyses produce an identical topology (Fig. S3, Supporting information), which differs from the concatenated phylogenies in the relationships within Iurida. The short internodes, likely due to an ancient and rapid radiation, account for these incongruences between data sets and previous analyses.

Comparison of UCE sequences across close relatives within spiders, harvestmen and scorpions indicates that our probe set will have applications beyond deep-level phylogenetics. The UCE loci show promise for use in species-level phylogenetics and species delimitation because the flanking regions show increasing variability as distance from the core UCE increases (Fig. 3). UCE enrichment produced 480 loci with more than 8000 variable sites across two members of the *Antrodiaetus riversi* complex, 292 loci with more than 25 000 variable sites across *Briggsus* species and 585 loci with more than 7000 variable sites across two *Centruroides* species (Table 3). These patterns of UCE variability are correlated with divergence in mitochondrial cytochrome oxidase I (Fig. 4). The proportion of polymorphic loci (minimum 0.9) and average variable sites per polymorphic locus (minimum 14) seen across these pairs is greater than those reported in the study of Smith *et al.* (2014; maximum 0.77 and 3.2, respectively), which demonstrated the utility of UCEs for species-level phylogenetic and species delimitation analyses in birds. UCEs may be advantageous for shallow-level studies in Arachnida because other techniques, such as RADseq, are susceptible to locus dropout in arachnid species with relatively deep levels of divergence (Bryson *et al.* 2016; Derkarabetian *et al.*

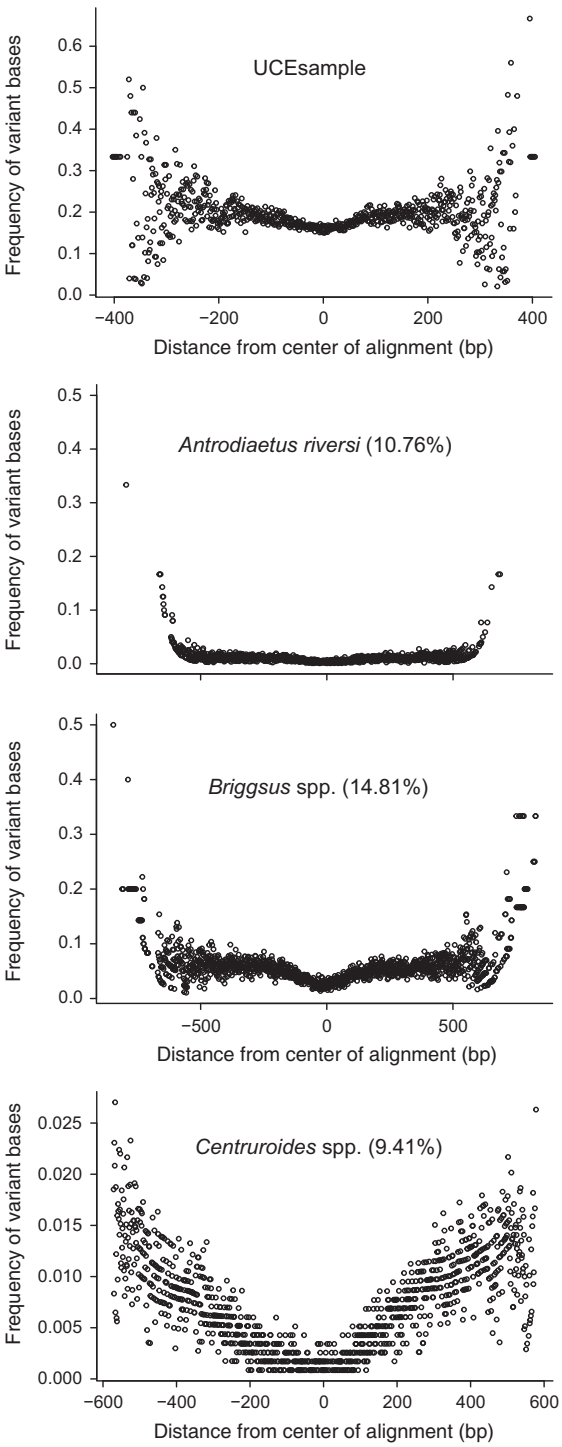

Fig. 4 Variability increases as distance from the core UCE increases. Data points with no variability were removed. For *Antrodiaetus* (turret spiders) and *Briggsus* (harvestmen) comparisons, the scale of *x*-axis is identical. Numbers in parentheses are uncorrected COI divergence values. In *Antrodiaetus*, an outlier was removed for better viewing. Cytochrome oxidase subunit I (COI) distances for *Centruroides* (bark scorpions) were calculated from GenBank Accession nos. AY995831.1 (*C. sculpturatus*) and AY995833.1 (*C. exilicauda*).

2016). Our UCE probe set will also provide a complementary/alternative resource to the recently developed spider anchored-enrichment loci (Hamilton *et al*. 2016b), which were used to produce a comparable number of loci (455) and a well-resolved phylogeny for the North American tarantula genus *Aphonopelma* (Hamilton *et al.* 2016a).

The arachnid-specific UCEs have utility and phylogenetic informativeness at all levels of Arachnida, spanning extremely ancient divergence times between orders (at least 400 MYA) to more recent congeneric divergences (<10 MYA for *Antrodiaetus*, Hedin *et al.* 2013). Thus, these markers have the potential to be an integral component of future comparative studies. Arachnids are an extremely diverse group, and yet it is estimated that less than half of arachnid species have been formally described (Chapman 2009). Species delimitation increasingly relies on phylogenomic data that can be sampled consistently across diverse taxa (Leaché *et al*. 2014; Smith *et al*. 2014; Rannala 2015; Harvey *et al*. 2016). The recovery of homologous UCEs across Arachnida makes these loci a valuable resource for discovery of new species and for inferring phylogenetic relationships in understudied arachnid groups.

Acknowledgements

For help during laboratory work, we thank Carl Oliveros and Rich Harrington. John Klicka provided other assistance. Tim Burkhart, Ken MacNeil, Peter Clausen and Cor Vink provided assistance with specimen collection. Live specimen photographs were provided by Mardon Erbland and Warren E. Savary. Funding for this research was provided by NSF Grant DEB-1354558 to MH. Startup funds from LSU and DEB-1242260 to BCF supported this work.

References

Alexander RP, Fang G, Rozowsky J, Snyder M, Gerstein MB (2010) Annotating non-coding regions of the genome. *Nature Reviews Genetics*, **11**, 559–571.

Bejerano G, Pheasant M, Makunin I *et al*. (2004) Ultraconserved elements in the human genome. *Science*, **304**, 1321–1325.

Blaimer BB, Lloyd MW, Guillory WX, Brady SG (2016) Sequence capture and phylogenetic utility of genomic ultraconserved elements obtained from pinned insect specimens. *PLoS ONE*, **11**, e0161531. doi:10.1371/journal.pone.0161531.

Bond JE, Garrison NL, Hamilton CA, Godwin RL, Hedin M, Agnarsson I (2014) Phylogenomics resolves a spider backbone phylogeny and rejects a prevailing paradigm for orb web evolution. *Current Biology*, **24**, 1765–1771.

Bryson RW, Savary WE, Zellmer AJ, Bury RB, McCormack JE (2016) Genomic data reveal ancient microendemism in forest scorpions across the California Floristic Province. *Molecular Ecology*, **25**, 3731–3751.

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, **17**, 540–552.

Chapman AD (2009) *Numbers of Living Species in Australia and the World*, 2nd edn. Australian Biodiversity Information Services, Canberra.

Crawford NG, Faircloth BC, McCormack JE, Brumfield RT, Winker K, Glenn TC (2012) More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. *Biology Letters*, **8**, 783–786.

Crawford NG, Parham JF, Sellas AB, et al. (2015) A phylogenomic analysis of turtles. *Molecular Phylogenetics and Evolution*, **83**, 250–257.

Derkarabetian S, Burns M, Starrett J, Hedin M (2016) Population genomic evidence for multiple refugia in montane-restricted harvestmen (Arachnida, Opiliones, *Sclerobunus robustus*) from the southwestern United States. *Molecular Ecology*, **25**, 4611–4631.

Derti A, Roth FP, Church GM, Wu C (2006) Mammalian ultraconserved elements are strongly depleted among segmental duplications and copy number variants. *Nature Genetics*, **38**, 1216–1220.

Dimitrov D, Benavides LR, Arnedo MA *et al.* (2016) Rounding up the usual suspects: a standard target-gene approach for resolving the interfamilial phylogenetic relationships of ecribellate orb-weaving spiders with a new family-rank classification (Araneae, Araneoidea). *Cladistics*, doi:10.1111/cla.12165 (in press).

Dunlop JA (1996) A trigonotarbid arachnid from the Upper Silurian of Shropshire. *Palaeontology*, **39**, 605–614.

Dunlop JA, Tetlie OE, Prendini L (2008) Reinterpretation of the Silurian scorpion *Proscorpius osborni* (Whitfield): integrating data from Palaeozoic and Recent scorpions. *Palaeontology*, **51**, 303–320.

Faircloth BC (2013) Illumiprocessor: a Trimmomatic wrapper for parallel adapter and quality trimming. doi: 10.6079/J9ILL.

Faircloth BC (2015) PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics*, **32**, 786–788.

Faircloth BC (2016) Identifying conserved genomic elements and designing universal probe sets to enrich them. bioRxiv preprint doi:10.1101/077172, bioRxiv 077172.

Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT, Glenn TC (2012) Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*, **61**, 717–726.

Faircloth BC, Sorenson L, Santini F, Alfaro ME (2013) A phylogenomic perspective on the radiation of ray-finned fishes based upon targeted sequencing of ultraconserved elements (UCEs). *PLoS ONE*, **8**, e65923. doi:10.1371/journal.pone.0065923.

Faircloth BC, Branstetter MG, White ND, Brady SG (2015) Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera. *Molecular Ecology Resources*, **15**, 489–501.

Fernández R, Hormiga G, Giribet G (2014) Phylogenomic analysis of spiders reveals nonmonophyly of orb weavers. *Current Biology*, **24**, 1772–1777.

Garrsion NL, Rodriguez J, Agnarsson I *et al*. (2016) Spider phylogenomics: untangling the spider tree of life. *PeerJ*, **4**, e1719. doi: 10.7717/peerj.1719.

Gilbert PS, Chang J, Pan C *et al*. (2015) Genome-wide ultraconserved elements exhibit higher phylogenetic informativeness than traditional gene markers in percomorph fishes. *Molecular Phylogenetics and Evolution*, **92**, 140–146.

Giribet G, Edgecombe GD, Wheeler WC, Babbitt C (2002) Phylogeny and systematic position of opiliones: a combined analysis of chelicerate relationships using morphological and molecular data1. *Cladistics*, **18**, 5–70.

Glenn TC, Nilsen R, Kieran TJ *et al.* (2016) Adapterama I: universal stubs and primers for thousands of dual-indexed Illumina libraries (iTru & iNext). *bioRxiv*, doi:10.1101/049114.

Grabherr MG, Haas BJ, Yassour M *et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, **29**, 644–652.

Hamilton CA, Hendrixson BE, Bond JE (2016a) Taxonomic revision of the tarantula genus *Aphonopelma* Pocock, 1901 (Araneae, Mygalomorphae, Theraphosidae) within the United States. *ZooKeys*, **560**, 1–340.

Hamilton CA, Lemmon AR, Lemmon EM, Bond JE (2016) Expanding anchored hybrid enrichment to resolve both deep and shallow relationships within the spider tree of life. *BMC Evolutionary Biology*, **16**, 212.

Harvey MS (2002) The neglected cousins: what do we know about the smaller arachnid orders? *Journal of Arachnology*, **30**, 357–372.

Harvey MG, Smith BT, Glenn TC, Faircloth BC, Brumfield RT (2016) Sequence capture versus restriction site associated DNA sequencing for shallow systematics. *Systematic Biology*, **65**, 910–924.

Hedin M, Starrett J, Akhter S, Schönhofer AL, Shultz JW (2012) Phylogenomic resolution of paleozoic divergences in harvestmen (Arachnida, Opiliones) via analysis of next-generation transcriptome data. *PLoS ONE*, **7**, e42888. doi:10.1371/journal.pone.0042888.

Hedin M, Starrett J, Hayashi C (2013) Crossing the uncrossable: novel trans-valley biogeographic patterns revealed in the genetic history of low-dispersal mygalomorph spiders (Antrodiaetidae, Antrodiaetus) from California. *Molecular Ecology*, **22**, 508–526.

Jeram AJ, Selden PA, Edwards D (1990) Land animals in the Silurian: arachnids and myriapods from the Shropshire, England. *Science*, **250**, 658–661.

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, **30**, 772–780.

Laurie M (1899) On a Silurian scorpion and some additional eurypterid remains from the Pentland Hills. *Transactions of the Royal Society of Edinburgh*, **39**, 575–590.

Leaché AD, Fujita MK, Minin VN, Bouckaert RR (2014) Species delimitation using genome-wide SNP data. *Systematic Biology*, **63**, 534–542.

Manthey JD, Campillo LC, Burns KJ, Moyle RG (2016) Comparison of target-capture and restriction-site associated DNA sequencing for phylogenomics: a test in cardinalid tanagers (Aves, Genus: *Piranga*). *Systematic Biology*, **65**, 640–650.

Masta SE, Longhorn SJ, Boore JL (2009) Arachnid relationships based on mitochondrial genomes: asymmetric nucleotide and amino acid bias affects phylogenetic analyses. *Molecular Phylogenetics and Evolution*, **50**, 117–128.

McCormack JE, Faircloth BC, Crawford NG, Gowaty PA, Brumfield RT, Glenn TC (2012) Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. *Genome Research*, **22**, 746–754.

McCormack JE, Harvey MG, Faircloth BC, Crawford NG, Glenn TC, Brumfield RT (2013) A phylogeny of birds based on over 1500 loci collected by target enrichment and high-throughput sequencing. *PLoS ONE*, **8**, e54848 doi: 10.1371/journal.pone.0054848.

McCormack JE, Tsai WLE, Faircloth BC (2015) Sequence capture of ultraconserved elements from bird museum specimens. *Molecular Ecology Resources*, **16**, 1189–1203.

Meiklejohn KA, Faircloth BC, Glenn TC, Kimball RT, Braun EL (2016) Analysis of a rapid evolutionary radiation using ultraconserved elements (UCEs): evidence for a bias in some multispecies coalescent methods. *Systematic Biology*, **65**, 612–627.

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

Mirarab S, Reaz R, Bayzid MS, Zimmermann T, Swenson MS, Warnow T (2014) ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics*, **30**, i541–i548.

Rannala B (2015) The art and science of species delimitation. *Current Zoology*, **61**, 846–853.

Regier JC, Shultz JW, Zwick A *et al.* (2010) Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature*, **463**, 1079–1083.

Rehm P, Pick C, Borner J, Markl J, Burmester T (2012) The diversity and evolution of chelicerate hemocyanins. *BMC Evolutionary Biology*, **12**, 19 doi: 10.1186/1471-2148-12-19.

Rohland N, Reich D (2012) Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research*, **22**, 939–946.

Rota-Stabelli O, Daley AC, Pisani D (2013) Molecular timetrees reveal a Cambrian colonization of land and a new scenario for ecdysozoan evolution. *Current Biology*, **23**, 392–398.

Sharma PP, Giribet G (2014) A revised dated phylogeny of the arachnid order Opiliones. *Frontiers in Genetics*, **5**, 1–13.

Sharma PP, Kaluziak ST, Pérez-Porro AR *et al.* (2014) Phylogenomic interrogation of arachnida reveals systematic conflicts in phylogenetic signal. *Molecular Biology and Evolution*, **31**, 2963–2984.

Sharma PP, Fernández R, Esposito LA, González-Santillán E, Monod L (2015) Phylogenomic resolution of scorpions reveals multilevel discordance with morphological phylogenetic signal. *Proceedings of the Royal Society of London B*, **282**, 20142953. doi:doi.org/10.1098/rspb.2014.2953.

Sharma PP, Wheeler WC (2014) Cross-bracing uncalibrated nodes in molecular dating improves congruence of fossil and molecular age estimates. *Frontiers in Zoology*, **11**, 57.

Shultz JW (2007) A phylogenetic analysis of the arachnid orders based on morphological characters. *Zoological Journal of the Linnean Society*, **150**, 221–265.

Smith BT, Harvey MG, Faircloth BC, Glenn TC, Brumfield RT (2014) Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. *Systematic Biology*, **63**, 83–95.

Stamatakis A (2014) RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.

Stamatakis A, Hoover P, Rougemont J (2008) A rapid bootstrap algorithm for the RAXML web servers. *Systematic Biology*, **57**, 758–771.

Streicher JW, Schulte JA II, Weins JJ (2016) How should genes and taxa be sampled for phylogenomic analyses with missing data? An empirical study in Iguanian lizards. *Systematic Biology*, **65**, 128–145.

Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*, **56**, 564–577.

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

Zhang Z-Q (2011) Phylum Arthropoda von Siebold, 1848 *In:* Zhang, Z.-Q. (Ed.) Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness. *Zootaxa*, **3148**, 99–103.

All authors contributed to the conception and design of the experiments; J.S., S.D., M.H. and R.W.B. contributed samples; J.S., S.D., R.W.B. and B.C.F. conducted laboratory work; J.S., S.D. and B.C.F. analysed data; all authors edited and commented on the manuscript.

Data accessibility

Raw sequence reads are available in the NCBI Short Read Archive Accession no. SRP078995 (BioProject ID #PRJNA328972). Untrimmed contigs and trimmed concatenated alignments are available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.5k3j3.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1 Sequencing results of sample pooling.

Fig. S2 Maximum-likelihood phylogeny based on 90% minimum taxon coverage for the 2perArachnid, Araneae + OUT, Opiliones + OUT, and Scorpiones + OUT data sets. Nodes have 100% bootstrap support, unless otherwise indicated.

Fig. S3 ASTRAL species trees based on 50%, 70%, and 90% minimum taxon coverage for the Araneae + OUT, Opiliones + OUT, and Scorpiones + OUT data sets. Nodes have 100% support, unless otherwise indicated.

Table S1 Statistics for genomes used in probe design and phylogenetic analyses. Target loci refer to number of loci targeted in the probe array.

[Correction added on 15th December, 2016, after first online publication: Present address was added to the first author James starret]