2010. The Journal of Arachnology 38:9–20

The phylogenetic utility of the nuclear protein-coding gene EF-1α for resolving recent divergences in Opiliones, emphasizing intron evolution

Marshal Hedin, **Shahan Derkarabetian**, **Maureen McCormack** and **Casey Richart**: Department of Biology, San Diego State University, San Diego, California 92182-4614, USA. E-mail: mhedin@sciences.sdsu.edu

Jeffrey W. Shultz: Department of Entomology, University of Maryland, College Park, Maryland 20742-4454, USA

Abstract. Our focus was to design harvestmen-specific PCR primers to target both introns and exons of the nuclear protein-coding gene Elongation Factor -1 alpha (EF-1α). We tested this primer set on ten genera representing all primary lineages of Opiliones, with sets of close phylogenetic relatives (i.e., sets of several congeners) included to specifically assess utility at shallow phylogenetic levels. Our research also included the collection of parallel mitochondrial protein-coding DNA sequence datasets for the congeneric sets to compare relative rates of evolution and gene tree congruence for EF-1α versus mitochondrial data. The harvestmen primers resulted in successful amplification for nine of ten tested genera. Exon sequences for these nine genera appear orthologous to previously-reported EF-1α Opiliones sequences, which were generated using RT-PCR methods. Newly-generated exon sequences are interrupted by three separate spliceosomal introns; two introns are restricted to one or two genera, but a third intron is conserved in position across all surveyed genera. Phylogenetic analyses of EF-1α nucleotide data for congeneric sets result in gene trees that are generally congruent with mitochondrial gene trees, with EF-1α phylogenetic signal coming from both intron and exon sites, and resolving apparently recent divergences (e.g., as recent as one million years ago). Overall, the combination of gene orthology, conserved intron position, and gene tree congruence at shallow levels suggest that this gene region will prove generally useful for both phylogeographic and species-level phylogenetic analyses in Opiliones, complementing already-documented utility at higher taxonomic levels.

Keywords: Molecular systematics, phylogeography, gene tree, orthology, elongation factor -1 alpha

Both phylogenetics and taxonomy obviously have been impacted by advances in molecular biology, and most modern systematic analyses published today include some molecular component. In the arachnid order Opiliones, several molecular systematic studies have been published that consider relatively ancient phylogenetic divergences (e.g., Giribet et al. 1999, 2002; Shultz & Regier 2001). Fewer studies have considered molecular phylogenetic divergence within a species, or between closely related species, and those published have relied mostly on mitochondrial DNA sequence data. While informative, a mitochondrial-only perspective has limitations (Ballard & Whitlock 2004; Rubinoff & Holland 2005), and is quickly becoming obsolete in a molecular systematics world dominated by multigenic datasets.

A handful of nuclear genes have been used to address recent divergences within Opiliones. Those utilized offer a mix of pros and cons. Thomas & Hedin (2008) used ribosomal ITS sequences at the intraspecific level in *Fumontana deprehendor* Shear (Laniatores), and although well-behaved in this case, this gene region has been shown to be problematic in other arthropod taxa because of a lack of concerted evolution (e.g., Harris & Crandall 2000; Bower et al. 2009). Ribosomal 28S sequences have been used in various harvestmen taxa, but in our experience these data show limited variability at the shallowest levels (e.g., see Hedin & Thomas 2010). Shultz & Regier (2009) used reverse transcriptase polymerase chain reaction (RT-PCR) methods to generate exon nucleotide data for the nuclear protein-coding genes EF-1α and POL II, and applied these data to species-level relationships in the genus *Caddo*. While these data are informative at shallow levels, gathering the data requires the extra cost and expertise needed to conduct RT-PCR. Finally, Sharma & Giribet (2009) used three novel nuclear genes (H3, H4, U2 snRNA) to address relationships in sandokanid laniatoreans. Although the focus was at slightly higher taxonomic levels, several congeneric comparisons were made. Again, despite the clear utility in the multigenic perspective used in this study, these gene fragments individually are small (330, 160, 130 basepairs (bp), respectively), and our examination of these data suggests minimal variation among congeneric species.

Here we assess the phylogenetic utility of the nuclear protein-coding gene Elongation Factor-1 alpha (EF-1α) for resolving relatively shallow opilion phylogenetic divergences. In metazoans, EF-1α is a core gene involved in protein synthesis – this fact facilitates use in molecular phylogenetics (e.g., the gene is expected to be found in all taxa, have conserved function, etc). This gene is commonly used in hexapod systematics (summarized in Caterino et al. 2000), and has also seen limited use in arachnids (Shultz & Regier 2001, 2009; Hedin & Maddison 2001). Given our interest in resolving recent divergences, we expect most phylogenetic information to come from either silent substitutions at third codon exon positions (e.g., Cho et al. 1995; Reed & Sperling 1999), or from introns (e.g., Danforth et al. 1999; Hedin & Maddison 2001). Here we present the results of a commonly employed strategy: to develop PCR primers that reside in relatively conserved exons, but span variable introns that are conserved in position (see Palumbi 1996). We have discovered several EF-1α introns in Opiliones, some of which are universally conserved in position, but variable at shallow phylogenetic levels. The primers developed permit consistent and robust amplification of a fast-evolving gene region using standard PCR methods, allowing potentially rapid collection of population-level samples (e.g., many individuals from

Table 1.—PCR primers. Primers marked with an asterisk were designed after sequence collection using more general primers. The relative position (POS) of primers (5′ end) is shown based on alignment with exon sequences of Shultz & Regier (2001). Ambiguity codes are standard.

		POS
Forward Primers		
EF1-OP1	5′ -CGTGGTATYACCATYGATATCAC -3′	28
EF1-OP2	5′ -GATTTCATCAARAACATGATYAC -3′	112
EF1-OP2SCLER	5′ -GATTTCATCAAGAACATGATTAC -3′	112
*EF1-OP2ASAB	5′ -GCTGTGCTTATTGTTGCTGCYGG -3′	157
*EF1-OP2BSAB	5′ - GGTACTGGTGAGTTTGAAGCTGG-3′	178
EF1-OP3	5′ - TTTGARGAAATCCARAARGAAGT-3′	322
EF1-OP3PHAL	5′ - TTTGAAGAAATCCAAAAGGAAGT-3′	322
EF1-OP3SCLER	5′ - TTTGAGGAAATCCAGAAGGAAGT-3′	322
EF1-OP4	5′ - TACATYAAGAAGATTGGTTA-3′	352
EF1-OP4SCLER	5′ - TACATCAAGAAGATCGGTTA-3′	352
*EF1-OP5LEIO	5′ - GGAGATAACATGTTGGAACAAAG-3′	415
EF1-OP5PHAL	5′ - AACATGTTGGAACAAAGTACCCA-3′	421
EF1-OP5LAN	5′ - AACATGYTGGAAGCTTCCTCC-3′	421
EF1-OP5ISCH	5′ - AACATGTTGGARGCCAGYGC-3′	421
EF1-OP6PHAL	5′ - CATCACCACTGAAGTTAAATCTG-3′	672
Reverse Primers		
*EF1-OP5LEIORC	5′ - CTTTGTTCCAACATGTTATCTCC-3′	437
EF1-OPRC1PHAL	5′ - CAGATTTAACTTCAGTGGTGATG -3′	692
EF1-OPRC1SCLER	5′ - CGGACTTGACCTCAGTGGTGATG -3′	694
EF1-OPRC2	5′ - GANACGTTCTTNACRTTGAA -3′	767
*EF1-OPRC2LEIO	5′ - GAAACGTTCTTAACATTGAA -3′	767
EF1-OPRC2PHAL	5′ - ACGGAAACGTTCTTAACGTTGAA -3′	770
EF1-OPRC3PHAL	5′ - ATGACCTGGGCRGTGAATTCTTC -3′	854
*EF1-OPRC3LEIO	5′ - ATAACCTGGGCAGTAAATTCTTC -3′	854
EF1-OPRC3SCLER	5′ - ATGACCTGAGCCGTGAACTCTTC -3′	854
EF1-OPRC4	5′ - GAACTTGCANGCAATGTGAGC -3′	935
*EF1-OPRC4LEIO	5′ - GAACTTGCAAGCAATGTGAGC -3′	935
EF1-OPRC5PHAL	5′ - GGTTGTCTTCCAATTTCTTGCC -3′	992

multiple populations from many species). We present this primer set as a resource to be explored, modified, and utilized by other opilion systematists.

METHODS

Primer design.—Shultz & Regier (2001) generated nearly full-length EF-1α exon nucleotide sequences using RT-PCR methods for a comprehensive sample of Opiliones. These sequences were downloaded from GenBank, then compiled and translated in MacClade 4.08 (Maddison & Maddison 2003). We designed primers in conserved exon regions that might also amplify introns, using known intron positions from spiders and hexapods as a guide (Table 1, Figure 1). Within conserved regions, we manually (i.e., without software) designed 20–25 bp oligonucleotides in regions with approximately equal base composition, and GC-rich 3′ nucleotides corresponding to strictly conserved second codon positions. Most primers were designed to work for all Opiliones, but in some cases we also designed clade-specific primers (e.g., EF1-OP3PHAL is a primer that best matches the phalangioid sequences from Shultz & Regier 2001, etc.). We designed additional internal, clade-specific primers after initial sequencing with our more general primers.

Samples.—We tested primer utility on ten genera (Table 2) that represent all primary higher-level clades within Opiliones (Fig. 2). Specimens were either preserved directly in 100% EtOH in the field, kept cold, and later transferred to a −80° C freezer (or −20° C freezer in the case of *Caddo*), or preserved directly in 100% EtOH in the lab and stored in a −80° C freezer. All specimens are currently housed in the Arthropods Genomics Collection at San Diego State University. We extracted genomic DNA from leg tissues using the Qiagen DNeasy Kit; these extracts were either suspended in 200 microliters (μl) of Qiagen AE buffer, or dried down using a Speedvac and resuspended in a smaller volume of AE buffer.

Testing primer utility.—For some samples (*Siro, Caddo, Ortholasma, Dendrolasma* and *Bishopella),* we assessed primer utility in a preliminary manner. Here we conducted PCR experiments to test whether our primers resulted in amplicons of expected size, and if so, we directly sequenced these amplicons. These sequences inform us about primer success, intron position, and relative intron size, but do not address variability of the nucleotide data within a taxon (see below). We conducted PCR experiments using various forward/reverse primer combinations (Table 2). All PCR reactions included 0.8 μl of genomic DNA with 0.08 μl *Ex Taq* polymerase (Takara Bio Inc.), and 2.5 μl each of Takara dNTP mix, Takara buffer (with Mg^{2+}) and primers (at 2–4 μM concentrations). We used tissue culture water (Sigma-Aldrich Co.) to bring reaction volumes to 25 μl. Experiments were conducted on two different PCR machines, including a MJ Research PTC-100 (95° C, 2 min; 35× (95° C, 30 s; 53° C, 45 s; 72° C, 45 s); 72° C, 10 min) and an Eppendorf Mastercycler Personal (94° C, 3 min; 35× (94° C, 45 s; 45° C, 45 s; 72° C, 1 min 15 s); 72° C, 5 min). Amplification products were visualized on agarose gels, purified via polyethylene glycol (PEG) precipitation, and sequenced directly at the SDSU Micro-

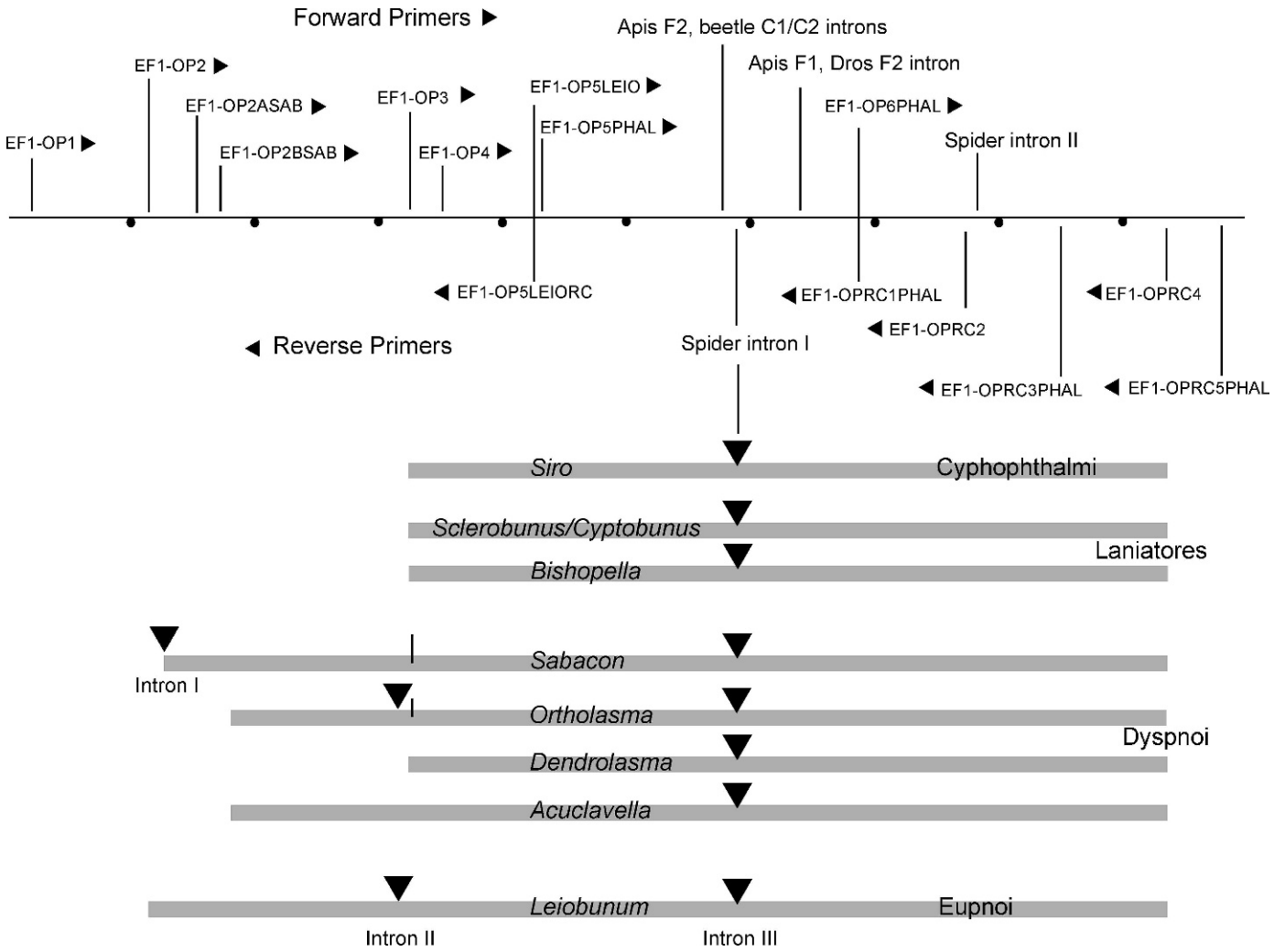

Figure 1.—Upper: Relative primer positions (see also Table 1). Spider intron positions from Hedin & Maddison (2001), hexapod intron positions from Brady & Danforth (2004). Small circles designate 100-bp intervals. Lower: Amplicon sizes and relative intron positions for opilion sample. Small hashes above amplicon boxes for *Ortholasma* and *Sabacon* indicate alternative 5′ primers used for some templates (see also Table 2).

chemical Core Facility (http://www.sci.sdsu.edu/dnacore/sdsu_dnacore.html) on an ABI Prism 3100 capillary machine.

Both strands were determined for most templates using PCR primers in sequencing reactions. For some *Leiobunum* templates, we used additional internal primers for sequencing. We assembled and edited sequence contigs using Sequencher version 4.5. For some taxa (see Results), AT-rich regions (and possible allelic length heterozygosity) within introns caused premature stops in some sequencing reads; sequences for these templates were thus determined using single reads from opposite sides of the intron. Also, for *Acuclavella* we were able to generate high-quality 780-bp sequence reads for the entire exon plus intron amplicon, and thus used only a single sequence read for several of these templates. Sites with two peaks of equal intensity on chromatograms were interpreted as representing nucleotide heterozygosity, and scored as such in data matrices using standard ambiguity codes.

To assess phylogenetic utility at shallow levels more rigorously, we collected EF-1α sequences for sets of close phylogenetic relatives (both within and between species) for *Leiobunum*, *Sabacon*, *Sclerobunus*/*Cyptobunus,* and *Acuclavella*. For these same taxon sets we also collected parallel mitochondrial sequence datasets, allowing us to gauge the relative variability of nuclear versus mitochondrial gene data, and ask if independent phylogenetic analyses resulted in generally congruent results. For *Sabacon*, *Sclerobunus*/*Cyptobunus,* and *Acuclavella,* we collected partial cytochrome oxidase I (CO1) mitochondrial sequences, following standard techniques as in Thomas & Hedin (2008). Forward primers included C1-N-1510 or C1-N-1718S, combined with C1-J-2568 or C1-J-2776S (see Hedin & Thomas 2010). For *Leiobunum*, we collected nearly full-length mitochondrial ND1 sequences using the custom primers LR-N-12945LEI (5′- TGACCTCGATGTTGAATTAA -3′) and CB-J-11638LEI (5′ - CCTWATAAACTAATCATTTAGC - 3′).

We conducted phylogenetic analyses at two different hierarchical levels. First, we conducted an exon-only analysis that included the Shultz & Regier (2001) exon data, plus newly-generated exon-only data (i.e., introns removed). This exon matrix was aligned manually, and subject to a heuristic

Table 2.—Taxon sample, including voucher and collection location information, mitochondrial sample, and successful EF-1α primers.

Taxon	Voucher no.	Collection location (N lat., W long.)	mtDNA	EF-1α primers
Sironidae				
Siro cf. kamiakensis	OP 2350	Idaho: Idaho Co., FS 311, S Route 14 (45.6853, −115.5427)	no	OP3, OPRC4LEI
Caddoidea: Caddidae				
Caddo agilis	OP2565	New Hampshire: Cheshire Co., Pisgah State Park (42.862, −72.428)	no	failed
Phalangioidea: Sclerosomatidae				
Leiobunum aldrichi	OP 829	Mississippi: Tishomingo Co., Tishomingo State Park (34.6054, −88.1927)	yes	OP2, OPRC4LEI
L. aldrichi	OP 1069	Missouri: Calhoun Co., Marshall (42.3012, −84.9674)	yes	OP2, OP5LEIORC, OPRC4LEI
L. vittatum	OP 835	Tennessee: Cumberland Co., Cumberland Mtn SP (35.9013, −84.9958)	yes	OP2, OP4, OP5LEIORC, OPRC4LEI
L. speciosum	OP 1405	Tennessee: Davidson Co., west Nashville (36.1227, −86.9061)	yes	OP2, OP4, OP5LEIORC, OPRC4LEI
L. calcar	OP 1394	North Carolina: Clay Co., Big Tuni Creek (35.1025, −83.7007)	yes	OP2, OP4, OP5LEIORC, OPRC4LEI
L. calcar	OP 814	Tennessee: Cocke Co., road to Cosby CG (35.7633, −83.2115)	yes	OP2, OPRC4LEI
L. undescribed species	OP 1383	Virginia: Grayson Co., Grayson Highlands SP (36.6247, −81.5013)	yes	OP2, OPRC4LEI
L. serratipalpe	OP 1080	Maryland: Montgomery Co., Unity, Tusculum Farm (39.2427, −77.0872)	yes	OP2, OPRC4LEI
Troguloidea: Nemastomatidae				
Ortholasma rugosum	OP 807	California: San Benito Co., NE of Pinacate Peak, off Hwy 101 (36.8602, −121.6126)	no	OP2BSAB, OP3, OPRC4LEI
Ortholasma rugosum	OP 808	California: San Mateo Co., San Bruno Mtn SP (37.6943, −122.4529)	no	OP3, OPRC4LEI
Dendrolasma mirabile	OP 1000	Oregon: Curry Co., Hwy 33, E of Gold Beach (42.5465, −124.1273)	no	OP3, OPRC4LEI
Ischryopsalidoidea: Ceratolasmatidae				
Acuclavella cf. *quattuor*	OP2233	Idaho: Idaho Co., Eagle Mt Trailhead, S side of Lochsa River (46.4292, −115.1335)	yes	OPRC4 only
A. merickeli	OP2251	Idaho: Idaho Co., along headwaters of Red River, Red River Rd (45.7853, −115.2026)	yes	OPRC4 only
A. quattuor	OP2255	Idaho: Idaho Co., Tributary of Crooked Creek, FS Rd 222 (45.5791, −115.4431)	yes	OPRC4 only
A. cosmetoides	OP2278	Idaho: Idaho Co., Canyon Creek Trailhead, US 12 (46.2101, −115.5442)	yes	OPRC4 only
A. cf. *merickeli*	OP2345	Washington: Jefferson Co., Cedar Creek, E of US (101 47.7105, −124.4095)	yes	OP2BSAB, OPRC4
A. cf. *merickeli*	OP2347	Washington: Lewis Co., tributary of Iron Creek, FS Rd 25 (46.4033, −121.9902)	yes	OP2BSAB, OPRC4
Ischryopsalidoidea: Sabaconidae				
Sabacon simoni	OP2522	Italy: Piemont, Prov. Cuneo, near Monesi di Triora (44.07381, 7.75028)	yes	OP2BSAB, OPRC4
S. cavicolens	OP721	Tennessee: Anderson Co., near Norris Dam Cave (36.221, −84.09)	yes	OP2BSAB, OPRC4
S. cavicolens	OP657	Virginia: Lee Co., Cave Spring Rec. Area (36.803, −82.92)	yes	OP2BSAB, OPRC4
S. cf. cavicolens	OP1283	North Carolina: Macon Co., Bullpen Bridge (35.015, 83.126)	yes	OP2BSAB, OPRC4
S. cf. cavicolens	OP691	North Carolina, Haywood Co., below Hebo Mtn. (35.686, −82.906)	yes	OP2BSAB, OPRC4
S. cf. cavicolens	OP1290	North Carolina: Transylvania Co., Looking Glass Creek (35.297, −82.767)	yes	OP2BSAB, OPRC4
S. cf. cavicolens	OP660	Virginia: Scott Co., Hwy 23/58, near Weber City (36.6341, −82.5604)	no	OP2, OPRC4

Table 2.—Continued.

Taxon	Voucher no.	Collection location (N lat., W long.)	mtDNA	EF-1α primers
S. cf. cavicolens	OP1535	North Carolina: Avery Co., Henson Creek, N of Ingalls (36.0374, −82.0420)	no	OP2, OPRC4
Travunioidea: Sclerobuninae				
Sclerobunus nondimorphicus	OP1056	Oregon: Clatsop County, Ecola SP (45.9221, −123.9767)	yes	OP3, OP4SCLER
Cyptobunus cavicolens	OP2143	Montana: Jefferson County, Lewis and Clark Caverns (45.8386, −111.8668)	yes	OPRC3SCLER; OPRC4
Sclerobunus robustus robustus	OP885	New Mexico: Sandoval County, Jemez Mtns (35.8384, −106.4044)	yes	OP4SCLER, OPRC3SCLER
Sclerobunus r. robustus	OP1149	Colorado: Dolores County, along Dolores River (37.7679, −107.9871)	yes	OP4SCLER, OPRC3SCLER
Sclerobunus r. robustus	OP1122	Colorado: Gunnison County, S of Gothic (38.9397, −106.4072)	yes	OP3, OPRC4
Sclerobunus r. robustus	OP1164	Colorado: Custer County, Wet Mtns (38.1335, −1105.1791)	yes	OP4SCLER, OPRC3SCLER
Grassatores: Phalangodidae				
Bishopella laciniosa	OP320	Tennessee: Johnson Co., Backbone Rock Rec Area, S Damascus (36.5948, −81.8173)	no	OP3, OPRC4LEI
Bishopella laciniosa	OP108	North Carolina: Jackson Co., Upper Falls, Whitewater River (35.0336, −83.0141)	no	OP3, OPRC4LEI

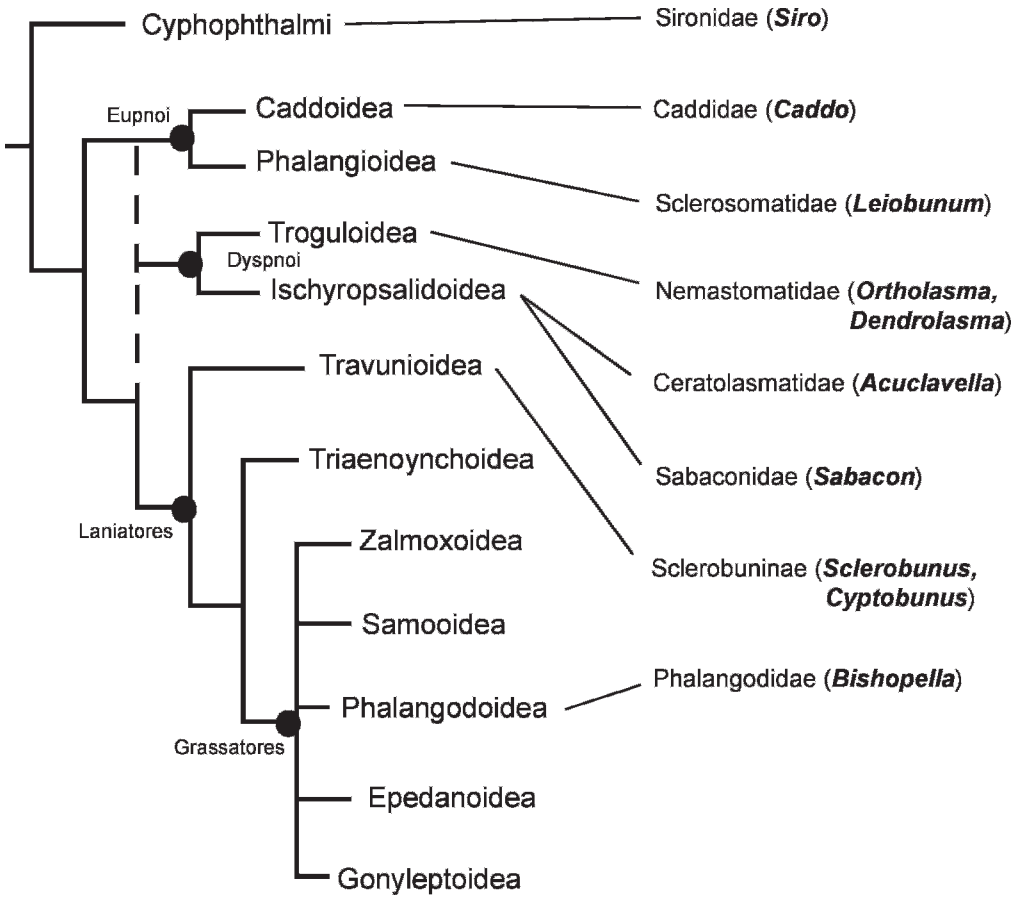

Figure 2.—Opiliones phylogeny with taxon sample. General tree structure follows Giribet & Kury (2007, fig. 3.4). The uncertain phylogenetic placement of Dyspnoi is indicated by a dashed line, representing a sister clade relationship to either Eupnoi (following Shultz & Regier 2001), or to Laniatores (e.g., Giribet et al. 1999).

parsimony search in PAUP v4.06 (Swofford 2002), using stepwise taxon addition from 1000 random replicates with tree bisection-reconnection (TBR) branch swapping. For the congeneric sets, we conducted exhaustive parsimony searches on both nuclear and mitochondrial matrices, conservatively treating EF-1α intron gaps as missing. Branch support was assessed using nonparametric bootstrap analyses (Felsenstein 1985), comprising 1000 pseudoreplicates of a branch-and-bound parsimony search for each congeneric set matrix.

RESULTS

Data availability, primer utility.—All sequences have been deposited to GenBank (accession numbers GQ870643–GQ870668; GQ872152–GQ872185). Both intron and exon alignments are available at www.treebase.org (study accession number S2469).

Using various primer combinations (Fig. 1, Table 2), we were able to amplify and generate EF-1α sequences for all surveyed genera, except for the caddoid *Caddo agilis*. We attempted eight different primer combinations for a single extraction from *Caddo*, using primers that worked well for other taxa. We suspect that the failure of this sample reflects a difference in sample preservation, as noted above. Overall, the most successful primers included the forward primers EF1-OP2, EF1-OP2BSAB, and EF1-OP3SCLER, combined with the reverse primer EF1-OPRC4LEI. The most 5′ (EF1-OP1) and 3′ (EF1-OPRC5PHAL) of our designed primers failed in all PCR experiments.

Features of the data.—Newly generated sequences were manually aligned with the exon-only nucleotide data of Shultz & Regier (2001) to identify the reading frame, intron insertions, and intron/exon boundaries. As a preliminary assessment of sequence orthology, we first removed all introns. The remaining exon data translate to expected amino acids, and with a single exception (see below) can be aligned without gaps. A strict consensus tree resulting from parsimony analysis of the exon-only data recovers expected phylogenetic placements for the newly-generated EF-1α sequences (Fig. 3), suggesting orthology of the gene copies used in this study. Ours is the first study to include the genus *Acuclavella* in a higher-level molecular analysis; we recover *Acuclavella* as sister to *Ceratolasma*, consistent with the hypothesis of Shear (1986, fig. 1).

Patterns of exon variability across divergent opilion taxa (e.g., *among* representatives of primary clades) have been addressed previously (Shultz & Regier (2001). Patterns of exon evolution *within* congeneric sets are summarized in Table 3. As expected, almost all exon variation occurs at third codon positions, with a total of 116, 14, and 7 variable sites at third, first, and second positions, respectively. Heterozygosity at exon positions is minimal, with less than 10 total sites scored as heterozygous for all newly-generated sequences.

We discovered three different introns in the harvestmen sample, here named introns I–III. The relative position of these introns, intron size, and intron/exon nucleotide boundaries are summarized in Figures 1 and 4. Intron I appeared only in two *Sabacon* samples (those amplified with EF1-OP2 primers; most *Sabacon* were amplified using downstream OP2BSAB primers). We also used the EF1-OP2 for *Leiobunum* (Fig. 1), but these sequences lack this intron. Examination of EF-1α genomic sequences for the tick *Ixodes scapularis* (http://iscapularis.vectorbase.org/, gene ISCW020299) indicates that intron I is shared in position between *Ixodes* and *Sabacon*. Intron I is relatively small in *Sabacon* (less than 100 bp), and includes 5′ (GT) and 3′ (AG) splice site signal sequences (see Fig. 4) consistent with proposed metazoan consensus signal sequences (Senapathy et al 1990; Mount et al. 1992).

Intron II was found in all *Leiobunum* and one *Ortholasma* sample (OP 807, amplified with OP2BSAB; *Ortholasma* OP 808 was amplified using downstream primers). These taxa are not phylogenetic relatives, and differ considerably at the exon level (see Fig. 3). We did not detect this intron in *Acuclavella* and some *Sabacon*, despite using primers that spanned this region (Fig. 1). This somewhat random phylogenetic pattern of intron presence/absence is most likely the result of differential intron loss (a similar argument applies for intron I), although we cannot rule out independent intron gains (e.g., see Roy & Penny 2006). Intron II varies in size (~ 100–200 bp), and to infer canonical 5′ (GT) and 3′ (AG) splice site signal sequences in *Ortholasma* requires a two amino acid deletion (see Fig. 4). We return to this inference in the Discussion.

Finally, intron III is found in all opilion samples that we have sequenced. This intron is shared in position with spiders (*Habronattus*, see Hedin & Maddison 2001), but not with *Ixodes*. Intron III varies considerably in size among harvestmen taxa, ranging from 60 to > 500 basepairs in length; all include canonical 5′ (GT) and 3′ (AG) splice site signal sequences (Fig. 4).

We generated high-quality intron I sequences for the taxa reported here, but in other *Sabacon* taxa a long (6-bp) poly-T region proved difficult to sequence through (data not reported). This provided our motivation to design *Sabacon*-specific primers (OP2BSAB) downstream of intron I. Intron II includes at least two long poly-T regions, but we had little difficultly in sequencing through this intron for several *Leiobunum* and a single *Ortholasma*. Our sample size is largest for intron III, where there is considerable variation in intron size (see above), nucleotide composition, and sequencing difficulty. The short (Fig. 4) *Acuclavella* introns were not difficult to sequence through, although these possess both poly-T and poly-A regions. Also, despite the long *Siro* intron III, and several simple-sequence regions, sequence reads were clean. Other taxa presented mixed success, even within congeneric sets, where some samples resulted in clean bi-directional reads, whereas other reactions resulted in clean reads that failed abruptly at poly-A/T regions. The sequence contigs for these latter samples comprise single strand reads from opposite directions, with minimal terminal overlap. Regions of terminal overlap that proved ambiguous to score were excluded from the phylogenetic analyses reported below.

Comparative phylogenetic utility.—The protein-coding mitochondrial data (CO1, ND1) translate to amino acids without stop codons, and can be aligned without gap insertions. We compared these mitochondrial data to intron plus exon EF-1α data for the four sets of close phylogenetic relatives. Introns II (*Leiobunum* only) and III (all taxa) were aligned manually, but for the six-taxon *Sclerobunus*/*Cyptobunus* set, introns from two specimens (OP1056, OP2143) were too divergent to align reliably. Here we conducted phylogenetic analyses on exons-only for the six-taxon set, and analyses of introns-only for four of six taxa.

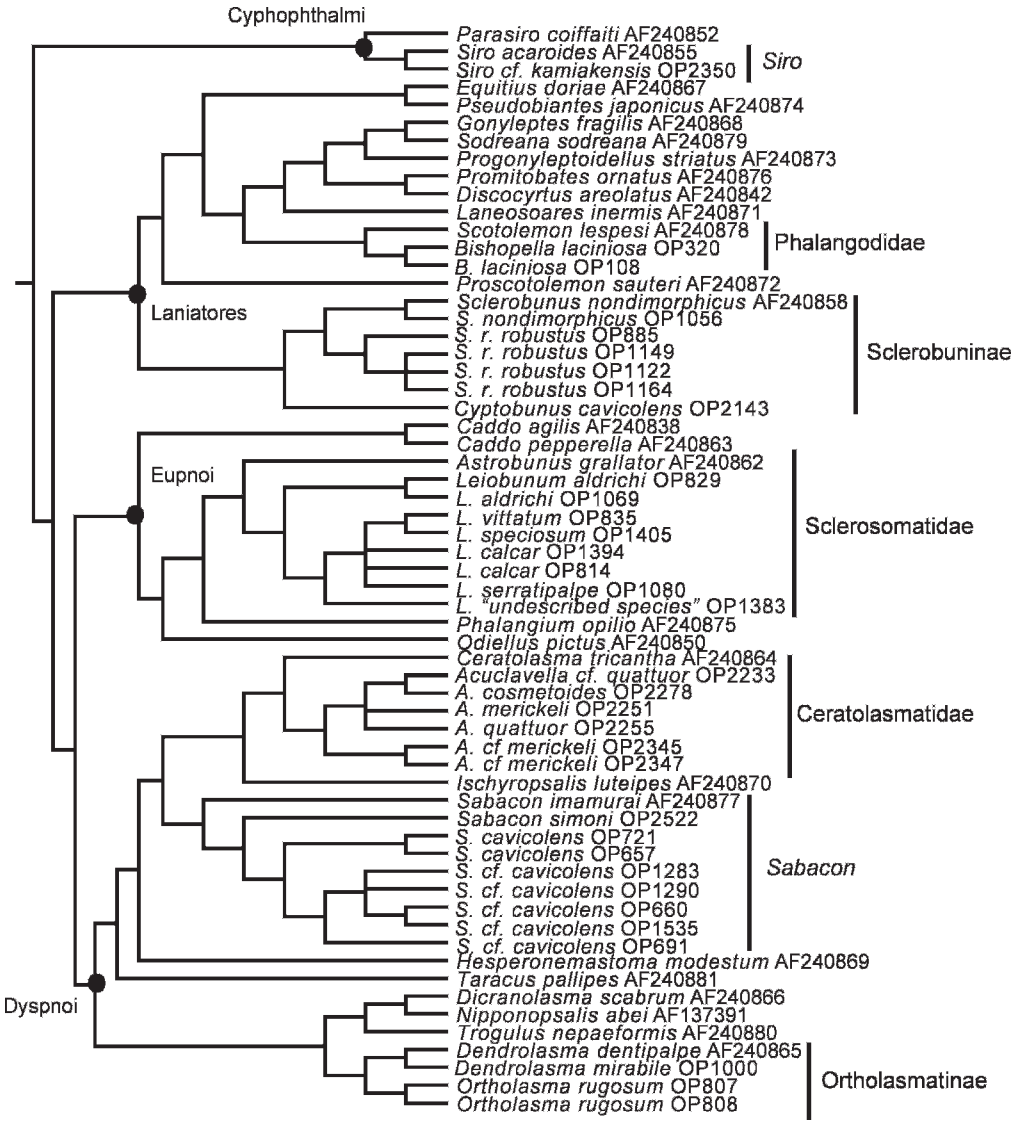

Figure 3.—Exon-only majority-rule consensus parsimony tree (N = 920, L = 2760). Higher level clade names are indicated, following taxonomy of Pinto-da-Rocha & Giribet (2007). Sequences generated in this study are shown with OP numbers; those from Shultz & Regier (2001) are shown with associated GenBank numbers. The *Sclerobunus* sequence (AF240858) generated by Shultz & Regier (2001) was originally misidentified as *S. robustus* - this sequence actually corresponds to *S. nondimorphicus*.

Results of the nine separate exhaustive parsimony searches (three for *Sclerobunus*/*Cyptobunus*, two for all other congeneric sets) are summarized in Fig. 5. These comparative analyses reveal several patterns of interest. First, all trees are well-resolved, with all but one of the searches resulting in a single most-parsimonious tree. Second, there is considerably more divergence depth in mitochondrial trees as compared to EF-1α trees, as indicated by differences in total tree length and individual branch length estimates. This reflects the well-documented higher rates of molecular evolution in animal mitochondrial genomes as compared to "average" nuclear genes (e.g., Moriyama & Powell 1997; Lin & Danforth 2004). Third, phylogenetic signal in the EF-1α data is coming from both intron and exon sites. This is most-clearly illustrated in the *Sclerobunus*/*Cyptobunus* taxon set where we conducted individual analyses on exons versus introns, but is also seen when comparing intron versus exon substitutions for specific "shallow" pairwise comparisons (see Table 3, Fig. 5). Finally, analyses of the independently-evolving mitochondrial versus EF-1α data result in generally similar gene tree estimates. Visual inspection indicates an overall pattern of topological and branch length similarity across data partitions. Moreover, we used the qualitative framework of Wiens (1998) to assess topological congruence. Among all trees, thirteen taxon bipartitions are strongly supported (bootstrap proportion values > 70; see Hillis & Bull 1993) by the mitochondrial data. Of these 13 bipartitions, 10 are also strongly supported by the EF-1α data (Fig. 5). Importantly, there is no evidence for "strong incongruence" (following Wiens 1998), where conflicting taxon bipartitions are strongly supported by the parallel datasets.

DISCUSSION

A general lack of genomic resources for Arachnida has hampered development of the type of multigenic molecular

Table 3.—Patterns of EF-1α sequence evolution. Variable intron sites were only counted at unambiguously aligned positions without indels. The following samples were not included in intron variation counts: *Sabacon* (OP660, OP1535, OP2522), *Sclerobunus*/*Cyptobunus* (OP1056, OP2143). Values for specific "shallow" pairwise comparisons (involving samples with mitochondrial divergence values between 2–3.2%) are shown in parentheses (see also Fig. 5).

	Variable Intron sites		Variable Exon sites		
Taxon	Intron II	Intron III	Pos 1	Pos 2	Pos 3
Sclerobunus/*Cyptobunus*	–	50 (10)	4	2	44 (1)
Acuclavella	–	16 (1)	2	2	16 (0)
Leiobunum	19 (1)	11 (3)	3	2	20 (1)
Sabacon	–	9 (1)	5	1	36 (4)

systematics perspective that is now seen in many other organismal groups (e.g., see Brito & Edwards 2008). In Opiliones, a core set of genes have been used at higher taxonomic levels (e.g., various mtDNA genes, 18S, 28S, histone 3, EF-1α), but nuclear gene choices at lower levels are extremely limited. We need to develop additional rapidly-evolving, informative nuclear genes for harvestmen. These will allow more accurate reconstructions of population and species' history, providing the necessary framework for addressing "shallow" systematic questions that abound in this diverse group (e.g., fine-scale historical biogeography, cryptic speciation, rapid character evolution, etc.).

The primer set and data reported here provide a starting point for other researchers to explore, modify, and utilize. The exon matrix now available (www.treebase.org, study accession number S2469) includes many conserved regions for primer design, and because the matrix includes taxonomic coverage for all higher-level harvestmen groups, it is easy to modify primers to match a group of interest. Furthermore, we have discovered an intron (intron III) that is most-parsimoniously reconstructed as being ubiquitous in Opiliones, with obviously conserved exon regions flanking this conserved position intron. With minor experimentation, it should be easy for researchers with access to standard PCR tools to generate both intron and exon EF-1a data for Opiliones.

The most conspicuous problem that we have identified is the presence of short simple sequence regions within introns, which because of apparent length heterozygosity, sporadically interrupt direct sequence reads. Here we would first recommend designing custom primers that span only a single intron

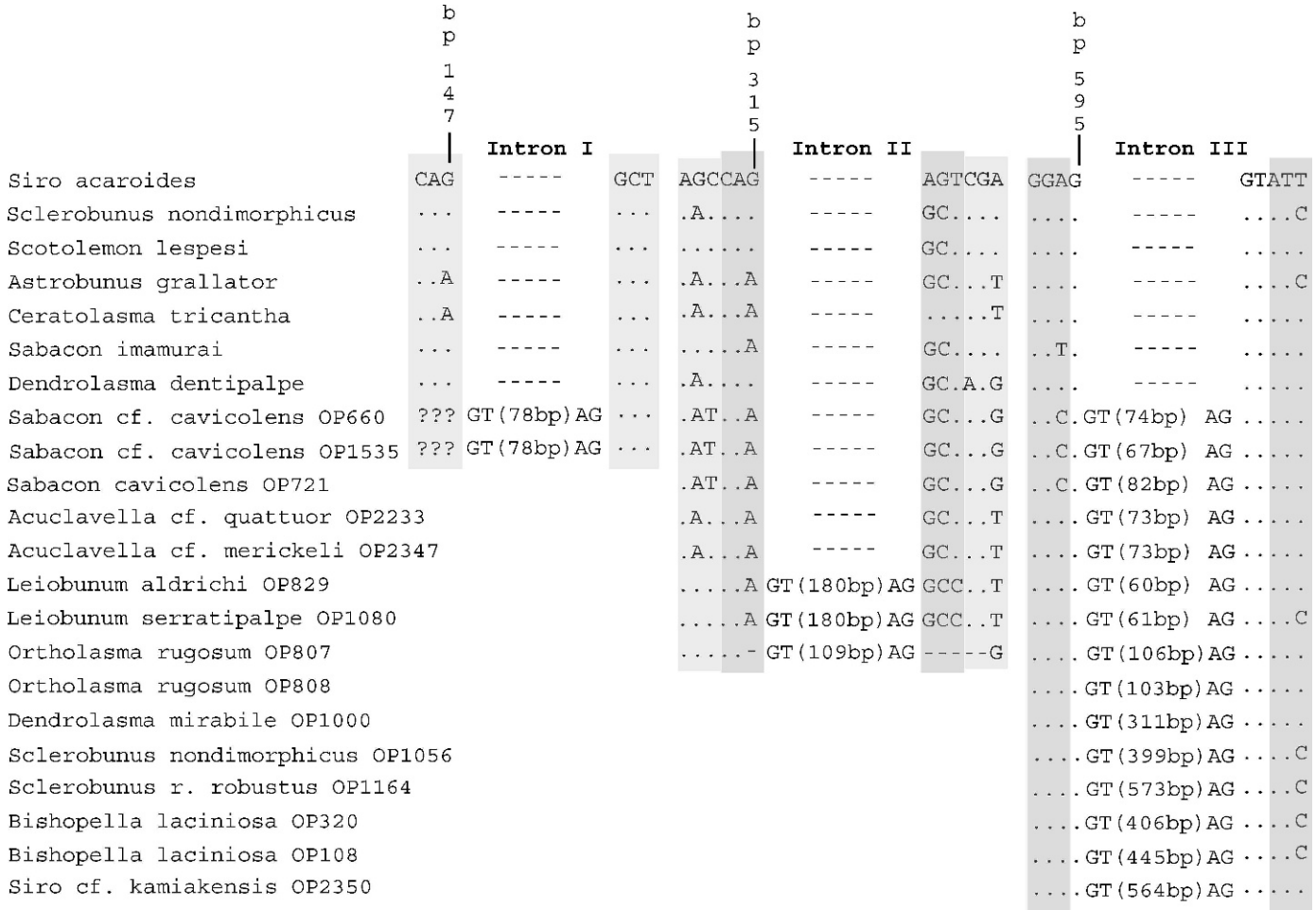

Figure 4.—Intron/Exon boundaries. Sequences generated in this study are shown with OP numbers, all those without are from Shultz & Regier (2001). Exon nucleotides are shaded in grey. Base positions are shown relative to alignment with exon sequences of Shultz & Regier (2001).

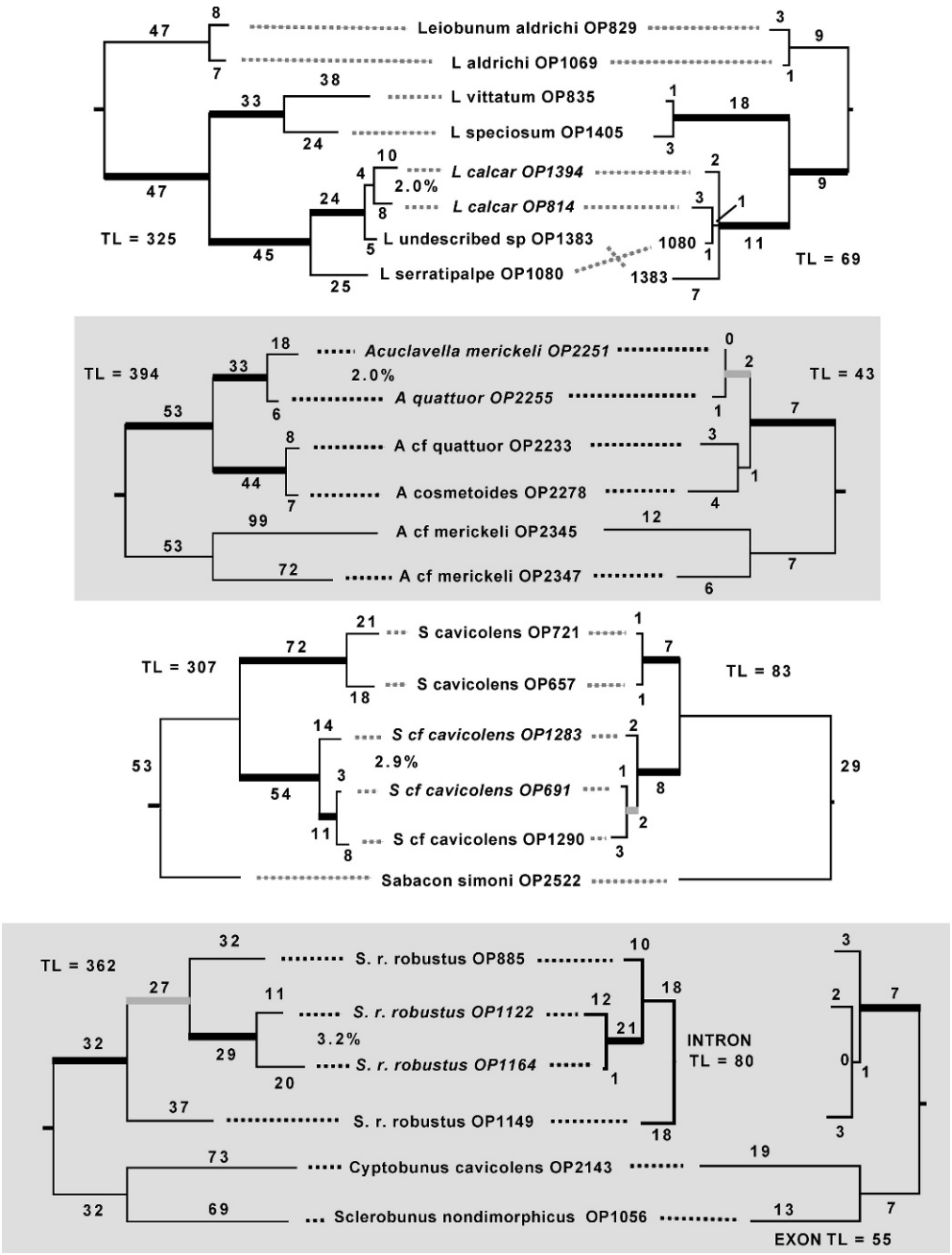

Figure 5.—Comparative mitochondrial (left) and EF-1α (right) parsimony trees. Number of parsimony inferred changes shown adjacent to branches. TL = length of most-parsimonious tree. Bootstrap proportion values above 90 are indicated by black thick branches; proportion values between 70–90 indicated by grey thick branches. All searches resulted in a single most-parsimonious tree, except for the *Leiobunum* EF-1α matrix, which resulted in 3 most-parsimonious trees (we illustrate only one of these). For mitochondrial trees, pairwise comparisons with divergence values between 2–3.2% (Kimura 2-parameter distances; Kimura 1980) are highlighted with bold, italicized text.

(after preliminary exploration of intron structure for the taxon of interest). Depending upon the severity of length heterozygosity, it may be necessary to clone PCR products, which is obviously more expensive and labor-intensive, but is still a widely used lab method. Other alternatives include allele-specific PCR, or bioinformatics approaches, such as the use of software programs that allow phase determination of length variable alleles resulting from direct sequencing (see Flot et al 2006; Flot 2007).

The data reported here reveal general congruence between nuclear versus mitochondrial gene trees (Fig. 5). Gene tree congruence, or lack thereof, is a reflection of several interacting variables, including sampling density, the depth and divergence history of the group of interest, and patterns of gene evolution (see Maddison 1997). For example, older, well-spaced speciation events are expected to result in more gene tree congruence than more-recent, ''compressed-in-time'' radiations. Gene flow across species boundaries in recently

separated species can cause incongruence among gene trees, as can comparisons of paralogous gene copies in multi-gene families.

Given our sample of taxa and exemplars within these taxa, we observed minimal incongruence, but again, this is an empirical issue that is expected to vary from group to group. In this context, one relevant question is how old are the harvestmen taxa within congeneric sets? We consider two indirect lines of evidence for estimating the ages of these groups. First, we have highlighted on Fig. 5 pairwise sample comparisons that differ by 2–3 percent for mitochondrial protein-coding genes. Applying a ''standard'' arthropod mitochondrial molecular clock (e.g., 2.3% pairwise per million years; Brower 1994; Pons et al. 2006) across trees for these individual groups would indicate divergences ranging from 1–8 million years ago (MYA). These are, of course, very rough estimates, with many assumptions which may or may not hold (a standard clock applies to Opiliones, no among-lineage rate variation, etc). Second, there is strong evidence for ancient vicariance within *Acuclavella*, leading to the primary phylogenetic separation of Olympic and Cascade mountain populations (OP2345, OP2347) from all other populations in the Idaho Rockies. In many other taxa from this region (e.g., Carstens et al. 2004; Steele et al. 2005), this west/east vicariance results from the Cascadian orogeny at approximately 5 MYA, which suggests a divergence time window generally consistent with the above molecular clock estimates.

The above divergence time estimates suggest that EF-1α data might be used to resolve divergences that are quite recent in absolute time. Although not extensive, all pairwise sample comparisons at the ''one million year horizon'' also show at least some divergence in EF-1α (see Table 3). Furthermore, we note that our phylogenetic treatment of insertion and deletion (indel) variation in introns was very conservative, as several indel sites appeared phylogenetically informative. More sophisticated treatment of this class of variation is expected to further increase the phylogenetic informativeness of this gene region for recent divergences (e.g., Simmons & Ochoterena 2000; Kawakita et al. 2003; Benavides et al. 2007).

In terms of gene tree congruence, one variable that seems not to be influencing our data is the presence of multiple gene copies, and potential mixing of paralogous copies. The paralogy/orthology issue is one of the most important problems negatively impacting the phylogenetic utility of nuclear protein-coding genes (reviewed in Sanderson & Shaffer 2002). Multiple gene copies are in fact known for EF-1α in other arthropods (see Danforth et al. 1999), including spiders (Hedin & Maddison 2001). ''Deep paralogy,'' or the inadvertent comparison of *divergent* gene copies (e.g., gene copies resulting from gene duplication before the divergence of Opiliones), is apparently not an issue in our data, as mixing such paralogs would not result in the expected systematic relationships that we see in the exon-only analysis (Fig. 3). An alternative is that harvestmen species are carrying multiple copies of the EF-1α gene that are kept similar by concerted evolution (as hypothesized for salticid spiders, Hedin & Maddison 2001); distinguishing such closely-related paralogs would be more difficult in a phylogenetic analysis.

The only hint of such ''cryptic'' paralogy is the *Ortholasma* OP807 sequence, which requires a 2 amino-acid exon gap to align intron II splice site signal sequences (see Fig. 4). Such a gap is never present in other exon sequences, even from close phylogenetic relatives (e.g., *Dendrolasma*). This sequence may represent a paralog, but the exon appears functional (no stop codons), and is placed as expected phylogenetically. A perhaps more plausible explanation is that we have misaligned the intron boundary, but this requires non-canonical splice sites for this intron. More data is needed to distinguish between these alternative hypotheses. In general, there is little to no evidence for multiple copies of EF-1α in harvestmen. Researchers in prior studies (Shultz & Regier 2001, 2009) have not suggested this problem, relationships inferred from exons provide expected phylogenetic results, and gene trees for congeneric sets are largely congruent with independent data. However, the possibility of such paralogy should not be dismissed or overlooked in future studies.

Various technologies are now available for the generation of comparative nuclear sequence data. These include expressed sequence tag (EST) libraries, next-generation sequencing (Mardis 2008), development of anonymous nuclear loci (e.g., Noonan & Yoder 2009), etc. Ultimately, these technologies may supercede studies that focus on a single ''candidate gene,'' for example, if we can ultimately generate whole genomic data for all samples of interest. However, these technologies are still too expensive for the average systematist, particularly in geographic areas where resources for systematics research are limited. Until then, we argue that studies of the phylogenetic behavior of ''candidate genes,'' such as the research reported here, still have considerable utility in the arachnological and broader systematics communities.

ACKNOWLEDGMENTS

We thank Joe Deas, Damian Elias, Dalton Hedin, Lars Hedin, Robin Keith, Jim Starrett, Steven Thomas and Jeff Underwood for assistance in collecting specimens. Axel Schönhofer provided important European *Sabacon* specimens. Personnel at Lewis & Clark Caverns State Park were helpful in our collections of *Cyptobunus cavicolens*. Specimens collected in the Great Smoky Mountains National Park were collected under a permit issued to Steven Thomas. Fred Coyle, Steve Perlaky, and the Abbott family provided excellent accommodations while collecting in southern Appalachia. J.W.S. was supported by the Maryland Agricultural Experiment Station. Funding was provided by a collaborative NSF grant awarded to J. Shultz (Award No. 0640179) and M. Hedin (Award No. 0640173).

LITERATURE CITED

Ballard, J.W. & M.C. Whitlock. 2004. The incomplete natural history of mitochondria. Molecular Ecology 13:729–744.

Benavides, E., R. Baum, D. McClellan & J.W. Sites, Jr. 2007. Molecular phylogenetics of the lizard genus *Microlophus* (Squamata: Tropiduridae): aligning and retrieving indel signal from nuclear introns. Systematic Biology 5:776–797.

Bower, J.E., R.D. Cooper & N.W. Beebe. 2009. Internal repetition and intra-individual variation in the rDNA ITS1 of the *Anopheles punctulatus* group (Diptera: Culicidae): multiple units and rates of turnover. Journal of Molecular Evolution 68:66–79.

Brady, S.G. & B.N. Danforth. 2004. Recent intron gain in Elongation factor-1α of colletid bees (Hymenoptera: Colletidae). Molecular Biology and Evolution 21:691–696.

Brito, P.H. & S.V. Edwards. 2008. Multilocus phylogeography and phylogenetics using sequence-based markers. Genetica 135: 439–455.

Brower, A.V.Z. 1994. Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. Proceedings of the National Academy of Sciences USA 91:6491–6495.

Carstens, B.C., A.L. Stevenson, J.D. Degenhardt & J. Sullivan. 2004. Testing nested phylogenetic and phylogeographic hypotheses in the *Plethodon vandkei* species group. Systematic Biology 53:781–792.

Caterino, M.S., S. Cho & F.A.H. Sperling. 2000. The current state of insect molecular systematics: a thriving Tower of Babel. Annual Review of Ecology and Systematics 45:1–54.

Cho, S., A. Mitchell, J.C. Regier, C. Mitter, R.W. Poole, T.P. Friedlander & S. Zhao. 1995. A highly conserved nuclear gene for low-level phylogenetics: elongation factor–1α recovers morphology-based tree for Heliothine moths. Molecular Biology and Evolution 12:650–656.

Danforth, B.N. & S. Ji. 1998. Elongation factor-1α occurs as two copies in bees: Implications for phylogenetic analysis of EF-1α sequences in insects. Molecular Biology and Evolution 15:225–235.

Danforth, B.N., H. Sauquet & L. Packer. 1999. Phylogeny of the bee genus *Halictus* (Hymenoptera: Halictidae) based on parsimony and likelihood analyses of nuclear EF-1α sequence data. Molecular Phylogenetics and Evolution 13:605–618.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791.

Flot, J.-F. 2007. Champuru 1.0: a computer software for unraveling mixtures of two DNA sequences of unequal lengths. Molecular Ecology Notes 7:974–977.

Flot, J.-F., A. Tillier, S. Samadi & S. Tillier. 2006. Phase determination from direct sequencing of length-variable DNA regions. Molecular Ecology Notes 6:627–630.

Giribet, G., M. Rambla, S. Carranza, J. Baguna, M. Riutort & C. Ribera. 1999. Phylogeny of the arachnid order Opiliones (Arthropoda) inferred from a combined approach of complete 18S and partial 28S ribosomal DNA sequences and morphology. Molecular Phylogenetics and Evolution 11:296–307.

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

Giribet, G. & A.B. Kury. 2007. Phylogeny and biogeography. Pp. 62–87. *In* Harvestmen: The Biology of Opiliones. (R. Pinto-da-Rocha, G. Machado & G. Giribet, eds.). Harvard University Press, Cambridge, Massachusetts.

Harris, D.J. & K.A. Crandall. 2000. Intragenomic variation within ITS1 and ITS2 of freshwater crayfishes (Decapoda: Cambaridae): implications for phylogenetic and microsatellite studies. Molecular Biology and Evolution 17:284–291.

Hedin, M.C. & W.P. Maddison. 2001. Phylogenetic utility and evidence for multiple copies of elongation factor-1α in the spider genus *Habronattus* (Araneae: Salticidae). Molecular Biology and Evolution 18:1512–1521.

Hedin, M.C. & S.M. Thomas. 2010. Molecular systematics of eastern North American Phalangodidae (Arachnida: Opiliones: Laniatores), demonstrating convergent morphological evolution in caves. Molecular Phylogenetics and Evolution. 54:107–121.

Hillis, D.M. & J.J. Bull. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Systematic Biology 42:182–192.

Kawakita, A., T. Sota, J.S. Ascher, M. Ito, H. Tanaka & M. Kato. 2003. Evolution and phylogenetic utility of alignment gaps within intron sequences of three nuclear genes in bumble bees (*Bombus*). Molecular Biology and Evolution 20:87–92.

Kimura, M. 1980. A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16:111–120.

Lin, C.-P. & B.N. Danforth. 2004. How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets. Molecular Phylogenetics and Evolution 30:686–702.

Maddison, W.P. 1997. Gene trees in species trees. Systematic Biology 46:523–536.

Maddison, D.R. & W.P. Maddison. 2003. MacClade 4, Release Version 4.07. Sinauer Associates, Sunderland, Massachusetts.

Mardis, E.R. 2008. Next-generation DNA sequencing methods. Annual Review of Genomics & Human Genetics 9:387–402.

Moriyama, E.N. & J.R. Powell. 1997. Synonymous substitution rates in *Drosophila*: mitochondrial versus nuclear genes. Journal of Molecular Evolution 45:378–391.

Mount, S.M., C. Burks, G. Hertz, G.D. Stormo, O. White & C. Fields. 1992. Splicing signals in *Drosophila*: intron size, information content, and consensus sequences. Nucleic Acids Research 20:4255–4262.

Noonan, B.P. & A.D. Yoder. 2009. Anonymous nuclear markers for Malagasy plated lizards (*Zonosaurus*). Molecular Ecology Resources 9:402–404.

Palumbi, S.R. 1996. Nuclei acids II: the polymerase chain reaction. Pp. 205–247. *In* Molecular Systematics, Second edition. (D.M. Hillis, C. Moritz & B.K. Mable, eds.). Sinauer Associates, Sunderland, Massachusetts.

Pinto-da-Rocha, R. & G. Giribet. 2007. Taxonomy. Pp. 88–246. *In* Harvestmen: The Biology of Opiliones. (R. Pinto-da-Rocha, G. Machado & G. Giribet, eds.). Harvard University Press, Cambridge, Massachusetts.

Pons, J., T.G. Barraclough, J. Gomez-Zurita, A. Cardoso, D.P. Duran, S. Hazell, S. Kamoun, W.D. Sumlin & A.P. Vogler. 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Systematic Biology 55:595–609.

Roy, S.W. & D. Penny. 2006. Smoke without fire: most reported cases of intron gain in nematodes instead reflect intron losses. Molecular Biology and Evolution 23:2259–2262.

Rubinoff, D. & B.S. Holland. 2005. Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. Systematic Biology 54:952–961.

Sanderson, M.J. & H.B. Shaffer. 2002. Troubleshooting molecular phylogenetic analyses. Annual Review of Ecology and Systematics 33:49–72.

Senapathy, P., M.B. Shapiro & N.L. Harris. 1990. Splice junctions, branch point sites, and exons: sequence statistics, identification, and application to genome project. Methods in Enzymology 183:252–278.

Sharma, P. & G. Giribet. 2009. Sandokanid phylogeny based on eight molecular markers —the evolution of a southeast Asian endemic family of Laniatores (Arachnida, Opiliones). Molecular Phylogenetics and Evolution 52:432–477.

Shear, W.A. 1986. A cladistic analysis of the Opilionid superfamily Ischyropsalidoidea, with descriptions of the new family Ceratolasmatidae, the new genus *Acuclavella*, and four new species. American Museum Novitates 2844:1–29.

Shultz, J.W. & J.C. Regier. 2001. Phylogenetic analysis of Phalangida (Arachnida, Opiliones) using two nuclear protein-encoding genes supports monophyly of Palpatores. Journal of Arachnology 29:189–200.

Shultz, J.W. & J.C. Regier. 2009. Pattern and timing of phylogenetic events in *Caddo* and implications for the biogeography of Caddidae (Arachnida: Opiliones). Journal of Arachnology 37:238–241.

Simmons, M.P. & H. Ochoterena. 2000. Gaps as characters in sequence-based phylogenetic analyses. Systematic Biology 49:369–381.

Steele, C.A., B.C. Carstens, A. Storfer & J. Sullivan. 2005. Testing hypotheses of speciation timing in *Dicamptodon copei* and

Dicamptodon aterrimus (Caudata: Dicamptodontidae). Molecular Phylogenetics and Evolution 36:90–100.

Swofford, D.L. 2002. PAUP* Phylogenetic Analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Thomas, S.M. & M. Hedin. 2008. Multigenic phylogeographic divergence in the paleoendemic southern Appalachian opilionid *Fumontana deprehendor* Shear (Opiliones, Laniatores, Triaenonychidae). Molecular Phylogenetics and Evolution 46:645–658.

Wiens, J.J. 1998. Combining data sets with different phylogenetic histories. Systematic Biology 47:568–581.

Manuscript received 20 May 2009, revised 18 August 2009.