Molecular systematics of sclerosomatid harvestmen (Opiliones, Phalangioida, Sclerosomatidae): Geography is better than taxonomy in predicting phylogeny

Molecular Phylogenetics and Evolution 62 (2012) 224–236

Molecular systematics of sclerosomatid harvestmen (Opiliones, Phalangioida, Sclerosomatidae): Geography is better than taxonomy in predicting phylogeny

Marshall Hedin a,*, Nobuo Tsurusaki b, Rogelio Macías-Ordóñez c, Jeffrey W. Shultz d

a Department of Biology, San Diego State University, San Diego, CA 92182–4614, USA
b Laboratory of Biology, Faculty of Regional Sciences, Tottori University, Tottori 680-8551, Japan
c Red de Biología Evolutiva, Instituto de Ecología A.C., Xalapa, Veracruz 91070, Mexico
d Department of Entomology, University of Maryland, College Park, MD 20742, USA

ARTICLE INFO

Article history:
Received 4 March 2011
Revised 24 September 2011
Accepted 9 September 2011
Available online 7 October 2011

Keywords:
Morphological homoplasy
Boreotropic Concept
Biogeography
Phylogenetics
Arachnida

ABSTRACT

Phylogenetic relationships within the Sclerosomatidae, the largest family of harvestmen, are explored using molecular data from four nuclear genes (28S and 18S rRNA, Histone 3 and Elongation factor-1α) and two mitochondrial gene regions (COI-COII, 16S and 12S rRNA). The taxon sample includes representative species from all families in Phalangioida and all subfamilies of Sclerosomatidae (Gagrellinae, Gymnae, Leiothuninae, Sclerosomatinae). Our results solve several major taxonomic problems, including placement of Gymnae sensu stricto in Phalangiidae, the monophyly of the Metopilio group and its exclusion from Sclerosomatidae, and reaffirmation of the familial rank of Protolophidae. However, most major groups of sclerosomatids (Leiothuninae, Gagrellinae, Leiothunum, Neolima) are recovered as polyphyletic, although with a phylogenetic structure suggesting a strong association between geography and monophyly as well as notable morphological convergence in traditional diagnostic characters. Phylogenetic affinities between biotas of the New World and Asian tropics, as well as between temperate North American and East Asia, suggest that sclerosomatid historical biogeography may conform with the Boreotropic Concept. Finally, we discuss how the many problems that remain in sclerosomatid systematics might be addressed.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

The arachnid order Opiliones is taxonomically rich, comprising 45 families, approximately 1500 genera, and more than 6400 described species (summarized in Machado et al., 2007; Kury, 2011). Multiple lines of evidence support the partitioning of Opiliones diversity into four primary clades (Fig. 1A), including the 2011). Multiple lines of evidence support the partitioning of Opiliones diversity into four primary clades (Fig. 1A), including the

1800 described species in over 200 genera), are abundant and conspicuous members of forest arthropod communities, and are clearly the most recognizable Opiliones in the Northern Hemisphere, there has been remarkably little higher-level systematic research focused on this clade of arachnids.

Researchers have not yet reached a consensus regarding the classification of Phalangioida; for purposes of this paper, we will use the classification scheme outlined in Table 1 (mostly following Crawford, 1992), which also summarizes the diversity and distribution of currently classified groups. A basal assemblage of southern temperate phalangioids from southern South America, South Africa, southern Australia, and New Zealand (Hunt and Cokendolpher, 1991; Cokendolpher and Taylor, 2007; Taylor, 2011) is relatively species-poor and lacks spiracular closing mechanisms (=“non-entapophysate”). These Gondwanan phalangioids have historically been classified into two families, the Monoscutitae and Neopilionidae (Hunt and Cokendolpher, 1991). Based on morphological phylogenetic analyses, (Taylor, 2011) recently transferred all non-entapophysate taxa into the Neopilionidae. The more diverse tropical and northern temperate entapophysate phalangioids possess respiratory spiracles with an internal closing mechanism. Recent workers have recognized three entapophysate families (Phalangiidae, Protolophidae, Sclerosomatidae) and a group of uncertain status, the Metopilio group (Gruber, 1969; Cokendolpher, 1984a). Within the Sclerosomatidae we recognize the subfamilies Gymnae, Sclerosomatinae, Gagrellinae and Leiothuninae, acknowledging that these groups are poorly delimited (see also Cokendolpher et al., 2007). For example, past researchers have commented...
Table 1
Taxonomic Summary of the Phalangioidea.

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Number of described genera/species</th>
<th>Geographic distribution</th>
<th>Defining features</th>
<th>Key references</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neopilionidae</td>
<td>13/49</td>
<td>Southern South America, South Africa, Australia, New Zealand</td>
<td>Spiracle with grate</td>
<td>Forster, 1948; Hunt and Cokendolpher, 1991; Cokendolpher, 2007a; Taylor, 2011</td>
</tr>
<tr>
<td>Phalangiidae</td>
<td>47/408</td>
<td>Palearctic, Nearctic, Africa (introduced elsewhere)</td>
<td>Pedipalpal claw smooth, soft body, spiny legs, penis morphology</td>
<td>Summarized in Tsurusaki, 2007</td>
</tr>
<tr>
<td>Protolophidae</td>
<td>1/8</td>
<td>Western North America</td>
<td>Structure of male palps; penis morphology</td>
<td>Cokendolpher, 2007b</td>
</tr>
<tr>
<td>Sclerosomatinae</td>
<td>3/10</td>
<td>Palearctic</td>
<td>Coxal II lobes; penis morphology</td>
<td>Summarized in Tourinho, 2007</td>
</tr>
<tr>
<td>Gagrellinae</td>
<td>117/1004</td>
<td>Neotropics; Indo-Malayan tropics</td>
<td>Most with nodules on leg femora</td>
<td>Šilhavý, 1946; Martens, 1982</td>
</tr>
<tr>
<td>Leiobuninae</td>
<td>14/192</td>
<td>Holarctic, to Central America</td>
<td>Weakly sclerotized dorsum, no femoral nodules (most)</td>
<td>Gruber, 1969; Cokendolpher, 1984a</td>
</tr>
<tr>
<td>Metopilio group</td>
<td>5/32</td>
<td>Western United States to central America</td>
<td>Alate penis with moveable stylus (some); scent pores visible from above, palpal claws smooth</td>
<td></td>
</tr>
</tbody>
</table>

Notes:
- Taxonomic classification following Crawford (1992), except for family-level status of Protolophidae (following Cokendolpher and Lee, 1993) and configuration of Neopilionidae (following Taylor, 2011).
- Taxonomic richness data from Table 4.1 of Pinto-da-Rocha et al., 2007, except for Metopilio group data from Tourinho (2005).
- Defining features from Cokendolpher et al., 2007 and listed primary references.
on morphological intermediacy among Sclerosomatinae, Gagrellinae and Leiobuninae (Martens, 1973, 1982).

Our systematic knowledge of entapophysate phalangioids varies geographically. The very diverse gagrelline faunas of the Asian and New World tropics have received limited systematic attention, with the bulk of early revisionary work conducted by Roewer (1910, 1923, 1953, 1954, 1955). Unlike modern researchers, Roewer did not consider male genital (penis) morphology in his work, and we expect these classifications to be artificial to varying degrees. Referring to Neotropical gagrellines, Tourinho and Kury (2001) state that “the classification system by Roewer is arbitrary, and the generic groups … are formed by species with morphological patterns that do not make any phylogenetic sense.” In contrast, the faunas of North America, western Europe, Japan, and the Himalayas are comparatively well-known, having been studied by multiple researchers who have explicitly considered penis morphology and other non-Roewerian features for classification purposes (e.g., Davis, 1934; Suzuki, 1976; Cokendolpher, 1984a; Starega, 1984; Tsurusaki, 1985; Martens, 1987). We expect classified groups in these regions to correspond more closely to natural groups, but do not rule out the possibility of misclassification due to either somatic or genitalic homoplasy. The basic structure of the penis in Sclerosomatidae is relatively simple (Macías-Ordóñez et al., 2010), and therefore likely prone to homoplasy.

Molecular phylogenetic research involving phalangioids has been limited. Giribet et al. (1999) included two phalangioid genera in their ribosomal RNA-based research, recovering these as sister taxa. Shultz and Regier (2001) used two nuclear protein-coding genes to also resolve a monophyletic Phalangiidea, but sampled only three genera. Giribet et al. (2002) and Garwood et al. (2011) used both morphology and molecules (18S, 28S rRNA) to resolve a monophyletic Phalangiidea, with independent lineages conforming to the Metopilio group (represented by Dalquestia), Phalangiidae (three genera sampled), and the Sclerosomatidae (four genera sampled). The most comprehensive work to date is that of Giribet et al. (2010), who sampled 16 total phalangioid genera, representing all primary hypothesized groups except for the Gyinae (Sclerosomatidae). Using multiple genes (although not all genes were available for all taxa), these authors recovered Phalangiidae, Monoscutidae (=Neopilionidae), and a Protolophus plus Sclerosomatidae clade, all with strong support (Fig. 1B). Several other hypothesized groups were not recovered as monophyletic, including the Metopilio group, and the sclerosomatid subfamilies Leio- buninae, Gagrellinae, and Sclerosomatinae. All four latter groups failed the test of monophyly even with small generic samples.

The research reported here focuses on entapophysate phalangioids, with a sampling of over 30 genera that emphasizes northern temperate sclerosomatids. We do not rigorously test the monophyly of Phalangiidae, but have sampled the genus Gyas to test a possible Gyinae plus Phalangiidae relationship (Martens, 1978). We have sampled four of five recognized genera of the New World Metopilio group, allowing a more robust test of its monophyly. We include multiple members of the genus Protolophus to assess the monophyly of the Protolophidae. Most conspicuously, we have sampled essentially all described sclerosomatid genera from North America and Japan. These faunas include a mix of mostly gagrellines and leiobunines, which allows us to comprehensively test whether these traditional subfamilies are monophyletic (i.e., span continents), or whether clades are defined more by biogeographical affinities. Similarly, these faunas also include genera (e.g., Leiobunum, Nelima) that span multiple continents, including Europe. Our sample allows us to address whether these genera are true trans-continental taxa, or conversely, whether superficially similar taxa have evolved independently on multiple continents. Our results clarify the phylogenetic structuring of Japanese and North American faunas, clearly illustrating a pattern of regional diversification, gagrelline to leiobunine paraphyly, and massive generic-level polyphyly. Several cases of homoplasy in somatic and penis morphology are revealed. Although the high generic richness and worldwide distribution of the Sclerosomatidae precludes a comprehensive phylogenetic analysis at present, our results point in obvious directions for future research.

2. Material and methods

Specimen identity and collection information is summarized in Supplementary Table I. All specimens were identified by the authors using primary literature, unless noted otherwise (see Supplementary Table I). We attempted to use only adult specimens for DNA study, but in some cases used immature specimens; these immature specimens were collected and identified to species by regional experts (see Supplementary Table I). For Mexican taxa in particular, some species determinations are tentative, and some taxa are identified to genus or subfamily only; this taxonomic uncertainty reflects a poorly known Mexican fauna, and a taxonomic literature that is sometimes difficult to interpret (e.g., inadequate original descriptions, no study of genitalia, etc.; Kury and Cokendolpher, 2000). All specimens have been assigned a unique specimen identification number, and are currently housed in the SDSU Terrestrial Arthropods collection. To facilitate future work, all voucher specimens have been digitally photographed, with digital images deposited at Morphbank (http://www.morphbank.net/).

Basal relationships within the Phalangioidea remain unclear. Taylor (2011, Fig. 43) recovered a sister–taxon relationship between non-entapophysate (=Neopilionidae) and entapophysate phalangioids in certain implied weights morphological phylogenetic analyses. However, Taylor (2011) included only three entapophysate taxa in his analyses (e.g., no Protolophidae, no Metopilio group), and in other analyses an entapophysate clade was derived from within a grade of non-entapophysate taxa (Fig. 41). Similar results were seen in molecular analyses of Giribet et al. (2010), where again non-entapophysate and entapophysate phalangioids are not recovered as sister clades (see Fig. 1B). Given this lack of resolution at the base of Phalangioidea we used published sequences of more distant Dyspnoi and Caddoiidea (see Fig. 1A) to root our molecular phylogenetic trees; we stress that our taxon sampling at these deep levels is mostly opportunistic and very sparse.

2.2. Gene data collection, sequence alignment and phylogenetic analysis

Genomic DNA was extracted from leg tissues using the Qiagen DNeasy Blood & Tissue kit, utilizing standard procedure. Polymerase chain reaction (PCR) was used to amplify six gene regions, including 18S ribosomal RNA, 28S ribosomal RNA, Histone 3 (H3), Elongation factor – 1 alpha (EF-1χ), contiguous mitochondrial Cytochrome oxidase subunits I and II (COI and COII), and contiguous mitochondrial 16S and 12S ribosomal genes. Information regarding amplification conditions and primers used in amplification and/or direct sequencing is included in Supplementary Table II. PCR amplicons were purified using either PEG precipitation or on Millipore plates. Purified amplicons were directly sequenced at the SDSU Microchemical Core Facility or at Macrogen USA. Sequencher software was used to edit and assemble sequence contigs, with all ambiguous sites scored as heterozygosity using standard ambiguity codes. GenBank sequences were incorporated into respective matrices as appropriate (see Section 3 and Supplementary Table I).
Nucleotide data for five of six gene regions was aligned manually in MacClade 4.0 (Maddison and Maddison, 2000). This included 18S and 28S rRNA data, which was straightforward to align manually (see Section 3). The mitochondrial 16S, 12S gene regions were aligned using MAFFT version 6 online (http://mafft.cbrc.jp/alignment/software/), applying the E-INS-I manual strategy with default parameters. To explore alternative alignment strategies, we also aligned the 28S rRNA data using this MAFFT approach. Models of DNA sequence evolution were chosen using jModelTest 0.1.1 (Posada, 2008); model likelihoods were calculated under three substitution schemes (JC, HKY, GTR) on a fixed BIONJ tree, allowing for unequal base frequencies and among-site rate variation. From these likelihood scores model selection was based on the Akaike Information Criterion (AIC).

Bayesian phylogenetic analyses were conducted using MrBayes v3.1 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) on six individual data partitions, and a matrix of combined data. For Bayesian analysis, non-coding gene regions were analyzed using a single best-fit model (i.e., unpartitioned). Protein-coding data were analyzed using a partitioned strategy (see Ronquist and Huelsenbeck, 2003; Nylander et al., 2004), using a three partitions model for the mitochondrial COI/COII data. Because relatively few second codon position sites were variable in the EF-1x and H3 data (six and four, respectively), we used a simpler two partitions model for these data (i.e., P1 = first and second codon sites, P2 = third codon sites). The same sequence models used in separate Bayesian analyses were used in a partitioned analysis of the combined data. For partitioned and combined analyses, estimated parameters (revmat, statefreq, gamma shape, pinvar) for each partition were “unlinked”. Default cold and heated chain parameters were used in all analyses. Searches were run until the average standard deviation of split frequencies (compared across two independent searches) dropped below 0.01 (Ronquist et al., 2005). We also assessed convergence by examining the potential scale reduction factor (PSRF) for model parameters, which is expected to approach one with convergence (Brooks and Gelman, 1998). The first 20–40% of tree topologies were discarded as burn-in; from this post burn-in tree set we generated a majority rule consensus tree with mean branch-length estimates. Split frequencies were interpreted as posterior probabilities of clades.

The combined data matrix was also analyzed using a heuristic parsimony search, using 1000 random addition sequence replicates, tree-bisection-reconnection (TBR) branch-swapping, and treating gaps as missing (PAUP v 4.0b10, Swoford, 2000). Parsimony support was assessed using 1000 nonparametric bootstrap pseudoreplicates, each pseudoreplicate consisting of 10 TBR random addition sequence replicates.

3. Results

All newly generated DNA sequences have been deposited to GenBank (www.ncbi.nlm.nih.gov/Genbank/), with accession numbers available in Supplementary Table I. We have included alignments at GenBank as popsets, and have uploaded parsimony and Bayesian consensus trees to the Interactive Tree of Life page (Letunic and Bork, 2006, 2011; http://itol.embl.de/shared/mhedin). In our presentation below we first discuss the 28S data, for which we have the largest taxon sample. We then present results of the combined analysis; clades recovered in the 28S and combined analyses form a logical framework for presentation of individual gene region results with smaller taxon sample sizes.

3.1. 28S analyses

We generated 94 new 28S sequences, and supplemented these with six sequences reported by Giribet et al. (2010). In preliminary analyses, we found problems with the 28S sequence reported by Giribet et al. (2010) for Eurybunus (DNA voucher #100532, GQ912759). These sequences are very divergent from the Eurybunus brunneus 28S sequences that we generated, differ from all other phalangiid sequences, and appear more related to Laniatores 28S sequences. We suspect that this sequence is a contaminant, and removed this from all subsequent analyses.

The 28S sequences range in length from 1153 to 1199 basepairs (bp), with a final manual alignment length of 1224 bp. Alignment length from MAFFT analyses (L = 1224) is identical, although MAFFT and manual alignments differ slightly in length-variable regions. Three sequences were particularly long (Gyas, Nelima doriae, Nelima semprionii), accounting for most of the difference between sequence length and alignment length. A 28S MAFFT Bayesian consensus phylogram includes many well-supported clades (with posterior probability values exceeding 0.95), with less support for interrelationships among these clades (Fig. 2). Using Hesperonomastoma (Dyspnio) as a distant outgroup, the 28S topology includes a relatively early-diverging Neopilinidae, Metopilidae group, and Phalangiidae (including the Gyinae species Gys Titamus). Interrelationships amongst these three lineages are unresolved. Members of the genus Protolophus (Protolophidae) are sister to a clade of all other sclerosomatids (i.e., all but Gyas). European sclerosomatids are weakly supported as a grade related to a clade of “New World Gagrellines”. Species of US Nelima, plus Leiobunum exilipes from the western US, together form a clade. East Asian taxa, including a mix of both gagrellines and leiobunines most of which are from Japan (all but Marthana from Thailand), fall into three well-supported clades, two of which partly correspond to previously defined morphology-based species groups (Suzuki, 1976). East Asian taxa together do not form a clade (but see below). A well-supported “North American” clade includes Togwoteeus, two separate clades of mostly Mexican taxa, and a weakly supported group of “US leiobunines”, mostly including species from eastern North America, but also the Mexican Leiothrum rolandii. The 28S MAFFT tree implies obvious and strongly supported polyphyly of the Gagrellinae, Leiobiuninae, Nelima, and Leiothrum.

Phylogenetic analyses of the 28S manual alignment recovery taxonomic relationships and PP support values very similar to those outlined above (Fig 2; http://itol.embl.de/shared/mhedin), with minor differences between closely related sequences within the above-mentioned clades. New World gagrellines are weakly supported (PP = 0.77) as a clade exclusive of European taxa in 28S manual analyses.

3.2. Combined analyses

Taxa were included in combined analyses when sequences were available for three of six gene regions, except for Caddo and Thysanotus (only 2 gene regions available). In total, the combined matrix included data for 65 terminal taxa. With four exceptions (see Supplementary Table I), we concatenated sequences generated from the same specimen. The combined data Bayesian consensus phylogram (using the 28S MAFFT alignment) includes many well-supported clades, but differs from 28S topologies in also including mostly strong support for interrelationships among these clades (Fig. 3A). Using Hesperonomastoma and Caddo as outgroups, the combined data topology includes a relatively early-diverging Neopilinidae, Metopilidae group, and Phalangiidae (including Gyas). A fourth major well-supported lineage includes a monophletic Protopilidae, sister to all remaining sampled taxa (=Sclerosomatidae, excluding Gyas). Within the sclerosomatids a well-supported clade includes European sclerosomatids sister to the “New World Gagrellines”. East Asian (=Japanese) taxa are again partitioned into three clades, and together form a well-supported clade sister to the “US Nelima” clade. Remaining taxa fall into the well-supported “North
Fig. 2. 28S MAFFT alignment Bayesian phylogram. Tree rooted using *Hesperonemastoma* (Dyspnoi), not shown. Branch subtending Protolophidae plus Sclerosomatidae clade not drawn to scale = one-half actual length. Thickened branches = PP values above 0.95. Posterior probability values above branches from MAFFT alignment analyses, below branches from manual alignment analyses.
Fig. 3. (A) Combined evidence Bayesian phylogram. Tree rooted using *Hesperonemastoma* (Dyspncoi) and Caddo, not shown. (B) Strict consensus parsimony tree, with parsimony bootstrap values.
American” clade, with Togwoteus sister to a clade including two clades of Mexican taxa, and a well-supported “US Leio bunines” clade. Again, the combined data tree implies obvious and strongly supported polyphyly of the Gagrellinae, Leio buninae, Nelima, and Leio bunum (Fig. 3A).

A heuristic parsimony search of the combined data resulted in two most-parsimonious trees (l = 1065); the strict consensus of these trees, with parsimony bootstrap values, is shown in Fig. 3B. Basically all of the primary groups recovered in Bayesian analyses are also well supported (nonparametric bootstrap values >70; Hillis and Bull, 1993) in a parsimony framework, with a few important exceptions. The Protopholids are placed in a more nested position by parsimony, allied with East Asian (=Japanese) taxa plus the “US Nelima” clade. East Asian taxa together do not form a clade. A “North American” clade is recovered in the strict consensus parsimony tree, but is not well supported in bootstrap analyses.

3.3. Other individual gene trees

We generated 18S sequences, each ~1700 bp in length, for 47 terminal taxa. To these data we added 16 18S sequences from Giribet et al. (2010). The 18S data are trivial to align, with only a single indel site in the alignment. In preliminary analyses, we found problems with the Eurybunus (DNA voucher #100532, GG912709) 18S sequence reported by Giribet et al. (2010) – the first part of this sequence seems authentic, but the latter two-thirds is very divergent, with many unique singleton mutations. We removed this suspected mosaic sequence from all subsequent analyses. Bayesian analysis of the 18S data strongly supports recovery of the following clades, consistent with those recovered in 28S and combined data analyses (see Fig. 4, http://itol.embl.de/shared/mhedin): Neopilionidae, Metopilio group, Phalangiidae (including Gyas), New World Gagrellines, East Asian Clade, East Asian Clade sister to US Nelima, and a “North American” clade (PP = 0.88). Although there are some 28S and combined evidence clades that are not separated by the more slowly-evolving 18S data (e.g., Mexican Clades I and II, two clades within the East Asian Clade), we emphasize that there are no well-supported 18S groups that directly conflict with the 28S or combined evidence results.

We generated 37 EF-1α sequences (~740 in length), and supplemented these with six GenBank sequences. Both exon and intron regions were amplified and sequenced, but only exon data were considered in phylogenetic analysis. Bayesian analysis of the EF-1α data recovers the following clades, with smaller samples sizes, corresponding to those seen in 28S and combined analyses (see Fig. 4, http://itol.embl.de/shared/mhedin): US Nelima, New World Gagrellines, East Asian (=Japanese) Gagrellines, Mexican Clades I and II, and the “US Leio bunines” clades. A “North American” clade is strongly supported, but does not include Togwoteus, whose placement is more-early diverging (but uncertain) on the EF-1α tree. We sampled only a single member of the Metopilio group, and lacked EF-1α sequences for Protolophus, so could not assess the monophyly of these groups with this gene region.

Histone 3 sequences were generated for 22 taxa (288 bp in length), and supplemented with six GenBank sequences. Although this sparse taxon sample precludes an assessment of congruence with most 28S and combined analyses clades, the H3 data recover relationships consistent with the following groups (see Fig. 4, http://itol.embl.de/shared/mhedin): US Nelima, New World Gagrellines, Mexican Clades I and II, and the “US Leio bunines” clades. The Mar thana specimen (DNA voucher #100613) reported by Giribet et al. (2010), which allies with the East Asian Gagrelline clade in both 28S and 18S analyses, has a reported H3 sequence that is very different from two Japanese gagrelline sequences that we generated; this sequence was retained in our analyses, but should be confirmed. We generated mitochondrial cytochrome oxidase sequence for 26 taxa, amplified and sequenced as a contiguous region that spans the 3’ end of COI (~610 bp) and the 5′ end of COI (~270 bp). Homologous nucleotides from the complete mitochondrial genome of Phalangium opilo (Masta, 2010) were also included in this matrix. Again, a relatively sparse taxon sample does not allow strong congruence assessment, but the following recovered relationships are consistent with 28S and combined analysis clades (see Fig. 4, http://itol.embl.de/shared/mhedin): Metopilio group, Phalangidae (including Gyas), Japanese Leio bunum curvipalpe group, Japanese Leio bunum rubrum group, Japan (PP = 0.74), Mexican C lade I, and a “US Leio bunines” clade.

Finally, we generated 58 mitochondrial 16S_12S sequences, amplified and sequenced as a contiguous region that spans most of 16S (~1230 bp), an intervening tRNA-Val, and the 3′ end of 12S (~430 bp). Some of these sequences are partial (see Supplementary Table I). Homologous nucleotides from the complete mitochondrial genome of P. opilo (Masta, 2010) were included, allowing us to confidently identify gene boundaries. Bayesian analysis of the 16S_12S data recovers the following 28S and combined analyses clades (see Fig. 4, http://itol.embl.de/shared/mhedin): Metopilio group, Protolophidae, New World Gagrellines, US Nelima, Japanese L. curvipalpe group, Japanese L. rubrum group, Mexican Clades I and II, and the “US Leio bunines” clade.

4. Discussion

The molecular phylogenetic results presented here support the monophyly of only two currently recognized phalangioid taxa, namely the Protolophidae and the Metopilio group. The more striking pattern is a lack of support for a majority of subfamily- and generic-level classification groups. Instead, biogeographic origin appears to play a more important predictive role in defining molecular clades, and when a priori taxa are recovered, these groups also correspond to regional classification groups (e.g., Protolophidae of the southwestern United States, Metopilio group of western North America). Conversely, when a priori classification groups are found to be paraphyletic or polyphyletic, these mostly correspond to subfamilies and genera that are hypothesized to span multiple continents (e.g., Gagrellinae, Leio buninae, Leio bunum, Nelima). In general, our results expose a perhaps unexpected pattern of regionally dominated phylogenetic structuring in this group of long-legged harvestmen. Furthermore, our results suggest that morphology has evolved conservatively and/or convergently, leading to difficulties in taxonomic classification. In the most recent and comprehensive classification of the group, Crawford (1992) commented that “In the Phalangioidea … the existing generic classification is almost wholly artificial.” Our data provide support for this contention.

In drawing these two general conclusions regarding biogeographic endemic and morphological conservatism/homoplasy, we need to be wary of the possibility of errors in molecular data or molecular phylogenetic analyses. Possible sources of error include genomic or sequence contamination, misalignment of length-variable data, saturation of the mitochondrial data, and gene paralogy in the nuclear protein-coding data. Although we cannot definitively rule out such sources of error, we argue that five general observations lend support to our molecular results. First, in this paper we have emphasized only those groups recovered in a large percentage (>95%) of taxon bifurcations, i.e., well-supported clades. Second, these highlighted molecular clades are either supported by multiple genes (all of which are providing some phylogenetic signal), or if not supported by all genes, then single gene analyses do not strongly conflict with such clades. Third, in combined data analysis, alternative reconstruction methods (parsimony and Bayesian approaches) recover the same
higher-level clades and most clade interrelationships, except for the placement of Protolophidae. Fourth, some recovered molecular clades also correspond to regional morphological groupings, to greater or lesser degrees. This implies that our emphasized molecular clades sometimes also have morphological support. Finally and perhaps most importantly, basically all highlighted molecular clades “make sense” from a biogeographic standpoint – molecular clades are clearly not biogeographically random.

4.1. Biogeographic regionalism and historical biogeography

In the past 5–10 years, molecular studies of non-phalangioid harvestmen taxa have clearly illustrated the biogeographic information content of these lineages. Studies of various lineages in the Cyphophthlami, Laniatores, Dyspnoi, and Caddoidea (see Fig. 1A) have revealed biogeographic endemism at multiple spatial scales and phylogenetic levels (e.g., Boyer et al., 2007a, 2007b; Shultz and Regier, 2009; Derkarabetian et al., 2010; Hedin and Thomas, 2010; Schönhofer and Martens, 2010). Because many non-phalangioid harvestmen are small-bodied, short-legged, and have very specific habitat preferences, both limited dispersal abilities and biogeographic endemism are unsurprising findings. Prior to our current research, phalangioids had been largely untouched from a molecular phylogenetic perspective. Because phalangioids are generally active, wandering, long-legged animals, one might predict less biogeographic signal than seen in non-phalangioids. Also, the current classification, with many transcontinental taxa, obviously suggests less biogeographic regionalism.

Our results instead indicate that phalangioid clades do indeed show biogeographic regionalism, although not necessarily as “fine-scaled” as seen in non-phalangioids. The majority of phalangioid clades appear to be continental endemics, but some of these clades are very broad-ranging within continents. Here we summarize these patterns of regionalism. Three separate “East Asian” clades are recovered, although with a single exception (Marthana sp. from Thailand) all of the taxa included in this study were sampled from Japan. The genus Marthana itself (presuming mono-phyly) includes more than 20 species distributed in insular and
mainland Southeast Asia. 28S and 18S sequences of *Marthana* are very similar to sequences from Japanese gagrellines (i.e., low divergence over a large area), suggesting a perhaps wide-ranging clade of Southeast Asian gagrelline taxa. Increased sampling of the Oriental fauna is obviously needed to test this contention.

A second Asian clade, including sampled taxa *Leiobunum japonense*, *L. ribrum*, and Japanese *Nelima*, includes species that generally inhabit low- to mid-elevation forests on the Japanese Islands. Although we have not sampled outside of Japan, we predict that this clade has a slightly broader distribution in the region. For example, *L. japonense* (Japanese Islands) and *Leiobunum maximum* (Ryukyu Islands, Taiwan, and southeastern China) are apparently related species having a distributional range that corresponds to the so-called “Sino-Japanese floral region” in phytogeography (*Suzuki, 1976*). In addition to the Japanese Islands, East Asian *Nelima* are known from Korea, Sakhalin Island, the Ryukyu Islands, and Taiwan. Finally, the third “East Asian” clade corresponds to the *L. curvivalpe* group (including *Gagrelopsis nodulifera*). This clade is endemic to the Japanese Islands (a single exception is *L. oharai* of the *curvivalpe* group from Taiwan: *Tsurusaki, 1991*). The distribution of harvestmen in this clade coincides well with the distribution of high-elevation Japanese beech (*Fagus crenata*) forests from western Japan. Although several *Fagus* species are also found in southeastern China, there are no literature records of related sclerosomatid taxa from this region. Also, one of the authors (NT) has examined unidentified sclerosomatids deposited in the Zoological Museum of Chinese Academy of Sciences, Beijing; no species related to the *curvivalpe* group were discovered.

We recovered several regional clades endemic to either North America or the New World. Some of these groups were predicted by current classification. Protolophids are restricted to the western United States, with a center of distribution in California (*Goodnight and Goodnight, 1942*). The apparently monophyletic *Metopillo* group includes five described genera (we sampled all but *Diguetinus*), with sampled taxa that range from California to Costa Rica. Mexico appears to represent the center of distribution for this group. Other novel North American endemic groups emerge from molecular phylogenetic analysis. Two such clades are centered in the highlands of Mexico (Mexican Clades I and II), with Mexican Clade I dominating the central and western portion of the country and extending into the southwestern United States. Sampled members of Mexican Clade II are found in eastern Mexico at lower elevations. The Sierra Madre Oriental may serve to separate these two large clades. A “US *Nelima*” clade includes *Nelima* species from temperate habitats in the northern United States and southern Canada, plus *L. exilipes* from the western US. This clade is a good example of a continent-spanning, but ultimately continent-endemic clade. The “US Leiobunines” *clade* is similar in this respect, with sampled taxa found principally in the eastern and central US. Although there is some distributional overlap between the “US *Nelima*” clade, Mexican, and “US Leiobunines” clades, the latter clade does appear to dominate at intermediate latitudes. Finally, a more broad-ranging “New World gagrelline” clade spans two continents, including taxa in our sample ranging from the central and southwestern United States (*Trachyrinus*), through Mexico and Central America (*Geaça, Prionoostema*) to Brazil (*Holocubanus*). Other gagrellines that are possibly part of this clade occur even further south in South America. This clade appears to include mostly tropical species, with limited distributional extent in northern and southern temperate habitats.

Our results show a clear relationship between phylogeny and geography among well-supported terminal clades, but there are fewer such associations deeper in the tree. The clade comprising European sclerosomatids and a subset of Neotropical gagrellines is an exception. These findings may simply reflect poor phylogenetic signal at deeper branches, and indeed, many ‘trunk’ nodes are weakly supported. Furthermore, omission of representative species from tropical and central Asia and scant sampling from the Mediterranean Region and the Neotropics represent substantial gaps in both geographic and phylogenetic data. However, despite the potential weaknesses of our results, the existence of unexpected endemism or phylogenetic affiliations is not unique to sclerosomatids, and it is possible that our current results are consistent with known patterns in historical biogeography.

The Boreotropic Concept (*Wolfe, 1975; Tiffney, 1985; Lavin and Luckow, 1993*) is potentially relevant to sclerosomatid biogeography. It is derived from evidence that global temperatures were unusually high during the late Paleocene and early Eocene, that the northern subpolar regions had tropical to subtropical climates during this time, and that there was substantial biotic interchange between continents via Beringian and North Atlantic terrestrial corridors. As the northern climate cooled during the Tertiary, the tropical elements within each continent retreated southward, losing their connections to each other due to latitudinal climatic and longitudinal oceanic barriers. An interchange of temperate groups persisted to varying degrees until it was terminated by Pleistocene glaciation. This scenario appears to explain certain similarities in the biotas of the New World and Asian tropics as well as between temperate North American and East Asia. The Boreotropic Concept has been developed and applied most thoroughly by plant systematists (*Xiang et al., 2000; Azuma et al., 2001; Smedmark and Anderberg, 2007; Mercx et al., 2008; Thomas et al., 2011*), although there are zoological examples as well (*Janis, 1993; Sammartin et al., 2001; Wallace and Wang, 2004; Min et al., 2005*), including non-sclerosomatid harvestmen (*Suzuki et al., 1977; Shultz and Regier, 2009*).

The existence of circumboreal tropics in the early Tertiary may have allowed movements of ancestral gagrelline-like sclerosomatids between Eurasia and North America. These taxa would then have followed the tropics southward, producing the modern disjunction between the largely tropical New World and Asian “Gagrelline” of today. If this separation had been superimposed on pre-existing phylogenetic diversity, phylogenetic analysis might thus reveal geographically distant sclerosomatids as close relatives (e.g., the European plus New World *Gagrellina* clade). Coupled with our phylogenetic results, this scenario also suggests that tropical forms then radiated multiple times in the newly emergent temperate regions, which would explain the existence of mixed regional clades of “gagrellines” (tropical ancestral forms) and “leiobunines” (subtropical/temperate descendant forms) in both North American and Asia. This scenario may be testable through phylogenetic analysis of a more thorough sample of Asian and New World gagrellines and the use of molecular dating to search for correlations in the timing of geoclimatic and phylogenetic events.

Our phylogenetic results provide more clarity to historical biogeographic patterns in phalangioid harvestmen, but several major biogeographic puzzles remain. First, do European sclerosomatids (excluding *Gyas*) represent a monophyletic regional fauna, despite the heterogeneous taxonomic nature of this group (e.g., including leiobunines, sclerosomatines, the enigmatic *Diceranopalus* group, etc.)? Our small sample hints at this possibility (e.g., Fig. 3B), but a much more comprehensive taxon sample is needed to test this hypothesis. Second, what are the biogeographic origins and affinities of the diverse central Asia fauna (e.g., Himalayas; *Martens, 1973, 1982, 1987*)? Finally, larger samples are needed to test the possibility of biogeographic regionalism in the phalangiids. Viewed as a single family, this clade includes several transcontinental taxa, e.g., *Mitopus morio* (*Martens, 1978*), the essentially cosmopolitan species *Phalangium opinol* (*Grabner and Hlunt, 1973*), etc. Are phalangiids atypical in lacking biogeographic regionalism altogether, or will smaller clades within the family exhibit more biogeographic endemism?
4.2. Morphological conservatism and convergence

Taken literally, our results indicate that recent classifications of sclerosomatids (Pinto-da-Rocha et al., 2007) are flawed, with most subfamilies and large genera emerging as poly- or paraplethypyls (Figs. 2–4). The failure of Roewer's typological classification will not surprise harvestman systematists, who appear to have retained problematic systems due primarily to the absence of alternatives. However, the number and magnitude of taxonomic inconsistencies between recent systems and the system implied by our phylogenetic results may be met with some skepticism, if not alarm, by some workers. Thus, it is important to explore why the morphological characters upon which previous systems were based do not necessarily predict phylogenetic relationships.

It bears repeating that Roewer's (1923) system of sclerosomatid taxonomy was typological; he focused on erecting formal classifications based on nested sets of idealized diagnostic characters. Yet, it is hard to imagine how else a worker of the early 20th century would deal with a taxonomic problem of this magnitude. Even now, criticisms of Roewer's system often focus on the imperfection of his diagnostic characters, not his basic approach. Indeed, each alternative proposal has been a more-refined typological system; there have been virtually no phylogenetic analyses or even argumentation (e.g., polarizing characters through outgroup comparison). This, too, reflects the persistent magnitude of the problem and our still-early position in the history of sclerosomatid systematics (Cokendolpher et al., 2007). It is important to note that the conflict between traditional classifications and the one indicated by our analysis is not a conflict between morphology and molecules per se but between typological and phylogenetic approaches. Phylogenetic methods using morphology might have produced results consistent with those derived from molecules had they been applied.

There is a remarkable vacuum of basic comparative biology that has certainly contributed to stagnation in sclerosomatid taxonomy. The systematic value of a character, whether in a typological or phylogenetic context, is closely related to the character's propensity for homoplasy. Heterochronic parallelism, for example, is a major potential source of morphological homoplasy, yet there is no way at present to evaluate its impact on sclerosomatid taxonomy. Very little is known about the postembryonic development of sclerosomatids (Gnaspinii, 2007). In general, specimens of early instars cannot be identified to species, or even genus in many cases, and are routinely discarded by systematists. Yet, many traits used in the classification of adults change during development, including color pattern, relative leg length, development of patellar apophyses on the palp, palpal setation, relative size of eyes, cuticular armature and so forth. For example, the presence of pre- and/or retro-lateral rows of coxal denticles appears to be a plesiomorphic feature among sclerosomatids. The widespread leibounine genus Nelima is diagnosed primarily by the absence of these denticles, a feature shared with early Leibounum instars (original obs.). Thus, the apparent polyphyly of Nelima indicated by our analysis may simply reflect multiple independent cases of paedomorphosis on different continents. If heterochrony is common in sclerosomatids, we should expect degradation of the phylogenetic signal offered by the affected character.

Adaptive convergence is another potential source of homoplasy. Yet, studies on the functional role of traditional diagnostic characters in sclerosomatids are virtually non-existent, and the potential for adaptive homoplasy can only be guessed at. The relative length of legs with respect to the body is often used as a diagnostic character, even though it is probably under significant selection and thus has substantial potential for homoplasy. Ground-dwelling harvestmen generally have shorter legs than scansorial species as well as heavier dorsal sclerotization, enlarged scuta and enhanced armature (i.e., spines, horns or processes) of the scutal, ocular and preoculus regions. Using these similarities to suggest close phylogenetic relationships could be misleading; for example, we found no evidence of a close relationship between the neotropical gagreline Carmania and Asian gagrellines suggested by Cokendolpher (1984b) based on such characters (Fig. 2 and 3). Similar arguments may also apply to the Gagrellinae more generally, which are historically defined by the presence of femoral nodules (pseudocariculations) that are sometimes expressed only on the elyrae and non-sacculate males (Fig. 5). These structures show a tendency toward reduction and loss in temperate forms, suggesting some mechanical or sensory function that varies with the environment. If so, similarity in the presence, absence or arrangement of femoral nodules may reflect functional similarities more than taxonomic relatedness. In general, the tendency to rely on single diagnostic characters when circumscribing taxa may be problematic without information on the biological role of these characters.

Reproductive structures are widely thought to be insulated from ecological effects and more immune to homoplasy than somatic structures, because features used in mating reflect sexual selection (i.e. sperm competition and female choice) rather than natural selection by the environment. Indeed, the systematic utility of male reproductive features spans the full taxonomic hierarchy of the Opiliones (Pinto-da-Rocha et al., 2007). However, it is not at all clear that the sclerosomatid penis is divorced from the demands of the environment (Macías-Ordóñez et al., 2010). The lateral wings or alae of the ground-plan penis are cuticular sacs that contain a secretion, even in preserved specimens ("opaque body" of Tourinho-Davis and Kury, 2003), that is delivered to the female during mating (Shultz, 2005; Macías-Ordóñez et al., 2010). Thus, the male brings a material investment to mating that presumably represents a cost in energy and foraging time, and its cost to the male (and value to the female) may increase under conditions of resource limitation. One such limitation may be the duration of the mating season, which can be long in the tropics but short in the temperate regions. In the sclerosomatids of temperate North America, the penial sacs have been reduced or lost at least five times, with associated development of enhanced clasp- ing mechanisms in males and defensive elaborations of the pregenital openings in females (Burns, Hedin and Shultz, unpublished obs.). A comparable phenomenon appears to have occurred in Japan in an entirely different lineage of sclerosomatids, the L. curvipalpe group (Tsursusaki, 1985), with southern species retaining sacculate penes and simple palps and northern species having non-sacculate males with strong claspung palps (Fig. 5). In short, even convergences among reproductive structures may occur and thereby confound morphology-based systems of classification, especially those that rely on key diagnostic characters.

These examples are not intended as an indictment against morphology's role in phylogenetic analysis nor as a claim that our molecule-based phylogenetic results are entirely correct. Rather, they show how a long-standing system of classification based on anachronistic taxonomic principles and very limited data can be suddenly overturned. We suggest that the long history of certain traditional taxonomic groups (Sclerosomatidae, Gagrellinae, Leibounine, Leibounum, Nelima) cannot be attributed to a compelling list of supporting characters but to the absence of alternatives. The results of our analysis are the first well-supported, compelling list of supporting characters but to the absence of alternatives.
arachnids, with approximately 1800 described species classified in over 200 genera. The contained family Sclerosomatidae is considered to be the single largest family in the order Opiliones. In this study we have assembled a relatively comprehensive sample of North American and Japanese genera, but our sampling in other areas has been basically opportunistic. Large biogeographic regions (e.g., Central and South America, Himalayas, southeast Asia) with very diverse faunas remain essentially unstudied. Moreover, very few systematists are actively working on phalangioids.

Working on this group of harvestmen is difficult because a “divide and conquer” strategy based on current taxonomy is likely to fail. Although obviously not universally true (e.g., the homogeneous genus Protolophus, etc.), for some faunas it is clearly impossible to predict phylogenetic placement based on current generic or subfamilial classification (e.g., the megadiverse tropical gagrelline faunas, mostly revised by Roewer). This means that we cannot take an exemplar phylogenetic approach, using a single taxon to represent a larger monophyletic group (e.g., using a species to represent a genus). Instead, we argue that a “sequence everything” model might be most appropriate at this time for the phalangioids, whereby we sample as many species as possible for relatively few genes. The 285 gene region might be a good place to start, as this gene amplifies easily, provides ample phylogenetic information, and is not problematic from an alignment perspective. In essence, we are arguing for something akin to a DNA barcoding approach in a largely unknown fauna. With a robust sample (e.g., from ~ 500 species) we predict that larger natural groups would become evident. These larger groups could then become targets for exemplar sampling and multigene phylogenetics, and the careful post hoc search for morphological synapomorphies defining such groups. This type of iterative systematics might ultimately allow researchers to build a predictive classification system for this hyperdiverse arachnid clade.

Acknowledgments

Funding was provided by a collaborative NSF grant awarded to J. Shultz (Award No. 0640179) and M. Hedin (Award No. 0640173). J. Shultz was also supported by the Maryland Agricultural Experimentation Station. N. Tsurusaki was supported by Grants-in-aid for Science Research from the Japan Society for the Promotion of Science (19570086 and 22570092). M. McCormack, J. Underwood and J. Deas helped in the collection of molecular data. Many persons helped to collect or donated specimens, including: A. Bailey, A. Bautista, P. Berea, J. Bond, J. Camacho, J. Carrel, T. Cekalovic, J.C. Cokendolpher, L.M. Cuenca, S. Derkarabetian, N. Duperre, A. Fusek, R. Keith, S. Masta, P. Miller, R.A. Moore, G. Morales, P. Nunez, D. Palmer, P. Paquin, Y. Perroni, R. Pinto da Rocha, C. Richart, J. Starrett, G. Stratton, B. Suter, K. Takenaka, S. Thomas, B. Tomberlin, H. Wilson, H. Wijnhoven, and D. Wytrykush. G. Giribet provided 18S and 28S data alignments. Comments of G. Giribet, A. Schönhofer, and an anonymous reviewer helped to improve the manuscript.

Appendix A. Supplementary material


References


