Spiders spin a diverse array of silk fibers that are predominately composed of repetitive proteins (spidroins) encoded by a gene family. Characterization of this gene family has focused on spidroins synthesized by the Araneomorphae (true spiders), whereas only a single sequence is known from the Mygalomorphae (tarantulas and their kin). To better understand the diversity and evolution of the spidroin gene family, we surveyed the silk gland transcriptomes of 4 divergent mygalomorph species. Through expressed sequence tag screening and probing of silk gland cDNA libraries, we discovered 6 novel mygalomorph spidroins and an ~8-kb cDNA of the previously reported *Euagrus chisoseus* fibroin 1. Mygalomorph spidroin cDNAs encode tandem iterations of sequence repeats, followed by a nonrepetitive carboxy-terminal domain. Though highly homogenized at the nucleotide level within a cDNA (89–100% identical), these repeats exhibit extensive variation across spidroins, consistent with intragenic repeats evolving in concert. Extreme homogeneity of intragenic repeats is also characteristic of araneomorph spidroins, suggesting that modular architecture and its maintenance through concerted evolution have persisted since the mygalomorph/araneomorph split (~240 MYA).

Phylogenetic analyses of C-terminal sequences grouped all mygalomorph spidroins, except *Alatypus* fibroin 1, in a clade. *Alatypus* fibroin 1 was instead more closely related to a subset of araneomorph spidroins, including those used in prey wrapping. Our results suggest that spidroin paralogs existed prior to the divergence of mygalomorphs and araneomorphs, followed by a far greater expansion of this gene family in araneomorphs, paralleling the dramatic functional diversification of their silk gland anatomy.

**Introduction**

With more than 39,000 described species and a 400-Myr-old fossil record, spiders (order Araneae) constitute an ancient and extremely prolific animal radiation (Shear et al. 1989; Coddington et al. 2004; Platnick 2007). The evolutionary success of spiders is often attributed to their utilization of high-performance silks for a variety of ecological functions, along with a spectacular diversification of silk-spinning organs, fibers, and constructs (Schulz 1987; Craig 2003). Spiders spin silks from proteins secreted in specialized abdominal glands that vary in number and morphology across species. For example, an individual araneoid spider (ecribellate orb weaver) can possess up to 7 morphologically distinct types of silk glands, each of which synthesizes a different set of proteins that assemble into a task-specific silk (e.g., web-building silks, egg case silk, prey-wrapping silk; Guerette et al. 1996). Nearly all spider silk cDNAs characterized to date have the common structural organization of long (~4 to 16 kb), highly repetitive sequence flanked by short, nonrepetitive amino and carboxy-termini encoding regions (Gatesy et al. 2001; Motriuk-Smith et al. 2005). These transcripts generate proteins, termed spider fibroins or spidroins, that not only share sequence similarity in their nonrepetitive termini but also exhibit extreme divergence in their repetitive regions. This pattern is consistent with spidroins being encoded by members of a gene family that have undergone functional specialization (Beckwitt and Arcidiacono 1994; Hayashi et al. 2004). Characterization of this gene family has focused on spidroins from the Araneoidea (~10,000 species; e.g., orb-web and cob-web weavers), a lineage within the infraorder Araneomorphae (true spiders: ~37,000 species), whereas little is known of spidroins from the araneomorph sister group, the Mygalomorphae (tarantulas and their kin: ~2,500 species).

The infraorder Mygalomorphae is composed of large, relatively sedentary spiders that primarily dwell in silk-lined burrows (Hedin and Bond 2006). Mygalomorphs share many morphological and ecological similarities with the Mesothelae, the spider lineage that is sister to Opisthothelae (Mygalomorphae + Araneomorphae; Platnick and Gertsch 1976; Coddington et al. 2004). Mesotheles and mygalomorphs possess a number of primitive features, including those associated with silk production (Haupt and Kovoor 1993). In contrast to araneomorphs, mygalomorphs have a comparatively undifferentiated spinning apparatus consisting of uniform spigots that lead to 1–3 types of globular silk glands (Palmer et al. 1982; Palmer 1985). Yet mygalomorphs utilize silk for a variety of ecological functions, including burrow construction, egg protection, and prey capture (Coyle 1986; Schulz 1987). Thus, unlike araneomorph spiders that have evolved a multitude of task-specific fiber types, mygalomorphs appear to draw upon a small number of generalized silks for different purposes.

To date, a single spidroin cDNA has been identified from one mygalomorph species, *Euagrus chisoseus*, based on sequence similarity of its carboxy (C)-terminus to araneomorph spidroin C-termini (Gatesy et al. 2001). This finding established the widespread distribution and antiquity of the spidroin gene family, which must predate the divergence of mygalomorphs and araneomorphs at least 240 MYA (Selden and Gall 1992). The sequence of this spidroin (*E. chisoseus* fibroin 1) suggests a repetitive organization, also similar to the tandem arrays that typify araneomorph spidroins. Within a sequence, araneomorph spidroins contain iterations of highly homogenized repeats (e.g., 98–100% identical at the nucleotide level), a pattern generally explained as a consequence of concerted evolution resulting from nonreciprocal recombination (gene conversion) or unequal crossing over among intragenic repeats (Beckwitt et al. 1998; Hayashi et al. 2004).
Repeat length (aa) dominated by short amino acid motifs (e.g., A (2–3 different residues accounting for spidroin ensemble repeats are low-complexity sequences within a spidroin sequence, repeat modules are extremely similar. Across spidroins, ensemble repeats widely vary in primary sequence: each contains an amino acids it contains (# diff. aa). sequence length and complexity (i.e., diversity of amino acids). Each ensemble is followed by its length in amino acids and the number of different amino acids it contains (# diff. aa).

In contrast to the striking uniformity of repeats within a spidroin, repeats from paralogs often radically differ in their sequence characteristics (Gatesy et al. 2001; Hayashi et al. 2004). An individual spidroin can be represented by its consensus ensemble repeat unit (a consensus of its translated intragenic repeats), that across paralogs can range in length from −28 to 357 amino acids (Garb et al. 2006; fig. 1). Some spidroin ensemble repeats are low-complexity sequences (2–3 different residues accounting for >80% of each repeat), dominated by short amino acid motifs (e.g., An, GAin, GGX, and/or GPXn, where X = a subset of amino acids). These motifs are poorly represented in other ensemble repeats, which instead contain longer sequences (180–357 amino acids) with less compositional bias (Hayashi et al. 2004). The mygalomorph spidroin E. chisoseus fibroin 1 has an ~345-amino acid ensemble unit that most closely resembles this latter class of repeats in its overall length and sequence complexity. At this stage, it is unknown whether the characteristics of this sequence are unique to Euagrus or representative of other mygalomorph spidroins.

To better understand the diversity and evolution of spidroins, we surveyed the silk gland transcriptomes of 4 mygalomorph species distributed in divergent families. As a result, we characterized 6 novel mygalomorph spidroins from 3 black widow spider (Latrodectus hesperus) spidroins (dragline silk proteins, MaSp1 and MaSp2; egg case silk protein, TuSp1) compared with the mygalomorph spidroin, Euagrus chisoseus fibroin 1. Within a spidroin sequence, repeat modules are extremely similar. Across spidroins, ensemble repeats widely vary in sequence length and complexity (i.e., diversity of amino acids). Each ensemble is followed by its length in amino acids and the number of different amino acids it contains (# diff. aa). An individual spidroin can be represented by its consensus ensemble repeat unit (a consensus of its translated intragenic repeats), that across paralogs can range in length from −28 to 357 amino acids (Garb et al. 2006; fig. 1).

Materials and Methods

Taxon Sampling

We obtained multiple individuals of Bothriocurtum californicum (Ctenizidae) from Mission Trails, San Diego, CA (San Diego Co.); Aliatypus plutonis (Antrodiaetidae) and Aptostichus sp. (Cycruchteniidae) from the Box Spring Mountains, Riverside, CA (Riverside Co.); and E. chisoseus (Dipluridae) from the Santa Rita Mountains, upper Madera Creek, AZ (Santa Cruz Co.). Current phylogenetic hypotheses for the Mygalomorphae indicate that these families broadly span the infraorder (Raven 1985; Goloboff 1993; Hedin and Bond 2006; Ayoub et al. 2007). All studies agree that Antrodiaetidae belongs to the Atypoida (or Atypoidina sensu Raven 1985), the sister lineage to the Orthopalpae, a group that includes most other mygalomorph families. Within the Orthopalpae, morphological characters place Ctenizidae and Cyrtuachteniidae in the Rastelloidina (Raven 1985; Goloboff 1993). Moreover, both morphological (Goloboff 1993) and molecular data (Hedin and Bond 2006; Ayoub et al. 2007) suggest that Dipluridae is in a basal position within the Orthopalpae.

cDNA library construction required us to combine multiple individuals of each species. However, the collected Aptostichus specimens could not be unambiguously diagnosed at the species level using morphological characters (Hedin M, personal communication). Because multiple Aptostichus species have been recorded from our collecting locality, we amplified and sequenced fragments of mitochondrial 16S rRNA from each individual. Two Aptostichus specimens having identical 16S sequences were combined for the library’s starting material. Vouchers were retained at −80 °C in the personal collection of the authors.

Molecular Methods

We dissected total silk gland tissue from live spiders anesthetized with CO2. Silk glands were immediately frozen in liquid nitrogen and stored at −80 °C. Total RNA was extracted from the tissue by homogenization in Trizol (Invitrogen, Carlsbad, CA), followed by purification using an RNeasy Mini Kit (Qiagen, Valencia, CA).
Oligo-(dT)$_{25}$-tagged magnetic beads (Dynal Biotech, Brown Deer, WI) were used to isolate mRNA from total RNA. From mRNA extractions, cDNA was synthesized by the Super-Script II Choice protocol (Invitrogen) using an anchored oligo-(dT)$_{12}$-V (V = A or C or G) primer. cDNA fragments were size-fractionated with a ChromaSpin 1,000 column (Clontech, Mountain View, CA) to enrich for transcripts ≥1,000 bp. Size-selected cDNAs were ligated into EcoRV (NEB, Ipswich, MA)-digested pZEoR-2 plasmids (Invitrogen) and transformed into TOP10 Escherichia coli cells (Invitrogen) by electroporation. For each species, ∼1,800 recombinant E. coli colonies were arrayed in long-term storage plates.

Each library was replicated onto nylon filters for screening with $^{32}$P-labeled oligonucleotide probes. All libraries were probed with GCGDGCDGCDGCDGCGDGC and CCWGCGCCWCWGCGCWGC (sequences shown 5'–3'), which have previously been employed to locate silk cDNAs (Gatesy et al. 2001; Garb and Hayashi 2005). To identify additional spidroin cDNAs not found through initial rounds of probing, ∼30% of each library was scored for insert size, and cDNAs ≥500 bp were sequenced with the T7 or Sp6 universal primers. Library-specific probes were designed from putative spidroin cDNAs (identified from the mygalomorph libraries through size screening) and used to screen the libraries. These new probes were 1) CGGATGGAAACTCCCTCACGC, 2) CGCCGATTTTACTAAGTGGAAAGC, 3) TGTGCTTGCCAAATGGAAGC, 4) ATGAGGCTGAAGCGACGAGATG, 5) ATGCSAGATGGTCGGTCTTC, 6) TTGCTGCTGGCTGGCTTC, 7) CAGAGATAGCGTCGGCGTTAGCGGTGTTCAATC, 8) TTGCTGCTTGTGTTGTTG, 9) CAAAGGCAGAAATGAAATC (1–9 for B. californicum), 10) GTTGGCAGAAGACGCTTC, 11) CAGATAGGGAATGTGCTGT, 12) AGAGGTGAGTATGGCCGTCTCT, 13) CTACTGAAAGATCGGTATGGCC (10–13 for A. trifasciata sp.); and 14) GAGCCTATTCCGGTAACATCGTCA, 15) CCGCGTATCCTCTTCCAC, 16) GGGCTTAGCGGTTGTTCAATC, 17) CTTGATGCTGCTCCCGATGCTGAT (14–17 for A. plitonis).

Restriction enzyme digests were utilized to determine the longest cDNA insert of each spidroin type. Because each clone contained highly repetitive sequence, it was not possible to “walk” down entire inserts with internal primers. Instead, clones were completely sequenced in both directions using the transposon-based GPS-1 Genome Priming System (NEB); a complete contig of each clone was then assembled.

Sequence Analyses

Expressed sequence tags (ESTs) were initially subjected to translated-Blast queries (BlastX; Altschul et al. 1997) against the National Center for Biotechnology Information protein database. All nucleotide sequences were subsequently grouped into nonredundant clusters using BLASTCLUST (http://www.ncbi.nlm.nih.gov/blast/docs/blastclust.html; similarity threshold = 60, minimum length coverage = 0.5). From the longest clone in each spidroin sequence cluster, we searched for tandem repeats using the EMBOSS programs, eqtandem and etandem (Rice et al. 2000). Etandem searches were conducted by screening for short tandem repeats (3–39 bp) and for longer repeats in the size range identified by eqtandem (extending search range by 50 bp upstream and downstream of estimated sizes). Subsequences retrieved from tandem searches were considered intragenic repeats if they were over 85% identical to their consensus (Verstrepen et al. 2005). Amino acid repeats were determined from conceptually translated cDNAs by visual searches and by using the RADAR program (http://www.ebi.ac.uk/Radar/; Heger and Holm 2000). Repeat ensembles were generated by aligning translated terminal repeats of sequences within a spidroin and determining the majority-rule consensus, with an “X” denoting equivocal positions. Predicted amino acid compositions were computed with MacVector 7.2 (Accelrys Inc., San Diego, CA).

To determine the phylogenetic relationships of mygalomorph spidroins relative to other gene family members, the sequences reported here were analyzed with published spidroin cDNAs. Phylogenetic analyses of spidroins have largely been conducted with sequences containing the non-repetitive C-terminal domain (~100 amino acids before stop codon), as repetitive sequences are problematic to align (Gatesy et al. 2001). The nonrepetitive N-terminal domain also provides conserved sequence for phylogenetic analyses (Motriuk-Smith et al. 2005; Rising et al. 2006). However, N-termini are much less common than C-termini among published cDNAs. Spidroin cDNAs appear to be more reliable than some sequences obtained via polymerase chain reaction because of occasional contamination (e.g., Tai et al. 2004; see supplementary fig. S1, Supplementary Material online). For these reasons, we selected cDNAs containing C-termini from all available major araneomorph lineages. These sequences included every functionally distinct type of spidroin and all unique repetitive organizations reported to date. Specifically, we sampled Argiope trifasciata (A.t.), A.t. AcSp1 (GenBank accession number: AY426339), A.t. Flag (AF350264); Nephila clavipes (N.c.), N.c. MiSp1 (AF027735); Latrodectus hesperus (L.h.), L.h. MaSp1 (AY953074), L.h. MaSp2 (AY953075), and L.h. TuSp1 (AY953070); Deinopis spinosa (D.s.), D.s. TuSp1 (AY953073), D.s. Flag (DQ399325), D.s. MaSp2a (DQ399328), D.s. MiSp1 (DQ399324), D.s. fibroin 1a (DQ399326), and D.s. fibroin 2 (DQ399323); Uloborus diversus (U.d.), U.d. MaSp1 (DQ399331) and U.d. AcSp1 (DQ399333); Agelenopsis aperta (A.a.), A.a. fibroin 1 (AY566305); Dolomedes tenebrosus (D.t.), D.t. fibroin 1 (AF350269); Plectreurys tristis (P.t.), P.t. fibroin 1–P.t. fibroin 4 (AF350281–AF350284); and E. chisoeus (E.c.), E.c. fibroin 1 (AF350271). The name of a spidroin gene or protein is designated by an abbreviation of the gland that it was initially isolated from (Ma = major ampullate, Mi = minor ampullate, Tu = tubuliform, Ac = aciniform, and Flag = flagelliform), preceding “Sp” for spidroin. Spidroin genes and proteins that cannot be assigned to these ortholog groups (e.g., E. chisoeus fibroin 1) are labeled “fibroin #, where numbers indicate different paralogs.

Amino acid sequences of C-terminal domains were first aligned with ClustalW (Higgins et al. 1994), implemented in MacVector, using default settings. This initial alignment was subsequently visually refined. For
phylogenetic analyses, alignment gaps were treated as either missing data or as additional presence/absence characters recoded using the “simple” method of Simmons and Ochoterena (2000) in SEQSTATE (Müller 2005). Maximum parsimony analyses were conducted with heuristic searches in PAUP* 4.0b10 (Swofford 2006), including 10,000 random-taxon-addition (RTA) replicates. Nodal support was assessed by 10,000 bootstrap (BS) replicates with 10 RTA per replicate and decay indices (DIs). Bayesian analyses of the C-termini alignment were performed with MrBayes 3.1.2 (Huelsenbeck and Ronquist 2001), employing the Jones, Taylor, and Thornton (JTT) substitution model with a gamma parameter for among-site rate variation. The JTT model had the highest posterior probability (PP). Bayesian tree searches were executed for 3 × 10⁶ generations, sampling trees every 1,000 generations (split frequencies fell below 0.01). Clade PP values were determined from a 50% majority-rule consensus of post burn-in trees (750,000 generations). Tree files for every possible root placement were generated in MacClade 4.0 (Maddison DR and Maddison WP 2000). Using the program GeneTree 1.3 (Page 1998), all rooting scenarios for the spidroin tree(s) were compared with reference to a fixed species phylogeny of the sampled taxa; our estimate of the species tree was based on Coddington et al. (2004). The number of gene duplication and loss events required to reconcile each root placement with the species tree was computed using the “show gene tree costs” command. Topologies that minimized gene duplications and losses were selected as preferred rooting scenarios.

Results and Discussion
Mygalomorph Spidroin Diversity

Using oligonucleotide probing and EST sequencing, we identified spidroin transcripts from the 4 mygalomorph spider silk gland cDNA libraries (GenBank accession number EU117159-EU117165). Each library yielded one or more sequence clusters, representatives of which could be conceptually translated into distinct spidroin-like polypeptides. All of these sequences had a single open reading frame that translated into repeating blocks of amino acids, followed by a non-repetitive C-terminal domain (fig. 2). BlastX searches conducted with these C-terminal sequences returned previously reported spidroin sequences with C-termini as top matches.

In total, we found 6 novel mygalomorph spidroins and sequenced an ~8-kb cDNA of E. chisoseus fibroin 1. From A. plutonis, we identified a single cluster of spidroin cDNAs, the longest of which was 4,729 bp (A. plutonis fibroin 1). The Aptostichus sp. silk gland library contained 2 spidroin sequence clusters, corresponding to at least 2 paralogous cDNAs: Aptostichus sp. fibroin 1 (3,388 bp) and fibroin 2 (2,411 bp). Screens of the B. californicum library revealed 3 paralogous spidroins: B. californicum fibroin 1 (1,873 bp), fibroin 2 (2,871 bp), and fibroin 3 (3,731 bp). The E. chisoseus silk gland cDNA library constructed by Gatesy et al. (2001) had 191 probe-positive clones. All sequenced Eua grus probe-positive cDNAs, including the 7,987-bp transcript that we report here, were extremely similar to the published E. chisoseus fibroin 1. This new cDNA extends the available data for E. chisoseus fibroin 1 by ~5.8 kb and overlaps with the sequence of Gatesy et al. (2001) over 2,144 bp. Where they overlap, these 2 sequences differ at 14 sites (8 synonymous, 3 non-synonymous, and 1 trinucleotide insertion/deletion [indel]), well within the reported range of allelic variation of spidroin genes (Beckwitt et al. 1998).

Analyses using etandem found that each mygalomorph spidroin cDNA contained long, tandem repeats ranging in length from 507 to 1,026 bp across sequences. Similarly, RADAR identified repeats at the amino acid level corresponding to the same dimensions and periodocities of those found by etandem. Like E. chisoseus fibroin 1, all other mygalomorph spidroins are characterized by a long ensemble repeat (169–183 amino acids). Eua grus chisoseus fibroin 1’s ensemble is approximately twice the length of the others (342 amino acids; fig. 2b), but RADAR also recognized subrepeats within E. chisoseus fibroin 1, each one half the length of the larger ensemble (fig. 2c). When aligned, these subrepeats share 56% identity over 178 sites, suggesting that the greater length of E. chisoseus fibroin 1’s ensemble may be explained by a duplication of an ancestral repeat similar in length to those found in other mygalomorph spidroins, followed by homogenization of this larger unit throughout the gene. Alternatively, a shift may have occurred in the periodicity of intragenic recombination tracts (i.e., from ~500 to ~1000 bp) in an ancestral silk gene already composed of serial repeats in the size range of those found in other mygalomorph spidroins (~500 bp).

Ensemble repeats of the different mygalomorph spidroins could not be readily decomposed into the simple amino acid sequence motifs (GA)n, GGX, and/or GPXn, that dominate certain araneomorph spidroins (e.g., the dragline silk proteins MaSp1 and MaSp2, the orb-web temporary scaffold protein MiSp, and the orb-web capture spiral protein Flag). However, the polyalanine (A)n motifs, found in MaSp1, MaSp2, and MiSp (Gatesy et al. 2001), are well represented in Eua grus and Bathoricrytum sequences but are less abundant in Aliatypus and Aptomostichus spidroins (fig. 2). A)n motifs are hypothesized to impart high tensile strength to dragline silk due to their assembly into β-crystalline sheets aligned in parallel to the fiber (Simmons et al. 1996; Parkhe et al. 1997) and may similarly contribute to the mechanical properties of mygalomorph silks. As previously noted from Eua grus, the mygalomorph spidroins contain homopolymeric runs of serine (S)n and threonine (T)n, in common with the araneomorph egg case spidroin TuSp1 (Garb and Hayashi 2005; Hu et al. 2005; Tian and Lewis 2005, 2006; Zhao et al. 2005) and spidroins of the haplogyne spider Plectreurys (Gatesy et al. 2001). Polyserine (S)n, while also found in the araneomorph prey-wrapping silk protein AcSp1 (Hayashi et al. 2004), is virtually absent in all other araneomorph silks characterized to date. Despite their shared A)n, S)n, and T)n motifs, mygalomorph ensembles could not be easily aligned across spidroins due to high length and sequence variability.

Mygalomorph spidroin amino acid compositions (predicted from cDNA translations) indicate that alanine and serine are by far the most abundant residues. Alanine and serine account for 50% or more of each spidroin, excepting those...
from *Aptostichus*, which are comparatively deficient in alanine. Although serine rich, mygalomorph spidroins have relatively little glycine (3.7–10%), a residue that constitutes 30–40% of MaSp1 and MaSp2 (fig. 3a). Within and across species, mygalomorph spidroins are compositionally similar.

By contrast, araneomorphs (e.g., orb-web and cob-web weavers) synthesize multiple, compositionally distinctive spidroins. Of these different araneomorph spidroins, the amino acid makeup of egg case silk protein (TuSp1) is most similar to mygalomorph spidroins (fig. 3a).

---

**Fig. 2.** (a) Schematic organization of mygalomorph spidroins: all are partial sequences encoding tandem-iterated repeat modules and a nonrepetitive C-terminal domain. The ellipsis (…) before the first depicted repeat module indicates predicted additional sequence not found in cDNA clones. (b) Consensus ensemble repeat sequence (majority-rule consensus of tandem repeats, an X indicating equivocal positions) of each spidroin followed by its repeat length and the length of the entire cDNA. Homopolymer runs of alanine (A<sub>n</sub>), serine (S<sub>n</sub>), and threonine (T<sub>n</sub>) are colored red, purple, and green, respectively. A portion of *Euagrus* fibroin 1’s ensemble sequence is bordered in gray to indicate the subrepeat referred to in (c). (c) An alignment of subrepeats within *Euagrus* fibroin 1 identified by RADAR; asterisks above alignment indicate identical positions. (d) Alignment of the C-termini from mygalomorph spidroins in interleaved format.
Predicted values of the amino acid composition of *E. chisoseus* fibroin 1 also closely match those determined by Palmer (1985) from silk fibers collected from the web of a *Euagrus* species (fig. 3b). The tight correspondence between these values suggests *E. chisoseus* fibroin 1 is the major component of *Euagrus* silk fibers.

By screening thousands of silk gland cDNAs, we found a single type of spidronin from *Aliatypus*, 1 from *Euagrus*, 2 from *Aptostichus*, and 3 from *Bothriocyrtum*. However, identical sampling procedures recovered 8 distinct types of spidroins from the derived araneomorph *D. spinosa* (Garb et al. 2006). Because of differential expression and sampling effects, we may not have identified all spidroins synthesized by the examined taxa. Nevertheless, our results suggest that mygalomorphs have a reduced diversity of spidroins relative to araneomorphs. These findings are consistent with the simple silk gland morphologies of mygalomorphs and the homogeneous amino acid composition of their silks compared with those of araneomorphs (fig. 3a).

**Intragenic Homogenization of Mygalomorph Spidroins**

Mygalomorph spidroins are composed of tandem repeats that are remarkably homogenized within a cDNA sequence (etandem identity scores of 87–100%). The previously reported *E. chisoseus* fibroin 1 was an ~2-kb partial sequence containing 2 repeat units and therefore provided minimal information regarding the repetitive organization of mygalomorph spidroins (Gatesy et al. 2001). The new ~8-kb fragment of *E. chisoseus* fibroin 1 is impressive in terms of its modular organization (fig. 4). This transcript contains 8 tandem repeats that are 1,026 or 1,032 bp in length. On average, these repeats are 98.5% identical and include 2 repeats (4 and 7 in fig. 4) that are 100% identical over 1,026 bp. When compared with their consensus sequence, *Euagrus* repeats vary at just 35 sites (24 synonymous and 11 nonsynonymous differences). Repeats in other mygalomorph spidroins show even greater levels of intragenic homogeneity. The *Aliatypus* cDNA contains 7 repeats of 543 bp that on average are 99% identical (see supplementary fig. S2, Supplementary Material online). These repeats differ at 18 sites (5 synonymous and 13 nonsynonymous differences) from their consensus. Moreover, the first 4 repeats in *Aptostichus* sp. fibroin 1 (507 bp each) are 100% identical, as are 2 of 3 repeats in *Aptostichus* sp. fibroin 2. Tandem repeats within the *Bothriocyrtum* cDNAs were similarly homogenized, having between 89 and 100% identity within sequences.
In addition to their highly similar nucleotide sequences, intragenic repeats of mygalomorph spidroins exhibited a relatively minor length variation. For instance, the ~1-kb *Eugagus* repeats contained just three 6 bp indels. These indel positions always corresponded to doublets of GCA, which always followed GCA doublets, a pattern that is suggestive of slip-strand replication error (fig. 4b). Within the *Aliatypus* sequence, tandem identified short subrepeats of 18 bp (93% average identity) encoding the motif “SAS-GAA” that was iterated 3 times in the first 7 ensemble repeats but 6 times in the eighth repeat (see supplementary fig. S2, Supplementary Material online). The 3 extra repetitions of this motif in this last repeat could be accounted for by 1 duplication event and 2 point mutations. All *Bothriocyrtum* sequences also showed minor variation in repeat length, almost always associated with indels of 1–3 codons that represent replicates of preceding codons.

The extreme homogeneity among intragenic repeats found within individual mygalomorph spidroins, coupled with divergence of these repeats across sequences, is suggestive of highly dynamic evolutionary processes. Homogeneity of tandem repeats is consistent with selection for identical repeat units in an ancestrally nonrepetitive sequence. Alternatively, this pattern may result from concerted evolution of ancestrally repetitive sequences, due to unequal crossing over events or through nonreciprocal recombination involving gene conversion (Charlesworth et al. 1994; Graur and Li 2000). Given the near uniformity of synonymous third codon positions in intragenic repeat alignments (e.g., fig. 4b), selection on ancestrally repetitive sequences is an unlikely explanation because third positions should exhibit greater variation unless constrained by selection for extreme codon bias. Such homogeneity of third codon positions is instead more compatible with a history of intragenic concerted evolution. Moreover, because all mygalomorph spidroins are composed of sequence repeats, the simplest explanation for their shared repetitive architecture is that it was inherited from a repetitive, ancestral sequence.

Concerted evolution has been proposed to explain similar patterns of modular architecture observed in araneomorph spidroins (Hayashi and Lewis 2000; Garb and Hayashi 2005), where intragenic repeats are homogenized but highly divergent across paralogs. These shared molecular features suggest that intragenic concerted evolution has been operating on the spidroin gene family prior to the divergence of mygalomorphs and araneomorphs (minimally dated to the early Triassic, ~240 MYA; Selden and Gall 1992). This further implies that spidroins synthesized by the common ancestor of mygalomorphs and araneomorphs were already composed of highly homogenized tandem repeats. Though convergently derived, insect silk proteins secreted by Lepidoptera, Trichoptera, and Diptera are also highly repetitive (Case and Byers 1983; Fedić et al. 2003; Yonemura et al. 2006), suggesting that this type of molecular organization is a fundamental prerequisite for silk fiber function.

**Spidroin Gene Family Evolution**

Phylogenetic analyses of the spidroin gene family were conducted using an amino acid alignment of the spidroin nonrepetitive C-terminal domain (122 characters for the 27 sampled sequences). Parsimony searches executed with gaps treated as missing data or considered as additional recoded characters retrieved a single most parsimonious tree (fig. 5). Majority-rule consensus trees resulting from Bayesian analyses of these data, though not identical to the parsimony tree, contained many of the same clades (clades supported by PP ≥ 0.50 shown in fig. 5). There were 51 possible root placements for the single most parsimonious tree. The minimal cost required to reconcile a particular gene tree with the given species tree was 16 duplications and 40 losses. Three of the 51 root placements shared this minimal cost, and these alternative rootings are depicted in figure 5.

Previous spidroin phylogenies were rooted with the mygalomorph sequence *Eugagus* fibroin 1 because mygalomorphs are sister to araneomorphs (from which all other sequences were known) (Garb and Hayashi 2005; Garb et al. 2006). However, inclusion of additional mygalomorph spidroins grouped neither araneomorph nor mygalomorph sequences as monophyletic (fig. 5). Instead, the resulting trees united all mygalomorph spidroins, except *Aliatypus* fibroin 1, in a clade receiving moderate support (BS = 61, DI = 5, PP = 1.0). *Aliatypus* fibroin 1 was placed in another clade that included the araneomorph aciniform gland silk protein AcSp1 (used in prey wrapping and reproduction) and *Plectreurys* fibroin 4 (*P.t.* fibroin 4), a spidroin that is yet to be functionally characterized.

Additional parsimony searches constraining all mygalomorph spidroins to monophyly found most parsimonious trees that were 7 steps longer than the unconstrained tree, a difference that was nonsignificant by both Kishino and Hasegawa (1989) and Templeton (1983) nonparametric tests. Although we cannot statistically reject monophyly of mygalomorph spidroins, our results suggest the occurrence of spidroin paralogs prior to the divergence of mygalomorph and araneomorph spiders. Proliferation of the spidroin gene family at this early stage would explain certain mygalomorph spidroins being more closely related to those in araneomorphs.

Within the clade of mygalomorph spidroins, relationships among sequences do not simply reflect expectations from species level phylogeny ([*Apostostichus*, *Bothriocyrtum*],...
may be indicative of lineage specific gene expansion within this species. Alternatively, the similarity of these sequences could be due to intergenic recombination events that would obscure orthology-paralogy relationships.

The large number of spidroin paralogs in some araneomorphs (e.g., 8 in the orbicularian D. spinosa) implicates widespread gene duplication and divergence in the diversification of spider silks. Some spidroin paralogs (e.g., Flag, MiSp1, MaSp1, and MaSp2) are associated with silk glands that are restricted to particular araneomorph lineages, indicating that these genes may have specialized along with the glands in which they are primarily expressed (Hayashi and Lewis 1998). In the spidroin gene family, minimally 10 duplication events have occurred in araneomorphs, whereas 4 are estimated within mygalomorphs (duplications computed by GeneTree 1.3, Page 1998). The reduced level of gene family expansion in mygalomorphs is consistent with the relative uniformity of their silk glands in morphology and function. By contrast, araneomorphs have experienced a greater diversification of spidroins in parallel with the functional specialization of their spinning apparatus. Even so, there is a wide diversity of mygalomorph spidroins (fig. 2), suggesting the need to further investigate the functional and evolutionary significance of these spider silk sequences.

**Supplementary Material**

Supplementary figures S1 and S2 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

**Acknowledgments**

This work was funded by National Science Foundation grants nos DEB-0236020 and MCB-9806999 and US Army Research Office grant nos DAAD19-02-1-0358 and W911NF-06-1-0455. We thank J. Bond, J. Gatesy, M. Hedin, N. Nguyen, C. Vink, V. Vo, and J. Woods for assistance with spider collections and laboratory work. N. Ayoub, J. Gatesy, M. Hedin, M. McGowen, and J. Starrett provided critical feedback on drafts of this manuscript.

**Literature Cited**


Adriana Briscoe, Associate Editor

Accepted August 17, 2007