

Piriform Spider Silk Sequences Reveal Unique Repetitive Elements

David J. Perry,^{†,*} Daniela Bittencourt,^{†,‡} Jessica Siltberg-Liberles,[†] Elibio L. Rech,[§] and Randolph V. Lewis[†]

Department of Molecular Biology, University of Wyoming, Laramie, Wyoming 821071-3944, United States, Laboratory of Molecular Biology, EMBRAPA Western Amazonia, Manaus, AM, Brazil, and Laboratory of Gene Transfer, Biotechnology Unit, EMBRAPA Genetic Resources and Biotechnology, Brasília, DF, Brazil

Received July 16, 2010; Revised Manuscript Received September 17, 2010

Orb-weaving spider silk fibers are assembled from very large, highly repetitive proteins. The repeated segments contain, in turn, short, simple, and repetitive amino acid motifs that account for the physical and mechanical properties of the assembled fiber. Of the six orb-weaver silk fibroins, the piriform silk that makes the attachment discs, which lashes the joints of the web and attaches dragline silk to surfaces, has not been previously characterized. Piriform silk protein cDNAs were isolated from phage libraries of three species: *A. trifasciata*, *N. clavipes*, and *N. cruentata*. The deduced amino acid sequences from these genes revealed two new repetitive motifs: an alternating proline motif, where every other amino acid is proline, and a glutamine-rich motif of 6–8 amino acids. Similar to other spider silk proteins, the repeated segments are large (>200 amino acids) and highly homogenized within a species. There is also substantial sequence similarity across the genes from the three species, with particular conservation of the repetitive motifs. Northern blot analysis revealed that the mRNA is larger than 11 kb and is expressed exclusively in the piriform glands of the spider. Phylogenetic analysis of the C-terminal regions of the new proteins with published spidroins robustly shows that the piriform sequences form an ortholog group.

Introduction

Orb-weaving spiders are capable of producing as many as six different types of silk fibers as well as glue-like proteins for a variety of tasks from capturing and wrapping prey to covering the egg case.^{1,2} The major ampullate silk is used to form the dragline from which spiders can suspend themselves as well as the primary structural frame of the web. Minor ampullate silk is responsible for the auxiliary spiral that further stabilizes the web and the flagelliform, or capture silk, is the highly elastic glycoprotein coated fiber capable of stopping and immobilizing flying prey. Tubuliform silk forms the tough outer shell of the egg case and aciniform silk is used for prey-wrapping. The proteins that make up all of these silks have been identified by cloning technologies, mainly based on mRNA.^{3–11} To date the protein(s) that comprises piriform silk, which forms the attachment discs that lash the joints of the web together, as well as attach dragline silk to solid surfaces, has not been identified from modern orb-weaving spiders.

The spider silk spidroins are composed of extremely large proteins, generally larger than 200kD, with unique repetitive arrangements of amino acids.^{10–12} The spidroin genes sequenced to date all encode a similar organization of highly conserved N-terminal and C-terminal regions of approximately 100 amino acids each flanking a large central segment of multiply repeated regions on the order of 30–200 amino acids per repeat. In the case of the major ampullate, minor ampullate, and flagelliform silks, the large repetitive regions are composed of short repeated sequence motifs.^{4,7,8,11–13} The aciniform and tubuliform genes,

on the other hand, encode long, complex repeats with no obvious simple internal repeated motifs.^{6,14}

The short, repetitive motifs have been of great interest for the creation of biomimetic fibers, as each of the short sequence motifs may provide a structural element to account for the basic properties of the fibers. Hence, new fibers of desired structural properties can be created by mixing and matching the various sequence motifs in a recombinant protein.¹⁵ The mechanical properties of most previously characterized spider silk proteins can be attributed to different combinations and arrangements of four repetitive amino acid motifs: polyalanine (A_n), alternating Gly-Ala (GA_n), amino acid triplets of two glycines followed by a variable residue (GGX_n), and Gly-Pro-Gly modules ($GPGXX_n$).¹ For example, MaSp1, one of the two proteins in major ampullate silk, contains repeated sequence motifs of GGX and poly-A; the first can form a 3_{10} -helix and the latter forms crystalline β -sheets.¹³ Both motifs contribute to the stiffness of the fiber. Flagelliform silk, on the other hand, contains multiple repeats of the GGX motif as well as a $GPGXX$ motif.⁷ The $GPGXX$ motif, also found in MaSp2, forms a highly elastic β -spiral that accounts for the elasticity of the capture silk^{16,17} and contributes to the supercontraction of the fiber when wet.¹⁸

To expand our knowledge of silk gene family members and provide other potential structural elements, we sought to sequence the protein(s) of the last uncharacterized silk type from orb-weavers, piriform. To that end, we constructed individual cDNA libraries from piriform gland-enriched fractions of *Argiope trifasciata* and *Nephila clavipes*. cDNAs were isolated that are expressed exclusively in the piriform glands of the spiders. A piriform cDNA was also identified from a third species, *Nephilengys cruentata*, by sequence similarity. The piriform cDNAs encode proteins with two new repetitive motifs, a glutamine-rich (QQ) sequence of Gln-Gln-Ser-Ser-Val-Ala (QQSSVA) and regions of alternating prolines (PXPXP), whose

* To whom correspondence should be addressed. Tel.: 307-766-5534. Fax: 307-766-5098. E-mail: dperry@uwyo.edu.

[†] University of Wyoming.

[‡] EMBRAPA Western Amazonia.

[§] EMBRAPA Genetic Resources and Biotechnology.

structure and physical properties may provide new elements to include in the biomimetic toolkit for designer nanofibers.

Materials and Methods

Spider Collection and Dissection. Female adult *N. clavipes* were purchased from Hatari Invertebrates (Portal, AZ). *A. trifasciata* were collected in Platte County, WY. *N. cruentata* were collected in the Atlantic Forest regions of Brazil. The spiders were maintained at room temperature in individual chambers with normal light/dark cycles and fed frozen/thawed crickets twice weekly. Spiders were anesthetized with CO₂ before dissection under a stereomicroscope. Individual silk glands were flash frozen in liquid N₂ and stored at -80 °C. Due to the difficulty in separating aciniform from piriform glands, libraries were constructed from mRNA isolated from piriform-enriched posterior regions.

mRNA Extraction. A total of 50–100 mg of individual glands or posterior regions was pooled and homogenized in TRI reagent, and total RNA was extracted according to the manufacturer's protocol (Molecular Research Center, Cincinnati, OH). mRNA was isolated using Dynabeads oligo(dT)₂₅ (Dyna, Inc., Lake Success, NY) for *N. clavipes* and *A. trifasciata* or by the Oligotex kit (Qiagen, Germany) for *N. cruentata*. The mRNA concentration was determined spectrophotometrically.

cDNA Library Construction. cDNA from *N. clavipes* and *A. trifasciata* was synthesized with Superscript II Choice System (Invitrogen, Carlsbad, CA), and Lambda ZAP II libraries (Stratagene, La Jolla, CA) were constructed essentially as in Tian and Lewis.¹⁰ Phage titering, library amplification, and plaque lifts were performed according to the manufacturer's protocols (Stratagene, La Jolla, CA) onto Hybond NX nylon membranes (GE Healthcare). Hybridization conditions and washes followed the manufacturer's suggestions. *N. cruentata* libraries were constructed in the Superscript II plasmid system with Gateway Technology (Invitrogen) and screened as in Bittencourt et al.¹⁹

Positive plaques from the phage libraries were subjected to at least one round of plaque purification before phagemid excision with the aid of the ExAssist helper phage (Stratagene). Phagemid transformants were grown in ampicillin-containing media, and plasmid DNA preparations (QIAprep miniprep kit) were screened for the presence of inserts by digestion with various restriction endonucleases followed by agarose gel electrophoresis. Plasmids with inserts larger than 1 kB were sequenced by the University of Wyoming Nucleic Acid Exploration Facility. The *N. cruentata* sequence was from one of the 960 plasmid clones randomly sequenced, but not identified, in Bittencourt et al.¹⁹

Hybridization Probes. Initial rounds of plaque screening used an oligonucleotide probe based on the *A. trifasciata* major ampullate sequence: 5'-CCW GGW CCT TGT CCW GGW CC corresponding to the complement of the reading frame for the peptide sequence GPGQQGPG. Later rounds of screening used a 579 bp *Pst*I restriction fragment from the initial *A. trifasciata* positive clone to identify a larger clone. The *N. cruentata* clone was identified by sequence homology to the *A. trifasciata* clone. When no positive signals were obtained from the *N. clavipes* library using *A. trifasciata* probes, that library was successfully rescreened using restriction fragments from the homologous *N. cruentata* clone. Restriction fragment probes were labeled with ³²P-dCTP by the Prime-It II Random Primer Labeling Kit (Stratagene, La Jolla, CA); oligonucleotides were end-labeled with ³²P-ATP by polynucleotide kinase (New England Biolabs).

DNA Sequencing. Plasmids were submitted to the University of Wyoming Nucleic Acid Exploration Facility for cycle sequencing with the BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystems). The purified reactions were subjected to capillary electrophoresis on an Applied Biosystems model 3130 Genetic analyzer. Standard primers for pBluescript SK (T3 promoter, T7 promoter, M13 forward, and M13 reverse) were utilized for the reactions. The *N. clavipes* sequence was completed by opening the pBluescript vector with *Apa*I and *Xho*I for sequential digestion and ligation by the Erase-a-Base protocol (Promega). Sequences were searched against the NCBI nonredundant database by BLAST (<http://blast.ncbi.nlm.nih.gov/>²⁰) and unknown sequences

were scrutinized further for silk-like characteristics. Sequences were edited with 4Peaks (<http://mekentosj.com>) and alignments, translations, and other manipulations were performed online with the Biology Workbench suite of programs (<http://workbench.sdsc.edu>) at the San Diego Supercomputer Center or locally with MacVector.

Polymerase Chain Reaction. Because the *N. clavipes* cDNA clone contained only repeated motifs and did not extend into the C-terminal region, gDNA was extracted from whole spiders according to the manufacturer's protocol for the Wizard Genomic DNA purification kit (Promega) for use as a template for PCR. The upstream primer, based on the known sequence, was 5'-GGACAAGAACAGAGTAT TTC. The downstream primers (5'-CAAYAAAGCRGCCARRGTTTCYAAGAG and 5'-TTTATTTCCCATTAACAAAAA) were based on the two known sequences from *N. cruentata* and *A. trifasciata*, with the first primer falling within the 3' coding region and the second primer placed 95 bp downstream of the known stop codons. PCR reactions were assembled according to the manufacturer's instructions for GoTaq Green (Promega). Cycling conditions, after an initial denaturation step of 4 min at 95 °C were 30 s at 95 °C, 30 s at 43.5 °C, and 2 min at 72 °C, repeated 30 times. A 774 and 975 bp band were purified by agarose gel electrophoresis, and sequenced with the PCR primers.

Northern Blot. Total RNA was purified as described above. 10 µg aliquots were electrophoresed on a denaturing formaldehyde gel and capillary transferred to Hybond NX nylon membrane (GE Healthcare). Membranes were hybridized with a ³²P-labeled gel-purified PvuII/SacI piriform restriction fragment overnight at 55 °C, washed stringently, and exposed to X-ray film (Fuji Medical). Blots were stripped with boiling 0.1 X SSC and 0.1% SDS and reprobed with ³²P-labeled mouse rRNA (DECAtemplate 18S-Mouse, Ambion). A BLAST search of the NCBI database had previously found that the mouse rRNA is 85% identical to the 28S rRNA of the closely related spider, *Argiope argentata* (accession no. AY230889). Signals were quantified with the UVP Biospectrum Imaging system and associated software. Signals were normalized to the corresponding rRNA.

Amino Acid Analysis. We observed when silking spiders for major ampullate silk, the spiders appeared to begin with a short burst of piriform silk before releasing the major ampullate fiber. This short fibrous burst was collected and sent to the Proteomics Core Facility at the University of California, Davis, for protein hydrolysis in 6 N HCl/0.1% phenol at 110 °C for 24 h and amino acid analysis on the Hitachi L-8800 amino acid analyzer (Table SI). The similarity of the resulting analysis to previously reported amino acid compositions from the luminal contents of piriform glands^{21,22} confirmed our hypothesis.

Phylogenetic Analysis. A phylogenetic analysis based on the C terminal amino acid sequence from the herein sequenced genes and a recently published piriform sequence from the cob weaving spider, *L. hesperus*, as well as other spider silk sequences from the orb-weaving spiders was performed. The fibroin sequence of *Euagrus chisoseus* was included as an outgroup to root the tree. The GenBank accession numbers were as follows: *N. clavipes*, MaSp1 (U20329), MaSp2 (M92913), MiSp1 (AF027735), MiSp2 (AF027736), TuSp (AY855102), Flag (AF027973), PiSp (GQ980330); *A. trifasciata*, MaSp1 (AF350266), MaSp2 (AF350267), Flag (AF350264), AcSp (AY426339), PiSp (GQ980328); *N. cruentata*, MaSp-like (EF638446), MiSp-like (EF638447), TuSp-like (EF638445), Flag-like (EF638444), PiSp (GU062417); *L. hesperus*, PySp (FJ973621); and *E. chisoseus*, Fibroin 1 (AF350271). The sequences were aligned using Muscle (Figure S2), and model testing was performed using Prottest.²³ Based on the best Prottest model (WAG²⁴ with four gamma categories), the tree was constructed by MrBayes3 software.^{25,26} The tree was run for 25165000 generations and the consensus tree was constructed with a default burn-in phase.

Results

Three species of spiders, *Argiope trifasciata*, *Nephila clavipes*, and *Nephilengys cruentata*, were collected or purchased from

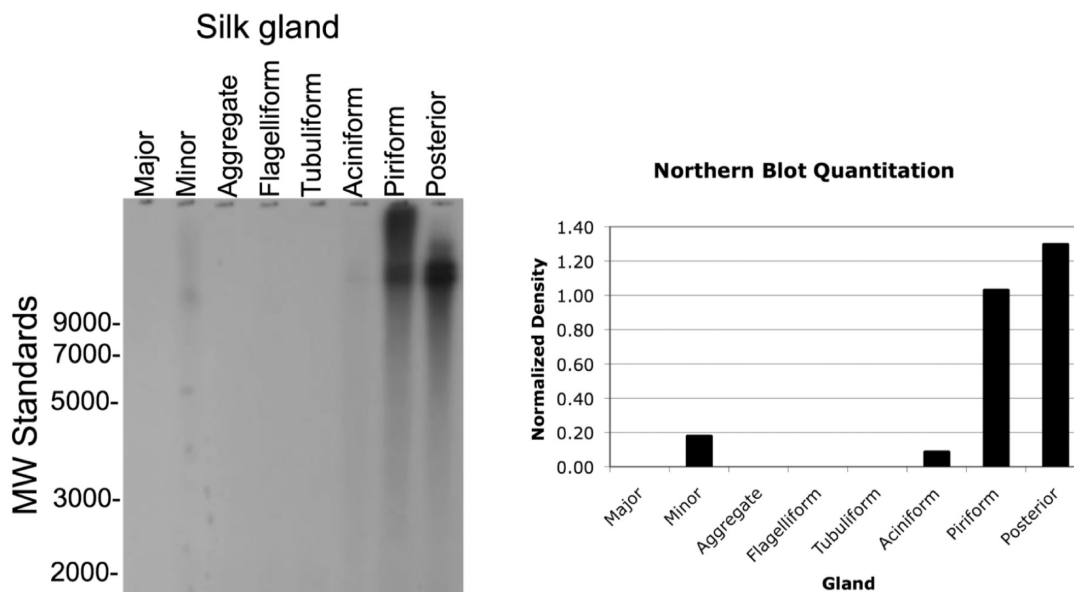


Figure 2. Northern blot of glandular RNAs. Total RNA purified from various silk glands of *A. trifasciata* was probed with a ^{32}P -labeled restriction fragment from the piriform cDNA clone. Signals were quantified with the UVP Biospectrum imaging system and normalized to the corresponding 18S rRNA signal on the same blot.

Discussion

We report here for the first time cDNA sequences of piriform silk protein from three different species of orb-weaving spiders. As with other spider silk spidroin proteins, the mRNAs are extremely large and are organized into multiply repeated sequence segments that contain repetitive motifs within them.^{5,13,30} The piriform repeated segments are highly homogenized within any one species such that the two complete *N. cruentata* repeated segments are 98% identical at the nucleic acid and amino acid levels with one repeat containing one more eight amino acid QQ motif than the other (Figure 3A). The three *N. clavipes* repeated segments are approximately 93% identical at both levels with the repeated segments varying in the number of QQ-motifs by one or two (Figure 3B).

Comparison across orthologs of the three species shows the cDNA of the repeated segments to be 43% identical (Figure 4). However, the *N. cruentata* and *N. clavipes* sequences are more closely related to each other than either one is to *A. trifasciata* in that they share 74% identity. Repeated segments from both retain only 52–54% sequence identity with the *A. trifasciata* sequence.

This is the first report of two new repetitive sequence motifs, referred to here as the PXP and QQ motifs that have not been found in any other spider silk protein sequences. In other spider silk proteins, repetitive motifs are thought to correspond to specific structural elements and physical properties. For example, the repeated poly alanine motifs of MaSp1 form β -sheets that provide strength,³¹ whereas the repeated GPGXX motifs in MaSp2 are expected to form biological nanosprings that account for the flexibility in the major ampullate fiber.^{17,32} In this case, the functional significance of the repetitive motifs has yet to be structurally defined. The PXP motifs are obviously structurally constrained by the alternating fixed angles of the proline residues, but there is no obvious periodicity to the intervening amino acids to suggest an ordered placement of hydrophobic versus hydrophilic residues. The seemingly random arrangement of the basic amino acid, arginine, among the nonpolar residues, alanine and leucine, would also appear to rule out a simple interchain superhelical structure.

The QQ motifs, on the other hand, display a degree of periodicity alternating between the polar residues of glutamine and serine followed by nonpolar residues. If the purpose of piriform silk is essentially to stick to foreign surfaces and other spider silks, this periodic arrangement of polar and nonpolar residues might present an ideal surface for both hydrophobic as well as hydrophilic interactions.

There are other examples of proteins that have repeated peptide motifs beginning with QQ, but there is no structural data for them. For example wheat gluteins tend to have high glutamine content and one high molecular weight glutenin contains the consensus hexapeptide repeat QPQGQ.^{33–36} While these glutamine-rich repetitive sequences are thought to provide extensive networks of interchain hydrogen bonds, the elasticity of the gluten is attributed to the presence of the proline residues.³⁶ Another protein, rat α -casein contains a motif of QQASLA repeated nine times, but this appears to be an insertion of no special structural relevance since it does not appear in orthologous proteins from other species.³⁷ In contrast, the conservation of the QQSSVA motif across spider species separated evolutionarily by at least 125 million years⁵ would imply this is a structurally relevant theme.

Virtually all algorithms for predicting protein secondary structure identify the PXP motifs as random coil. The model peptide studies of Rauscher et al.³⁸ would suggest that the proline-rich motifs might promote elastomeric properties. The QQ motifs are predicted to be α -helix by some structure prediction methods³⁹ and β -sheets by others.⁴⁰ The inverse dependence on glycine content of the folding index proposed by Dicko et al.⁴¹ would also imply that these sequences might adopt a relatively ordered structure. Algorithms that purport to predict global regions of protein order and disorder variously predict the PXP regions to be disordered bookends to a more globular domain for the QQSSVA regions (GlobPlot,⁴² DisEMBL by loops/coil definition⁴³), exactly the opposite with the PXP regions ordered and the QQSSVA regions disordered (DisEMBL by Remark-465 definition⁴³), or completely disordered in virtually all regions (PONDR⁴⁴). All of these disagreements serve to highlight the difficulties of using algorithms based

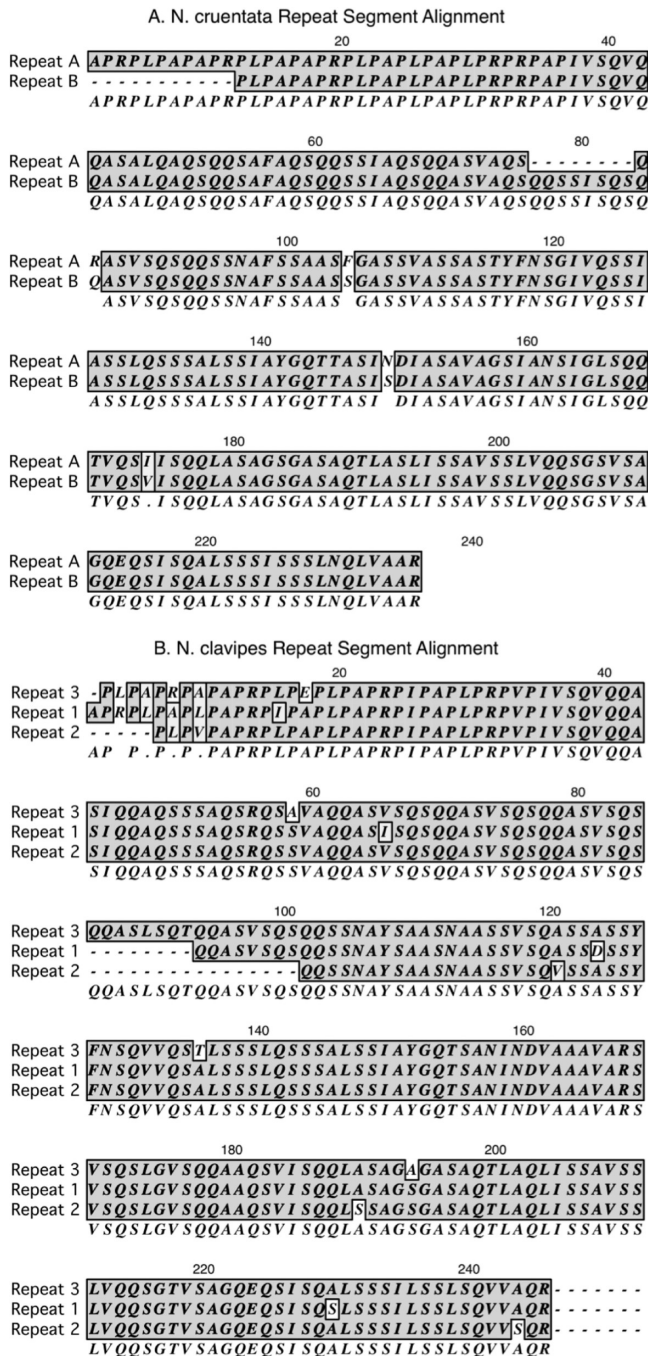


Figure 3. Alignment of repeated protein segments. The repeated segments of the piriform protein sequences were aligned by ClustalW. (A) Alignment of the *N. cruentata* repeated segments and (B) alignment of the *N. clavipes* repeated segments. Conserved and similar amino acids are boxed and shaded in gray; the consensus sequence appears below. Note the primary repeat segment length differences for both species are due to varying numbers of QQ motifs.

primarily on soluble proteins for predicting the structure of insoluble proteins.

Previously, the final approximately 100 amino acids of the C-terminal sequences were found to be much more highly conserved across both orthologous and paralogous genes than the repetitive regions and that is the case here as well. Across all three species, the C-terminal amino acid sequences were 71% identical (Figure 5). As with the repetitive sequences, the amino acid sequences from *N. clavipes* and *N. cruentata* were much more closely related (92% identity) than either was to the *A.*

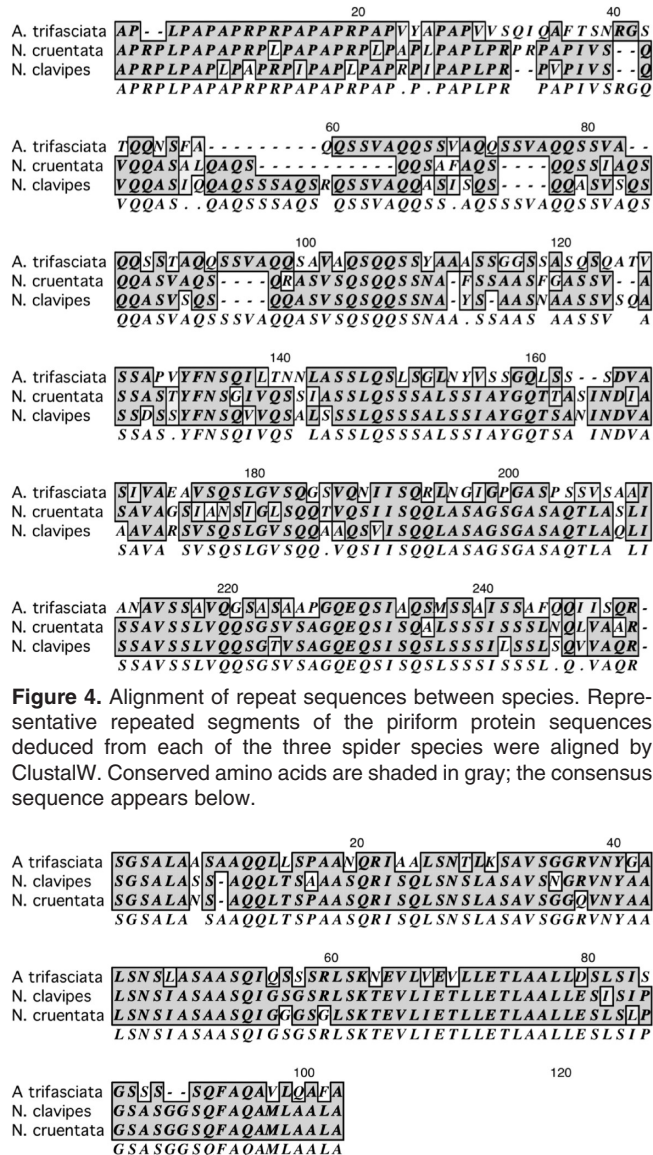


Figure 4. Alignment of repeat sequences between species. Representative repeated segments of the piriform protein sequences deduced from each of the three spider species were aligned by ClustalW. Conserved amino acids are shaded in gray; the consensus sequence appears below.

Figure 5. Piriform C-terminal alignments. The C-terminal region of approximately 100 amino acid residues from each of the three piriform protein sequences was aligned by ClustalW. Conserved and similar amino acids are boxed and shaded in gray; the consensus sequence appears below.

trifasciata sequence. A similar degree of conservation was found at the nucleic acid level as well with identity across the three cDNAs of 68%.

During the preparation of this manuscript, another piriform protein sequence was reported from the black widow spider, *L. hesperus*.⁴⁵ The sequence of the *L. hesperus* protein is drastically different from the three reported here with a repetitive motif of AAARAQAQAE and nothing resembling the QQSSVA motifs found here and completely lacking the amino acid proline. This sequence results in a calculated amino acid composition for the ensemble repeat of over 45% alanine (Table 1). The calculated amino acid compositions of neither our three protein sequences nor the *L. hesperus* sequence correspond well to the published experimentally determined amino acid compositions of the luminal contents of piriform glands (Tables 1 and SI). In the calculated amino acid compositions, only the repetitive regions were used since partial sequences are likely to give too much weight to the amino acids in the C-terminal region when compared to the complete protein. These amino acid composi-

Table 1. Representative Amino Acid Composition of Piriform Silk Proteins^a

amino acid	partial amino acid composition of piriform silk proteins							
	experimental amino acid analysis				calculated amino acid composition			
	<i>N. clavipes</i>	<i>A. diadematus</i>	<i>A. aurantia</i>	<i>L. hesperus</i>	<i>A. trifasciata</i>	<i>N. clavipes</i>	<i>N. cruentata</i>	<i>L. hesperus</i>
Ala	18.0	9.9	13.1	19.4	16.9	16.0	17.5	45.9
Ser	16.9	14.9	16.1	10.2	27.7	26.9	26.0	6.1
Pro	6.4	7.8	5.0	1.9	8.8	8.3	8.7	0.0
His	0.9	2.8	1.4	4.5	0.2	0.0	0.0	4.0
Arg	2.9	3.6	3.5	5.2	2.3	3.0	2.8	9.3

^a The compositions of a representative set of the amino acids determined experimentally from *A. diadematus*,²¹ *A. aurantia*,²² *L. hesperus*,⁴³ and *N. clavipes* (Table SI) are compared to the calculated amino acid compositions of the repetitive regions of *A. trifasciata*, *N. clavipes*, *N. cruentata* (all from this paper), and *L. hesperus*.⁴⁵

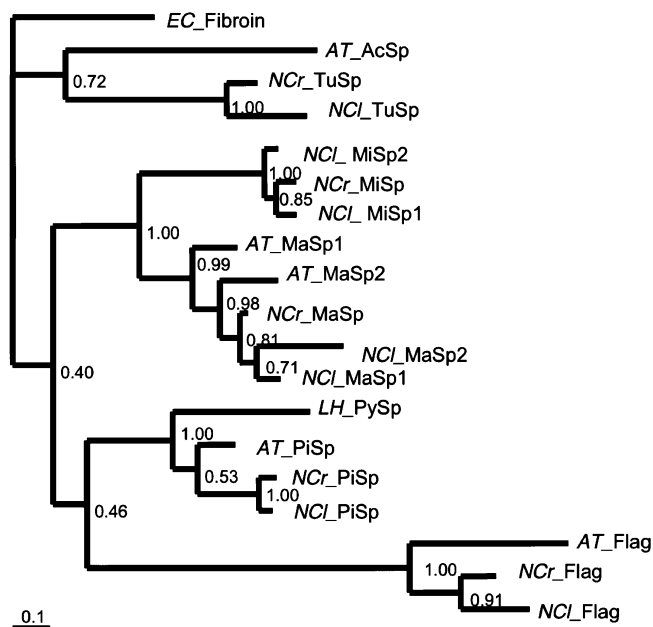


Figure 6. Phylogenetic analysis of piriform protein C-terminals. A phylogenetic tree was constructed from the C-terminal protein sequences of all known spidroins from three species of spiders: *A. trifasciata*, *N. clavipes*, and *N. cruentata*. In addition, the recently published C-terminal sequence from *L. hesperus* was included to show its relationship to the piriform proteins reported in this paper. For brevity, *E. chosiseus* is abbreviated *EC*; *A. trifasciata*, *AT*; *N. clavipes*, *NCl*; *N. cruentata*, *NCr*; and *L. hesperus*, *LH*. The numbers at the nodes in the tree reflect the posterior probabilities.

tion comparisons are not completely ideal as there is undoubtedly species variation and since piriform glands are notoriously difficult to dissect cleanly from spiders, factors that probably account for the degree of variation between the different experimentally determined amino compositions shown in Table 1. Furthermore, the three sequences reported here are from orb-weaving spiders, whereas *L. hesperus* is a cob-weaving spider. It is entirely plausible that the homologues of both proteins are in piriform silk. For instance, piriform proteins of the family reported here could account for the levels of serine and proline found experimentally in luminal contents, whereas a protein similar to the *L. hesperus* sequence could account for the histidine and arginine levels found (Table 1). If the repetitive motifs are as conserved across all spider species as well as they are in the three species reported here (Figure 4), the comparisons strongly suggest piriform silk is comprised of more than one protein.

A phylogenetic tree was constructed from the conserved C-terminal regions of the known silk fibroin sequences of the three species studied here with the addition of the piriform protein from *L. hesperus* (Figure 6). The tree is stable in

separating all different spider silk types in this set, but does not really resolve the relationships between the different types. However, this analysis reveals that the three piriform sequences of the orb-weaving spiders reported here are indeed close homologues to piriform sequences from the cob-weaving spider, *L. hesperus*, with a posterior probability of 1.00 despite the major differences that exist in the repetitive regions of the proteins. This remarkable degree of relationship between the proteins makes the idea that homologues of both proteins may be included in the piriform fiber even more plausible.

In summary, we have sequenced the first cDNA for the last uncharacterized orb-weaving spider silk protein, piriform. The DNA sequence encodes a protein that fits well within the canonical structure of spider silks in possessing large repeated segments containing within them highly repetitive, short sequence motifs and a C-terminal region that is conserved across species. Because piriform silk is extruded from multiple spinnerets in relatively short lengths, it has been impossible to obtain natural material for structural studies. Based on these partial cDNAs, it will be possible to construct recombinant proteins to elucidate the structural properties of piriform silk. Finally, the two repetitive motifs unique to these sequences may provide new elements to be included in the toolbox of cassettes that can be used for the creation of bioinspired nanofibers with desirable mechanical properties.

Acknowledgment. We thank Melinda Creager for her skills in dissecting spiders. We also thank Drs. Michael Hinman and Holly Steinkraus for critical reading of the manuscript. The Table of Contents graphic was created by WebLogo, version 2.8.2.⁴⁶ This study was supported in part by NIH and AFOSR grants to R.V.L., EMBRAPA and CNPq (486492/2006-0) to E.L.R., and J.S.-L. is supported by an NIH-INBRE grant awarded to the University of Wyoming.

Supporting Information Available. Partial cDNA sequences from which the piriform silk protein sequences were deduced (Figure S1), the alignment of the piriform protein C-terminal amino acid sequences (Figure S2) that were used to construct the phylogenetic tree in Figure 6, and the complete amino acid analysis of the *N. clavipes* piriform silk (Table SI). This material is available free of charge via the Internet at <http://pubs.acs.org>.

References and Notes

- (1) Lewis, R. V. Spider silk: Ancient ideas for new biomaterials. *Chem. Rev.* **2006**, *106* (9), 3762–74.
- (2) Vollrath, F. Spider webs and silks. *Sci. Am.* **1992**, *266*, 70–76.
- (3) Choresch, O.; Bayarmagnai, B.; Lewis, R. V. Spider web glue: Two proteins expressed from opposite strands of the same DNA sequence. *Biomacromolecules* **2009**, *10* (10), 2852–6.
- (4) Colgin, M. A.; Lewis, R. V. Spider minor ampullate silk proteins contain new repetitive sequences and highly conserved “non-silk-like” spacer regions. *Protein Sci.* **1998**, *7* (3), 667–72.

- (5) Gatesy, J.; Hayashi, C.; Motriuk, D.; Woods, J.; Lewis, R. Extreme diversity, conservation, and convergence of spider silk fibroin sequences. *Science* **2001**, *291* (5513), 2603–5.
- (6) Hayashi, C. Y.; Blackledge, T. A.; Lewis, R. V. Molecular and mechanical characterization of aciniform silk: Uniformity of iterated sequence modules in a novel member of the spider silk fibroin gene family. *Mol. Biol. Evol.* **2004**, *21* (10), 1950–9.
- (7) Hayashi, C. Y.; Lewis, R. V. Evidence from flagelliform silk cDNA for the structural basis of elasticity and modular nature of spider silks. *J. Mol. Biol.* **1998**, *275*, 773–784.
- (8) Hinman, M. B.; Lewis, R. V. Isolation of a clone encoding a second dragline silk fibroin. *Nephila clavipes* dragline silk is a two-protein fiber. *J. Biol. Chem.* **1992**, *267* (27), 19320–4.
- (9) Motriuk-Smith, D.; Smith, A.; Hayashi, C. Y.; Lewis, R. V. Analysis of the conserved N-terminal domains in major ampullate spider silk proteins. *Biomacromolecules* **2005**, *6* (6), 3152–9.
- (10) Tian, M.; Lewis, R. V. Molecular characterization and evolutionary study of spider tubuliform (eggcase) silk protein. *Biochemistry* **2005**, *44* (22), 8006–12.
- (11) Xu, M.; Lewis, R. V. Structure of a protein superfiber: spider dragline silk. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87* (18), 7120–4.
- (12) Hayashi, C. Y.; Lewis, R. V. Molecular architecture and evolution of a modular spider silk protein gene. *Science* **2000**, *287* (5457), 1477–9.
- (13) Ayoub, N. A.; Garb, J. E.; Tinghitella, R. M.; Collin, M. A.; Hayashi, C. Y. Blueprint for a high-performance biomaterial: full-length spider dragline silk genes. *PLoS ONE* **2007**, *2* (6), e514.
- (14) Garb, J. E.; Hayashi, C. Y. Modular evolution of egg case silk genes across orb-weaving spider superfamilies. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (32), 11379–84.
- (15) Blackledge, T. A.; Hayashi, C. Y. Silken toolkits: biomechanics of silk fibers spun by the orb web spider *Argiope argentata* (Fabricius 1775). *J. Exp. Biol.* **2006**, *209* (Pt 13), 2452–61.
- (16) Hayashi, C. Y.; Shipley, N. H.; Lewis, R. V. Hypotheses that correlate the sequence, structure, and mechanical properties of spider silk proteins. *Int. J. Biol. Macromol.* **1999**, *24* (2–3), 271–5.
- (17) Liu, Y.; Shao, Z.; Vollrath, F. Elasticity of spider silks. *Biomacromolecules* **2008**, *9* (7), 1782–6.
- (18) Liu, Y.; Spöner, A.; Porter, D.; Vollrath, F. Proline and processing of spider silks. *Biomacromolecules* **2008**, *9* (1), 116–21.
- (19) Bittencourt, D.; Souto, B. M.; Verza, N. C.; Vinecky, F.; Dittmar, K.; Silva, P. I., Jr.; Andrade, A. C.; da Silva, F. R.; Lewis, R. V.; Rech, E. L. Spidroins from the Brazilian spider *Nephilengys cruentata* (Araneae: Nephilidae). *Comp. Biochem. Physiol., Part B: Biochem. Mol. Biol.* **2007**, *147* (4), 597–606.
- (20) Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25* (17), 3389–402.
- (21) Andersen, S. O. Amino acid composition of spider silks. *Comp. Biochem. Physiol.* **1970**, *35* (3), 705–711.
- (22) Tillinghast, E. K. The chemical fractionation of the orb web of *Argiope* spiders. *Insect Biochem.* **1984**, *14* (1), 115–120.
- (23) Abascal, F.; Zardoya, R.; Posada, D. ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics* **2005**, *21* (9), 2104–5.
- (24) Whelan, S.; Goldman, N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* **2001**, *18* (5), 691–9.
- (25) Huelsenbeck, J. P.; Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **2001**, *17* (8), 754–5.
- (26) Ronquist, F.; Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **2003**, *19* (12), 1572–4.
- (27) Spöner, A.; Schlott, B.; Vollrath, F.; Unger, E.; Grosse, F.; Weisshart, K. Characterization of the protein components of *Nephila clavipes* dragline silk. *Biochemistry* **2005**, *44* (12), 4727–36.
- (28) Beckwith, R.; Arcidiacono, S. Sequence conservation in the C-terminal region of spider silk proteins (Spidroin) from *Nephila clavipes* (Tetragnathidae) and *Araneus bicentenarius* (Araneidae). *J. Biol. Chem.* **1994**, *269* (9), 6661–3.
- (29) Guerette, P. A.; Ginzinger, D. G.; Weber, B. H.; Gosline, J. M. Silk properties determined by gland-specific expression of a spider fibroin gene family. *Science* **1996**, *272* (5258), 112–5.
- (30) Hayashi, C. Y. Evolution of spider silk proteins: insight from phylogenetic analyses. In *Molecular Systematics and Evolution: Theory and Practice*; DeSalle, R., Giribet, G., Wheeler, W., Eds.; Birkhäuser Verlag: Switzerland, **2002**, 209–23.
- (31) Gosline, J. M.; Guerette, P. A.; Ortlepp, C. S.; Savage, K. N. The mechanical design of spider silks: from fibroin sequence to mechanical function. *J. Exp. Biol.* **1999**, *202* (Pt 23), 3295–303.
- (32) Hayashi, C. Y.; Lewis, R. V. Spider flagelliform silk: lessons in protein design, gene structure, and molecular evolution. *Bioessays* **2001**, *23* (8), 750–6.
- (33) Tatham, A. S.; Shewry, P. R.; Mifflin, B. J. Wheat gluten elasticity: a similar molecular basis to elastin. *FEBS Lett.* **1984**, *177* (2), 205–208.
- (34) Sugiyama, T.; Rafalski, A.; Peterson, D.; Soll, D. A wheat HMW glutenin subunit gene reveals a highly repeated structure. *Nucleic Acids Res.* **1985**, *13* (24), 8729–8737.
- (35) Tatham, A. S.; Shewry, P. R. Comparative structures and properties of elastic proteins. *Philos. Trans. R. Soc., B* **2002**, *357* (1418), 229–34.
- (36) Shewry, P. R.; Halford, N. G.; Belton, P. S.; Tatham, A. S. The structure and properties of gluten: an elastic protein from wheat grain. *Philos. Trans. R. Soc., B* **2002**, *357* (1418), 133–42.
- (37) Hobbs, A. A.; Rosen, J. M. Sequence of rat α - and γ -casein mRNAs: evolutionary comparison of the calcium-dependent rat casein multigene family. *Nucleic Acids Res.* **1982**, *10* (24), 8079–98.
- (38) Rauscher, S.; Baud, S.; Miao, M.; Keeley, F. W.; Pomes, R. Proline and glycine control protein self-organization into elastomeric or amyloid fibrils. *Structure* **2006**, *14* (11), 1667–76.
- (39) Garnier, J.; Gibrat, J. F.; Robson, B. GOR method for predicting protein secondary structure from amino acid sequence. *Methods Enzymol.* **1996**, *266*, 540–53.
- (40) Holley, L. H.; Karplus, M. Protein secondary structure prediction with a neural network. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86* (1), 152–6.
- (41) Dicko, C.; Porter, D.; Bond, J.; Kenney, J. M.; Vollrath, F. Structural disorder in silk proteins reveals the emergence of elastomericity. *Biomacromolecules* **2008**, *9* (1), 216–21.
- (42) Linding, R.; Russell, R. B.; Neduva, V.; Gibson, T. J. GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res.* **2003**, *31* (13), 3701–8.
- (43) Linding, R.; Jensen, L. J.; Diella, F.; Bork, P.; Gibson, T. J.; Russell, R. B. Protein disorder prediction: implications for structural proteomics. *Structure* **2003**, *11* (11), 1453–9.
- (44) Obradovic, Z.; Peng, K.; Vucetic, S.; Radivojac, P.; Brown, C. J.; Dunker, A. K. Predicting intrinsic disorder from amino acid sequence. *Proteins* **2003**, *53* Suppl 6, 566–72.
- (45) Blasingame, E.; Tuton-Blasingame, T.; Larkin, L.; Falick, A. M.; Zhao, L.; Fong, J.; Vaidyanathan, V.; Visperas, A.; Geurts, P.; Hu, X.; La Mattina, C.; Viera, C. Pyriform spidroin 1, a novel member of the silk gene family that anchors dragline silk fibers in attachment discs of the black widow spider, *Latrodectus hesperus*. *J. Biol. Chem.* **2009**, *284* (42), 29097–108.
- (46) Crooks, G. E.; Hon, G.; Chandonia, J. M.; Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **2004**, *14* (6), 1188–90.