Available online at www.sciencedirect.com
ScienceDirect
Comparative Biochemistry and Physiology, Part B 147 (2007) 597–606
CBP
www.elsevier.com/locate/cbpb

Spidroins from the Brazilian spider *Nephilengys cruentata* (Araneae: Nephilidae)

D. Bittencourt [a,d,*], B.M. Souto [a,d], N.C. Verza [d], F. Vinecky [d], K. Dittmar [b], P.I. Silva Jr. [c], A.C. Andrade [d], F.R. da Silva [d], R.V. Lewis [e], E.L. Rech [d]

[a] *Instituto de Ciências Biológicas, Departamento de Biologia Celular, Universidade de Brasília, Brasília—DF, Brazil*
[b] *Department of Molecular Biology, Bioinformatics Group, University of Wyoming, Laramie—WY, USA*
[c] *Laboratório Especial de Toxicologia Aplicada de Artrópodes-Instituto Butantan, São Paulo—SP, Brazil*
[d] *Núcleo de Biotecnologia, EMBRAPA Recursos Genéticos e Biotecnologia, Brasília—DF, Brazil*
[e] *Department of Molecular Biology, University of Wyoming, Laramie—WY, USA*

Received 31 January 2007; received in revised form 28 March 2007; accepted 31 March 2007
Available online 6 April 2007

Abstract

Spiders produce up to six different kinds of silk, each one for a specific biological function. Spider silks are also known for their unique mechanical properties. The possibility of producing new materials with similar properties motivated research on these silk proteins (spidroins). Using expression sequence tags, we identified four spidroins produced by major ampullate, minor ampullate, flagelliform and tubuliform silk glands from the Brazilian spider *Nephilengys cruentata* (Araneae: Nephilidae). The new protein sequences showed substantial similarity to other spidroins previously described, with high content of alanine and glycine due to the presence of the highly repetitive motifs (polyAla, $(GA)_n$, $(GGX)_n$, $(GPGGX)_n$). Similarities among sequences were also observed between the different spidroins with the exception of tubuliform spidroin, which presents a unique complex amino acid sequence with high amounts of serine and low amounts of glycine.

Keywords: EST; Sequence; Silk; Spider; Spidroins

1. Introduction

Orb-web spiders can produce a variety of silks with unique mechanical properties for diverse practical purposes (Foelix, 1996). The fibers constituents are proteins (spidroins) which are synthesized in the epithelial cells of specialized glands and secreted into the glandular lumen where they are stored in a highly concentrated (50% w/v) liquid crystalline spin dope (Hijirida et al., 1996; Scheibel, 2004). Fiber assembly is believed to occur during the passage of the spin dope through the spinning duct, where extraction of water, sodium and chloride is accompanied by a decrease of pH from 6.9 to 6.3 until extrusion from spinnerets as insoluble fibers (Vollrath and Knight, 2001; Rising et al., 2005; Vollrath, 2005). Despite knowledge available on some of the details, the process of spinning fibers is still not clearly understood.

Orb-web weavers have up to seven different types of glands, six producing a specific silk and one the glue (Table 1) (Vollrath, 1992). Studies on various Araneoid spidroin cDNAs sequences from different silk glands have shown that they share common structural features (Xu and Lewis, 1990; Gatesy et al., 2001). All cDNAs are large transcripts consisting of a non-repetitive and conserved N-terminal and C-terminal, and an internal highly repetitive region rich in alanine, glycine and serine amino acids. The repetitive region of most spidroins is formed by assembly of up to four amino acid motifs responsible for distinct structural modules as follows: poly-alanine (A_n), alternating glycine and alanine $(GA)_n$, amino acid triplets composed of two glycines and a third variable amino acid $(GGX)_n$, and glycine-proline-glycine modules $(GPGXX)_n$

* Corresponding author. Instituto de Ciências Biológicas, Departamento de Biologia Celular, Universidade de Brasília, Brasília—DF, Brazil. Tel./fax: +55 61 3448 4694.
E-mail address: dmatias@cenargen.embrapa.br (D. Bittencourt).

1096-4959/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.cbpb.2007.03.013

Table 1
Spider glands, proteins and their uses

Spider gland	Product	Function
Major ampullate gland	Major ampullate spidroins 1 and 2	Drag line, structural silk
Minor ampullate gland	Minor ampullate spidroins 1 and 2	Auxiliary spiral
Flagelliform gland	Flagelliform spidroin	Core fibers of capture spiral
Tubuliform gland	Tubuliform spidroin	Egg sac
Aciniform gland	Aciniform spidroin	Swathing pray
Piriform gland	Piriform spidroin	Cement for joints and attachments
Aggregate gland	Adhesive molecules	Aqueous coating of the capture spiral

(Gatesy et al., 2001). Different combinations of these modules forming a larger ensemble repeat have been suggested to be responsible for the distinct mechanical properties of the fibers produced by silk glands, on the basis of several studies (Scheibel, 2004; Vollrath, 2005; Hayashi et al., 1999; Fahnestock et al., 2000) A_n and $(GA)_n$ form crystalline β-sheets thought to provide tensile strength, $(GGX)_n$ probably forms a Gly-II helical structure, and $(GPGXX)_n$ forms β-spirals; the last may be involved in forming an amorphous matrix which would provide elasticity. More recently, spidroin genes were identified which were composed of a novel type of consensus assembly, with long and complicated repeats containing high levels of serine (Garb and Hayashi, 2005; Hu et al., 2005a,b; Tian and Lewis, 2005; Huang et al., 2006).

The N-terminal region analyzed from different spiders' silks has been shown to be the most conserved part of the silk proteins (Smith et al., 2005). In one instance it has additional Met codons downstream of the first one creating possible additional translation starts (Smith et al., 2005). Although the N-terminal function is related to the transport of the spidroins into the glandular lumen due the presence of a signal peptide (Hayashi and Lewis, 1998; Smith et al., 2005), the function of the mature N-terminal remains unknown. The C-terminal region of 100 amino acids is also a highly conserved domain among spidroins. Due to its high conservation this region was proposed to play an important role in fiber assembly (Jin and Kaplan, 2003; Spooner et al., 2005). Recently, it was demonstrated that the proteins' organization into a macromolecular fiber structure may depend on the C-terminal domain for the correct fiber density and orientation (Ittah et al., 2006).

The majority of the described spidroin sequences were generated from EST (expressed sequence tag) (Xu and Lewis, 1990; Hinman and Lewis, 1992; Hayashi et al., 2004; Tian and Lewis, 2005). Although there are limitations mainly using large sized mRNAs, EST has proven to be a rapid and economical technique for the identification of genes of biological interest (Adams et al., 1991). In this article we describe four partial cDNA sequences that encode spidroins produced by major ampullate, minor ampullate, flagelliform and tubuliform silk glands from the Brazilian spider *Nephilengys cruentata* (Araneae: Nephilidae). Aiming at possible fiber production using biotechnology approaches capable of mimicing the natural properties of the silks produced by spiders, our results contribute to understanding the effective mechanical properties of fiber silks by providing important information about their protein structure.

2. Materials and methods

2.1. RNA isolation

N. cruentata spiders were collected in the Atlantic Forest native regions (Brazil). The spider silk glands were isolated in the laboratory under a stereomicroscope and immediately frozen in liquid nitrogen. The isolated glands included the major ampullate, minor ampullate, tubuliform, flagelliform, aciniform, and pyriform glands. After pulverization, total RNA extraction was performed using TRIZOL reagent (Invitrogen, USA) following the manufacturer's recommendations. The yield, purity and integrity of total RNA were determined by measuring absorbance at 260/280 nm and by agarose gel electrophoresis. The Oligotex kit (Qiagen, Germany) was used for mRNA purification according to the technical manual and the concentration and purity of the isolated mRNA was also determined by measuring absorbance at 260/280 nm.

2.2. Construction of the cDNA Library

The cDNA library was made using the "SUPERSCRIPT II Plasmid System with GATEWAY Technology" (Invitrogen), according to the manufacturer's instructions. After *Escherichia coli* DH5α transformation using electro-transformation (Sambrook and Russell, 2001), the libraries were plated on selective media, the positive clones for cDNA insertion were picked and transferred to 96 well plates. Plasmid DNAs was used for sequencing after alkaline lyses plasmid DNA extraction (Sambrook and Russell, 2001) and quantification of templates. The sequencing reactions were performed using the Big Dye kit (Applied Biosystems, USA), following the manufacturer's instructions. The purified reactions were sequenced with an ABI 3700 DNA sequencer. The resulting chromatograms were directly transferred to a central data base similar to the one described by Telles et al. (2001) for processing and analysis.

2.3. Screening and gene identification

BLASTX (Altschul et al., 1997) was used to determine potential positive clones by searching for homologous repetitive and C-terminal sequences. Several combinations of restriction enzymes were used to check the size of the inserts. Clones with inserts bigger than 1.5 kb were treated with exonuclease III (Erase-a-Base kit, Promega, USA) and used in a nested deletion strategy for sequencing. The positive clones for MaSp1 and flagelliform cDNAs were then used as probes by random primer labeling with [α-^{32}P]dCTP in Southern Blots analysis (Sambrook and Russell, 2001) for identification of additional related sequences in the silk gland libraries.

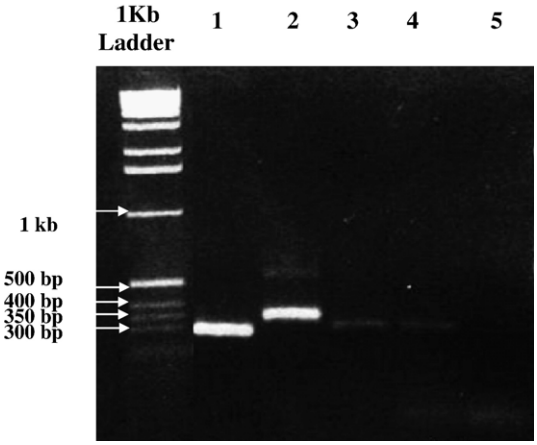

Fig. 1. RT-PCR was performed using total RNA from silk glands. 1, Flagelliform spidroin cDNA (287 bp); 2, MaSp cDNA (356 bp); 3, MiSp cDNA (304 bp); 4, TuSp cDNA (302 bp); 5, negative control. 1 Kb DNA ladder used was obtained from Invitrogen. The oligonucleotides used in the PCR are described in the Materials and methods.

2.4. RT-PCR

In order to verify the positives clones found in the library, silk gland total RNA was used to perform reverse transcription. Superscript II (Invitrogen, USA) was used in the reactions following the manufacturer's instructions. Polymerase chain reaction (PCR) analysis were conducted using Taq Polymerase (Invitrogen, USA) in the following conditions: initial template denaturation was set for 2 min at 94°, 35 repeated cycles were 30 s at 94°, 1 min at 56° and 30 s at 68°, the final extension at 72° for 10 min. Oligonucleotides used were designed according to the C-terminal sequence of the spider silk cDNA positive clones. The respective forward and reverse primers used for each spider silk cDNA were: Major ampullate spidroin foward-TGCAGGTCAAGGTGGATATGGTG, Major ampullate spidroin reverse-CCTAGAGCTTGGTAAACCGATTGAC, Minor ampullate spidroin forward-CTCTGCGGGTGTAGGTGTTGG, Minor ampullate spidroin reverse-TGCAGCAGACGAACCAACAGA, Flagelliform spidroin forward-GCATATGGAGCTGGTTCTGGTACAC, Flagelliform spidroin reverse-CTTCGAACAGAGTTGGAATATGCAT, Tubuliform spidroin foward-CATCTTCC GGCTTAGCATC and Tubuliform spidroin reverse-ACCAGCGAGAGCAGTTGC.

2.5. Sequence analysis and phylogenetic tree

Base calling and quality assignment of individual bases were done through the use of Phred (Ewing and Green, 1998; Ewing et al., 1998). Ribosomal poly(A) tails, low-quality sequences, vector and adapter regions were removed as described by Telles and da Silva (2001) with minor adaptations to this project. The resulting sets of cleaned sequences were assembled into clusters of overlapping sequences using the cap3 (Huang and Madan, 1999) or phrap (http://www.phrap.org/phredphrap/phrap.html) assembler, with individual base quality and default parameters. BLAST (Altschul et al., 1997) and FASTA (Pearson, 1990) were also used to analyze the sequences. Multiple sequence alignments were performed using ClustalW version 1.8 (Thompson et al., 1994) with default settings and refined by examination. Ensemble repeats within each spidroin were aligned, and a consensus repeat was generated for each translated protein. Codon usage analysis was made utilizing the software CodonUsage with standard genetic code (Stothard, 2000).

Phylogenetic analysis was conducted using the alignment of the amino acid C-terminal sequences performed in MAFFT (5.8), under the accuracy oriented E-INS-i algorithm (mafft-genafpair-maxiterate 1000 input_file>output_file), as implemented on the online server of Kyushu University, Japan (http://align.bmr.kyushu-u.ac.jp/mafft/online/server/) (Katoh et al., 2005). Phylogenetic analysis was performed on the computational cluster of the David Liberles laboratory, University of Wyoming. The topology was reconstructed using maximum likelihood as implemented in PHYML (Guindon and Gascuel, 2003), using the Blosum 62 model of protein evolution. Non-parametric bootstrap values were assessed with 1000 bootstrap replicates.

3. Results

3.1. cDNA library

In order to identify novel sequences encoding different spider silk proteins from the spider *N. cruentata*, 960 random clones from the silk gland cDNA library were partially sequenced and analyzed. Ninety-six more positive clones selected from another 24 96-well plates were also partially sequenced and analyzed after Southern Blot analysis. Of these, 21 clones were selected after BLAST searches according to their size and their amino acid translation by comparing their similarity with previously described spidroins. All selected clones in the library were partial transcripts as noted by 5′ end heterogeneity. Four clones containing the C-terminal region and repetitive sequence were classified as being MaSp1-like cDNAs and the largest clone for this silk group was 3241 bp. Three other clones were classified as MiSp1 cDNAs, where two of them contained only the repetitive sequence, and the largest one with 3486 bp contained both the repetitive sequence and the non-repetitive C-terminal. The most highly represented cDNA from the library encoded flagelliform spidroin, with nine partial cDNAs also containing the repetitive sequence and C-terminal, the longest of which was 3277 bp long. Although having the smallest cDNAs clones, tubuliform spidroin was also identified in the *N. cruentata* cDNA library with two clones of 1600 bp and 1636 bp. Three other clones containing novel repetitive sequences were selected for further study.

RT-PCR was performed in order to verify the presence of the spidroin cDNAs in *N. cruentata* silk glands. Total RNA from all silk glands combined was used as a template for polymerization of the cDNA first strand and primers designed from spidroin sequences found in the cDNA library were used to amplify the C-terminal region. All cDNAs analyzed were positive for their presence in the silk glands (Fig. 1), eliminating the possibility of transcript contamination or clone recombination.

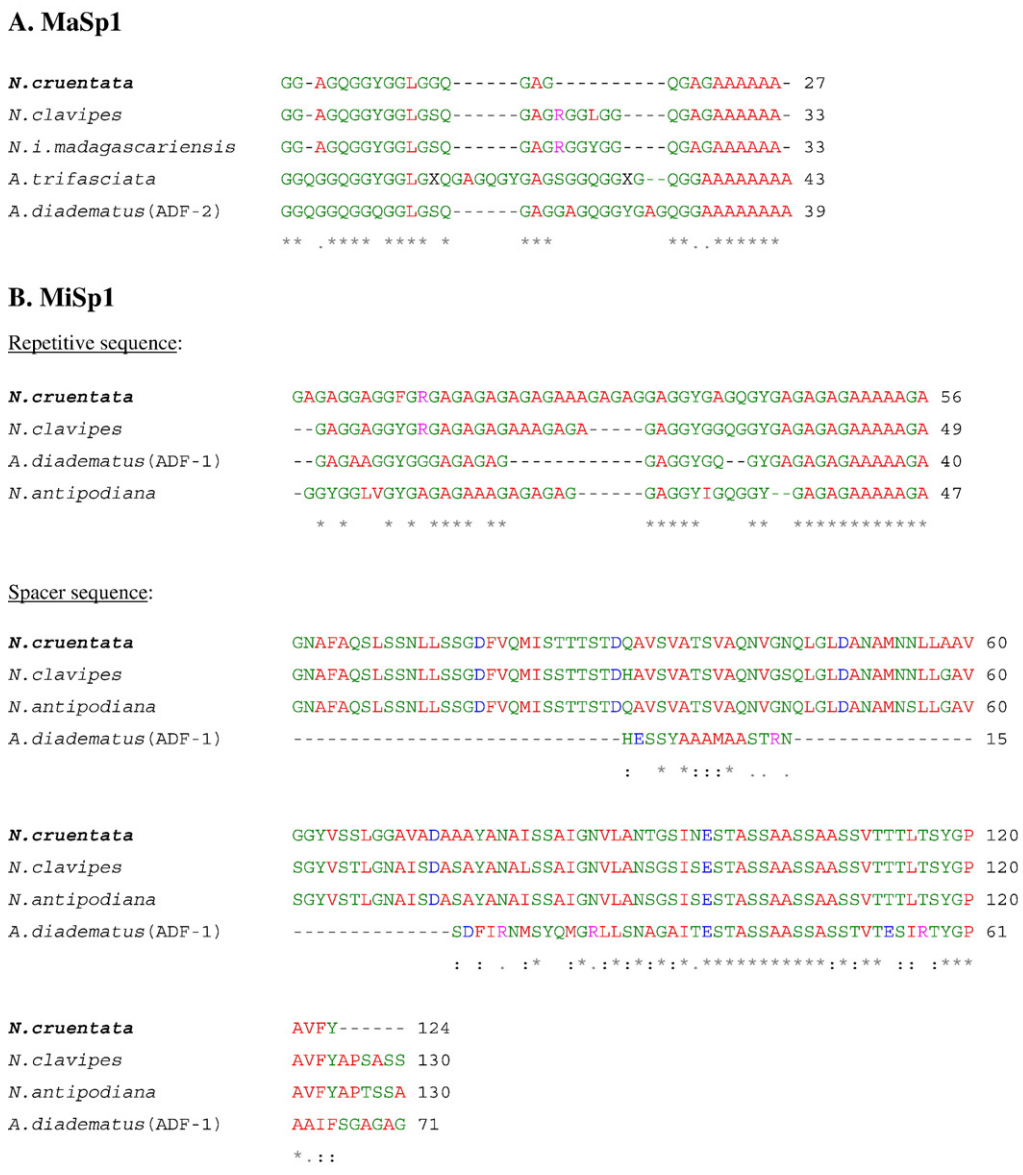

Fig. 2. ClustalW alignment of amino acid sequences of the consensus repetitive sequence of different orb-web weavers spider silk proteins. (A) Repetitive sequence of major ampullate spidroins. (B) Repetitive sequence and spacer of minor ampullate spidroins. Amino acids are indicated by one letter abbreviations. Hyphens indicate gaps introduced to obtain the best alignment. "*" means that the residues or nucleotides in that column are identical in all sequences in the alignment, ":" means that conserved substitutions have been observed, according to the color (red—Small aa (small+hydrophobic (incl. aromatic-Y)), blue—Acidic aa, magenta—Basic aa, green—Hydroxyl+Amine+Basic-Q), and "." means that semi-conserved substitutions are observed. Sequences from *N. cruentata* are identified in bold. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2. Sequence analysis

All selected cDNA clones had an open reading frame for a spider silk spidroin with a polyadenylation signal (data not shown). Following the nomenclature for published spider silk spidroins the identified genes were named as NCMaSp1-like (*N. cruentata* major ampullate spidroin-like), NCMiSp1-like (*N. cruentata* minor ampullate spidroin-like), NCFlag-like (*N. cruentata* flagelliform spidroin-like) and NCTuSp-like (*N. cruentata* tubuliform spidroin-like) according to their transcripts. *N. cruentata* spidroins

Fig. 3. Flagelliform spidroin schematic of consensus sequences of different orb-web weavers spiders. Hyphens indicate gaps introduced to obtain the best alignment. Sequence from *N. cruentata* is identified in bold.

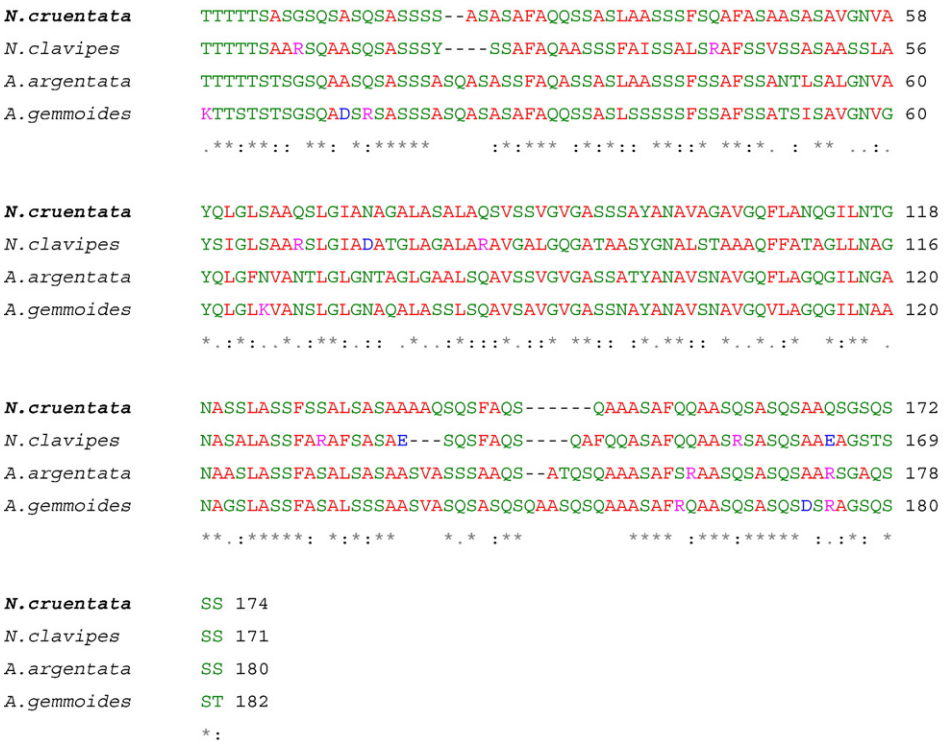

Fig. 4. ClustalW alignment of the consensus repetitive amino acid sequence of tubuliform proteins from different orb-web weaver spiders. Sequence alignment abbreviations are the same as in Fig. 2. Sequence from *N. cruentata* is identified in bold.

encode repetitive alanine and glycine rich proteins dominated by simple amino acid motifs. The NCMaSp1-like and NCMiSp1-like spidroins are highly repetitive protein sequences with the amino acids motifs $(GGX)_n$ and polyAla, this being replaced by $(GA)_n$ repeats in NCMiSp1-like protein (Fig. 2). We also identified a non-repetitive spacer between the repetitive sequences from NCMiSp1-like spidroin (Fig. 2B). NCFlag-like spidroin also contained a repetitive sequence with $(GPGGX)_n$ motifs separated by a non-repetitive sequence (Fig. 3). Unlike the previous *N. cruentata* silk proteins, NCTuSp-like spidroin shows few of the common spidroin amino acid motifs; instead the sequence contains repeats of 174 amino acids, largely composed by alanine and serine, forming new motifs such as S_n, $(SX)_n$ (X represents Q, L, A, V and F), and GX (X represents Q, N, I, L, A and V) (Fig. 4).

The codon usage for *N. cruentata* spidroins' most abundant amino acids (Table 2) follows the preference for adenine (A) and thymine (T) as the wobble base encoding each amino acid, with the exception of NCTuSp-like spidroin. NCMaSp1-like and NCMiSp1-like had a high content of glycine, alanine and glutamine in their amino acid sequences, all of them using A or T as the third nucleotide in their codon usage. The codon bias for glutamine and alanine was CAA and GCA/T respectively, with CAA presented in 100% of the cases in NCMaSp1-like and 96% in NCMiSp1-like. Glycine, the most prevalent amino acid

Table 2
Codon frequencies for *N. cruentata* silk proteins most abundant amino acids

Am Acid	Codon	Frequency (%)				Am Acid	Codon	Frequency (%)			
		FLAG	MiSp	MaSp	TuSp			FLAG	MiSp	MaSp	TuSp
Ala	GCG	5	1	1	4	Pro	CCG	2	0	0	10
Ala	GCA	22	24	48	30	Pro	CCA	27	33	50	20
Ala	GCT	72	61	33	33	Pro	CCT	60	33	50	50
Ala	GCC	1	14	18	34	Pro	CCC	11	33	0	20
Gln	CAG	25	4	0	50	Ser	AGT	4	19	21	15
Gln	CAA	75	96	100	50	Ser	AGC	2	4	11	11
Glu	GAG	31	0	0	0	Ser	TCG	4	2	11	11
Glu	GAA	69	100	100	100	Ser	TCA	30	12	13	11
Gly	GGG	1	2	0	0	Ser	TCT	49	48	32	31
Gly	GGA	47	45	52	46	Ser	TCC	11	15	13	21
Gly	GGT	42	45	41	31	Thr	ACG	0	0	0	0
Gly	GGC	10	9	8	23	Thr	ACA	91	20	0	31
Tyr	TAT	41	67	90	40	Thr	ACT	4	70	67	38
Tyr	TAC	59	33	10	60	Thr	ACC	4	10	33	31

in both proteins, showed a preference of 93% and 90% for GGA/T in NCMaSp1-like and NCMiSp1-like, in that order. Glycine and alanine were also present in large amounts in the NCFlag-like protein sequence. Their codon usage follows the same bias found for NCMaSp1-like and NCMiSp1-like cDNA sequences, with 89% for GGA/T and 94% for GCA/T. Compared with the high frequency of codon bias toward A and T at the wobble position in the sequences encoding NCMaSp1-like, NCMiSp1-like and NCFlag-like proteins, the codon usages for alanine, serine and glutamine, the three most

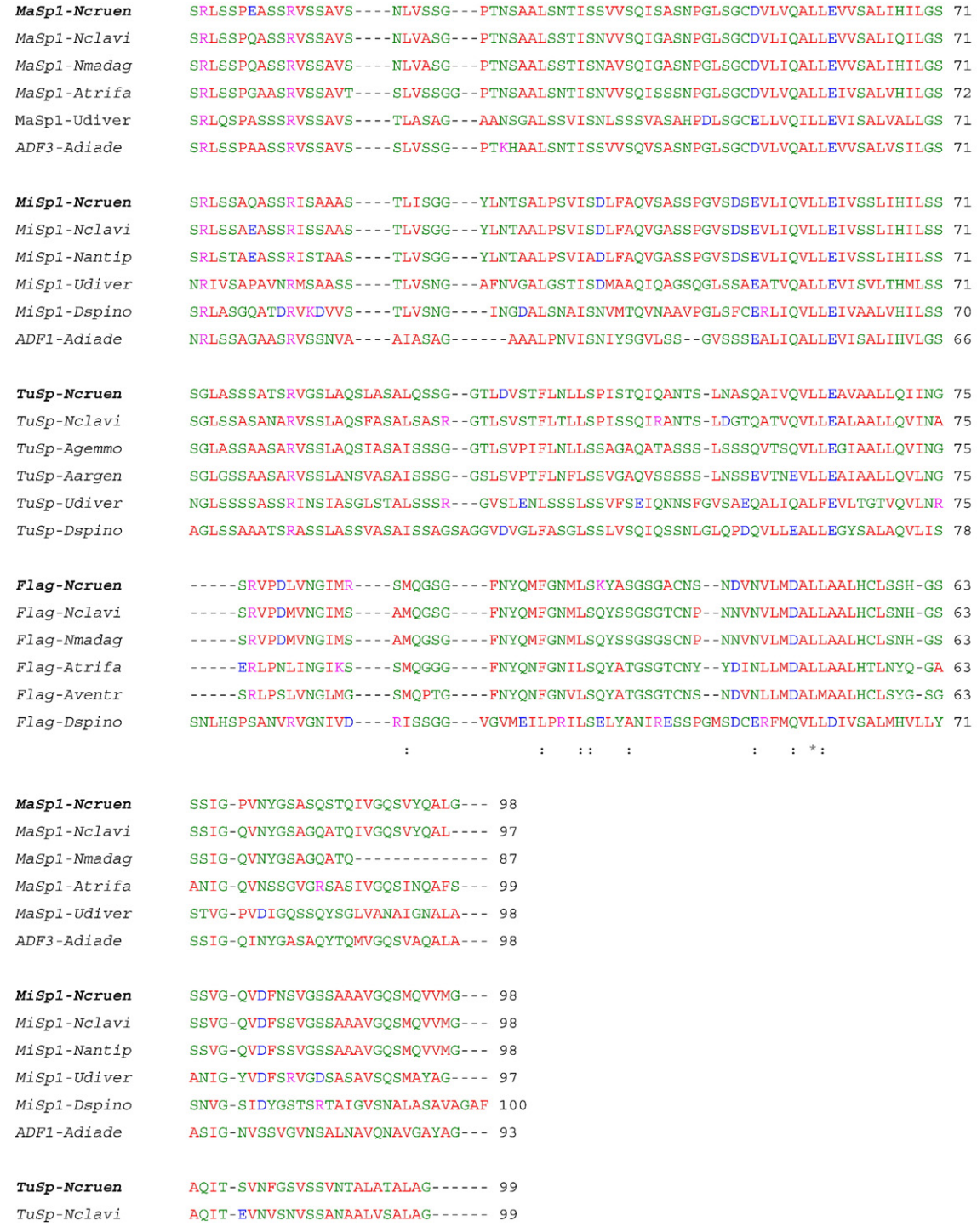

Fig. 5. ClustalW alignment of C-terminal amino acid sequences. Amino acids are indicated by one letter abbreviations and numbered from N- to C-terminal. Sequences alignment abbreviations are the same as in Fig. 2. Sequences from *Nephilengys cruentata* are identified in bold. *Araneus diadematus* ADF-3 was added as a representative of MaSp2 group. Abbreviations of spider species used in this figure (from top to bottom): Ncruen, *N. cruentata*; Nclavi, *Nephila clavipes*; Nmadag, *Nephila inaurata madagascariensis*; Atrifa, *Argiope trifasciata*; Udiver, *Uloborus diversus*; Adiade, *A. diadematus*; Nantip, *Nephila antipodiana*; Agemmo, *Araneus gemmoides*; Aargen, *Argiope argentata*; Dspino, *Deinopis spinosa*; Aventr, *Araneus ventricosus*. Abbreviations used for silk spidroins: MaSp1, major ampullate spidroin 1; ADF3, fibroin 3 (major ampullate spidroin 2); MiSp1, minor ampullate spidroin 1; ADF1, fibroin 1; TuSp, tubuliform spidroin; Flag, flagelliform spidroin.

```
TuSp-Agemmo      AQIR-SVNLANVPNVQQALVSALSG------ 99
TuSp-Aargen      AQIT-SVNLRNV------------------- 86
TuSp-Udiver      GQTS-FVSVSSPTVISSSF------------ 93
TuSp-Dspino      SQIS-SVSVSSSSALGPALLNYLVG------ 102

Flag-Ncruen      PSFGSSPTPSAMNAYSNSVRRMFQF------ 87
Flag-Nclavi      SSFAPSPTPAAMSAYSNSVGRMFAY------ 87
Flag-Nmadag      SSFAPSPTPAAMSAYSNSVGRMFAY------ 87
Flag-Atrifa      SYVPSYPSPSEMLSYTENVRRYF-------- 85
Flag-Aventr      S-VPSTPTYSAMSAYNQSIRRMFTY------ 86
Flag-Dspino      EDVRRGIPGDTAEAVANAVAGVVLSVV---- 98
```

Fig. 5 (*continued*).

frequent amino acid in the NCTuSp-like protein, are only moderately biased toward A and T, with 63% for alanine, 57% for serine and 50% for glutamine.

3.3. Phylogenetic analysis

Maximum likelihood (ML) analysis examined the relationship between the C-terminal amino acid sequences from *N. cruentata* spidroins and previously reported spidroin genes from different spider silk glands and species. Alignment of the C-terminal region from all identified spidroins with known sequences from other silk proteins generated using ClustalW is shown in Fig. 5. In order to perform the phylogenetic analysis an alignment of the amino acid C-terminal sequences was also performed in MAFFT (5.8) (data not shown). The topology resulting from the ML analysis (-lnL: 16456.89) is depicted in Fig. 6, and rendered as an unrooted tree. The results showed that the spidroins found in the *N. cruentata* library belong to the spidroin gene family, and all of them were correctly classified according to their ortholog group.

4. Discussion

For many years spider silks have piqued the interest of mankind because of their extreme mechanical properties. With advances in biotechnology, the possibility arose of producing new materials based on spider silk polymers. In this work we identified different silk genes from the spider *N. cruentata*. Using cDNAs from the spider silk glands we were able to identify partial transcripts from major ampullate, minor ampullate, flagelliform and tubuliform glands. Although this strategy has been shown to have its limitations in obtaining full spider silk transcripts, it is commonly used to identify novel spider silk genes (Hayashi et al.,

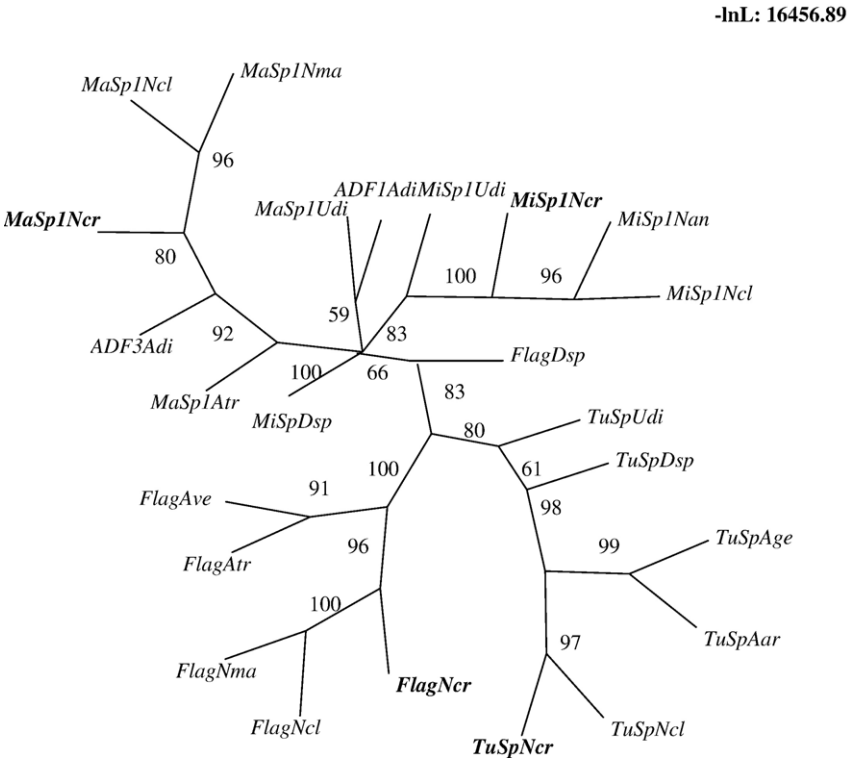

Fig. 6. Unrooted tree from spider silk C-terminal sequence phylogenetic analysis. *Nephilengys cruentata* spidroins are identified in bold. Abbreviations of spider species used in this figure: Ncr, *N. cruentata*; Ncla, *Nephila clavipes*; Nma, *Nephila inaurata madagascariensis*; Atr, *Argiope trifasciata*; Udi, *Uloborus diversus*; Adi, *Araneus diadematus*; Nan, *Nephila antipodiana*; Age, *Araneus gemmoides*; Aar, *Argiope argentata*; Dsp, *Deinopis spinosa*; Ave, *Araneus ventricosus*. Abbreviations used for silk spidroins: MaSp1, major ampullate spidroin 1; ADF3, fibroin 3; MiSp1, minor ampullate spidroin 1; ADF1, fibroin 1; TuSp, tubuliform spidroin; Flag, flagelliform spidroin. GenBank accessions: *MaSp1Ncl*—P19837, *MaSp1Nma*—AAK30606, *MaSp1Atr*—AAK30595, *MaSp1Udi*—ABD61596, *ADF3Adi*—AAC47010, *MiSp1Ncl*—AAC14589, *MiSp1Nan*—ABC72645, *MiSp1Udi*—ABD61597, *MiSp1Dsp*—ABD61589, *ADF1Adi*—AAC47008, *TuSpNcl*—AAX45295, *TuSpAge*—AAX45293, *TuSpAar*—AAY28932, *TuSpUdi*—AAY28933, *TuSpDsp*—AAY28934, *FlagNcl*—AAC38847, *FlagNma*—AAF36092, *FlagAtr*—AAK30594, *FlagAve*—AAT36347 and *FlagDsp*—ABD61590.

2004; Lawrence et al., 2004; Pouchkina-Stantcheva and McQueen-Mason, 2004; Tian et al., 2004).

In agreement with previously described sequences encoding spider silk proteins (Xu and Lewis, 1990; Hinman and Lewis, 1992), the codon usage for *N. cruentata* spidroins' most abundant amino acids, with the exception of NCTuSp-like spidroin, follows the preference for adenine (A) and thymine (T) as the third base of three encoding each amino acid. This difference in the biased codon usages were also found for *Araneus gemmoides* and *Nephila clavipes* TuSp1 coding sequences, although they have a higher amount of glycine instead of glutamine in their amino acid sequences than found in NCTuSp-like spidroin (Tian and Lewis, 2005). The prevalence of T and A in major ampullate, minor ampullate and flagelliform coding sequences may prevent the formation of numerous hairpin loops in nearby G/C-rich regions present in the codons of their repetitive sequences (polyAla, $(GGX)_n$, $(GPGGX)_n$) (Hinman and Lewis, 1992). Such regions do not exist in tubulifom proteins.

Like previously described spider silk proteins (Xu and Lewis, 1990; Gatesy et al., 2001; Tian et al., 2004; Lewis, 2006), our findings also demonstrated proteins with high content of alanine and glycine amino acids due to the presence of highly repetitive motifs. NCMaSp1-like protein is composed of similar repeats with polyAla and $(GGX)_n$ motifs, where the X residues are A, Y, L or Q. Its ensemble repeats are very similar to the ones found in *N. clavipes* and *Nephila inaurata madagascariensis* proteins with a 96% identity. As all of these spiders belong to the Nephilidae family, similarity is not entirely unexpected. However *N. cruentata* repeats vary in length lacking GGX motifs. Those motifs were proposed to adopt a Gly-II helix forming an amorphous matrix that connects crystalline regions and provides elasticity. It was also found using FTIR microscopy that the secondary structure of major ampullate spidroins has predominantly (38%) helical structures in comparison with sheets and turns (Winkler and Kaplan, 2000; Van Beek et al., 2002; Dicko et al., 2004). On the other hand this helix is too rigid to be elastic; it is more likely that the $(GGX)_n$ motifs are responsible for another interprotein link similar to the β-sheet. Although major silks present a high content of $(GGX)_n$ motifs, they are known for their high tensile strengths and toughness provided by the polyAla stretches, with strength values in the range of 1–2 GPa (Gosline et al., 1999, 2002; Blackledge and Hayashi, 2006). In contrast, minor ampullate silk has decreased strength but increased extensibility in comparison with major silks, which can be related to the presence of fewer polyAla stretches and higher numbers of $(GA)_n$ and $(GGX)_n$ motifs in the minor protein ensemble repeats (Blackledge and Hayashi, 2006). Those motifs were also found in NCMiSp1-like spidroin sequence. Although consensus repeats showed very similar organization with previously described minor ampullate spidroins, the number of $(GA)_n$ repeats varies in comparison between those proteins. However, a highly conserved non-repetitive 124 amino acid sequence (spacer) was found interrupting the repetitive regions of NCMiSp1-like spidroin and it is almost identical with the spacer from *N. clavipes* and *N. i. madagascariensis* minor ampullate spidroins with few amino acid substitutions and deletions (89% and 91% identity respectively). The main function of this region remains unknown, but it may serve to separate crystalline regions as well as participate in inter-chain protein associations through charged residues (Lewis, 2006).

NCFlag-like protein is also composed of highly repetitive sequence formed by the $(GPGGX)_n$ (X represents A, V, S and Y) and three GGX (X represents A, S and T) motifs separated by a non-repetitive sequence with many charged and hydrophilic amino acids. This sequence is also very similar to previously described flagelliform spidroins (Hayashi and Lewis, 2000). Flagelliform silk, the most extensible silk (Blackledge and Hayashi, 2006), is responsible for the formation of the capture spiral in the orb-web. The GPGGX motif has been suggested to conform to a β-spiral secondary structure similar to a spring that could easily contribute to the elastic mechanism of the fiber, with the proline bonds generating force for retraction of the silk after stretching (Hayashi et al., 1999).

Tubuliform and aciniform silks are unique among spider silk spidroins due to their complex amino acid composition and NCTuSp-like protein was similar. It is composed of large repeats rich in alanine and serine amino acids with several new motifs (S_n, $(SX)_n$, and GX). In contrast to other silk spidroins, NCTuSp-like spidroin contains high amounts of serine and low amounts of glycine. In addition it also showed more amino acids with large side chains such as valine and leucine. Secondary structure predictions of tubuliform spidroins using X-ray diffraction indicated the presence of large amounts of β-sheet. It also showed that eggcase silk has a larger *b* dimensional value in the β-sheet than major and minor ampullate silks, which indicates the presence of large-side-chain amino acids (Parke et al., 1997). TEM (transmission electron microscopy) diffraction data also agrees with that obtained from X-ray diffraction (Barghout et al., 1999). The lack of the usual repeats rich in alanine and/or glycine motifs found in most spiders spidroins is probably related to function. Tubuliform spidroins are used to construct structures for reproduction different from major or minor ampullate or flagelliform silks that serve as structural fibers for prey capture. However, recent mechanical data from *Argiope argentata* tubuliform silks showed that these fibers perform quite well in comparison with structural ones, with a 0.47 GPa strength value (Blackledge and Hayashi, 2006).

Alignment of the repeat units from NCTuSp-like with tubuliform repeats from different spider species showed a high similarity between them. Surprisingly the *N. cruentata* repeat unit sequence was more similar to *A. gemmoides* and *A. argentata* unit repeats than to *N. clavipes* repeats, even though they belong to different families according to morphological evidence. NCtuSp-like sidroin also showed sequence similarities with fibroin 1 from the mygalomorphae spider *Euagrus chisoseus* (AF350271). Previously described tubuliform spidroins exhibited a remarkable degree of homogeneity between consecutive intragenic repeats (Garb and Hayashi, 2005; Tian and Lewis, 2005). Unfortunately we were unable to verify this evidence of concerted evolution in NCtuSp1-like spidroin because of the small size of the obtained transcript in the cDNA library which showed only one complete repeat.

The nonrepetitive and highly conserved C-terminal region was also identified in all described *N. cruentata* spidroins. Alignment of the C-terminal amino acid sequence described for these spidroins with those of the known fibroin genes showed great sequence conservation, with the more highly conserved sequences among ortholog gene groups than among paralogous genes. Most spider silk proteins share 30% identity among their C-terminal, with flagelliform spidroins the most highly diverged. However all sequences share a particular conserved region at the QALLE amino acid sequence even between orb-web and non-orb-web weavers spiders. This region corresponds to the region with higher hydrophobicity in the C-terminal, and secondary structure predictions suggest that this QALLE region is also responsible for the formation of α-helices (Spooner et al., 2005; Challis et al., 2006). Since the C-terminal sequence is much more conserved than the repetitive region among different spiders' species, it is suggested that it might play an important role with certain amino acid motifs being preserved by selection against mutations that could disrupt the C-terminal function during evolution. Several functions have been attributed to the C-terminal from silk proteins. It might be responsible for the formation of irregular sized micelles in the spinning dope in order to prevent premature fiber formation (Jin and Kaplan, 2003) or have a function in correct protein folding since it has been demonstrated that the C-terminal is retained in the fiber (Sponner et al., 2004; Ittah et al., 2006).

Because of its high sequence conservation, C-terminal region alignment was used for phylogenetic analysis. The result shows that different ortholog genes cluster together rather than by species, agreeing with previously phylogenetic analysis of the spidroin gene family (Garb and Hayashi, 2005; Challis et al., 2006; Garb et al., 2006); all identified *N. cruentata* spidroins grouped together in their specific gene order. The observation that silk proteins cluster according to their type suggests that their evolution may be due to a divergent evolution involving a common ancestor, rather than by a model of concerted evolution, where one would expect genes to cluster within species and not within the same gene order (Challis et al., 2006; Garb et al., 2006). Based on fossil evidence, a possible common ancestor could be the extinct *Macryphants*, dating from the lower cretaceous at least 136 million years ago (Selden, 1989).

In summary, we have studied different silk spidroins from the Brazilian spider *N. cruentata*. We were able to identify the protein sequences from silks responsible for orb-web building and prey capture (MaSp, MiSp and Flag), as well as for the construction of the eggcase (TuSp). Throughout our study it was possible to demonstrate a high degree of similarity between these sequences and sequences from other spider species. It was also observed similarities among sequences from different gene groups with the exception of tubuliform spidroin that present a complete different protein structure, and which play a distinct function in the spiders' life (Hu et al., 2005a; Tian and Lewis, 2005). These results support the argument that mechanical properties are correlated to their protein sequences (Hayashi et al., 1999; Rising et al., 2005). Further mechanical and structural study of *N. cruentata* spidroins should also provide more evidence for this correlation.

Acknowledgements

We would like to thank Dr Michael Hinman and Dr David Perry (University of Wyoming) for reviewing this work and Kelly Martins de Brito for technical assistance. This research was funded by CNPq (Conselho Nacional de Desenvolvimento Cientifico e Tecnologico-Brazil), grant: 486492/2006-0.

References

Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. Science 252, 1651–1656.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Barghout, J.Y.T., Thiel, B.L., Viney, C., 1999. Spider *Araneus diadematus*. cocoon silk: a case of non-periodic lattice crystals with a twist? Int. J. Biol. Macromol. 24, 211–217.

Blackledge, T.A., Hayashi, C.Y., 2006. Silken toolkits: biomechanics of silk fibers spun by the orb web spider *Argiope argentata* (Fabricius 1775). J. Exp. Biol. 209, 2452–2461.

Challis, R.J., Goodacre, S.L., Hewitt, G.M., 2006. Evolution of spider silks: conservation and diversification of the C-terminal. Insect. Mol. Biol. 15, 45–56.

Dicko, C., Knight, D., Kenney, J.M., Vollrath, F., 2004. Secondary structures and conformational changes in flagelliform, cylindrical, major, and minor ampullate silk proteins. Temperature and concentration effects. Biomacromolecules 5, 2105–2115.

Ewing, B., Green, P., 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 8, 186–194.

Ewing, B., Hillier, L., Wendl, M.C., Green, P., 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 8, 175–185.

Fahnestock, S.R., Yao, Z., Bedzyk, L.A., 2000. Microbial production of spider silk proteins. J. Biotechnol. 74, 105–119.

Foelix, R.F., 1996. Biology of Spiders 1996 Oxford University Press Inc. and Georg Thieme Verlag New York.

Garb, J.E., Hayashi, C.Y., 2005. Modular evolution of egg case silk genes across orb-weaving spider superfamilies. Proc. Natl. Acad. Sci. U. S. A. 102, 11379–11384.

Garb, J.E., DiMauro, T., Vo, V., Hayashi, C.Y., 2006. Silk genes support the single origin of orb webs. Science 312, 1762.

Gatesy, J., Hayashi, C., Motriuk, D., Woods, J., Lewis, R.V., 2001. Extreme diversity, conservation, and convergence of spider silk fibroin sequences. Science 291, 2603–2605.

Gosline, J.M., Guerette, P.A., Ortlepp, C.S., Savage, K.N., 1999. The mechanical design of spider silks: from fibroin sequence to mechanical function. J. Exp. Biol. 202, 3295–3303.

Gosline, J., Lillie, M., Carrington, E., Guerette, P., Ortlepp, C., Savage, K., 2002. Elastic proteins: biological roles and mechanical properties. Philos. Trans. R. Soc. Lond., B Biol. Sci. 357, 121–132.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.

Hayashi, C.Y., Lewis, R.V., 1998. Evidence from flagelliform silk cDNA for the structural basis of elasticity and modular nature of spider silks. J. Mol. Biol. 275, 773–784.

Hayashi, C.Y., Lewis, R.V., 2000. Molecular architecture and evolution of a modular spider silk protein gene. Science 287, 1477–1479.

Hayashi, C.Y., Shipley, N.H., Lewis, R.V., 1999. Hypotheses that correlate the sequence, structure, and mechanical properties of spider silk proteins. Int. J. Biol. Macromol. 24, 271–275.

Hayashi, C.Y., Blackledge, T.A., Lewis, R.V., 2004. 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. 21, 1950–1959.

Hijirida, D.H., Do, K.G., Michael, C., Wong, S., Zax, D., Jelinkis, L.W., 1996. C13 NMR of *Nephila clavipes* major ampullate silk gland. Biophys. J. 71, 3442–3447.

Hinman, M.B., Lewis, R.V., 1992. Isolation of a clone encoding a second dragline silk protein. J. Biol. Chem. 267, 19320–19324.

Hu, X., Kohler, K., Falick, A.M., Moore, A.M., Jones, P.R., Sparkman, O.D., Vierra, C., 2005a. Egg case protein-1. A new class of silk proteins with fibroin-like properties from the spider *Latrodectus hesperus*. J. Biol. Chem. 280, 21220–21230.

Hu, X., Lawrence, B., Kohler, K., Falick, A.M., Moore, A.M., McMullen, E., Jones, P.R., Vierra, C., 2005b. Araneoid egg case silk: a fibroin with novel ensemble repeat units from the black widow spider, *Latrodectus hesperus*. Biochemistry 44, 10020–10027.

Huang, X., Madan, A., 1999. CAP3: A DNA sequence assembly program. Genome Res. 9, 868–877.

Huang, W., Lin, Z., Sin, Y.M., Li, D., Gong, Z., Yang, D., 2006. Characterization and expression of a cDNA encoding a tubuliform silk protein of the golden web spider *Nephila antipodoana*. Biochimie 88, 849–858.

Ittah, S., Cohen, S., Garty, S., Cohn, D., Gat, U., 2006. An essential role for the C-terminal domain of a dragline spider silk protein in directing fiber formation. Biomacromolecules 7, 1790–1795.

Jin, H.J., Kaplan, D.L., 2003. Mechanism of silk processing in insects and spiders. Nature 424, 1057–1061.

Katoh, K., Kuma, K., Toh, H., Miyata, T., 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33, 511–518.

Lawrence, B.A., Vierra, C.A., Moore, A.M., 2004. Molecular and mechanical properties of major ampullate silk of the black widow spider, *Latrodectus hesperus*. Biomacromolecules 5, 689–695.

Lewis, R.V., 2006. Spider silk: ancient ideas for new biomaterial. Chem. Rev. 106, 3762–3734.

Parke, A.D., Seeley, S.K., Gardner, K., Thompson, L., Lewis, R.V., 1997. Structural studies of spider silk proteins in the fiber. J. Mol. Recognit. 10, 1–6.

Pearson, W.R., 1990. Rapid and sensitive sequence comparison with FASTP and FASTA methods. Enzymology 183, 63–98.

Pouchkina-Stantcheva, N.N., McQueen-Mason, S.J., 2004. Molecular studies of a novel dragline silk from a nursery web spider, *Euprosthenops* sp. Pisauridae. Comp. Biochem. Physiol. Part B Biochem. Mol. Biol. 138, 371–376.

Rising, A., Nimervoll, H., Grip, S., Fernandez-Arias, A., Storckenfeldt, E., Knight, D., Vollrath, F., Engstron, W., 2005. Spider silk proteins—mechanical property and gene sequence. Zoolog. Sci. 22, 273–281.

Sambrook, J., Russell, D.W., 2001. Molecular Cloning: A Laboratory Manual, 3th ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Scheibel, T., 2004. Spider silks: recombinant synthesis, assembly, spinning, and engineering of synthetic proteins. Microb. Cell Fact. 3, 14.

Selden, P.A., 1989. Orb weaving spiders in the early Cretaceous. Nature 340, 711–713.

Smith, D.M., Smith, A., Hayashi, C.Y., Lewis, R.V., 2005. Analysis of the conserved N-terminal domains in major ampullate spider silk proteins. Biomacromolecules 6, 3152–3159.

Sponner, A., Unger, E., Grosse, F., Weisshart, K., 2004. Conserved C-termini of spidroins are secreted by the major ampullate glands and retained in the silk thread. Biomacromolecules 5, 840–845.

Spooner, A., Vater, W., Rommerskirch, W., Vollrath, F., Unger, E., Grosse, F., Weisshart, K., 2005. The conserved C-termini contribute to the properties of spider silk fibroins. Biochem. Biophys. Res. Commun. 338, 897–902.

Stothard, P., 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques 28, 1102–1104.

Telles, G.P., da Silva, F.L., 2001. Trimming and clustering sugarcane ESTs. Genet. Mol. Biol. 24, 17–23.

Telles, G.P., Braga, M.D.V., Dias, Z., Lin, T.Z., Quitzau, J.A.A., da Silva, F.R., Meidanis, J., 2001. Bioinformatics of the sugarcane EST project. Genet. Mol. Biol. 24, 8–15.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Tian, M., Lewis, R.V., 2005. Molecular characterization and evolutionary study of spider tubuliform eggcase silk protein. Biochemistry 44, 8006–8012.

Tian, M., Liu, C., Lewis, R.V., 2004. Analysis of major ampullate silk cDNAs from two non-orb-weaving spiders. Biomacromolecules 5, 657–660.

van Beek, J.D., Hess, S., Vollrath, F., Meier, B.H., 2002. The molecular structure of spider dragline silk: folding and orientation of the protein backbone. Proc. Natl. Acad. Sci. U. S. A. 99, 10266–10271.

Vollrath, F., 1992. Spider web and silks. Sci. Am. 266, 70–76.

Vollrath, F., 2005. Spider's web. Curr. Biol. 15, R364–R365.

Vollrath, F., Knight, D., 2001. Liquid crystalline spinning of spider silk. Nature 410, 541–548.

Winkler, S., Kaplan, D.L., 2000. Molecular biology of spider silk. J. Biotechnol. 74, 85–93.

Xu, M., Lewis, R.V., 1990. Structure of a protein superfiber: spider dragline silk. Proc. Natl. Acad. Sci. U. S. A. 87, 7120–7124.