



UNIVERSIDADE DE BRASÍLIA  
INSTITUTO DE BIOLOGIA  
DEPARTAMENTO DE BIOLOGIA CELULAR  
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA MOLECULAR

# Caracterização molecular de proteínas de sedas de aranhas da biodiversidade brasileira

Por

Daniela Matias de Carvalho Bittencourt

Tese apresentada ao Departamento de Biologia  
Celular curso de pós-graduação em Biologia  
Molecular como parte do requisito à obtenção do  
título de DOUTOR EM BIOLOGIA  
MOLECULAR

Orientador: Prof. Dr. Elíbio Leopoldo Rech Filho

Brasília,

Julho de 2007



*“Ao meu querido esposo, Flávio  
Bittencourt, por fazer dos meus  
sonhos também os seus.”*

## AGRADECIMENTOS

Primeiramente eu gostaria de agradecer a minha família pelo apoio constante durante toda a minha formação.

Ao meu orientador Dr. Elíbio Rech, que foi o responsável pelo o meu sucesso, pelos conselhos e incentivos.

Aos meus colegas do Laboratório de Transferência de Genes da Embrapa Recursos Genéticos e Biotecnologia, em especial Betúlia, Paula e Natália.

Ao Dr. Randolph Lewis e sua equipe da Universidade do Wyoming, por me receberem de portas abertas em seu laboratório.

Ao IBAMA e CEGEN por autorizarem a coleta do material biológico para o desenvolvimento desta pesquisa.

Ao Instituto Butantan, pelo fornecimento de parte do material genético estudado.

Ao Departamento de Biologia Celular da Universidade de Brasília pela oportunidade de desenvolver este trabalho.

Ao Conselho Nacional de Pesquisa (CNPq), pelo suporte financeiro.

À Embrapa Recursos Genéticos e Biotecnologia pela disponibilidade das instalações físicas, equipamentos e pela capacitação de recursos humanos.

Enfim, a todos aqueles que de uma maneira ou de outra, estiveram presentes e participaram desta grande conquista, o meu muito obrigada.

## ÍNDICE

Índice	ii
Índice de figuras e tabelas	v
Resumo	vii
Abstract	viii
<b>Capítulo 1 – Introdução</b>	1
Relacionamentos evolutivos	2
Arquitetura molecular	5
Tipos de seda	6
Ampolada principal	6
Ampolada secundária	9
Tubuliforme	9
Aciniforme	10
Flageliforme	10
Agregata	11
Piriforme	11
Biologia da seda	14
Propriedades mecânicas	18
Produção de seda sintética	22
<b>Capítulo 2 - Espidroínas da aranha <i>Nephilengys cruentata</i></b>	25
Introdução	26
Resultados	28
Biblioteca de cDNA	28
Análise das seqüências	29
Frequência de codons	31
Análise filogenética	33
Discussão	37
<b>Capítulo 3 - Propriedades Mecânicas da fibra da glândula Ampolada principal de <i>N. cruentata</i></b>	42

Introdução	43
Resultados	44
Diâmetro da fibra	44
Testes mecânicos	45
Microscopia de força atômica	46
Discussão	48
<b>Capítulo 4 - Espidroínas da aranha <i>Avicularia juruensis</i></b>	51
Introdução	52
Resultados	53
Biblioteca de cDNA	53
Análise das seqüências	54
Frequência de codons	57
Discussão	57
<b>Capítulo 5 - Relacionamento evolutivo da MaSp entre as diferentes espécies de aranhas</b>	60
Introdução	61
Resultados e discussão	62
<b>Referências</b>	67
<b>Anexo 1 - Material e métodos</b>	78
Extração de RNA	78
Construção das bibliotecas de cDNA	78
Seleção e análise de seqüências	79
RT-PCR	80
Árvore filogenética de <i>N. cruentata</i>	80
Coleta da fibra ampolada principal de <i>N. cruentata</i>	81
Medida do diâmetro da fibra	81
Testes mecânicos	81
Microscopia de força atômica	82
Análise filogenética de <i>A. juruensis</i> , mapeamento de caracteres e reconciliação da árvore	82
<b>Anexo 2</b>	85
Seqüências de cDNA das espidroínas de <i>N. cruentata</i>	85
Seqüências de cDNA das espidroínas de <i>A. juruensis</i>	96

<b>Anexo 3</b> – Artigo: Spidroins from the Brazilian spider <i>Nephilengys cruentata</i> (Araneae: Nephilidae)	105
<b>Anexo 4</b> – Artigo: How old are major ampullate silks?	116

## ÍNDICE DE FIGURAS E TABELAS

### Capítulo 1

Figura 1 - Representação da relação filogenética de Araneae	3
Figura 2 - Glândulas de sedas de aranha	4
Figura 3 - Desenho esquemático da microestrutura da fibra ampolada principal	8
Figura 4 - Representação artística da estrutura glandular geral	14
Figura 5 - Histologia da glândula sericígena	17
Tabela 1 - Motivos de aminoácidos encontrados em 3 diferentes tipos de espidroínas e sua estrutura adotada	6
Tabela 2 - Comparação entre diferentes propriedades mecânicas de sedas de aranha, seda do bicho-da-seda, fibras naturais e polímeros sintéticos	7
Tabela 3 - Comparação do teor de aminoácidos, do desempenho mecânico e da função ecológica de cinco sedas identificadas até o momento	13
Tabela 4 - Compilação de alguns atributos materiais, propriedade correspondente e unidades de medida	20

### Capítulo 2

Figura 1 - RT-PCR com RNA total da(s) glândula(s) sericígena(s) de <i>N.cruentata</i>	29
Figura 2 - Alinhamento das regiões repetitivas de duas espidroínas de <i>N. cruentata</i> com espidroínas de diferentes espécies de aranhas	30
Figura 3 - Esquema da região repetitiva de espidroínas da glândula Flagelliforme de diferentes espécies de aranhas	31
Figura 4 - Alinhamento da região repetitiva da espidroína da glândula Tubuliforme de <i>N. cruentata</i> com as de diferentes espécies de aranhas	32
Figura 5 - Alinhamento da região C-terminal das espidroínas de <i>N. cruentata</i> com espidroínas de diferentes espécies de	35

aranhas	
Figura 6 - Árvore sem raiz da topologia resultante da análise de ML (- lnL: 16456.89) das seqüências da região C-terminal de várias espidroínas	36
Tabela 1 - Escolha de codons para os aminoácidos mais freqüentes nas espidroínas de <i>N. cruentata</i>	33
<b>Capítulo 3</b>	
Figura 1 - Desempenho mecânico da fibra ampolada principal de <i>N. Cruentata</i>	46
Figura 2 - Imagens obtidas da fibra ampolada principal de <i>N. cruentata</i> por meio da microscopia de força atômica	47
Tabela 1 - Resultados dos testes mecânicos das fibras de duas aranhas <i>N. cruentata</i>	44
Tabela 2 - Propriedades mecânicas (valores de engenharia) da seda ampolada principal de diferentes espécies de aranha	45
<b>Capítulo 4</b>	
Figura 1 - RT-PCR com RNA total da(s) glândula(s) sericígena(s) de <i>A. juruensis</i>	54
Figura 2 - Região repetitiva das Espidroínas 1 e 2 expressas pela <i>A. juruensis</i>	55
Figura 3 - Alinhamento da região C-terminal das três traduções supostamente parálogas para o gene da <i>Espidroína 1</i> (1A, 1B e 1C) de <i>A. juruensis</i>	55
Figura 4 - Alinhamento da região C-terminal da Espidroína 2 de <i>A. juruensis</i> com proteínas MaSp2 de diferentes espécies de aranhas	56
Tabela 1 - Escolha de codons para os aminoácidos mais freqüentes nas espidroínas de <i>A. juruensis</i>	57
<b>Capítulo 5</b>	
Figura 1 – A) Relacionamentos filogenéticos entre as espidroínas de <i>A. juruensis</i> com diferentes seqüências de espidroínas publicamente disponíveis. B) Árvore filogenética de aranha com dados da ML de Ayoub <i>et al.</i> , 2007.	64

## RESUMO

As aranhas produzem até seis tipos diferentes de seda, cada um para uma função biológica específica. As sedas de aranhas também são conhecidas por suas exclusivas propriedades mecânicas. A possibilidade de produzir novos materiais com propriedades semelhantes motivou a pesquisa sobre essas proteínas da seda (espidroínas). Neste trabalho foram identificados diferentes espidroínas produzidas pelas glândulas sericígenas das aranhas *Nephilengys cruentata*, produtora de teias em orbital, e *Avicularia juruensis*, da família das caranguejeiras, aranhas rudimentares possuidoras de apenas uma ou duas glândulas produtoras de seda; ambas encontradas na fauna brasileira. As seqüências das espidroínas de *N. cruentata* mostraram uma considerável semelhança com outras espidroínas anteriormente descritas, com alto teor de alanina e glicina devido à presença dos motivos altamente repetitivo (poli-Ala, (GlyGlyX)<sub>n</sub>, (GlyProGlyGlyX)<sub>n</sub>). Estudos mecânicos e estruturais da fibra ampolada principal, uma das produzidas pela aranha *N. cruentata*, também foram conduzidos no intuito de esclarecer as correlações entre estrutura e função desta espidroína. Nos últimos anos, a maioria das pesquisas sobre as proteínas de seda de aranhas se concentrou nas aranhas de fiação orbicular e em suas intrigantes teias. Outras aranhas de construção não orbicular, tais como a "primitiva" Mygalomorphae, foram em grande parte negligenciadas, assim criando uma nítida lacuna de conhecimento na evolução da seda de aranhas. Neste estudo, foram identificadas duas espidroínas produzidas pela glândula globular da aranha migalomorfa *Avicularia juruensis*, uma aranha nativa da Amazônia brasileira. A análise das seqüências e da filogenia usando 77 C-terminais a partir de 35 espécies de aranhas classificaram, de modo evidente, uma das espidroínas da *Avicularia* dentro da classe da ampolada maior (MaSp2), contribuindo para o melhor entendimento de vários aspectos evolutivos das sedas de aranhas.

**Palavras-chave:** aranhas, espidroínas, evolução, fauna brasileira, estrutura molecular, propriedades mecânicas

## ABSTRACT

Spiders are able to produce up to six different kinds of silk, each one for a specific biological function. Spiders' silks are also known for their unique mechanical properties. The possibility to produce new materials with similar properties led to an advance in the studies about the silks' proteins (spidroins). In this work were identified spidroins produced by the silk glands from the Brazilian spiders *Nephilengys cruentata*, an orb-weaver spider, and *Avicularia juruensis*, from the tarantula family which has only one or two silk glands. *N. cruentata* spidroins sequence showed great similarity to other spidroins previously described, with high content of alanine and glycine amino acids due to the presence of highly repetitive motifs (poly-Ala, (GlyGlyX)<sub>n</sub>, (GlyProGlyGlyX)<sub>n</sub>). In order to establish a comparison between the amino acid sequence motifs found in the major ampullate silk from *N. cruentata* and its mechanical properties, structural and mechanical analyses were also done. Most research on spider silk proteins in recent years has concentrated on orb weaving spiders and their intriguing webs. Other non-web building spiders, such as the "primitive" Mygalomorphae have been largely neglected, creating a clear knowledge gap in spider silk evolution. In this study we have identified two spidroins produced by the globular gland of the mygalomorph spider *Avicularia juruensis*, a native spider from the Brazilian Amazon. Sequence and phylogenetic analysis using 77 C-termini from 35 spider species clearly place one of the *Avicularia* spidroins within the major ampullate (MaSp2) clade, contributing for the better understanding of several evolutionary aspects of the spider silks.

**Key words:** spiders, spidroins, evolution, Brazilian fauna, molecular structure, mechanical properties

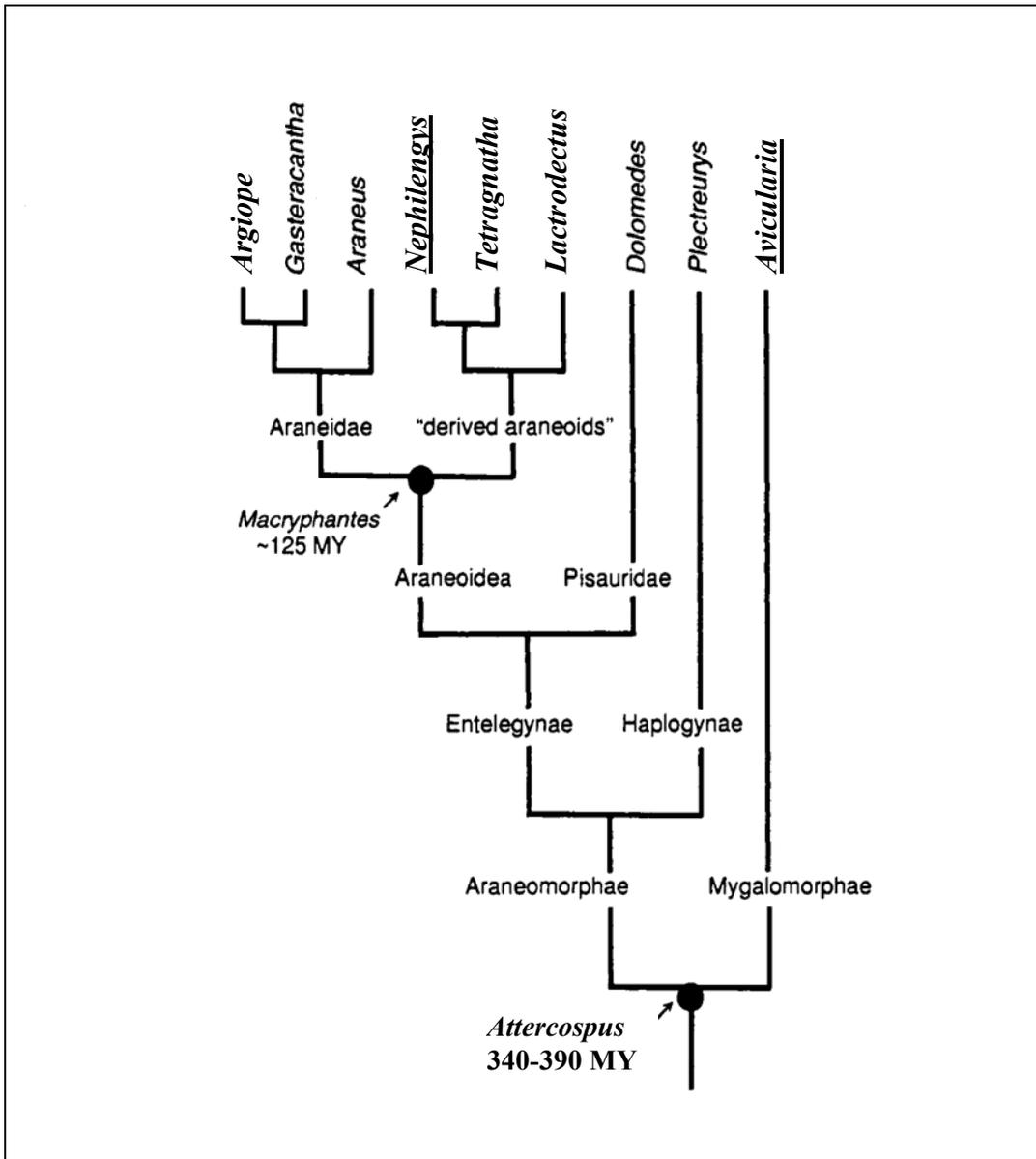
# Capítulo 1

## Introdução



## Relacionamentos evolutivos

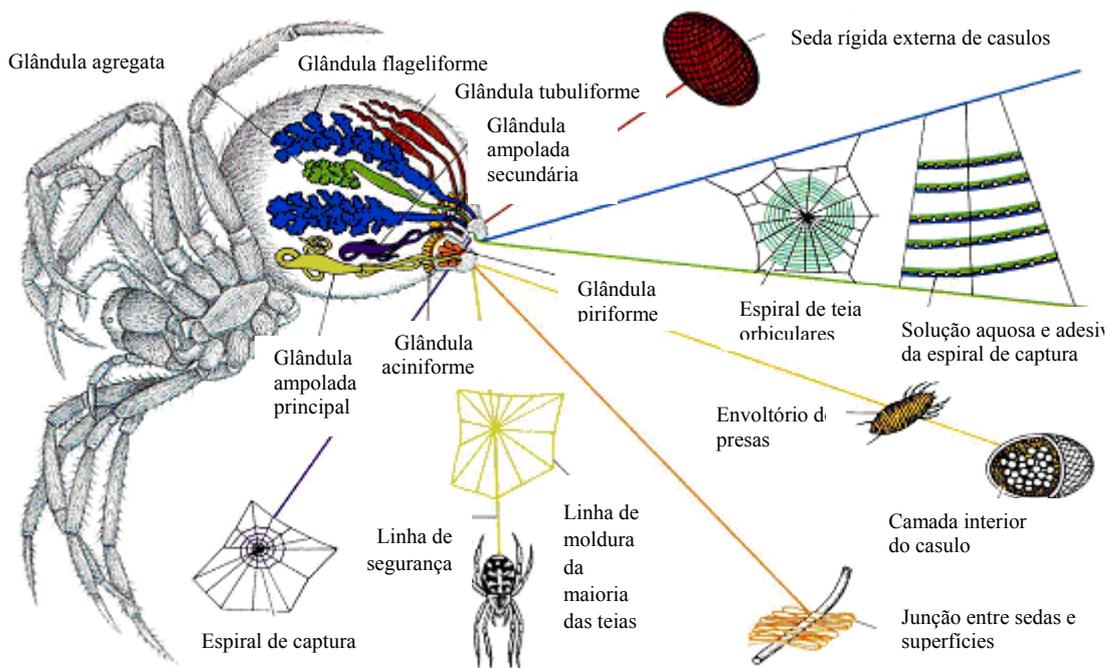
As sedas podem ser produzidas por pelo menos 39.725 espécies de aranhas até o momento identificadas, e agrupadas em 108 famílias (Platnick, 2007). O primeiro fóssil de aranha data de 380 milhões de anos e é a mais antiga espécie de que se tem conhecimento até o momento (Shear *et al.*, 1989, Selden *et al.*, 1991). As aranhas formam um grupo monofilético, sua monofilia é apoiada basicamente por caracteres morfológicos que incluem apêndices abdominais modificados, tais como fiandeiras, glândulas sericígenas e fúsulas. Há duas infraordens monofiléticas atualmente reconhecidas na ordem Araneae: Mesothelae e Opisthotheleae, sendo a última constituída pelas aranhas Mygalomorphae (aranhas rudimentares) e Araneomorphae (aranhas mais especializadas) (**Figura 1**). Mesothelae é o menor grupo e é representado por aranhas que possuem caracteres antigos (ex. abdômen segmentado); nenhuma seqüência de proteína da seda de quaisquer dessas aranhas foi descrita. O grupo Mygalomorphae (caranguejeiras e seus parentes) tem aproximadamente 2.500 famílias reconhecidas (Platnick, 2007) e este grupo inclui a aranha *Euagrus chisoseus* cuja seqüência da proteína de sua seda foi relatada (Gatesy *et al.*, 2001). A Araneomorphae, também conhecida como aranha verdadeira, é a mais diversa em número de famílias (94), e compreende o grande grupo de aranhas tecedoras de teias orbiculares (Orbiculariae: Araneoidea e Deinopoidea) (Foelix, 1996), a maior parte dos dados descritos até o momento sobre as proteínas das sedas de aranha vem desse grupo (Xu e Lewis, 1990; Hinman e Lewis, 1992; Hayashi *et al.*, 2004; Huang *et al.*, 2006). As seqüências parciais das proteínas de sedas descritas nesta dissertação vêm de aranhas das famílias Nephilidae (*Nephilengys cruentata*) e Theraphosidae (*Avicularia juruensis*), pertencentes às subordens Araneomorphae e Mygalomorphae, respectivamente. Com base no mais antigo fóssil de aranha conhecido, *Attercospus* (Selden *et al.*, 1991), a divergência das migalomorfos para as araneomorfos teria ocorrido há 340-390 milhões de anos atrás (Ayoub *et al.*, 2007).



**Figura 1:** Representação da relação filogenética de Araneae (Coddington e Levi, 1991; Hormiga *et al.*, 2000; Gatesy *et al.*, 2001; Ayoub *et al.*, 2007). Espécies extintas, *Attercospus* e *Macryphantes* estão indicadas por círculos pretos. O gênero de aranhas utilizado neste trabalho está sublinhado.

A diversibilidade de espécies se deve, em grande parte, à capacidade das aranhas de produzir sedas durante seu ciclo de vida como também a utilização especializada dos diferentes tipos de sedas produzidos. Durante sua evolução e colonização em diferentes ecossistemas, as aranhas desenvolveram uma estrutura glandular mais especializada capaz de produzir diferentes tipos de seda. Atualmente, de acordo com a espécie, aranhas podem possuir até sete tipos diferentes de glândulas sericígenas, seis delas produtoras de sedas, sendo cada uma utilizada para um propósito específico e

caracterizadas por um impressionante desempenho mecânico (Blackledge e Hayashi, 2006); e uma glândula capaz de produzir uma substância glicoproteica adesiva (Vollrath, 1992). Embora a organização geral das glândulas responsáveis pela produção dos mais diferentes tipos de seda seja muito similar entre as espécies de aranhas, cada uma delas possui uma morfologia única (**Figura 2**).



**Figura 2:** Glândulas de sedas de aranha (Vollrath, 1992). Aranhas podem produzir até seis tipos de sedas. Cada tipo é responsável por uma função e é produzida por um sistema único de glândulas especializadas localizadas no abdômen da aranha como mostrado acima. Cada tipo de glândula se apresenta em pares.

Nos últimos anos, as relações controversas entre os diferentes gêneros e ordens de aranhas estão sendo analisadas por meio da geração de dados a partir do seqüenciamento de DNA e da construção de árvores filogenéticas. Por exemplo, as tentativas para dirimir as relações amplamente debatidas entre as viúvas-negras (gênero *Lactrodectus*) foram relatadas com base no DNA mitocondrial (Garb *et al.*, 2004) e DNA ribossomal (Zhang *et al.*, 2004). Mais recentemente, relações filogenéticas das migalomorfas foram baseadas no fator de alongamento-1 $\gamma$  e nos genes nucleares dos RNAs ribossomais 18S e 28S (Hedin e Bond, 2006; Ayoub *et al.*, 2007). As seqüências das proteínas de sedas (espidroínas) também podem ser usadas para as análises

filogenéticas e para a determinação das relações evolutivas entre elas (Hayashi *et al.*, 2004; Garb e Hayashi, 2005; Challis *et al.*, 2006, Garb *et al.*, 2006).

### **Arquitetura Molecular**

Independente do tipo, todas as sedas de aranha são compostas por grandes proteínas, calculadas em até 500 kD. Essas proteínas podem ser examinadas de acordo com a hierarquia da sua arquitetura molecular. A análise da seqüência de aminoácidos possibilita não apenas uma perspectiva de evolução, mas também de estrutura e função das espidroínas. As diferentes sedas produzidas pelas espécies de aranha que fiam uma teia orbicular aérea, são de importância decisiva na luta pela sobrevivência e, portanto, as mudanças evolutivas são raras e de grande importância. O exame das seqüências de aminoácidos da espidroína da seda da glândula ampolada principal de uma variedade de diferentes espécies de aranha que produzem a teia orbicular, revelou não apenas um alto grau de conservação das seqüências, mas também identificou um subconjunto repetitivo de seqüências-motivo (Gatesy *et al.*, 2001). Além das considerações evolutivas, as séries concatenadas de motivos repetitivos de aminoácidos expõem uma estrutura de nanocomponentes na qual os domínios ricos em alanina e glicina (poli-Ala e (Gly-Ala)<sub>n</sub>) adotam uma estrutura em folhas-β responsável pela formação de cristais (Lucas, 1964; Parkhe *et al.*, 1997; Qin *et al.*, 1997; Hayashi e Lewis, 1998; Grubb e Ji, 1999; Hayashi e Lewis, 2000; Riekel *et al.*, 2000); e a região GlyGlyX (X=Leu, Tyr, Ser, Ala) forma 3<sub>10</sub>-hélices que conectam os cristais de modo a estabilizar e orientar a estrutura da fibra (Tamburro *et al.*, 1991; Tatham e Shewry, 2000). Hipoteticamente, a base para a elasticidade da fibra deve ser a prolina presente no motivo GlyProGlyXX (X=Gly, Gln, Tyr, Ala, Ser), que é considerado responsável por formar uma matriz móvel e um tanto amorfa (Dong *et al.*, 1991; Hayashi e Lewis, 1998; Gosline *et al.*, 2002) (**Tabela 1**). As implicações de tal disposição de motivos podem ser deduzidas a partir de estudos moleculares dinâmicos baseados em outros nanocomponentes de polímero que sugerem que o movimento das regiões internas de um nanocomponente contribui significativamente para o aumento da sua resistência (Gersappe, 2002).

**Tabela 1**

*Motivos de aminoácidos encontrados em 3 diferentes tipos de espidroínas e sua estrutura adotada*

Proteína	Espiral-β elástica * GPGXX		Folhas-β * Rico em Ala		3 <sub>10</sub> - hélice*	Espaçador** Estrutura desconhecida	C- terminal** Estrutura desconhecida
	GPGGX	GPGQQ	(GA) <sub>n</sub>	A <sub>n</sub>	GGX	Único	Único
MaSp1			X	X	X		X
MaSp2	X	X		X			X
MiSp1			X	X	X	X	X
MiSp2			X	X	X	X	X
Flag	X				X	X	X

X - indica a presença do motivo na estrutura.

\*Cada motivo de aminoácido é responsável por uma determinada propriedade mecânica na espidroína.

\*\* As regiões C-terminal e região espaçadora não repetitivos ainda não foram identificadas como responsáveis por adicionar propriedade mecânica as espidroínas.

### **Tipos de Seda**

***Ampolada principal:*** A maioria dos estudos científicos foi realizada com base na fibra da ampolada maior que serve como linha de segurança da aranha e como a seda de moldura estrutural para sua teia. As razões para importância da fibra ampolada maior são duas: -primeiro, a glândula ampolada maior é a mais volumosa no abdômen da aranha e sua forma distinta de ampola faz com que seja fácil identificar e dissecar; - segundo, as propriedades físicas desta seda, tais como o diâmetro da fibra e a facilidade de coleta (por meio da confecção forçada da seda), sem mencionar as propriedades mecânicas (**Tabela 2**) dessa super-resistente fibra, faz com que seja relativamente fácil realizar uma variedade de experiências.

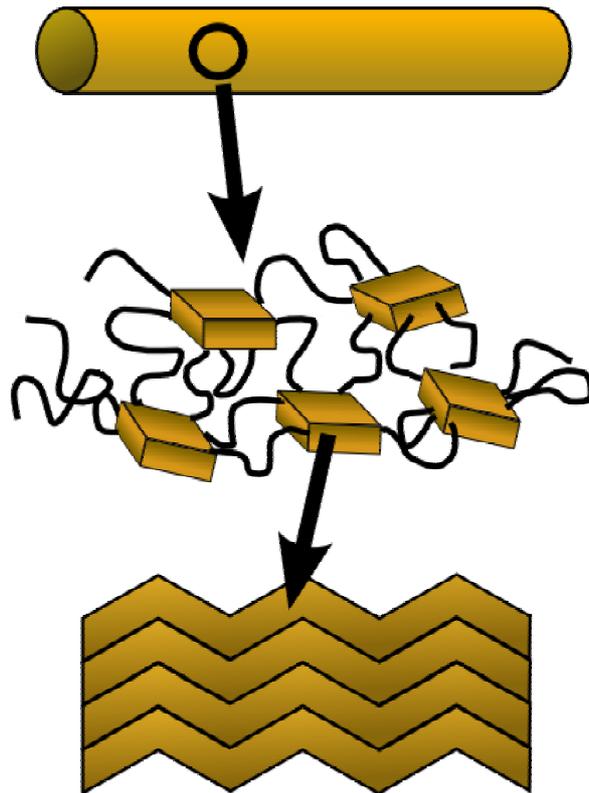
**Tabela 2**

*Comparação entre diferentes propriedades mecânicas de sedas de aranha, seda do bicho-da-seda, fibras naturais e polímeros sintéticos*

<b>Material</b>	<b>Tensão (Pa)</b>	<b>Extensibilidade (%)</b>	<b>Energia para romper (J/kg)</b>
Seda ampolada principal	$4 \times 10^9$	35	$1 \times 10^5$
Seda ampolada secundária	$1 \times 10^9$	5	$3 \times 10^4$
Seda flagelliform	$1 \times 10^9$	>200	$1 \times 10^5$
Seda do bicho-da-seda	$1.3 \times 10^9$		
Kevlar	$4 \times 10^9$	5	$3 \times 10^4$
Borracha	$1 \times 10^6$	600	$8 \times 10^4$
Tendão	$1 \times 10^9$	5	$5 \times 10^3$
Nylon, Tipo 6	$7 \times 10^7$	200	$6 \times 10^4$

Hinman *et al.*, 2000

A modelagem molecular da seda ampolada maior seca utiliza extensas quantidades de dados apresentando o modelo aceito de folhas- $\beta$  incorporadas em uma matriz móvel, um tanto amorfa, que formam um composto de fibra nano-estruturado (Termonia, 1994) (**Figura 3**). A base para as duas fases distintas (isto é, rígida e móvel) dessa fibra de biopolímero reside em sua amalgamação de duas proteínas distintas: MaSp1 e MaSp2. Embora a natureza da interação de MaSp1 com MaSp2 não seja identificada, o fato de que ambas as proteínas estão sempre acopladas na fibra da ampolada maior, independente da espécie da tecedora da teia orbicular, indica que a associação entre as duas proteínas é conservada (Tian e Lewis, 2004). Esta relação estabelece as bases para uma série de conclusões com relação ao propósito do alto grau de conservação das proteínas, como também uma base para as correlações entre as seqüências-motivo de aminoácidos e das propriedades mecânicas da fibra sólida.



**Figura 3:** Desenho esquemático da microestrutura da fibra ampolada principal. Dentro da fibra, blocos de aminoácido formam regiões de folhas- $\beta$  (abaixo), que promovem a organização de cristais. Estes cristais são intercalados por regiões amorfas que criam um material nanoestruturado (meio).

Independentemente das funções da seda ou da ecologia da aranha e, apesar das diferenças na seqüência de aminoácidos das fibras, todas apresentam um impressionante desempenho mecânico (Blackledge e Hayashi, 2006). Ademais, certos motivos (isto é, poli-Ala) na seqüência de aminoácidos da fibra foram mantidos, independente da estratégia predatória da aranha (Gatesy *et al.*, 2001; Tian e Lewis, 2004). Essas observações no contexto de correlação das propriedades mecânicas da fibra com sua seqüência de aminoácidos são o alicerce para a sugestão de que a evolução de uma fibra forte e resistente antecede a teia orbicular (Swanson *et al.*, 2006).

As fibras da glândula ampolada maior são conhecidas como possuidoras de resistência na mesma ordem de magnitude daquela da fibra Kevlar de alto desempenho, mas com aproximadamente sete vezes sua elasticidade. Grande parte dos dados é de espécies ecologicamente similares. Para constatar o potencial da seda de aranha para a aplicação comercial e a fim de verdadeiramente entender os princípios de concepção da

sua natureza, são necessários estudos mais extensos sobre as propriedades mecânicas dessa superconsistente fibra (**Tabela 2**).

**Ampolada secundária:** Restrições contraditórias de *design*: a necessidade de ser forte o suficiente para parar a vítima voadora e manter ainda sua elasticidade e a necessidade simultânea de poder absorver a energia cinética do impacto da presa, fazem da teia orbicular com sua organização de múltiplas fibras uma estrutura aperfeiçoada durante a evolução das aranhas (Lucas, 1964; Denny, 1976; Wirth e Barth, 1992). A seda da ampolada secundária forma a espiral auxiliar da teia. As propriedades mecânicas da seda da ampolada secundária (**Tabela 2**) também precisam ser totalmente elucidadas, mas as curvas preliminares de tensão indicam que ela tem cerca de um quarto da resistência da seda da ampolada maior e, em linhas gerais, a mesma elasticidade de fibra Kevlar (Gosline *et al.*, 1999).

A análise das seqüências das ampolada secundárias revela que da mesma forma que a seda da ampolada maior, esta é uma combinação de duas proteínas. Entretanto, uma comparação mais detalhada demonstra uma ausência do motivo GlyProGlyXX, como também a presença de uma região espaçadora única não repetitiva (Colgin e Lewis, 1998). Embora não haja uma função designada para esta região, reconhece-se como sendo conservada entre as espécies produtoras da teia orbicular. Além disso, a total falta de GlyProGlyXX em qualquer uma das proteínas das ampolada secundárias, quando consideradas no contexto das propriedades mecânicas, proporciona crédito adicional à suposição de que GlyProGlyXX confere elasticidade à fibra.

**Tubuliforme:** Embora os aracnídeos não sejam os únicos artrópodes a ter evoluído a seda, certamente são os mais adeptos e especializados em seu uso. Ao contrário de muitos outros usuários de seda, ao longo de suas vidas, as aranhas evoluíram a capacidade de produzir muitas de suas sedas; entretanto, a seda da tubuliforme, utilizada na construção do casulo, só é produzida durante o período reprodutivo do ciclo de vida de uma aranha. As comparações de seqüências a partir de uma variedade de espécies indicam que apesar das origens antigas desta seda, as proteínas ortólogas das tubuliformes (TuSp1) possuem repetições de seqüências altamente conservadas (Tian e Lewis, 2005; Garb e Hayashi, 2005). A análise da composição de aminoácidos revela apenas uma proteína da tubuliforme com alto teor de serina e baixo teor de glicina devido a uma série de motivos repetitivos ( $\text{Ala}_{n \leq 3}$ ,  $\text{Ser}_n$ ,  $(\text{SerAla})_n$ ,  $(\text{SerGln})_n$ , GlyX) não

vistos em quaisquer outras sedas de aranhas (Tian e Lewis, 2005; Garb e Hayashi, 2005; Hu *et al.*, 2005b; Hu *et al.*, 2006).

Embora a seqüência da proteína da tubuliforme seja conhecida, as informações com relação à mecânica da seda do casulo das aranhas orbiculares são escassas (Blackledge e Hayashi, 2006). Entretanto, as observações no contexto da função biológica da seda tubuliforme sugerem que essa fibra em particular teria que ser suficientemente forte para proteger os ovos, ainda frágeis, de modo a fazer com que seja possível a prole ser chocada. Na realidade, a seda da tubuliforme tem um módulo de Young mais elevado quando comparado à seda da ampolada maior, mas um limite de ruptura e resistência inferiores (Blackledge e Hayashi, 2006).

**Aciniforme:** Ao contrário de muitas outras sedas, a análise de aminoácido indica que a seda da aciniforme tem um percentual relativamente baixo de glicina e alanina (Hayashi *et al.*, 2004). Além disso, a seda da aciniforme é produzida como uma secreção fibrosa por múltiplas fúsculas localizadas nas fiandeiras medianas posteriores. Essa seda é utilizada não apenas para envolver a presa, mas a aranha também a usa para decorar a teia (visto nas espécies de *Argiope*), construir teias de espermatozóides e serve como uma camada primária na confecção do casulo. A seda da aciniforme contém apenas uma proteína conhecida, AcSp1, com unidades repetitivas únicas: poli-Ser e ThrGlyProSerGly não vista na seda da ampolada maior, na seda da ampolada secundária, na seda da tubuliforme ou na seda da flageliforme (Hayashi *et al.*, 2004).

Apesar das diferenças na subestrutura das seqüências, a arquitetura molecular da seda da aciniforme mantém a extraordinária resistência, demonstrando um aumento maior do que 50% na resistência quando comparada à seda da ampolada maior. Esta espantosa resistência é consequência de uma extensibilidade que é mais alta do que a da seda da ampolada principal ou secundária; e uma resistência que é menor do que a da seda da ampolada maior, mas maior do que a da seda da ampolada secundária (Hayashi *et al.* 2004; Blackledge e Hayashi, 2006).

**Flageliforme:** Grande parte das informações disponíveis sobre a base molecular da elasticidade natural nas fibras da seda de aranhas foi produzida a partir de estudos da seda da flageliforme. A observação da composição natural da teia e da mecânica nos informa que a espiral de captura, que é uma combinação de produtos de duas glândulas: fibra da glândula flageliforme e cola líquida de glândula agregata, é a parte mais

extensível da teia. Embora a solução adesiva da aggregate seja extremamente importante para a habilidade da espiral de captura e enlaçar da presa, é a fibra da flageliforme que deve ser extensível o suficiente para manter a integridade da teia. Na realidade, testes mecânicos demonstram que a seda da flageliforme tem uma elasticidade superior a 200%, além do desempenho de qualquer borracha natural (Hayashi *et al.*, 1999).

Analisando a seqüência da proteína da flageliforme, a composição de motivos apóia os estudos mecânicos e empresta credibilidade à suposição de que o motivo pentapeptídeo, GlyProGlyXX, é a base da elasticidade (Hayashi e Lewis, 1998). A arquitetura da seqüência mostra que o motivo GlyProGlyXX forma uma unidade repetitiva da proteína da flageliforme com até 63 repetições concatenadas antes da interrupção por outro motivo. Com base em uma extensiva pesquisa de outras proteínas contendo repetições de Pro-Gly na sua estrutura primária e na corroboração de estudos estruturais sobre o glúten de trigo e a elastina W4, a estrutura secundária resultante do motivo pentâmero iterado é hipoteticamente voltas- $\beta$  do tipo II (Hutchinson e Thornton, 1994; Urry *et al.*, 1995; Urry e Parker, 2002). Além disso, há uma evidente falta de domínios ricos em alanina que corresponde à resistência à tensão diminuída da fibra quando comparada às fibras da ampelada maior (Hayashi e Lewis, 1998).

**Agregata:** A glândula agregata produz a cola pegajosa da seda flageliforme que reveste o núcleo da espiral de captura da teia. As modificações glicoproteicas, não vistas em outras proteínas das sedas de aranhas, são cogitadas por serem responsáveis pelas qualidades adesivas às proteínas da agregata (Tillinghast *et al.*, 1991). Há também evidência de que esta cola glicoproteica contenha sais que agem como bactericidas (Vollrath e Knight, 2001) e compostos orgânicos que atraem a água atmosférica (Edmonds e Vollrath, 1992). Essa água serve para plastificar a fibra flageliforme que esta cola reveste (Vollrath e Edmonds, 1989).

**Piriforme:** Embora a seda da piriforme seja fundamental à teia orbicular aérea, nem a seqüência ou as propriedades mecânicas dessa seda responsável pela adesão da teia a superfícies é conhecida. A análise da composição de aminoácido desta glândula mostra uma abundância de resíduos polares e carregados (isto é, ácido glutâmico, serina, lisina). De fato, a composição da seda da piriforme é mais similar ao produto protéico da glândula aggregate do que as fibras de teias mais tradicionais (Andersen, 1970).

Identificar as excelentes propriedades mecânicas da seda de aranha é fácil, porém a fonte de tal extraordinário *design* permanece desconhecida. Há três explicações possíveis para o conjunto de atributos materiais, aparentemente incomparáveis. Primeiro, a resposta reside na arquitetura molecular das proteínas, em sua seqüência (**Tabela 3**). Segundo, a chave encontra-se no aparato de fiação da aranha. Por último e talvez o mais realista, uma combinação de ambas, da seqüência das proteínas junto com o inconfundível aparato de fiação, responderia pelo equilíbrio entre resistência e elasticidade.

Considerando a extensa variação das propriedades mecânicas resultantes de um simples subconjunto de motivos de aminoácidos que se combinam de modo a criar uma pequena variedade de proteínas, essa variação propicia uma forte evidência do potencial significativo de se adaptar um biomaterial à base de seda (**Tabela 3**). Por conseguinte, entender como uma única fibra de um décimo até um centésimo do diâmetro de um único fio de cabelo humano pode ser mais forte que aço e mais elástica que qualquer outra borracha natural estimula a imaginação e leva a uma pesquisa no intuito de produzir uma fibra sintética de alto desempenho que possa emular e, potencialmente, melhorar as capacidades da seda de aranha para aplicações médicas (isto é, substituição de tendões e suturas ultrafinas de cirurgia óptica e neurocirurgia), militares (isto é, uniformes de campanha à prova de balas leves) e outras aplicações comerciais (isto é, *airbags* (bolsas de ar) e pneus). A utilização da seda de aranha como base para as potenciais aplicações descritas acima, só será possível a partir do momento que a natureza básica dessas fibras for compreendida detalhadamente, não apenas a sua seqüência de aminoácidos, mas também de todos os caminhos de produção da seda como uma base para o controle das diversas e altamente desejáveis propriedades mecânicas.

**Tabela 3**

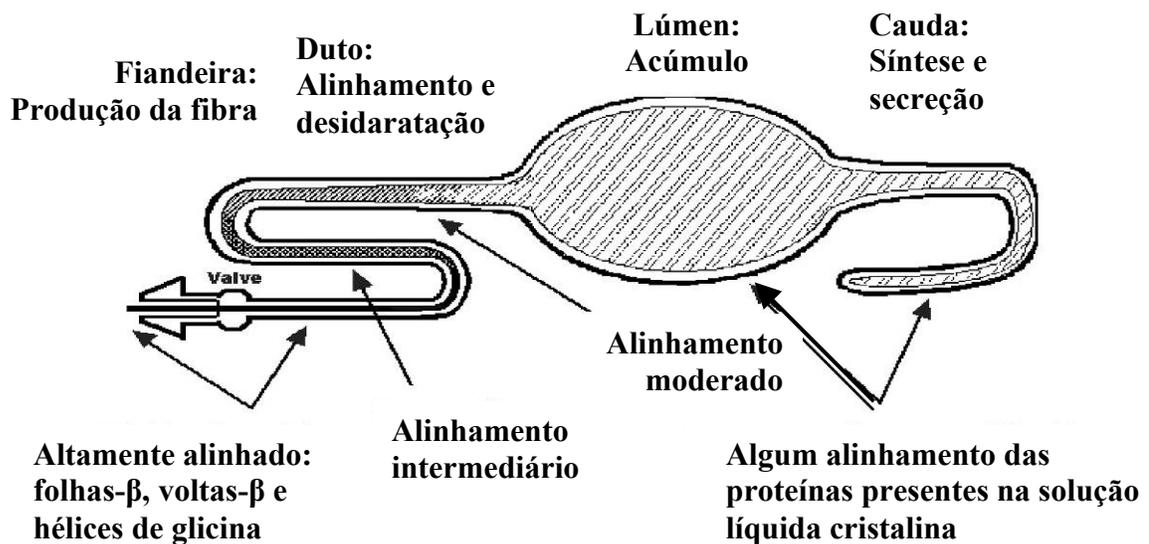
*Comparação do teor de aminoácidos, do desempenho mecânico e da função ecológica de cinco sedas identificadas até o momento (modificado de Blackledge e Hayashi, 2006).*

Seda	Elementos moleculares	Desempenho mecânico	Função ecológica
Ampolada principal	Composição: motivos GA e poli(A), repetições curtas de motivos GPGX <sub>n</sub> <sup>1,2</sup> Estrutura 2 <sup>a</sup> : cristais de folha-β posicionados ao longo do eixo da fibra e incorporados em matriz amorfa <sup>3,4</sup>	Alta resistência à tração, baixa extensibilidade, exhibe supercontração quando umedecida.	Linha de segurança, elementos estruturais primários secos da maioria das teias de captura.
Tubuliforme	Composição: rica em A e S, repetições longas e complexas faltando motivos de sub-repetições <sup>5,6</sup> Estrutura 2 <sup>a</sup> : folhas -β serpenteiam em paralelo à fibra e incorporadas em matriz amorfa <sup>3,7,8,9</sup>	Módulo de Young alto, baixa força, muito pequeno endurecimento após o rendimento da fibra, diâmetro grosso.	Seda externa rígida de casulos.
Ampolada secundária	Composição: motivos (GA) <sub>n</sub> com ausência de motivos poli(A) <sup>1</sup> Estrutura 2 <sup>a</sup> : cristais de folha-β posicionados ao longo do eixo da fibra e incorporados em matriz amorfa <sup>9</sup>	Módulo de Young e extensibilidade altos, resistência à tração moderada e robustez.	Espiral temporária do orbe.
Aciniforme	Composição: rica em A e S, repetições longas e complexas faltando motivos de sub-repetições <sup>10</sup> Estrutura 2 <sup>a</sup> : não caracterizada	Módulo de Young, alta extensibilidade e robustez, folha de multi-filamentos de fibras finas.	Envoltório de presas, decorações de teia, camada interior de casulos.
Espiral de captura	Composição: fibra da espiral da teia (glândula sericígena flageliforme) revestida com cola de glicoproteínas, a fibra do núcleo possui longas repetições de motivos GPGX <sub>n</sub> <sup>11</sup> Estrutura 2 <sup>a</sup> : falta β-folha, sub-repetições se duplicam em “nanomolas” moleculares, fibra plastificada <sup>12,13,14,15</sup> .	Extremamente extensível e elástica altamente complacente, fibra úmida revestida de cola.	Espiral pegajosa de teias orbiculares da <i>ecribellate</i>

Os aminoácidos estão indicados por abreviações de uma letra: A, alanina; G, glicina; P, prolina; S, serina; X, glicina ou outro aminoácido.<sup>1</sup> (Gatesy *et al.*, 2001),<sup>2</sup> (Xu e Lewis, 1990),<sup>3</sup> (Parkhe *et al.*, 1997),<sup>4</sup> (Thiel *et al.*, 1997),<sup>5</sup> (Garb e Hayashi, 2005),<sup>6</sup> (Tian e Lewis, 2005),<sup>7</sup> (Barghout *et al.*, 2001),<sup>8</sup> (Barghout *et al.*, 1999),<sup>9</sup> (Dicko *et al.*, 2004),<sup>10</sup> (Hayashi *et al.*, 2004),<sup>11</sup> (Hayashi e Lewis, 2000),<sup>12</sup> (Gosline *et al.*, 1984),<sup>13</sup> (Hayashi e Lewis, 1998),<sup>14</sup> (Hayashi e Lewis, 2001),<sup>15</sup> (Vollrath e Edmonds, 1989).

## **Biologia da seda**

As aranhas desenvolveram um sistema complexo de glândulas distintas especializadas em secretar seda. Essas glândulas não apenas mantêm reservatórios separados de proteínas de seda individuais, mas as observações da estrutura total do aparato de fição revelam que cada glândula se conecta as fúsculas das fiandeiras. Esta organização não apenas permite que a aranha reserve certas sedas para usos específicos, mas também proporciona a produção de múltiplas fibras simultaneamente. Compreender, e assim, reproduzir como a aranha tece a seda exige conhecimento sobre como a seda é secretada e armazenada na glândula; e também sobre as forças necessárias para produzir e expelir uma fibra sólida a partir de uma solução líquida cristalina de fição, sem usar solventes extremos ou outras condições ambientais (Vollrath e Knight, 2001). Uma vez que a glândula ampolada principal é a mais volumosa e a mais acessível de todas as glândulas produtoras de seda, ela foi usada como modelo para decifrar as tênues características morfológicas e funcionais desse avançado mecanismo de fição.

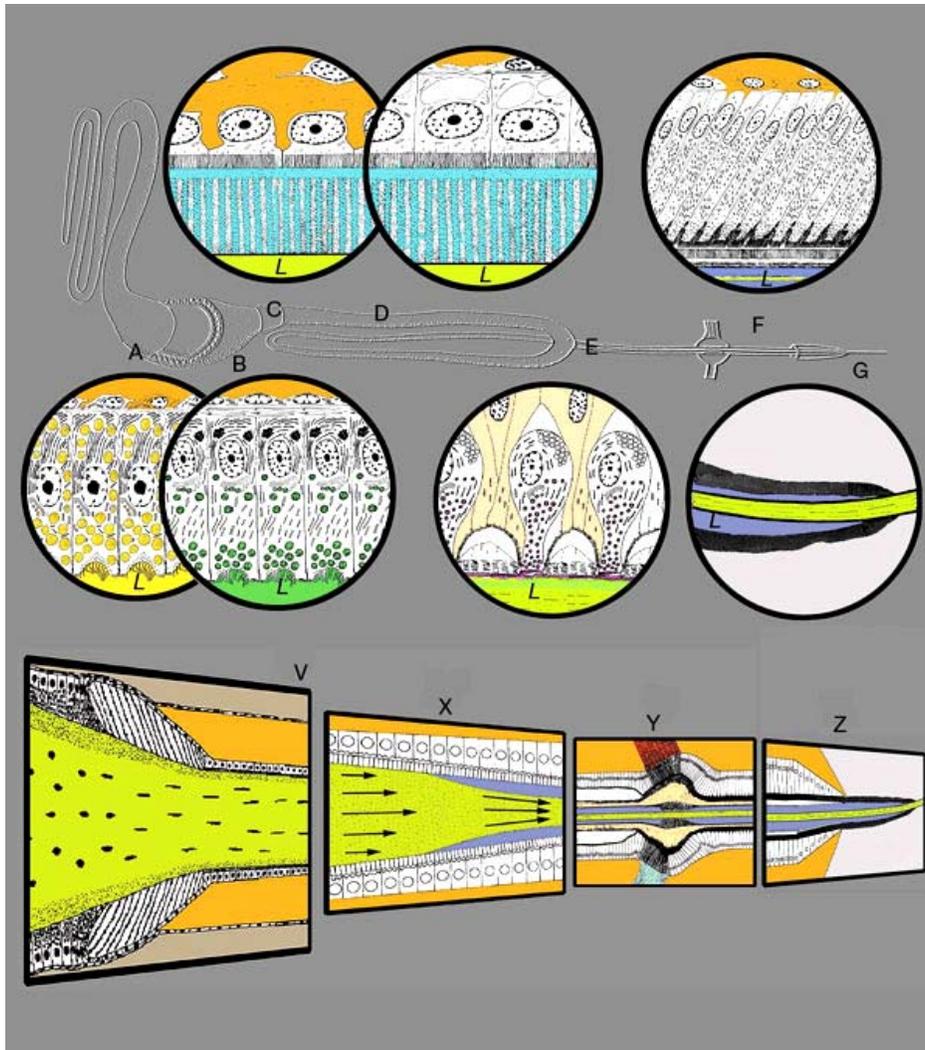


**Figura 4:** Representação artística da estrutura glandular geral. As funções hipotéticas foram designadas a áreas específicas da glândula. Reprodução cortesia do Dr. Michael Hinman.

A partir da observação inicial, a glândula pode ser dividida em 4 seções diferentes: o lúmen que serve como repositório de armazenamento para a solução de fiação; a cauda que abriga as células epiteliais especializadas que secretam as espidroínas (proteínas constituintes da seda); o duto que orienta as proteínas da seda e também reabsorve a água da solução de fiação; e a fiandeira que funciona como válvula e possui funções de controle da produção da fibra final (**Figura 4**). Porém, um exame mais minucioso revela duas zonas transversais distintas na porção da cauda da glândula (Vollrath e Knight, 2001) (**Figura 5**). A zona A é composta por células colunares altas que secretam gotas de mucopolissacarídeos (Hijirida *et al.*, 1996; Rising *et al.*, 2005) na matriz aquosa de espidroínas que nesta fase possui uma concentração de 25-30% (p/v) (Chen *et al.*, 2002). À medida que a solução de fiação caminha em direção à fiandeira, as gotas de mucopolissacarídeos crescem e se aglutinam. Enquanto a coalescência acontece, o maior obstáculo a tão alto conteúdo de proteínas se torna aparente. A habilidade de tecer uma fibra sólida a partir de uma solução altamente viscosa, porém aquosa, é única e incomparável na química dos polímeros. Independentemente se em um sistema biológico natural ou em uma situação sintética, evitar a auto-organização e a interação das proteínas antes de abandonar o abdômen da aranha é um assunto que deve ser elucidado.

Compreender como a aranha confronta e supera este assunto exige um olhar retrospectivo para as seqüências-motivo das espidroínas. As seqüências-motivo que são típicas para a seda de aranha podem ser divididas em duas categorias: domínios hidrofóbicos e domínios hidrofílicos. Não é apenas a presença de blocos hidrofílicos e hidrofóbicos de aminoácidos, mas o mais importante é o arranjo dos domínios que permite que uma solução líquida e cristalina de fiação seja processada em uma fibra sólida, ao mesmo tempo em que se impede as interações prematuras que fariam com que as proteínas da seda se precipitassem. Os aminoácidos hidrofóbicos são agrupados para excluir a água à medida que a solução flui através da glândula afunilada, com isso criando cristais de folhas- $\beta$  para confeccionar fios fortes e insolúveis (Knight e Vollrath, 2002). Para manter a solubilidade da solução de fiação, domínios hidrofílicos flanqueiam um centro principalmente hidrofóbico (Bini *et al.*, 2004). Além da interação entre os domínios hidrofílicos, domínios hidrofóbicos e o ambiente aquoso, a contribuição da região C-terminal não repetitiva e altamente conservada das proteínas MaSp na formação

de fibras não pode ser negligenciada. Uma análise do domínio C-terminal das principais proteínas das glândulas sericígenas ampullate revela um único resíduo conservado de cisteína que poderia permitir a formação de dímeros, o que, conseqüentemente, contribuiria substancialmente para as propriedades mecânicas da fibra tecida. Não obstante o papel do resíduo de cisteína é evidente que a C-terminal encontra-se presente nos conteúdos glandulares, como também na fibra final (Sponner *et al.*, 2005). Estudos adicionais ilustram, de maneira conclusiva, a necessidade do domínio C-terminal para a adequada orientação e polimerização da fibra (Sponner *et al.*, 2005; Ittah *et al.*, 2006), como também sugerem uma função na preservação da solubilidade ou na promoção da agregação das espidroínas, dependendo das condições ambientais (Sponner *et al.*, 2005). Notavelmente, observe-se que a região C-terminal não é necessária para a fiação de fibras sintéticas. Depois de abandonar a zona A e entrar na zona B, a matéria-prima da seda é revestida por outra solução viscosa da zona B que é secretada a partir das células colunares, que diferem ligeiramente em sua morfologia quando comparadas às células colunares da zona A (**Figura 5**; Vollrath e Knight, 1999). A fase líquida cristalina, que é fundamental para a velocidade com a qual a aranha pode tecer uma fibra de seda, possibilita que a suspensão de espidroínas flua pela glândula ao mesmo tempo em que mantém uma orientação simultaneamente similar àquela encontrada em um cristal com as moléculas sendo organizadas em paralelo ao longo do eixo (Vollrath e Knight, 2001).



**Figura 5:** Histologia da glândula sericígena. Os painéis de A-G destacam a histologia da parte de glândula responsável pelo o processo de tecelagem (lúmen, L). A parte inferior detalha as funções que ocorrem em partes específicas da glândula. A produção da solução de fição ocorrendo nas zonas A e B são mostradas nos painéis A e B, respectivamente. Note a diferença na morfologia das células de secretoras. Os painéis C e D mostram a histologia de epitélio no primeiro e segundo membros, respectivamente. Os painéis V-Z demonstram a natureza afilada do extenso duto. O processo de polimerização é principalmente interno e começa no terceiro membro do duto (painel X), que é aproximadamente 4mm da fúsula de saída. As funções do estreitamento da glândula são reter água, alinhar as proteínas, produzir a fibra final sob tensão no passo final de polimerização, e finalmente produzir uma fibra mecanicamente equilibrada com custos metabólicos mínimos. Reproduzido de Vollrath e Knight, 2001.

Entender como a aranha pode armazenar tal solução de proteínas aquosa altamente concentrada é apenas parte da equação que explica a base para a compreensão das excelentes propriedades mecânicas da fibra da glândula ampolada principal (**Tabela 1, Tabela 2**). A outra parte da equação consiste no processo de extrusão. A força de cisalhamento envolvida na compactação da solução de fiação extremamente viscosa à medida que prossegue em direção à fiandeira propicia a evidência de que a separação de fases induzida por este mecanismo desempenha um papel essencial na formação de uma fibra sólida (Chen *et al.*, 2002).

Combinar as forças necessárias para tecer uma fibra sólida com a alta viscosidade da solução de fiação proporciona à aranha a flexibilidade de alterar a velocidade do processo natural de fiação e produzir uma fibra sólida com forças externas relativamente pequenas que se distendem sobre a fibra (isto é, gravidade ou as pernas da aranha). Até o momento, recriar não apenas as condições da solução de fiação altamente concentrada, mas também reproduzir as forças de fiação no laboratório foi relativamente mal sucedido. A fiação sintética ou de laboratório emprega um conjunto completamente diferente de forças, sendo necessário separar o processo de organização e desidratação das espidroínas extraindo a solução de fiação por meio de uma agulha com um diâmetro pequeno (produz o alinhamento das proteínas) em um banho de coagulação (imita a desidratação). Antecipa-se que a discrepância entre as forças naturais e sintéticas usadas para produzir uma fibra de seda possui alguma influência nas propriedades mecânicas da fibra produzida sinteticamente.

### **Propriedades Mecânicas**

Os testes de tensão da seda da ampolada principal da *Nephila clavipes* são o princípio básico para a maioria dos estudos sobre as propriedades mecânicas das sedas de aranha. Historicamente, com base em estudos rudimentares, os resultados de testes da seda da aranha *N. clavipes*, que podem ser aplicados a outras tecedoras de teia orbicular, embora haja diferenças entre as espécies (Brooks *et al.*, 2005), demonstram a mistura ecologicamente ideal das propriedades mecânicas para realizar uma ampla variedade de funções exigidas durante a vida da aranha (Termonia, 1994). Porém, compreender o

*design* prototípico desta fibra necessita de um entendimento das funções biológicas da seda em particular (Gosline *et al.*, 2002). A teia precisa ser forte o bastante para deter um inseto voador, e ainda, não tão elástico para evitar que o inseto seja catapultado fora da mesma devido ao recuo elástico (Denny, 1976; Wirth e Barth, 1992; Gosline *et al.*, 2002). Conseqüentemente, a maioria dos estudos se concentra na resistência da seda, como medida pela capacidade de tensão (definida como força dividida pela área da seção transversal original da fibra, supondo um corte transversal circular (Lucas, 1964), em um volume constante (Vollrath e Knight, 2001; Guinea *et al.*, 2006), e na extensibilidade da seda medida pela tração (a tração de engenharia é a mudança no comprimento dividida pelo comprimento original da fibra). Observe-se que a suposição de volume constante contradiz o comportamento observado de afinamento ou estreitamento que ocorre durante o teste de tensão. A maioria dos estudos que considera essas duas propriedades (isto é, tensão e tração) com a adição da energia para romper (a quantidade de energia exigida para fragmentar a fibra como descrito pela área sob a curva de tensão-tração), e o módulo de Young (rigidez quantificada pelo declive do segmento linear inicial da curva de tensão-tração antes do primeiro ponto de rendimento ou transição de fase/alinhamento das proteínas da fibra) propicia um quadro relativamente abrangente dos princípios de *design* que governam as funções mecânicas da seda da ampolada principal (**Tabela 4**).

Entretanto, quantificar determinadas propriedades, tais como o módulo de Young, que é normalmente usado como uma medida da elasticidade da fibra, não propicia um entendimento abrangente sobre a verdadeira elasticidade da fibra. A maioria dos estudos que utilizam o módulo de Young e a extensibilidade para definir a elasticidade dá um quadro incompleto da elasticidade apenas medindo a deformação recuperável ou a deformação elástica. A fim de completar o conceito, é necessário considerar a deformação plástica ou a deformação irreversível. É a interação tanto da deformação elástica quanto da plástica que proporciona à fibra natural o devido equilíbrio de extensibilidade e o recuo elástico (Beard, 1992). Historicamente, um número limitado de histerese ou experiências cíclicas de carregamento/descarregamento foi concluído. Basicamente, esses estudos foram realizados para revelar a energia que foi perdida. Estudos adicionais contribuiriam substancialmente para a compreensão da elasticidade.

**Tabela 4**

*Compilação de alguns atributos materiais, propriedade correspondente e unidades de medida.*

<b>Atributo funcional</b>	<b>Propriedade material</b>	<b>Unidade</b>
Rigidez	Módulo de elasticidade	Nm <sup>-2</sup> ,
	Módulo de Young	Pascals
Força	Tensão de fratura, $\sigma_{\max}$	Nm <sup>-2</sup> ,
		Pascals
Dureza	Energia necessária para quebrar a fibra	Jm <sup>-3</sup> , Jm <sup>-2</sup> , Pascals
Extensibilidade	Tração de fratura, $\varepsilon_{\max}$	mm/mm, %

Uma vez que há um corpo da literatura que defende o uso dessa medida alternativa de força e elasticidade para avaliar a seda da aranha, uma completa compreensão do desempenho mecânico das fibras da seda da ampolada principal exige o conhecimento dos parâmetros relacionados com as medições da real tensão e tração. Em contraste com a tensão e tração de engenharia, há algumas reivindicações de que a real tensão e tração seriam uma medida mais precisa das forças e, portanto, deveria ser usada para todas as comparações dos dados do teste mecânico (Blackledge e Hayashi, 2006; Guinea *et al.*, 2006). No entanto, calcular a real tensão e tração para as fibras de seda é tecnicamente desafiante, uma vez que uma única fibra da ampolada principal possui um diâmetro muito pequeno, ainda que potencialmente variável. O cálculo da real tensão necessita conhecer a área da seção transversal instantânea em cada valor da extensão. Os benefícios de tal intenso teste mecânico e as análises resultantes são discutíveis. Felizmente, a suposição de volume constante da fibra durante o teste de tensão foi recentemente confirmada de maneira experimental (Guinea *et al.*, 2006) e possibilita que a tensão ou tração de engenharia seja matematicamente relacionada com a real tensão ou tração, por meio das seguintes equações onde  $\sigma_{tr}$  é a real tensão,  $\sigma_{eng}$  é a tensão de

engenharia,  $\epsilon_{tr}$  é a real tração, e  $\epsilon_{eng}$  é a tração de engenharia. (Brooks *et al.*, 2005; Blackledge e Hayashi, 2006; Guinea *et al.* 2006):

$$\sigma_{tr} = \sigma_{eng}(1 + \epsilon_{eng}) \quad \text{Equação 1}$$

$$\epsilon_{tr} = \ln(1 + \epsilon_{eng}) \quad \text{Equação 2}$$

A controvérsia sobre as propriedades de tensão e tração não é a única questão para complicar o teste e a comparação das fibras. O assunto bem mais urgente da variação intra e entre as espécies nas propriedades físicas, como também a natureza viscoelástica da seda de aranha, estabelece generalidades concernentes às propriedades mecânicas e, dessa maneira, faz com que a avaliação das diferenças entre as espécies seja imprecisa (Madsen *et al.*, 1999; Brooks *et al.*, 2005). Ter as características tanto de um sólido quanto de um líquido (viscoelasticidade) possibilita que a fibra possua algumas incríveis propriedades físicas, mas também traduz a possibilidade de variação sendo introduzida no próprio sistema de teste. As condições ambientais, como também a taxa de tração, são todos os parâmetros decisivos de testes que podem introduzir variação aos dados (Gosline *et al.*, 1986, Guess e Viney, 1998; Shao *et al.*, 1999; Liu *et al.*, 2005; Perez-Rigueiro *et al.*, 2005). Inversamente, a viscoelasticidade e a supercontração, a habilidade de uma fibra ampolada principal não estendida de contrair até a metade de seu comprimento original na presença da água (Bell *et al.*, 2002; Elices *et al.*, 2004), foram informadas como uma maneira de reduzir a variação nas propriedades mecânicas (Guinea *et al.*, 2005). Não obstante, a extensa variação inata é evidente a partir dos desvios padrão produzidos a partir de uma coorte de indivíduos (Madsen *et al.*, 1999; Brooks *et al.*, 2005). O primeiro potencial para divergência ocorre no nível fundamental da anatomia do aparato de fição. A diminuição gradual do duto no ponto de polimerização permite que a aranha controle o diâmetro da fibra produzida (Rising *et al.*, 2005). Essa diminuição gradual não apenas permite a aranha controlar o diâmetro de uma fibra naturalmente fiada, mas o efeito é ainda mais aparente nas fibras que são obtidas por meio da produção induzida da seda. Uma vez que todos os tipos de testes mecânicos contam com a área da

seção transversal da fibra para transformar força em tensão, as variações no diâmetro impactarão a tensão de ruptura da fibra. Por conseguinte, a maioria dos testes usa um diâmetro médio calculado a partir do diâmetro em vários pontos ao longo do comprimento da fibra. A variação no diâmetro da fibra pode apenas explicar os desvios nas propriedades da fibra até certo ponto. É necessário que haja fontes adicionais de discrepância nas propriedades mecânicas de fibras isoladas a partir de aranhas de uma mesma espécie, ainda que a diversidade nas propriedades físicas de sedas de diferentes espécies tenha sido relacionada às diferenças na seqüência das proteínas (Brooks *et al.*, 2005; Swanson *et al.*, 2006). Independentemente da base da diversidade mecânica amplas variações impedem o uso de sedas de aranhas nativas como uma fonte de biomateriais, e traz a necessidade de se desenvolver um biomimético da seda sintético para aproveitar a capacidade de alto desempenho da seda de aranha.

Embora os princípios de *design* fundamentais que regem as propriedades mecânicas da seda de aranha sejam conhecidos, estabelecer a função conferida por um motivo de aminoácido é decisivo para a capacidade de explorar e reproduzir as propriedades mecânicas das fibras da seda de aranha e, desse modo, ser capaz de adaptar essas propriedades de forma a criar uma fibra sintética que se ajuste a certos critérios.

### **Produção de Seda Sintética**

Uma única aranha pode produzir até 6.100 metros de seda em um minuto, sob confecção induzida; trabalho de toda uma vida de cerca de 5.000 aranhas para produzir seda suficiente para fazer um vestido (Toner, 1992; Vollrath *et al.*, 2002). Infelizmente, a natureza territorial, o limitado tempo de vida, a baixa densidade e a falta de consistência no processo de fiação das aranhas tornam impossível aproveitar a seda nativa e adaptá-la a um novo biomaterial. Desse modo, mesmo que as sedas de aranhas sejam uma fonte cobiçada para novos biomateriais, utilizar a seda natural não é prático.

Além disso, a falta de dados de seqüências genéticas das aranhas e a natureza essencial da seda de aranha, impossibilitam o uso de alguns dos recentes avanços tecnológicos na biologia molecular (isto é, RNAi (RNA interferente)) para descobrir a contribuição específica de cada proteína de seda aos atributos físicos destas fibras.

Assim, as técnicas de clonagem clássica e de produção de proteínas foram exploradas em uma tentativa de aproveitar e manipular as propriedades mecânicas de seda da ampolada principal de aranhas.

A produção de fibras de seda de aranha está sendo investigada atualmente em cinco sistemas diferentes: bactérias, fungos, células de insetos, plantas e mamíferos. Há cinco áreas principais de comparação entre os sistemas: capacidade comercial, facilidade de expressão de proteínas, limitações de tamanho das proteínas, técnicas disponíveis e expressão estável. Produzir seda em fungos tem quatro vantagens básicas: proteínas maiores podem ser produzidas de forma a imitar mais acuradamente as proteínas da seda natural, a produção em grande escala é relativamente acessível, a purificação de proteínas é simplificada em *Pichia pastoris* através da secreção eficiente, as cepas produtoras de seda podem ser mantidas com estabilidade para, pelo menos, 100 duplicações (Fahnestock e Bedzyk, 1997). Apesar dessas atrativas características a maioria dos pesquisadores ainda escolhe produzir a seda sintética em bactérias, uma vez que o sistema de clonagem é mais definido, a expressão de proteína em grande escala é conveniente e factível, e as condições de purificação são bem estabelecidas (Lewis *et al.*, 1996; Fahnestock e Irwin, 1997; Brooks e Lewis, 2004). A produção de seda sintética usando o sistema de expressão de células de insetos está em seus primeiros estágios de desenvolvimento, embora a expressão da proteína da seda da flageliforme tenha sido recentemente realizada com sucesso (Miao *et al.*, 2006). A principal vantagem de tal sistema é a clonagem rápida e a produção de uma linha de células estáveis. A produção de seda sintética em plantas tem grande potencial para produzir uma grande quantidade de proteínas de seda. Também possui a capacidade de expressar proteínas maiores e mais similares às naturais. Além disso, os protocolos desenvolvidos para o isolamento de proteínas propiciam a esse sistema um benefício ecológico distinto (Schellar e Conrad, 2005). Porém, um eficiente isolamento de proteínas para o uso comercial ainda não é possível. Talvez o sistema mais promissor, embora o de custo mais proibitivo, é a cultura de células de mamíferos. Usando as proteínas de seda sintética, produzidas e isoladas a partir do leite de cabra, as primeiras fibras de seda biomimética leve foram tecidas (Lazaris *et al.*, 2002). A pesquisa básica que acontece em todos esses sistemas está rapidamente convergindo para produzir uma versão necessária e totalmente sintética da

seda de aranhas com propriedades adaptadas de modo a revelar e manipular a relação estrutura/função.

## CAPÍTULO 2

### Espidroínas da aranha *Nephilengys cruentata*



**Este capítulo foi baseado na seguinte publicação (Anexo 3):**

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, 2007. Spidroins from the Brazilian spider *Nephilengys cruentata* (Araneae: Nephilidae). *Comp Biochem Physiol B Biochem Mol Biol.* 4, 597-606.

## Introdução

Aranhas de teia orbicular podem produzir uma variedade de sedas com exclusivas propriedades mecânicas para os mais diversos propósitos práticos (Foelix, 1996). Os componentes das fibras são proteínas (espidroínas) que são sintetizadas nas células epiteliais de glândulas especializadas e secretadas no lúmen glandular, onde são armazenadas como uma solução de fiação líquida e cristalina altamente concentrada (Hijirida *et al.*, 1996; Scheibel, 2004). Acredita-se que o agrupamento das fibras aconteça durante a passagem da solução de fiação pelo duto da glândula sericígena onde a extração de água, sódio e cloreto são acompanhados por uma diminuição do pH de 6,9 a 6,3 até a sua extrusão pelas fiandeiras como fibra insolúvel (Vollrath e Knight, 2001; Rising *et al.*, 2005; Vollrath, 2005). Apesar do conhecimento disponível sobre alguns detalhes, o processo de fiação de fibras ainda não é claramente entendido.

As tecedoras de teia orbicular possuem até sete tipos diferentes de glândulas, seis que produzem uma seda específica e uma a cola (Vollrath, 1992). A maioria das espidroínas descrita até o momento foi gerada a partir de EST (seqüência expressa identificada) (Xu e Lewis, 1990; Hinman e Lewis, 1992; Hayashi *et al.*, 2004; Tian e Lewis, 2005). Embora haja limitações principalmente quanto ao uso de mRNAs (RNA mensageiros) de tamanho grande, a produção de cDNAs a partir das ESTs provou ser uma técnica rápida e econômica para a identificação de genes de interesse biológico (Adams *et al.*, 1991). Estudos moleculares das várias seqüências de cDNAs (DNAs complementares) produzidas a partir das glândulas sericígenas de diferentes aranhas demonstraram que elas compartilham características estruturais comuns (Xu e Lewis, 1990; Gatesy *et al.*, 2001). Independente do tipo, todas as sedas de aranha são compostas por grandes proteínas, calculadas como sendo entre 300-500 kD, constituídas por um N-terminal e C-terminal não repetitivos e conservados, e uma região interna altamente repetitiva rica em aminoácidos alanina, glicina e serina. A região repetitiva da maioria das espidroínas é formada pelo agrupamento de até quatro motivos de aminoácidos responsáveis por módulos estruturais distintos como segue: polialanina (poli-Ala), glicina e alanina alternadas (GlyAla)<sub>n</sub>, grupos de três aminoácidos composto de duas glicinas e um terceiro aminoácido variável (GlyGlyX)<sub>n</sub>, e módulos de glicina-prolina-glicina

(GlyProGlyXX)<sub>n</sub> (Gatesy *et al.*, 2001). Baseado em vários estudos, as diferentes combinações desses módulos formando um bloco repetitivo maior são responsáveis pelas distintas propriedades mecânicas das fibras (Hayashi *et al.*, 1999; Fahnestock *et al.*, 2000; Scheibel, 2004; Vollrath, 2005). Os módulos poli-Ala e (GlyAla)<sub>n</sub>, por exemplo, formam folhas-β cristalinas que propiciam resistência à tensão, (GlyGlyX)<sub>n</sub> provavelmente forma uma estrutura helicoidal Gly-II, e (GlyProGlyXX)<sub>n</sub> forma as espirais-β; que podem estar envolvidas na formação de uma matriz amorfa responsável pela elasticidade. Mais recentemente, novas espidroínas foram identificadas compostas por um novo tipo de módulo repetitivo, com longas e complicadas repetições contendo altos níveis de serina (Garb e Hayashi, 2005; Hu *et al.*, 2005a; Hu *et al.*, 2005b; Tian e Lewis, 2005; Huang *et al.*, 2006).

A região N-terminal de diferentes espidroínas é a parte mais conservada das proteínas da seda (Smith *et al.*, 2005). Esta possui codons de iniciação da transcrição adicionais abaixo do primeiro criando, possivelmente, diferentes inícios de tradução (Smith *et al.*, 2005). Embora a função da região N-terminal esteja relacionada ao transporte das espidroínas para o lúmen glandular devido a presença de um peptídeo sinal (Hayashi e Lewis, 1998; Smith *et al.*, 2005), a sua função na estrutura protéica permanece desconhecida. A região C-terminal constituída por 100 aminoácidos também é um domínio altamente conservado entre as espidroínas. Devido a sua alta conservação, sugeriu-se que essa região desempenhe um importante papel na polimerização das fibras (Jin e Kaplan, 2003; Spooner *et al.*, 2005). Recentemente, demonstrou-se que a organização das proteínas em uma estrutura macromolecular com correta densidade e orientação das fibras deve depender do domínio C-terminal (Ittah *et al.*, 2006).

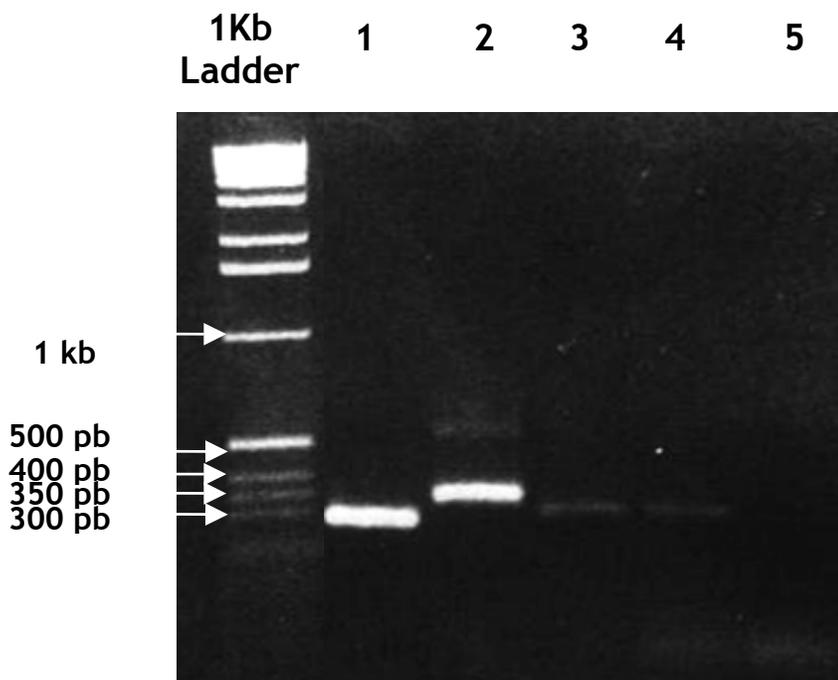
Neste capítulo descrevemos quatro seqüências parciais de cDNA que codificam as espidroínas produzidas pelas glândulas sericígenas ampolada principal, ampolada secundária, flageliforme e tubuliforme da aranha *Nephilengys cruentata* (Araneae: Nephilidae). Objetivando a possível produção de fibras usando as abordagens de biotecnologia capazes de imitar as propriedades naturais das sedas produzidas por aranhas, nossos resultados mostram proteínas possuídoas de uma seqüência de aminoácidos altamente conservada e contribuem para o entendimento entre estrutura e função destas proteínas.

## Resultados

### *Biblioteca de cDNA*

A fim de identificar as novas seqüências que codificam as diferentes proteínas da seda da aranha *N. cruentata*, 960 clones aleatórios da biblioteca de cDNA (DNA complementar) das glândulas sericígenas foram parcialmente seqüenciados e analisados (material e métodos – Anexo 1). Outros noventa e seis clones selecionados a partir do *Southern Blot* feito em 24 placas de 96 poços foram parcialmente seqüenciados e analisados. Desses, 21 clones foram selecionados de acordo com o tamanho e semelhança com espidroínas previamente descritas. Todos os clones selecionados eram transcrições parciais na direção 5' terminal. Quatro clones contendo a região C-terminal e a seqüência repetitiva foram classificados como sendo cDNAs correspondentes a MaSp1, e o clone maior para este grupo de seda foi de 3241 pares de base (pb). Três outros clones foram classificados como cDNAs MiSp1, onde dois deles continham apenas a seqüência repetitiva, e o maior com 3486 pb continha tanto a seqüência repetitiva quanto o C-terminal não repetitivo. O cDNA da espidróina da glândula flageliforme foi o mais altamente representado na biblioteca, com nove clones parciais também contendo a seqüência repetitiva e C-terminal, o mais longo dos quais tinha 3277 pb de comprimento. Embora tendo os menores clones de cDNAs, a espidroína da tubuliforme também foi identificada na biblioteca de cDNA da *N. cruentata* com dois clones de 1600 pb e 1636 pb (Anexo 2). Três outros clones contendo novas seqüências repetitivas foram selecionados para estudo adicional.

RT-PCR (reação em cadeia da polimerase via transcriptase reversa) foi executada a fim de verificar a presença de cDNAs das espidroínas identificadas nas glândulas sericígenas da *N. cruentata*. O RNA total da(s) glândula(s) sericígena(s) foi usado como um molde para a polimerização da primeira fita de cDNA, e os iniciadores escolhidos a partir das seqüências das espidroínas encontradas nas bibliotecas de cDNA foram usados para amplificar a região C-terminal. Todos os cDNAs analisados foram positivos devido a sua presença na(s) glândula(s) sericígena(s) (**Figura 1**), eliminando a possibilidade de contaminação de transcritos ou recombinação de clone.



**Figura 1:** RT-PCR com RNA total da(s) glândula(s) sericígena(s) de *N.cruentata*. 1, Espidroína Flageliforme cDNA (287 pb); 2, MaSp cDNA (356 pb); 3, MiSp cDNA (304 pb); 4, TuSp cDNA (302 pb); 5, controle negativo. 1Kb DNA ladder (Invitrogen).

#### *Análise das seqüências*

Todos os clones de cDNA selecionados possuem uma fase de leitura para uma espidroína de seda de aranha, com um sinal de poliadenilação. Seguindo a nomenclatura para as espidroínas de seda de aranha publicadas, os genes identificados em *N. cruentata* foram denominados: NCMaSp1-like (*N. cruentata* espidroína ampolada principal 1-like), NCMiSp1-like (*N. cruentata* espidroína ampolada secundária 1-like), NCFlag-like (*N. cruentata* espidroína flagelliforme-like) e NCTuSp-like (*N. cruentata* espidroína tubuliforme-like), de acordo com suas traduções. As espidroínas da *N. cruentata* codificam proteínas repetitivas constituídas por simples motivos de aminoácidos ricos em alanina e glicina. As espidroínas NCMaSp1-like e a NCMiSp1-like possuem seqüências com os motivos de aminoácidos (GlyGlyX)<sub>n</sub> e com poli-Ala, este sendo substituído por repetições (GlyAla)<sub>n</sub> na NCMiSp1-like (**Figura 2**). Também foi identificada uma região

## A. MaSp1

```

N. cruentata          GG-AGQGGYGG LGGQ-----GAG-----QGAGAAAAA- 27
N. clavipes           GG-AGQGGYGG LGSQ-----GAGRGGLGG---QGAGAAAAA- 33
N. i. madagascariensis GG-AGQGGYGG LGSQ-----GAGRGGYGG---QGAGAAAAA- 33
A. trifasciata       GGQGGQGGYGG LGXQAGQGYGAGSGGQGGXG--QGGAAAAAAA 43
A. diadematus (ADF-2) GGQGGQGGQGG LGSQ-----GAGGAGQGGYAGQGGAAAAAAA 39
** . ***** *          ***          ** . *****

```

## B. MiSp1

### Região repetitiva:

```

N. cruentata          GAGAGGAGGFGRGAGAGAGAGAAAGAGAGGAGGYGAGQGYGAGAGAGAAAAAGA 56
N. clavipes           --GAGGAGGYGRGAGAGAGAAAGAGA-----GAGGYGGQGGYAGAGAGAAAAAGA 49
A. diadematus (ADF-1) --GAGAAGGYGGGAGAGAG-----GAGGYGQ--GYGAGAGAGAAAAAGA 40
N. antipodiana       -GGYGGLVGYGAGAGAAAGAGAGAG-----GAGGYIGQGGY--GAGAGAAAAAGA 47
* *   * * ***** *          *****   *   *****

```

### Região espaçadora:

```

N. cruentata          GNFAFQSLSSNLLSSGDFVQMISSITTTSTDQAVSVATSVAQNVGNQLGLDANAMNNLLAAV 60
N. clavipes           GNFAFQSLSSNLLSSGDFVQMISSITTTSDHAVSVATSVAQNVGSQLGLDANAMNNLLGAV 60
N. antipodiana       GNFAFQSLSSNLLSSGDFVQMISSITTTSTDQAVSVATSVAQNVGNQLGLDANAMNSLLGAV 60
A. diadematus (ADF-1) -----HESSYAAMAASRN----- 15
                        : * *::* . . .

N. cruentata          GGYVSSLGGAVADAAYANAISSAIGNVLANTGSINESTASSAASSAASSVTTTLTSYGP 120
N. clavipes           SGYVSTLGNAI SDASAYANALSSAIGNVLANS GISESTASSAASSAASSVTTTLTSYGP 120
N. antipodiana       SGYVSTLGNAI SDASAYANAISSAIGNVLANS GISESTASSAASSAASSVTTTLTSYGP 120
A. diadematus (ADF-1) -----SDFIRNMSYQMGRLLSNAGAITESTASSAASSASSTVTESIRTYGP 61
                        : : . .* :*::*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

```

```

N. cruentata          AVFY----- 124
N. clavipes           AVFYAPSASS 130
N. antipodiana       AVFYAPTSSA 130
A. diadematus (ADF-1) AAIFSGAGAG 71
* .::

```

**Figura 2:** Alinhamento das regiões repetitivas de duas espiroínas de *N. cruentata* com espiroínas de diferentes espécies de aranhas. (A) região repetitiva de NCMaSp. (B) Região repetitiva e espaçadora de NCMiSp. Aminoácidos estão abreviados com uma letra. “-”, indica “gaps” nas seqüências para um melhor alinhamento; “\*”, indica que o aminoácido presente naquela coluna é idêntico em todas as seqüências; “.”, indica que uma substituição conservada ocorreu de acordo com as cores (vermelho-aa pequeno, azul- aa ácido, rosa- aa básico, verde-hidroxila+amino+básico-Q); e “:”, indica que substituições semi-conservativas ocorreram. A seqüência de *N. cruentata* está identificada em negrito.

espaçadora não repetitiva entre as seqüências repetitivas de NCMiSp1-*like* (**Figura 2B**). A NCFlag-*like* também é constituída por uma seqüência repetitiva com motivos (GlyProGlyXX)<sub>n</sub> interespaçada por uma seqüência não repetitiva (**Figura 3**). Ao contrário das proteínas acima, a NCTuSp-*like* mostrou pouco dos motivos de aminoácidos comumente encontrados na maiorias das espidroínas. A seqüência contém repetições de 174 aminoácidos, compostos em grande maioria por alanina e serina, formando novos motivos tais como Ser<sub>n</sub>, (SerX)<sub>n</sub> (X representa Gln, Leu, Ala, Val e Phe), e GlyX (X representa Gln, Asn, Ile, Leu, Ala e Val) (**Figura 4**).

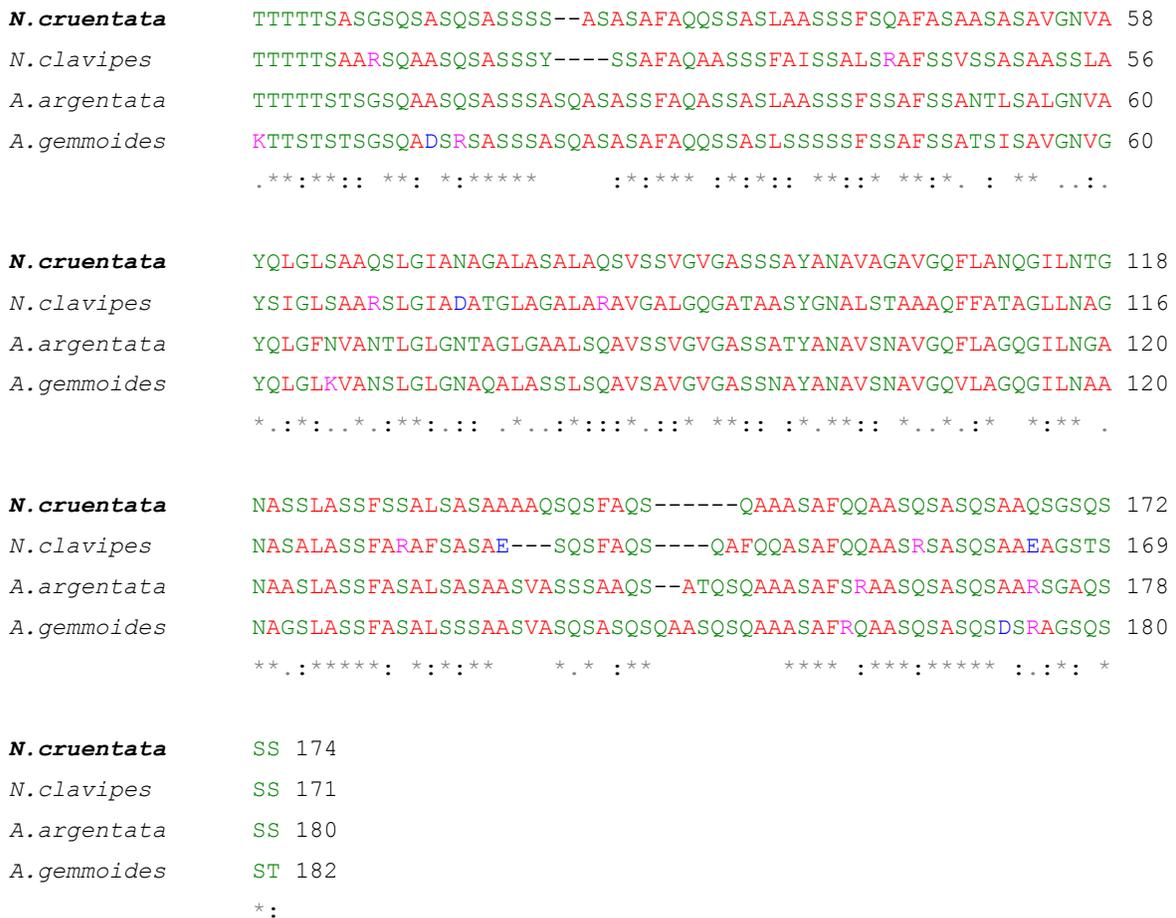
<i>N. cruentata</i>	[GPGGX] <sub>19</sub> [GGX] <sub>3</sub>	TVIEDLDITVNGPGGPITISEELTVGGPAGGS	[GPGGX] <sub>24</sub>
<i>N. clavipes</i>	[GPGGX] <sub>41</sub>	TIIEDLDITIDGADGPPITISEELTIS-GAGGS	[GPGGX] <sub>26</sub>
<i>N. i.madagascariensis</i>	[GPGGX] <sub>36</sub> [GGX] <sub>7</sub>	TVIEDLDITIDGADGPITISEELTIGGAGAGGS	[GPGGX] <sub>19</sub>
<i>A. trifasciata</i>	[GPGGX] <sub>6</sub>	GPVTVDVDSVGGAPGG	[GPGGX] <sub>5</sub> [GGX] <sub>6</sub> [GPGGX] <sub>7</sub>

**Figura 3:** Esquema da região repetitiva de espidroínas da glândula Flageliforme de diferentes espécies de aranhas. “-”, indica “gaps” nas seqüências para um melhor alinhamento. A seqüência de *N. cruentata* está identificada em negrito.

#### *Frequência de codons*

Os codons utilizados com maior frequência para os aminoácidos mais abundantes das espidroínas da *N. cruentata* (**Tabela 1**) segue a preferência para adenina (A) e timina (T) como o terceiro nucleotídeo entre os três codificadores de cada aminoácido, com a exceção da espidroína NCTuSp-*like*. A NCMaSP1-*like* e a NCMiSp1-*like* possuem um alto teor de glicina, alanina e glutamina em suas seqüências de aminoácidos, todas elas usando A ou T como o terceiro nucleotídeo no uso de codons. A preferência de codons para glutamina e alanina foi CAA e GCA/T respectivamente, com CAA presente em 100% dos casos na NCMaSP1-*like* e em 96% na NCMiSp1-*like*. Glicina, o aminoácido mais preponderante em ambas as proteínas, mostrou uma preferência de 93% e 90% para GGA/T na NCMaSP1-*like* e na NCMiSp1-*like*, nessa ordem. Glicina e alanina também estavam presentes em grandes quantidades na seqüência da proteína NCFlag-*like*. O uso de codons delas segue a mesma preferência encontrada para as seqüências de cDNA da

NCMaSp1-like e da NCMiSp1-like, com 89% para GGA/T e 94% para GCA/T. Comparado com a alta frequência da preferência de codons para A e T na posição de oscilação nas seqüências codificando as proteínas NCMaSp1-like, NCMiSp1-like e NCFlag-like, os usos de códon para alanina, serina e glutamina, o três aminoácidos mais freqüentes na proteína NCTuSp-like são apenas moderadamente tendenciosos para A e T, com 63% para alanina, 57% para serina e 50% para glutamina.



**Figura 4:** Alinhamento da região repetitiva da espidroína da glândula Tubuliforme de *N. cruentata* com as de diferentes espécies de aranhas. Abreviações são as mesmas descritas para o alinhamento da figura 2.

**Tabela 1***Escolha de codons para os aminoácidos mais freqüentes nas espidroínas de N. cruentata*

AmAc	Codon	Frequência (%)				AmAc	Codon	Frequência (%)			
		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

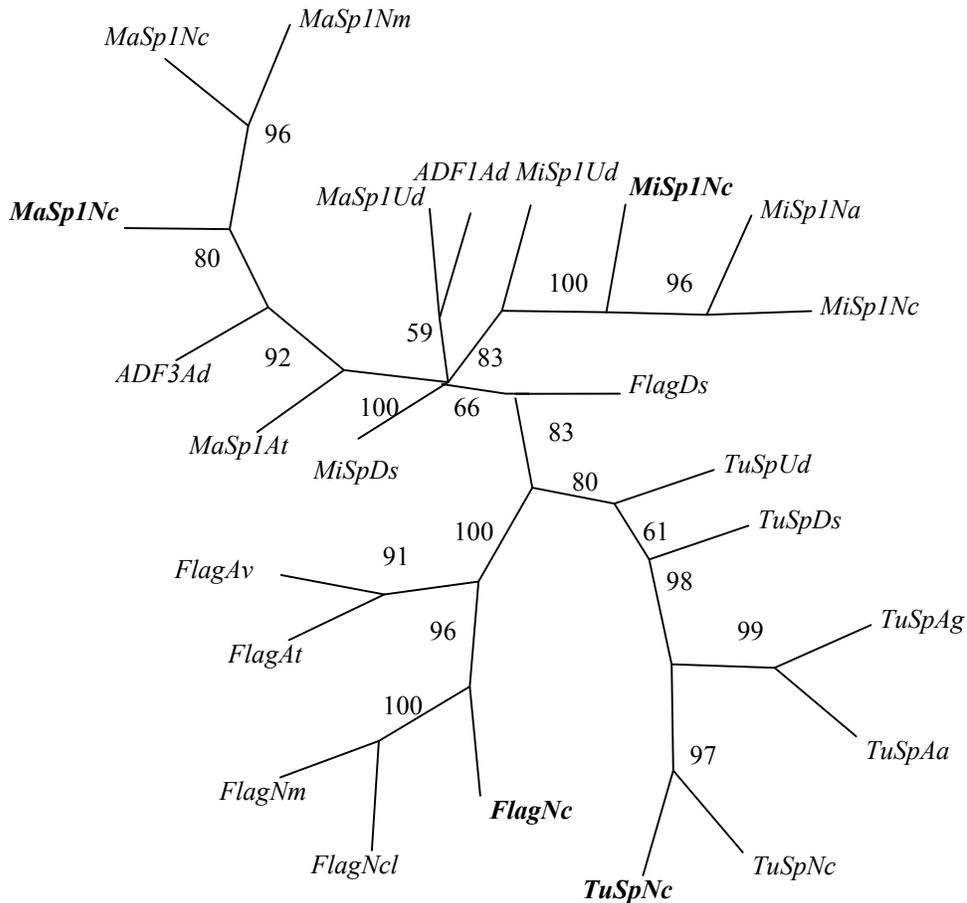
*Análise filogenética*

A análise de probabilidade máxima (ML) examinou a relação entre as seqüências de aminoácidos da região C-terminal a partir das espidroínas da *N. cruentata* e dos genes de espidroínas anteriormente informados a partir de diferentes glândulas sericígenas e espécies de aranha. O alinhamento da região C-terminal de todas as espidroínas identificadas com seqüências conhecidas de outras proteínas de seda gerado usando CLUSTALW encontra-se ilustrado na **Figura 5**. A fim de executar a análise filogenética, um alinhamento das seqüências de aminoácidos do C-terminal também foi executado em MAFFT (5.8) (dados não mostrados). A topologia resultante da análise de ML (- lnL: 16456.89) encontra-se descrita na **Figura 6**, e representada como uma árvore sem raiz. Os resultados mostraram que as espidroínas encontradas na biblioteca da *N. cruentata* pertencem à família de genes das espidroínas, e todas elas foram corretamente classificadas de acordo com seu grupo ortólogo.

**MaSp1-Ncruen** SRLSSPEASSRVSSAVS----NLVSSG---PTNSAALSNTISSVVSQISASNPGLSGCDVLVQALLEVVSALIHILGS 71  
 MaSp1-Nclavi SRLSSPQASSRVSSAVS----NLVSSG---PTNSAALSSTISNVVSQIGASNPGLSGCDVLIQALLEVVSALIQILGS 71  
 MaSp1-Nmadag SRLSSPQASSRVSSAVS----NLVSSG---PTNSAALSSTISNAVSQIGASNPGLSGCDVLIQALLEVVSALIHILGS 71  
 MaSp1-Atrifa SRLSSPGAASRVSSAVT----SLVSSGG---PTNSAALSNTISNVVSQISSNPGLSGCDVLVQALLEIVSALVHILGS 72  
 MaSp1-Udiver SRLQSPASSRVSSAVS----TLASAG---AANSALSSVISNLSSSVASAHFDLSCGELLVQILLEVISALVALLGS 71  
 ADF3-Adiade SRLSSPAASSRVSSAVS----SLVSSG---PTKHAALSNTISSVVSQVSAASNPGLSGCDVLVQALLEVVSALVSIILGS 71  
  
**MiSp1-Ncruen** SRLSSAQASSRISAAAS----TLISGG---YLNTSALPSVISDLFAQVSAASSPGVSDSEVLIQVLEIVSSLIHILSS 71  
 MiSp1-Nclavi SRLSSAEASSRISAAAS----TLVSGG---YLNTAALPSVISDLFAQVGAASSPGVSDSEVLIQVLEIVSSLIHILSS 71  
 MiSp1-Nantip SRLSTAASSRISTAAS----TLVSGG---YLNTAALPSVIADLFAQVGAASSPGVSDSEVLIQVLEIVSSLIHILSS 71  
 MiSp1-Udiver NRIVSAPAVNRMSAASS----TLVSNNG---AFNVGALGISTDMAAQIQAGSQGLSSAEATVQALLEVISVLTHTLSS 71  
 MiSp1-Dspino SRLASGQATDRVKDVVS----TLVSNNG---INGDALSNAINVMTQVNAAVPGLSFCERLIQVLEIVAALVHILSS 70  
 ADF1-Adiade NRLSSAGAASRVSSNVA----AIASAG-----AAALPNVISNIYSGVLSS--GVSSSEALIQALLEVISALIHVILGS 66  
  
**TuSp-Ncruen** SGLASSSATSRVGS LAQSLASALQSSG--GTLDVSTFLNLLSPISTQIQANTS--LNASQAIQVLEAVALLQIING 75  
 TuSp-Nclavi SGLSSASANARVSSLAQSFASALSASR--GTLVSVTFLTLLSPISSQIRANTS--LDGTQATVQVLEALALALQVINA 75  
 TuSp-Agemmo SGLASSAASARVSSLAQSIASAISSSG--GTLVSVPIFLNLLSSAGAQTASSS--LSSSQVTSQVLEGIALLQVING 75  
 TuSp-Aargen SGLGSSAASARVSSLANSVASAISSSG--GSLSVPTFLNFLSSVGAQVSSSSS--LNSSEVTNEVLEIAIALLQVLNG 75  
 TuSp-Udiver NGLSSSSASSRINSIASGLSTALSSSR---GVSLNLSLSSVFSSEIQNNSFGVSAEQALIQALFEVLVTGTQVQLNR 75  
 TuSp-Dspino AGLSSAAATSRASSLASSVASAISSAGSAGVDVGLFASGLSSLVSQIQSSNLGLQPDQVLEALLEGYALAQVLIS 78  
  
**Flag-Ncruen** -----SRVPDLVNGIMR-----SMQGS-----FNYQMFGNMLSKYASGSGACNS--NDNVLMDALLAALHCLSSH-GS 63  
 Flag-Nclavi -----SRVPDMVNGIMS-----AMQGS-----FNYQMFGNMLSQYSSGSGTCNP--NNNVLMDALLAALHCLSNH-GS 63  
 Flag-Nmadag -----SRVPDMVNGIMS-----AMQGS-----FNYQMFGNMLSQYSSGSGTCNP--NNNVLMDALLAALHCLSNH-GS 63  
 Flag-Atrifa -----ERLPNLINGIKS-----SMQGG-----FNYQNFNGNLSQYATGSGTCNY--YDINLLMDALLAALHTLNYQ-GA 63  
 Flag-Aventr -----SRLPSLVNGLMG-----SMQPTG-----FNYQNFNGNLSQYATGSGTCNS--NDNVLLMDALMAALHCLSYG-SG 63  
 Flag-Dspino SNLHSPSANVRVGNIVD----RISSGG---VGVMEILPRILSELYANIRESSPGMSDCERFMQVLLDIVSALMHVLLY 71  
  
 : : : : : : \*:  
**MaSp1-Ncruen** SSIG-PVNYGSASQSTQIVGQSVYQALG--- 98  
 MaSp1-Nclavi SSIG-QVNYGSAGQATQIVGQSVYQAL--- 97  
 MaSp1-Nmadag SSIG-QVNYGSAGQATQ----- 87  
 MaSp1-Atrifa ANIG-QVNSSGVRSASIVGQSIHQAFS--- 99  
 MaSp1-Udiver STVG-PVDIGQSSQYSGLVANAIGNALA--- 98  
 ADF3-Adiade SSIG-QINYGASAQYTMVQGSVAQALA--- 98  
  
**MiSp1-Ncruen** SSVG-QVDFNSVGSAAAVGQSMQVVMG--- 98  
 MiSp1-Nclavi SSVG-QVDFSSVGSAAAVGQSMQVVMG--- 98  
 MiSp1-Nantip SSVG-QVDFSSVGSAAAVGQSMQVVMG--- 98  
 MiSp1-Udiver ANIG-YVDFSRVGDASAVSQSMAYAG--- 97  
 MiSp1-Dspino SNVG-SIDYGSTSRTAIGVSNALASAVAGAF 100  
 ADF1-Adiade ASIG-NVSSVGVNSALNAVQNAVGAAYAG--- 93  
  
**TuSp-Ncruen** AQIT-SVNFSGVSSVNTALATALAG----- 99  
 TuSp-Nclavi AQIT-EVNVSNVSSANAALVSALAG----- 99  
 TuSp-Agemmo AQIR-SVNLANVPNVQQALVSALSG----- 99  
 TuSp-Aargen AQIT-SVNLARNV----- 86  
 TuSp-Udiver GQTS-FVSVSSPTVISSSF----- 93  
 TuSp-Dspino SQIS-SVSVSSSALGPALLNYLVG----- 102

<b>Flag-Ncruen</b>	<b>PSFGSSPTPSAMNAYSNSVRRMFQF</b> ----- 87
Flag-Nclavi	SSFAPSPTPAAMSAYSNSVGRMFAY----- 87
Flag-Nmadag	SSFAPSPTPAAMSAYSNSVGRMFAY----- 87
Flag-Atrifa	SYVPSYPSPSEMLSYTENVRRYF----- 85
Flag-Aventr	S-VPSTPTYSAMSAYNQSIIRRMFTY----- 86
Flag-Dspino	EDVRRGIPGDTAEAVANAVAGVVLSSVV----- 98

**Figura 5:** Alinhamento da região C-terminal das espidroínas de *N. cruentata* com espidroínas de diferentes espécies de aranha. Aminoácidos estão abreviados com uma letra e numerados no sentido N- para o C-terminal. Abreviações são as mesmas descritas para o alinhamento da figura 2. As seqüências de *N. cruentata* estão identificadas em negrito. *Araneus diadematus* ADF-3 foi adicionada como representante do grupo da MaSp2. As abreviações das espécies de aranha usadas nesta figura (de cima para baixo): 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*. Abreviações utilizadas para as espidroínas: MaSp1, espidroína ampolada principal 1; ADF3, fibroína 3 (espidroína ampolada principal 2); MiSp1, espidroína ampolada secundária 1; ADF1, fibroína 1; TuSp, espidroína tubuliforme; Flag, espidroína flageliforme. Acessos do GenBank: MaSp1-Nclavi (P19837), MaSp1-Nmadag (AAK30606), MaSp1-Atrifa (AAK30595), MaSp1-Udveri (ABD61596), ADF3-Adiade (AAC47010), MiSp1-Nclavi (AAC14589), MiSp1-Nantip (ABC72645), MiSp1-Udiver (ABD61597), MiSp1-Dspino (ABD61589), ADF1-Adiade (AAC47008), TuSp-Nclavi (AAX45295), TuSp-Agemmo (AAX45293), TuSp-Aargen (AAY28932), TuSp-Udiver (AAY28933), TuSp-Dspino (AAY28934), Flag-Nclavi (AAC38847), Flag-Nmadag (AAF36092), Flag-Atrifa (AAK30594), Flag-Aventr (AAT36347) and Flag-Dspino (ABD61590).



\_1

**Figura 6:** Árvore sem raiz da topologia resultante da análise de ML (- lnL: 16456.89) das seqüências da região C-terminal de várias espiroínas. As seqüências de *N. cruentata* estão identificadas em negrito. As abreviações das espécies de aranha usadas nesta figura: Ncr, *N. cruentata*; Ncla, *N. clavipes*; Nma, *N. i. madagascariensis*; Atr, *A. trifasciata*; Udi, *U. diversus*; Adi, *A. diadematus*; Nan, *N. antipodiana*; Age, *A. gemmoides*; Aar, *A. argentata*; Dsp, *D. spinosa*; Ave, *A. ventricosus*. Abreviações utilizadas para as espiroínas e números de acesso do GenBank são as mesmas utilizadas no alinhamento da figura 5.

## Discussão

Durante muitos anos, as sedas de aranhas têm suscitado o interesse da humanidade por causa de suas extremas propriedades mecânicas. Com os avanços na biotecnologia, surgiu a possibilidade de produzir novos materiais com base nos polímeros da seda de aranhas. Neste trabalho, identificamos diferentes genes de seda da aranha *N. cruentata*. Usando os cDNAs das glândulas sericígenas das aranhas pudemos identificar transcrições parciais das glândulas ampolada principal, ampolada secundária, flageliforme e tubuliforme de *N. cruentata*. Embora essa estratégia tenha demonstrado possuir suas limitações para obter transcrições completas da seda de aranhas, ela é usada com frequência para identificar novos genes da seda de aranhas (Hayashi *et al.*, 2004; Lawrence *et al.*, 2004; Pouchkina-Stantcheva e McQueen-Mason, 2004; Tian *et al.*, 2004).

De acordo com as seqüências previamente descritas que codificam as proteínas de seda de aranhas (Xu e Lewis, 1990; Hinman e Lewis, 1992), o uso de códon para os aminoácidos mais abundantes nas espidroínas da *N. cruentata*, com a exceção da NCTuSp-*like*, seguem a preferência por adenina (A) e timina (T) como a terceira base de três que codifica cada aminoácido. Esta diferença nos usos de códon entre as espidroínas também foi encontrada nas seqüências codificadoras de TuSp1 da *Araneus gemmoides* e da *Nephila clavipes* (Tian e Lewis, 2005). A prevalência de T e A nas seqüências que codificam as espidroínas NCMaSp1-*like*, NCMiSp1-*like* e NCFlag-*like* ocorre no intuito de evitar a formação de numerosos grampos nas regiões próximas ricas em guanina e citosina presentes nos codons de suas seqüências repetitivas (poli-Ala, (GlyAla)<sub>n</sub>, (GlyGlyX)<sub>n</sub>, (GlyProGlyXX)<sub>n</sub>) (Hinman e Lewis, 1992). Tais regiões não existem na proteína NCTuSp-*like*.

Como as proteínas de sedas de aranha anteriormente descritas (Xu e Lewis, 1990; Gatesy *et al.*, 2001; Tian *et al.*, 2004; Lewis, 2006), nossos achados também demonstraram proteínas com alto teor dos aminoácidos alanina e glicina devido à presença de motivos altamente repetitivos. A proteína análoga NCMaSp1 é composta de repetições similar com motivos poli-Ala e (GlyGlyX)<sub>n</sub>, onde os resíduos de X são Ala, Tyr, Leu ou Gln. Seus blocos repetitivos são muito semelhantes àqueles encontrados nas

proteínas da *N. clavipes* e da *Nephila inaurata madagascariensis* com uma identidade de 96%. Como todas essas aranhas pertencem à família Nephilidae, a semelhança não é totalmente inesperada. Entretanto, as repetições da *N. cruentata* variam no comprimento e na quantidade de motivos GlyGlyX. Foi proposto que esses motivos adotam uma estrutura secundária na forma de hélice proporcionando elasticidade. Usando a microscopia FTIR (infravermelho com transformada de Fourier) foi observado que a estrutura secundária das espidroínas da glândula ampolada principal possui predominantemente estruturas helicoidais (38%) em comparação com estruturas em folhas e voltas (Winkler e Kaplan, 2000; van Beek *et al.*, 2002; Dicko *et al.*, 2004). Por outro lado, essa hélice é muito rígida para ser elástica; é mais provável que os motivos (GlyGlyX)<sub>n</sub> sejam responsáveis por outra estrutura proteica similar à folha-β. Embora as sedas da ampolada principal apresentem um alto teor de motivos (GlyGlyX)<sub>n</sub>, elas são conhecidas por suas alta resistência à tração e robustez proporcionadas pelo motivo poli-Ala, com valores de força entre 1-2 GPa (Gosline *et al.*, 1999; Gosline *et al.*, 2002; Blackledge e Hayashi, 2006). Em contrapartida, a seda da glândula ampolada secundária possui uma resistência diminuída, mas uma maior extensibilidade em comparação com as sedas da principal, o que pode estar relacionado com a menor presença de motivos poli-Ala e maior quantidade de (GlyAla)<sub>n</sub> e (GlyGlyX)<sub>n</sub> em seus blocos repetitivos (Blackledge e Hayashi, 2006). Esses motivos também foram encontrados na seqüência da espidroína NCMiSp1-like. Embora suas repetições tenham demonstrado organização muito similar com as espidroínas da ampolada secundária descritas anteriormente, o número de repetições (GlyAla)<sub>n</sub> é bastante variável entre essas proteínas. Entretanto, a região espaçadora de NCMiSp1-like constituído por uma seqüência não repetitiva de 124 aminoácidos é quase idêntica a da MiSp de *N. clavipes* e *N. i. madagascariensis*, com poucas substituições e deleções de aminoácidos (89% e 91% de identidade, respectivamente). A principal função dessa região permanece desconhecida, mas acredita-se que possa servir para separar as regiões cristalinas, como também para participar nas associações entre as cadeias proteicas através de resíduos carregados (Lewis, 2006).

A proteína NCFlag-like também é composta por uma seqüência altamente repetitiva formada pelos motivos (GlyProGlyXX)<sub>n</sub> (X representa Ala, Val, Ser e Tyr) e

três motivos GlyGlyX (X representa Ala, Ser e Thr), separados por uma seqüência não repetitiva constituída por aminoácidos carregados e hidrofílicos. A seqüência da NCFlag-like também é muito similar com outras espidroínas da glândula flageliforme previamente descritas (Hayashi e Lewis, 2000). A seda da glândula flageliforme é a seda mais elástica (Blackledge e Hayashi, 2006), e é responsável pela formação da espiral de captura na teia orbicular. Foi proposto que o motivo GlyProGlyXX possua uma estrutura secundária similar a uma espiral- $\beta$  similar a uma mola que poderia facilmente contribuir para o mecanismo elástico da fibra, com as ligações entre prolinas gerando força para a retração da seda após seu estiramento (Hayashi *et al.*, 1999).

As sedas das glândulas tubuliforme e da aciniforme são únicas entre as espidroínas de aranhas devido a sua complexa composição de aminoácidos, e a proteína NCTuSp-like não é diferente. Ela é composta de grandes repetições ricas em aminoácidos alanina e serina com vários novos motivos ( $\text{Ser}_n$ ,  $(\text{SerX})_n$ , e GlyX). Em contraste com outras espidroínas, NCTuSp-like contém altas quantidades de serina e baixas quantidades de glicina. Além disso, ela também apresenta uma maior quantidade de aminoácidos com cadeias laterais grandes como valina e leucina. As predições de estruturas secundárias das espidroínas da glândula tubuliforme usando difração de raio-X indicaram a presença de grandes quantidades de folha- $\beta$ . A difração de raio-X também mostrou que a seda do casulo tem um maior valor *b* dimensional na folha- $\beta$  que as sedas da ampolada principal e secundária, o que indica a presença de aminoácidos de cadeia lateral grande (Parkhe *et al.*, 1997). Os dados de difração em TEM (microscopia eletrônica de transmissão) também concordam com os obtidos pela difração de raios-X (Barghout *et al.*, 1999). A falta das repetições habituais ricas em motivos de alanina e/ou glicina encontradas na maioria das espidroínas de aranhas está provavelmente relacionada com a função. As espidroínas da tubuliforme são usadas para construir as estruturas para a reprodução, diferentemente das sedas da ampolada principal e secundária ou da flageliforme, que servem como fibras estruturais para a captura de presas. Porém, dados mecânicos recentes de sedas da glândula tubuliforme da *Argiope argentata* mostraram que essas fibras também possuem boas propriedades em comparação com as estruturais, com um valor de resistência de 0,47 GPa (Blackledge e Hayashi, 2006).

O alinhamento das unidades repetitivas da NCTuSp-*like* com as repetições da tubuliforme de diferentes espécies de aranhas demonstrou uma alta similaridade entre elas. Surpreendentemente o bloco repetitivo da *N. cruentata* foi mais semelhante às repetições da *A. gemmoides* e da *A. argentata* do que as repetições da *N. clavipes*, mesmo que pertençam a famílias diferentes de acordo com evidências morfológicas. NCTuSp-*like* também mostrou semelhanças com a Fibroína 1 da aranha Mygalomorphae *Euagrus chisoseus* ([AF350271](#)). As espidroínas da tubuliforme, previamente descritas, exibiram um extraordinário grau de homogeneidade entre as consecutivas repetições intragênicas (Garb e Hayashi, 2005; Tian e Lewis, 2005). Entretanto, não conseguimos verificar esta evidência de evolução coordenada na NCTuSp1-*like* por causa do pequeno tamanho do transcrito obtido na biblioteca de cDNA que mostrou apenas uma repetição completa.

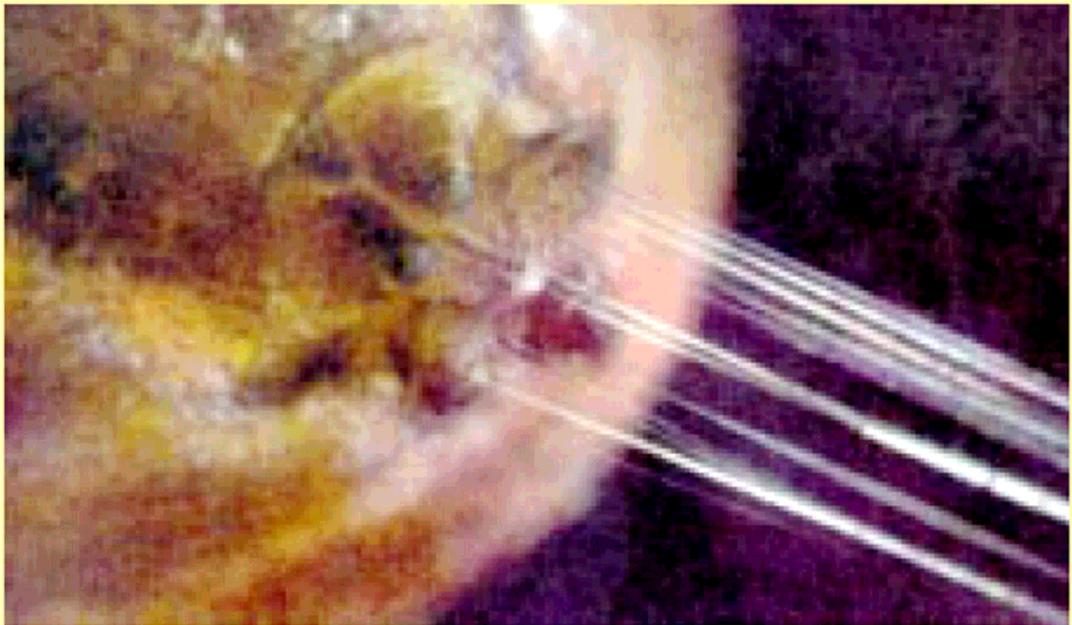
A região C-terminal não repetitiva e altamente conservada também foi identificada em todas as espidroínas descritas de *N. cruentata*. O alinhamento da seqüência de aminoácidos da região C-terminal descrito para essas espidroínas com aquelas de espidroínas previamente descritas mostrou que esta é uma região altamente conservada entre as espécies. A maioria das proteínas de seda de aranhas compartilha 30% de identidade entre a sua região C-terminal, sendo as flageliformes as mais divergentes. Porém, todas as seqüências compartilham uma determinada região conservada na seqüência de aminoácidos QALLE, até mesmo entre as aranhas tecedoras de teia orbicular e não orbicular. Essa região corresponde àquela com maior hidrofobicidade na região C-terminal, e as predições de estrutura secundária sugerem que essa região QALLE é também responsável pela formação de  $\alpha$ -hélices (Spooner *et al.*, 2005; Challis *et al.*, 2006). Uma vez que a seqüência da região C-terminal é muito mais conservada que a região repetitiva entre as diferentes espécies de aranha, sugere-se que ela possa desempenhar um importante papel, de maneira que os motivos de aminoácidos importantes para o desempenho de uma determinada função foram preservados pela seleção natural. Várias funções foram atribuídas a região C-terminal das proteínas de sedas. Ela pode ser responsável pela formação de micelas de tamanho irregular na solução de fiação a fim de evitar a formação prematura da fibra (Jin e Kaplan, 2003) ou ter uma função no correto dobramento das proteínas, uma vez que foi demonstrado que a

região C-terminal fica retida na fibra polimerizada (Sponner *et al.*, 2004; Ittah *et al.*, 2006).

Em resumo, foram estudadas as diferentes espidroínas da seda das aranhas brasileiras *N. cruentata*. Conseguimos identificar as seqüências de proteínas das sedas responsáveis pela construção da teia orbicular e pela captura de presas (MaSp, MiSp e Flag), como também pela construção do casulo (TuSp) de *N. cruentata*. Durante todo o nosso estudo foi possível demonstrar um alto grau de semelhança entre essas seqüências e as seqüências de outras espécies de aranhas. Similaridades também foram observadas entre as seqüências de diferentes grupos de genes, com a exceção da espidroína da tubuliforme de *N. cruentata*, que desempenha uma função distinta na vida da aranha (Hu *et al.*, 2005a; Tian e Lewis, 2005). O estudo adicional mecânico e estrutural das diferentes sedas de *N. cruentata* propiciará importantes informações sobre o argumento de que as propriedades mecânicas das diferentes espidroínas estão correlacionadas com suas seqüências de aminoácidos (Hayashi *et al.*, 1999; Rising *et al.*, 2005).

## Capítulo 3

### Propriedades mecânicas da fibra da glândula Ampolada principal de *N. cruentata*



## Introdução

A fibra produzida pela glândula ampolada principal, também conhecida pelo nome da glândula que a produz, utilizada como linha de segurança e moldura estrutural para teia, é a mais estudada entre todos os tipos de sedas produzidas por aranhas devido às suas excepcionais propriedades mecânicas. A seda ampolada principal de aranhas orbiculárias são conhecidas por serem super-resistentes com dureza superior à do aço e até mesmo do kevlar (Gosline *et al.*, 1999). Estudos moleculares da MaSp de *N. cruentata* (NCMaSp1-like, Capítulo 2) mostraram que, semelhante a outras MaSp1 descritas anteriormente em diferentes espécies de aranhas, esta proteína também é composta por módulos repetitivos constituído pelos motivos de aminoácidos poli-Ala e (GlyGlyX)<sub>n</sub>. Tais motivos estruturais são considerados como os responsáveis por proporcionar as características mecânicas à fibra (Hayashi *et al.*, 1999; Rising *et al.*, 2005).

No intuito de estabelecer uma comparação entre os motivos de aminoácidos encontrados na seqüência de NCMaSp1-like e suas propriedades mecânicas, curvas de tensão/tração foram obtidas, sendo que “tensão” (GPa) define a força máxima exercida na fibra momentos antes do seu rompimento, e “tração” (%) a extensibilidade da fibra. Os valores do módulo de Young (GPa) ou rigidez também foram quantificados pelo declive do segmento linear inicial da curva de tensão/tração antes do primeiro ponto de rendimento ou transição de fase da fibra. A análise da fibra por meio de microscopia de força atômica também foi realizada no intuito de esclarecer a organização nanoestrutural da fibra e adicionar dados para o desenvolvimento de hipóteses que correlacionem não apenas a seqüência primária da proteína, mas também a estrutura tri-dimensional com as propriedades mecânicas da seda.

A partir do momento em que houver uma melhor compreensão da relação entre estrutura molecular e função da fibra de aranha, será possível determinar se a seda ampolada principal proporciona ou não um bom modelo para a produção de biomateriais por meio da engenharia genética.

## Resultados

### *Diâmetro da fibra*

As fibras da glândula ampolada principal de *N. cruentata* possuíam em média  $4,58 \pm 0,73 \mu\text{m}$  de diâmetro. A fibra #1 da aranha N1 (N1masp#1) teve o maior diâmetro observado de  $5,90 \mu\text{m}$ , entretanto a aranha N2 apresentou a fibra com menor diâmetro, a fibra #2 (N2masp#2) possuía um diâmetro de  $3,26 \mu\text{m}$  (**Tabela 1**).

Uma análise geral mostra que a média dos diâmetros encontrados para as fibras da aranha N1 foi maior do que para a aranha N2,  $4,93 \pm 0,60\mu\text{m}$  e  $4,04 \pm 0,60\mu\text{m}$  respectivamente. Claramente, observa-se uma variação entre as os diâmetros das fibras da mesma aranha e entre as aranhas, por isso a necessidade de utilizar várias amostras de diferentes aranhas para a obtenção de análises mecânicas mais próximas do valor médio.

**Tabela 1**

*Resultados dos testes mecânicos das fibras de duas aranhas N. cruentata.*

Amostra	Diâmetro ( $\mu\text{m}$ )	Tensão (GPa)	Tração (%)	Rigidez (GPa)
N1masp1	5,90	1,25	28,23	7,45
N1masp2	4,53	1,74	33,30	13,60
N1masp3	4,86	1,08	33,87	3,71
N1masp4	5,60	1,38	24,89	12,50
N1masp5	5,10	1,42	33,98	8,74
N1masp6	4,06	1,92	10,73	17,90
N1masp7	5,26	1,43	45,12	3,80
N1masp8	4,86	1,46	33,15	8,55
N1masp9	4,26	1,33	33,11	9,24
N2masp1	4,13	0,59	21,16	5,43
N2masp2	3,26	2,15	27,47	22,80
N2masp3	4,13	1,19	22,69	15,00
N2masp4	3,73	1,83	28,63	11,10
N2masp5	3,93	1,24	17,28	19,00
N2masp6	5,10	1,13	24,10	16,90
MÉDIA	4,58	1,40	27,84	11,71
DESVIO PADRÃO	0,73	0,38	8,26	5,76

**Tabela 2**

*Propriedades mecânicas (valores de engenharia) da seda ampolada principal de diferentes espécies de aranha*

	RIGIDEZ (GPa)	TENSÃO (GPa)	TRAÇÃO (%)
<i>N. cruentata</i>	11,71 ± 5,76	1,40 ± 0,38	27,84 ± 8,26
<i>N. clavipes</i> *	13,8 ± 0,76	1 ± 0,004	20 ± 1,1
<i>A. trifasciata</i> *	8,2 ± 0,63	1,2 ± 0,003	23 ± 0,6
<i>A. diadematus</i> *	8.3 ± 0.54	1,06 ± 0,005	29 ± 2,4
<i>L. geometricus</i> **	12,91 ± 7,38	0,83 ± 0 ,19	14 ± 6

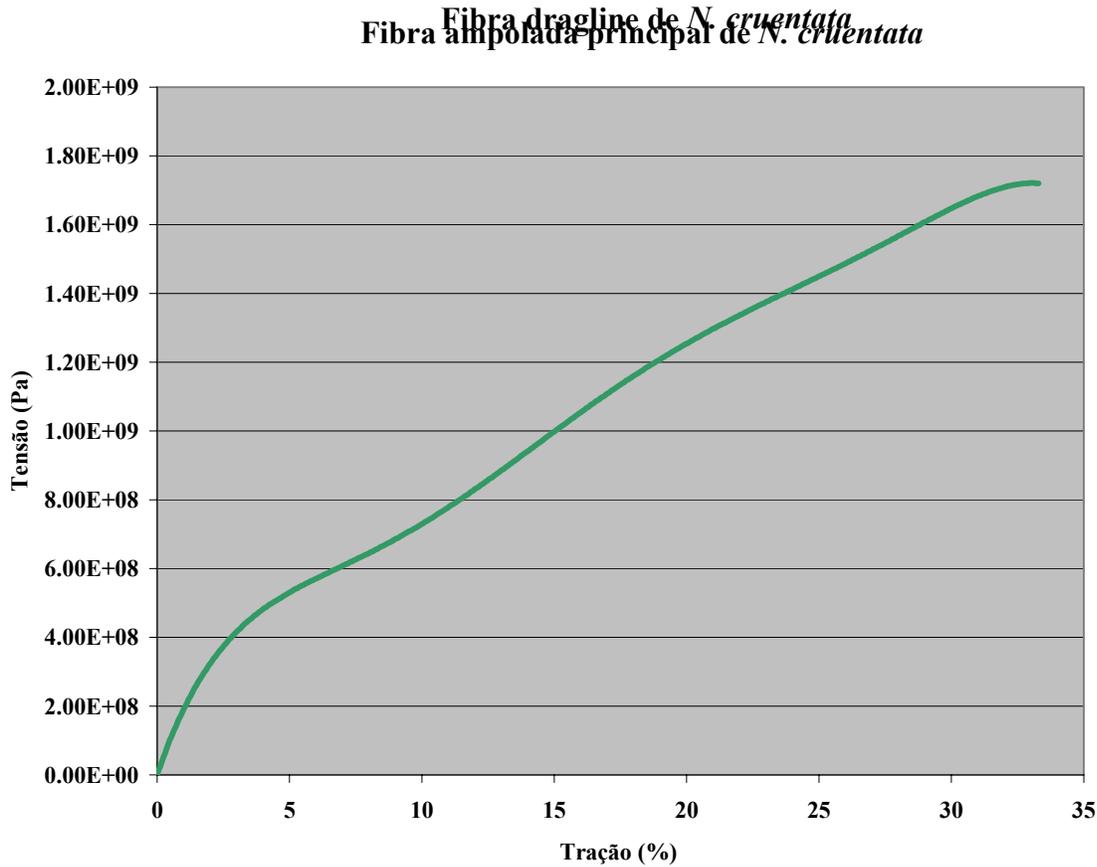
\*Swanson *et al.*, 2006

\*\*Motriuk-Smith e Lewis, 2004

### *Testes mecânicos*

Os valores de engenharia obtidos para a tensão das amostras analisadas, variaram de 0,59 GPa para a fibra N2masp1 até 2,15 GPa para N2masp2, o valor mais elevado para a aranha N1 foi de 1,92 GPa. A média de força para todas as fibras testadas foi de 1,40 ± 0,38 GPa. Os valores de engenharia para a tração em porcentagem, isto é quanto a fibra se estendeu em relação ao seu comprimento inicial, foi em média de 27,84 ± 8,26 %. O menor valor encontrado para a tração da fibra foi de 10,73 % para N1masp6 e o maior valor foi de 33,89% para a amostra N1masp5. O valor médio da rigidez (módulo de Young) encontrado para as amostras analisadas foi de 11,71 ± 5,76 GPa, o menor e o maior valor encontrado foi 3,71 GPa e 22,80 GPa, para N1masp3 e N2masp2, respectivamente. A correlação existente entre os valores de tensão e tração obtidos pelo teste mecânico da fibra ampolada principal de *N. cruentata* (N1masp2 – desempenho mecânico com valores similares às médias obtidas), pode ser visto na **Figura 1**. Os valores reais de tensão ( $\sigma_{tr}$ ) e tração ( $\epsilon_{tr}$ ) também foram calculados a partir dos valores de engenharia, onde a média dos respectivos valores encontrados foram 1,79 ± 0,49 GPa e 10,42 ± 0,02 %.

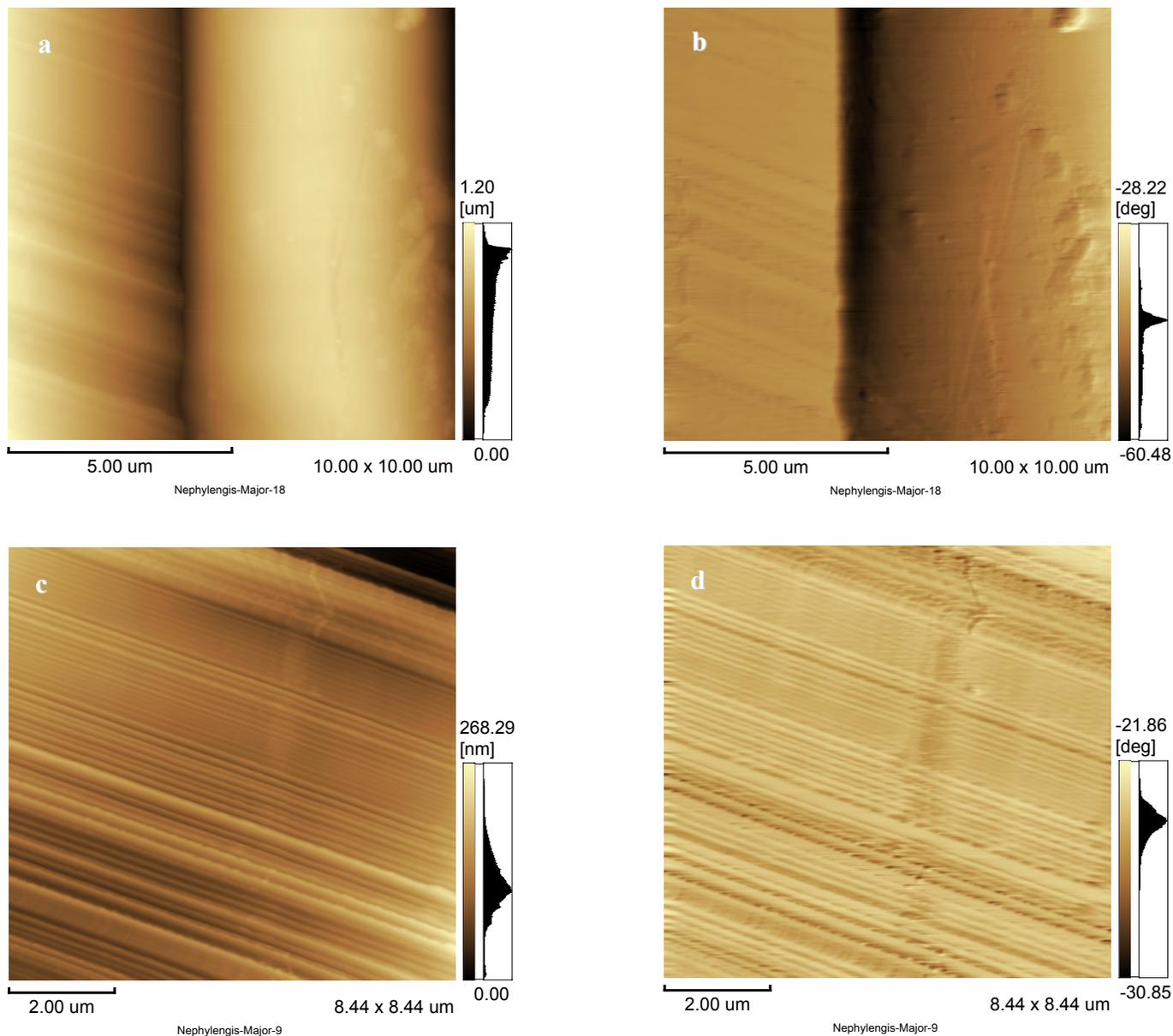
A **Tabela 2** também mostra os valores obtidos pela análise mecânica das fibras de *N. cruentata* em comparação com os valores obtidos da seda ampolada principal de diferentes espécies de aranha.



**Figura 1:** Desempenho mecânico da fibra ampolada principal de *N. cruentata*.

#### *Microscopia de força atômica*

Por meio da análise da fibra ampolada principal utilizando a microscopia de força atômica (MFA) foi possível observar que esta é uma fibra cilíndrica, constituída por nanofibras arranjadas obliquamente ao seu eixo longitudinal (**Figura 2**). Foi também possível observar a intercalação de nanofibras mais macias (escuras), e mais duras (claro), e a presença de imperfeições ao longo da mesma (**Figura 2b, 2d**).



**Figura 2:** Imagens obtidas da fibra ampolada principal de *N. cruentata* por meio da microscopia de força atômica. (a, c) Topografia da fibra ampolada principal de *N. cruentata* em superfície de grafite investigada por meio da microscopia de força atômica no modo contato. (b, d) Imagens da modulação de força das mesmas regiões de topografia obtidas “a” e “c”, respectivamente. As regiões mais macias da fibra são mostradas em preto.

## Discussão

A análise da fibra ampolada principal de *N. cruentata* por meio da MFA mostrou que esta é composta por regiões (nanofibras) com diferentes graus de dureza (regiões macias alternadas por regiões mais rígidas), o que indica diferentes interações moleculares entre as proteínas constituintes da fibra ou diferenças nas composições das nanofibras. Tal resultado vai de acordo com o modelo aceito da estrutura molecular adotada pelas espidroínas de modo a formar um composto de fibra nano-estruturado. De acordo com os motivos de aminoácidos presentes nas proteínas constituintes da Ampolada principal, folhas- $\beta$  formadas pelos motivos ricos em resíduos de alanina estariam incorporados em uma matriz móvel, um tanto amorfa, constituída pelos motivos  $(\text{GlyGlyX})_n$  (Termonia, 1994). Oroudjev *et al.* (2002) também propôs um modelo similar para a organização nanoestrutural de fibras sintéticas baseadas na seqüência de aminoácidos de MaSp1 produzidas por *E. coli*.

Os resultados obtidos por meio do teste mecânico da fibra ampolada principal de *N. cruentata* também concorda com os dados da MFA, uma vez que esta possui duas características extremas, força e elasticidade, também atribuídas a diferentes estruturas moleculares presentes na fibra (Hayashi *et al.*, 1999). Entretanto, nenhum tipo de estrutura molecular pôde ser determinada claramente por meio da resolução obtida na MFA.

Trabalhos anteriores discutem a presença de uma fina camada presente ao redor da seda ampolada principal de *N. clavipes* caracterizando a “pele” da fibra (Work, 1985; Augsten *et al.*, 2000). Dessa forma, pode-se especular que a região mais escura longitudinal, e provavelmente mais macia, vista na fibra de *N. cruentata* nas imagens de MFA corresponda a esta estrutura. Li *et al.* (1994) também observou uma potencial região que caracterizaria a “pele” da fibra em resoluções obtidas em MFA de cortes transversais ( $45^\circ$ ) da fibra ampolada principal de *N. clavipes*.

Imperfeições na fibra também podem ser vistas na análise por MFA, o que não é surpreendente. Apesar das extremas propriedades mecânicas das fibras de teia de aranha, aranhas são inconsistentes tecedoras. Tal afirmação pode ser confirmada apenas pela observação da variação do diâmetro e propriedades mecânicas das fibras tecidas por

diferentes aranhas da mesma espécie, e até mesmo pela mesma aranha (Madsen *et al.*, 1999; Motriuk-Smith e Lewis, 2004, Brooks *et al.*, 2005). O diâmetro das fibras coletadas tanto da aranha N1 quanto da N2, apresentaram uma variação de  $\pm 0,6\mu\text{m}$  de diâmetro entre elas. Da mesma forma, a fibra #6 da aranha N1 era mais forte do que a amostra de #7, entretanto esta apresentou uma capacidade de tração muito maior do que a anterior. A variação das propriedades mecânicas entre fibras produzidas por diferentes indivíduos também foi observada em *N. cruentata*, com uma variação de até  $\pm 0,38$  GPa na força máxima exercida para o rompimento da fibra. Entretanto, a possibilidade destas imperfeições serem produto da manipulação da fibra durante o preparo da amostra para a análise não é totalmente descartado.

A seda ampolada principal de *N. cruentata* apresentou excelentes propriedades mecânicas, embora seja mais forte do que as de *A. diadematus*, *A. trifasciata* e *N. clavipes*, a fibra de *N. cruentata* é menos elástica do que a fibra das duas primeiras, entretanto a de *N. clavipes* ainda é mais rígida do que todas (Motriuk-Smith e Lewis, 2004; Swanson *et al.*, 2006). Elasticidade é definida como sendo a habilidade de deformação reversível de uma fibra sem a perda de energia, ou a habilidade de deformação em valores altos de tração, mas baixos de tensão. Desta maneira, proteínas elásticas devem possuir baixos valores de rigidez (módulo de Young ou módulo de elasticidade) (Gosline *et al.*, 2002). De uma maneira geral, podemos concluir que a seda ampolada principal de *N. cruentata* possui propriedades mecânicas com ordem de magnitude similar à de outras espécies de aranhas, inclusive não orbiculares (Blackledge e Hayashi, 2006; Swanson *et al.*, 2006).

Em contraste com a tensão e tração de engenharia, pode-se também calcular os reais valores de tensão e tração, os quais seriam uma medida mais precisa das forças e, portanto, devem ser usados para todas as comparações dos dados do teste mecânico (Blackledge e Hayashi, 2006; Guinea *et al.*, 2006). O cálculo da real tensão necessita conhecer a área da seção transversal instantânea em cada valor da extensão da fibra, entretanto, a suposição de volume constante da mesma durante o teste de tensão foi recentemente confirmada de maneira experimental (Guinea *et al.*, 2006), possibilitando o cálculo dos valores de  $\sigma_{tr}$  e  $\varepsilon_{tr}$  utilizando simples fórmulas (Capítulo 1). Os valores médios de  $\sigma_{tr}$  e  $\varepsilon_{tr}$  encontrados para a fibra de moldura de *N. cruentata* mostram que esta

ainda é mais forte do que previsto pelo cálculo de engenharia, entretanto o valor real de tração mostra que esta possui uma extensibilidade menor do que a apresentada pela tração de engenharia.

A produção de novos materiais baseados na estrutura da ampolada principal de aranha é o objetivo final da maioria dos estudos sobre esta fibra. Este trabalho forneceu indícios sobre as relações entre estrutura e função desta fibra, entretanto o conhecimento das propriedades mecânicas por si só não nos dá certeza destas relações. Análises físicas mais aprofundadas deste material devem ser realizadas para a melhor compreensão do comportamento mecânico desta fibra, para então sim, identificar o *design* responsável pelas propriedades deste extremo biomaterial.

## CAPÍTULO 4

### Espidroínas da aranha

#### *Avicularia juruensis*



**Este capítulo foi baseado no artigo a ser submetido para publicação (Anexo 4):**

Bittencourt, D., Dittmar, K., Lewis, R.V., Rech, E.L. How old are ampolada principal silks?

## Introdução

As aranhas (Araneae) consistem de duas linhagens principais, a saber, a Mesothelae, e a Opisthothelae, que contém a Mygalomorphae, (caranguejeiras e semelhantes), e a Araneomorphae (aranhas verdadeiras). Recente prova filogenética mostra nitidamente a relação de monofilia e de grupo-irmão das infraordens migalomorfas e araneomorfas (Coddington *et al.*, 2004; Hedin e Bond, 2006, Ayoub *et al.*, 2007). Os cálculos da divergência de tempo estimam suas respectivas origens de 392 até 340 milhões de anos atrás (Ayoub *et al.*, 2007).

Todas as aranhas produzem seda, mas usam grupos de proteínas da seda de tarefas específicas para tecer as teias, construir casulos ou dispersar (Garb e Hayashi, 2005). Acredita-se que a síntese de diferentes proteínas das sedas esteja intimamente ligada a diferenciação morfológica das glândulas sericígenas; motivo pelo qual as sedas são normalmente classificadas de acordo com sua glândula de origem (Vollrath, 1992, Challis *et al.*, 2006). As migalomorfas possuem apenas um ou dois tipos de glândulas, elas não constroem teias orbiculares, e acredita-se que tecam apenas um tipo de seda "ancestral" (Vollrath, 1992). Por outro lado, as aranhas de fiação orbicular (Araneomorphae: Orbiculária) podem ter até sete glândulas morfológicamente distintas que, conseqüentemente, sintetizam tipos diferentes de seda. Por exemplo, considera-se que três tipos de glândulas sejam responsáveis pela produção de fibras usadas na teia orbicular para a captura de presas: as glândulas ampoladas principais (espidroínas da ampolada principal 1 e 2 - MaSp), as glândulas ampoladas secundárias (espidroínas da ampolada secundária 1 e 2 - MiSp) e as glândulas flageliformes (espidroínas das flageliformes - Flag) (Gatesy *et al.*, 2001). Embora as proteínas análogas de seda MaSp1 sejam conhecidas a partir de aranhas não orbiculárias (Tai *et al.*, 2004, Tian *et al.*, 2004), a presença de MaSp, MiSP e Flag na base da divergência *Deinopidae* (tecedoras orbiculares com cribelo)- *Araneidae* (tecedoras orbiculares sem cribelo) seja interpretada como apoio pela origem única das aranhas orbiculares (Garb *et al.*, 2006).

Porém, a maioria dos estudos com referência às características moleculares das sedas de aranhas é baseada na Orbiculariae (Araneomorphae), apenas algumas seqüências são conhecidas das espidroínas de migalomorfas (Gatesy *et al.*, 2001, Tai *et*

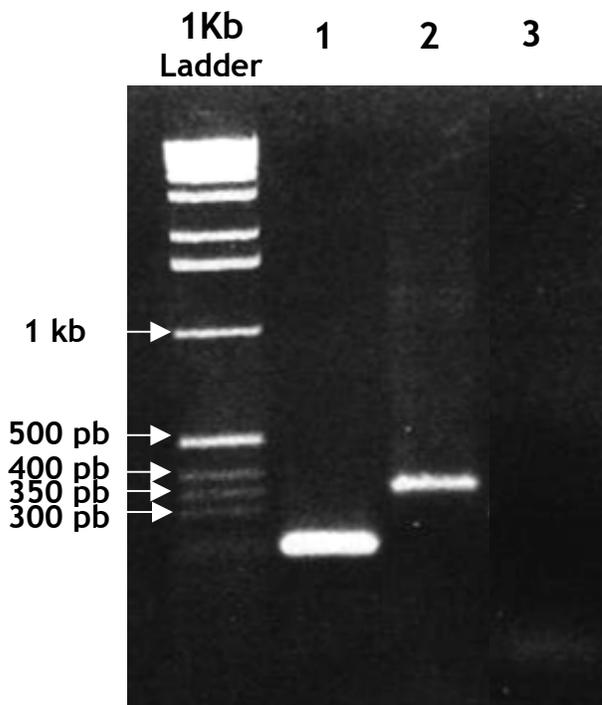
al., 2004). Aqui relatamos as seqüências de cDNA de dois genes de seda de uma aranha migalomorfa (*Avicularia juruensis*) encontrada na região da Amazônia, Brasil. Embora um gene seja típico de outras espidroínas isoladas de migalomorfos (por exemplo, *Euagrus chisoseus*), o outro, curiosamente, mostra uma clara relação entre as sedas da ampolada principal.

## Resultados

### *Biblioteca de cDNA*

Apesar da presença de apenas uma ou duas glândulas sericígena na *A. juruensis*, interessantemente, foi possível agrupar os trinta e quatro clones positivos, entre os 1.248 clones analisados na biblioteca de cDNA, em dois grupos distintos. O transcrito mais abundante foi chamado de *Espidroína 1* (3.154 pb), e o segundo cDNA de espidroína foi denominado *Espidroína 2* (1.874 pb) (Anexo 2). Todos os clones de cDNA seqüenciados foram parciais na direção 5' e como as outras espidroínas suas traduções caracterizou unidades repetidas de aminoácidos seguidas por uma região C-terminal não repetitiva. Foram descobertas três transcrições supostamente parálogas para o gene da *Espidroína 1* (*Espidroína 1A* - 2 clones, *1B* - 9 clones e *1C* - 17 clones). Outros cinco clones contendo apenas a seqüência repetitiva da *Espidroína 1* também foram identificados. O maior número de clones representando a *Espidroína 1* em comparação com o gene da *Espidroína 2* (somente um clone), sugere que esse gene seja mais expresso.

A fim de confirmar a expressão de ambas as espidroínas na glândula sericígena da *A. juruensis*, a RT-PCR foi realizada usando-se iniciadores flanqueando a região C-terminal das transcrições da *Espidroína 1* (219 pb) e da *Espidroína 2* (356 pb) (**Figura 1**). Todos os cDNAs analisados foram positivos devido a sua presença na(s) glândula(s) sericígena(s) (Figura 1), eliminando a possibilidade de contaminação de transcritos ou recombinação de clone.



**Figura 1:** RT-PCR com RNA total da(s) glândula(s) sericígena(s) de *A. juruensis*. 1, Espidroína 1 cDNA (219 pb); 2, Espidroína 2 cDNA (356 pb); 3, controle negativo. 1Kb DNA ladder (Invitrogen).

#### *Análise das seqüências*

Entre todos os transcritos analisados foi possível identificar traduções para duas proteínas distintas (**Figura 2**). A seqüência de aminoácidos da Espidroína 1 apresentou uma seqüência longa repetitiva (183 pb através da tradução do transcrito *1C*) rica em serina e alanina, com uma cadeia de treoninas. Entretanto, a Espidroína 2 apresentou uma seqüência altamente repetitiva com os motivos (GlyAla)<sub>n</sub> e (GlySer)<sub>n</sub>. Ambas proteínas também possuíam uma região C-terminal não repetitiva altamente conservada em comparação com espidroínas descritas previamente (**Figuras 3 e 4**).

## Espidroína 1

```
YSLASSIASAASSASSAAAASSSSAAAGAAAASEAAASAAATSTTTTTSTSRAAAASAAAAASASGAAG
AAGAASAASAASASSLQQSLGSALAQSSSFAAAFAQASSAASAAAIAYALAQTVANQIGFSSYSSAFARAA
SSAVYSIGGLASASAYAFAFASAFSQVLSNYGLLNINNA
```

## Espidroína 2

```
GAG (A/S) GSGSGSGS
```

**Figura 2:** Região repetitiva das Espidroínas 1 e 2 expressas pela *A. juruensis*. Os motivos de aminoácidos encontrados nas espidroínas estão representados como: verde-A<sub>n</sub>, vermelho-GA e azul-GS.

```
1A      GLLPPLFVLPNSATERISSMVSSLLSAVSSNGLDASSFGDTIASLVSQISVNNSDLSSS 60
1B      GLLPPLSILPDSANERISSVVSSLLAAVSSNGLDASSLGDNLASLVSQISANNADLSSS 60
1C      NLLPPLSVLPDSANERISSVVSSLLSAISSNGLDASSLGGTIASLVSQISVSNAKLSSS 60
      .***** :***:*.*****:*****:*.*****:*.*****:*.*****:*.*****
1A      QVLEALLEILSGMVQILSYAEVGTVNTKTVSSTSAAVAQAISSAFSGNQNS 112
1B      QVMVEALLEVLSGIVQILSYAEVGAVNTETVSSTSSAVAQAISSAVLG---- 108
1C      QVFLEALLEVLSGMVQILSYAEVGAVNTDTVISTSSAVAQAISSAVSG---- 108
      **..*****.***.*****.*** ** ***.***** *
```

**Figura 3:** Alinhamento da região C-terminal das três traduções supostamente parálogas para o gene da *Espidroína 1* (1A, 1B e 1C) de *A. juruensis*. Aminoácidos estão abreviados com uma letra. “-”, indica “gaps” nas seqüências opara um melhor alinhamento; “\*”, indica que o aminoácido presente naquela coluna é idêntico em todas as seqüências; “:”, indica que uma substituição conservada ocorreu de acordo com as cores (vermelho-aa pequeno, azul - aa ácido, rosa - aa básico, verde-hidroxila+amino+básico-Q); e “.”, indica que substituições semi-concervativas ocorreram.

```

Espidroína2 A. jur  ARLSSPQASSRVSSAFFSLVSSGPTSPGALSNAISSVVSQVVSANPGLSGCDVLVQALLE 60
MaSp2a D. spi       SRMSTPGSGSRISNAVSNILSSGVSSSSGLSNAISNISSSISANPGLSGCDVLVQVLE 60
MaSp2 A. amo       SRLSSPQASSRVSSAVSSLVSSGPTNPAALSNAMSSVVSQVVSANPGLSGCDVLVQALLE 60
MaSp2 A. tri       SRLSSPQASSRVSSAVSTLVSSGPTNPAALSNAISSVVSQVSSSNPGLSGCDVLVQALLE 60
MaSp2 A. aur       SRLSSPQASSRVSSAVSTLVSSGPTNPAALSNAISSVVSQVVSANPGLSGCDVLVQALLE 60
MaSp2 A. bic       SRLSSSAASSRVSSAVSSLVSSGPTTTPAALSNTISSAVSQISANPGLSGCDVLVQALLE 60
MaSp2 L. hes       SALSSPTTHARISSHASTLLSSGPTNAAALSNVISNAVSQVVSANPGLSSSCDVLVQALLE 60
MaSp2 L. geo       SALSSPTTHARISSHASTLLSSGPTNSAAISNVISNAVSQVVSANPGLSSSCDVLVQALLE 60
MaSp2 N. cla       SRLASPDSGARVASAVSNLVSSGPTSSAALSSVISNAVSQIGASNGLSGCDVLVQALLE 60
MaSp2 N. mad       SRLASPDSGARVASAVSNLVSSGPTSSAALSSVISNAVSQIGASNGLSGCDVLIQALLE 60
MaSp2 N. sen       SRLASPDSGARVASAVSNLVSSGPTSSAALSSVIXNAVSQIGASNGLSGCDVLIKALLE 60
MaSp2 G. mam       SRLSSPQAGARVSSAVSALVASGPTSPAASVSSAISNVASQISANPGLSGCDVLVQALLE 60
MaSp2 U. div       SRLNSPASTSRVASAVSSLASAGAPSVGSLSSVISLSSSVSASNPGLSGCELLTQVLE 60
MaSp2 A. ven       NRLSSSGAANRVSSNVAIASGG---AAALPNVMSNIYSGVLGS--GVSSSEALIQALLE 55
ADF-3 A. dia      SRLSSPAASSRVSSAVSSLVSSGPTKHAALSNTISSVVSQVVSANPGLSGCDVLVQALLE 60
ADF-4 A. dia      SVYLRLLQPRLEVSSAVSSLVSSGPTNGAAVSGALNSLVSQISANPGLSGCDALVQALLE 60
      . . . . . : : . * . . . . : . * : . * * * . . : * . ***

Espidroína2 A. jur  IVS-----QYASLVGQ-SVNQALRY 79
MaSp2a D. spi       VISALVHILGASAVGQVG--SSPQNAQMVAANAVANAFS- 97
MaSp2 A. amo       IVSALVHILGSSSIGQINYAASSQYQMVGQ-SVAQALA- 98
MaSp2 A. tri       IVSALVHILGSSSIGQINYAASSQYQLVGQ-SLTQALG- 98
MaSp2 A. aur       LVSAVHILGSSSIGQINYAAS----- 82
MaSp2 A. bic       VVSALVHILGSSSVGQINYGASAQYQMV----- 89
MaSp2 L. hes       IITALISILDSSSVGQVNYGSSGQYQIVGQ-SMQQAMG- 98
MaSp2 L. geo       LITALISIVDSSNIGQVNYGSSGQYQMVG----- 90
MaSp2 N. cla       IVSACVTILSSSIGQVNYGAASQFAQVVGQ-SVLSAF-- 97
MaSp2 N. mad       IVSACVTILSSSIGQVNYGAA----- 82
MaSp2 N. sen       IVSACVTILSSSIGQVNYGAA----- 82
MaSp2 G. mam       IVSALVSISSASIGQINYGASGQYAAAMI----- 89
MaSp2 U. div       VVSALVALLGSAVGPVDSVSSSQYAGLVSS-AIAQAL-- 97
MaSp2 A. ven       VISALMHVLGSASIGNVSSAGLDSTLNVVQN-AVSQYAG- 93
ADF-3 A. dia      VVSALVSISSSIGQINYGASAQYQMVGQ-SVAQALA- 98
ADF-4 A. dia      LVSAVAILSSASIGQVNVSSVSQSTQMISQ-ALS----- 94
      . . .

```

**Figura 4:** Alinhamento da região C-terminal da Espidroína 2 de *A. juruensis* com proteínas MaSp2 de diferentes espécies de aranhas. Abreviações são as mesmas descritas para o alinhamento da figura 2. Abreviações das espécies de aranhas e número de acesso no GenBank (de cima para baixo): *A. jur*, *Avicularia juruensis*; *D. spi*, *Deinopsis spinosa* (ABD61593); *A. amo*, *Argiope amoena* (AAR13813); *A. tri*, *Argiope trifasciata* (AAK30596); *A. aur*, *Argiope aurantia* (AAK30592); *A. bic*, *Araneus bicentenarius* (AAC04503); *L. hes*, *Latrodectus hesperus* (AAY28936); *L. geo*, *Latrodectus geometricus* (AAK30603); *N. cla*, *Nephila clavipes* (AAT75315); *N. mad*, *Nephila inaurata madagascariensis* (AAZ15322); *N. sen*, *Nephila senegalensis* (AAK30609); *G. mam*, *Gasteracantha mammosa* (AAK30601); *U. div*, *Uloborus diversus* (ABD61599); *A. ven*, *Araneus ventricosus* (AAN85281); *A. dia*, *A. diadematus* (AAC47010 e AAC47011).

### *Frequência de codons*

A *Espidroína 2* da *A. juruensis* mostrou uma preferência por A (adenina) ou T (timina) como a terceira base de três para codificar os aminoácidos mais prevalentes em sua seqüência de proteínas (**Tabela 1**). A preferência de codons para glicina foi GGA/T em 82% dos casos, para serina a preferência para A e T foi de 91%, e os codons de alanina também usaram A ou T na posição de oscilação em 86% dos casos. Por outro lado, semelhante à seqüência das espidroínas da glândula tubuliforme, os codons da *Espidroína 1* para os aminoácidos alanina e serina são apenas moderadamente tendenciosos para A e T, com valores de 67 % e 56 %, respectivamente.

**Tabela 2**

*Escolha de codons para os aminoácidos mais freqüentes nas espidroínas de A. juruensis*

AmAc	Codon	Frequência (%)		AmAc	Codon	Frequência (%)	
		Espidroína 1	Espidroína 2			Espidroína 1	Espidroína 2
Ala	GCG	10	06	Gly	GGC	10	14
Ala	GCA	45	61	Ser	AGT	06	34
Ala	GCT	22	25	Ser	AGC	12	03
Ala	GCC	23	07	Ser	TCG	13	02
Gly	GGG	13	04	Ser	TCA	18	41
Gly	GGA	33	39	Ser	TCT	32	16
Gly	GGT	44	43	Ser	TCC	19	03

### **Discussão**

As aranhas da espécie *A. juruensis* possuem apenas uma ou duas glândula sericígena indiferenciadas, entretanto foram identificadas duas espidroínas diferentes, Espidroínas 1 e 2. Embora não seja comum, diferentes proteínas de seda produzidas pela mesma glândula foram também encontradas nas aranhas orbiculáreas com glândulas especializadas. Por exemplo, a MaSp1 foi identificada como sendo produzida pelas glândulas tubuliformes das aranhas *Aranues diadematus* e *Latrodectus hesperus* (Guerette *et al.*, 1996; Garb e Hayashi, 2005). Foi igualmente demonstrado que duas

proteínas podem estar presentes na mesma seda, como a MaSp1 e a MaSp2 na seda ampolada principal (Xu e Lewis, 1990), sugerindo que a combinação de proteínas deve possuir aplicações funcionais com relação às qualidades mecânicas da seda. Entretanto, não se sabe até que ponto as espidroínas produzidas por *A. juruensis* são ou não usadas simultaneamente na produção de suas fibras.

A maioria das proteínas da seda de aranhas é conhecida como sendo composta de combinações de quatro motivos simples de aminoácidos (poli-Ala, (GlyAla)<sub>n</sub>, GlyGlyX (onde X representa um pequeno subconjunto de aminoácidos), e GlyProGlyX(X)<sub>n</sub>. Entretanto, esses três motivos são deficientemente representados nas espidroínas encontradas na biblioteca de cDNA da glândula sericígena de *A. juruensis*.

As traduções dos três transcritos encontrados para Espidroína 1 mostraram alta similaridade entre as suas repetições e as regiões C-terminais com poucas substituições e deleções de aminoácidos, com identidades de até 85% em ambos os casos. A proteína da seda Fibroína 1 de *Euagrus chioseus*, também uma aranha migalomorfa, possui unidades repetitivas de ~180 aminoácidos de comprimento, e semelhantemente a Espidroína 1, é composta por uma seqüência rica em serina e alanina, inclusive com uma cadeia de treoninas (Gatesy *et al.*, 2001). Embora as regiões ricas em alanina sejam encontradas em diferentes sedas de aranhas, a treonina é um aminoácido raro nas espidroínas de Araneidae (tecedoras de teia orbicular). Análises utilizando o programa BLASTX (Altschul, 1997) mostraram que a seqüência repetitiva da Espidroína 1 também é similar a *cylindrical silk protein 1* (BAE54450) da aranha *Nephila clavata*.

A Espidroína 2 mostrou um padrão completamente novo em sua seqüência de aminoácidos se comparada com outras proteínas de seda descritas de aranhas migalomorfas. Embora tenhamos encontrado o motivo (GlyAla)<sub>n</sub> na composição repetitiva de aminoácidos da Espidroína 2, um motivo importante na composição das espidroínas de aranhas tecedoras de teia orbicular, também achamos em uma grande quantidade de motivo (GlySer)<sub>n</sub>, até o momento um motivo descrito apenas nas espidroínas das aranhas de tecedoras de teia não orbicular *Kukulcania hibernalis* e *Agelenopsis aperta* (Tian *et al.*, 2004). Outra característica interessante é que a região C-terminal da Espidroína 2 é bem parecida com aquela das aranhas Araneomorphae, a MaSp2. O alinhamento entre as seqüências de aminoácidos da região C-terminal da

Espidroína 2 da *A. juruensis* com a MaSps 2 de diferentes aranhas mostrou valores de identidade de 86% para a *Argiope amoena* e 83% para a *Argiope trifasciata*. Tai *et al.* (2004) também descobriram uma alta similaridade entre as regiões C-terminal de um gene análogo da MaSp1 da aranha migalomorfa *Macrothele holsti* com a MaSp1 de aranhas de tecedoras de teia orbicular. As aranhas migalomorfas são conhecidas por possuir apenas uma glândula sericígena não diferenciada e suas fiandeiras pouco desenvolvidas (Foelix, 1996), sugerindo que a região C-terminal das principais espidroínas foi conservada desde antes da separação filogenética das aranhas araneomorfas e migalomorfas há aproximadamente 340-390 milhões de anos (Ayoub *et al.*, 2007). Uma razão para a conservação da região C-terminal durante evolução entre as espidroínas de diferentes aranhas poderia estar relacionada com uma função importante. Sugere-se que a função da região C-terminal poderia evitar a prematura formação de fibras enquanto a proteína ainda estivesse na glândula, ou estar relacionada com o enovelamento das proteínas, uma vez que essa região ainda encontra-se presente na fibra polimerizada (Beckwitt e Arcidiacono, 1994; Sponner, 2004).

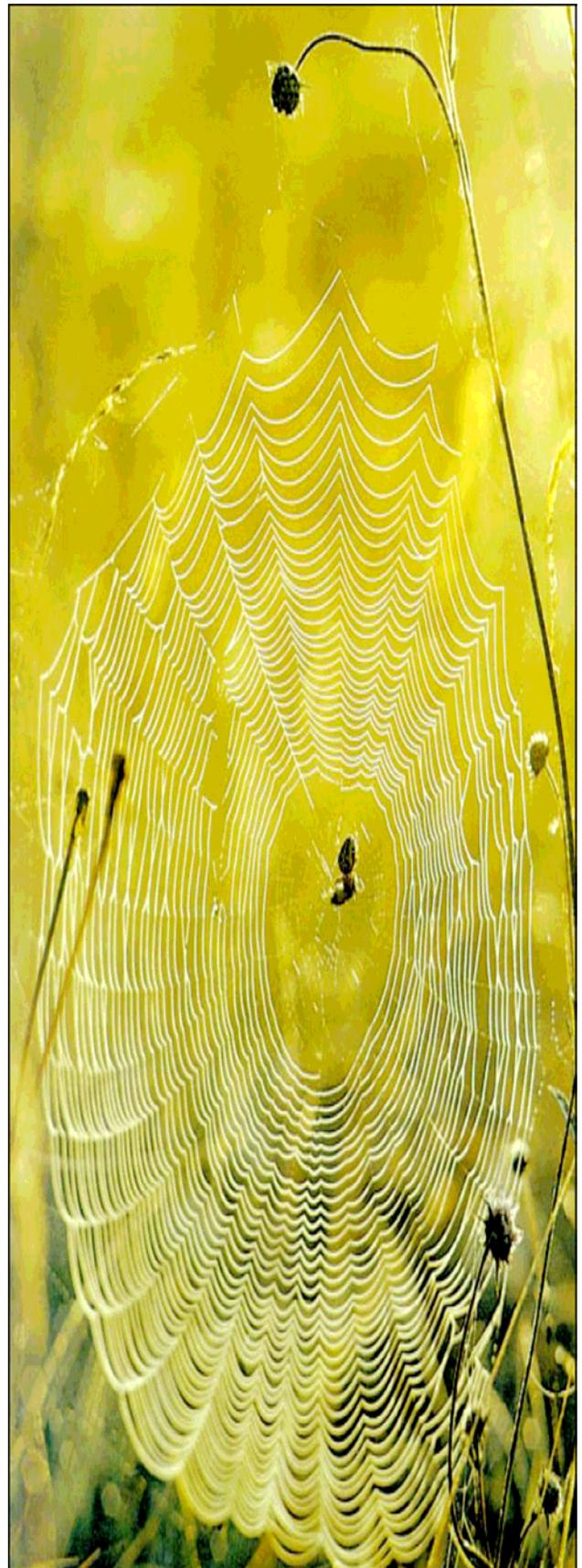
A falta dos motivos poli-Ala, (GlyAla)<sub>n</sub>, GlyGlyX e GlyProGlyXX encontrada na maioria das proteínas de aranhas tecedores de teia orbicular nas espidroínas da *A. juruensis*, uma aranha rudimentar que não constrói teias orbiculares, apóia a correlação entre as seqüências primárias de proteínas, estrutura secundária e propriedades mecânicas onde as repetições habituais ricas em motivos de alanina e/ou glicina pode ser fundamental para sedas envolvidas na captura de presas (Hayashi *et al.*, 1999). Embora a Espidroína 1 apresente o motivo poli-Ala e a Espidroína 2 (GlyAla)<sub>n</sub> em suas seqüências de aminoácidos, a presença de apenas um desses motivos não deve ser suficiente para produzir uma teia para a captura de presas, uma vez que a combinação de diferentes sedas compostas por proteínas com motivos variáveis é necessária para a produção de uma teia orbicular (Vollrath, 1994).

## Capítulo 5

### Relacionamento evolutivo da MaSp entre as diferentes espécies de aranhas

**Este capítulo foi baseado no artigo a ser  
submetido para publicação (Anexo 4):**

Bittencourt, D., Dittmar, K., Lewis, R.V.,  
Rech, E.L. How old are ampolada principal  
silks?



## Introdução

As mais elaboradas teias de aranha são contruídas pelas aranhas orbiculárias do grupo das Araneomorphae, as Araneoidea e Deinopoidea. Suas teias possuem espirais de captura envoltas por uma linha de suporte, tais fibras são extremamente resistentes, capazes de absorver a energia cinética produzido por uma presa voadora. Acredita-se que ambas linhagens tenham evoluído a partir de um ancestral em comum, e duas de suas fibras (Espidroína ampolada principal 2 – MaSp2, e espidroína flageliforme – Flag) foram identificadas como suporte da monofilia das orbiculárias (Garb *et al.*, 2006). Entretanto, as aranhas Mygalomorphae, o grupo irmão das Araneomorphae, não produzem teia em orbital, e usam suas sedas apenas para a construção de casulos e tocas, diferentemente das orbiculárias que as utilizam para caça.

Por este motivo não apenas as espidroínas das migalomorfos são pouco estudadas, mas estas são também conhecidas como aranhas “primitivas”. É bem sabido que aranhas orbiculárias utilizam uma variedade de sedas específicas para realizar diferentes tarefas, e assim, presumem-se que aranhas mygalomorfos não produzam tais fibras. Adicionalmente, esta linha de pensamento leva a suposição que estas aranhas apenas possuam sedas “ancestrais” (basal filogeneticamente). No intuito de identificar a seda ancestral a partir de uma aranha migalomorfa, *Avicularia juruensis* (Theraposidae), uma descoberta interessante foi feita, que desafia pareceres comuns sobre a evolução das sedas de aranha.

Três tipos de sedas, produzidos por glândulas diferentes, são responsáveis pela produção de teias em orbital; MaSp 1 e 2, MiSp e espidroína flageliforme (Gatesy *et al.*, 2001). A partir dos cDNAs de *A. juruensis*, nós identificamos duas espidroínas sendo expressas. Enquanto a Espidroína 1 mais abundante, é bem semelhante a Fibroína 1 da migalomorfa *Euagrus chisoseus*, a Espidroína 2 possui semelhanças claras com a MaSp2 do grupo orbiculárias da família Araneomorphae.

A fim de validar estes resultados, foi investigada a posição evolutiva das seqüências da *Avicularia* a partir da análise da filogenia molecular da seqüência de aminoácidos de 77 C-terminais de diferentes espidroínas de 35 espécies de aranha, e

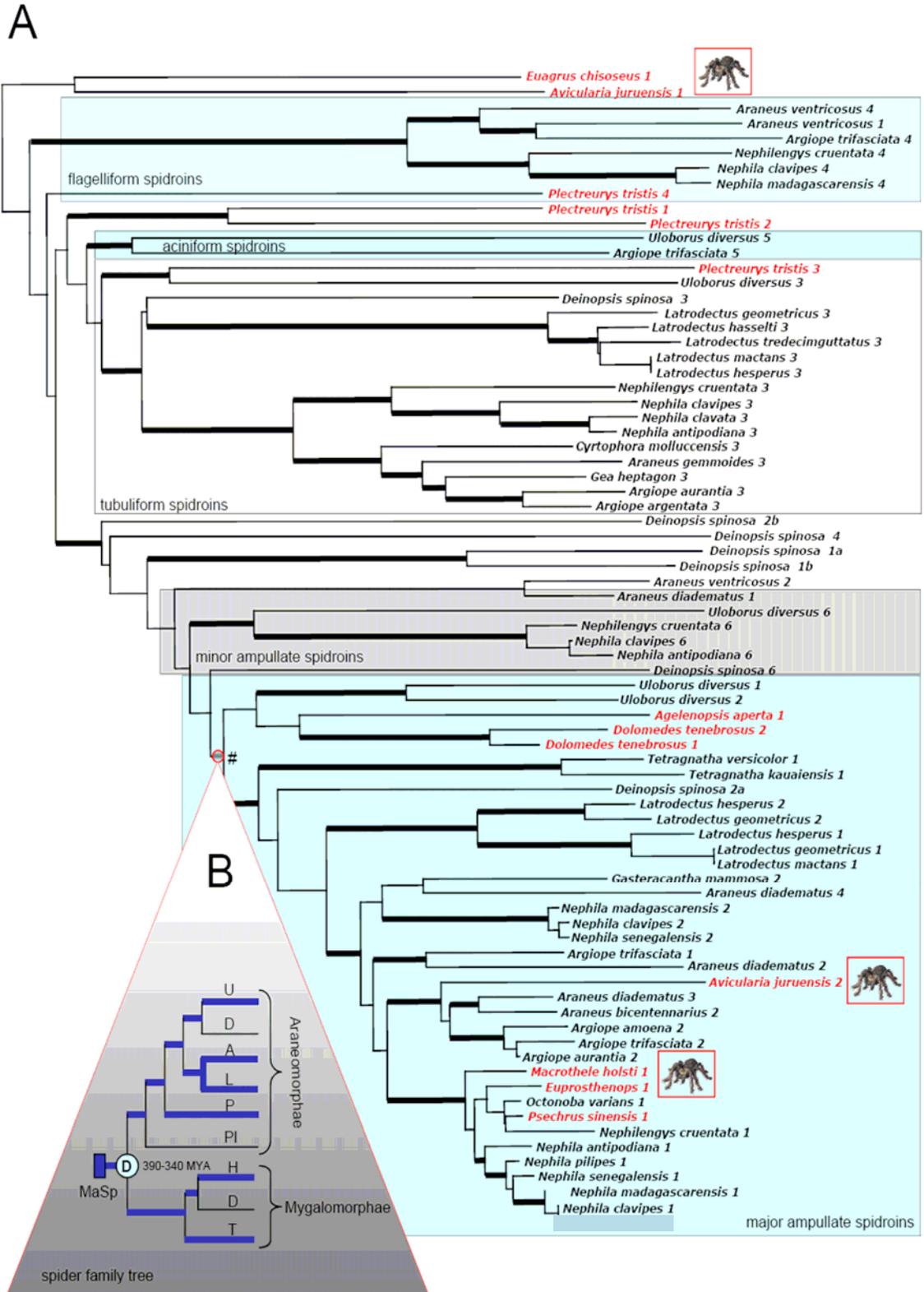
exploramos suas implicações para a origem das sedas de aranha em geral, e para a fiação orbicular em particular.

## Resultados e Discussão

A análise da filogenia molecular das regiões C-terminais examinou a relação das duas espidroínas isoladas da *A. juruensis* com outras espidroínas conhecidas. O resultante alinhamento de nucleotídeos consistiu em 366 pb. Todas as abordagens analíticas resultaram em topologias similares. Apresentamos a árvore de ML (Maximum likelihood) (**Figura 1**). A análise de MP gerou 25 árvores mais parcimoniosas (comprimento: 4074; IC: 0,423; IR: 0,602). O rigoroso consenso desintegrou todos os nós basais. A melhor árvore de ML ( $-\ln L = -7671.60611$ ) foi computado implementando-se o modelo GTR+I+G identificado por Modeltest v3.7 (Posada e Buckley, 2004). Os valores *bootstrap* (PB) não indicam nenhum apoio na maioria das ramificações basais (**Figura 1**).

Consistente com as análises anteriores, a maioria das seqüências que foram classificadas funcionalmente como MaSps, MiSps, Flag, AcSps e TuSps formam agrupamentos ortólogos discerníveis, intercaladas com terminais que não parecem pertencer a qualquer grupo funcional (Garb e Hayashi, 2005; Challis *et al.*, 2006). Dentro de algumas dessas classes, os padrões de especiação de aranhas são claramente preservados (por exemplo, TuSp), indicando uma origem a partir da duplicação para todas as principais classes funcionais.

Em geral, nenhum apoio foi encontrado para quaisquer dos nós basais, contudo, várias classes individuais recebem altos PBs. Assim, não se pode chegar a nenhuma conclusão quanto à sucessão de descendência entre as classes da principal seda funcional, e a noção comum de que as espidroínas são as produtoras de sedas mais "antigas" não pode ser confirmada. Isto pode vir a mudar com a adição de outros grupos de aranhas, atualmente subamostrados, como as migalomorfas ou araneomorfas não orbiculárias.



**Figura 1:** A) Relacionamentos filogenéticos entre as espidroínas de *A. juruensis* com diferentes seqüências de espidroínas publicamente disponíveis. As Araneomorphae incluídas: *Plectreurys tristis* [1 (AAK30610); 2 (AAK30611); 3 (AAK30612); 4 (AAK30613)], *Deinopsis spinosa* [1b (ABD61592); 1a (ABD61591); 2a (ABD61593); 2b (ABD61594); 4 (ABD61590)], *Dolomedes tenebrosus* [1 (AAK30598); 2 (AAK30599)], *Argiope aurantia* [2 (AAK30592); 3 (AAX45292)], *Argiope trifasciata* [1 (AAK30595); 2 (AAK30596); 4 (AAK30593); 5 (AAR83925)], *Argiope amoena* [2 (AAR13813)], *Argiope argentata* [3 (AAY28932)], *Latrodectus hasselti* [3(AAY28941)], *Latrodectus hesperus* [1(AAY28935); 2(ABD66603); 3 (AAY28931)], *Latrodectus geometricus* [1 (AAK30602), 2 (AAK30604), 3 (AAY28940)], *Latrodectus mactans* [3 (AAY28938)], *Cyrtophora moluccensis* [3 (AAY28944)], *Psechrus sinensis* [1 (AAV48939)], *Octonoba varians* [1 (AAV48931)], *Uloborus diversus* [ 1 (ABD61596), 2 (ABD61599), 3 (AAY28933), 5 (ABD61598), 6 (ABD61597)], *Araneus diadematus* [1 (AAC47008), 2 (AAC47009), 3 (AAC47010), 4 (AAC47011)], *Araneus ventricosus* [1 (AAN85280), 2 (AAN85281), 4 (ABK00016)], *Araneus bicentenarius* [2 (AAC04503)], *Araneus gemmoides* [3 (AAX45294)], *Gea heptagon* [3(AAY28943)], *Gasteracantha mammosa* [2 (AAK30601)], *Agelenopsis aperta* [ 1 (AAT08436)], *Nephila clavipes* [1 (AAT75312), 2 (AAT75315), 3 (AAX45295), 4 (AAF36089), 6 (AAC14589)], *Nephila clavata* [3 (BAE54450)], *Nephila antipodiana* [1 (ABC72644), 3 (AAY90151), 6 (ABC72645)], *Nephila madagascarensis* [1 (AAK30606), 2 (AAK30607), 4 (AAF36092)], *Nephila pilipes* [1 (AAV48946)], *Nephila senegalensis* [1 (AAK30608), 2 (AAK30609)], *Nephilengys cruentata* [ 1 (EF638446), 3 (EF638445), 4 (EF638444), 6 (EF638447)], *Euprosthenoops* sp. [1 (Pouchkina-Stanteva e McQueen-Mason, 2004)], *Tetragnatha versicolor* [1 (AAK30615)], *Tetragnatha kauaiensis* [1 (AAK30614)], Mygalomorphae: *Macrothele holsti* [1 (AAV48940)], e *Euagrus chisoseus* [1 (AAK30600)]. B) Árvore filogenética de aranha com dados da ML de Ayoub *et al.*, 2007. U = Uloboridae, D = Deinopidae, A = Araneidae, L = Latrodectidae, P = Pisauridae, Pl = Plectreuridae, T = Theraphosidae, D = Dipluridae, H = Hexathelidae. D – separação filogenética entre araneomorfos e migalomorfos (Ayoub *et al.*, 2007). Ramos azuis – presença de MaSps.

Entretanto, o que é mais importante, a Espidroína 2 da *A. juruensis* se aninha claramente dentro de um grupo de seqüências da MaSp2, e encontra-se posicionado no mesmo agrupamento que contém uma seqüência de outra aranha migalomorfa - a *Macrothele holsti*, que foi classificada uma seda análoga da MaSp1 (Tai *et al.*, 2004).

Nossa evidência sugere que, pelo menos, duas aranhas migalomorfas expressam C-terminais que possuem uma alta identidade de seqüências e afinidade filogenética com as sedas da ampolada principal. Seguindo a prática comum, uma função similar às outras MaSps poderia ser deduzida. Porém isso deve ser testado, e o fato de que as repetições são ligeiramente diferentes das repetições da MaSp das tecedoras orbiculares, combinado com o estilo de vida das tecedoras não orbiculares das migalomorfas sugere uma função diferente de outras MaSps.

Porém, a partir de uma perspectiva evolutiva, a topologia recuperada possibilita duas linhas de raciocínio: A) as aranhas migalomorfas evoluíram a ampolada principal como as regiões C-terminais (e sedas) independentemente depois de sua ramificação a partir das araneomorfias, o que seria um exemplo patente, mas improvável da evolução paralela; ou B) A região C-terminal da análoga MaSp das migalomorfas é a sobrevivente de duplicações de genes de seda na base da divergência migalomorfa-araneomorfa. Isto também insinuaria que a origem das MaSps de fiação orbicular pode ser colocada em um ponto de tempo antes da origem da aranha orbiculária e da fiação orbicular, o que será discutido abaixo.

O mapeamento da presença ou ausência das regiões C-terminais na análoga ampolada principal na árvore de espécies de aranhas tem como resultado a identificação do nó mais basal que une as araneomorfias e migalomorfas como o ponto de origem da MaSp, tanto sob a otimização ACCTRAN quanto DELTRAN. Com base nos dados disponíveis, a reconciliação da árvore resultou na identificação de 23 eventos de duplicações (62 perdas, contagem D/L: 77) por toda a árvore da C-terminal. Algumas dessas duplicações deduzidas estão conectadas as relações debilmente apoiadas, e assim, é provável que mudem com uma amostragem ampliada de táxons, contudo todos os grupos principais - MaSp, MiSp, TuSp, Flag, como também MaSp + MiSp - registram uma duplicação. Entretanto, nosso interesse específico estava na possibilidade de MaSp2 (PB=85%) ser uma seda comum, em vez de uma característica específica da orbiculária.

NOTUNG leva em consideração a estimativa de limites mais baixos no tempo de duplicação, isto sendo descrito como as espécies/especiação mais antigas na qual a duplicação deve ter estado presente. Sob todos os cenários possíveis de raízes mais parcimoniosa, o limite mais baixo para a MaSp foi calculado como sendo a divisão araneomorfa-migalomorfa. De maior importância, o mesmo limite também é calculado para a classe contendo a seqüência da Espidroína 2 da *Avicularia* (classe da MaSp2), e da *Macrothele* (MaSp1). Portanto, MaSp2, e as sedas da ampolada principal, em geral, não podem ser classificadas como uma sinapomorfia para a Orbiculária (Garb e Hayashi, 2005), mas têm que ser consideradas plesiomórficas, e apesar de não refutar a monofilia da orbiculária, eles certamente não a apoiam de modo único. Aparentemente, o único grupo de seda de aranhas que é autapomórfico a Orbiculária são as sedas da tubuliforme, e até mesmo isso pode mudar, dado a amostragem mais intensiva.

O padrão filogenético geral combinado com nossos resultados confirma uma importante influência da duplicação de genes na evolução de seda de aranhas, com exceção de outros padrões, tais como conversão de genes e a homogeneização intragênica (Garb e Hayashi, 2005; Challis *et al.*, 2006). Devido à aparente importância da duplicação de genes para a evolução de novas funções biológicas sobre todas as escalas de tempo evolutivas, isso faz sentido para as sedas de aranhas, e é concebível que a base funcional para seus diferentes usos tenha evoluído no princípio da radiação de aranhas, há 340 – 390 milhões de anos.

## REFERÊNCIAS

- 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.
- Andersen, S.O., 1970. Amino Acid Composition of Spider Silks. *Comp. Biochem. Physiol.* 35: 705-711.
- Augsten, k., Mühlig, P., Herrmann, C., 2000. Glycoproteins and skin-core structure in *Nephila clavipes* spider silk observed by light and electron microscopy. *Scanning* 22, 12-15.
- Ayoub, N.A., Garb, J.E., Hedin, M., Hayashi, C.Y., 2007. Utility of the nuclear protein-coding gene, elongation factor-1 gamma (EF-1 $\gamma$ ), for spider systematics, emphasizing family level relationships of tarantulas and their kin (Araneae:Mygalomorphae). *MPE* 42, 394-409.
- Barghout, J.Y.J., Czernuszka, J.T., Viney, C., 2001. Multiaxial anisotropy of spider (*Araneus diadematus*) cocoon silk fibres. *Polymer* 42, 5797-5800.
- 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.
- Beckwitt, R., Arcidiacono, S., 1994. Sequence conservation in the C-terminal region of spider silk proteins (Spidroin) from *Nephila clavipes* (Tetragnathidae) and *Araneus bicentenarius* (Araneidae). *J. Biol. Chem.* 269, 6661-6663.
- Beard, J., 1992. Warding off Bullets by a Spider's Thread. *New Scientist* 14, 18.
- Bell, F.I., McEwen, I.J., Viney, C., 2002. Fiber science: supercontraction stress in wet spider dragline. *Nature* 416, 37.
- Bini, E., Knight, D.P., Kaplan, D.L., 2004. Mapping domain structures in silks from insects and spiders related to protein assembly. *J Mol Biol.* 335, 27-40.

- 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.
- Brooks, A.E., Lewis, R.V., 2004. Probing the elastic nature of spider silk in pursuit of the next designer fiber. *Biomed Sci Instrum.* 40, 232-237.
- Brooks, A.E., Steinkraus, H.B., Nelson, S.R., Lewis, R.V., 2005. An investigation of the divergence of ampolada principal silk fibers from *Nephila clavipes* and *Argiope aurantia*. *Biomacromolecules* 6, 3095-3099.
- 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.
- Chen, X., Knight, D.P., Vollrath, F., 2002. Rheological characterization of nephila spidroin solution. *Biomacromolecules* 3, 644-648.
- Coddington, J.A., Giribet, G., Harvey, M.S., Prendini, L., Walter, D.E., 2004. In: *Assembling the Tree of Life*, Cracraft J, Donoghue MJ. (eds.) (Oxford University Press, New York).
- Coddington, J.A., Levi, H.M., 1991. Systematics and evolution of spiders (Araneae). *Ann. Rev. Ecol. Syst.* 22, 565-592.
- Colgin, M.A., Lewis, R.V., 1998. Spider ampolada secundária silk proteins contain new repetitive sequences and highly conserved non-silk-like "spacer regions". *Protein Sci.* 7, 667-672.
- Denny, M., 1976. The Physical Properties of Spider's Silk and Their Role in the Design of Orb-Webs. *J.Exp.Biol.* 65, 483-506.
- Dicko, C., Knight, D., Kenney, J.M., Vollrath, F., 2004. Secondary structures and conformational changes in flagelliform, cylindrical, major, and ampolada secundária silk proteins. Temperature and concentration effects. *Biomacromolecules* 5, 2105-2115.
- Dong, Z., Lewis, R.V., Middaugh, C.R., 1991. Molecular Mechanism of Spider Silk Elasticity. *Arch Biochem Biophys.* 284, 53-57.
- Durand, D., Halldorsson, B.V., Vernot., B., 2006. A hybrid micro-macroevolutionary approach to gene tree reconstruction. *J Comput Biol.* 13, 320-335.

- Edmonds, D., Vollrath, F., 1992. The contribution of atmospheric water vapour to the formation and efficiency of a spider's web. *Proc. R. Soc. Lond.* 248, 145-148.
- Elices, M., Perez-Rigueiro, J., Plaza, G., Guinea, G.V., 2004. Recovery in Spider Silk Fibers. *J. Appl. Pol. Sci.* 92: 3537-3541.
- 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., Bedzyk, L.A., 1997. Production of synthetic spider dragline silk protein in *Pichia pastoris*. *Appl Microbiol Biotechnol* 47, 33-39.
- Fahnestock, S.R., Irwin, S.L., 1997. Synthetic spider dragline silk proteins and their production in *Escherichia coli*. *Appl Microbiol Biotechnol.* 47, 23-32.
- 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*. Oxford University Press, Inc. and Georg Thieme Verlag, New York.
- Garb, J.E., Dimauro, T., Vo, V., Hayashi, C.Y., 2006. Silk genes support the single origin of orb webs. *Science* 312, 1762.
- Garb, J.E., Gonzales, A., Gillsepe, R.G., 2004. The black widow spider genus *Lactrodectus* (Araneae:Theridiidae): phylogeny, biogeography, and invasion history. *MPE* 31, 1127-1142.
- Garb, J.E., Hayashi, C.Y., 2005. Modular evolution of egg case silk genes across orb-weaving spider superfamilies. *Proc. Natl. Acad. Sci. USA* 102, 11379-11384.
- 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.
- Gersappe, D., 2002. Molecular mechanisms of failure in polymer nanocomposites. *Phys Rev Lett.* 89, 058301.
- Gosline, J.M., DeMont, M.E., Denny, M.W., 1986. The Structure and Properties of Spider Silk. *Endavour.* 10: 37-43.

- Gosline, J.M., Denny, M.W., Demont, M. E., 1984. Spider silk as rubber. *Nature* 309, 551-552.
- 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.M., 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.
- Grubb, D.T., Ji, G., 1999. Molecular Chain Orientation in Supercontracted and Re-extended Spider Silk. *Int. J. Biol. Macromol.* 24, 203-210.
- Guerette, P.A., Ginzinger, D.G., Weber, B.H., Gosline, J.M., 1996. Silk properties determined by gland-specific expression of a spider fibroin gene family. *Science* 272, 112-115.
- Guess, K.B., Viney, C., 1998. Thermal analysis of ampolada principal (dragline) spider silk: the effect of spinning rate on tensile modulus. *Thermochimica Acta.* 315, 61-66.
- Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 52, 696-704.
- Guinea, G.V., Elices, M., Perez-Rigueiro, J., Plaza, G.R., 2005. Stretching of supercontracted fibers: a link between spinning and the variability of spider silk. *J. Exp. Biol.* 208, 25-30.
- Guinea, G.V., Perez-Rigueiro, J., Plaza, G.R., Elices, M. 2006. Volume constancy during stretching of spider silk. *Biomacromolecules* 7, 2173-7.
- 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., Lewis, R.V., 2001. Spider flagelliform silk: lessons in protein design, gene structure, and molecular evolution. *BioEssays* 23, 750-756.

- 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.
- Hedin, M., Bond, J.E., 2006. Molecular phylogenetics of the spider infraorder Mygalomorphae using nuclear rRNA genes (18S and 28S): conflict and agreement with the current system of classification. *Mol. Phylogenet Evol.* 41, 454-71.
- Hijirida, D. H., Do, K.G., Michal, C., Wong, S., Zax, D., Jelinski, L.W., 1996. C13 NMR of *Nephila clavipes* ampolada principal silk gland. *Biophys. J.* 71, 3442-3447.
- Hinman, M.B., Jones, J.A., Lewis, R.V., 2000. Synthetic spider silk: a modular fiber. *Trends Biotechnol.* 18, 374-379.
- Hinman, M.B., Lewis, R.V., 1992. Isolation of a clone encoding a second dragline silk protein. *J. Biol. Chem.* 267, 19320-19324.
- Hormiga, G., Scharff, N., Coddington, J.A., 2000. The phylogenetic bases of sexual size dimorphism in orb-weaving spiders (Araneae, Orbiculariae). *Systematic Biol.* 49, 435-462.
- 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.
- Hu X., Kohler K., Falick A.M., Moore A.M., Jones P.R., Vierra C. 2006. Spider egg case core fibers: trimeric complexes assembled from TuSp1, ECP-1, and ECP-2. *Biochemistry.* 45, 3506-3516.

- 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.
- Huang, X., Madan, A., 1999. CAP3: A DNA sequence assembly program. *Genome Res.* 9, 868–877.
- Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754-755.
- Hutchinson, E.G., Thornton, J.M., 1994. A revised set of potentials for beta-turn formation in proteins. *Protein Sci.* 3, 2207-16.
- 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.
- Knight, D.P., Vollrath, F., 2002. Biological liquid crystal elastomers. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 357, 155-63.
- Lawrence, B.A., Vierra, C.A., Moore, A.M., 2004. Molecular and mechanical properties of ampolada principal silk of the black widow spider, *Latrodectus hesperus*. *Biomacromolecules* 5, 689-695.
- Lazaris, A., Arcidiacono, S., Huang, Y., Zhou, J.F., Duguay, F., Chretien, N., Welsh, E.A., Soares, J.W., Karatzas, C.N., 2002. Spider silk fibers spun from soluble recombinant silk produced in mammalian cells. *Science* 295, 472-6.
- Lewis, R.V., Hinman, M., Kothakota, S., Fournier, M.J., 1996. Expression and purification of a spider silk protein: a new strategy for producing repetitive proteins. *Protein Exp.r Purif.* 7, 400-6.
- Lewis, R.V., 2006. Spider silk: ancient ideas for new biomaterial. *Chem. Rev.* 106, 3762-3734.
- Li, S.F., McGhie, A.J., Tang, S.L., 1994. New internal Structure of Spider Dragline Silk Revealed by Atomic Force Microscopy. *Biophys J.*, 66, 1209-1212.

- Liu, Y., Shao, Z., Vollrath, F., 2005. Extended wet-spinning can modify spider silk Properties. *Chem. Commun (Camb)*, 19, 2489-2491.
- Lucas, F., 1964. Spiders and Their Silks. *Discovery* 25, 20-26.
- Maddison, D.R., Maddison, W.P., 2006. Sinauer Associates, Inc. Mac Clade Version 4.08.
- Madsen, B., Shao, Z.Z., Vollrath, F., 1999. Variability in the mechanical properties of spider silks on three levels: interspecific, intraspecific and intraindividual. *Int J Biol Macromol.* 24, 301-6.
- Miao, Y., Zhang, Y., Nakagaki, K., Zhao, T., Zhao, A., Meng, Y., Nakagaki, M., Park, E.Y., Maenaka, K., 2006. Expression of spider flagelliform silk protein in Bombyx mori cell line by a novel Bac-to-Bac/BmNPV baculovirus expression system. *Appl Microbiol Biotechnol.* 71, 192-9.
- Motriuk-Smith, D., Lewis, R.V., 2004. Brown Widow (*Latrodectus geometricus*) ampolada principal silk protein and its material properties. *Biomed. Sci. Instrum.* 40, 64-69.
- Oroudjev, E., Soares, J., Thompson, J.B., Fossey, S.A., Hansma, H.G., 2002. Segmented nanofibers of spider dragline silk: atomic force microscopy and single-molecule force spectroscopy. *Proc. Natl. Acad. Sci. USA* 99, 6460-6465.
- Parkhe, 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.
- Perez-Rigueiro, J., Elices, M., Plaza, G., Real, J.I., Guinea, G.V., 2005. The effect of spinning forces on spider silk properties. *J. of Exp. Biology* 208, 2633-2639.
- Platnick, N.I., 2007. The world spider catalog, version 6.5. American Museum of Natural History, online at <<http://research.amnh.org/entomology/spiders/catalog/index.html>>
- Posada, D., Buckley, T.R., 2004. Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Syst. Biol.* 53, 793-808.
- 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. B. Biochem. Mol. Biol.* 138, 371-376.

- Qin, X., Coyne, K.J., Waite, H.J., 1997. Tough Tendons. Mussel Byssus Has Collagen With Silk Like Domains. *J. of Biological Chemistry*. 272, 32623-32627.
- Riekel C., Madsen B., Knight D., Vollrath F., 2000. X-Ray Diffraction on Spider Silk During Controlled Extrusion Under a Synchrotron Radiation X-ray Beam. *Biomacromolecules* 1, 622-626.
- Rising A., Nimmervoll H., Grip S., Fernandez-Arias A., Storckenfeldt E., Knight D.P., Vollrath F., Engstrom W. 2005. Spider silk proteins--mechanical property and gene sequence. *Zoolog. Sci.* 22, 273-81.
- Sambrook, J., Russell, D.W., 2001. Molecular Cloning: A Laboratory Manual. 3<sup>th</sup> 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.
- Scheller J., Conrad U. 2005. Plant-based material, protein and biodegradable plastic. *Curr. Opin. Plant Biol.* 8, 188-96.
- Selden, P.A., Shear, W.A., Bonamo, P.M., 1991. A spider and other arachnids from the Devonian of New York, and reinterpretation of Devonian Araneae. *Paleontology* 34, 241-245.
- Shao, Z., Young, R.J., Vollrath, F., 1999. The effect of solvents on spider silk studied by mechanical testing and single-fibre Raman spectroscopy. *Int. J. Biol. Macromol.* 24, 295-300.
- Shear, W.A., Palmer, J.A., Coddington, J.A., Bonamo, P.M., 1989. A Devonian spinneret: early evidence of spiders and silk use. *Science* 246, 479-481.
- Smith, D.M., Smith, A., Hayashi, C.Y., Lewis, R.V., 2005. Analysis of the conserved N-terminal domains in ampelada principal spider silk proteins. *Biomacromolecules* 6, 3152-3159.
- Sponner, A., Unger, E., Grosse, F., Weisshart, K., 2004. Conserved C-termini of Spidroínas are secreted by the ampelada principal 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.

- Stauffer, S.L., Coguill, S.L., Lewis, R.V., 1994. Comparison of physical properties of three silks from *Nephila clavipes* and *Araneus gemmoides*. *J. Arachnol.* 22, 2437-2446.
- Stothard, P., 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* 28, 1102-1104.
- Swanson, B.O., Blackledge, T.A., Beltan, J., Hayashi, C.Y., 2006. Variation in the Material Properties of Spider Dragline Silk Across Species. *Appl. Phys. A* 82, 213-218.
- Swofford, D.L., 2002. PAUP\* version 4.0b10. Sinauer Associates, Sunderland, Mass.
- Tai, P.L., Hwang, G.Y., Tso, I.M., 2004. Inter-specific sequence conservation and intra-individual sequence variation in a spider silk gene. *Int. J. Biol. Macromol.* 34, 295-301.
- Tamburro, A.M., Guantieri, V., Scopa, A., Drabble, J.M., 1991. Polypeptide Models of Elastin: CD and NMR Studies on Synthetic Poly(X-Gly-Gly). *Chirality* 3, 318-323.
- Tatham, A.S., Shewry, P.R., 2000. Elastomeric Proteins: Biological roles, Structures and Mechanisms. *Trends Biochem. Sci.* 25, 567-571.
- 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.
- Telles, G.P., da Silva, F.L., 2001. Trimming and clustering sugarcane ESTs. *Genet. Mol. Biol.* 24, 17-23.
- Termonia, Y. 1994. Molecular Modeling of Spider Silk Elasticity. *Macromolecules* 27, 7378-7381.
- Thiel, B. L., Guess, K. B., Viney, C., 1997. Non-periodic lattice crystals in the hierarchical microstructure of spider (ampolada principal) silk. *Biopolymers* 41, 703-719.
- 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 ampolada principal silk cDNAs from two non-orb-weaving spiders. *Biomacromolecules* 5, 657-660.
- Tillinghast, E.K., Townley, M.A., Bernstein, D.T., Gallagher, K.S., 1991. Comparative study of orb web hygroscopicity and adhesive spiral composition in three araneid spiders. *J. Exp. Zool.* 25, 154-165.
- Toner, M., 1992. Spin Doctor. *Discover*. May: 32-36.
- Urry, D.W., Luan, C.H., Peng, S.Q., 1995. Molecular biophysics of elastin structure, function and pathology. *Ciba Found Symp.* 192, 4-22; discussion 22-30.
- Urry, D.W., Parker, T.M., 2002. Mechanics of Elastin: Molecular Mechanism of Biological Elasticity and its Relationship to Contraction. *J Muscle Res Cell Motil.* 23, 543-559.
- 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. USA* 99, 10266-10271.
- Vollrath, F., 1992. Spider web and silks. *Sci. Am.* 266, 70-76
- Vollrath, F., 1994. General Properties of some spider silks. In: Kaplan, D., Adams, W.w., farmer, B., Viney, C. Silk polymers materials science and biotechnology. American chemical Society, Washington, DC.
- Vollrath, F., 2005. Spider's web. *Curr. Biol.* 15, R364-5.
- Vollrath, F., Barth, P., Basedow, A., Engstrom, W., List, H., 2002. Local tolerance to spider silks and protein polymers in vivo. *In Vivo* 16, 229-34.
- Vollrath, F., Edmonds, D., 1989. Modulation of the mechanical properties of spider silk coating with water. *Nature* 34, 305-307.
- Vollrath, F., Knight, D., 1999. Structure and function of the silk production pathway in the spider *Nephila edulis*. *Int. J. Biol. Macromol.* 24, 243-249.
- 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.
- Wirth, E., Barth, F.G., 1992. Forces in the Spider Orb Web. *J.Comp.Physiol A.* 171, 359-371.
- Work, R.W., 1985. Viscoelastic behavior and wet supercontraction of ampolada principal silk fibers of certain orb-web-building spiders (Araneae, Araneidae). *J. Arachnol.* 15, 65-80.
- Xu, M., Lewis, R.V., 1990. Structure of a protein superfiber: spider dragline silk. *Proc. Natl. Acad. Sci. USA* 87, 7120-7124.
- Zhang, D., Cook, W.D., Horner, N.V., 2004. ITS2 rDNA variation of two black widow species, *Lactrodectus mactans* and *Lactrodectus hesperus* (Araneae, Theridiidae). *J. Arachnol.* 32, 349-352.

## ANEXO 1

### Materiais e Métodos

#### *Extração de RNA*

As aranhas *N. cruentata* e *A. juruensis* foram coletadas nas regiões nativas da Floresta Atlântica (São Paulo/SP, 23 ° 33' 54" e W 46° 46' 09") e Floresta Amazônica (Monte Negro/RO S 10 ° 17' 40" e W 60 ° 19' 31"), no Brasil respectivamente. As aranhas *N. cruentata* são orbiculárias e possuem todos os diferentes tipos de glândulas sericígenas, para a construção da biblioteca de cDNA o RNA total de todas as glândulas foi coletado em conjunto. As aranhas *A. juruensis* por sua vez, são caranguejeiras (aranhas arborícolas) e possuem apenas uma ou duas glândulas produtoras de seda indiferenciadas. As glândulas sericígenas foram isoladas sob um estereomicroscópio e, imediatamente, congeladas em nitrogênio líquido. Após pulverização, a extração de RNA total foi realizada usando o reagente TRIZOL (Invitrogen, EUA) seguindo as recomendações do fabricante. O rendimento, pureza e a integridade do RNA total foram determinados pela medição da absorbância em 260/280 nm e por meio de eletroforese em gel de agarose. O kit Oligotex (Qiagen, Alemanha) foi usado para a purificação de mRNA de acordo com o manual técnico, e a concentração e pureza do mRNA isolado também foram determinadas através da absorbância (260/280 nm).

#### *Construção das bibliotecas de cDNA*

As bibliotecas de cDNA foram construídas utilizando o kit "SUPERScript II Plasmid System with GATEWAY Technology" (Invitrogen, EUA.), de acordo com as instruções do fabricante. Após a transformação de *Escherichia coli* DH5 $\alpha$  utilizando a eletroporação (Sambrook e Russell, 2001), as bibliotecas foram cultivadas em placas com meio seletivo e os clones positivos para inserção de cDNA foram escolhidos e transferidos para placas de 96 poços. Os plasmídeos foram usados para seqüenciamento após a sua extração das bactérias por lise alcalina (Sambrook e Russell, 2001) e

quantificação. As reações de seqüenciamento foram realizadas usando-se para tal o kit Big Dye (Applied Biosystems, EUA), seguindo as instruções do fabricante, e o seqüenciador de DNA ABI 3700. Os cromatogramas resultantes foram diretamente transferidos para um banco de dados central semelhante ao descrito por Telles *et al.* (2001) para processamento e análise.

### *Seleção e análise de seqüências*

Potenciais clones positivos foram determinados utilizando BLASTX (Altschul, 1997) através da identificação de seqüências homólogas às regiões repetitiva e C-terminal. Várias combinações de enzimas de restrição foram usadas para conferir o tamanho das inserções. Os clones com inserções maiores que 1,5 kb foram tratados com exonuclease III (kit Erase-a-Base, Promega, EUA.) e usados para seqüenciamento. Fragmentos dos clones positivos para os cDNAs das espidroínas das glândulas ampolada principal e flageliforme de *N. clavipes* foram usados como sondas na análise *Southern Blot* (Sambrook e Russell, 2001), por meio de marcação aleatória do iniciador com [ $\alpha$ -<sup>32</sup>P] dCTP, para a identificação de novos clones positivos.

O *Base calling* e *quality assignment* de bases individuais foram realizados através do uso de Phred (Ewing e Green, 1998; Ewing *et al.*, 1998). As caudas poli(A) ribossômicas, seqüências de baixa qualidade, vetor e regiões de adaptadores foram removidos conforme descrito por Telles e da Silva (2001) com adaptações secundárias para este projeto. Os conjuntos resultantes de seqüências de nucleotídeos foram reunidos em agrupamentos de seqüências sobrepostas usando-se os programas cap3 (Huang e Madan, 1999) ou phrap (<http://www.phrap.org/phredphrap/phrap.html>), com qualidade de base individual e parâmetros pré-definidos. Alinhamentos múltiplos de seqüências foram executados usando ClustalW versão 1.8 (Thompson *et al.*, 1994) com valores pré-definidos e refinados através de exame. O conjunto de repetições dentro de cada espidroína foi alinhado, e uma repetição de consenso de aminoácidos foi gerada para cada proteína traduzida. A análise do uso de codons foi realizada utilizando o software CodonUsage com o código genético padrão (Stothard, 2000).

## RT-PCR

A fim de verificar os clones positivos encontrados na biblioteca, o RNA total da(s) glândula(s) sericígena(s) foi usado como molde para a transcrição reversa. *Superscript* II (Invitrogen, EUA) foi utilizado nas reações, seguindo as instruções do fabricante. A análise da reação em cadeia da polimerase (PCR) foi conduzida usando-se a Taq Polimerase (Invitrogen, EUA) nas seguintes condições: a desnaturação inicial do molde foi ajustada para 2 minutos a 94°, os 35 ciclos repetidos foram 30 segundos a 94°, 1 minuto a 56° e 30 segundos a 72°, a extensão final foi de 72° por 10 minutos. Os oligonucleotídeos usados foram designados de acordo com a seqüência C-terminal dos clones positivos de cDNA. Os respectivos *primers* reverso e direto usados para cada cDNA da seda de aranha foram, para *N. cruentata*: MaSp forward- 5'-TGCAGGTCAAGGTGGATATGGTG-3', MaSp- 5'-CCTAGAGCTTGGTAAACCGATTGAC-3', MiSp- 5'-CTCTGCGGGTGTAGGTGTTGG-3', MiSp reverse- 5'-TGCAGCAGACGAACCAACAGA-3', Flag Spidroína forward- 5'-GCATATGGAGCTGGTTC TGGTACAC-3', Flag Spidroína reverse- 5'-CTTCGAACAGAGTTGGAATATGCAT-3', TuSp forward- 5'-CATCTTCCGGCTTAGCATC-3' and TuSp reverse- 5'-ACCAGC GAGAGCAGTTGC-3'; para *A. juruensis*: Spidroína 1 forward- 5'-TGTCGCAGTTT CTTCCAATGG-3', Spidroína 1 reverse- 5'-CCACAGCAGCGGAAGTTGAAC-3', Spidroína 2 forward- 5'-GGACCAGCACGCCAAGGACC-3' e Spidroína 1 reverse- 5'-CGAAGAGCTTGATTTACAGATTGGCC-3'.

## Árvore filogenética de *N. cruentata*

A análise filogenética foi conduzida usando o alinhamento das seqüências de aminoácidos da região C-terminal realizado em MAFFT (5.8), sob o algoritmo E-INS-i (mafft --genafpair --maxiterate 1000 input\_file > output\_file) orientado para a precisão, conforme implementado no servidor *online* da Universidade de Kyushu, Japão (<http://align.bmr.kyushu-u.ac.jp/mafft/online/server/>) (Kato *et al.*, 2005). A análise filogenética foi executada no grupo de computadores do laboratório do Dr. David Liberles, Universidade de Wyoming. A topologia foi reconstruída usando a probabilidade

máxima como implementada em PHYML (Guindon e Gascuel, 2003), utilizando o modelo Blossum 62 de evolução das proteínas. Os valores não paramétricos *bootstrap* (método de simulação) foram avaliados com 1000 réplicas *bootstrap*.

#### *Coleta da fibra ampolada principal de N. cruentata*

As aranhas *N. cruentata* (N1 e N2) foram anestesiadas com dióxido de carbono, imobilizadas numa superfície plástica com pedaços de fita adesiva com as fiandeiras voltadas para cima. A fibra ampolada principal foi identificada a partir da localização da espinereta anterior sob um estereomicroscópio. Dez amostras com aproximadamente 5 cm de comprimento foram coletadas de cada aranha, e montadas num cartão para análise mecânica como descrito anteriormente (Stauffer *et al.*, 1994).

#### *Medida do diâmetro da fibra*

As fibras foram analisadas utilizando o microscópio Nikon Eclipse E200 equipado com uma câmera fotográfica. Cada fibra foi observada num aumento de 800X e visualizadas por meio do programa SPOT Basic. Os diâmetros foram medidos usando o programa ImageJ versão 1.32 (<http://rsb.info.nih.gov/ij/>), os valores finais foram determinados pela média de cinco medidas tomadas ao longo da fibra.

#### *Testes mecânicos*

Cada fibra foi testada por meio do aparelho MTS (Sistema Mecânico de Prova) Synergie 100 usando uma célula com 10g de carga. A fibra foi esticada com uma velocidade de 2 mm/min, e cada ponto analisado foi coletado num índice de 35 Hz. Os dados foram coletados usando Testworks 4 *software* (Corporação de Sistemas de MTS, Cary, NC) e curvas de tensão-tração foram construídas usando Microsoft Office Excel 2003. As seguintes equações foram usadas para calcular valores de engenharia de tensão, tração, e rigidez:

$\sigma_{\text{eng}}$  (tensão)=F/A, onde F é a força aplicada e A é a área do corte transversal da fibra.

$\epsilon_{\text{eng}}$  (tração)= $\Delta L/L_0$ , onde  $\Delta L$  é a mudança de comprimento da fibra e  $L_0$  é o comprimento inicial.

Y (rigidez ou modulo de young)=  $\sigma / \epsilon$

As seguintes equações foram usadas para calcular valores reais de tensão e tração:

$$\sigma_{\text{tr}} (\text{tensão}) = \sigma_{\text{eng}} (1 + \epsilon_{\text{eng}})$$

$$\epsilon_{\text{tr}} (\text{tração}) = \ln (1 + \epsilon_{\text{eng}})$$

### *Microscopia de força atômica*

As fibras de teia de aranha avaliadas por meio de microscopia de força atômica foram diretamente enroladas na superfície de grafite. O microscópio de força atômica foi operado no modo contato e imagens de modulação de força também foram obtidas. O equipamento utilizado foi um microscópio de força atômica Shimadzu SPM-9600 com um scanner apresentando dimensão de varredura máxima de 125  $\mu\text{m}$  x 125  $\mu\text{m}$ . Ponteiros de nitreto de silício ( $\text{Si}_3\text{N}_4$ ) piramidais com uma constante de mola de aproximadamente 0,032 N/m foram utilizadas. A força aplicada e a velocidade de varredura nas amostras durante o procedimento foram determinados nunca excedendo 10 nN. A resolução adotada foi de 512 x 512 linhas a uma frequência de 1 Hz. Repetidos procedimentos de varredura (incluindo traço-retraço e variações de ângulo da amostra) foram realizados na mesma região no intuito de confirmar que nenhuma alteração morfológica estivesse ocorrendo durante o procedimento de análise. As imagens foram processadas e as características nanoestruturais foram aferidas utilizando os softwares de aquisição e análise que acompanham os equipamentos.

### *Análise filogenética de A. juruensis, mapeamento de caracteres e reconciliação da árvore*

As seqüências de nucleotídeo da região C-terminal de 77 genes de seda a partir de

35 espécies de aranhas foram baixadas do GenBank ou extraídas da literatura. Essas seqüências incluíram todas as proteínas distintas conhecidas funcionalmente: a) flageliforme (Flag, espiral de captura de teia orbicular), b) ampolada principal 1+2 (MaSp, linha de moldura e segurança), c) Aciniforme (AcSp, teia de esperma, captura de presas, decoração), d) Ampolada secundária (MiSp, espiral temporária), e) tubuliforme (TuSp, casulo), e f) funcionalmente indeterminada (por exemplo, "fibroínas"). Todas as 78 seqüências (76 + 2 x Avicularia) foram traduzidas em suas respectivas seqüências de aminoácidos, e alinhadas usando-se o algoritmo G-Ins-i orientado para a precisão, que utiliza as informações globais de alinhamento informação de alinhamento formado em pares, conforme implementado em MAFFT 5.861 (desvantagem de abertura de lacuna (gap open penalty) = 3) (Kato et al., 2005). As seqüências foram então novamente traduzidas para seus respectivos alinhamentos de nucleotídeos (documento de biopython (módulo de python para Biologia Molecular), e submetidas à análise filogenética por meio das abordagens de Máxima Parcimônia (MP), Probabilidade Máxima (ML) (PAUP\* v4.0b10, Swofford, 2002), e abordagem bayesiana com MCMC (MrBayes v.3.1.2, Huelsenbeck e Ronquist, 2001). As pesquisas heurísticas de MP foram conduzidas com *branch swapping* do tipo TBR e 10.000 adições de seqüências aleatórias (RSA), as lacunas foram tratadas como o 5° estado. A análise de ML foi realizada usando-se o modelo de melhor ajuste de evolução conforme estimado pelo Modeltest v3.7 (AIC Criterion, Posada e Buckley, 2004), combinado com 100 réplicas de RSA (algoritmo de encriptação de dados). O suporte *bootstrap* foi avaliado através de 1000 pseudo-réplicas/100 RSA, e 100 pseudo-réplicas/1 RSA para MP e ML, respectivamente. A análise bayesiana incluiu três execuções independentes a  $3 \times 10^6$  gerações, cada uma com quatro cadeias (temperatura: 0.15), e árvores aleatórias de partida; a frequência da amostragem foi 1000. Os pontos de amostra antes de atingirem o patamar de contagens de probabilidade foram descartados como debug. As probabilidades posteriores de 95% foram consideradas como apoiadoras da relação topológica. Consistente com o trabalho anterior (Garb e Hayashi, 2005; Challis *et al.*, 2006), as árvores foram enraizadas para a Fibroína 1 da *Euagrus chisoseus* e para a Espidroína 1 da seda da aranha *Avicularia juruensis* (migalomorfás, irmãos das araneomorfás).

Para localizar a presença [1] ou ausência [0] de sedas da ampolada principal (1

e/ou 2) em uma árvore da espécie, as topologias das aranhas foram construídas no MacClade 4.08 (Maddison e Maddison, 2006). Nossa amostragem das aranhas migalomorfas incorpora três famílias: Dipluridae (*Euagrus chisoseus*), Hexathelidae (*Macrothele holsti*), ambas comumente denominadas como aranhas de teia de folhas, teia de funil ou teia de cortina, e a Theraphosidae (*Avicularia juruensis*). A recente pesquisa filogenética mostrou que a Dipluridae e a Hexathelidae fazem parte de um agrupamento parafilético na base do clado das não Atypidae (Hedin e Bond, 2006; Ayoub *et al.*, 2007), enquanto que as Theraphosidae são monofiléticas e derivadas. Por conseguinte, sob nosso cenário de três táxons, relações dipluridae-hexathelidae foram invocadas como o grupo-irmão para a *Avicularia* na árvore das espécies. A localização seguiu a otimização ACCTRAN e DELTRAN.

A fim de avaliar a contribuição relativa de duplicações de gene, e suas colocações topológicas contra os eventos de especiação por toda a evolução de sedas de aranhas, a reconciliação de árvores de genes com base nas regiões C-terminais com árvores de espécies de aranhas e baseada nas recentes filogenias de aranhas (Hedin e Bond, 2006; Ayoub *et al.*, 2007) foi conduzida em NOTUNG v.2.1. (Durand *et al.*, 2006). Apenas os táxons representados através das regiões C-terminais das aranhas foram incluídos. Os pesos limites foram atribuídos pelos valores *bootstrap* de ML, o custo de duplicação foi ajustado para 2.0, e o custo de perda para 0.5. O programa também foi usado para testar o enraizamento otimizando o número de duplicação e perdas sob diferentes cenários de raízes, e para avaliar o menor limite de duplicação.



301 T V N G P G G P I T I S E E L T V G G P  
 961 GGTGCTGGAGGTTTCAGGACCTGGTGGTCTGGACCAGGAGGCGCAGGACCCGGTGGTGCA  
 321 G A G G S G P G G A G P G G A G P G G A  
 1021 GGACCAGGAGGAGTAGGACCTGGTGGTCTGGAGGCCCTGGTGGTCTGGAGGACCTTTC  
 341 G P G G V G P G G A G G P G G A G G P F  
 1081 GGTCCAGGTGGTTCCGGACCCGGAGGTGCAGGCGGCGCTGGAGGACCTTATGGACCTGGT  
 361 G P G G S G P G G A G G A G G P Y G P G  
 1141 GGAGCTTACGGACCTGGTGGACCTGGAGGGCCTGGTGGTCTGGAGGACCCGGTTCTGGC  
 381 G A Y G P G G P G G P G G P G G P G S G  
 1201 GGACCTTACGGACCTGGTGGTCTTACGGACCTGGTGGTCTTACGGACCTGGTGGTCT  
 401 G P Y G P G G A Y G P G G A Y G P G G A  
 1261 TATGGACCTGGTGGAGCTGGTGGACCATATGGACCCGGTGGACCTTACGGACCAGGTGGT  
 421 Y G P G G A G G P Y G P G G P Y G P G G  
 1321 CCATACGGACCTGGTGGAGCTGGTGGACAGGTGGTCTGGTGGACCTTATGGACCAGGA  
 441 P Y G P G G A G G P G G A G G P Y G P G  
 1381 GGTGCTGGACCTGGCGGATACGGACCTGGAGGCGCTGGACCTGGTGGATACGGACCTGGC  
 461 G A G P G G Y G P G G A G P G G Y G P G  
 1441 GGTTCTGGACCTGGAGGCGCTGGACCTGGTGGATACGGACCTGGCGGTTCTGGACCTGGT  
 481 G S G P G G A G P G G Y G P G G S G P G  
 1501 GGTGCTGGACCCGGTGGATATGGACCTGGTGGTCTGGACCCGGTGGATACGGACCTGGT  
 501 G A G P G G Y G P G G A G P G G Y G P G  
 1561 GGTGCTGGACCTGGCGGTGCTGGACCTGGTGGTCTGGTCTGGTGGATACGGACCTGGC  
 521 G A G P G G A G P G G A G P G G Y G P G  
 1621 GGTTCTGGACCAGGTGGTCTGGATCTGGTGGCCCAGGCGGAGCGGGAGGTTTCAGGTGGA  
 541 G S G P G G P G S G G P G G A G G S G G  
 1681 ACAACAGTCATAGAAGATTTGGACATAACAGTTAATGGTCCAGGAGGCCCGATAACAATC  
 561 T T V I E D L D I T V N G P G G P I T I  
 1741 TCAGAAGAGCTAACAGTTGGTGGTCCAGGTGCTGGAGGTTTCAGGACCTGGTGGTCTGGA  
 581 S E E L T V G G P G A G G S G P G G A G  
 1801 CCAGGAGGCGCAGGACCCGGTGGTGCAGGACCAGGAGGCGCAGGACCTGGTGGTGCAGGA  
 601 P G G A G P G G A G P G G A G P G G A G  
 1861 CCAGGAGGAGTAGGACCTGGTGGTCTGGAGGCCCTGGTGGTCTGGAGGACCTTTCGGT  
 621 P G G V G P G G A G G P G G A G G P F G  
 1921 CCAGGTGGTTCCGGACCCGGAGGTGCAGGCGGCGCTGGAGGACCTTATGGACCTGGTGG  
 641 P G G S G P G G A G G A G G P Y G P G G  
 1981 GCTTACGGACCAGGTGGACCCGGAGGACCTGGAGGGCCTGGTGGTCTGGAGGACCCGGT  
 661 A Y G P G G P G G P G G P G G P G G P G  
 2041 TCTGGCGGACCTTACGGACCTGGTGGTCTTACGGACCTGGTGGTCTTACGGACCTGGT

681 S G G P Y G P G G A Y G P G G A Y G P G  
2101 GGTGCTTATGGACCTGGTGGAGCTGGTGGACCAGGTGGAGCTGCTGGACCATATGGACCC  
701 G A Y G P G G A G G P G G A A G P Y G P  
2161 GGTGGACCTTACGGACCAGGTGGTCCATACGGACCTGGTGGAGCTGGTGGACCAGGTGGT  
721 G G P Y G P G G P Y G P G G A G G P G G  
2221 GCTGGTGGACCTTATGGACCAGGAGGTGCTGGACCTGGCGGATACGGACCTGGAGGCGCT  
741 A G G P Y G P G G A G P G G Y G P G G A  
2281 GGACCTGGTGGATACGGACCTGGCGGTTCTGGACCTGGAGGCGCTGGACCTGGTGGATAC  
761 G P G G Y G P G G S G P G G A G P G G Y  
2341 GGACCTGGCGGTTCTGGACCTGGTGGTGGTGGACCCGGTGGATATGGACCTGGTGGTGGT  
781 G P G G S G P G G A G P G G Y G P G G A  
2401 GGACCCGGTGGATACGGACCTGGTGGTGGTGGACCTGGCGGTTGGTGGTGGTGGTGGTGGT  
801 G P G G Y G P G G A G P G G A G P G G A  
2461 GGTCCTGGTGGATACGGACCTGGCGGTTCTGGACCAGGTGGTCTGGATCTGGTGGCCCA  
821 G P G G Y G P G G S G P G G P G S G G P  
2521 GGCGGAGCGGGAGGTTTCAGGTGGAACAACAGTAATAGAAGATTTGGACATAACACTTAAT  
841 G G A G G S G G T T V I E D L D I T L N  
2581 GGTCCAGGAGGCCCGATAACAATCTCAGAAGAGCTAACAGTTGGTGGTCCAGGTGCTGGA  
861 G P G G P I T I S E E L T V G G P G A G  
2641 GGTTTCAGGACCTGGTGGTGGTGGACCCAGGAGGCGCAGGACCCGGTGGTGCAGGACCAGGA  
881 G S G P G G A G P G G A G P G G A G P G  
2701 GGCGCAGGACCAGGAGGAGTAGGACCTGGTGGTGGTGGAGGACCTTATGGTTCTGGTGGT  
901 G A G P G G V G P G G A G G P Y G S G G  
2761 TTCGGATTCGGAGGTGCAGGCGGCTCTGGAGGACCTTATGTACCTGGTGGAGCATATGGA  
921 F G F G G A G G S G G P Y V P G G A Y G  
2821 GCTGGTTCTGGTACACCATCTTATAGTGGATCTCGTGTTCCTGATTTGGTGAATGGTATA  
941 A G S G T P S Y S G S R V P D L V N G I  
2881 ATGCGTTCGATGCAAGGCTCTGGTTTCAACTATCAAATGTTTGGCAACATGTTATCGAAA  
961 M R S M Q G S G F N Y Q M F G N M L S K  
2941 TATGCCTCCGGATCAGGTGCATGCAATTCAAATGATGTTAATGTTTTAATGGATGCTCTT  
981 Y A S G S G A C N S N D V N V L M D A L  
3001 CTTGCGGCTTTGCACTGTCTCAGTAGCCATGGATCCCCATCATTTGGGTCTTCTCCAACC  
1001 L A A L H C L S S H G S P S F G S S P T  
3061 CCTTCTGCAATGAATGCATATTCCAACCTCTGTTTCAAGAATGTTCCAATTCTAAGGTTAT  
1021 P S A M N A Y S N S V R R M F Q F \* G Y  
3121 ACTCCTTTAAACTTGAATTTATTTTCAAATCATTTTGGATGAACCTTAGTTACTCATTGGA  
1041 T P L N L N L F S N H F D E P \* L L I \*  
3181 AGAAAAAATAAATATCTTTTTAGCAGAAAAA

1061 R K K \* I S F \* Q K K K K K K K K K K K K K  
 3241 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  
 1081 K K K K K K K K K K K K K

*Seqüência do cDNA parcial de NCMA<sub>Sp</sub>-like*

1 CAAGGTGCCGGAGCAGCAGCAGCAGCAGCAGCCGCAGGTGGAGCTGGACAAGGCGGATAT  
 1 Q G A G A A A A A A A G G A G Q G G Y  
 61 GGAGGTCTTGGTGGCCAAGGAGCTGGAGCCGCAGCTGCAGCAGCTGGTGGTCTGGACAA  
 21 G G L G G Q G A G A A A A A G G A G Q  
 121 GGAGGATATGGAGGTCAAGGTGCTGGACAAGGTGCAGCCGCAGCAGCAGCTAGTGGTGCC  
 41 G G Y G G Q G A G Q G A A A A A A S G A  
 181 GGACAAGGAGGATATGAAGGTCCAGGTGCCGGACAAGGTGCAGGTGCAGCCGCAGCAGCA  
 61 G Q G G Y E G P G A G Q G A G A A A A A  
 241 GCTGGTGGTGCCGGACAAGGAGGATATGGAGGTCTTGGTGGCCAAGGAGCTGGACAAGGA  
 81 A G G A G Q G G Y G G L G G Q G A G Q G  
 301 GCTGGAGCCGCAGCTGCAGCAGCTGGTGGTGCCGGACAAGGAGGATATGGAGGTCTTGGT  
 101 A G A A A A A A G G A G Q G G Y G G L G  
 361 GGCCAAGGAGCTGGACAAGGAGCTGGAGCCGCAGCTGCAGCAGCTGGTGGTGGCCGGACAA  
 121 G Q G A G Q G A G A A A A A A G G A G Q  
 421 GGAGGATATGGAGGTCAAGGTGCTGGACAAGGTGCAGCAGCAGCAGCAGCTGGTGGTCT  
 141 G G Y G G Q G A G Q G A A A A A A G G A  
 481 GGACAAGGAGGATATGGAGGCCTAGGTTCTGGACAAGGCGGATATGGTAGACAAGGTGCC  
 161 G Q G G Y G G L G S G Q G G Y G R Q G A  
 541 GGAGCAGCAGCAGCAGCAGCAGCCGCAGGTGGAGCTGGACAAGGCGGATATGGAGGTCTT  
 181 G A A A A A A A A G G A G Q G G Y G G L  
 601 GGTGGCCAAGGAGCTGGAGCCGCAGCTGCAGCAGCTGGTGGTCTGGACAAGGAGGATAT  
 201 G G Q G A G A A A A A A G G A G Q G G Y  
 661 GGAGGTCAAGGTGCTGGACAAGGTGCAGCCGCAGCAGCAGCTAGTGGTGCCGGACAAGGA  
 221 G G Q G A G Q G A A A A A A S G A G Q G  
 721 GGATATGGAGGTCCAGGTGCCGGACAAGGTGCAGGTGCAGCCGCAGCAGCAGCTGGTGGT  
 241 G Y G G P G A G Q G A G A A A A A A G G  
 781 GCCGGACAAGGAGGATATGGAGGTCTTGGTGGCCAAGGAGCTGGACAAGGAGCTGGAGCC  
 261 A G Q G G Y G G L G G Q G A G Q G A G A  
 841 GCAGCTGCAGCAGCTGGTGGTGCCGGACAAGGAGGATATGGAGGTCAAGGTGCTGGACAA  
 281 A A A A A G G A G Q G G Y G G Q G A G Q  
 901 GGTGCAGCAGCAGCAGCAGCTGGTGGTGCCGGACAAGGAGGATATGGAGGCCTAGGTTCT  
 301 G A A A A A A G G A G Q G G Y G G L G S



181 GGAGGTTACGGAAGAGGTGCTGGAGCCGGAGCTGGAGCTGGAGCTGGAGCTGCTGCCGGA  
61 G G Y G R G A G A G A G A G A G A A A G  
241 GCTGGAGCTGGAGCTGCTGCTGGAGCAGGAGCTGGTGGTGGTGGAGGATATGGCGCTGGA  
81 A G A G A A A G A G A G G A G G Y G A G  
301 CAAGGCTATGGTGCAGGTGCAGGAGTTGGTGGTGGTGGAGCAGGTGCAGGA  
101 Q G Y G A G A G V G A A A A A G A G A G  
361 GTTGGTGGTGGTGGAGGTTACGGAAGAGGTGCTGGAGCTGGAGCTGGTGGTGGTGGTGGT  
121 V G G A G G Y G R G A G A G A G A G A G  
421 GGTGCAGGAGGTTACGGAAGAGGTGCTGGTGGTGGAGCTGGAGCTGGTGCAGGAGCTGGT  
141 G A G G Y G R G A G A G A G A G A G A G  
481 GGTGCTGGAGGATATGGCGCTGGACAAGGCTATGGTGCAGGTGCAGGAGCTGGTGCAGCC  
161 G A G G Y G A G Q G Y G A G A G A G A A  
541 GCCGCTGCTGGAGCCGGTGCAGGTGCTGGTGGTGGTGGAGGTTACGGAAGAGGTGCTGGT  
181 A A A G A G A G A G G A G G Y G R G A G  
601 GCTGGAGCTGGAGCTGGTGCAGGAGCTGGTGGTGGTGGAGGATATGGCGCTGGACAAGGC  
201 A G A G A G A G A G G A G G Y G A G Q G  
661 TATGGTGCAGGTGCAGGAGCTGGTGCAGCCGGCTGGAGCCGGTGCAGGTGCTGGT  
221 Y G A G A G A G A A A A A G A G A G A G  
721 GGTGCTGGAGGTTACGGAAGAGGTGCTGGTGGTGGTGGAGCTGGAGCTGGTGCAGGAGCTGGT  
241 G A G G Y G R G A G A G A G A G A G A G  
781 GGTGCTGGAGGATATGGCGCTGGACAAGGCTATGGTGCAGGTGCAGGAGCTGGTGCAGCC  
261 G A G G Y G A G Q G Y G A G A G A G A A  
841 GCCGCTGCTGGAGCCGGTGCAGGTGCTGGTGGTGGTGGAGGTTACGGAAGAGGTGCTGGT  
281 A A A G A G A G A G G A G G Y G R G A G  
901 GCTGGAGCTGGAGCTGGTGCAGGTGCTGGTGGTGGTGGAGGTTACGGAAGAGGTGCTGGA  
301 A G A G A G A G A G G A G G Y G R G A G  
961 GCTGGAGCTGGGGCTGGAGCTGCTGCAGGAGCAGGAGCTGGTGGTGGTGGACGATATGGC  
321 A G A G A G A A A G A G A G G A G R Y G  
1021 GCTGGACAAGGCTATGGTGCAGGTGCAGGAGCTGGGGCTGCCCGGCTGCAGGAGCAGGT  
341 A G Q G Y G A G A G A A A G A G A G  
1081 GGTGCTGGAGGATATGGCGCTGGACAAGGCTATGGTGCAGGTGCAGGAGCTGGTGGTGGC  
361 G A G G Y G A G Q G Y G A G A G A G A A  
1141 GCCGCTGCTGGAGCAGGTGCAGGAGTTGGTGGTGGTGGAGGTTACGGAAGTGGTGGTGGT  
381 A A A G A G A G V G G A G G Y G S G A G  
1201 GCCGAGCTGGAGCTGGAGCTGGAGCTGCTTCTGGAGCTGCTGCTGGAGCTGCTGCTGGT  
401 A G A G A G A G A A S G A A A G A A A G  
1261 GCAGGAGCTGGTGGTGGTGGAGGATATGGCACTGGACAAGGCTATGGTGCAGGTGCAGGT  
421 A G A G G A G G Y G T G Q G Y G A G A G

1321 GCTGGAGCTGGTGCTGGTGCTGGTGCTGGTGCTGGTGCTGGTGCTGGTGCTGGTGCTGGAGCTGGA  
441 A G A G A G A G G A G G Y G R G A G A G  
  
1381 GCTGGTGCAAGGAGCTGGTGCTGGAGGATATGGCGCTGGACAAGGCTATGGTGCAAGT  
461 A G A G A G G A G G Y G A G Q G Y G A G  
  
1441 GCAGGAGCTGGTGCAAGCCCGCTGGTGCTGGAGACGGTGCAAGTGGTGCTGGTGCTGGAGT  
481 A G A G A A A A A G D G A G A G G A G G  
  
1501 TACGGAAGAGGTGGTGCTGGAGCTGGAGCTGGGGCTGGAGCTGGTGCAAGGAGCAGGAGCTGGT  
501 Y G R G A G A G A G A G A A A G A G A G  
  
1561 GGTGGTGCAAGGATATGGCGCTGGACAAGGCTATGGTGCAAGTGGTGCAAGGAGCTGGGGCTGGC  
521 G A G G Y G A G Q G Y G A G A G A G A A  
  
1621 GCCGGTGCAAGGAGCAGGTGGTGCTGGAGGATATGGCGCTGGACAAGGCTATGGTGCAAGT  
541 A G A G A G G A G G Y G A G Q G Y G A G  
  
1681 GCAGGAGCTGGTGCTGGCGCCGCTGGTGCAAGGAGTGGTGCTGGAGTGGTGCTGGAGT  
561 A G A G A A A A A G A G A G V G G A G G  
  
1741 TACGGAAGAGGTGGTGCAAGCCGAGCTGGAGCTGGTGCTGGCGAGCTGGAGCTGGAGCTGGT  
581 Y G R G A G A G A G A A A G A G A G A A  
  
1801 GCTGGAGCAGGAGCTGGTGCTGGAGGATATGGCACTGGACAAGGCTATGGTGCAAGT  
601 A G A G A G G A G G Y G T G Q G Y G A G  
  
1861 GCAGGTGGTGCTGGTGCTGGTGCTGGTGCTGGTGCTGGTGCTGGTGCTGGTGCTGGAGT  
621 A G A G A G A G G A G G Y G R G A G A G  
  
1921 GCTGGAGCTGGTGCAAGGAGCTGGTGCTGGAGAATATGGCGCTGGACAAGGCTATGGT  
641 A G A G A G A G G A G E Y G A G Q G Y G  
  
1981 GCAGGTGGTGCTGGTGCAAGCCCGCTGGTGCAAGCCGGTGCAAGTGGTGCTGGAGTGGT  
661 A G A G A G A A A A A G A G A G A G G A  
  
2041 GGAGGTGGTGCTGGAGCTGGAGCTGGGGCTGGAGCTGGTGCAAGGAGCAGG  
681 G G Y G R G A G A G A G A A A G A G  
  
2101 GCTGGTGCTGGAGGATATGGCGCTAGACAAGGCTATGGTGCAAGTGGTGCAAGGAGCTGGG  
701 A G G A G G Y G A R Q G Y G A G A G A G  
  
2161 GCTGGCGCCGGTGCAAGGAGCTGGAGTGGTGCTGGAGGATATGGCGCTGGACAAGGCTATGGT  
721 A A A G A G A G G A G G Y G A G Q G Y G  
  
2221 GCAGGTGGTGCTGGTGCAAGCCCGCTGGTGCAAGCCGGTGCAAGTGGTGCTGGTGCTGGAGT  
741 A G A G A G A A A A A G A G A G A G G A  
  
2281 GGAGGTGGTGCTGGAGCTGGAGCTGGAGCTGGGGCTGGAGCTGGTGCAAGGAGCAGG  
761 G G Y G R G A G A G A G A G A A A G  
  
2341 CAAGGCTATGGTTCAGGTGGTGCTGGAGCTGGTGCCAGTGGTGCTGGTGCTGGAGGAGT  
781 Q G Y G S G A G A G A G A S A G G A G S  
  
2401 TACGGAAGAGGTGGTGCTGGCGCCGCTGGTGCAAGCCGGTGCTGGAGGATAT  
801 Y G R G A G A G A A A A S G A G A G G Y

2461 GCGCTGGACAAGGCTATGGTGCAGGTGCAGGAGCTGTTGCTTCTGCCGCTGCTGGAGCC  
 821 G A G Q G Y G A G A G A V A S A A A G A  
  
 2521 GGTTCCAGGAGCTGGTGGTGCTGGAGGTTACGGAAGAGGTGCCGTTGCTGGTTCTGGAGCT  
 841 G S G A G G A G G Y G R G A V A G S G A  
  
 2581 GGTGCCGGAGCAGGAGCTGGTGGTGCTGGAGGATATGGTGCAGGAGCTGGTGGTGGCC  
 861 G A G A G A G G A G G Y G A G A G A G A  
  
 2641 GCTGCTGGAGCAGTTGCTGGTGGTTCTGGAGGATATGGCGGCAGACAAGGCGGTTATAGC  
 881 A A G A V A G G S G G Y G G R Q G G Y S  
  
 2701 GCAGGTGCGGGAGCTGGTGCAGGCTGCTGCTGGAGCAGGTGCAGGTGGAAGTGGAGGC  
 901 A G A G A G A A A A A G A G A G G T G G  
  
 2761 TACGGAAGAGGTTCTGGTGGTGCTGGAGCCGCAGCTGGTGGTGCTGCTGGAGCTGGTGGTGCT  
 921 Y G R G S G A G A A A G A A A G A G A A  
  
 2821 GGAGGATATGGTGGCTATGGCGCAGGTGCTGGAGCTGGTGGCGGTGGTGCTAGAGGTTAC  
 941 G G Y G G Y G A G A G A G A G G A R G Y  
  
 2881 GGAGGAGGTGCTGGTGGTGCTGGAGCAGGTGCCGCTGCTGGAGGTTATGGAAGAAGAGCAGGT  
 961 G G G A G A G A G A A A G G Y G R R A G  
  
 2941 GGATCCATTGTAGGAACTGGAATAAGTCAATTTCTTCTGGAAGTGGTTCTAGCTATTCC  
 981 G S I V G T G I S A I S S G T G S S Y S  
  
 3001 GTTCTTCCGGCGGTTACGCCTCTGCGGGTGTAGGTGTTGGATCCACTGTTGCATCTACC  
 1001 V S S G G Y A S A G V G V G S T V A S T  
  
 3061 ACATCTCGTTTTGAGTTCAGCACAAAGCATCTTCTAGAATATCTGCTGCTGCTTCTACTTTA  
 1021 T S R L S S A Q A S S R I S A A A S T L  
  
 3121 ATATCTGGAGGTTACTTGAATACATCTGCCTTACCATCAGTCATTTCTGATTTGTTTGGC  
 1041 I S G G Y L N T S A L P S V I S D L F A  
  
 3181 CAAGTCAGTGCTTCATCCCCTGGTGTATCAGATAGTGAAGTTTTGATTCAAGTTTTGTTG  
 1061 Q V S A S S P G V S D S E V L I Q V L L  
  
 3241 GAAATTGTTTCTTCCCTTATCCATATTCTCAGTTCTTCCAGTGTAGGGCAAGTTGACTTC  
 1081 E I V S S L I H I L S S S S V G Q V D F  
  
 3301 AATTCTGTTGGTTCGTCTGCTGCAGCTGTTGGACAGTCCATGCAAGTTGTCATGGGTTAA  
 1101 N S V G S S A A A V G Q S M Q V V M G \*  
  
 3361 TTAATGAGTGTCTCTCCCAATTAATTCTTTAAATACAGTTAAGCATTAAATAATAA  
 1121 L K W L S L P N \* F F K Y S \* A F K N K  
  
 3421 AAATAATGTAAATTTTTCTGCATAAATAAATAATTTTTCTGCTTGGAAAAAAAAAAAAA  
 1141 K \* C K I F C I N K N I F L L G K K K K  
  
 3481 AAAAA  
 1161 K

*Clone 11F12*

1 TACGGAAGAGGTGCTGGTGCTGGAGCCGGAGCTGGTGCTGGAGCTGCCGCCGGAGCAGGA  
1 Y G R G A G A G A G A G A G A A A G A G  
61 GCTGGTGGTGGTGGAGGATATGGCGCTGGACAAGGCTATGGTGCAGGTGCAGGAGCTGGT  
21 A G G A G G Y G A G Q G Y G A G A G A G  
121 GCAGCCGCCGCTGCTGGAGCCGGTGCAGGTGCTGGTGGTGGTGGAGGTTTCGGAAGAGGT  
41 A A A A A G A G A G A G G A G G F G R G  
181 GCTGGTGCTGGAGCCGGAGCTGGTGCTGGAGCTGCCGCCGGAGCAGGAGCTGGTGGTGGT  
61 A G A G A G A G A G A A A G A G A G G A  
241 GGAGGATATGGCGCTGGACAAGGCTATGGTGCAGGTGCAGGAGCTGGTGCAGCCGCCGCT  
81 G G Y G A G Q G Y G A G A G A G A A A A  
301 GCTGGAGCTGGTGCCGGTGGTGGTGGTGGTGGAGGTTACGGAAGAGGTGCTGGTGCTGGA  
101 A G A G A G A G G A G G Y G R G A G A G  
361 GCCGGAGCTGGTGCTGCCGCCGCTGCTGGAGCAGGAGCTGGTGGTGGTGGAGGATATGGC  
121 A G A G A A A A A G A G A G G A G G Y G  
421 GCTGGACAAGGCTATGGTGCAGGTGCAGGAGCTGGTGCTGCCGCCGCTGCTGGAGCCGGT  
141 A G Q G Y G A G A G A A A A A G A G  
481 GCAGGAACTGGTGGTGGTGGAGGATATGGTGCAGGACAAGGCTATGGTGCAGGTGCTGGA  
161 A G T G G A G G Y G A G Q G Y G A G A G  
541 GCTGGTGCAGGTGCTGGAGCTGGTGCCGGTGGTGGTGGTGGTGGAGGTTACGGAAGAGGT  
181 A G A G A G A G A G A G G A G G Y G R G  
601 GCTGGTGCTGGAGCTGGGGCCGGAGCAGGAGCTGGTGGTGGTGGAGGATATGGCGTCGGA  
201 A G A G A G A G A G A G G A G G Y G V G  
661 CAAGGCTATGGTGCAGGTGCTGGAGCTGGTGCTGGAGCAGGAAGTGCAGCTGGAAACGCT  
221 Q G Y G A G A G A G A G S A A G N A  
721 TTCGACAATCCCTCTCATCCAATCTCCTCTCATCAGGAGATTTTGTTCAAATGATCTCC  
241 F A Q S L S S N L L S S G D F V Q M I S  
781 ACCACAACTTCTACTGATCAGGCTGTGAGCGTTGCAACGAGCGTTGCTCAAAATGTTGGA  
261 T T T S T D Q A V S V A T S V A Q N V G  
841 AATCAACTTGGCCTCGATGCCAATGCTATGAACAATTTACTGGCTGCTGTTGGTGGGTAT  
281 N Q L G L D A N A M N N L L A A V G G Y  
901 GTTTCATCATTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT  
301 V S S L G G A V A D A A A Y A N A I S S  
961 GCTATTGGCAATGTTTTAGCTAACACTGGTTCTATTAATGAAAGTACCGCATCTTCCGCT  
321 A I G N V L A N T G S I N E S T A S S A  
1021 GCTTCGAGTGCTGCTTCGTGCGTTACAACAACATTGACATCTTATGGGCCAGCCGTATTT  
341 A S S A A S S V T T T L T S Y G P A V F

1081 TACGGAAGAGGTGCTGGAGCTGGTGCTGCAGCCGCTGCTGGAGCAGGTGCAGGTGGCGCT  
 361 Y G R G A G A G A A A A G A G A G G A  
  
 1141 GGAGGATACGGAAGTGGTGCTGGAGCTGGTGGTGCTGGAGGATATGGTGCAGGAGCTGGT  
 381 G G Y G S G A G A G G A G G Y G A G A G  
  
 1201 GCTGGTGGCGCTGCTGGAGCAGGAGCTGGTGGTTCTGGAGGATATGGCGGCAGACAAGGT  
 401 A G A A A G A G A G G S G G Y G G R Q G  
  
 1261 GGTTATGGCGCAGGTGCTGGAGCTGGTTCC  
 421 G Y G A G A G A G S

*Seqüência do cDNA parcial de NCTuSp-like*

1 GCAAGCCAGAGCGCTAGCAGCAGCAGTGCTTCGGCCTCTGCATTTCGCACAACAGTCCTCT  
 1 A S Q S A S S S A S A S A F A Q Q S S  
  
 61 GCTTCCCTTGCAGCCTCCTCTTTTTCAGCCAGGCCTTCGCTTCGGCCGCTTCCGCTCT  
 21 A S L A A S S S F S Q A F A S A A S A S  
  
 121 GCCGTGCGAAACGTTGCTTACCAGCTAGGCTTATCCGCAGCACAATCTCTCGGAATAGCC  
 41 A V G N V A Y Q L G L S A A Q S L G I A  
  
 181 AATGCTGGAGCACTCGCTAGTGCTTTAGCTCAGTCTGTTTCTTCTGTAGGCGTTGGAGCC  
 61 N A G A L A S A L A Q S V S S V G V G A  
  
 241 AGTTCAAGTGCCTACGCCAATGCAGTCGCCGGTGCCGTTGGACAGTTCTTAGCCAATCAG  
 81 S S S A Y A N A V A G A V G Q F L A N Q  
  
 301 GGTATTTTGAACACAGGCAATGCATCTTCCCTAGCCTCCTCGTTCTCCAGTGCCCTCTCC  
 101 G I L N T G N A S S L A S S F S S A L S  
  
 361 GCCTCGGCAGCAGCCGCGCAATCCCAATCATTTCGCACAGAGTCAAGCAGCAGCTTCGGCC  
 121 A S A A A A Q S Q S F A Q S Q A A A S A  
  
 421 TTCCAACAAGCAGCATCACAGAGTGCTAGCCAGAGTGCTGCCCAATCTGGTTCTCAGTCC  
 141 F Q Q A A S Q S A S Q S A A Q S G S Q S  
  
 481 TCTTCCACCACTACCACCACCTCGGCCTCAGGAAGTCAATCCGCAAGCCAGAGCGCTAGC  
 161 S S T T T T S A S G S Q S A S Q S A S  
  
 541 AGCAGCAGTGCTTCGGCCTCTGCATTTCGCACAACAGTCCTCTGCTTCCCTTGCAGCCTCC  
 181 S S S A S A S A F A Q Q S S A S L A A S  
  
 601 TCTTCTTTTCAGCCAGGCCTTCGCTTCGGCCGCTTCCGCTCTGCCGTGCGAAACGTTGCT  
 201 S S F S Q A F A S A A S A S A V G N V A  
  
 661 TACCAGCTAGGCTTATCCGCAGCACAATCTCTCGGAATAGCCAATGCTGGAGCACTCGCT  
 221 Y Q L G L S A A Q S L G I A N A G A L A  
  
 721 AGTGCTTTAGCTCAGTCTGTTTCTTCTGTAGGCGTTGGAGCCAGTTCAAGTGCCTACGCC  
 241 S A L A Q S V S S V G V G A S S S A Y A

781 AATGCAGTCGCCGGTGCCGTTGGACAGTTCTTAGCCAATCAGGGTATTTTGAACACAGGC  
261 N A V A G A V G Q F L A N Q G I L N T G  
  
841 AATGCATCTTCCCTAGCCTCCTCGTTTTCTAATGCGCTTTCGTTCATCCGCCGCTAATTCA  
281 N A S S L A S S F S N A L S S S A A N S  
  
901 GTTGGTTCTGGATTGTTATTGGGTCCTTACAATACGTTGGAAGTATTGCTCCAAGTATA  
301 V G S G L L L G P S Q Y V G S I A P S I  
  
961 GGAGGTGCTGCTGGAATATCAATCGCTGGTCCTGGAATTTTATCATACTTACCTCCTGTT  
321 G G A A G I S I A G P G I L S Y L P P V  
  
1021 TCTCCGCTGAATGCACAGATAATCTCCTCTGGTTTTACTTGCTTCTTTGGCACCAGTATTA  
341 S P L N A Q I I S S G L L A S L A P V L  
  
1081 TCATCTTCCGGCTTAGCATCATCCAGTGCGACTTCTAGAGTTGGCAGTTTGTCTCAATCT  
361 S S S G L A S S S A T S R V G S L A Q S  
  
1141 TTGGCATCCGCATTGCAATCTTCCGGGAGGTACACTGGATGTTTTCGACCTTCTTGAATCTT  
381 L A S A L Q S S G G T L D V S T F L N L  
  
1201 CTGTCTCCCATTTCTACACAAATTCAAGCCAATACTTCTCTAAATGCATCACAGGCGATT  
401 L S P I S T Q I Q A N T S L N A S Q A I  
  
1261 GTCCAAGTTTTACTTGAAGCTGTAGCTGCTCTGCTGCAAATTATCAACGGAGCTCAAATA  
421 V Q V L L E A V A A L L Q I I N G A Q I  
  
1321 ACTTCTGTCAATTTTGGCAGTGTCTCCAGCGTAAACACAGCCTTGGCAACTGCTCTCGCT  
441 T S V N F G S V S S V N T A L A T A L A  
  
1381 GGTTGATTTTTATGTCCCTTTTGAAGAATTCTTTCCTACTGGAACTTATAAAATGTAT  
461 G \* F L C P F \* R I L C L L E T Y K M Y  
  
1441 AATCTTTATTTTATTTCTGATTTCAACTCAATATTGCATTTCGTTATCTTTGCTTGTCTGT  
481 N L Y F I S D F N S I L H S L S L L V C  
  
1501 GCAATTATTGATTTTTAAAATTATATTATGAACCCTGAAATTTTCCTGATAATAAATATT  
501 A I I D F \* N Y I M N P E I F L I I N I  
  
1561 CTTGAATGCCAA  
521 L E C Q K  
  
1621 AAAAAAAAAAAAAAAAAA  
541 K K K K K

## Seqüências de cDNA das espidroínas de *A. juruensis*

### Seqüência do cDNA parcial de Espidroína 1

#### Espidroína 1A

1 TACTCTCTAGCAAGCTCCATTGCAAGCGCTGCATCCTCGAGTGCATCTTCGGCAGCAGCA  
1 Y S L A S S I A S A A S S S A S S A A A

61 GCGGCGTCATCTTCTTCCGCGAGCAGCAGGAGCAGCCGCGGCTTCGGAAGCAGCAGCTTCT  
21 A A S S S S A A A G A A A A S E A A A S

121 GCCGCCGCCACTTCCACGACAACAACAAGAAGTACTTCTCGTGCCGCGAGCAGCAGCATCC  
41 A A A T S T T T T T S T S R A A A A A S

181 GCCGCAGCCGCGGCTCTGCCTCGGGAGCCGCCGGCGCAGCGGGAGCAGCATCAGCCGCT  
61 A A A A A S A S G A A G A A G A A S A A

241 AGCGCTGCTTCAGCTTCTTCGTCCTTGCAACAGTCTCTGGGATCTGCCTTAGCACAAAGT  
81 S A A S A S S S L Q Q S L G S A L A Q S

301 AGCTCATTTCGAGCAGCCTTCGCCCCAAGCAAGTAGCGCTGCTTCTGCAGCAGCCATAGCA  
101 S S F A A A F A Q A S S A A S A A A I A

361 TATGCTCTTGCACAGACCGTGGCGAATCAAATCGGTTTCTTCTTCTACTCCTCAGCTTTC  
121 Y A L A Q T V A N Q I G F S S Y S S A F

421 GCAAGAGCAGCTTCATCAGCCGTATACAGCATAGGGGGCTTGGCTTCTGCATCTGCATAT  
141 A R A A S S A V Y S I G G L A S A S A Y

481 GCCTTTGCTTTTGGCAGCGCCTTTTACAAGTTCTCTCAAATTACGGTTTACTTAAACATA  
161 A F A F A S A F S Q V L S N Y G L L N I

541 AATAACGCGTACTCTCTAGCAAGCTCCATTGCAAGCGCTGCATCCTCGAGTGCATCTTCG  
181 N N A Y S L A S S I A S A A S S S A S S

601 GCAGCAGCAGCAGCGGCGTCATCTTCTTCCGCGAGCAGCAGGAGCAGCCGCGGCTTCAGGT  
201 A A A A A A S S S S A A A G A A A A S G

661 ACAGCAGCTTCTGCCGCCGCCACTTCCACCACCACAACAAGAAGTACTTCTAGAGCCGCT  
221 T A A S A A A T S T T T T T S T S R A A

721 GCAGCAGCATCCGCCGCGAGCCGCGGCTCTGCCTCGGGAGCCGCCGACGCAGCGGGAGCA  
241 A A A S A A A A A S A S G A A D A A G A

781 GCATCAGCCGCTAGCGCTGCTTCAGCTTCTTCGTCCTTGCAACAATCTCTGGGATCTGCC  
261 A S A A S A A S A S S S L Q Q S L G S A

841 TTAGCACAAAGTAGCTCATTTCGAGCAGCCTTCGCCCCAAGCAAATAGCGCTGCTTCTGCA  
281 L A Q S S S F A A A F A Q A N S A A S A

901 GCAGCCATAGCATATGCTCTTGCACAGACCGTGGCAAATCAAATCGGTTTCTTCTCTAC

301           A A I A Y A L A Q T V A N Q I G F S S Y  
961           TCCTCAGCTTTTCGCAAGCGCAGCTTCTTCAGCCGTATCCAGCTTAGGGGGCTTCGCTTCT  
321           S S A F A S A A S S A V S S L G G F A S  
1021          GCATCTGCATATGCCTTTGCTTTTGCCAGCGCCTTTTACAAGTTCTCTCAAATTACGGT  
341           A S A Y A F A F A S A F S Q V L S N Y G  
1081          TTACTTAACATAAATAACGCCTACTCTCTAGCAAGCTCCATTGCAAGCGCTGCATCCTCG  
361           L L N I N N A Y S L A S S I A S A A S S  
1141          AGTGCATCTTCGGCAGCAGCAGCGGCATCATATTCCTTCTCAGCAACAGGAGCAGCCTCT  
381           S A S S A A A A A S Y S F S A T G A A S  
1201          TCGGCAGCAGTAGGTGCGGCAGCGACATCTGGTGCAGCGACATCTGGTGCAGCGACTTCC  
401           S A A V G A A A T S G A A T S G A A T S  
1261          TCTAGCTCTGCGACGGGTGTCGGAGGAAGTGTCTCCTCTGGAGCATCACCCGCTTCCGCT  
421           S S S A T G V G G S V S S G A S P A S A  
1321          GGAAGTCAACAGGTGGCGGTATCTCATTTCTACCTGTCCAGACACAACGTGGTTTCGGG  
441           G T A T G G G I S F L P V Q T Q R G F G  
1381          CTTGTGCCCTCTCCTTCAGGTAATATTGGTGCAAATTTTCCTGGTTCTGGTGAATTTGGT  
461           L V P S P S G N I G A N F P G S G E F G  
1441          CCATCACCTTTGACATCACCAGTTTATGGTCCGGGTATTCTTGGCCCTGGGCTTGTCTG  
481           P S P L T S P V Y G P G I L G P G L V V  
1501          CCCTCATTACAGGGGCTGTTGCCACCTTTATTTGTTTTACCATCGAATTCGGCAACTGAA  
501           P S L Q G L L P P L F V L P S N S A T E  
1561          AGAATTTTCGTCCATGGTATCGTCTTTGTTGTCCGCAGTTTCTTCCAATGGATTGGATGCT  
521           R I S S M V S S L L S A V S S N G L D A  
1621          TCTTCTTTTGGTGATAACCATAGCTTCCCTGGTTTCGCAGATATCCGTGAATAATTCCGAT  
541           S S F G D T I A S L V S Q I S V N N S D  
1681          CTTTCTTCGTCACAAGTCTTGCTTGAGGCGCTCCTTGAAATTTTGTCTGGAATGGTACAA  
561           L S S S Q V L L E A L L E I L S G M V Q  
1741          ATCCTTTCTTATGCTGAAGTCGGGACTGTTAATACGAAGACCGTGAGTTCAACTTCCGCT  
581           I L S Y A E V G T V N T K T V S S T S A  
1801          GCTGTGGCTCAAGCTATCTCTTCGGCTTTTTTCGGGAAATCAGAATTCTTGAGCTGCCTAA  
601           A V A Q A I S S A F S G N Q N S \* A A \*  
1861          TGAAGTTTTTTTTTTCATCAAATATTTTTAAAATATTATGACCCACTGATTTAATTTTTAT  
621           \* R F F F I K Y F \* N I M T H \* F N F Y  
1921          TACTATCAATATTGGAAGTGAATTTAATAGGTGTTGTTATTTCTGCTGTGTAATGTTGG  
641           Y Y Q Y W K \* N L I G V V I S A V \* C W  
1981          TAATGGTTGTAATGTAAGTAGTATGGTATTGGTAAAAAAAAAAAAAAAAAAAAAAAAAAAA  
661           \* W L \* M \* L V W Y W \* K K K K K K K K  
2041          AAAAAAAAAAAAAAAAAAAAAAAAAAAA

681 K K K K K K K K K

*Espidroína 1B*

1 ACAACAAGCACTTCTACAGCCGCGAGCAGCAGCCGCGAGCAGCGGACTTCGGCTCGGGAGCC  
1 T T S T S T A A A A A A A A D F G S G A  
61 GCCCGCGCAGCGCAAACAGCATCAGCCGCTAGCGCTGCTTCAGCTTCTTCGTCCTTGCAA  
21 A R A A Q T A S A A S A A S A S S S L Q  
121 CAGTCTTTGGGATCTGCCTTAGCACAAAGTAGCTCATTTGCAGCAGCCTTCGACCAAGCA  
41 Q S L G S A L A Q S S S F A A A F D Q A  
181 AATAGCGCTGCTTCTGCAGCAGCCATAGCATAACGCTCTTGCACAGTCCGCGGCGAATCAA  
61 N S A A S A A A I A Y A L A Q S A A N Q  
241 GTCGGTTTGTCTTCTACTCCGCGAGCTATCTCAAACGCAGCTGCAGCAGCCGTAGGAAGC  
81 V G L S S Y S A A I S N A A A A A V G S  
301 GTAGGTGGCTACGCTTCTGCATCTGCCTATGCCTTTGCTTTTGCCAGCGGCGTTTCACAA  
101 V G G Y A S A S A Y A F A F A S G V S Q  
361 GTTCTATCAAATTACGGTTTAATTAACCTAAGTAACGCCTTATTTTTAGCAAGTTCGATA  
121 V L S N Y G L I N L S N A L F L A S S I  
421 GCAAACGCTGCATCGGCGAGTGCATCTTCGGCAGCAGCAGCGGCGTCATCTTCTTCCGCA  
141 A N A A S A S A S S A A A A A S S S S A  
481 GCAACAGGAGCAGCCGCGGCTTTGGGAGGCGCTGGTTCTGCCGCCGCACTTCCACCACC  
161 A T G A A A A L G G A G S A A A T S T T  
541 ACAATAACAAGCACTTCTACAGCCGTAGCAGCAGCCTCTGGCTCGGGAGCCGCCCGCGCA  
181 T I T S T S T A V A A A S G S G A A R A  
601 GCGCAAACAGCATCAGCCGCTAGCGCTGCTTCAGCTTCTTCGTCCCTTGACAGTCTTTG  
201 A Q T A S A A S A A S A S S S L A Q S L  
661 GGATCTGCCTTAGCACAAAGTAGCTCATTTGCAGCAGCCTTCGACCAAGGCAATAGCGCT  
221 G S A L A Q S S S F A A A F D Q G N S A  
721 GCTTCTGCAGCAGCCATAGCATAACGTTCTTGCACAGTCCGCGGCGAATAAAGTCGGTTTG  
241 A S A A A I A Y V L A Q S A A N K V G L  
781 TCTTCTACTCCGCGAGCTATCTCAAACGCTGCTTCAGCAGCCGTAGAAAGCGTAGGTGGC  
261 S S Y S A A I S N A A S A A V E S V G G  
841 TACGCTTCCGCATCTGCCCATGCCTTTGCTTTTGCCAGCGCCGTTTCACAAGTTCTATCA  
281 Y A S A S A H A F A F A S A V S Q V L S  
901 AATTACGGTTTAATTAACCTAAGTAACGCCTTGTCCCTAGCAAGTTCGATAGCAAACGCT  
301 N Y G L I N L S N A L S L A S S I A N A  
961 GTATCGGCGAGTGCATCTTCGGCAGCAGCTGTGTCATCTGCTGCAGCAGCAACAGGTGCA  
321 V S A S A S S A A A V S S A A A A T G A

1021 ACCTCTTCGGCAGCAGTAGGTGCAGCAGCGACATGTGGGGCAGCGACTTCCGCTAGTTCT  
341 T S S A A V G A A A T C G A A T S A S S  
  
1081 GCGACGGGGCGTCCGAGAAACTGTTGCCTGTGCAACATCGCCCGGTCCACTGGAACCGCG  
361 A T G V G E T V A C A T S P A S T G T A  
  
1141 GCAGGTGGCGGTATCTCATCTTTACCTGTTTCAGACACAACCTGGTTTTGGGTTTTTGCTC  
381 A G G G I S S L P V Q T Q P G F G F L L  
  
1201 TCTCCCTCAGGTAATATTGGTCCAAGTGTCTGGTTCTGGTGGGTTTGGTCCATCACCT  
401 S P S G N I G P S V S G S G G F G P S P  
  
1261 TTGCCATCTCCAGCTTCTGACGGATTTAGCCCATCGCCTTTGCCATCACAAGTTTATGGT  
421 L P S P A S D G F S P S P L P S Q V Y G  
  
1321 CCTGGTATTCTTGGTCCCGGTCTCGTCGCACCTTCGTTAGAAGGGCTGTTGCCACCTTTA  
441 P G I L G P G L V A P S L E G L L P P L  
  
1381 TCAATTTTGCCATCGGATTCTGCAAATGAAAGAATTTTCGTCTGTAGTATCTTCTTTGTTG  
461 S I L P S D S A N E R I S S V V S S L L  
  
1441 GCCGCCGTTTCTTCCAATGGATTGGATGCTTCTTCTTGGCGATAACTTAGCTTCACTG  
481 A A V S S N G L D A S S L G D N L A S L  
  
1501 GTTTCGCAGATATCCGCGAATAATGCCGATCTTTCTTCGTCACAAGTTATGGTTGAGGCT  
501 V S Q I S A N N A D L S S S Q V M V E A  
  
1561 CTTCTTGAAGTTTTGTCTGGAATAGTTCAGATCCTTTCTTATGCTGAAGTTGGGGCTGTT  
521 L L E V L S G I V Q I L S Y A E V G A V  
  
1621 AATACGGAAACCGTAAGTTCAACTTCTCTGCTGTGGCTCAAGCTATTTCTTCGGCTGTT  
541 N T E T V S S T S S A V A Q A I S S A V  
  
1681 TTGGGATAATCAAAATTCTTGAGCTGCCTAATGAAACTGTTTTTTTTTTAAACAAATATTT  
561 L G \* S K F L S C L M K L F F F \* Q I F  
  
1741 TAAAAATATTATGGCCCACTGATTTAATTTTCATTAGTATCAATGTTGGAAGTGGGAATT  
581 \* K Y Y G P L I \* F S L V S M L E V G I  
  
1801 TAATATGTTTTGTTTATTTCTGCTGTGTAATGTTGTTAATGGTTGTATATGTAAC TAGTA  
601 \* Y V L F I S A V \* C C \* W L Y M \* L V  
  
1861 TGGTATTGGTAATAAAAAACATTGCATTTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  
621 W Y W \* \* K H C I \* K K K K K K K K K K  
  
1921 AAAAA  
641 K

*Espidroína 1C*

1 TCAAATTACGGTTTGCTTAACATAAATAACGCTTACTCTCTAGCAAGTTCGATTGCAAAC  
1 S N Y G L L N I N N A Y S L A S S I A N

61 GCTGCATCAGCGAGTGCATCTTCTGCAGCAGCGGCAGCGGCGTCATCGTCTTCCGCAGCA  
21 A A S A S A S S A A A A A S S S S A A  
121 GCAGCAGCAGCCGCGGCTTCGGGAGCAGCAGGTTCTGCCGCCGCACTTCCACCACAACA  
41 A A A A A A S G A A G S A A A T S T T T  
181 TCAACAAGCTCTTCTACAGCCGCAGCAGCAGCATCCGCCGCGGCTTCAGCCGCCGCTTCC  
61 S T S S S T A A A A A S A A A S A A A S  
241 GCTTCGGGAGCCGCCGCGCTGCGGGAGCGTCATCAGCCGCTAGCGATGCTTCAGCTTCT  
81 A S G A A R A A G A S S A A S D A S A S  
301 TCGTCCTTGCAACAGTCTCTGGGATCTGCTTTAGCACAAAGTAGCTCATTTGCAGCAGCC  
101 S S L Q Q S L G S A L A Q S S S F A A A  
361 TTCGCCAAGCAAATAGCGCTGCTTCTGCAGCAGCCATAGCATATGCTCTTGCACAGATC  
121 F A Q A N S A A S A A A I A Y A L A Q I  
421 GTGGCGAATCAAATCGGTTTCTTCTTCTACTCCTCAGCTTTTGCAAGCGCCGCTCATCA  
141 V A N Q I G F S S Y S S A F A S A A S S  
481 GCCGTATCCAGCTTAGGGAGCTTCGCTTCTGCATCTGCATATGCCTTTGCTTTTGCCAGC  
161 A V S S L G S F A S A S A Y A F A F A S  
541 GCCTTTTCACAAGTTCTCTCAAATTACGGTTTGCTTAACATAAGTAACGCTTACTCTCTA  
181 A F S Q V L S N Y G L L N I S N A Y S L  
601 GCAAGTTCGATTGCAAGCGCTGCATCAGCAAGTGCATCTTCTGCAGCAGCGGCAGCGGCG  
201 A S S I A S A A S A S A S S A A A A A A  
661 TCATCGTCTGCCGAGCAGCAGGAGCAGCCGCGGCTTCGGGAGCAGCAGGTTCTGCCGCC  
221 S S S A A A A G A A A A S G A A G S A A  
721 TCCACTTCCACCACAACATCAACAAGCACTTCTACAGCCGCAGCAGCAGCATCCGCCGCG  
241 S T S T T T S T S T S T A A A A A S A A  
781 GCTGCAGCCGCCGCTCCGCTTCGGGAGCCGCCGCGCTGCGGGAGCGTCATCAGCCGCT  
261 A A A A A S A S G A A R A A G A S S A A  
841 AGCGCTGCTTCAGCTTCTTCGTCCTTGCAACAGTCTCTGGGATCTGCTTTAGCACAAAGT  
281 S A A S A S S S L Q Q S L G S A L A Q S  
901 AGCTCATTTGCAGCAGCCTTCGCCAAGCAAATAGCGCTGCTTCTGCAGCAGCCATAGCA  
301 S S F A A A F A Q A N S A A S A A A I A  
961 TATGCTCTTGCACAGACCGTGGCGAATCAAATCGGTTTCTTCTTCTATTCTCAGCTTTT  
321 Y A L A Q T V A N Q I G F S S Y S S A F  
1021 GCAACCGCCGCTCATCAGCCGATCCAGCTTAGGGAGCTTCGCTTCTGCATCTGCATAT  
341 A T A A S S A V S S L G S F A S A S A Y  
1081 GCCTTTGCTTTTGCCAGCGCCTTTTCACAAGTTCTCTCAAATTACGGTTTGCTTAACATA  
361 A F A F A S A F S Q V L S N Y G L L N I  
1141 AGTAACGCTTACTCTCTAGCAAGTTCGATTGCAAACGCTGCATCAGCGAGTGCATCTTCT  
381 S N A Y S L A S S I A N A A S A S A S S

1201 GCAGCAGCGGCAGCGGCGTCATCTTCTTCCGCAGCAGCAGGAGCAGCCGCAGCTTCGGGA  
401 A A A A A A S S S S A A A G A A A A S G  
1261 GCAGCAGGTTCTGTGCGCCGCACTTCCACCACCACAACAGCAAGCACTTCGACAGCCGCA  
421 A A G S V A A T S T T T T A S T S T A A  
1321 GCAGCAGCATCAGCCGCGGCTGCAGCCGCGCCTCGGCTTCGGGAGCCGCCCGCGCTGCG  
441 A A A S A A A A A A S A S G A A R A A  
1381 GGAGCGTCATCAGCCGCTAGCGCTGCTTTCAGCTTCTTCGTCCTTGCAACAGTCTCTGGGA  
461 G A S S A A S A A S A S S S L Q Q S L G  
1441 TCTGCTTTAGCACAAAGTAGCTCATTTGCAGCAGCCTTCGCCCAAGCAAATAGCGCTGCT  
481 S A L A Q S S S F A A A F A Q A N S A A  
1501 TCTGCAGCAGCTATAGCATATGCTCTTGCACAGACCGTGGCGAATCAAATCGGTTTCTCT  
501 S A A A I A Y A L A Q T V A N Q I G F S  
1561 TCATACTCCTCAGCTTTTGCAAGCGCCGCTCATCAGCCGTATACAGCTTAGGCAGCTTC  
521 S Y S S A F A S A A S S A V Y S L G S F  
1621 GCTTCTGCATCTGCATATGCCTTTGCTTTTGCCAGCGCCTTTTACAAGTTCTCTCAAAT  
541 A S A S A Y A F A F A S A F S Q V L S N  
1681 TACGGTTTGCTTAACATAAATAACGCTTACTCTCTAGCAAGTTGATTGCAAACGCTGCA  
561 Y G L L N I N N A Y S L A S S I A N A A  
1741 TCAGCGAGTGCATCTTCTGCAGCAGCGGCAGCGGCGTCATCGTCTTCCGCAGCAGCAGGA  
581 S A S A S S A A A A A A S S S S A A A G  
1801 GCAGCCGCGGCTTCGGGAGCAGCAGGTTCTGCCGCTCCACTTCCACCACAACATCAACA  
601 A A A A S G A A G S A A S T S T T T S T  
1861 AGCACTTCTACAGCCGAGCAGCAGCATCCGCCGCTCCGCTTCGGGAGCCGCCCGCGCT  
621 S T S T A A A A A S A A S A S G A A R A  
1921 GCGGGAGCGTCATCAGCCGCTAGCGCTGCTTTCAGCTTCTTCGTCCTTGCAACAGTCTCTG  
641 A G A S S A A S A A S A S S S L Q Q S L  
1981 GGATCTGCTTTAGCACAAAGTAGCTCATTTGCAGCAGCCTTCGCCCAAGCAAATAGCGCT  
661 G S A L A Q S S S F A A A F A Q A N S A  
2041 GCTTCTGCAGCAGCCATAGCATATGCTCTTGCACAGGCCGTGGCGAATGAAATCGGTTTC  
681 A S A A A I A Y A L A Q A V A N E I G F  
2101 TCTTTCTACTCCTCAGCTTTTGCTAGCGCAGCTTCATCAGCCGTATTTCGGCTTAGGGAGC  
701 S F Y S S A F A S A A S S A V F G L G S  
2161 TTCGTTTCTGCATCTGCATATGCCTTTGCTTTTGCCAGCGCCTTTTACAAGTTCTCTCA  
721 F A S A S A Y A F A F A S A F S Q V L S  
2221 AATTACGGTTTGCTTAACATAAATAACGCCTACTCTCTAGCAAGCTCGATTGCAAACGCT  
741 N Y G L L N I N N A Y S L A S S I A N A  
2281 GCATCAACGAGTGCATCTTCCGCAGCAGCAGCAGTAGTGGCGTCATCTTCTTCTGCA  
761 A S T S A S S A A A A A V V A S S S S A

2341 GCAACAGCTGCAGGCGCGGCATCGACATCTGTTCCCGCGACTTCCTTGAGTTCTGCGACC  
 781 A T A A G A A S T S V P A T S L S S A T  
 2401 AGAGTCGGTGGAAGTCTCTCTCTGCAGTTTTACCCGCTTCTGCTAGAACC GCAACAGGT  
 801 R V G G S L S S A V S P A S A R T A T G  
 2461 GACGGTACCACATATCTACCTGTCCAGATACAACCTGGTATCGGGTTTGTGCCTTCTCTT  
 821 D G T T Y L P V Q I Q P G I G F V P S L  
 2521 TCAGGTGATATTGGCCCAAATGTTCTGGTTCTGGTGGATTTGGTTCACCAGCTTTGCCA  
 841 S G D I G P N V P G S G G F G S P A L P  
 2581 TCCCCAGTTTATGGTCCTGCTATTCTTGGTCCC GGTCTTGTTCGCACCTGCATTAGCGAAT  
 861 S P V Y G P A I L G P G L V A P A L A N  
 2641 CTGTTGCCACCATTATCAGTTTTTACCATCGGATTCTGCAAATGAAAGAATTTCTCCGTC  
 881 L L P P L S V L P S D S A N E R I S S V  
 2701 GTATCTTCTTTGTTGTCCGCAATTTCTTCCAATGGATTGGATGCTTCTTCTTCTGGCGGT  
 901 V S S L L S A I S S N G L D A S S L G G  
 2761 ACCATAGCTTCACTAGTTTTCGCAGATATCCGTGAGTAATGCCAAACTTTCTTCGTCACAA  
 921 T I A S L V S Q I S V S N A K L S S S Q  
 2821 GTCTTTCTTGAGGCTCTTCTTGAAGTTTTGTCTGGAATGGTTCAGATTCTTTCCTATGCT  
 941 V F L E A L L E V L S G M V Q I L S Y A  
 2881 GAAGTTGGGGCTGTTAATACAGACACCGTAATTTCAACTTCGTCTGCTGTGGCTCAGGCT  
 961 E V G A V N T D T V I S T S S A V A Q A  
 2941 ATCTCTTCGGCTGTTTCGGGATAAACTGTATTTTTTTTTCAACGAATGTTTATAAAACGTT  
 981 I S S A V S G \* T V F F F N E C L \* N V  
 3001 ATGTTCCACTCATTTAATTTTTAATTGTACCAATATTGGAAGTGGGAATTTAGTATGTGT  
 1001 M F H S F N F \* L Y Q Y W K W E F S M C  
 3061 TGTTATTTCTGCTGTGTTATGTTGGTAATGGTTGTAAATGTAAGTACTAGTGTGGTATTGATA  
 1021 C Y F C C V M L V M V V N V T S V V L I  
 3121 ATAAAAACATTGCATTTAAAAA AAAAAAAAAAAAAA  
 1041 I K T L H L K K K K K

*Seqüência do cDNA parcial de Espidroína 2*

1 TCGGGATCTGGTTCTGGAAGTGGAGCAGGGTCTGGTGGAGGAAGTGGTGCGGGATCTGGT  
 1 S G S G S G S G A G S G G G S G A G S G  
 61 TCTGGAAGCGGAGCAGGAGCAGGATCTGGTAGTGGCTCAGGTT CAGGAAGTGGAGCAGGA  
 21 S G S G A G A G S G S G S G S G S G A G  
 121 TCTGGTAGTGGCTCAGGTT CAGGAAGTGGAGCAGGTT CAGGTAAGTGGCTCAGGTT CAGGA  
 41 S G S G S G S G S G A G S G S G S G S G  
 181 AGTGGAGCACGATCTGGTAGTGGCTCAGGTT CAGGAAGTGGAGCAGGTT CGGGTAGTGGC

61 S G A R S G S G S G S G S G A G S G S G  
241 TCAGGTTTCAGGAAGTGGGGCAGGTGCAGGTAGTGGTTTCAGGTTTCAGGAAGTGGAGCAGGT  
81 S G S G S G A G A G S G S G S G S G A G  
301 TCAGGTAGTGGAGCAGGTTTCAGGTAGTGGCTCAGGTTTCAGGAAGTGGGGCAGGTTTCAGGT  
101 S G S G A G S G S G S G S G S G A G S G  
361 AGTGGCTCAGGTTTCAGGAAGTGGGGCAGGCGCAGGTAGTGGCTCAGGTTTCAGGACGTGGA  
121 S G S G S G S G A G A G S G S G S G S G R G  
421 GCAGGTTTCAGGTAGTGGTTTCAGGTTTCAGGAAGTGGAGCAGGTTTCAGGTAGTGGCTCAGGT  
141 A G S G S G S G S G S G A G S G S G S G  
481 TCAGGACGTGGAGCAGGTTTCAGGTAGTGGCTCAGGTTTCAGGAAGTGGGGCAGGTGCAGGT  
161 S G R G A G S G S G S G S G S G A G A G  
541 AGTGGCTCAGGTTTCAGGAAGTGGAGCAGGTTTCAGGTAGTGGCTCAGGTTTCAGGAAGTGGGA  
181 S G S G S G S G A G S G S G S G S G S G  
601 GCAGGTTTCAGGTAGTGGCTCAGGTTTCAGGAAGTGGAGCTGGATCTGGTAGTGGCTCAGGT  
201 A G S G S G S G S G S G A G S G S G S G  
661 TCAGGAAGTGGAGCAGGATCTGGTAGTGGCTCAGGTTTCAGGAAGTGGAGCAGGTTTCAGGT  
221 S G S G A G S G S G S G S G S G A G S G  
721 AGTGGCTCAGGTTTCAGGAAGTGGAGCAGGTTTCAGGTAGTGGCTCAGGTTTCAGGAAGTGGGA  
241 S G S G S G S G A G S G S G S G S G S G  
781 GCAGGCTCAGGTAGTGGCTCAGGTTTCAGGAAGTGGAGCAGGAGCTGGTAGTGGTTTCAGGT  
261 A G S G S G S G S G S G A G A G S G S G  
841 TCATGTAGAAAAGATGCAGGTGGTCATGATGGCGGATATGGGAAAAGCTTGGTTTTGAA  
281 S C R K D A G G H D G G Y G K K L G F E  
901 TTCGGTACGCCTGCAGCAGCAGCTGTTACCCTTGGACCTGGAGCTGGACAACAAGGCCCA  
301 F G T P A A A A V T L G P G A G Q Q G P  
961 GGTGGAGCTGGACAACAAGGACCAGGAGGCCAAGGACCATATGGACCAGTTGCTAGCGCC  
321 G G A G Q Q G P G G Q G P Y G P V A S A  
1021 GCCGCAGCTGTTGCTGGAGGTTATGGACCTGGAGCTTTACCACAAGGACCAGCAGCCAA  
341 A A A V A G G Y G P G A L P Q G P A R Q  
1081 GGACCTTCCGGTCCTGTTTCTTCCAGCACCAGTTGCATCGGCAGCTGCTGCTCGCCTTCT  
361 G P S G P V S S A P V A S A A A A R L S  
1141 TCTCCTCAGGCTAGTTCTAGAGTATCTTCCAGCTTTTTTTTTCTTTGGTATCAAGTGGTCCA  
381 S P Q A S S R V S S A F F S L V S S G P  
1201 ACTAGTCCTGGTGCACCTTTCTAATGCCATCAGTAGTGTGTTTTCACAAGTTAGTGCAAGC  
401 T S P G A L S N A I S S V V S Q V S A S  
1261 AATCCAGGTCTCTCTGGTTGCGATGTACTCGTGCAAGCATTGCTGGAAATTGTATCCGCC  
421 N P G L S G C D V L V Q A L L E I V S A  
1321 CTTGTATCTATCCTTGCCTCATCTAGTATCGGGCAAATCAACTATGGAGCTTCCGCTCAG

441 L V S I L A S S S I G Q I N Y G A S A Q  
1381 TATGCCTCTTTGGTTGGCCAATCTGTAAATCAAGCTCTTCGTTATTAATTTAGCAAATGA  
461 Y A S L V G Q S V N Q A L R Y \* F S K \*  
1441 TTTGCAAACCTTTTTTCAATGTTACTAACACATACTTTTTAAAATTTCTCAATAAATTTGAA  
481 F A N F F Q C Y \* H I L L K F L N K F E  
1501 GCATATTATATTTCTCTTGTGTTATTTATTTGTTACATGCGGAGATGAACATTGATCCT  
501 A Y Y I S S C V I Y L L H A E M N I D P  
1561 GTTACAAATTTATATTTAAAATTATTTCTTTAAATAATCGAAAGTGGATTAAAAGTACTT  
521 V T N L Y L K L F L \* I I E S G L K V L  
1621 TTACAAAACCTTTGCATTTAGATTTTCATGAAAAAATATTTGTTTCAGCGTTAGTAAACGAT  
541 L Q N F A F R F H E K I F V S A L V N D  
1681 AAACATTTTTGGTCCTATCAATTATTAATTTTTTTTTATAATCTTTTGATTGCCATATTTA  
561 K H F W S Y Q L L I F F I I F \* L P Y L  
1741 TAATTTCTTTAAAATTATTTACGATTTCTCCTACATTTTCTTTTTTAAATCCATTTTCAT  
581 \* F L \* N Y L R F L L H F L F \* I H F H  
1801 GTGTCTTTCAAGAATTTTGTGATTAAATGTGGTATTTTTTTCATGATAAAAAAAAAAAAAA  
601 V S F K N F V I K C G I F S \* \* K K K K  
1861 AAAAAAAAAAAAAA  
621 K K K K

### **ANEXO 3**

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, 2007. Spidroins from the Brazilian spider *Nephilengys cruentata* (Araneae: Nephilidae). *Comp Biochem Physiol B Biochem Mol Biol.* 4, 597-606.

Provided for non-commercial research and educational use only.  
Not for reproduction or distribution or commercial use.



This article was originally published in a journal published by Elsevier, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues that you know, and providing a copy to your institution's administrator.

All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at:

<http://www.elsevier.com/locate/permissionusematerial>



ELSEVIER

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

ScienceDirect

Comparative Biochemistry and Physiology, Part B 147 (2007) 597–606

CBP

[www.elsevier.com/locate/cbpb](http://www.elsevier.com/locate/cbpb)

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

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

<sup>a</sup> Instituto de Ciências Biológicas, Departamento de Biologia Celular, Universidade de Brasília, Brasília—DF, Brazil

<sup>b</sup> Department of Molecular Biology, Bioinformatics Group, University of Wyoming, Laramie—WY, USA

<sup>c</sup> Laboratório Especial de Toxicologia Aplicada de Artrópodes-Instituto Butantan, São Paulo—SP, Brazil

<sup>d</sup> Núcleo de Biotecnologia, EMBRAPA Recursos Genéticos e Biotecnologia, Brasília—DF, Brazil

<sup>e</sup> 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)<sub>n</sub>, (GGX)<sub>n</sub>, (GPGGX)<sub>n</sub>). 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.

© 2007 Elsevier Inc. All rights reserved.

**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)<sub>n</sub>, alternating glycine and alanine (GA)<sub>n</sub>, amino acid triplets composed of two glycines and a third variable amino acid (GGX)<sub>n</sub>, and glycine-proline-glycine modules (GPGXX)<sub>n</sub>

\* 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](mailto:dmatias@cenargen.embrapa.br) (D. Bittencourt).

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  $\beta$ -sheets thought to provide tensile strength,  $(GGX)_n$  probably forms a Gly-II helical structure, and  $(GPGXX)_n$  forms  $\beta$ -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 (Iftah 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* DH5 $\alpha$  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 [ $\alpha$ - $^{32}$ P]dCTP in Southern Blots analysis (Sambrook and Russell, 2001) for identification of additional related sequences in the silk gland libraries.

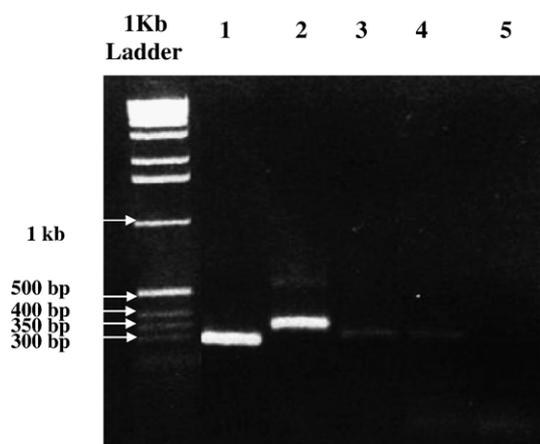


Fig. 1. RT-PCR was performed using total RNA from silk glands. 1, Flagelliform spidroin cDNA (287 bp); 2, MaSp cDNA (356 bp); 3, MiSp cDNA (304 bp); 4, TuSp cDNA (302 bp); 5, negative control. 1 Kb DNA ladder used was obtained from Invitrogen. The oligonucleotides used in the PCR are described in the 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 forward-TGCAGGTCAAGGTGGATATGGTG, Major ampullate spidroin reverse-CCTAGAGCTTGGTAAACCGATTGAC, Minor ampullate spidroin forward-CTCTGCGGTGTAGGTGTTGG, Minor ampullate spidroin reverse-TGCAGCAGACGAACCAACAGA, Flagelliform spidroin forward-GCATATG-GAGCTGGTTCTGGTACAC, Flagelliform spidroin reverse-CTTCGAACAGAGTTGGAATATGCAT, Tubuliform spidroin forward-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/>) (Kato 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.



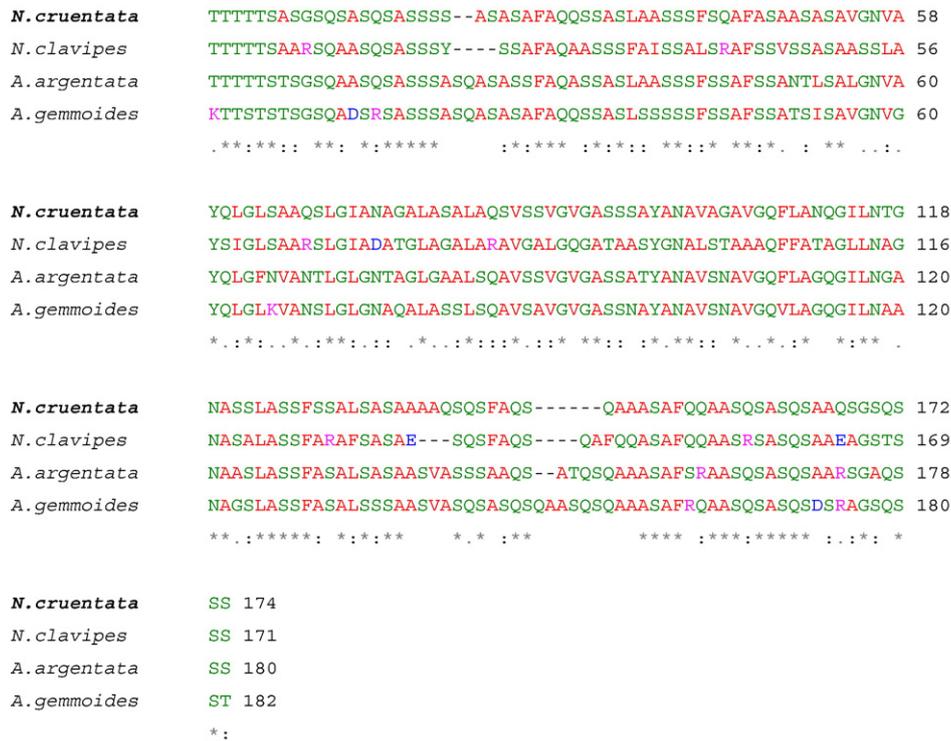


Fig. 4. ClustalW alignment of the consensus repetitive amino acid sequence of tubuliform proteins from different orb-web weaver spiders. Sequence alignment abbreviations are the same as in Fig. 2. Sequence from *N. cruentata* is identified in bold.

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)<sub>n</sub> and polyAla, this being replaced by (GA)<sub>n</sub> 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)<sub>n</sub> 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<sub>n</sub>, (SX)<sub>n</sub> (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 NCMAp1-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 NCMAp1-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 NCMAp1-like, NCMiSp1-like and NCFlag-like proteins, the codon usages for alanine, serine and glutamine, the three most

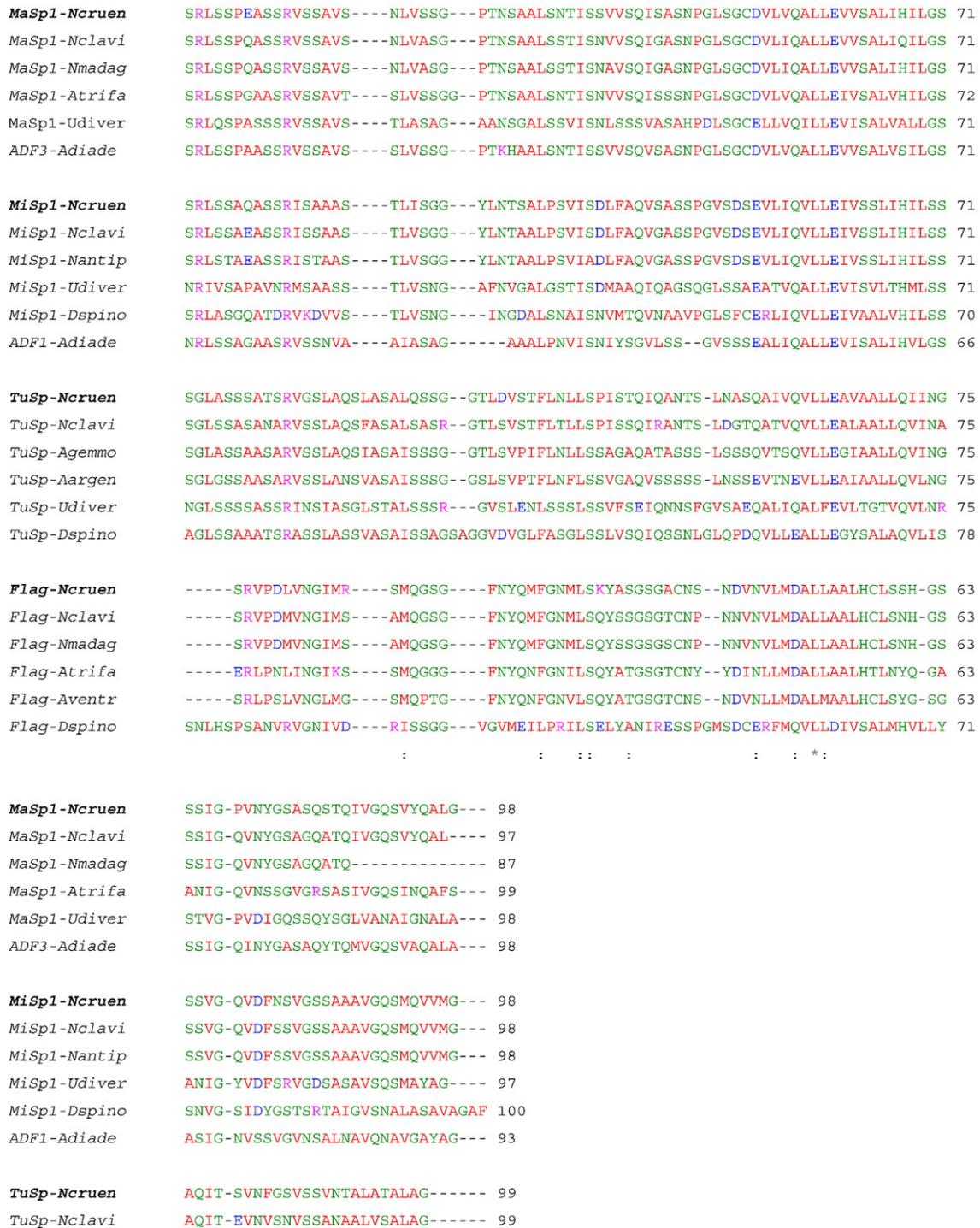


Fig. 5. ClustalW alignment of C-terminal amino acid sequences. Amino acids are indicated by one letter abbreviations and numbered from N- to C-terminal. Sequences alignment abbreviations are the same as in Fig. 2. Sequences from *Nephilengys cruentata* are identified in bold. *Araneus diadematus* ADF-3 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, *Deinopsis 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-SVNLANVPNVQQALVLSALSG-----	99
TuSp-Aargen	AQIT-SVNLARNV-----	86
TuSp-Udiver	GQTS-FVSVSSPTVISSSF-----	93
TuSp-Dspino	SQIS-SVSVSSSALGPALLNLYLVG-----	102
<b>Flag-Ncruen</b>	<b>PSFGSSPTPSAMNAYSNSVRRMPQF-----</b>	<b>87</b>
Flag-Nclavi	SSFAPSPPTPAAMSAYSNSVGRMFAY-----	87
Flag-Nmadag	SSFAPSPPTPAAMSAYSNSVGRMFAY-----	87
Flag-Atrifa	SYVPSYSPSEMLSYTENVRRYF-----	85
Flag-Aventr	S-VPSTPTYSAMSAYNQSI RRMFTY-----	86
Flag-Dspino	EDVRRGIPGDTAEAVANAVAGVVLVSVV----	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.,

-lnL: 16456.89

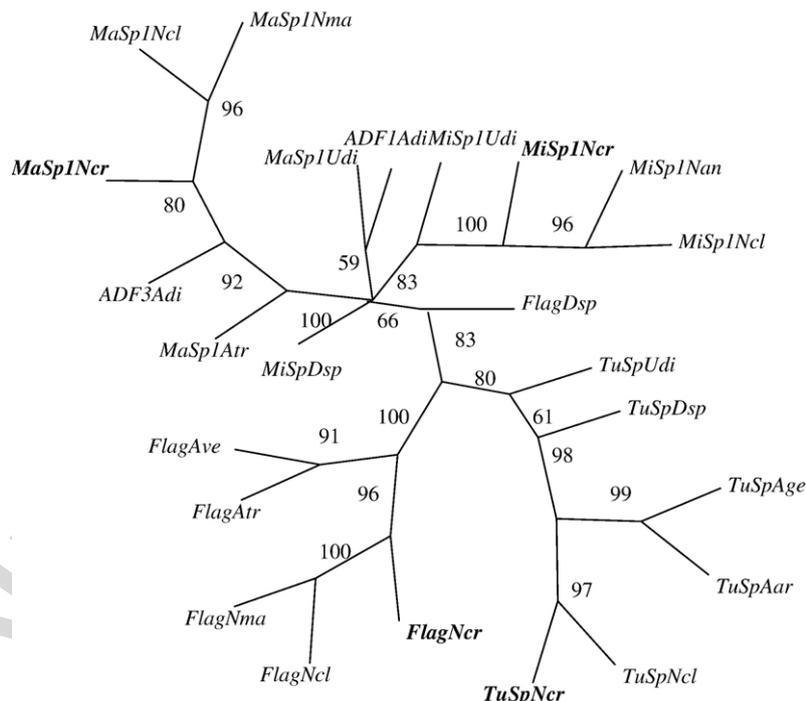


Fig. 6. Unrooted tree from spider silk C-terminal sequence phylogenetic analysis. *Nephilengys cruentata* spidroins are identified in bold. Abbreviations of spider species used in this figure: Ncr, *N. cruentata*; Ncla, *Nephila clavipes*; Nma, *Nephila inaurata madagascariensis*; Atr, *Argiope trifasciata*; 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)<sub>n</sub>, (GPGGX)<sub>n</sub>) (Hinman and Lewis, 1992). Such regions do not exist in tubuliform 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)<sub>n</sub> 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)<sub>n</sub> motifs are responsible for another interprotein link similar to the  $\beta$ -sheet. Although major silks present a high content of (GGX)<sub>n</sub> 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)<sub>n</sub> and (GGX)<sub>n</sub> 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)<sub>n</sub> 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)<sub>n</sub> (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  $\beta$ -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<sub>n</sub>, (SX)<sub>n</sub>, 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  $\beta$ -sheet. It also showed that eggcase silk has a larger *b* dimensional value in the  $\beta$ -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 spidroin 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  $\alpha$ -helices (Spohner 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 (Spohner 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 Científico e Tecnológico-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.
- Itah, 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., Engstrom, 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.
- Spohner, 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.

## ANEXO 4

### How old are spider major ampullate silks?

**Daniela Bittencourt<sup>a,c,b</sup>; Katharina Dittmar<sup>b,d</sup>; Randolph V. Lewis<sup>b</sup>; Elíbio L. Rech<sup>c</sup>**

- a) Departamento de Biologia Celular, Universidade de Brasília, Brazil
- b) Department of Molecular Biology, University of Wyoming
- c) Núcleo de Biotecnologia, EMBRAPA, Brazil
- d) Department of Biological Sciences, SUNY Buffalo

The most elaborate spider webs have evolved within the araneomorph orb-weaving spiders – the Araneoidea and the Deinopoidea. Their nets have intricate frames and radial support lines, which are spun from silks with the ability to absorb kinetic energy produced by flying and struggling prey. Both lineages are highly likely to have evolved from one common ancestor, and two spider silks (major ampullate spidroin 2 – MaSp2, and flagelliform spidroin - Flag) have been identified as supporting orbicularian monophyly uniquely (1). Mygalomorphae however, the sistergroup to the Araneomorphae, do not build orb webs, and only use their silk for seemingly simple tasks as protecting their eggs or lining burrows. Not only is this the reason why not much is known about mygalomorph spidroins, but also why they are often regarded as “primitive” spiders.

It is well known that spiders use specific assemblages of silks to accomplish different tasks, and accordingly, it is tacitly assumed that mygalomorph spiders generally

do not have silks similar to those used in orb web weaving. Additionally, this line of thought is equated with them having “ancestral” silks (phylogenetically basal). In our quest to find the original “ancestral” silk in a basal mygalomorph spider – *Avicularia juruensis* (Theraphosidae) – we came across an interesting discovery, which challenges common views about spider silk evolution.

Three types of silk, produced by different glands, are thought to be responsible for the production of fibers used in the orb-web; MaSp 1 and 2, minor ampullate spidroin (MiSp) and Flag (1). From the cDNAs of the mono-glandular *Avicularia juruensis*, we identified two expressed spidroins (7). While “spidroin1” was the more abundant transcript, and by initial BLAST searches most similar to the mygalomorph *Euagrus chisoseus* spidroin, “spidroin 2” showed clear similarities to MaSp2 from the orb-weaving araneoid clade (supplementary material).

To validate these findings, phylogenetic analyses involving 77 silk C-terminals from 35 spider species were conducted (7). Consistent with previous results, most sequences form orthologous clusters according to function, interspersed with terminals that do not seem to belong to either functional group (1,2,3) (Fig. 1). Within some of these clades patterns of spider speciation are clearly preserved. Generally, no support was found for any of the basal nodes, thus, no inferences can be made as to the succession of descent among the major functional silk clades, and the notion that mygalomorph spidroins are “ancestral” silks cannot be corroborated.

More importantly, however, mygalomorph “spidroin 2” clearly nests within orbicularian MaSp2 sequences, and is positioned in the same cluster encompassing a sequence from another mygalomorph spider – *Macrothele holsti* - which was classified as

a MaSp1 silk (4). From an evolutionary perspective, the most likely reasoning is that mygalomorph MaSp-like silks are the survivors of gene duplications (3), which would imply that the origin of orb weaver MaSps should be placed to a time point before the origin of Orbicularia. In fact, reconciling the species tree with the silk gene tree does just that; it pinpoints the oldest speciation in which MaSps must have been present on the mygalomorph-araneomorph split, 340 – 390 MYA (5,7). Therefore, while not refuting orb weaver monophyly, MaSp2s, and major ampullate silks in general cannot be classified as orbicularian synapomorphies (1), but have to be considered plesiomorphic for Opisthothelae.

The overall phylogenetic pattern attests to a major influence of gene duplication in silk evolution, aside from other following processes such as gene conversion or intragenic homogenization (2,3). Given the apparent importance of gene duplication for the evolution of new biological functions, this makes sense for spider silks, and it is conceivable that the functional basis for their different uses arose at the outset of spider radiation, even before gland or spinneret differentiation. In order to advance in the production of synthetic silk, it is imperative to fully understand the connection between silk sequence divergence and expression, gland anatomy, and function (6). Therefore, in the future more emphasis should be placed on sampling mygalomorph and araneomorph non-orbicularian taxa.

## References and Notes

1. J.E. Garb, T. DiMauro, V. Vo, C. Hayashi, *Science* **312**, 1762 (2006).

2. J.E. Garb, C. Hayashi, *PNAS* **102**, 11379-11384 (2005).
3. R.J. Challis, S.L. Goodacre, G.M. Hewitt, *Insect Molecular Biology* **15**, 45-56 (2006).
4. P.L. Tai, G.Y. Hwang, I.M. Tso, *Int. J. Biol. Macromol.* **34**, 295-301 (2004).
5. N.A. Ayoub, J.E. Garb, M. Hedin, C. Hayashi. *MPE* **42**, 394-409 (2007).
6. A. Spöner *et al.*, *Biochem. Biophys. Res. Commun.* **338**, 897 (2005).
7. Materials and Methods online.
8. Betúlia M. Souto provided assistance. Megan L. Porter commented on drafts. Work was supported by CNPq – Brazil (#486492/2006-0, DB), NIH (RVL) and AFOSR (RVL). Sequences were deposited with accession # X and X. IBAMA collecting permit: 1984-06

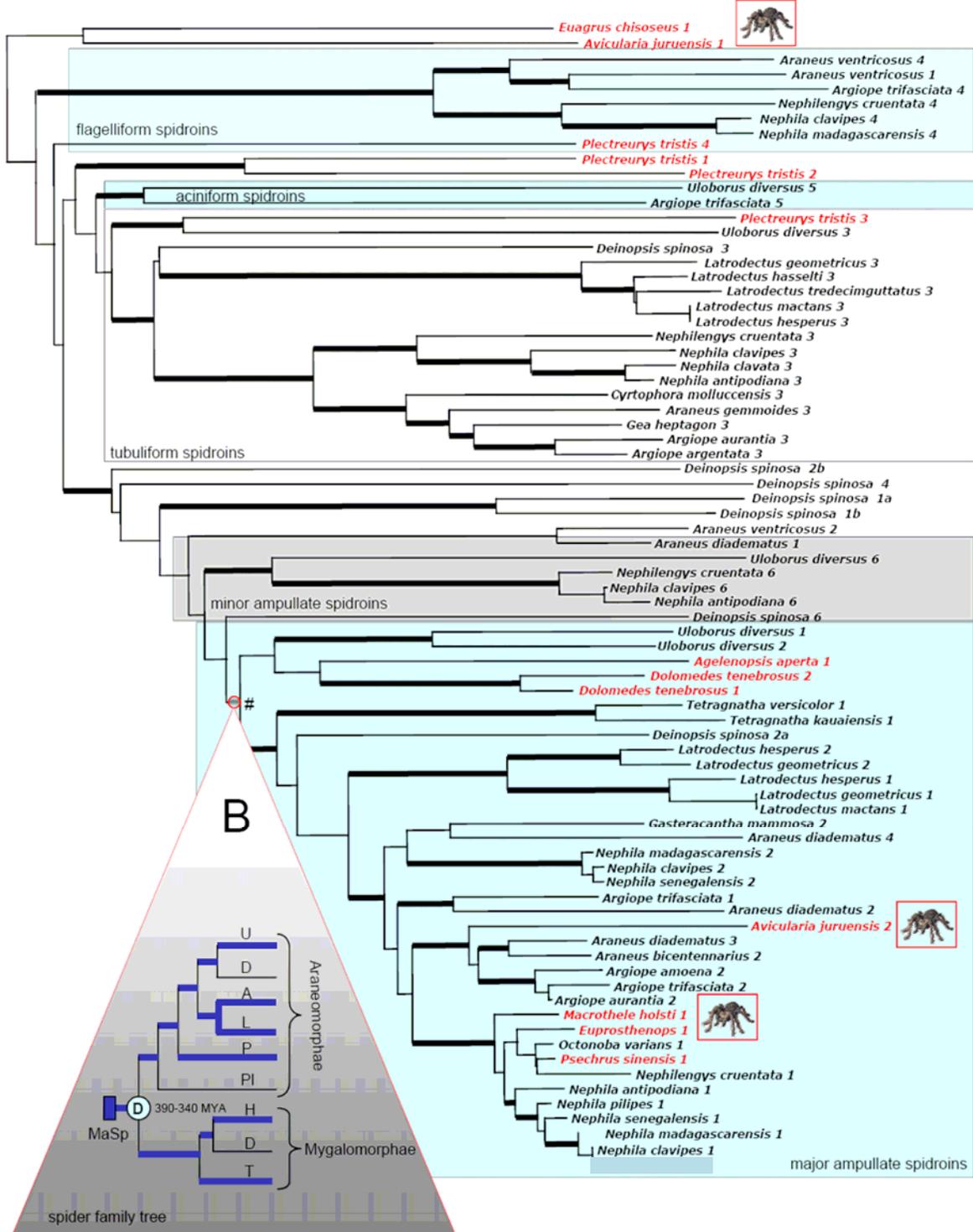
### Figure Legends:

#### Figure 1.

**A.** ML phylogram (lnL = -7671.60611). Bold branches = BP>75. Red Branches = Non-orbicularian taxa, black = Orbicularia. # = MaSp node estimated to be result of duplication (among 23 others, not shown).

**B.** Spider family tree using ML data from Ayoub et al. 2007. U = Uloboridae, D = Deinopidae, A = Araneidae, L = Latrodectidae, P = Pisauridae, Pl = Plectreuridae, T = Theraphosidae, D = Dipluridae, H = Hexathelidae). D = Lower bound of duplication at mygalomorph-araneomorph split, date from Ayoub et al. 2007. Blue branches: Presence of MaSPs.

A



**Supporting Online Material for**

**How old are major ampullate silks?**

Daniela Bittencourt; Katharina Dittmar\*; Randolph V. Lewis; Elíbio L. Rech

\*To whom correspondence should be addressed. E-mail: [katharinad@gmail.com](mailto:katharinad@gmail.com)

**This supporting material includes:**

Materials and Methods

References

Figs. S1 and S2

## Materials and Methods

Two *Avicularia juruensis* spiders were collected in the Amazon Forest native regions in Brazil (Monte Negro/RO S10°17'40"/W 60°19' 31"). *Avicularia* has only one gland, and each was isolated under a stereomicroscope and immediately frozen in liquid nitrogen. The cDNA libraries were constructed using “SUPERSCRIPT II Plasmid System with GATEWAY Technology” (Invitrogen, USA), according to the manufacturer’s instructions. From the clones we identified two distinct silk cDNAs called *Spidroin 1* (3154bp) and *Spidroin 2* (1874bp) (Fig. S1, S2). The respective forward and reverse primers used for each spider silk cDNA (RT-PCR) were: AJSpid1f: 5'-TGTC CGCAGTTTCTTCCAATGG-3', JSpid1r:5'CCACAGCAGCGGAAGTTGAAC-3', AJSpid2f:5'GGACCAGCACGCCAAGGACC-3' andAJSpid2r: 5' CGAAGAGCTTGATTACAGATTGGCC-3'.

Silks were classified into orthologous groups according to established criteria (S1). Top BLASTX hit to *A. juruensis* “spidroin 1” in open reading frame was *E. chisoseus* ( $E = 9e-06$ ), whereas *A. juruensis* “spidroin 2” blasted to *Argiope trifasciata* MaSp 2 ( $E = 5e-79$ ). Phylogenetic relationships of the spidroins reported here (GenBank accession# X and X) were analyzed with publicly available data. From the Araneomorphae we included *Plectreurys tristis* [1 (AAK30610); 2 (AAK30611); 3 (AAK30612); 4 (AAK30613)], *Deinopsis spinosa* [1b (ABD61592); 1a (ABD61591); 2a (ABD61593); 2b (ABD61594); 4 (ABD61590)], *Dolomedes tenebrosus* [1 (AAK30598); 2 (AAK30599)], *Argiope aurantia* [2 (AAK30592); 3 (AAX45292)], *Argiope trifasciata* [1 (AAK30595); 2 (AAK30596); 4 (AAK30593); 5 (AAR83925)], *Argiope amoena* [2 (AAR13813)],

*Argiope argentata* [3 (AAY28932)], *Latrodectus hasselti* [3(AAY28941)], *Latrodectus hesperus* [1(AAY28935); 2(ABD66603); 3 (AAY28931)], *Latrodectus geometricus* [1 (AAK30602), 2 (AAK30604), 3 (AAY28940)], *Latrodectus mactans* [3 (AAY28938)], *Cyrtophora moluccensis* [3 (AAY28944)], *Psechrus sinensis* [1 (AAV48939)], *Octonoba varians* [1 (AAV48931)], *Uloborus diversus* [ 1 (ABD61596), 2 (ABD61599), 3 (AAY28933), 5 (ABD61598), 6 (ABD61597)], *Araneus diadematus* [1 (AAC47008), 2 (AAC47009), 3 (AAC47010), 4 (AAC47011)], *Araneus ventricosus* [1 (AAN85280), 2 (AAN85281), 4 (ABK00016)], *Araneus bicentenarius* [2 (AAC04503)], *Araneus gemmoides* [3 (AAX45294)], *Gea heptagon* [3(AAY28943)], *Gasteracantha mammosa* [2 (AAK30601)], *Agelenopsis aperta* [ 1 (AAT08436)], *Nephila clavipes* [1 (AAT75312), 2 (AAT75315), 3 (AAX45295), 4 (AAF36089), 6 (AAC14589)], *Nephila clavata* [3 (BAE54450)], *Nephila antipodiana* [1 (ABC72644), 3 (AAY90151), 6 (ABC72645)], *Nephila madagascarensis* [1 (AAK30606), 2 (AAK30607), 4 (AAF36092)], *Nephila pilipes* [1 (AAV48946)], *Nephila senegalensis* [1 (AAK30608), 2 (AAK30609)], *Nephilengys cruentata* [ 1 (EF638446), 3 (EF638445), 4 (EF638444), 6 (EF638447)], *Euprosthops* sp. [1 (S2)], *Tetragnatha versicolor* [1 (AAK30615)], *Tetragnatha kauaiensis* [1 (AAK30614)], from the Mygalomorphae: *Macrothele holsti* [1 (AAV48940)], and *Euagrus chioseus* [1 (AAK30600)].

Spidroin C-terminals were translated into their respective amino acid sequences, and aligned using the accuracy oriented G-Ins-i algorithm, which utilizes global pairwise alignment information, as implemented in MAFFT 5.861 [gap opening penalty = 3] (S3). Sequences were then back-translated to their respective nucleotide alignments (Biopython

script, 366 bp), and subjected to phylogenetic analysis by Maximum Parsimony (MP), Maximum Likelihood (ML) [PAUP\* v4.0b10, S4], and Bayesian Markov Chain Monte Carlo (MrBayes v.3.1.2, S5) approaches. . Heuristic MP searches were conducted with TBR branch swapping and 10.000 random sequence additions (RSA), [gaps = 5<sup>th</sup> state, 25 trees, length: 4074]. ML analysis (presented, Fig. 1A) was performed using the best-fit model of evolution as estimated by Modeltest v3.7 (AIC Criterion, GTR+I+ $\Gamma$ , S6), combined with 100 replicates of RSA. Bootstrap support (BP) was evaluated by 1000 pseudoreplicates /100 RSA, and 100 pseudoreplicates/1 RSA for MP, and ML, respectively. Bayesian analysis included three runs at  $3 \times 10^6$  generations, implementing GTR+I+ $\Gamma$ , each with four chains (temperature: 0.15), and random starting trees; sampling frequency was 1000. Clade posterior probabilities ( $pP$ ) in the 95 percentile (after burn-in) were taken as supportive of a topological relationship. Consistent with previous work (1, 3) trees were rooted to mygalomorphs. Analyses reached similar overall topologies, but differed slightly in the position of individual terminals. The MP strict consensus collapses all basal nodes, and ML-BP and  $pP$  indicate no support.

The relative contribution of gene duplications, and their topological placement versus speciation events was tested in NOTUNG v.2.1 (S7) by reconciling gene trees with species trees (species tree according to (5), [edge weights = ML-BP, duplication = 2.0, loss cost = 0.5]). Additionally, different roots were tested by optimizing the number of duplications (D) and losses (L), and lowest bounds of duplications were evaluated. Presence [1] and absence [0], (ambiguous = missing) of MaSps was traced in MacClade 4.08 under ACCTAN and DELTRAN on the species tree (Fig. 1B, S8). Under all

possible most parsimonious root scenarios (= same # of Ds [23] and Ls [62]) the lower bound for MaSps was estimated to be the araneomorph-mygalomorph split, which recently was estimated at ca. 340-390 MYA (5). Character tracing clearly maps MaSps at the base of the spider tree.

### References:

- S1. J. Gatesy, C. Hayashi, D. Motriuk, J. Woods, R. Lewis, *Science* **291**, 2603 (2001).
- S2. N.N. Pouchkina-Stantcheva, S.J. McQueen-Mason, *CPB* **138**, 371–376 (2004).
- S3. K. Katoh, K. Kuma, H. Toh and T. Miyata, *Nucleic Acids Res.* **33**, 511-518 (2005).
- S4. D.L. Swofford, PAUP\* version 4.0b10. Sinauer Associates, Sunderland, Mass.  
(2002).
- S5. J.P. Huelsenbeck, F. Ronquist, *Bioinformatics* **17**, 754 (2001).
- S6. D. Posada D, T.R. Buckley. *Systematic Biology* **53**, 793-808 (2004).
- S7. D. Durand, B. V. Halldorsson, B. Vernet. *Journal of Computational Biology* **13**, 320-335 (2006).
- S8. D.R. Maddison, W.P. Maddison. Sinauer Associates, Inc. Mac Clade Version 4.08  
(2006)

## Supplementary Figure Legends

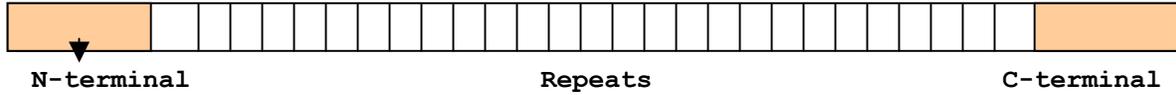
**Figure S1.** Spidroins expressed by *Avicularia juruensis*. (A) General molecular structure from spider silks proteins. (B) Ensemble repeats from Spidroin 1 and Spidroins 2. The amino acid motifs found in *A. juruensis* spidroins are represented by: green-A<sub>n</sub>, red-GA and blue-GS.

**Figure S2.** ClustalW alignment of C-terminal amino acid sequences from Spidroin 2 from *A. juruensis* (*A.jur*) and MaSp2 from *Argiope amoema* (*A.amo*), highlighted in green and yellow respectively (86% identity). Amino acids are indicated by one letter abbreviations and numbered from N- to C-terminal. 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.

## Figures

S1

A.



B.

### Spidroin 1

```
YSLASSIASAASSSSASSAAAASSSSAAAAGAAAASEAAASAAATSTTTTTSTSRAAAASAAAAASASGAAG  
AAGAASAASAASASSSLQQLGSALAQSSSFAAAFQASSAASAAAIAYALAQTVANQIGFSSYSSAFARAA  
SSAVYSIGGLASASAYAFAFASAFSQVLSNYGLLNINNA
```

### Spidroin 2

```
GAG (A/S) GSGSGSGS
```

S2

```
A. jur ARLSSPQASSRVSSAFFSLVSSGPTSPGALSNAISSVVSQVSASNPGLSGCDVLVQALLE 60  
A. amo SRLSSPQASSRVSSAVSSILVSSGPTNPAALSNAMSSVVSQVSASNPGLSGCDVLVQALLE 60  
:*****. *****. *. *****:*****
```

```
A. jur IVS-----QYASLVGQSVNQLRY 79  
A. amo IVSALVHILGSSSIGQINYAASSQYQMVGQSVQAALA- 98  
*** ***. ***** **
```